



基于多种机器学习算法的人血浆辐射敏感脂质代谢物模型探索

陈琦, 赵骅, 蔡恬静, 高义哲, 高玲, 刘青杰

引用本文:

陈琦,赵骅,蔡恬静,高义哲,高玲,刘青杰. 基于多种机器学习算法的人血浆辐射敏感脂质代谢物模型探索[J]. 中华放射医学与防护杂志, 2024, 44(6): 457-463.

在线阅读 View online: <https://doi.org/10.3760/cma.j.cn112271-20231228-00225>

您可能感兴趣的其他文章

Articles you may be interested in

^{60}Co γ 射线全身照射大鼠的小肠组织中辐射敏感脂质代谢物的筛选

Screening of radiosensitive lipid metabolites in rat small intestine after total body irradiation with ^{60}Co γ - rays

中华放射医学与防护杂志. 2021, 41(3): 172-177 <https://doi.org/10.3760/cma.j.issn.0254-5098.2021.03.003>

全身照射后大鼠血浆代谢特征分析

Analysis of the metabolic characteristics in rat plasma after total body irradiation

中华放射医学与防护杂志. 2021, 41(6): 401-406 <https://doi.org/10.3760/cma.j.issn.0254-5098.2021.06.001>

全胸照射引起的大鼠血浆辐射损伤早期代谢特征研究

Early metabolic characterization in plasma of rats exposed to whole thorax irradiation

中华放射医学与防护杂志. 2020, 40(12): 919-925 <https://doi.org/10.3760/cma.j.issn.0254-5098.2020.12.005>

禁食对 ^{137}Cs γ 射线照射诱导小鼠肠道辐射损伤的代谢组学研究

Metabonomic study on the effect of fasting on ^{137}Cs γ -ray radiation-induced intestinal injury in mice

中华放射医学与防护杂志. 2021, 41(8): 564-570 <https://doi.org/10.3760/cma.j.issn.0254-5098.2021.08.002>

需行全身照射患者辐射损伤血浆代谢谱的研究

Preliminary study on plasma metabolites of total body irradiation patients

中华放射医学与防护杂志. 2017, 37(1): 7-11 <https://doi.org/10.3760/cma.j.issn.0254-5098.2017.01.002>

基于多种机器学习算法的人血浆辐射敏感脂质代谢物模型探索

陈琦^{1,2} 赵骅¹ 蔡恬静¹ 高义哲¹ 高玲¹ 刘青杰¹

¹中国疾病预防控制中心辐射防护与核安全医学所 辐射防护与核应急中国疾病预防控制中心重点实验室, 北京 100088; ²湖北省疾病预防控制中心传染病防治所急性传染病预防控制部, 武汉 430079

通信作者: 刘青杰, Email: liuqingjie@nirp.chinacdc.cn

【摘要】 目的 通过脂质组学方法联合机器学习 (ML) 多种算法, 探索人外周血辐射敏感脂质代谢物分类模型。**方法** 收集 2023 年 3—9 月北京市某综合医院准备接受骨髓移植的 25 例白血病患者放射治疗病例, 照射前和照射后外周血样本 97 份, 其中对照组 24 份, 为照射前血液样本; 辐射组 73 份, 为 4、8、12 Gy 照射剂量下的 24、25 和 24 份血液样本。采用基于超高效液相色谱—串联质谱 (UPLC-MS/MS) 平台的靶向脂质组学方法, 分析辐射组与对照组差异脂质。线性回归筛选 0~12 Gy 的剂量范围内辐射剂量响应脂质。采用 5 种机器学习方法构建训练集辐射分类模型, 验证集进行模型的验证和评价。**结果** 与对照组相比, 辐射组敏感的脂质代谢物中 62 个脂质浓度变化差异有统计学意义 ($t = -4.91 \sim 4.74$, $P < 0.05$), 包括鞘磷脂 (SM)、胆固醇酯 (CE)、神经酰胺 (Cer)、磷脂酰肌醇 (PI)、己糖神经酰胺 (HexCer)、溶血磷脂酰胆碱 (LysoPC)、醚磷脂酰胆碱 (PCO)、磷脂酰乙醇胺 (PE)、溶血磷脂酰乙醇胺 (LysoPE) 这 9 大类。在 0~12 Gy 的剂量范围内, 筛选出 20 种具有良好剂量反应的脂质代谢物, 包括 11 个 SM, 7 个 CE, 1 个 Cer 和 1 个 PI。决策树 (DT)、支持向量机 (SVM)、轻量梯度增强机 (Light GBM)、随机森林 (RF)、K 最邻近 (KNN) 这 5 种机器学习训练模型拟合均较好 ($F1 = 0.69 \sim 1.00$), 灵敏度较高。通过评价验证指标, 辐射分类判定效果最好的模型为随机森林 (灵敏度 1.00、准确率 0.72、 $F1$ 值 0.80)。**结论** 通过靶向脂质组学分析, 发现人类样本中辐射响应的脂质代谢物和辐射剂量响应的脂质。机器学习方法中的 RF 模型可以为探索人类辐射脂质代谢物模型提供新的思路。

