



## 基于AI的多组学分析方法鉴别新型冠状病毒感染和社区获得性肺炎的价值

王晓兰, 赵建华

Value of AI-based Multiomics Analysis in Differentiating COVID-19 from Community-acquired Pneumonia

WANG Xiaolan and ZHAO Jianhua

在线阅读 View online: <https://doi.org/10.15953/j.ctta.2023.049>

### 您可能感兴趣的其他文章

#### Articles you may be interested in

#### 新型冠状病毒肺炎胸部CT影像学特征及检查优选策略初探

Novel Coronavirus Pneumonia Chest CT Features and Examination Strategy

CT理论与应用研究. 2020, 29(2): 163–168

#### 新型冠状病毒肺炎临床表现及CT影像学分析

Clinical Manifestations and CT Imaging Analysis of Corona Virus Disease 2019

CT理论与应用研究. 2021, 30(4): 525–532

#### COVID-19各临床分型对应CT表现及人工智能辅助应用价值

The Clinical Subtypes of Corona Virus Disease 2019 Correspond to CT Findings and the Value of Artificial Intelligence

CT理论与应用研究. 2020, 29(5): 534–542

#### 人工智能CT定量分析预测并评估COVID-19肺炎临床分型的研究

Study on Predicting and Evaluating Clinical Classification of COVID-19 Pneumonia by Artificial Intelligence CT Quantitative Analysis

CT理论与应用研究. 2021, 30(6): 743–751

#### 64层螺旋CT螺距的不同设置在新型冠状病毒感染肺炎中筛查应用探讨

Application of Different Settings of 64-slice Spiral CT in Screening of Novel Coronavirusinfected Pneumonia

CT理论与应用研究. 2020, 29(5): 543–550

#### 2019新型冠状病毒肺炎的临床与CT表现

Clinical and CT Manifestations of Corona Virus Disease 2019

CT理论与应用研究. 2020, 29(3): 281–287



关注微信公众号，获得更多资讯信息

王晓兰, 赵建华. 基于 AI 的多组学分析方法鉴别新型冠状病毒感染和社区获得性肺炎的价值[J]. CT 理论与应用研究, 2023, 32(3): 357-366. DOI:10.15953/j.ctta.2023.049.

WANG X L, ZHAO J H. Value of AI-based Multiomics Analysis in Differentiating COVID-19 from Community-acquired Pneumonia[J]. CT Theory and Applications, 2023, 32(3): 357-366. DOI:10.15953/j.ctta.2023.049. (in Chinese).

# 基于 AI 的多组学分析方法鉴别新型冠状病毒感染和社区获得性肺炎的价值

王晓兰<sup>1</sup>, 赵建华<sup>2,3</sup>

1. 内蒙古科技大学包头医学院研究生院, 内蒙古 包头 014040
2. 内蒙古自治区人民医院影像医学科, 呼和浩特 010017
3. 内蒙古医科大学内蒙古临床医学院/包头医学院内蒙古临床医学院, 呼和浩特 010017

**摘要:** 目的: 旨在评估基于影像组学特征和常规临床信息 (包括临床症状及临床检验数据) 的多组学模型在区分新型冠状病毒感染 (COVID-19) 和社区获得性肺炎 (CAP) 方面的分类性能。方法: 收集奥密克戎 (Omicron) 变异株引起的 COVID-19 确诊患者和其他病毒感染引起的 CAP 确诊患者临床及胸部 CT 影像资料, 基于数据集构建影像组学模型、临床特征模型、多组学模型, 通过受试者工作特性曲线 (ROC) 分析评估每个模型的性能。结果: 选择 8 个影像组学特征和 7 个临床特征来构建影像组学模型、临床特征模型、多组学模型。在测试集中, 影像组学模型受试者工作特征曲线下面积 (AUC) 为 0.759, 临床特征模型 AUC 为 0.853, 多组学模型 AUC 为 0.9。结论: 基于 AI 的多组学分析方法构建的多组学模型分类性能高于影像组学模型和临床特征模型, 对 COVID-19 和 CAP 的鉴别诊断具有可行性。

**关键词:** 人工智能; 新型冠状病毒感染; 社区获得性肺炎; 影像组学; 检验组学

DOI:10.15953/j.ctta.2023.049 中图分类号: O242; R814; R563.1 文献标识码: A

2019 年新型冠状病毒感染 (coronavirus disease 2019, COVID-19) 是由严重急性呼吸系统综合征冠状病毒 2 (severe acute respiratory syndrome coronavirus 2, SARS-CoV-2) 引起的病毒性传染病。随着时间的推移 SARS-CoV-2 不断变异、进化, 包括阿尔法变异株、贝塔变异株、伽马变异株、德尔塔变异株、奥密克戎 (Omicron) 变异株 5 种重点关注变异株, 以及艾普西龙变异株、泽塔变异株等 8 种一般关注变异株。

影像组学可以从 COVID-19 患者和社区获得性肺炎 (community-acquired pneumonia, CAP) 患者胸部 CT 图像中提取和分析大量具有高通量的定量成像特征, 临床实验室检测数据在 COVID-19 与 CAP 的诊断中起着至关重要的作用。Ferrari 等<sup>[1]</sup>展示了简单的血液检测如何帮助识别假阳性/假阴性 RT-PCR 检测; 2022 年, Wen 等<sup>[2]</sup>提出了检验组学 (clinlabomics) 的新概念, 将临床检验医学与 AI 相结合。在本项研究中, 我们旨在开发和评估一种基于影像组学特征、临床症状和临床检验的多组学模型在区分 Omicron 变异株引起的 COVID-19 和其他病毒感染引起的 CAP 的价值。

## 1 资料和方法

### 1.1 研究对象

回顾性收集内蒙古自治区人民医院 Omicron 变异株感染引起的 COVID-19 确诊患者以及其他病毒感染引起的 CAP 确诊患者的临床及影像资料。纳入标准: ① 图像质量清晰, 满足图像分析要求; ② 鼻咽或口咽拭子 RT-PCR 检测确诊为 COVID-19; ③ 免疫、支气管肺泡灌洗、NGS 检测确诊为 CAP。

收稿日期: 2023-03-13。

基金项目: 内蒙古自治区人民医院院内基金 (基于深度学习的病毒性肺炎不同临床转归胸部 CT 评价 (2020YN08)); 包头医学院研究生教育教学改革项目 (人工智能在放射影像学专业学位研究生教学中的初步应用 (B-YJSJG202303))。

