

基于血浆代谢组学与机器学习的缺血性卒中亚型判别及隐源性卒中病因预测

自胤余, 刘子森, 邓 镇

南方医科大学南方医院神经内科, 广东 广州 510515

摘要:目的 分析缺血性卒中不同亚型之间血浆代谢特征的差异,并通过10种机器学习算法构建识别缺血性卒中不同亚型的模型,为隐源性卒中的辨识提供客观化依据,以提高临床诊断率。方法 本研究共招募了来自3家医院的277例急性缺血性卒中(AIS)患者作为研究对象,并根据TOAST分型对患者进行分组。所有患者采集静脉血样并进行¹H-NMR代谢组检测,比较分析并鉴定了46种代谢物。采用代谢组学和孟德尔随机化分析来识别潜在的生物标志物,PyCaret用于开发和优化机器学习(ML)模型。开发血浆代谢物机器学习(PMML)模型用以区分缺血性卒中的不同亚型,并将隐源性卒中(SUE)队列用以外验证模型效能。结果 在大动脉粥样硬化(LAA)、小动脉闭塞(SVO)和心源性栓塞(CE)之间观察到代谢物谱的显著差异。孟德尔随机化分析表明,葡萄糖和脯氨酸可能作为SVO和LAA的潜在生物标志物。基于3种诊断良好的亚型(LAA、SVO和CE)的代谢谱,本研究开发了血浆代谢物机器学习(PMML)模型,该模型在区分上述缺血性卒中亚型方面表现出较高的预测准确性(准确率:0.8630;ROC曲线下面积:0.9586)。SUE患者被用作验证队列,其中,被预测为LAA的患者,23.9%出现易损斑块;而被预测为SVO或CE的患者有10.9%出现易损斑块。此外,5名被预测患有CE的患者通过长程心电图监测被诊断为阵发性心房颤动。结论 不同的血浆代谢特征与不同的缺血性卒中亚型相关。本研究的PMML模型基于诊断良好的卒中亚型的血浆代谢特征,在识别这些亚型方面表现出较高的准确率,并且在预测隐源性卒中的病因学方面具有巨大潜力。

关键词:急性缺血性卒中;隐源性卒中;血浆代谢物;机器学习;孟德尔随机化

Plasma metabolomics combined with machine learning for classifying ischemic stroke subtypes and predicting the etiology of cryptogenic stroke

ZI Yinyu, LIU Zimiao, DENG Zhen

Department of Neurology, Nanfang Hospital, Southern Medical University, Guangzhou 510515, China

Abstract: Objective To analyze plasma metabolic profiles of different ischemic stroke subtypes and construct machine learning (ML)-based models to identify these subtypes for accurate diagnosis of cryptogenic stroke. **Methods** A total of 277 patients with acute ischemic stroke (AIS) from 3 hospitals were prospectively enrolled and stratified according to the TOAST classification criteria. Venous blood samples were collected for ¹H-NMR metabolomics detection, resulting in the identification of 46 metabolites. Metabolomics and Mendelian randomization analyses were employed to identify the potential biomarkers, and PyCaret was used to develop and optimize the ML models. A Plasma Metabolite Machine Learning (PMML) model was developed to differentiate the ischemic stroke subtypes, and the efficacy of the model was validated using a cryptogenic stroke (SUE) cohort. **Results** Significantly different metabolite profiles were observed among patients with large artery atherosclerosis (LAA), small vessel occlusion (SVO), and cardioembolism (CE). Mendelian randomization analysis indicated that glucose and proline may serve as potential biomarkers for SVO and LAA. The PMML model developed based on the metabolic profiles of LAA, SVO, and CE demonstrated a high predictive accuracy for distinguishing these ischemic stroke subtypes with an accuracy of 0.8630 and an area under the ROC curve of 0.9586. In the SUE patients (the validation cohort), 23.9% of the patients predicted to have LAA exhibited vulnerable plaques, as compared with a rate of 10.9% in those predicted to have SVO or CE. Five patients predicted to have CE were diagnosed with paroxysmal atrial fibrillation via long-term electrocardiographic monitoring. **Conclusion** Different ischemic stroke subtypes have distinct plasma metabolic profiles. The PMML model developed based on plasma metabolic features of the 3 well-diagnosed stroke subtypes demonstrates a high accuracy in identifying these subtypes with a great potential for predicting the etiology of cryptogenic stroke.

Keywords: acute ischemic stroke; cryptogenic stroke; plasma metabolites; machine learning; Mendelian randomization

缺血性脑卒中(IS)占有所有卒中的70%以上,并且具有高发病率、高残疾率、高死亡率、高复发率,是一种严

重危害人类健康和生活质量的严重疾病^[1,2]。根据TOAST分型,缺血性脑卒中可分为大动脉粥样硬化(LAA)、小动脉闭塞(SVO)、心源性栓塞(CE)、原因未明的脑卒中(SUE)和其他特定原因的脑卒中(SOE)^[3]。每种亚型都需要不同的二级预防方案,并且预后各不相同^[4]。例如,LAA的预防主要侧重于抗血小板治疗和调节血脂^[5,6],而CE则更强调抗凝治疗^[7]。卒中亚型的错误识别可能导致治疗策略不当,从而增加卒中复发的风

收稿日期:2025-12-20

基金项目:国家自然科学基金(82271305);广东省自然科学基金(2021A1515011027);广东省自然科学基金(2022A1515012525)

Supported by National Natural Science Foundation of China (82271305).

