

doi: 10.12452/j.fxcxb.26060902

基于UPLC-HRMS技术结合化学计量学数据自动解析的肺炎患者血清代谢组学研究

杨丽华^{1,2}, 杨永刚¹, 马 阳¹, 李小花², 赵延博², 马成龙², 于永杰^{2*}

(1. 宁夏回族自治区第五人民医院, 宁夏 石嘴山 753000; 2. 宁夏医科大学 药学院, 宁夏 银川 750004)

摘要: 该研究对肺炎患者和健康人群的血清样本进行离心、冻干、复溶处理后, 取上清液过滤, 利用超高效液相色谱-高分辨质谱联用(UHPLC-HRMS)的非靶向代谢组学技术, 结合课题组开发的化学计量学数据自动解析方法(AntDAS)进行化合物信息提取、对齐、差异性成分筛查以及化合物鉴定, 最终筛选并鉴定出8个类别35个差异代谢物, 主要包括氨基酸及其衍生物、脂类、胆碱类、肉碱类、吲哚类、鞘氨醇类、胆汁酸类及嘌呤类。基于筛选的差异代谢物, 采用费希尔判别(Fisher)、K均值聚类(K-means)、偏最小二乘法判别分析(PLS-DA)及支持向量机(SVM)4种经典化学计量学方法构建诊断判别模型, 发现PLS-DA模型的拟合度及预测能力良好, 可准确区分肺炎组和健康组, 模型准确率达94.08%。研究结果有望为肺炎的精准诊断提供新的方案。

关键词: 肺炎; 血清; 超高效液相色谱-高分辨质谱(UHPLC-HRMS); AntDAS-DDA; 化学计量学; 代谢组学

中图分类号: O657.7; O652.9 **文献标识码:** A **文章编号:** 1004-4957(2026)11-0001-08

Metabolomics Study of Serum from Pneumonia Patients Based on UPLC-HRMS Technology Combined with Automated Chemometric Data Analysis

YANG Li-hua^{1,2}, YANG Yong-gang¹, MA Yang¹, LI Xiao-hua², ZHAO Yan-bo²,
MA Cheng-long², YU Yong-jie^{2*}

(1. The Fifth People's Hospital of Ningxia Hui Autonomous Region, Shizuishan 753000, China; 2. College of Pharmacy, Ningxia Medical University, Yinchuan 750004, China)

Abstract: Serum samples from patients with pneumonia and healthy controls were subjected to centrifugation, lyophilization, and reconstitution, followed by filtration of the supernatants. Ultra-high-performance liquid chromatography-high-resolution mass spectrometry (UHPLC-HRMS), combined with the automatic untargeted data analysis strategy AntDAS developed by the research team, was adopted for compound information extraction, alignment, screening of differential components, and compound identification. A total of 35 differential metabolites belonging to eight categories, including amino acids and their derivatives, lipids, cholines, carnitines, indoles, sphingosines, bile acids, and purines, were screened and identified. Based on these differential metabolites, four classical chemometric-based diagnostic discrimination models were evaluated: Fisher discriminant analysis (Fisher), k-means clustering (K-means), partial least squares discriminant analysis (PLS-DA), and support vector machine (SVM). Among these models, PLS-DA demonstrated satisfactory goodness-of-fit and predictive capability, accurately discriminating patients with pneumonia from healthy control group, with a model accuracy of 94.08%. These findings provide a potential new strategy for the precise diagnosis of pneumonia.

Key words: pneumonia; serum; ultra-high-performance liquid chromatography-high-resolution mass spectrometry; AntDAS-DDA software; chemometrics; metabolomics

肺炎是终末气道、肺泡和肺间质的炎症, 属于下呼吸道感染疾病, 具有高传染性和潜在的致死性,

收稿日期: 2026-06-09; 修回日期: 2026-07-25

基金项目: 宁夏回族自治区卫生健康青年科技培育项目(2025-NWQP-B024)

* 通讯作者: 于永杰, 博士, 教授, 研究方向: 化学计量学仪器分析新方法研究, E-mail: yongjie.yu@163.com

可由病原微生物、理化因素、免疫损伤、过敏及药物所致^[1]。肺炎按病因分为细菌性肺炎、病毒性肺炎、非典型病原体肺炎等^[1]。传统肺炎的诊断主要依据患者的临床体征、白细胞总数、C反应蛋白、降钙素原、血沉等生物标志物、病原学和肺部影像学等结果。然而,传统的病原检测技术、生物标志物、严重程度评分工具等存在局限性、滞后性、特异性差等问题^[2],比如在病原体不确定或病情复杂的情况下,存在肺部影像学诊断精度有限、生物标志物检测结果易出现假阴性或假阳性、病原学检测效率偏低、对病毒检测难度较大等问题,难以满足临床诊疗需求。因此,寻找新的肺炎诊断方法是十分必要的。

国内外学者围绕肺炎诊断已开展多维度、系统性的研究。在病原微生物检测方面,于婧等^[3]利用宏基因组二代测序建立了肺炎新诊断模式;在医学影像分析方面,刘艺文^[4]利用卷积神经网络(CNN)构建了肺炎诊断新方式,蔡娜等^[5]将机器学习算法用于肺炎的整合型鉴别诊断模型;在新型生物标志物方面,Arnold等^[6]研究了基于非侵入性操作的呼气生物标志物的肺炎诊断方法,Chen等^[7]运用代谢组学实现了细菌、真菌和病毒感染的肺炎诊断的判别分析,陈司琴^[8]利用代谢组学对社区获得性肺炎患者不同样本(肺泡灌洗液、血清、痰液)进行了新型代谢生物标记物研究。目前围绕肺炎诊断虽然开展了诸多研究,但多集中于影像技术、病原微生物方面,针对肺炎患者血清代谢组学的系统性研究相对较少,且缺乏具有临床转化价值的血清代谢标志物,不利于肺炎患者的早期识别和预后。