排除标准：① CT 检查结果阴性；② CT 影像表现为直径  $< 5\text{mm}$  磨玻璃结节；③ 图像模糊、呼吸运动伪影重，不能满足图像处理需求。

将患者分为 COVID-19 组（76 例，包括普通型和重型）和 CAP 组（50 例，包括单一病毒感染及病毒 + 病毒、病毒 + 细菌混合感染）。所有 COVID-19 病例于 2022 年 12 月 20 日至 2023 年 2 月 1 日期间获得。CAP 病例于 2018 年 11 月 1 日至 2022 年 6 月 1 日期间获得。

## 1.2 CT 检查方法

胸部 CT 图像使用 GE LightSpeed VCT 及宝石 CT、西门子 SOMATOM Definition Flash 双源 CT、东芝 Aquilion ONE 320 排 CT 四种不同扫描仪获取。

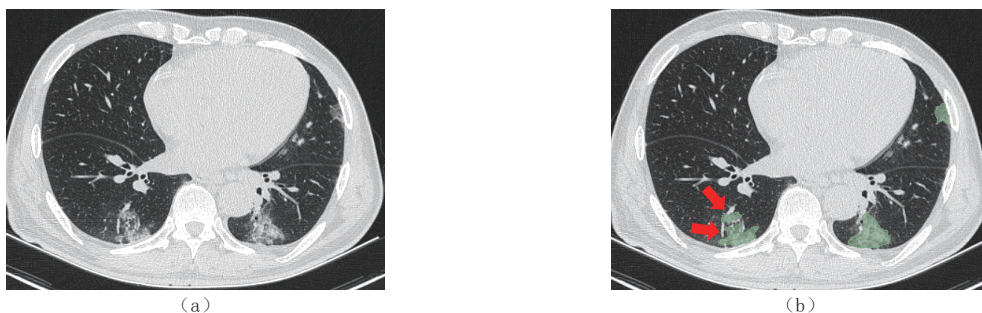
扫描参数：① GE LightSpeed VCT 及宝石 CT：管电压 120kV，自动管电流，重建层厚 1.25mm；② SOMATOM Definition Flash 双源 CT：管电压 120kV，自动管电流，重建层厚 1mm；③ Aquilion ONE 320 排 CT：管电压 120kV，自动管电流，重建层厚 0.8mm。扫描矩阵  $512 \times 512$ 。

所有患者经指导呼吸训练后，在吸气结束屏住呼吸时以仰卧位进行扫描。扫描范围从胸廓入口到横膈。

## 1.3 影像组学分析

图像预处理：影响特征的因素主要包括图像的预处理，目的是实现图像的同质化，减少图像质量本身引起的偏倚，提高特征识别率。首先对所有图像按照体素大小为  $1\text{mm} \times 1\text{mm} \times 1\text{mm}$  进行重采样，将图像转换为各向同性体素间距，以减少不同扫描仪和扫描参数的影响。IBSI（图像生物标志物标准化倡议，imaging biomarker standardization initiative）建议提取特征时特别是提取纹理特征的时候要保持体素各向同性。然后将 Bin Width 设置为 25（pyradiomics 官方建议值）进行灰度离散化，经过离散化处理后得到的计算机能够识别、处理的点阵图像，是一个用数字阵列来表示的图像。预处理过程由物理师编写代码完成。

肺炎病变三维分割：首先将从 PACS 中提取的 DICOM 格式图匿名化并转换为 nifty 格式，然后将图像导入到开源软件 ITK-snap 4.0 版本；随后由 1 名初级医师逐层手动绘制三维感兴趣区域（region of interest, ROI），勾画好的 ROI 由 1 位具有丰富工作经验的高年资影像科医生进行审核，分割不理想的情况下由两位放射科医生讨论后达成共识；最终采纳经高年资医师改进后的 ROI。将其保持在尽量贴合病灶的边缘，并从分析中排除相邻结构，例如血管或支气管分支，空腔或正常肺实质（图 1），从而更加准确的分割病灶。方法是包含轴位胸部 CT 图像上显示的所有肺炎病灶，CT 值固定到 1000 到 -1000 之间以显示更好对比度。为了使分割结果更为准确，我们同时分割了骨组织、软组织、肺动脉以及表现为正常的肺组织，只保留最后的病灶部分，以此达到病灶的完整分割；为更全面地提取特征，避免选择非病灶区信息，ROI 勾画时排除每例患者直径  $< 5\text{mm}$  的微小结节灶。



注：（a）和（b）分别为感兴趣区域勾画前后的胸部 CT 图像，红箭示增粗的血管。

图 1 COVID-19 患者的胸部 CT 图像感兴趣区域勾画示意图

Fig.1 Region of interest in the chest CT image of a patient with COVID-19

影像组学特征提取：使用 python 平台下的 Pyradiomic 开源工具包提取标准化特征，提取特征符合 IBSI 标准；总共提取 1288 个特征，主要的特征类别：（1）252 个一阶特征（first-order features）。（2）14 个形状特征（shape features）。（3）1022 个纹理特征（texture features），包括 ① 灰度依赖矩阵（gray-level dependence matrix, GLDM）特征；② 灰度共生矩阵（gray-level cooccurrence matrix, GLCM）特征；③ 灰度大小区域矩阵（gray-level size zone matrix, GLSZM）特征；④ 相邻灰度差矩阵（neighboring gray-tone difference matrix, NGTDM）特征；⑤ 灰度游程矩阵（gray-level run-length matrix, GLRLM）特征。提取了原始图像的一阶特征和纹理特征，使用了 2 个滤波器：小波滤波器、拉普拉斯高斯滤波器。

特征选择和降维：在特征选择和降维过程中，针对训练集，使用 Scikit-learn 工具包中基于 10 折交叉验证的最小绝对收缩和选择算子方法（least absolute shrinkage and selection operator, LASSO）回归选择最相关的组学特征。从 CT 图像中选出系数不为 0 的特征。其次为了使图像中冗余特征去除，我们采用相关性分析（pearson 相关系数，当两列特征的 pearson 相关系数大于 0.9 时，即判断一列特征可以由另一列来代替，去掉其中的一个，达到去冗余的效果）剔除与其余特征具有相关性系数大于 0.9 的冗余特征。