作者简介:自胤余,硕士,E-mail: yinyuzi_cn@163.com

通信作者:邓 镇,博士,副教授,硕士生导师,E-mail: dengzhencn83@smu.edu.cn

险^[8]。传统急性缺血性卒中的病因学诊断,特别是隐源性卒中的诊断需要依赖于长程心电图、经胸多普勒彩超、高分辨率MRI和其他耗时且昂贵的诊断程序^[9-11]。因此,对缺血性卒中病因的有效识别,对于优化临床工作者的有效诊疗和预防策略至关重要。

代谢组学是对参与代谢过程的小分子代谢物的研究,它直接反映了细胞表型并有助于理解中心法则的下游效应^[12]。鉴于以上特征,通过代谢组学分析可以通过研究细胞、组织和血液等生物样本,从而在发现生物标志物方面展现出了巨大的潜力^[13],包括IS^[14, 15]。先前的研究针对IS各亚型进行了代谢物分析,并取得了重大进展^[16]。例如,NT-ProBNP、肌钙蛋白和D-二聚体对CE的预测价值已得到充分证实^[17, 18],此外,一项孟德尔随机化研究确定了2种脂质和5种氨基酸,包括亚油酰甘油磷酸乙醇胺和天冬氨酸,是SVO或腔隙性梗死的潜在危险因素^[19]。此外,脂蛋白(a)[Lp(a)]已被确定为LAA发生的独立危险因素,在AIS患者的代谢组学分析中,LysoPC(22:6)在LAA组和非LAA组之间差异显著^[20, 21]。然而,虽然这些研究比较了特定IS亚型和健康个体的代谢特征,或描述单一IS亚型的代谢特征,但针对不同IS亚型患者的代谢特征仍鲜有研究。IS发生后,机体通常会因应激而引起代谢变化,这些变化是非特异性的,并且可能在不同的病因类型中表现出相似的表现,加大了临床对IS病因识别的复杂程度。因此,分析和比较不同卒中亚型的代谢特征是本研究的主要重点^[22]。

本研究分析了277例明确TOAST分型的AIS患者的血浆代谢物浓度,对各种病因亚型的血浆代谢特征进行了聚类 and 差异分析,以明确其代谢特征。随后,采用机器学习(ML)技术来检测和识别这些亚型之间代谢特征的差异。此外,本研究还开发了临床预测模型来表征急性期不同IS亚型相关的代谢物变化情况。最后,应用预测模型来分析TOAST分型为隐源性卒中的患者队列,并通过后续评估验证被预测为CE或LAA的患者。

1 资料和方法

1.1 研究对象

本研究纳入2020年1月1日~9月30日收治于中山大学附属第一医院、中山大学孙逸仙纪念医院、南方医科大学南方医院的277例急性缺血性脑卒中患者作为研究对象,并根据TOAST分型对患者进行分组。所有患者及其责任家属了解本研究,并签署书面知情同意书。该队列研究属于多中心前瞻性研究,所有诊断为急性缺血性卒中的患者均登记于在线数据库(注册号:ChiCTR1800015956)。南方医科大学南方医院医学伦

理委员会审议批准了该研究方案(伦理批号:NFE-2018-025)。

纳入标准:根据AHA/ASA急性缺血性卒中患者早期管理指南^[4]诊断为急性缺血性卒中。所有患者均接受了必要及全面检查:全血相关检查包括血常规检查、肝肾功能、凝血常规(至少禁食8h);24h心电图、超声心动图;颈动脉超声检查;主动脉弓上方的CTA或DSA。排除标准:无法接受全面检查的患者。最终,277例患者被纳入研究(图1)。

1.2 方法

1.2.1 数据收集 本研究收集并分析患者的人口统计学特征、临床信息(包括基线美国国立卫生研究院卒中量表(NIHSS)评分、高血压、房颤、糖尿病、高脂血症史、IS/短暂性脑缺血发作史、吸烟和饮酒)、实验室检查结果和住院期间血浆代谢物浓度,用于开发模型。

1.2.2 ¹H-NMR代谢组检测 本研究,所有患者的静脉血样均在卒中发作后72h内使用肝素抗凝管采集。样品在采集后1h内完成进行离心(4℃,1500 r/min,10 min),取出血浆储存于-80℃超低温冰箱。而后,通过一系列步骤制备样品并进行代谢组学分析,包括基于超滤膜的离心过滤,将滤液转移到新的标记离心管中,然后加入ACDSS试剂,混合并离心,并将上清液转移到标记的NMR管中进行NMR光谱收集。使用Chenomx NMR Suite软件(version 8.5)处理¹H-NMR游离感应衰减信号,以进行自动傅里叶变换,包括相位和基线校正。以三甲基硅烷基丙磺酸钠(DSS)为化学位移标准,采用逆卷积法调整光谱峰形(CSI)。比较分析鉴定了46种代谢物并确定了它们相应的绝对浓度,用于创建数据矩阵。