【关键词】 电离辐射; 脂质组学; 机器学习; 随机森林

基金项目: 国家自然科学基金 (82173463, 82003393)

Exploration of models of radiosensitive lipid metabolites of human plasma based on multiple machine learning algorithms

Chen Qi^{1,2}, Zhao Hua¹, Cai Tianjing¹, Gao Yizhe¹, Gao Ling¹, Liu Qingjie¹

¹Key Laboratory of Radiological Protection and Nuclear Emergency, China CDC, National Institute for Radiological Protection, Chinese Center for Disease Control and Prevention, Beijing 100088, China;

²Department of Acute Infectious Disease Prevention and Control, Institute of Infectious Disease Control and Prevention, Hubei Center for Disease Control and Prevention, Wuhan 430079, China

Corresponding author: Liu Qingjie, Email: liuqingjie@nirp.chinacdc.cn

【Abstract】 Objective To explore classification models for radiosensitive lipid metabolites in human peripheral blood by combining lipidomics with multiple machine learning (ML) algorithms.

Methods Totally 97 peripheral blood samples were collected from 25 leukemia cases admitted to a general

DOI: 10.3760/cma.j.cn112271-20231228-00225

收稿日期 2023-12-28 本文编辑 汤海澄

引用本文: 陈琦, 赵骅, 蔡恬静, 等. 基于多种机器学习算法的人血浆辐射敏感脂质代谢物模型探索[J]. 中华放射医学与防护杂志, 2024, 44(6): 457-463. DOI: 10.3760/cma.j.cn112271-20231228-00225.

Chen Q, Zhao H, Cai TJ, et al. Exploration of models of radiosensitive lipid metabolites of human plasma based on multiple machine learning algorithms[J]. Chin J Radiol Med Prot, 2024, 44(6): 457-463. DOI: 10.3760/cma.j.cn112271-20231228-00225.

hospital in Beijing from March to September 2023 who were ready to undergo bone marrow transplantation, including 0 Gy blood samples before irradiation in the control group ($n=24$), and 73 blood samples after irradiation at doses of 4, 8 and 12 Gy in the radiation group ($n=73$), and the targeted lipidomics based on the ultra-high performance liquid chromatography-mass spectrometry (UHPLC-MS) platform method to analyze the differences of different lipids between control and radiation groups. Then, lipids responsive to radiation doses of 0–12 Gy were identified using linear regression. Finally, classification models were constructed using five ML algorithms based on the training set, followed by the validation and evaluation of these models using the validation set. **Results** Compared with the control group, the differences in the concentration changes of 62 lipids in 9 classes of lipid metabolites sensitive to radiation group were statistically significant ($t = -4.91$ to 4.74 , $P < 0.05$), including sphingomyelins (SMs), cholesteryl esters (CEs), ceramides (Cers), phosphatidylinositols (PIs), hexosylceramides (HexCers), lysophosphatidylcholines (LysoPCs), phosphatidylcholines (PCOs), phosphatidylethanolamines (PEs), and lysophosphatidylethanolamines (LysoPEs). Twenty lipids responsive to radiation doses of 0–12 Gy were identified, namely 11 SMs, 7 CEs, 1 Cer, and 1 PI. The five models based on ML algorithms of decision tree (DT), support vector machine (SVM), light gradient boosting machine (Light GBM), random forest (RF), and K-nearest neighbors (KNN) all exhibited high goodness of fit ($F1 = 0.69-1.00$) and high sensitivity. The evaluation and validation metrics revealed that the RF-based model yielded the optimal radiation classification discrimination (sensitivity: 1.00; accuracy: 0.72; F1 score: 0.80). **Conclusions** Lipid metabolites responsive to radiation and lipids responsive to radiation dose in human samples were identified using targeted lipidomics. The RF-based model can provide new ideas for exploring models of human radiosensitive lipid metabolites.

[Key words] Ionizing radiation; Lipidomics; Machine learning; Random forest

Fund programs: National Natural Science Foundation of China (82173463, 82003393)

放射性或核紧急情况可能会导致数十人或者上百人的大规模伤亡,目前能实现短时间内简易分辨照射剂量的工具仍缺乏^[1-2]。染色体畸变分析和微核分析是个体剂量评估最有效的生物模拟技术,但需耗费大量的人力、物力和时间^[3],这使得在紧急情况下对人群进行辐射的评估和干预变得困难。因此,急需一种工具,可以从事故场景中快速、高效诊断未受影响或受照射剂量可控病例,以及需要紧急干预的辐射病例。辐射脂质组学是一种研究辐射诱导反应的脂质代谢反应新方法,可以在短时间内高通量地检测脂质分子的变化^[4-5],从而全面了解辐射对脂质代谢的影响。近年来在动物研究领域取得了许多新的发现和进展^[6-7],辐射标志物的新成果为研究辐射动物模型奠定了坚实的基础,并为非人灵长类动物或人类模型的研究提供了更多的可能性。本研究在前期已经建立的动物辐射诱导潜力标志物和剂量估算模型基础上,探索以全身照射的白血病病例血液样本开展生物模拟,筛选辐射扰动的脂质标志物。