#### 1.4 临床数据

回顾性收集 COVID-19 患者和 CAP 患者相关人口统计学、临床症状和常规实验室检测数据，包括年龄、性别、和发热、乏力、呼吸困难等 7 种临床症状以及血常规、C 反应蛋白、生化全项、血气分析、降钙素原、D2 聚体等 35 个实验室检测指标。临床特征筛选使用 R 软件 4.0 版本对临床相关特征进行单因素逻辑回归和多因素逻辑回归筛选最终变量。

#### 1.5 模型构建与评估

本研究采用的机器学习方法——随机森林（random forest, RF），RF 是一种决策树集成方法，使用称为 bagging（bootstrap aggregation）的重采样过程构建许多树，每棵树的节点都使用随机选择的树特征子集进行划分。当引入更多的树时，RF 不会过度拟合，甚至可以提供泛化误差的极限值。RF 在处理多维和非线性数据、高效并行处理的机会以及抗噪声能力方面比其他机器学习算法具有重要优势<sup>[3]</sup>。在训练集参数训练时，我们使用十折交叉验证，以避免模型的随机训练参数较高，同时获得模型平均接收者操作特征曲线下面积（area under curve, AUC）及接收者操作特征曲线（receiver operating characteristic curve, ROC）。有利于对模型的泛化能力进行更好解释，同时也从另一方面进一步避免模型出现过拟合情况。

#### 1.6 统计学分析

所有统计分析使用 Python 软件和 R 软件执行。连续变量表示为（平均值 ± 标准差）或中位数，通过独立样本 *t* 检验或曼-惠特尼 U 检验进行分析。分类变量以数字表示，使用  $\chi^2$  检验或 Fisher 精确检验进行分析。单因素逻辑回归和多因素逻辑回归分析用于分析因素与诊断之间的相关性。计算 AUC 以评估分类性能。

## 2 结果

### 2.1 人口统计学

COVID-19 队列和 CAP 队列中基于选定特征的患者临床基线资料比较见表 1。最终纳入 COVID-19 确诊病例 55 例（平均年龄（68.3 ± 10.5）岁），CAP 确诊病例 33 例（平均年龄（54.0 ± 17.1）岁）。 $P < 0.05$  为统计学具有显著差异。

### 2.2 临床特征

表 2 所示临床特征通过单因素逻辑回归和多因素逻辑回归进行筛选，最终 7 个临床特征纳入 RF



模型, 包括年龄 (age), 尿酸 (uric acid, URIC), 血氧分压 (partial pressure of oxygen, PaO<sub>2</sub>), 嗜酸性粒细胞 (eosinophil, EOS), 嗜碱性粒细胞 (basophil, BASO), 平均红细胞血红蛋白浓度 (mean corpuscular hemoglobin concentration, MCHC), 血小板分布宽度 (platelet distributionwidth, PDW)。

表1 人口统计学特征  
Table 1 Demographic characteristics

特征		CAP (n=33)	COVID-19 (n=55)	Overall (n=88)	P
年龄/岁	Mean (SD)	54.0 (17.1)	68.3 (10.5)	62.9 (15.0)	< 0.001
	Median (Min, Max)	57.0 (24.0, 86.0)	70.0 (35.0, 88.0)	66.5 (24.0, 88.0)	
UIRC/(μmol/L)	Mean (SD)	218 (102)	294 (168)	265 (151)	0.073
	Median (Min, Max)	240 (0, 429)	240 (0, 978)	240 (0, 978)	
PaO <sub>2</sub> /mmHg	Mean (SD)	45.5 (39.2)	70.4 (37.1)	61.1 (39.6)	0.016
	Median (Min, Max)	62.0 (0, 111)	71.7 (0, 169)	69.8 (0, 169)	
EOS/(10 <sup>9</sup> /L)	Mean (SD)	0.065 (0.058)	0.038 (0.050)	0.048 (0.054)	0.084
	Median (Min, Max)	0.050 (0, 0.240)	0.010 (0, 0.180)	0.030 (0, 0.240)	
BASO/(10 <sup>9</sup> /L)	Mean (SD)	0.291 (0.315)	0.202 (0.168)	0.235 (0.236)	0.230
	Median (Min, Max)	0.200 (0, 1.400)	0.200 (0, 0.600)	0.210 (0, 1.400)	
MCHC/(g/L)	Mean (SD)	329 (44.4)	343 (9.32)	338 (28.8)	0.083
	Median (Min, Max)	340 (90.0, 356)	344 (318, 365)	343 (90.0, 365)	
PDW/%	Mean (SD)	15.6 (2.85)	16.4 (0.350)	16.1 (1.80)	0.104
	Median (Min, Max)	16.1 (0, 17.0)	16.4 (15.7, 17.5)	16.4 (0, 17.5)	

表2 患者临床特征  
Table 2 Patients clinical characteristics

特征	B. x	SE. x	OR. x	CI. x	Z. x	P. x	B. y	SE. y	OR. y	CI. y	Z. y	P. y
年龄/岁	0.074	0.019	1.08	1.04~1.12	3.866	0.00	0.074	0.019	1.08	1.04~1.12	3.866	0.00
BASO/(10 <sup>9</sup> /L)	-30.983	15.791	0.00	0~0.97	-1.962	0.05	-30.983	15.791	0.00	0~0.97	-1.962	0.05
EOS/(10 <sup>9</sup> /L)	-9.056	4.216	0.00	0~0.45	-2.148	0.032	-9.056	4.216	0.00	0~0.45	-2.148	0.032
MCHC/(g/L)	0.063	0.024	1.07	1.02~1.12	2.655	0.008	0.063	0.024	1.07	1.02~1.12	2.655	0.008
PaO <sub>2</sub> /mmHg	0.017	0.006	1.02	1.01~1.03	2.765	0.006	0.017	0.006	1.02	1.01~1.03	2.765	0.006
PDW/%	2.188	0.703	8.92	2.25~35.38	3.113	0.002	2.188	0.703	8.92	2.25~35.38	3.113	0.002
URIC/(μmol/L)	0.004	0.002	1.00	1~1.01	2.199	0.028	0.004	0.002	1.00	1~1.01	2.199	0.028

注: x 为单因素逻辑回归, y 为多因素逻辑回归, B 为 β 值, SE 为标准误, OR 为风险比, CI 为 OR 的 95% 置信区间, Z 为统计量, P 为概率。