1.2.3 IS分型 颈动脉和椎动脉超声、计算机断层扫描血管造影(CTA)、数字减影血管造影(DSA)、24h动态心电图和经胸超声心动图在所有入组患者中进行常规采集。根据TOAST分型标准^[3],两位对研究设计不知情的经验丰富的神经科医生,通过审查临床信息和可用数据独立评估每位患者的卒中亚型。在意见不一的情况下,由第3位神经科医生进行独立审查以达成共识。大动脉粥样硬化型脑卒中的诊断标准:有相应颅内、外大动脉粥样硬化证据,颈动脉颅外段狭窄50%以上或斑块不稳定,皮质、小脑、脑干或皮质下梗死责任灶>1.5 cm。小动脉闭塞型脑卒中:有腔隙综合征临床特点,皮质下或脑干梗死灶<1.5 cm。患有潜在心源性栓塞或大动脉粥样硬化性卒中的患者被排除在该组之外。心源性栓塞:有二尖瓣狭窄伴/不伴心房纤颤、病态窦房结综合征、卵圆孔未闭、心脏瓣膜置换术后、心肌梗死/充血性心力衰竭、心房黏液瘤、扩张型心肌病等心源性栓塞

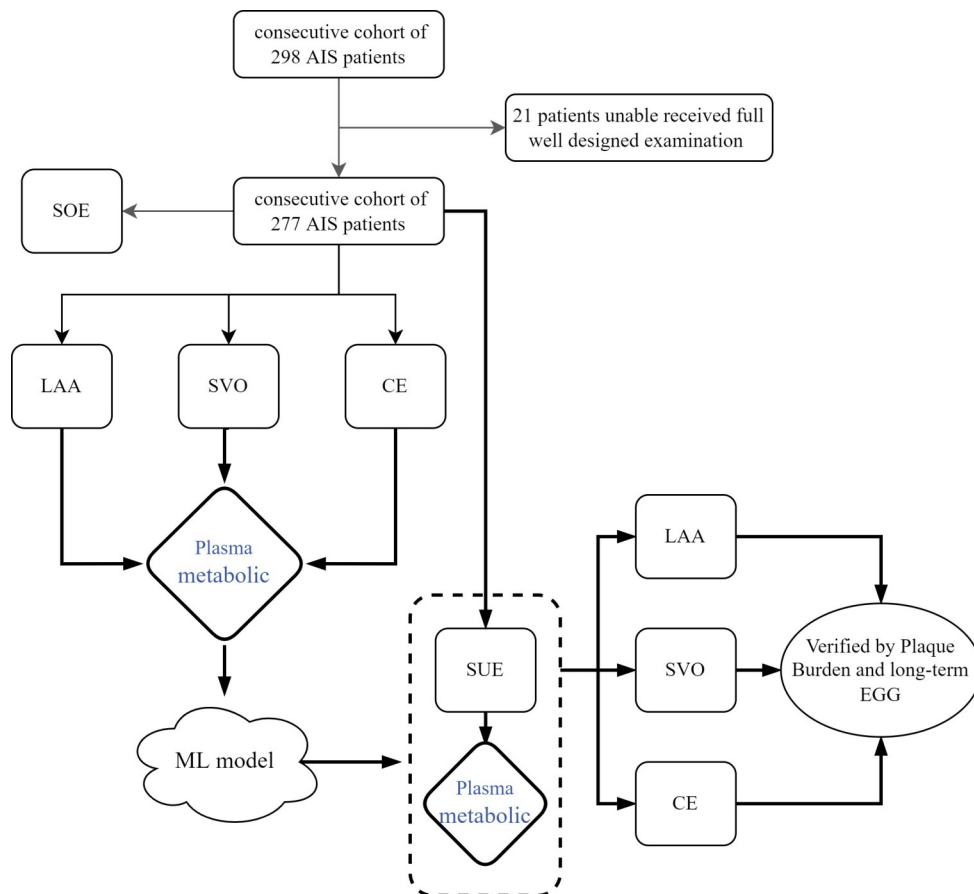


图1 本研究的流程图

Fig.1 Flowchart of the trial. AIS: Acute ischemic stroke; LAA: Large artery atherosclerosis; SVO: Small vessel occlusion; CE: Cardioembolism; ML: Machine learning.

证据,影像学存在散发、累及多支动脉供血区域,梗死责任灶 >1.5 cm,无相应颅内外大动脉粥样硬化证据并排除其他原因。原因未明的脑卒中:脑卒中的两个或两个以上根本原因;尽管进行了彻底评估,但未发现可能的原因;或由于评估不充分而无法确定确切原因。其他特定原因的脑卒中:患有其他已确定缺血性卒中原因的患者组成,例如非动脉粥样硬化性血管疾病或高凝状态(例如癌症相关)。原因未明的脑卒中的病因验证:首先,对于被归类为 SUE 的患者,均进行了颈动脉对比增强超声(CEUS)的检查。采用以下标准进一步评估 SUE 患者的斑块负荷:颈动脉 2 级新生血管;颈动脉斑块破裂;颈动脉或主动脉弓斑块表面凹陷深度在 1~2 mm(轻度溃疡性斑块);颈动脉或主动脉弓中的斑块厚度 ≥ 3 mm。此外,对于预测为 CE 的患者,进行为期两周的心电图监测。