机器学习(machine learning, ML)是一门多领域交叉学科^[8],位于人工智能(artificial intelligence, AI)的核心,通过大数据计算和模拟学习行为,获取新信息,不断训练和完善已有的成果。目前机器学习用于辐射领域的研究主要为肿瘤放射治疗的影像学和预后的研究^[9-10]。本研究

通过高通量辐射响应脂质代谢物与机器学习算法相结合,探索辐射剂量分类模型的建立和验证,有助于更好地应对辐射事故,并为应急响应提供数据支持。

材料与方法

1. 研究对象:2023年3—9月北京市某综合医院25例准备接受骨髓移植的白血病病例,收集病例接受0、4、8和12 Gy ^{60}Co γ 射线全身照射前后的血液样本。3例病例分别在0、4和12 Gy各缺失1份样本,累计收集到97份血液样本。其中,对照组为放射治疗前24份0 Gy血液样本,辐射组包括73份接受照射的样本(4、8和12 Gy剂量下分别为24、25和24份)。研究对象纳入标准为年龄 >18 岁,准备接受骨髓移植的确诊白血病病例,无其他纳入排除标准。本研究通过中国疾病预防控制中心辐射防护与核安全医学所伦理委员会批准(审批号:[2023]-1),所有参与者在纳入研究前均签署知情同意书。

2. 血浆制备:样本采集后转运至实验室,离心分离血浆,分装冻存于 -80°C 。检测前血浆样品于 4°C 解冻,提取100 μl 血浆至1.5 ml离心管,加入300 μl 异丙醇:乙腈=1:1。离心半径6 cm,12 000 r/min离心10 min,加入20 μl 内标,待检测。

3. 超高效液相色谱-质谱分析 (ultra-high performance liquid chromatography mass spectrometry, UPLC-MS): ACQUITY® UPLC I-CLASS 系统 (美国沃特世公司生产) 和 XEVO® TQ-S Micro 三重四极杆串联质谱仪 (美国沃特世公司生产) 以正离子和负离子分离模式洗脱。脂质在 2.1×100 mm, $1.7 \mu\text{m}$ BEH C8 色谱柱上, 毛细管电压为 3.0 kV。正离子流动相 A 为乙腈: 水 = 6: 4 和 5 mmol/L 甲酸铵 + 0.1% 甲酸; 流动相 B 为异丙醇: 乙腈 = 9: 1 和 5 mmol/L 甲酸铵 + 0.1% 甲酸。负离子流动相 A 为乙腈: 水 = 1: 9 和 1 mmol/L 乙酸铵 + 0.04% 乙酸, 流动相 B 为异丙醇: 乙腈 = 1: 1, 洗脱流速为 0.3 ml/min, 温度为 55℃。

4. 数据处理和统计学分析: 通过 Skyline 19.1 软件将原始数据进行校准峰值、变量积分和数据归一化, 使用多反应监测 (multiple reaction monitoring, MRM) 转换来量化脂质。Simca 14.1 软件开展正交偏最小二乘判别分析 (orthogonal partial least squares discriminant analysis, OPLS-DA), 根据变量投影重要度 (variable importance in projection, VIP) > 1、 $P < 0.05$ 来筛选对照组和辐射组组间辐射响应有统计学意义的脂质。通过拟合 0、4、8、12 Gy 这 4 个剂量组脂质浓度与辐射剂量线性模型, 根据 R^2 值 > 0.9 来筛选辐射诱导剂量响应脂质。

5. 机器学习建模和验证: 对照组和辐射组组间辐射响应有统计学意义的脂质作为变量。将对照组和辐射组 97 份样本均纳入数据集, 用简单随机抽样方法抽取 80% 作为训练集, 剩下 20% 作为验证集, 训练集和验证集样本无重叠。使用 R 4.3.2 软件, 调用机器学习数据包运行函数。分别采样决策树 (decision tree, DT)、支持向量机 (support vector machine, SVM)、轻量梯度增强机 (light gradient boosting machine, Light GBM)、随机森林 (random forest, RF)、K 最邻近 (K-nearest neighbors, KNN) 这 5 种机器学习方法, 识别独立的特征因子并拟合曲线, 建立分类预测模型。为确保稳健性, 本研究用不同的交叉验证种子重复了 20 次优化过程。一旦训练完成固定训练模型, 然后使用验证集数据开展模型验证和指标评价。根据灵敏度、特异度、准确度、精准确度、召回率和 F1 值来评价训练集建立的模型和验证集验证模型的性能。 $F1 = 2 \times (\text{精准确度} \times \text{召回率}) / (\text{精准确度} + \text{召$

回率), 取值范围是 0 到 1, 越接近 1, 模型性能越好。受试者工作特征 (receiver operating characteristic, ROC) 曲线计算曲线下面积 (area under the curve, AUC), AUC 范围在 0.5 和 1 之间, 越接近 1 模型分辨性能越高。