代谢组学是研究生物体(细胞、组织或生物体)在不同条件下产生的所有代谢物变化的科学,是基因组学和蛋白质组学的延伸和终端^[9]。代谢组学在恶性肿瘤、消化系统、代谢疾病、神经系统、呼吸系统、心血管类、风湿免疫等疾病的诊断和预后中发挥着重要作用^[10-13]。近年来,超高效液相色谱-高分辨质谱(UHPLC-HRMS)融合了液相色谱的高效分离和高分辨质谱的高质量准确度优势,已成为非靶向代谢分析中应用广泛、最具潜力的分析技术之一^[14]。然而在实际应用中,如何从采集的数据中有效提取化合物信息仍是亟待突破的关键瓶颈。目前基于UPLC-HRMS的代谢组学开发出的数据分析方法有XCMS^[15]、MS-DIAL^[16]、Mzmine2^[17]和Progenesis QI^[18]等,但在实际工作中存在的假阳性、假阴性、碎片离子识别等问题仍困扰研究人员。为解决这些挑战,作者课题组开发了色谱质谱数据智能解决方案AntDAS-DDA^[19-20],相较于现有分析软件,其在离子色谱图(EIC)提取、特征检测、MS/MS构建、碎片离子识别与全流程自动化方面优势显著,特别适合复杂样品的代谢组解析;不仅可以有效提取复杂样本的特征,还能充分利用MS/MS的优势,提高化合物的鉴定效率,为血清中代谢物的变化研究提供新的分析手段。

本研究首次将UHPLC-HRMS技术与AntDAS-DDA结合,用于人血清样本中代谢物的系统分析。以肺炎患者和健康人群的血清样本为研究对象,通过解析内源性代谢物的构成,利用化学计量学,筛查肺炎患者和健康人群的血清差异性代谢物,构建诊断判别模型,以期肺炎的快速诊断提供科学依据。

1 实验部分

1.1 仪器与试剂

仪器: Waters ACQUITY UPLC-AB SCIEX TripleTOF™ 5600 plus MS超高效液相色谱-高分辨质谱仪(美国沃特世公司和美国SCIEX公司); CV200真空离心浓缩仪(北京吉艾姆科技有限公司);自动涡旋混合器(天津药典标准仪器厂); Legend Micro 17型高速离心机、-80℃冰箱(赛默飞世尔科技公司)。

试剂: 甲醇、乙腈、异丙醇(均为色谱纯,赛默飞世尔科技有限公司); 甲酸(质谱级,德国默克公司); 蒸馏水(广州屈臣氏食品饮料有限公司)。

1.2 样本收集

本研究所需的血清样本来源于宁夏回族自治区第五人民医院检验科,经医院伦理委员会批准(伦理审批号: NWYLL202573),所有患者均签署知情同意书。样本于12 000 r/min离心10 min,取上清液分装到2 mL离心管中,封膜标记,置于-80℃冰箱,直至样本分析。

1.2.1 纳入标准 (1)肺炎患者: ①男女不限,18岁<年龄≤65岁,有高分辨率胸部影像学资料。②符合《中国成人社区获得性肺炎诊断和治疗指南(2016年版)》《呼吸疾病诊疗指南第3版》等疾病诊断标

准。③自愿且签署知情同意书。(2)健康人群: ①男女不限, 18岁<年龄≤65岁; ②实验室检验、影像学等检查结果未发现具有临床意义的阳性体征。

1.2.2 排除标准 (1)肺炎以外的其他诊断, 包括但不限于慢性阻塞性肺疾病、肺心病、哮喘者; (2)明显的意识障碍, 明显的神经肌肉病变, 如严重肌炎导致呼吸衰竭、肌肉感染、重症肌无力、脑梗塞偏瘫; (3)合并恶性肿瘤、血液系统疾病; (4)合并心、肝、肾等脏器功能衰竭, 肺心病不在此列; (5)妊娠、哺乳期患者; (6)经研究者判断无法配合完成研究的受试者; (7)资料不齐者。

1.3 样本前处理

取出-80℃冰箱中的血清样本40个(肺炎组20个, 健康组20个), 放置室温解冻。精密吸取100 μL置于离心管中, 加入300 μL乙腈, 涡旋30 s后置于-80℃冰箱冷冻沉淀蛋白1 h, 于4℃下15 000 r/min离心15 min, 吸取上清液置于离心管中, 离心冻干2 h后加入200 μL乙腈-水溶液(1:1), 涡旋30 s后在4℃下15 000 r/min离心15 min, 吸取上清液过滤后置于带有棕色内衬管的色谱瓶中进行分析。

从上述20个肺炎血清样本中各取100 μL均匀混合, 并加入甲醇, 在4℃下12 000 r/min离心10 min, 获得质量控制样品(QC)。进样过程中每5个样本中间穿插1个QC样本。样本总数为48个, 包括40个血清样本与8个QC样本。

1.4 仪器条件

超高效液相色谱条件: 色谱柱为Waters ACQUITY™ Premier BEH C₁₈(2.1 mm×100 mm, 1.7 μm); 柱温: 40℃; 进样量: 2 μL; 流速: 0.2 mL/min; 流动相A为0.1%甲酸-水, B为0.1%甲酸-乙腈。梯度洗脱程序: 0~7 min, 98%~19% A; 7~17 min, 19%~16% A; 17~25 min, 16%~10% A; 25~40 min, 10%~0% A; 40~45 min, 0% A; 45~48 min, 0%~98% A。

质谱条件: 使用MassLynx数据采集软件, 并在数据依赖性采集(DDA)模式下采集质谱信息。离子源使用电喷雾电离源(ESI), 正离子模式下采集。数据格式: 轮廓图; 采集模式: 灵敏度模式; 离子源温度: 100℃; 毛细管电压: 3.0 kV; 锥孔电压: 40 kV; 脱溶气温度: 450℃; 脱溶气流速: 800 L/h。质量轴校正液: 200 pg/μL亮氨酸脑啡肽(*m/z* 556.277 1 [M+H]⁺)。

一级质谱质量扫描范围为100~1 200 Da; 一级采集频率为0.2 s; 二级质谱质量扫描范围为50~1 000 Da; 二级采集频率为0.03 s, 每个循环监测的候选离子(TOP_n)选择6个母离子扫描二级质谱信息, 动态排除时间为6 s; MS/MS碰撞能量使用变化的碰撞能, 低质量碰撞能量: 10~20 V, 高质量碰撞能量: 40~60 V; 当MS1扫描的强度超过6 000 intensity/s时, 质谱仪自动切换到MS/MS模式, 且当扫描的强度达到40 000~60 000 intensity/s时停止。