### 2.3 肺炎分类

通过特征选择和降维最终选择了 8 个影像组学特征和 7 个临床特征分别构建了影像组学模型、临床特征模型、多组学模型, 通过 ROC 曲线分析评估 3 个模型的性能, 计算 AUC、准确率、灵敏度、特异度和 F1 score。所有模型中, 多组学模型 AUC 最高, 其次是临床特征模型, 最后是基于 CT 的影像组学模型。表 3 显示了 3 个模型在训练集和测试集中的分类性能。

图 2~图 4 显示 3 个模型基于 RF 算法预测 COVID-19 患者时, 选定的特征重要性在每个叶子节点的分布, 每个特征在每个节点都为个体预测提供不同的权重重要性。图 5 显示影像组学模型、临床特征模型、多组学模型分别在训练集和测试集中的 ROC 曲线图。基于单因素逻辑回归筛选具有统计学意义的临床独立危险因素和 Rad-score 构建 nomogram 图 (图 6)。

## 3 讨论

COVID-19 典型胸部 CT 征象可能有助于疑似病例的早期筛查, 但仅通过影像科医生视觉诊断高

表 3 影像组学模型、临床特征模型、多组学模型的性能  
Table 3 The performances of radiomics, clinical features, and multiomics models

性能	影像组学模型		临床特征模型		多组学模型	
	训练集	测试集	训练集	测试集	训练集	测试集
AUC	0.82	0.759	0.87	0.853	0.94	0.900
准确率	0.75	0.778	0.79	0.815	0.74	0.778
灵敏度	0.88	0.882	0.90	0.882	0.84	0.824
特异性	0.52	0.600	0.62	0.700	0.55	0.700
F1 score	0.75	0.833	0.79	0.857	0.74	0.833

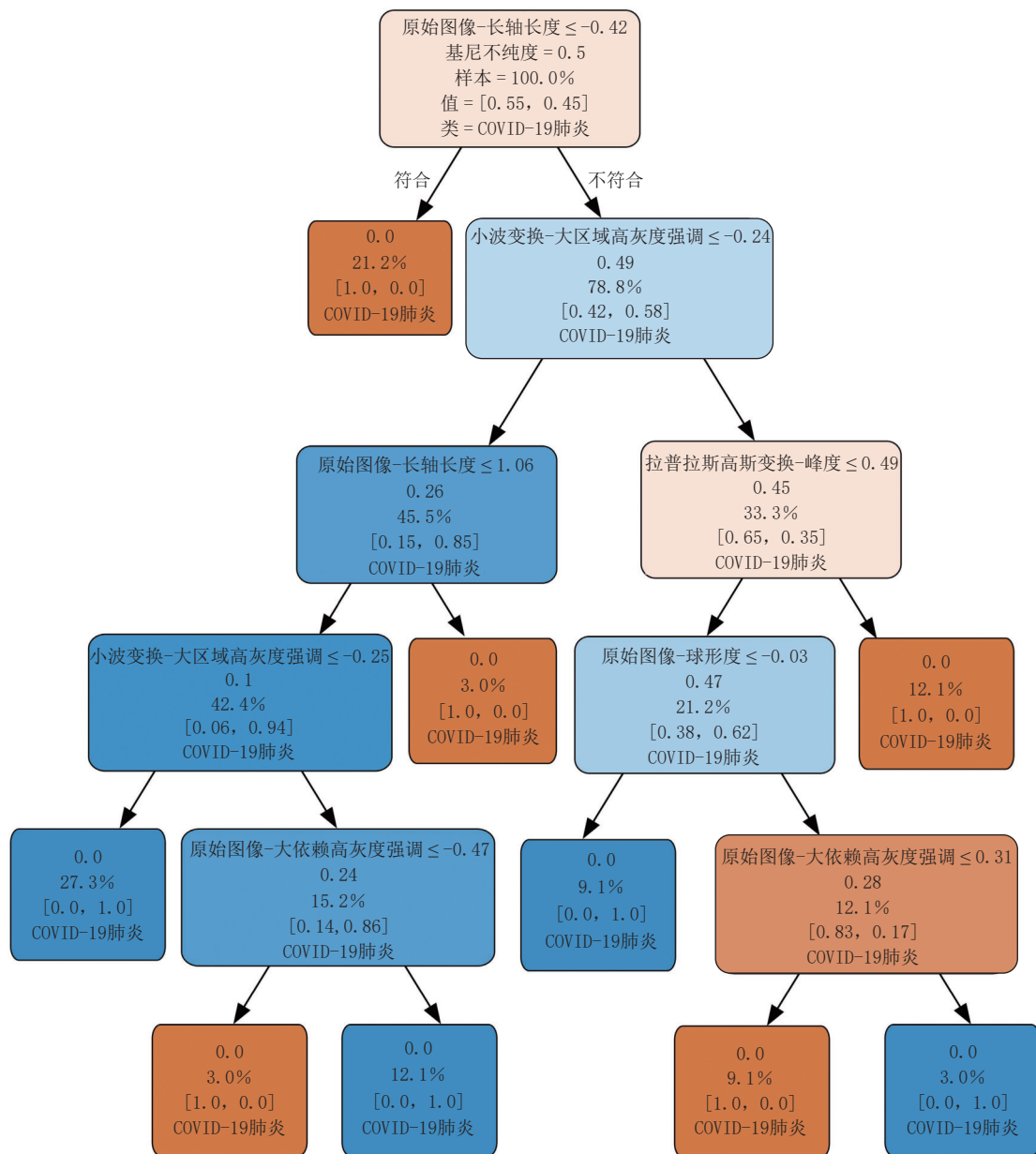


图 2 基于影像组学特征的 RF 模型预测 COVID-19 流程图

Fig.2 Flow chart for COVID-19 prediction by RF model based on radiomics features

度相似征象的胸部 CT 图像误诊是不可避免的。Li 等<sup>[4]</sup>研究结果显示使用 CT 诊断病毒性肺炎漏诊率较低，但对于识别特定病毒和区分病毒仍然有限。既往关于影像科医生在胸部 CT 图像中鉴别 COVID-

19 和其他肺炎的研究显示，灵敏度（70%~94%）和特异性（24%~100%）均存在很大差异，具有高灵敏度和中等特异性<sup>[5]</sup>，这种 CT 结果解释的可变性给临床医生的鉴别诊断带来了常规挑战。我们的研究表明，与仅使用影像科医生视觉诊断的方法相比，AI 可增强影像科医生在区分 COVID-19 和 CAP 方面的表现，产生了较高的准确率、灵敏度和特异性指标。