结 果

1. 辐射诱导脂质代谢物: 本研究通过高特异性和高灵敏度的超高效液相色谱质谱分析方法, 建立了人体辐射效应脂质数据库。采用监督模型偏最小二乘法判别分析, 辐射组与对照组相比, 62 个脂质代谢物差异有统计学意义 ($t = -4.91 \sim 4.74$, $P < 0.05$)。其中全身照射后浓度显著上升的脂质代谢物 50 个, 包括: 38 个鞘磷脂 (sphingomyelin, SM), 7 个胆固醇酯 (cholesteryl ester, CE), 2 个神经酰胺 (ceramide, Cer), 2 个磷脂酰肌醇 (phosphatidylinositol, PI), 1 个己糖神经酰胺 (hexosylceramide, HexCer); 浓度显著下降的脂质代谢物 12 个, 包括: 5 个溶血磷脂酰胆碱 (lysophosphatidylcholine, LysoPC), 3 个醚磷脂酰胆碱 (alkyl ether phosphatidylcholine, PCO), 1 个 Cer, 1 个磷脂酰乙醇胺 (phosphatidylethanolamine, PE), 1 个 PI, 1 个溶血磷脂酰乙醇胺 (lysophosphatidylethanolamine, LysoPE), 见图 1。7 个 CE 变化倍数 (fold change, FC) > 2, 提示辐射后有显著变化, 具有分辨样本是否受到照射的能力, 具备成为辐射诱导分类标志物潜力。分别为 CE 18: 1、CE 18: 2、CE 18: 3、CE 20: 3、CE 20: 4、CE 20: 5 和 CE 22: 6。

进一步采用热力图分析对照组与辐射组间变化程度最大的脂质差异代谢物, 主要为: CE 22: 6 (3.37 倍), CE 20: 3 (3.34 倍), CE 18: 2 (3.18 倍), CE 20: 4 (3.09 倍), CE 18: 1 (2.91 倍), CE 20: 5 (2.91 倍), CE 18: 3 (2.74 倍), LysoPC (20: 0) (0.68 倍), LysoPC (19: 0) (0.61 倍), LysoPC (20: 1) (0.59 倍)。图 2 热力图分析, 显示对照组与辐射组间差异最显著的 25 个脂质代谢物中 SM 占比为 76% (19/25), 其他 6 个脂质分别为 3 个 LysoPCs, 2 个 CEs 和 1 个 PI。其中辐射诱导后出现上调的脂质为 19 个 SMs, PI (18: 0/22: 6), CE 18: 1, CE 18: 3; 照射后出现下调的脂质 3 个, 为 LysoPC (20: 1), LysoPC (19: 0), LysoPC (17: 0)。其中, 热力

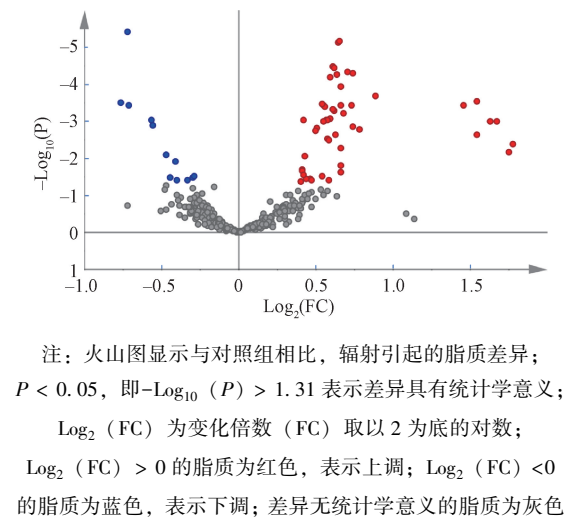


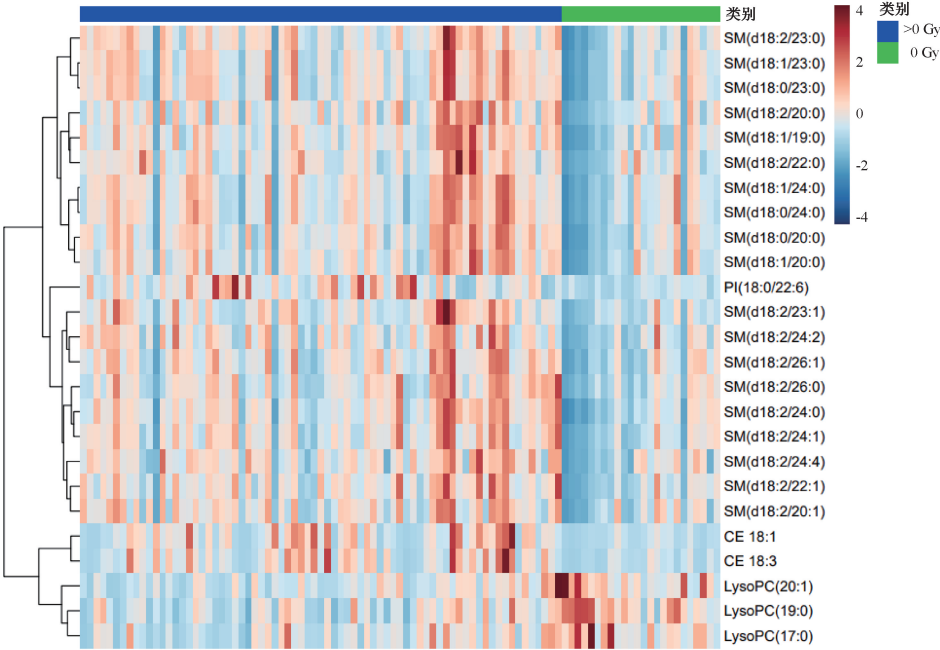
图1 对照组与辐射组426个血浆脂质代谢物火山图
Figure 1 Volcano plot of 426 plasma lipid metabolites in the control and radiation groups