1.5 数据处理

本研究采用MassLynx软件采集原始数据, 通过MS Convert(<https://proteowizard.sourceforge.io>)软件将原始数据转换为mzML格式并导入AntDAS-LCMS数据处理软件进行数据分析, 实现自动化的离子色谱(EIC)峰同组之间差异性化合物的分析, 结合偏最小二乘判别分析(PLS-DA)考察样本的聚类趋势与组间分离情况, 建立化学计量学诊断判别模型, 最后将筛选到的化合物质谱谱图与第三方数据库(<http://prime.psc.riken.jp/compms/msdial/main.html>)进行化合物匹配鉴定, 分析血清样本中代谢物的含量分布差异。

2 结果与讨论

2.1 基于AntDAS的化合物信息自动化解析

AntDAS能够直接读取UPLC-HRMS的原始数据并进行自动化分析, 提取其中的化学成分信息。图1为健康人群(A)和肺炎患者(B)血清样本的典型总离子流图(TIC), 由图可见血清样品丰富的化学信息, TIC峰较为均匀地分布在整个色谱图中。采用AntDAS-LCMS软件对原始数据进行预处理, 获得了26 101×48的EIC列表, 其中26 101为EIC峰数量, 48为样本数量。图2A给出了AntDAS提取其中1个EIC(*m/z* 312.326 1 Da)下所有色谱峰的示例, 共提取出10个EIC峰, *m/z*分布范围为312.326 0~312.326 3 Da, 极为接近。表明即使高分辨质谱能够提供高精度*m/z*值, 依然会出现多个离子, 使得研究人员难以准确筛选出目标化合物对应的离子, 给实际的分析工作带来挑战。图2A中的插图提取

EIC 对应的化合物注册信息表, 图 2B 和 2C 给出了色谱峰的局部放大图, 重点展示了流出时间在 7.5~8.5 min 和 12~20 min 区间内若干低信号峰的提取情况。由此可见, AntDAS 能够完成离子丰度较低的化合物中色谱峰的高质量提取, 并能够估计出色谱峰的流出区间。

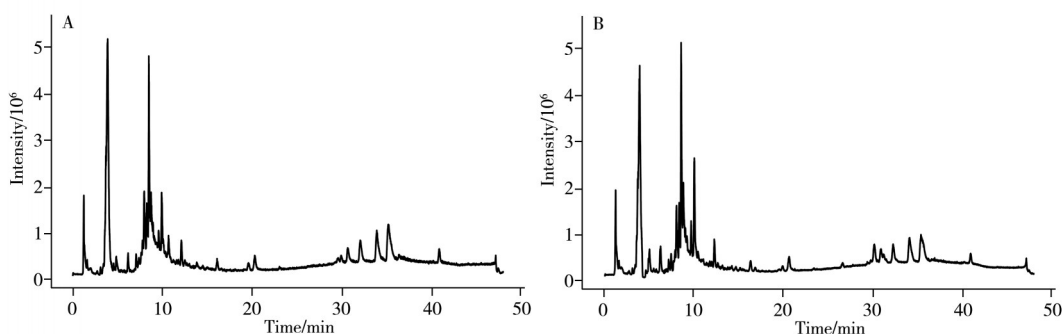


图 1 健康人群(A)和肺炎患者(B)血清样本的典型总离子流图

Fig. 1 Total ion chromatogram for serum samples from healthy individuals(A) and individuals with pneumonia(B)

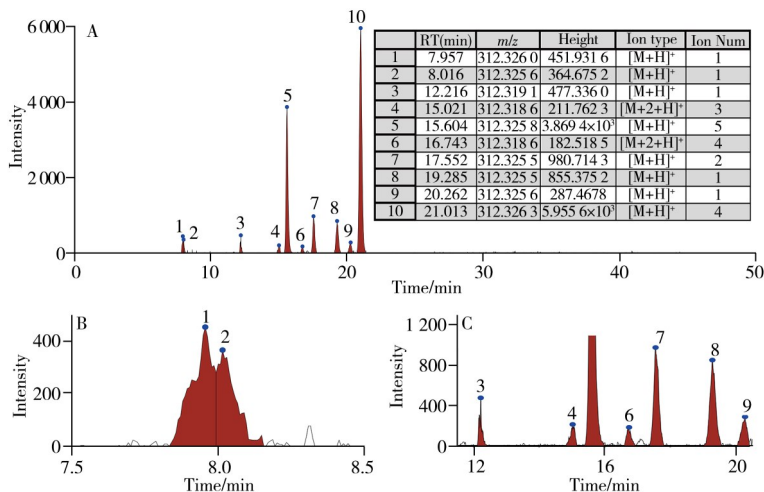


图 2 AntDAS 进行 EIC 峰提取及其细节图示例: EIC 色谱峰提取结果(A); EIC 色谱峰局部放大(B、C)

Fig. 2 Detailed diagram of EIC peak extraction with AntDAS: EIC peak extraction results(A); local magnification of the EIC peak(B, C)

2.2 化学计量学分析

利用方差分析(ANOVA)对已注册化合物的 EIC 峰进行分析, 筛选差异代谢物。本研究将置信水平设定为 0.05, 即 p 值(p -value) < 0.05; 从统计学角度来看, 具有显著性差异的成分应当是绝大多数组内样本均可检测到的物质。因此样本百分比(One group)设为 80%, 筛选出 11 502 个在不同组别间存在差异的 EIC 峰。随后基于差异性 EIC 峰进行主成分分析(PCA)和层次聚类分析(HCA), 结果如图 3 所示。HCA 分析结果(图 3A)显示, 肺炎组和健康组血清样本聚类效果明显, 各成一类。PCA 分析结果(图 3B)提示, 两组样本明显区分, 且两组样本明显地聚集在 95% 的置信区间内。PCA 分析得分图中第一主成分的贡献率为 29.2%, 第二主成分的贡献率为 13.7%, 主成分对样本信息的累计贡献率超过 40%。