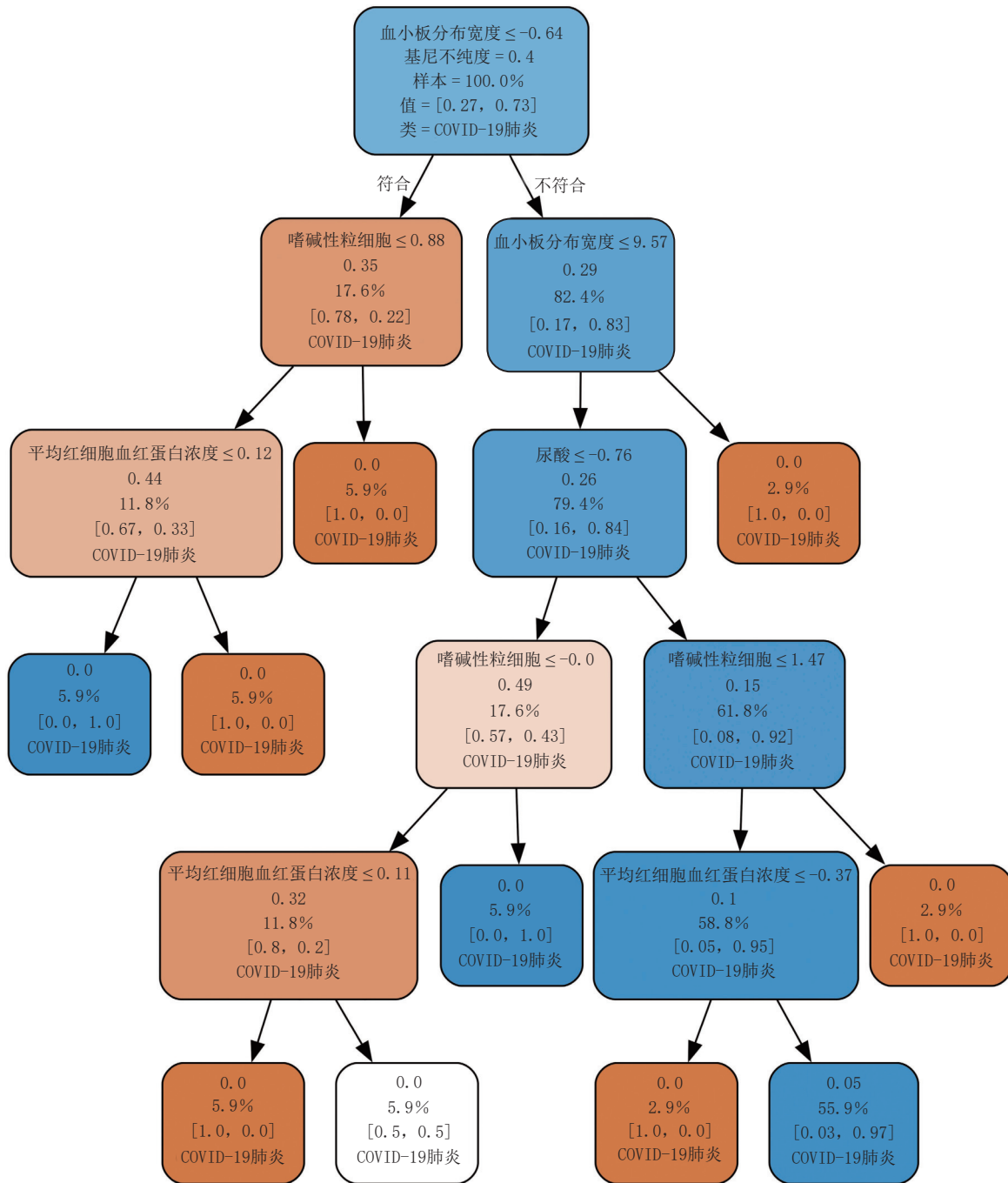


图 3 基于临床特征的 RF 模型预测 COVID-19 流程图

Fig.3 Flow chart for clinical features-based RF model COVID-19 prediction

在我们的研究中，发现年龄、URIC、PaO<sub>2</sub>、EOS、BASO、MCHC、PDW 是区分 COVID-19 和 CAP 的重要指标。Dai 等<sup>[6]</sup>研究显示 URIC, MCHC 等是区分 COVID-19 和 CAP 的重要指标。临床模型预测 COVID-19 时，我们发现 PDW 是个体预测的关键特征，包括 BASO 在内的特征在每个节点都为个体预测提供了不

同的权重重要性。有研究<sup>[7]</sup>结果证明当常规外周血检查的PDW 值为 $\leq 2.12$  fL 时向医生提示 SARS-CoV-2 感染。EOS 和 BASO 计数减少在 COVID-19 中较为常见。本研究 COVID-19 组 EOS 及 BASO 水平较低, 平均值均低于 CAP 组。EOS 具有潜在的抗病毒活性, 有研究<sup>[8]</sup>显示外周血 EOS 计数可能是 COVID-19 患者诊断、评估和预后监测的有效指标, 较低的 EOS 水平也与患者预后和死亡率有关<sup>[9]</sup>。本研究 COVID-19 患者队列和 CAP 患者队列平均 PaO<sub>2</sub> 水平低于正常值, 与机体感染肺炎导致肺泡和血管损伤有关<sup>[10-11]</sup>。滕晓蕾等<sup>[12]</sup>研究显示重症 COVID-19 患者的氧合功能较重症 CAP 患者更差。本研究 COVID-19 组平均年龄高于 CAP 组, 这可能与老年人感染 COVID-19 相对易感性较高有关<sup>[13]</sup>。