1.2.4 机器学习 LAA 和 SVO、CE 亚型的数据用于训练 ML 算法,SUE 亚型的数据则用于验证。在对 46 种代谢物的浓度进行归一化并进行成对比较后,利用 Python 3.9.12 和 Pycaret 2.2.3 将所得代谢物比率用作 ML 过程的输入。将数据以 7:3 的比例随机分为训练数据集和测试数据集,并应用五重交叉验证进行模型推导

和内部评估。开发并验证了轻量级梯度提升机(LightGBM)、随机森林(RF)、极限梯度提升(XGBoost)、类别型特征梯度提升(CatBoost)、极端随机树(ET)、梯度提升分类器(GBC)、岭回归(Ridge)、逻辑回归(LR)、决策树(DT)、以及支持向量机(SVM)、共 10 种分类算法。

采用精确率、召回率、F1 分数、曲线下面积(AUC)、准确率和其他相关性能指标的标准定义。使用真阳性(TP)、真阴性(TN)、假阳性(FP)和假阴性(FN)来描述预测 SUE 的正确或错误分类^[23, 24]。精确率:是针对预测结果而言,是指在所有预测为正的样本中,实际为正的样本的概率。召回率:又称敏感度,衡量模型正确识别正类样本的能力,敏感度越高,假阴性的情况越少。F1 分数:是衡量准确度和召回率的加权调和平均。AUC 值:反映模型在不同阈值下敏感度和特异度的平衡,AUC 值越接近 1,表示模型的分类性能越好。准确率:衡量模型预测正确的样本占总样本的比例,是评估模型整体性能的基本指标。马修斯相关系数(MCC):用于评估二元分类模型性能的指标,针对不平衡的数据。该研究确保了数据管理每个步骤的数据质量、机密性和适当的文档记录,并及时准确地将测试数据纳入报

告。

1.3 统计学分析

所有统计分析均使用 SPSS 26.0。连续性变量表示为均数±标准差、中位数、最大值、最小值和四分位数,而分类数据表示为*n*(%)。假设检验采用双侧检验,*P*<0.05被认为差异具有统计学意义。基线数据比较在 $\alpha=0.05$ 水平上使用双尾统计检验。使用卡方检验或分类数据的 Fisher 精确检验进行组间比较。采用单样本 Kolmogorov-Smirnov 检验评估连续性数据的正态性,采用 Mann-Whitney *U* 检验比较组间连续变量。对于多因素回归分析,纳入单因素分析中 *P*<0.1 的因素,并采

用正向逐步法确定最终影响因素。卡方检验用于统计比较 LAA 组和非 LAA 组之间的斑块发生率,评估每位患者大动脉斑块分布的差异。

2 结果

2.1 患者基线资料

本研究包括 277 例 AIS 患者,患者在年龄、性别和 IS 病因方面反映了各医院收治患者的基本情况(表 1)。在研究人群中,高血压患者 142 例(51.26%),有吸烟史患者 136 例(49.10%),糖尿病患者 66 例(23.83%),卒中史患者 42 例(15.16%),冠心病患者 14 例(5.05%)。

表 1 患者基线资料

Tab.1 Baseline characteristics of the patients

Characteristic	LAA (142)	SVO (76)	SUE (27)	CE (22)	SOE (10)	Total (277)	<i>P</i>
Age (year)	64.32±13.55	62.39±12.31	65.00±14.29	70.74±19.79	59.19±20.11	64.27±14.22	NS
Gender							
Male	112 (78.87%)	66 (86.84%)	22 (81.48%)	8 (36.36%)	8 (80.00%)	216 (77.98%)	<0.001
Female	30 (21.13%)	10 (13.16%)	5 (18.52%)	14 (63.64%)	2 (20.00%)	61 (22.02%)	<0.001
NIHSS	5.22 (3.31-8.34)	5.06 (4.00-8.00)	5.00 (3.82-7.95)	10.50 (5.00-17.58)	6.28 (1.50-8.00)	5.46 (3.72-8.59)	<0.001
Medical history							
Current smoker	70 (49.30%)	32 (42.11%)	14 (51.85%)	6 (27.27%)	4 (40.00%)	126 (45.49%)	NS
Former smoker	6 (4.23%)	4 (5.26%)	0 (0.00%)	0	0 (0.00%)	10 (3.61%)	NS
Current drinker	42 (29.58%)	26 (34.21%)	4 (14.81%)	0	2 (20.00%)	74 (26.71%)	<0.001
Former drinker	8 (5.63%)	4 (5.26%)	0	0	0	12 (4.33%)	NS
Stroke	18 (12.68%)	10 (13.16%)	2 (7.41%)	12 (54.55%)	0	42 (15.16%)	<0.001
Hypertension	84 (59.15%)	42 (55.26%)	10 (37.04%)	6 (27.27%)	0	142 (51.26%)	<0.001
Diabetes	48 (33.80%)	12 (15.79%)	4 (14.81%)	0	2 (20.00%)	66 (23.83%)	<0.001
Coronary Heart Disease (CHD)	6 (4.23%)	6 (7.89%)	0	2 (9.09%)	0	14 (5.05%)	NS

The grouping is based on the TOAST typing: LAA: Large-artery atherosclerosis; CE: Cardiac embolism; SVO: Small vessel occlusion, SOE: Stroke caused by other causes; SUE: Stroke of unknown etiology. Statistical methods using SPSS 26. 0 Statistical software for statistical analysis. The metrological data age and NIHSS are first tested for normality, which is expressed as "Mean±SD" when the normal distribution is met. The patients in this study had a positive skew distribution, so they were represented by median and quartile methods (range). Count data for sex, smoking history, alcohol consumption history, stroke history, diabetes, hypertension, coronary heart disease chi-square test.