2.3 血清差异代谢物的分析

在正离子模式下, 以酪氨酸为例展示 AntDAS-DDA 的化合物鉴别过程(图 4)。经 AntDAS DDA 解析获得多个 EIC 峰如 m/z 77.039 2、91.055 4、95.048 9、107.045 6、119.050 7、136.074 3、165.053 9、182.080 7 等, 图 4A 为 QC 样本在 m/z 182.080 7 下的 EIC 提取结果。AntDAS 能够实现源内裂解碎片的精准识别与归属, 有效避免将源内裂解碎片误判为独立化合物, 并构建化合物一级质谱谱图(MS1), 不仅降低了化合物的假阳性鉴定结果, 还能够作为辅助手段提高化合物鉴定的准确性。图 4C 为 AntDAS 构建的化合物一级质谱谱图, 由化合物源内裂解形成的碎片离子构成。图 4D 为 m/z 182.080 7 的二级质谱谱图(MS/MS), 该离子断裂后主要形成 m/z 136.074 3 和 165.061 4 的碎片离子, 与图 4C 中

MS 的离子一致, 表明 AntDAS 能够从复杂数据中准确构建化合物的 MS 和 MS/MS 谱图。图 4B 给出了保留时间为 2.247 5 min, m/z 为 182.080 7 的源内裂解识别过程, 可见各离子的 EIC 色谱峰保留时间一致, 色谱流出曲线高度接近, 表明可能来自同一个化合物。将图 4C 和图 4D 的质谱图与标准谱图进行比对, 根据匹配度将化合物最终鉴定为 *L*-酪氨酸(Tyrosine), MS1、MS/MS 的匹配度分别为 0.994 4、0.919 2, 显示出较高的可信度。

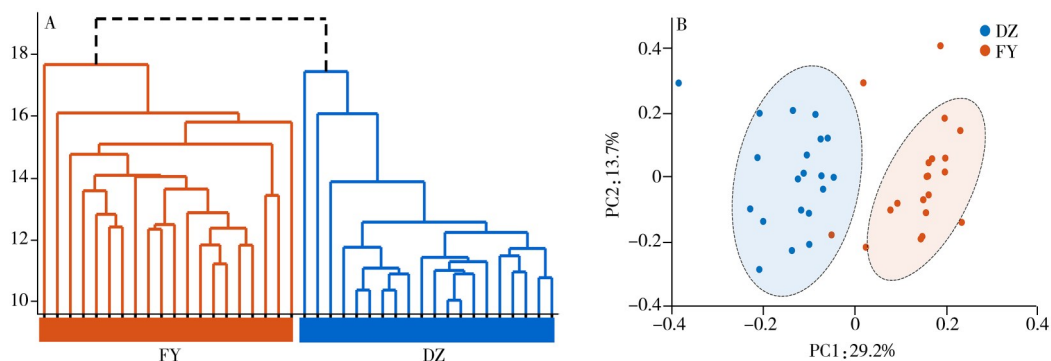


图 3 血清样本的 HCA(A)和 PCA(B)分析结果

Fig. 3 HCA(A) and PCA(B) analysis results of serum samples

DZ(Blue): healthy person; FY(Red): patients with pneumonia(DZ(蓝色): 健康人群; FY(红色): 肺炎患者)

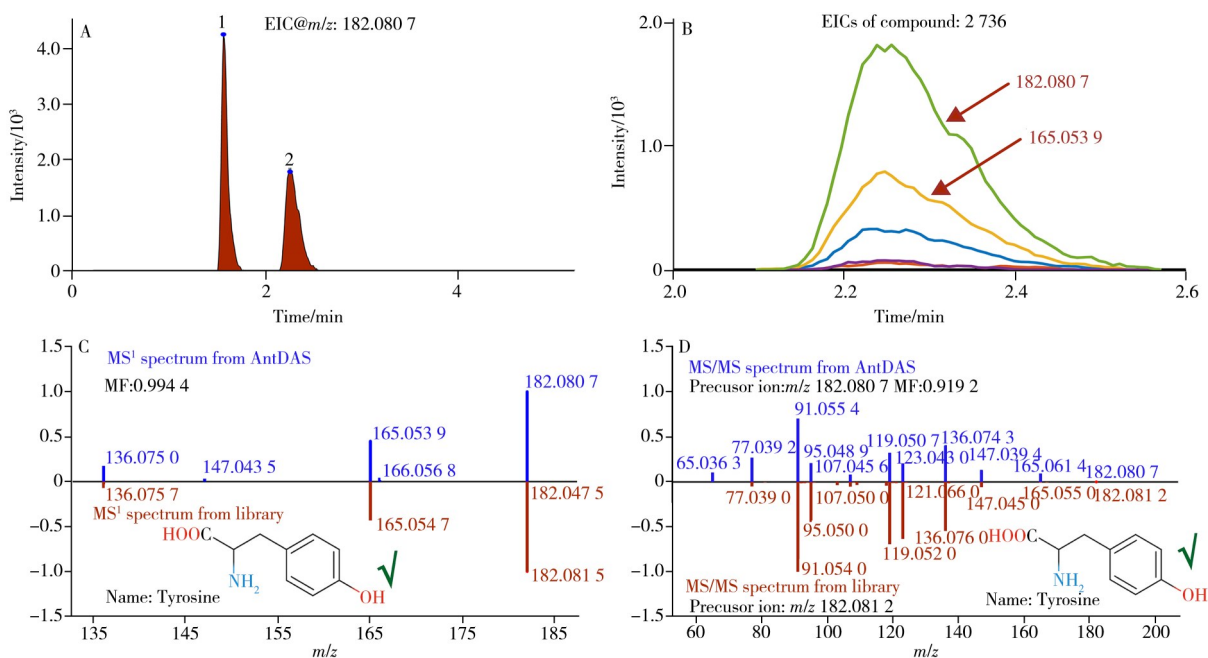


图 4 AntDAS 的化合物鉴定过程示例: QC 样本在 m/z 182.080 7 下的 EIC 提取结果(A); 化合物色谱峰流出曲线(B); 基于构建的 MS1 和 m/z 182.080 7 的 MS/MS 谱图的匹配结果(C、D)