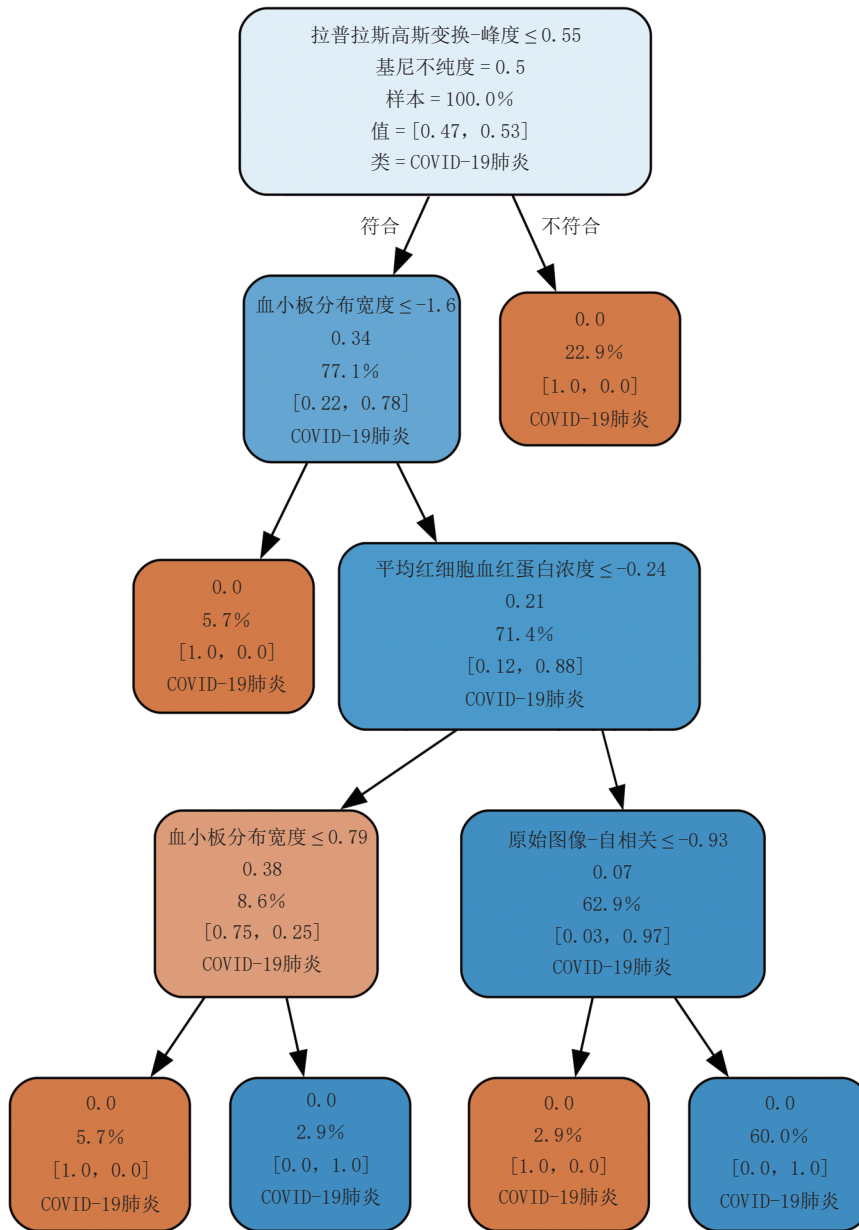


图 4 多组学模型预测 COVID-19 流程图

Fig.4 Flow chart for multiomics model COVID-19 prediction

COVID-19 与其他病毒感染 (H1N1、H5N1、流感、呼吸道合胞病毒、鼻病毒、腺病毒等) 在影像学检查结果上存在重叠<sup>[14-15]</sup>。影像组学是通过转换数字医学图像来生成可挖掘的高通量数据<sup>[16]</sup>。在既往研究中, 影像组学可以更深入地了解肺炎病变的异质性<sup>[17-19]</sup>。因此, 影像组学在理论上是区分 COVID-19 与其他病毒引起的 CAP 的可行方法。田斌等<sup>[20]</sup>基于 14 种机器学习算法构建影像组学模型分



析 COVID-19 和 CAP 的鉴别诊断效能，其中 RF 模型诊断效能最高。本研究使用 LASSO 选择 8 个最具预测性的影像组学特征构建基于 RF 算法的影像组学模型，其中部分是经小波滤波器和拉普拉斯高斯滤波器过滤的一阶或纹理特征，并且在多组学模型中，基于拉普拉斯高斯变换的峰度特征是预测 COVID-19 的关键特征。这可能表明，两种高度成像重叠肺炎之间的区别可能需要空间或频域中强调的特征或具有相对更高稳定性的高阶特征。