2.2 不同卒中亚型患者的血浆代谢谱存在明显差异

在各卒中亚型患者的血浆中浓度的排名前 10 的代谢物存在差异(图 2)。在 LAA 亚型中,3-羟基丁酸的浓度最高;而在 CE 亚型中,乳酸的浓度最高;在 SVO 亚型中,2-羟基异戊酸的浓度最高(图 2A~C);对于 SUE 和 SOE 亚型,排名靠前的代谢物 2-氨基丁酸表现出较高的一致性(图 2D、E)。

基于常见的 3 种 IS 亚型的代谢物热图分析(图 3),进一步验证了不同 IS 亚型患者的血浆代谢谱存在明显差异,且提示 LAA 和 SVO 的发病机制存在相似性^[25]。进而,本研究使用 CE 与非心源性栓塞患者(NCE=LAA+SVO)的数据进行生物信息学分析。富集和通路影响分析表明,CE 组呈现上调趋势的代谢物主要参

与三羧酸循环、糖酵解和发育不良相关代谢途径(图 4A、C),而 NCE 组富集的代谢物主要参与氨酰基 tRNA 合成、乙醛酸途径和二羧酸途径(图 4B、D)。

2.3 孟德尔随机化揭示了葡萄糖和脯氨酸对 SVO 和 LAA 的影响

通过拟合散点图获得线性回归参数 $\beta=0.44$, $se=0.19$ (图 5A)。森林图显示,每个 SNP 位于零的右侧,计算出的比值比(OR)为 1.55(95% *CI*: 1.08~2.26)(图 5B)。

对脯氨酸和 LAA 的发生进行了类似的分析,结果显示,当增加脯氨酸浓度被认为是暴露因素,LAA 的发生被认为是结果变量时,孟德尔随机化分析得出的 *P*<0.05。拟合线性回归参数 $\beta=1.49$, $se=0.59$, $OR=4.48$ (95% *CI*: 1.39~14.43)(图 5C、D)。