Fig. 4 An example of the compound identification process by AntDAS: EIC extraction results for the QC sample at m/z 182.080 7 (A); chromatographic elution profiles of the ions(B); matched results based on the constructed MS1 spectrum and MS/MS spectra of m/z 182.080 7(C, D)

将解析获得的差异化合物的 MS1 及 MS/MS 质谱信息导入第三方数据库进行化合物匹配, 并进行人工逐个筛查, 共鉴定出匹配程度较高的差异代谢物 35 个(见表 1), 包括胆碱类、肉碱类、吡啶类、鞘氨醇类、胆汁酸类、氨基酸及其衍生物、嘌呤类及脂类共 8 类。从代谢物数量看, 氨基酸及其衍生物最多, 有 11 个, 占总数的 31.43%, 其次为脂类(6 个)、肉碱类(5 个), 鞘氨醇类和嘌呤类均为 4 个。本研究以质谱匹配度 0.65 为界定阈值, 匹配度 ≥ 0.65 的化合物认定为结构初步确证成分, 低于该阈值的为仅推定候选成分; 低匹配度候选物仅作化学成分谱辅助参考, 不参与后续标志物筛选与诊断建模分析。

表 1 AntDAS 鉴定出的 35 个差异代谢物

Table 1 The 35 significant difference metabolites identified by AntDAS

No.	t_R/min	m/z	Adduct ion	MatchD	Formula	Compound	CAS No.	Classification
1	1.227	104.105 5	$[\text{M}]^+$	0.800 6	$\text{C}_5\text{H}_{14}\text{NO}$	Choline(胆碱)	62-49-7	Choline(胆碱类)
2	6.555	288.288 9	$[\text{M}+\text{H}]^+$	0.918 2	$\text{C}_{17}\text{H}_{37}\text{NO}_2$	C17-Sphinganine (<i>D</i> -赤型鞘氨醇 C17 碱)	32164-02-6	Sphingoid base (鞘氨醇类)
3	7.115	300.288 8	$[\text{M}+\text{H}]^+$	0.982 7	$\text{C}_{18}\text{H}_{37}\text{NO}_2$	Sphingosine(<i>D</i> -鞘氨醇)	123-78-4	Sphingoid base (鞘氨醇类)
4	12.964	284.294 2	$[\text{M}-\text{H}_2\text{O}+\text{H}]^+$	0.991 4	$\text{C}_{17}\text{H}_{37}\text{NO}$	1-Desoxymethylsphinganine(1-脱氧甲基 鞘氨醇 m17 : 0)	1219484- 98-6	Sphingoid base (鞘氨醇类)
5	6.205	318.300 2	$[\text{M}+\text{H}]^+$	0.671 6	$\text{C}_{18}\text{H}_{39}\text{NO}_3$	Phytosphingosine(糖脂)	554-62-1	Sphingoid base (鞘氨醇类)
6	3.051	146.059 2	$[\text{M}+\text{H}]^+$	0.592 9	$\text{C}_9\text{H}_7\text{NO}$	3-Formylindole(3-吲哚甲醛)	487-89-8	Indole(吲哚类)
7	4.736	176.070 1	$[\text{M}+\text{H}]^+$	0.995 1	$\text{C}_{10}\text{H}_9\text{NO}_2$	3-Indoleacetic acid(3-吲哚乙酸)	87-51-4	Indole(吲哚类)
8	1.235	130.048 4	$[\text{M}+\text{H}]^+$	0.972 4	$\text{C}_5\text{H}_9\text{NO}_3$	<i>L</i> -Pyroglutamic acid(<i>L</i> -焦谷氨酸)	98-79-3	Amino acids(氨基酸类)
9	1.251	162.111 8	$[\text{M}+\text{H}]^+$	0.820 5	$\text{C}_7\text{H}_{15}\text{NO}_3$	<i>L</i> -Carnitine(左旋肉碱)	541-15-1	Amino acids(氨基酸类)
10	1.266	118.085 0	$[\text{M}+\text{H}]^+$	0.978 1	$\text{C}_5\text{H}_{11}\text{NO}_2$	Betaine(甜菜碱)	107-43-7	Amino acids(氨基酸类)
11	1.336	116.069 7	$[\text{M}+\text{H}]^+$	0.998 7	$\text{C}_5\text{H}_9\text{NO}_2$	<i>L</i> -Proline(<i>L</i> -脯氨酸)	147-85-3	Amino acids(氨基酸类)
12	1.555	118.085 0	$[\text{M}+\text{H}]^+$	0.686 8	$\text{C}_5\text{H}_{11}\text{NO}_2$	5-Aminovaleric acid(5-氨基戊酸)	660-88-8	Amino acids(氨基酸类)
13	1.555	182.080 7	$[\text{M}+\text{H}]^+$	0.951 7	$\text{C}_9\text{H}_{11}\text{NO}_3$	Tyrosine(<i>L</i> -酪氨酸)	60-18-4	Amino acids(氨基酸类)
14	1.122	147.112 0	$[\text{M}+\text{H}]^+$	0.744 7	$\text{C}_6\text{H}_{14}\text{N}_2\text{O}_2$	Lysine(<i>L</i> -赖氨酸)	56-87-1	Amino acids(氨基酸类)
15	1.573	132.101 3	$[\text{M}+\text{H}]^+$	0.978 6	$\text{C}_6\text{H}_{13}\text{NO}_2$	<i>L</i> -Isoleucine(<i>L</i> -异亮氨酸)	73-32-5	Amino acids(氨基酸类)
16	1.573	294.153 9	$[\text{M}+\text{H}]^+$	0.743 1	$\text{C}_{12}\text{H}_{23}\text{NO}_7$	<i>N</i> -Fructosyl isoleucine(果糖异亮氨酸)	87304-79-8	Amino acids(氨基酸类)