图中筛选出的辐射响应脂质均包含于火山图分析中筛选出的差异脂质。

2. 辐射诱导剂量反应响应脂质：从辐射组与对照组中筛选的62个差异有统计学意义的脂质中，开展进一步剂量响应脂质筛选。共计20个脂质响应曲线 R^2 值 >0.9 ，包括11个SM：SM(d18：0/17：0)，SM(d18：0/18：0)，SM(d18：1/18：

0)，SM(d18：1/19：0)，SM(d18：1/20：0)，SM(d18：2/20：0)，SM(d18：2/22：1)，SM(d18：2/23：1)，SM(d18：2/24：0)，SM(d18：2/26：0)和SM(d18：2/26：1)，7个CE：CE 18：1，CE 18：2，CE 18：3，CE 20：3，CE 20：4，CE 20：5，CE 22：6，1个Cer(d18：2/18：0)和1个PI(18：2/18：2)。在0~12 Gy的剂量范围内，从4个剂量点样本不同的脂质代谢物中筛选出20种具有良好剂量反应的脂质代谢物。大多数脂质在12 Gy时变化最明显，这也验证了辐射剂量组在更高剂量时可以更好地分离和鉴定。随着剂量的增加，19种脂质代谢物的浓度显著增加，而PI(18：2/18：2)的浓度显著下降。为了更直观地展示这些脂质的变化模式，图3中选取了其中的4种具有代表性的脂质代谢物进行了图形展示。这20种脂质既可作为人体急性辐照后的分类标记，也可用于区分不同辐射剂量量组的标记。

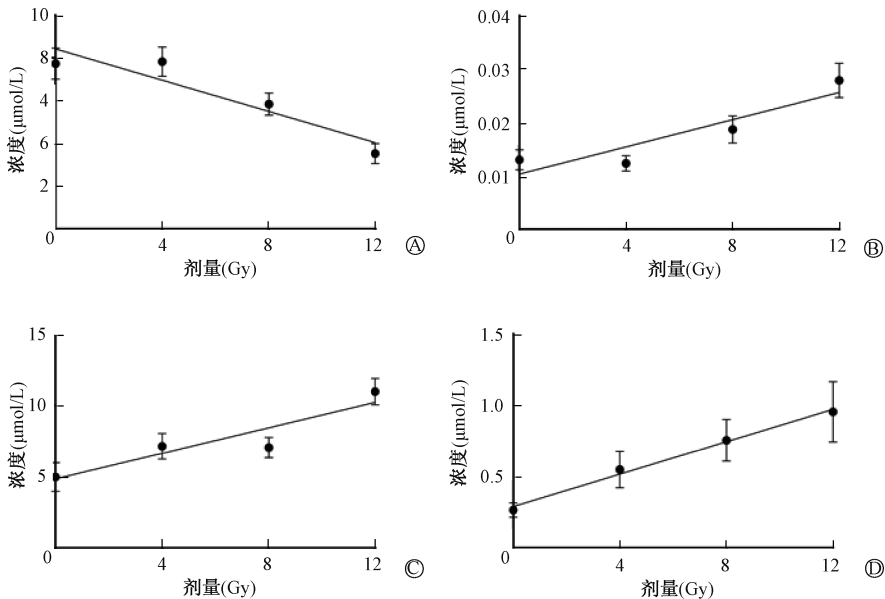
3. 机器学习建模和评价：通过辐射组与对照组筛选出的62个差异脂质代谢物作为建立模型的研究变量，分别调用5种机器学习方法。训练模型和模型验证得到以下参数(表1)，除KNN模型验证集灵敏度为0.90以外，其他4个模型的训练集



注：SM. 鞘磷脂；PI. 磷脂酰肌醇；CE. 胆固醇酯；LysoPC. 溶血磷脂酰胆碱。横坐标为纳入研究的97个样本，纵坐标为明显扰动的脂质代谢物。一种脂质代谢物在热力图上占据一行，通过颜色深浅来表示该脂质代谢物在不同样品中的表达水平。红色代表上调，蓝色代表下调，颜色越深表示脂质代谢物表达量越高