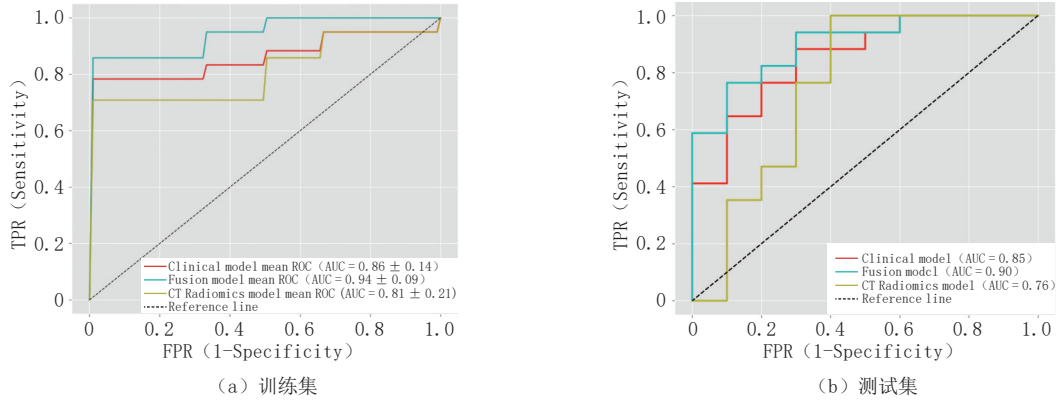


图 5 影像组学模型、临床特征模型、多组学模型的 ROC 曲线图

Fig.5 ROC curves of radiomics, clinical features, and multiomics models

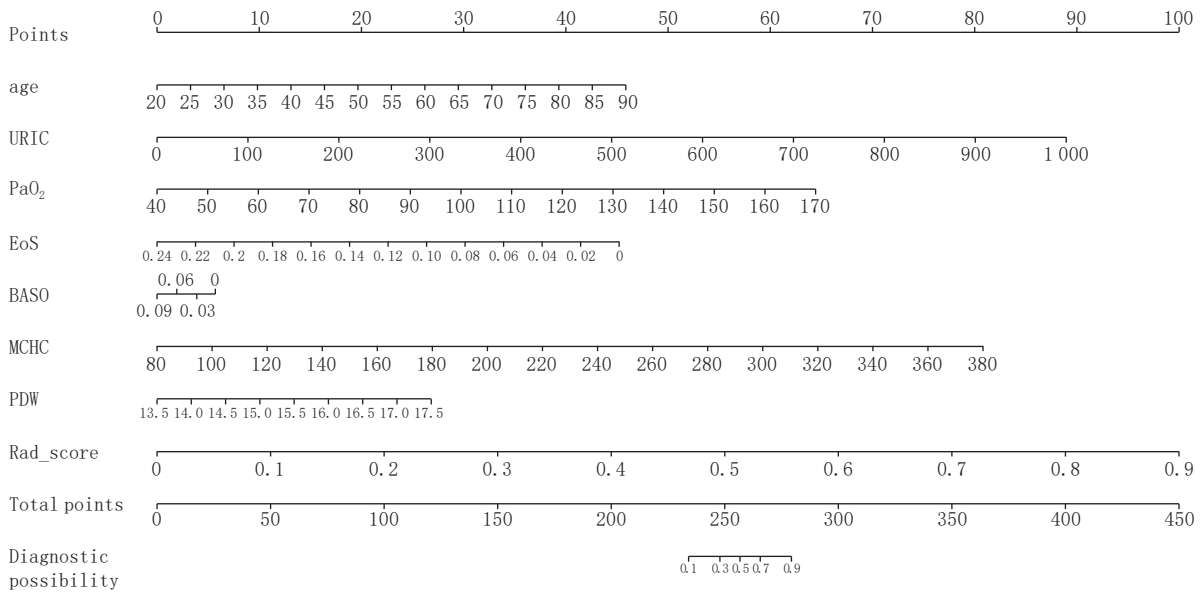


图 6 预测 COVID-19 的列线图

Fig.6 Nomogram for COVID-19 prediction

我们的研究表明，基于 RF 的多组学模型通过结合影像组学、临床症状、临床检验，在 COVID-19 和其他病毒感染引起的 CAP 的鉴别诊断方面取得了良好的效果，基于临床独立危险因素和 rad-score 构建 nomogram 图可以更好的表征 COVID-19 感染情况。这符合添加额外的临床信息可以显著提高影像组学性能的观点<sup>[21-22]</sup>。Shiri 等<sup>[21]</sup>研究显示，语义特征、临床特征和影像组学特征的结合可以有效预测 COVID-19 患者的生存率。同样，Chao 等<sup>[22]</sup>证明，影像学和非影像学数据的整合显著提高了 COVID-19 患者对 ICU 入院需求的预测性能。总而言之，整合多组学信息对 COVID-19 的诊断是有效的。

本项研究局限性：① 本研究是一项回顾性研究，未进行有助于了解模型在实践中表现的前瞻性验证。② 本研究中的患者队列相对较小，缺乏外部验证。③ 我们的 COVID-19 队列在症状发作、实

实验室指标和 CT 之间的时间分布上是异质性的, 早期 COVID-19 胸部 CT 的样本量有限。④最后, 所有患者都来自中国内蒙古自治区, 可能无法完全代表 COVID-19 Omicron 变异株表型谱系和 CAP 表型谱系。

## 4 结论

我们使用 RF 算法基于常规临床信息的多组学分析方法开发了一种新的 COVID-19 和 CAP 鉴别诊断模型。我们的研究表明, 基于 AI 的多组学分析方法可以有效识别 COVID-19 和其他病毒感染引起的 CAP。对模型结果的分析使我们能够识别一组具有明显高预测潜力的特征, 有助于改善护理决策和提高患者生存率。

## 参考文献

- [1] FERRARI D, MOTTA A, STROLLO M, et al. Routine blood tests as a potential diagnostic tool for COVID-19[J]. *Clinical Chemistry Laboratory Medicine*, 2020, 58(7): 1095-1099. DOI:10.1515/ccIm-2020-0398.
- [2] WEN X, LENG P, WANG J, et al. Clinlabomics: Leveraging clinical laboratory data by data mining strategies[J]. *BMC Bioinformatics*, 2022, 23(1): 387. DOI:10.1186/s12859-022-04926-1.
- [3] LEBEDEV A V, WESTMAN E, Van WESTEN G J, et al. Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness[J]. *Neuroimage Clinical*, 2014, 6: 115-125. DOI:10.1016/j.nicl.2014.08.023.
- [4] LI Y, XIA L. Coronavirus disease 2019 (COVID-19): Role of chest CT in diagnosis and management[J]. *American Journal of Roentgenol*, 2020, 214(6): 1280-1286. DOI:10.2214/AJR.20.22954.
- [5] BAI H X, HSIEH B, XIONG Z, et al. Performance of radiologists in differentiating COVID-19 from viral pneumonia on chest CT[J]. *Radiology*, 2020, 296: E46-E54. DOI:10.1148/radiol.2020200823.
- [6] DAI W, KE P F, LI Z Z, et al. Establishing classifiers with clinical laboratory indicators to distinguish COVID-19 from community-acquired pneumonia: Retrospective cohort study[J]. *Journal Medical Internet Research*, 2021, 23(2): e23390. DOI:10.2196/23390.
- [7] WANG Y, FAN Z, WANG S, et al. The diagnostic value of platelet distribution width in patients with mild COVID-19[J]. *Journal Clinical Laboratory Analysis*, 2021, 35(4): e23703. DOI:10.1002/jcla.23703.
- [8] XIE G, DING F, HAN L, et al. The role of peripheral blood eosinophil counts in COVID-19 patients[J]. *Allergy*, 2021, 76(2): 471-482. DOI:10.1111/all.14465.
- [9] TABACHNIKOVA A, CHEN S T. Roles for eosinophils and basophils in COVID-19?[J]. *Nature Reviews Immunology*, 2020, 20(8): 461. DOI:10.1038/s41577-020-0379-1.
- [10] PREDILETTO I, D'ANTONI L, CARONARA P, et al. Standardizing PaO<sub>2</sub> for PaCO<sub>2</sub> in P/F ratio predicts in-hospital mortality in acute respiratory failure due to COVID-19: A pilot prospective study[J]. *European Journal of Internal Medicine*, 2021, 92: 48-54. DOI:10.1016/j.ejim.2021.06.002.
- [11] TORRES A, CILLONIZ C, NIEDERMAN M S, et al. Pneumonia[J]. *Nature Reviews Disease Primers*, 2021, 7(1): 25. DOI:10.1038/s41572-021-00259-0.
- [12] 滕晓蕾, 谢云, 陈道南, 等. 重症新冠肺炎与重症社区获得性肺炎临床特征对比[J]. *中华危重病急救医学*, 2022, 34(5): 485-491. DOI:10.3760/cma.j.cn121430-20211013-01477.  
TENG X L, XIE Y, CHEN D N, et al. Comparative analysis of clinical features between severe coronavirus disease 2019 and severe community acquired pneumonia[J]. *Chinese Critical Care Medicine*, 2022, 34(5): 485-491. DOI:10.3760/cma.j.cn121430-20211013-01477. (in Chinese).
- [13] 杨莎, 李东泽, 郑艾宜, 等. 简易体征评分系统对新冠患者严重程度和预后预测价值的多中心观察性研究[J]. *中国胸心血管外科临床杂志*, 2023, 30(2): 167-172. DOI:10.7507/1007-4848.202203002.  
YANG S, LI D Z, ZHENG A Y, et al. Predictive value of the simplified signs scoring system for the severity and prognosis of patients with COVID-19: A multicenter observational study[J]. *Chinese Journal of Clinical Thoracic and Cardiovascular Surgery*, 2023, 30(2): 167-172. DOI:10.7507/1007-4848.202203002. (in Chinese).
- [14] SONG F, SHI N, SHAN F, et al. Emerging 2019 Novel Coronavirus (2019-nCoV) pneumonia[J]. *Radiology*, 2020, 295(1): 210-217. DOI:10.1148/radiol.2020200274.
- [15] KOO H J, LIM S, CHOE J, et al. Radiographic and CT features of viral pneumonia[J]. *Radiographics*, 2018, 38(3): 719-739. DOI:10.1148/rg.2018170048.
- [16] LAMBIN P, LEIJENAAR R T H, DEIST T M, et al. Radiomics: The bridge between medical imaging and personalized medicine[J]. *Nature Reviews Clinical Oncology*, 2017, 14(12): 749-762. DOI:10.1038/