17	2.555	166.086 0	$[\text{M}+\text{H}]^+$	0.944 1	$\text{C}_9\text{H}_9\text{NO}_2$	Phenylalanine(苯丙氨酸)	62056-68-2	Amino acids(氨基酸类)
18	2.660	245.186 2	$[\text{M}+\text{H}]^+$	0.711 9	$\text{C}_{12}\text{H}_{24}\text{N}_2\text{O}_3$	Isoleucylleucine(<i>L</i> -异亮酰胺- <i>L</i> -亮氨酸)	26462-22-6	Amino acids(氨基酸类)
19	8.023	400.346 5	$[\text{M}+\text{H}]^+$	0.956 2	$\text{C}_{25}\text{H}_{46}\text{ClNO}_4$	Palmitoyl- <i>DL</i> -carnitine chloride (±)-棕榈酰肉碱氯化物	6865-14-1	Carnitine(肉碱类)
20	5.817	316.248 7	$[\text{M}+\text{H}]^+$	0.588 4	$\text{C}_{17}\text{H}_{33}\text{NO}_4$	Decanoyl- <i>L</i> -carnitine(左旋肉碱癸酰酯)	3992-45-8	Carnitine(肉碱类)
21	6.571	344.278 3	$[\text{M}+\text{H}]^+$	0.770 6	$[\text{C}_{23}\text{H}_{46}\text{NO}_4]^+$	<i>DL</i> -Lauroylcarnitine(月桂酰左旋肉碱)	7023/3/2	Carnitine(肉碱类)
22	7.694	424.341 6	$[\text{M}+\text{H}]^+$	0.905 4	$\text{C}_{25}\text{H}_{45}\text{NO}_4$	Linoleyl carnitine(亚油酰基肉碱)	36816-10-1	Carnitine(肉碱类)
23	1.555	204.122 1	$[\text{M}]^+$	0.979 9	$\text{C}_9\text{H}_{18}\text{NO}_4$	Acetylcarnitine(乙酰基肉碱)	14992-62-2	Carnitine(肉碱类)
24	1.565	136.075 1	$[\text{M}+\text{H}]^+$	0.382 4	$\text{C}_5\text{H}_5\text{N}_5$	Adenine(腺嘌呤)	73-24-5	Purine(嘌呤类)
25	1.565	137.047 9	$[\text{M}+\text{H}]^+$	0.900 9	$\text{C}_5\text{H}_7\text{N}_4\text{O}$	Hypoxanthine(次黄嘌呤)	68-94-0	Purine(嘌呤类)
26	1.555	169.034 5	$[\text{M}+\text{H}]^+$	0.711 1	$\text{C}_5\text{H}_4\text{N}_4\text{O}_3$	Uric acid(尿酸)	69-93-2	Purine(嘌呤类)
27	2.289	181.025 1	$[\text{M}+\text{Na}]^+$	0.857 6	$\text{C}_4\text{H}_6\text{N}_4\text{O}_3$	Allantoin(尿囊素)	97-59-6	Purine(嘌呤类)
28	7.803	520.339 4	$[\text{M}+\text{H}]^+$	0.994 3	$\text{C}_{26}\text{H}_{50}\text{NO}_7\text{P}$	LPC 18 : 2(1-亚油酰-2-羟基-SN-甘油 -3-卵磷脂)	22252-07-9	Lipid(脂类)
29	7.957	478.292 2	$[\text{M}+\text{H}]^+$	0.925 4	$\text{C}_{23}\text{H}_{44}\text{NO}_7\text{P}$	LPE 18 : 2(溶血磷脂酰乙醇胺 18 : 2)	85046-18-0	Lipid(脂类)
30	11.416	305.247 7	$[\text{M}+\text{H}]^+$	0.687 5	$\text{C}_{20}\text{H}_{32}\text{O}_2$	Arachidonic acid(花生四烯酸)	506-32-1	Lipid(脂类)
31	10.837	326.304 7	$[\text{M}+\text{H}]^+$	0.821 1	$\text{C}_{20}\text{H}_{39}\text{NO}_2$	<i>N</i> -Oleoyl ethanolamine(油酰乙醇胺)	111-58-0	Lipid(脂类)
32	11.672	282.279 0	$[\text{M}+\text{H}]^+$	0.521 0	$\text{C}_{18}\text{H}_{35}\text{NO}$	Oleamide(油酰胺)	301-02-0	Lipid(脂类)
33	11.672	313.273 9	$[\text{M}-\text{H}_2\text{O}+\text{H}]^+$	0.921 9	$\text{C}_{19}\text{H}_{38}\text{O}_4$	2, 3-Dihydroxypropyl hexadecanoate (甘油单棕榈酸酯)	26657-96-5	Lipid(脂类)
34	5.600	450.320 8	$[\text{M}+\text{H}]^+$	0.878 2	$\text{C}_{26}\text{H}_{43}\text{NO}_5$	Glycochenodeoxycholic acid(甘氨鹅脱 氧胆酸)	640-79-9	Bile acid(胆汁酸类)
35	6.378	158.153 4	$[\text{M}+\text{H}]^+$	0.913 2	$\text{C}_{24}\text{H}_{40}\text{O}_5$	Hyochoic acid(猪胆酸)	547-75-1	Bile acid(胆汁酸类)

根据 35 种差异性代谢物的相对丰度进行热图分析(见图 5)。结果表明,肺炎组和健康组血清样本中代谢物的含量分布有显著差异。其中肺炎患者中胆碱、苯丙氨酸、花生四烯酸、果糖异亮氨酸、*D*-鞘氨醇、*L*-赖氨酸、*L*-酪氨酸、腺嘌呤等相对健康组显著富集。健康人群中左旋肉碱癸酰酯、1-亚油酰-2-羟基-SN-甘油-3-卵磷脂、次黄嘌呤、甘油单棕榈酸酯、甘氨酰脱氧胆酸等含量相对肺炎组较高。整体来看,肺炎患者血清样本中的氨基酸及其衍生物类、脂类代谢物含量高,健康人群中氨基酸和胆汁酸类代谢产物含量高。肺炎患者血清中鉴定出的代谢物 *D*-鞘氨醇属于鞘脂类代谢产物,其参与炎症反应和免疫调节,鞘脂类代谢异常可能与疾病的进展和发病机制密切相关^[21]。