图2 97份样本中对对照组与辐射组差异最显著的前25个脂质代谢物

Figure 2 Top 25 lipid metabolites with the most significant differences between the control and radiation groups in 97 samples



注：PI. 磷脂酰肌醇；Cer. 神经酰胺；SM. 鞘磷脂；CE. 胆固醇酯

图3 20个辐射响应的剂量响应脂质代谢物中的4个脂质 ($\bar{x} \pm s$) A. PI (18 : 2/18 : 2) ;
B. Cer (d18 : 2/18 : 0) ; C. SM (d18 : 0/17 : 0) ; D. CE 20 : 5

Figure 3 Four dose-responsive lipids from 20 radiation-responsive lipid metabolites ($\bar{x} \pm s$)
A. PI (18 : 2/18 : 2) ; B. Cer (d18 : 2/18 : 0) ; C. SM (d18 : 0/1770) ; D. CE 20 : 5

和验证集以及 KNN 模型的训练集灵敏度均为 1.00。特异度只有 KNN 和随机森林模型训练集模型达到 1.00，其他模型均低于 0.50。5 个模型训练集的准确度和精准度指标均高于 0.80，但是验证集出现下降。KNN 模型验证集召回率为 0.9，其他 5 个模型均处于高召回率状态。综合评价指标 F1 值显示，支持向量机、决策树、Light GBM 模型训练性能较好，但是验证集的 F1 值在 0.7~0.8 之间，而 KNN 的验证集 F1 值低于 0.7，效果不理想。5 种机器学习中模型训练和验证效果最好的模型为随机森林，训练集 F1 值达到最佳性能 1.00，验证集 F1 值为 0.80。图 4 中随机森林和支持向量

机 ROC 曲线显示，随机森林在训练集中 AUC 为 1.00，模型验证 AUC 为 0.81，均显示出优秀的建模能力和良好的分辨能力。支持向量机模型训练和验证 AUC 分别为 0.72 和 0.63，表明模型建立可以达到中等性能，而分辨能力不够好。随机森林模型在辐射分类判定中的表现，证明了其在处理复杂、高维度数据时的优良性能。

讨 论

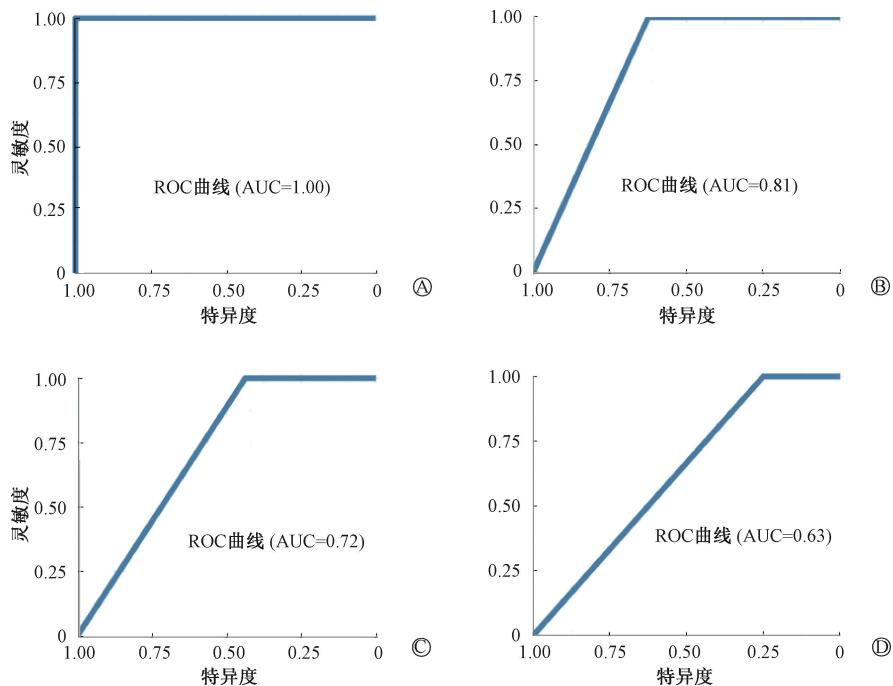
辐射脂质组学可以用于短时间高通量检测辐射诱导的脂质代谢反应。本课题组之前的研究表明，在大鼠模型的血浆样本和肠组织中筛选出脂肪酸 FA、SM、Cer、LysoPC、磷脂酰胆碱 (Phosphatidylcholine, PC)、PC-O、醚磷脂酰乙醇胺 (Alkyl ether Phosphatidylethanolamine, PE-O)、甘油三酯 (Triacylglyceride, TG) 等辐射后具有良好反应的脂质代谢物^[11-13]，具备成为辐射敏感标志物的潜质。Li 等^[14]在人组织中开展辐射研究，发现 6 种 TG 对辐射诱导效应明显。Unger 等^[15]分析了 11 例食管癌放疗病例血浆样本，发现 SM、PI、LysoPE、LysoPC、PC、FA 等脂质放疗后出现显著变化。本研究在人类样本中开展高通量脂质筛选，血浆中受到辐射扰动的脂质代谢物主要为 SM、CE、LysoPC、Cer、PI、HexCer、PCO、PE、