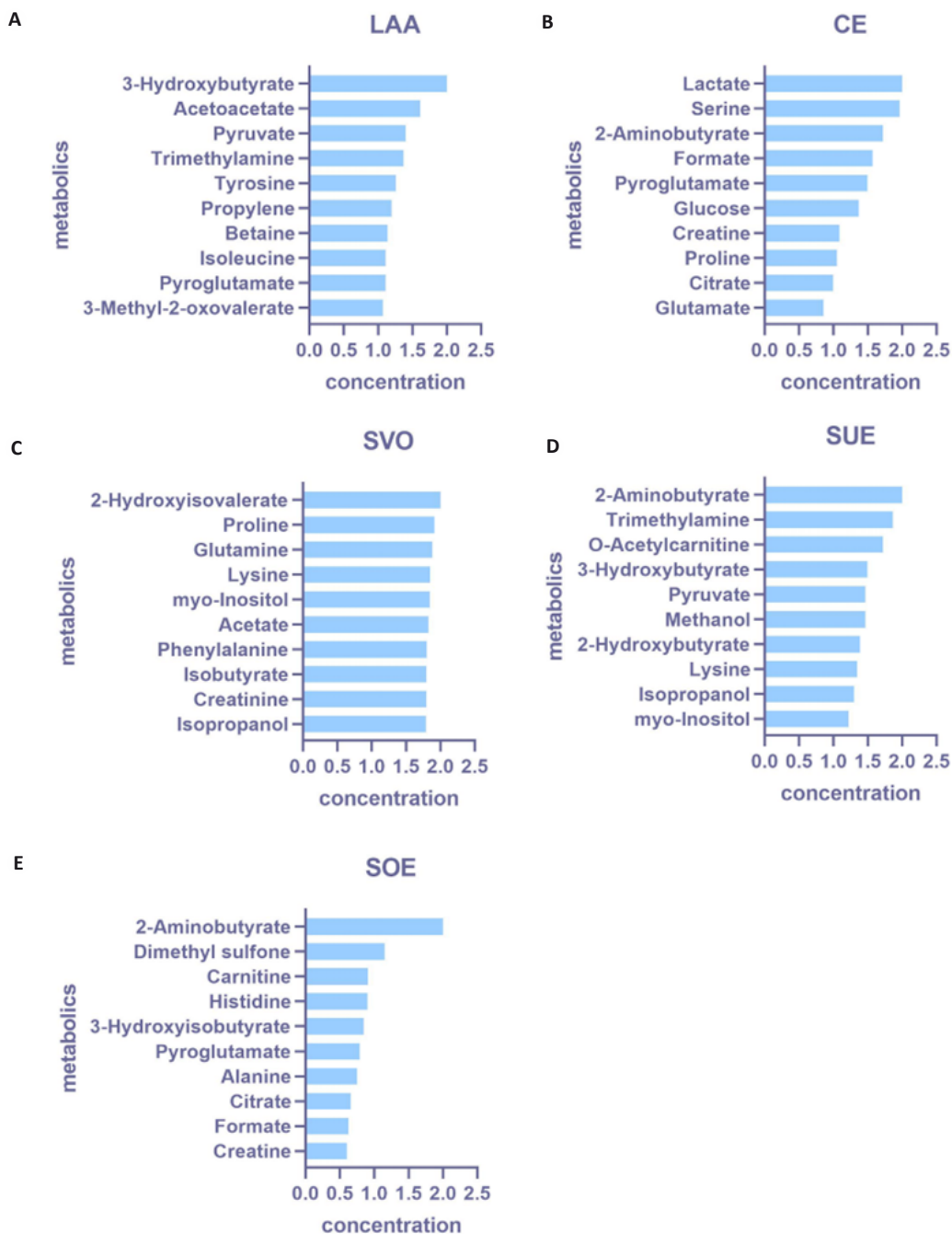


图2 在5种不同IS亚型患者血浆代谢特征中排名前10的代谢物及其相对浓度

Fig.2 The top 10 plasma metabolites in the plasma metabolic profiles of 5 different stroke subtypes and their relative concentrations. A: 3-hydroxybutyrate, acetoacetate, and pyruvate show the highest plasma concentrations in LAA subtype. B: Lactate, serine, and 2-aminobutyrate show the highest plasma concentrations in CE subtype. C: 2-hydroxyisovalerate, proline, and glutamine show the highest plasma concentrations in SVO subtype. D: 2-aminobutyrate, trimethylamine, and O-acetylcarnitine show the highest plasma concentrations in SUE subtype. E: 2-aminobutyrate, dimethyl sulfone, and carnitine show the highest plasma concentrations in SOE subtype.

2.4 ML模型根据代谢物比率准确预测卒中亚型

基于CE、LAA和SVO亚型在代谢特征上表现出明显的差异。因而,本研究对代谢物浓度进行了成对比较,并使用ML训练了各种分类算法。基于准确性、AUC、召回率、Kappa值、马修相关系数(MCC)和其他指标的综合评估对这些算法进行了排名。排名前三的算法是LightGBM、随机森林(RF)和XGBoost。其中,LightGBM准确预测CE、LAA和SVO分型,并显示出

最佳预测效果以及较高的训练和交叉验证分数。此外,其在测试数据集中也表现良好,预测准确度为0.8904, AUC为0.8911, Kappa值为0.7781。通过进一步优化LightGBM模型的参数,以增强其拟合度。最后,将LightGBM模型与RF和XGBoost相结合,使用堆叠分类器方法创建最终的血浆代谢物机器学习(PMML)模型。在测试数据中,PMML模型在CE和SVO的预测方面达到了100%的准确度,并且在LAA中也实现了