2.4 肺炎诊断模型的构建

基于鉴定的 35 个代谢物,采用费希尔判别(Fisher)、K 均值聚类(K-means)、偏最小二乘法判别分析(PLS-DA)及支持向量机(SVM)4 种算法构建肺炎诊断判别模型。从血清样本中随机选择 80% 的样本用于建模,剩余 20% 样本作为预测样本,随机抽取 1 000 次统计模型预测误差。本研究模型仅基于同批

次样本开展内部随机验证，无独立外部样本校验，模型跨批次泛化能力存在一定局限。后续可收集不同批次的独立样本进行外部验证，进一步完善并提升肺炎诊断判别模型的稳定性与实用价值。所得混淆矩阵结果如表 2 所示。各模型对肺炎血清样本的预测集识别平均准确率分别为 Fisher 模型 72. 23%、K-means 模型 89. 71%、PLS-DA 模型 94. 08%、SVM 模型 88. 16%，4 种算法均表现出良好的判别能力，其中 PLS-DA 模型的整体性能最优。在 PLS-DA 模型中，肺炎患者和健康人群的血清样本均得到准确识别，其中肺炎患者的识别率达 99. 57%，健康人群有小比例的判错情况。上述基于所鉴定差异性代谢物的诊断判别结果与所有差异性代谢物的聚类结果(图 3)相一致，表明本研究构建的基于 AntDAS 结合 UHPLC-HRMS 的策略能够有效区分健康人群和肺炎患者血清样本，充分验证了基于代谢组学结合多算法建模在肺炎诊断中的可靠性与适用性。

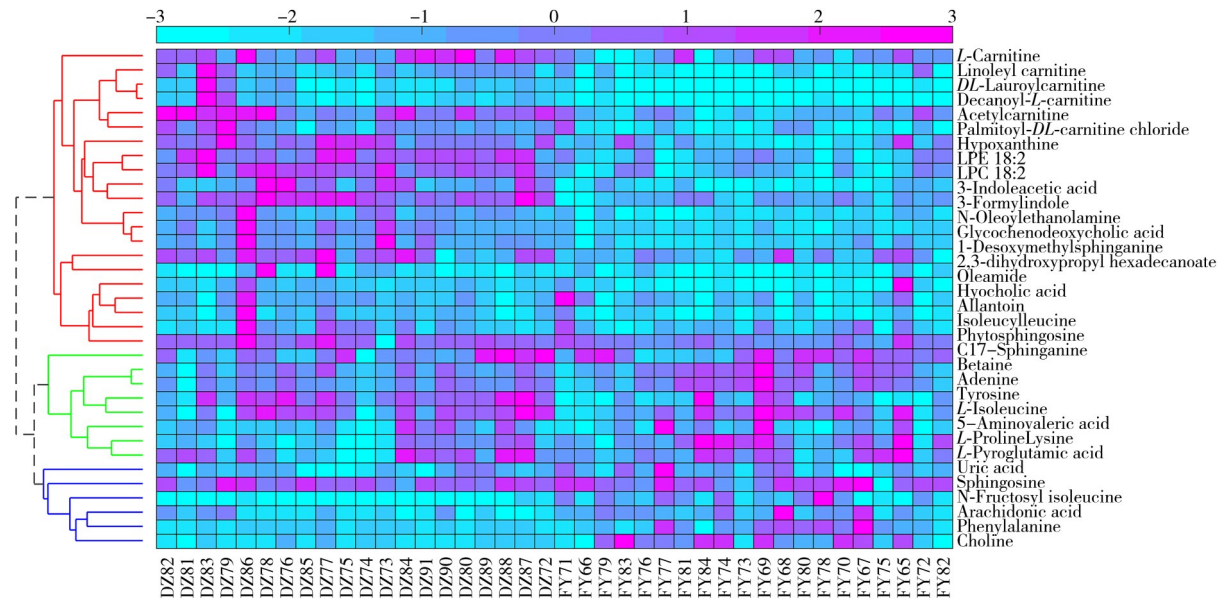


图 5 基于 AntDAS 血清样本差异代谢物的聚类热图分析

Fig. 5 Cluster heatmap analysis of metabolites differing between serum samples based on AntDAS

表 2 不同算法构建肺炎诊断模型的混淆矩阵结果

Table 2 Results of confusion matrix for pneumonia diagnosis model constructed using different algorithms

Discrimination model	Prediction sample	Real sample		Average accuracy rate/%
		Healthy population accuracy rate/%	Patients with pneumonia accuracy rate/%	
Fisher	健康人群	75. 65	24. 35	72. 23
	肺炎患者	31. 20	68. 80	
K-means	健康人群	92. 98	7. 03	89. 71
	肺炎患者	13. 58	86. 43	
PLS-DA	健康人群	88. 59	11. 41	94. 08
	肺炎患者	0. 43	99. 57	
SVM	健康人群	86. 33	13. 68	88. 16
	肺炎患者	10. 03	89. 98	

3 结 论

本文采用化学计量学液质联用数据解析软件 AntDAS 结合 UPLC-HRMS 非靶向代谢组学技术对血清样本中的差异代谢物进行分析，研究结果表明，肺炎患者和健康人群的血清代谢水平存在显著差异，并筛选鉴定出 35 种差异代谢物，涵盖 8 种化合物类别，胆碱、苯丙氨酸、花生四烯酸、果糖异亮氨酸、D-鞘氨醇、L-赖氨酸、L-酪氨酸、腺嘌呤等在肺炎患者血清中的含量相对较高。基于这些差异性代谢物构建 PLS-DA 模型，模型聚类清晰、组间分离好，可有效区分肺炎组和健康组。尽管本研究证实了 UHPLC-HRMS 结合 AntDAS 在区分肺炎人群方面具有一定的潜在应用价值，但有必要指出，受限于研究中的样本量，模型的优化和调整需通过更大规模的样本容量做进一步的临床验证。另一方面，鉴于目前缺乏统一的液质高分辨质谱数据库且第三方数据库覆盖化合物有限的实际情况，本研究中难免遗漏潜在的生物标志物。后续研究中，作者会进一步扩充样本容量，以提升判别模型泛化能力。同时亦