表 1 5 种机器学习模型训练集和验证集评价指标

Table 1 Evaluation metrics for the training and validation sets of five machine learning-based models

机器学习		灵敏度	特异度	准确度	精准度	召回率	F1 值
决策树	训练集	1.00	0.00	0.80	0.80	1.00	0.89
	验证集	1.00	0.00	0.56	0.56	1.00	0.71
KNN	训练集	1.00	1.00	1.00	1.00	1.00	1.00
	验证集	0.90	0.13	0.56	0.56	0.90	0.69
Light GBM	训练集	1.00	0.00	0.80	0.80	1.00	0.89
	验证集	1.00	0.00	0.56	0.56	1.00	0.71
随机森林	训练集	1.00	1.00	1.00	1.00	1.00	1.00
	验证集	1.00	0.38	0.72	0.67	1.00	0.80
支持向量机	训练集	1.00	0.44	0.89	0.88	1.00	0.93
	验证集	1.00	0.25	0.67	0.63	1.00	0.77

注：KNN. K 最邻近模型；Light GBM. 轻量梯度增强机模型



注：AUC. 曲线下面积；ROC. 受试者工作特征曲线

图 4 随机森林和支持向量机模型训练和模型验证 ROC 曲线 A. 随机森林模型训练；B. 随机森林模型验证；

C. 支持向量机模型训练；D. 支持向量机模型验证

Figure 4 ROC curves of the training and validation of the random forest and support vector machine-based models

A. Training of the RF-based model；B. Validation of the RF-based model；C. Training of the SVM-based model；

D. Validation of the SVM-based model

LysoPE。差异最显著的 25 个脂质代谢物中 SM 占比为 76%。同时，在 0~12 Gy 的剂量范围内，从 4 个剂量点样本不同的脂质代谢物中筛选出 SM、CE、Cer、PI 等 20 种具有良好剂量反应的脂质代谢物，其中含 11 个 SM。鞘磷脂在动物和人类样本中均表现出辐射诱导显著的效应。鞘磷脂在细胞膜的鞘磷脂代谢过程中发挥着关键作用，能够增加膜流动性、影响膜成分和功能。此外，鞘磷脂还参与了细胞内的信号转导过程，因此推测辐射可以通过破坏鞘磷脂代谢来影响细胞膜的稳定性及功能，同时影响信号转导过程，进一步导致辐射损伤。

当前放射领域的剂量估算模型建立主要基于 logistic 回归方法^[11-12]，但模型的验证和评价探索研究却显不足。机器学习模型在各个领域都有广泛的应用，包括线性回归、决策树、随机森林、支持向量机和朴素贝叶斯等模型。本研究通过前期筛选的辐射敏感脂质，探索建立人类样本急性电离辐射快速分类的评价模型的新思路。经过多个机器模型的评价，随机森林模型的灵敏度和准确度等各项评价指标，以及训练集和验证集 ROC

曲线均显示出明显的优势，但存在验证集特异度较低的问题。随机森林具有预测精度高和泛化能力强的特点，同时在特征选择和缺失值处理方面相对稳健，适合于人类复杂背景的样本分析^[8]。机器学习模型的训练和优化需要大量的数据和计算资源，样本量的不足也会造成训练不足，对预测结果产生一定的偏差。本研究初步验证了辐射组学技术与机器学习算法模型联合应用的可行性，下一步将收集多中心、大样本量来训练模型，优化参数和结构，提升剂量估算模型的准确性和可靠性。因医学伦理学的要求，健康人群无法进行全身照射实验，全身照射治疗的病人被视为最接近实际的最佳辐射生物模拟模型。放疗后脂质代谢物的变化揭示生物体对辐射的微妙反应，本研究发现了有潜力的辐射相关生物标志物，并探索了剂量响应趋势。

本研究表明，机器学习和脂质组学联合建立的预测模型，是未来开发应对短时间内大量个体中高通量、人工智能的快速分类鉴定工具的新思路。目前的模型可能还不够成熟和完善，但它提供了一个理论框架和工具，帮助我们更好地探索

和优化辐射剂量估算模型研究。随着人工智能的出现和机器学习等技术的不断发展^[16], 辐射脂质组学结合机器学习模型研究, 提示脂质组学在识别早期辐射风险、估算辐射照射剂量和开发精准辐射医疗等方面具有广阔的发展前景。