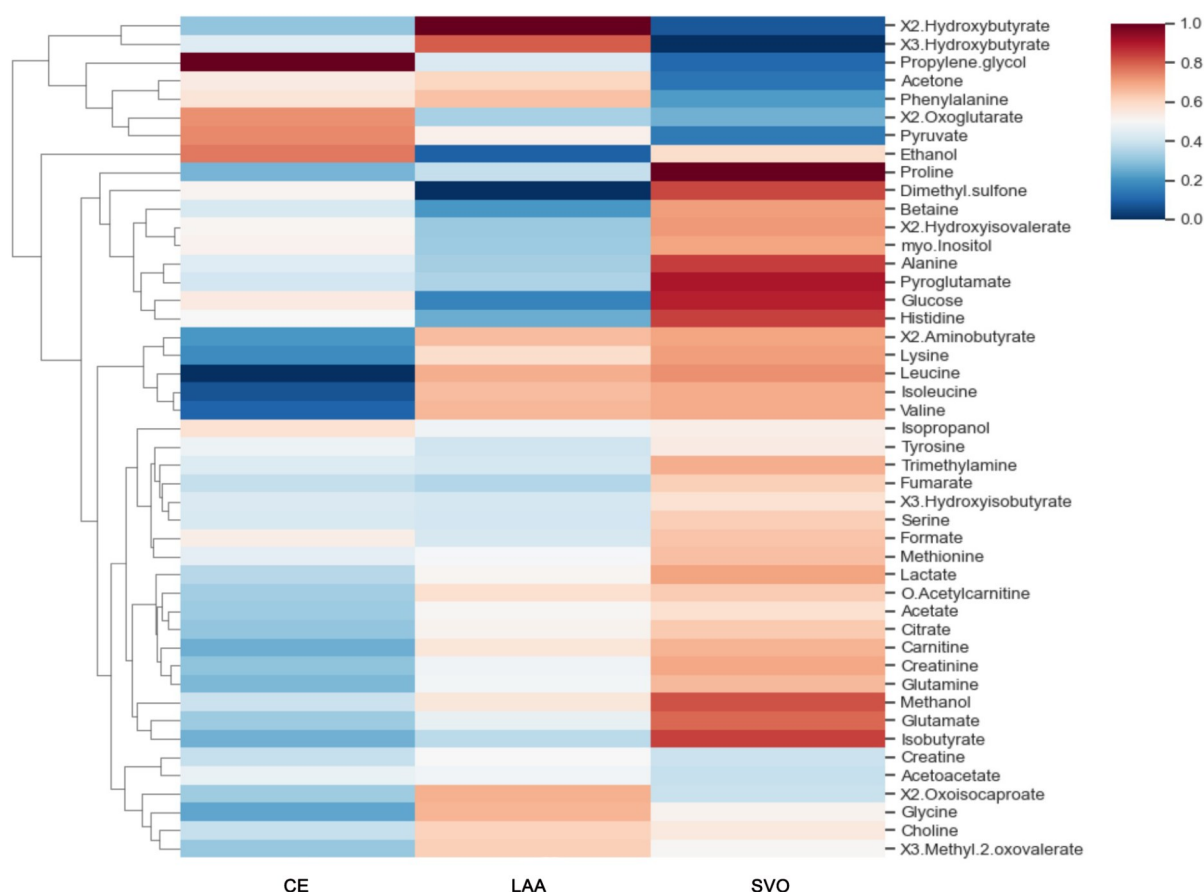


图3 CE、LAA、SVO患者的代谢物热图提示不同IS亚型患者的血浆代谢谱存在明显差异
Fig.3 Distinct metabolic profiles of CE, LAA and SVO patients.

84.90%的准确度(图6A),且该模型的AUC为0.96(图6B)。对堆叠分类器模型的深入分析表明,乙酸盐/丝氨酸和丙酮/肌酸比值对模型的影响最为显著(图6C)。利用排名前10的代谢物对,研究团队开发了一种高性能的堆叠分类器模型(图6D)。

2.5 基于SUE数据的PMML模型的探索性验证

SUE数据集由27名诊断为SUE患者的数据组成,用于探索性的验证该模型的病因预测分布准确率,以反映模型的临床预测价值。ML模型预测了5例CE患者(18.52%),11例LAA患者(40.74%),11例SVO患者(40.74%)。与非LAA组相比,预测为LAA的患者表现出更高的斑块负荷(图7A)。使用为期两周的长程心电图监测进一步评估预测患有CE的患者,在5名预测为CE患者的长程心电图中发现阵发性心房颤动(图7B)。

3 讨论

本研究基于AIS患者的代谢谱,通过识别3种卒中亚型的不同代谢物谱,开发机器学习模型来预测SUE的病因。该模型准确区分了3种主要的IS亚型,并识别了SUE的潜在原因,能更好的帮助临床医生定制个

性化检查和治疗策略。

上述研究结果表明,每种IS亚型都表现出独特的代谢组学特征,反映了其潜在病理生理机制的差异。以上结果提示,通过分析患者的代谢特征,可以准确预测IS的病因,从而有助于实现有效诊疗,提高临床效率,改善IS患者的治疗效果并优化卒中管理。此外,本研究的机器学习模型由血浆代谢谱开发而成,无需进行耗时且较为昂贵的传统检查,以期能够提升诊疗效率,并减轻患者和医疗保健系统的负担。

本研究分析显示,乙酸盐/丝氨酸比值以及丙酮/肌酸比值是本模型中最重要的两对预测因子。这些代谢物在能量代谢途径中起着至关重要的作用,包括糖异生、三羧酸循环、酮体合成和肌酸-磷酸肌酸转化。丝氨酸合成途径绕过糖酵解,涉及将3-磷酸甘油酸转化为丝氨酸,而乙酸盐是由丙酮酸分解产生的。本研究的结果表明,在IS条件下,细胞下调丝氨酸合成,同时上调糖酵解,突出了IS后发生的代谢适应。丙酮是一种酮体,是为了响应缺氧而合成的,它刺激脂肪动员和酮体的产生,以满足IS患者增加的能量需求^[26, 27]。肌酸水平的降低,表明肌酸激酶促进磷酸肌酸转化为肌酸,导致ATP产生^[28]。丙酮/肌酸比值升高与乙酸/