- nrclinonc. 2017. 141.
- [17] ZHANG T, YUAN M, ZHONG Y, et al. Differentiation of focal organising pneumonia and peripheral adenocarcinoma in solid lung lesions using thin-section CT-based radiomics[J]. *Clinical Radiology*, 2019, 74(1): 78.e23-78.e30. DOI:10.1016/j.crad.2018.08.014.
- [18] WANG B, LI M, MA H, et al. Computed tomography-based predictive nomogram for differentiating primary progressive pulmonary tuberculosis from community-acquired pneumonia in children[J]. *BMC Medical Imaging*, 2019, 19(1): 63. DOI:10.1186/s12880-019-0355-z.
- [19] YANLING W, DUO G, ZUOJUN G, et al. Radiomics nomogram analyses for differentiating pneumonia and acute paraquat lung injury[J]. *Scientific Reports*, 2019, 9(1): 15029. DOI:10.1038/s41598-019-50886-7.
- [20] 田斌, 余晖, 任基刚, 等. 基于机器学习的多种分类模型在新型冠状病毒肺炎与社区获得性肺炎鉴别诊断中的效能[J]. *放射学实践*, 2021, 36(5): 590-595. DOI:10.13609/j.cnki.1000-0313.2021.05.004.
- TIAN B, YU H, REN J G, et al. Effectiveness of multiple classification models based on machine learning in the differential diagnosis of COVID-19 and community-acquired pneumonia[J]. *Radiologic Practice*, 2021, 36(5): 590-595. DOI:10.13609/j.cnki.1000-0313.2021.05.004. (in Chinese).
- [21] SHIRI I, SOROURI M, GERAMIFAR P, et al. Machine learning-based prognostic modeling using clinical data and quantitative radiomic features from chest CT images in COVID-19 patients[J]. *Computers in Biology Medicine*, 2021, 132: 104304. DOI:10.1016/j.combiomed.2021.104304.
- [22] CHAO H, FANG X, ZHANG J, et al. Integrative analysis for COVID-19 patient outcome prediction[J]. *Medical Image Analysis*, 2021, 67: 101844. DOI:10.1016/j.media.2020.101844.

## Value of AI-based Multiomics Analysis in Differentiating COVID-19 from Community-acquired Pneumonia

WANG Xiaolan<sup>1</sup>, ZHAO Jianhua<sup>2,3✉</sup>

1. Graduate School, Baotou Medical College, Inner Mongolia University of Science and Technology, Baotou 014040, China
2. Department of Medical Imaging, Inner Mongolia People's Hospital, Hohhot 010017, China
3. Inner Mongolia Medical University Inner Mongolia Clinical Medical College/Baotou Medical College Inner Mongolia Clinical Medical College, Hohhot 010017, China

**Abstract:** Objective: To assess the effectiveness of a multiomics model that combines radiomics characteristics and routine clinical information (including clinical symptoms and laboratory data) to distinguish between coronavirus disease 2019 (COVID-19) and community-acquired pneumonia (CAP). Methods: Retrospective data of patients with confirmed COVID-19 caused by the Omicron variant and patients with CAP caused by other viral infections were collected, including chest CT imaging and clinical data. Radiomics, clinical features, and multiomics models were constructed using the entire dataset, and the performance of each model in distinguishing between COVID-19 and CAP was evaluated using receiver operating characteristic curve (ROC) analysis. Results: A total of 8 radiomics features and 7 clinical features were selected to construct the radiomics, clinical features, and multiomics models. The area under the subject operating characteristic curve (AUC) of the radiomics model was 0.759, that of the clinical model was 0.853, and that of the multiomics model was 0.9. Conclusions: The study suggests that AI-based multiomics model has a better performance in differentiating between COVID-19 and CAP compared with those of the radiomics and clinical features models.

**Keywords:** artificial intelligence; coronavirus disease 2019; community-acquired pneumonia; radiomics; clinlabomics



**作者简介:** 王晓兰, 女, 内蒙古科技大学包头医学院在读研究生, 主要从事深度学习在病毒性肺炎胸部 CT 的应用研究, E-mail: [w15124857513@163.com](mailto:w15124857513@163.com); 赵建华<sup>✉</sup>, 男, 内蒙古自治区人民医院影像医学科副主任医师、硕士生导师, 主要从事深度学习在病毒性肺炎胸部 CT 的应用研究, E-mail: [zjh2822yyjh@163.com](mailto:zjh2822yyjh@163.com)。