利益冲突 无

作者贡献声明 陈琦负责实验操作和论文撰写; 赵骅、蔡恬静、高义哲协助数据处理和技术指导; 高玲、刘青杰负责课题设计和论文修改

参 考 文 献

- [1] Li S, Lu X, Feng JB, et al. Developing gender-specific gene expression biodosimetry using a panel of radiation-responsive genes for determining radiation dose in human peripheral blood [J]. *Radiat Res*, 2019, 192 (4): 399-409. DOI: 10.1667/RR15355.1.
- [2] Sun JL, Li S, Lu X, et al. Identification of the differentially expressed protein biomarkers in rat blood plasma in response to gamma irradiation [J]. *Int J Radiat Biol*, 2020, 96 (6): 748-758. DOI: 10.1080/09553002.2020.1739775.
- [3] Swartz HM, Williams BB, Flood AB. Overview of the principles and practice of biodosimetry [J]. *Radiat Environ Biophys*, 2014, 53 (2): 221-232. DOI: 10.1007/s00411-014-0522-0.
- [4] Karahalil B. Overview of systems biology and omics technologies [J]. *Curr Med Chem*, 2016, 23 (37): 4221-4230. DOI: 10.2174/0929867323666160926150617.
- [5] Yan J, Risacher SL, Shen L, et al. Network approaches to systems biology analysis of complex disease: integrative methods for multi-omics data [J]. *Brief Bioinform*, 2018, 19 (6): 1370-1381. DOI: 10.1093/bib/bbx066.
- [6] 赵骅, 习聪, 田雪蕾, 等. 全身照射后大鼠血浆代谢特征分析 [J]. *中华放射医学与防护杂志*, 2021, 41 (6): 401-406. DOI: 10.3760/cma.j.issn.0254-5098.2021.06.001.
Zhao H, Xi C, Tian XL, et al. Analysis of the metabolic characteristics in rat plasma after total body irradiation [J]. *Chin J Radiol Med Prot*, 2021, 41 (6): 401-406. DOI: 10.3760/cma.j.issn.0254-5098.2021.06.001.
- [7] 田雪蕾, 刘海翔, 陆雪, 等. 环状 RNA hsa_circZDHHHC21_004 对⁶⁰Co γ 射线照射后人小肠上皮细胞 HIEC-6 增殖能力的影响 [J]. *中华放射医学与防护杂志*, 2023, 43 (5): 321-327. DOI: 10.3760/cma.j.cn112271-20230104-00002.
- [8] Tian XL, Liu HX, Lu X, et al. Effect of circular RNA hsa_circZDHHHC21_004 on the proliferation of the human small intestinal epithelial cells HIEC-6 exposed to ⁶⁰Co γ -rays [J]. *Chin J Radiol Med Prot*, 2023, 43 (5): 321-327. DOI: 10.3760/cma.j.cn112271-20230104-00002.
- [9] Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023 [J]. *N Engl J Med*, 2023, 388 (13): 1201-1208. DOI: 10.1056/NEJMr2302038.
- [10] Cheng J, Pan Y, Huang W, et al. Differentiation between immune checkpoint inhibitor-related and radiation pneumonitis in lung cancer by CT radiomics and machine learning [J]. *Med Phys*, 2022, 49 (3): 1547-1558. DOI: 10.1002/mp.15451.
- [11] Qin Q, Huang B, Wu A, et al. Chinese radiation intestinal injury research group. development and validation of a post-radiotherapy prediction model for bowel dysfunction after rectal cancer resection [J]. *Gastroenterology*, 2023, 165 (6): 1430-1442. e14. DOI: 10.1053/j.gastro.2023.08.022.
- [12] Zhao H, Xi C, Tian M, et al. Identification of potential radiation responsive metabolic biomarkers in plasma of rats exposed to different doses of cobalt-60 gamma rays [J]. *Dose Response*, 2020, 18 (4): 1559325820979570. DOI: 10.1177/1559325820979570.
- [13] Xi C, Zhao H, Lu X, et al. Screening of lipids for early triage and dose estimation after acute radiation exposure in rat plasma based on targeted lipidomics analysis [J]. *J Proteome Res*, 2021, 20 (1): 576-590. DOI: 10.1021/acs.jproteome.0c00560.
- [14] Liu HX, Lu X, Zhao H, et al. Enhancement of acylcarnitine levels in small intestine of abdominal irradiation rats might relate to fatty acid β -oxidation pathway disequilibrium [J]. *Dose Response*, 2022, 20 (1): 15593258221075118. DOI: 10.1177/15593258221075118.
- [15] Li W, Wang T, Zhang X, et al. Distinct lipid profiles of radiation-induced carotid plaques from atherosclerotic carotid plaques revealed by UPLC-QTOF-MS and DESI-MSI [J]. *Radiother Oncol*, 2022, 167: 25-33. DOI: 10.1016/j.radonc.2021.12.006.
- [16] Unger K, Li Y, Yeh C, et al. Plasma metabolite biomarkers predictive of radiation induced cardiotoxicity [J]. *Radiother Oncol*, 2020, 152: 133-145. DOI: 10.1016/j.radonc.2020.04.018.
- [17] Hosny A, Parmar C, Quackenbush J, et al. Artificial intelligence in radiology [J]. *Nat Rev Cancer*, 2018, 18 (8): 500-510. DOI: 10.1038/s41568-018-0016-5.