将深入探究代谢物的含量差异,为肺炎的临床诊断提供更完善的理论支撑。

参考文献:

- [1] Chen H Z, Zhong N S, Lu Z Y. *Internal Medicine*. 9rd ed. Beijing: People's Medical Publishing House(陈灏珠, 钟南山, 陆再英. 内科学. 第九版. 北京: 人民卫生出版社), **2019**: 41-45.
- [2] Zhang F, Song L, Han F F. *Chin. J. Respir. Crit. Care Med.* (张帆, 宋琳, 韩锋锋. 中国呼吸与危重监护杂志), **2024**, 23(10): 741-744.
- [3] Yu J, Zhao J, Fu L L, Cui T T, Pan W S. *Chin. J. Difficult Complicated Cases*. (于婧, 赵靖, 付琳琳, 崔彤彤, 潘文森. 疑难病杂志), **2026**, 25(3): 316-322.
- [4] Liu Y W. *Research on Intelligent Analysis Method and Clinical Application of Multimodal Imaging for Pneumonia*. Shanghai: Donghua University(刘艺文. 面向肺炎多模态影像的智能分析方法与临床应用研究. 上海: 东华大学), **2025**.
- [5] Cai N, Zhao W K. *Chin. J. Nosocomiol.* (蔡娜, 赵卫康. 中华医院感染学杂志), **2026**, (11): 1692-1697.
- [6] Arnold C G, McCartney M M, Davis C E, Mizgerd J P, Kenyon N J. *Expert Rev. Respir. Med.*, **2025**, (20): 491-500.
- [7] Chen L, Xue J B, He Y K, Zhao L L, Zhang Y, Yin L, Fu S N, Yu W Y, Ma X Q, Wang Y, Tang Y F, Gao Z C. *Ann. Med.*, **2024**, 56: 2399320.
- [8] Chen S Q. *Metabolic Biomarkers for Accurate Diagnosis and Severity Assessment in Community Acquired Pneumonia*. Guangzhou: Southern Medical University(陈司琴. 社区获得性肺炎精准诊断和病情分型的新型代谢生物标记物研究. 广州: 南方医科大学), **2024**.
- [9] Tian L, Yang B, Hu X L, Du J Y, Tang S R, Liu G S. *J. Instrum. Anal.* (田力, 杨宝, 胡雪黎, 杜桂艳, 唐生尧, 刘国生. 分析测试学报), **2023**, 42(5): 596-604.
- [10] Liu Y F, Fan X Y, Yang H T, Lu X D. *J. Liaoning Univ: Nat. Sci. Ed.* (刘玉峰, 范笑荧, 杨宏天, 卢晓丹. 辽宁大学学报: 自然科学版), **2020**, 47(2): 149-155.
- [11] Zhang H L, Si Y M. *Acta Chin. Med.* (张海龙, 司一妹. 中医学报), **2020**, 35(9): 1913-1917.
- [12] Jiang T, Deng J L, Hu X J, Yao D S, Chen Q G, Hu R M, Ma X X, Tu L P, Tan X, Yuan W, Ma L Z, Cui J, Lu H, Xu J T. *Front. Immunol.*, **2026**, 17: 1781013.
- [13] Mai D M, Li Z X, Cao Z R, Lin P Q, Tan J Q, Li R, Ye Q. *Phytother. Res.*, **2026**, 40(2): 829-843.
- [14] Li S F, Liu W B, Feng S H, Yang Y Q, Hu Y J, Hao X F, Cheng Y Y, Yu Y J. *J. Instrum. Anal.* (李淑芳, 刘文斌, 冯书惠, 杨亚琴, 胡永建, 郝学飞, 程莹莹, 于永杰. 分析测试学报), **2025**, 44(10): 2143-2152.
- [15] Giera M, Aisporna A, Uritboonthai W, Hoang L, Derks R J E, Joseph M K, Baker E S, Siuzdak G. *Mol. Syst. Biol.*, **2024**, 20: 1153-1155.
- [16] Takeda H, Matsuzawa Y, Takeuchi M, Takahashi M, Nishida K, Harayama T, Todoroki Y, Shimizu K, Sakamoto N, Oka T, Maekawa M, Chung M H, Kurizaki Y, Kiuchi S, Tokiyoshi K, Buyantogtokh B, Kurata M, Kvasnička A, Takeda U, Uchino H, Hasegawa M, Miyamoto J, Tanabe K, Takeda S, Mori T, Kumakubo R, Tanaka T, Yoshino T, Okamoto M, Takahashi H, Arita M, Tsugawa H. *Nat. Commun.*, **2024**, 15: 9903.
- [17] Schmid R, Heuckeroth S, Korf A, Smirnov A, Myers O, Dyrland T S, Bushuiev R, Murray K J, Hoffmann N, Lu M, Sarvepalli A, Zhang Z, Fleischauer M, Dührkop K, Wesner M, Hoogstra S J, Rudt E, Mokshyna O, Brungs C. *Nat. Biotechnol.*, **2023**, 41: 447-449.
- [18] Wang X C, Ma X L, Liu J N, Zhang Y, Zhang J N, Ma M H, Ma F L, Yu Y J, She Y. *Anal. Chim. Acta*, **2023**, 1254: 341127.
- [19] Wang X C, Zhang J N, Zhao J J, Guo X M, Li S F, Zheng Q X, Liu P P, Lu P, Fu H Y, Yu Y J, She Y. *Anal. Chem.*, **2023**, 95: 638-649.
- [20] Zhang J N, Du Q Y, Wang X C, Liu J N, Ma M H, Ma X L, Zhao J J, Guo X M, Zhang Y, Yu Y J. *J. Instrum. Anal.* (张佳妮, 都青钰, 汪兴财, 刘嘉楠, 马梦晗, 马兴玲, 赵娟娟, 郭晓萌, 张阳, 于永杰. 分析测试学报), **2022**, 41(8): 1178-1184, 1192.
- [21] Lee M, Lee S Y, Bae Y S. *Exp. Mol. Med.*, **2023**, 55(6): 1110-1130.

(责任编辑: 丁 岩)