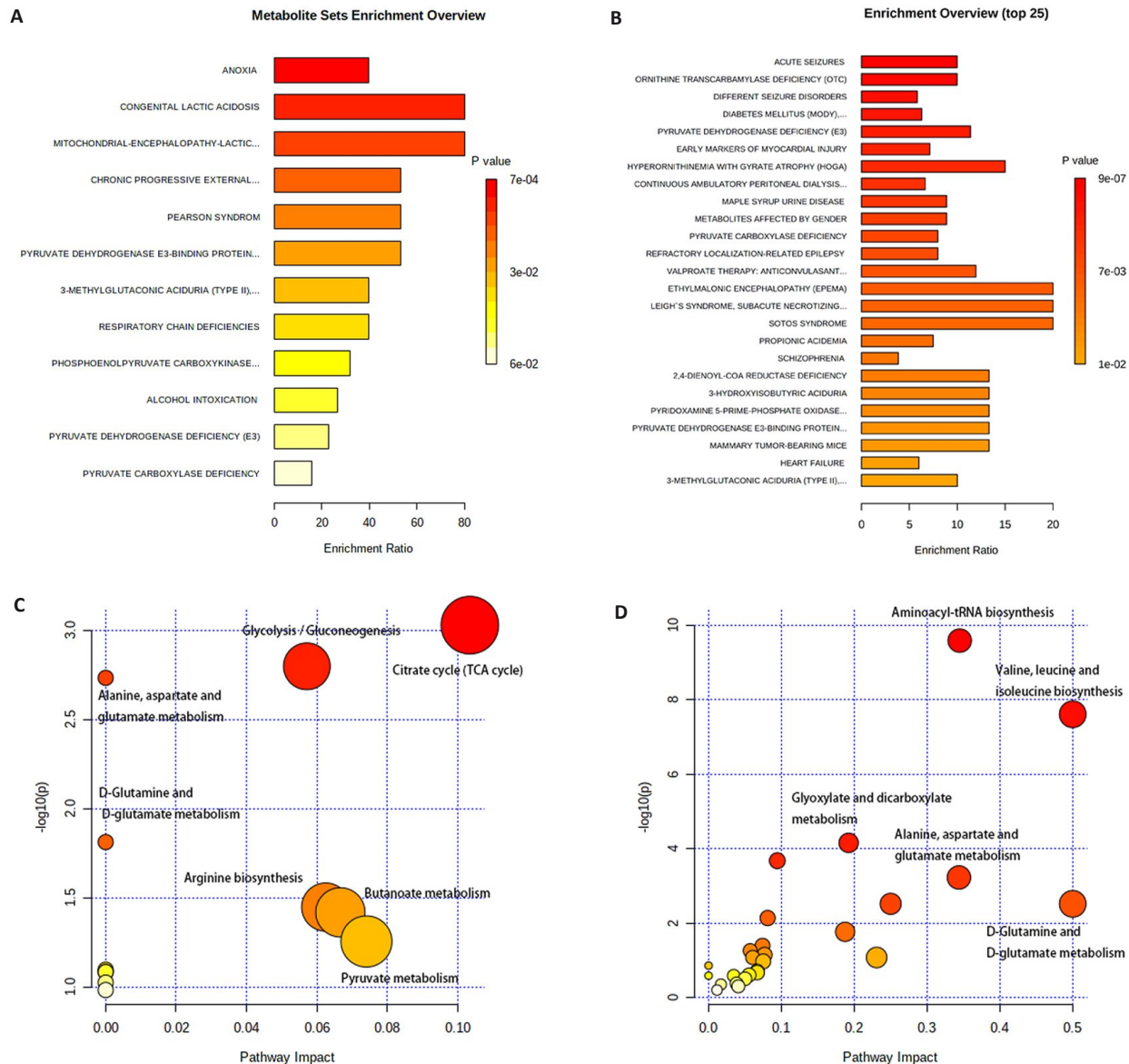


图4 CE患者和NCE患者的富集和通路分析差异

Fig.4 Differential enrichment and pathway analysis between CE and NCE patients. **A, C:** Metabolites with an upregulation trend in the CE group are primarily enriched in the Krebs cycle, glycolysis, and dysplasia-related metabolic pathways. **B, D:** Metabolites enriched in the NCE group are primarily involved in aminoacyl-tRNA biosynthesis, the glyoxylate pathway, and the dicarboxylate pathway. NCE=LAA+SVO.

丝氨酸比值增加类似,表明IS后患者通过上调无氧代谢途径(例如糖酵解和肌酸-磷酸肌酸转化),而不是增加需氧代谢来增强ATP的产生。这种适应性机制有可能减轻活性氧造成的损害,进一步强调了身体对缺血应激的代谢反应。

本研究的孟德尔随机化分析表明,葡萄糖水平与SVO的发生,以及脯氨酸水平与LAA的发生之间存在潜在的因果关系。已知血糖水平升高会损害内皮细胞^[29],特别是在小动脉和小静脉中,导致内皮损伤,血管壁硬化,并最终在这些血管供应的组织中缺氧^[30, 31]。葡萄糖已被公认为SVO的独立危险因素。此外,研究表明^[32],高浓度葡萄糖能独立于胰岛素,直接激活相关转

录因子,并显著上调糖酵解限速酶及整个糖酵解通路的基因表达,并同步促进酮体生成通路的代偿性上调,而通过ML模型筛选出的“丙酮/肌酸”、“乙酸盐/丝氨酸”代谢物对,也进一步反映出该通路扰动的核心表型。尽管对脯氨酸和LAA的研究有限,但研究表明,脯氨酸构成了近25%的胶原蛋白^[33],胶原蛋白是血管壁的关键组成部分。在慢性炎症(如动脉粥样硬化)导致血管壁损伤的情况下,储存在胶原蛋白中的脯氨酸可能会释放到血液中^[34]。但是脯氨酸在LAA发病机制中的具体作用,仍有待进一步的研究。

SUE是异质性的,由多种原因引起,此外,SUE的病因诊断通常依赖于长程心电图、经胸多普勒超声、高

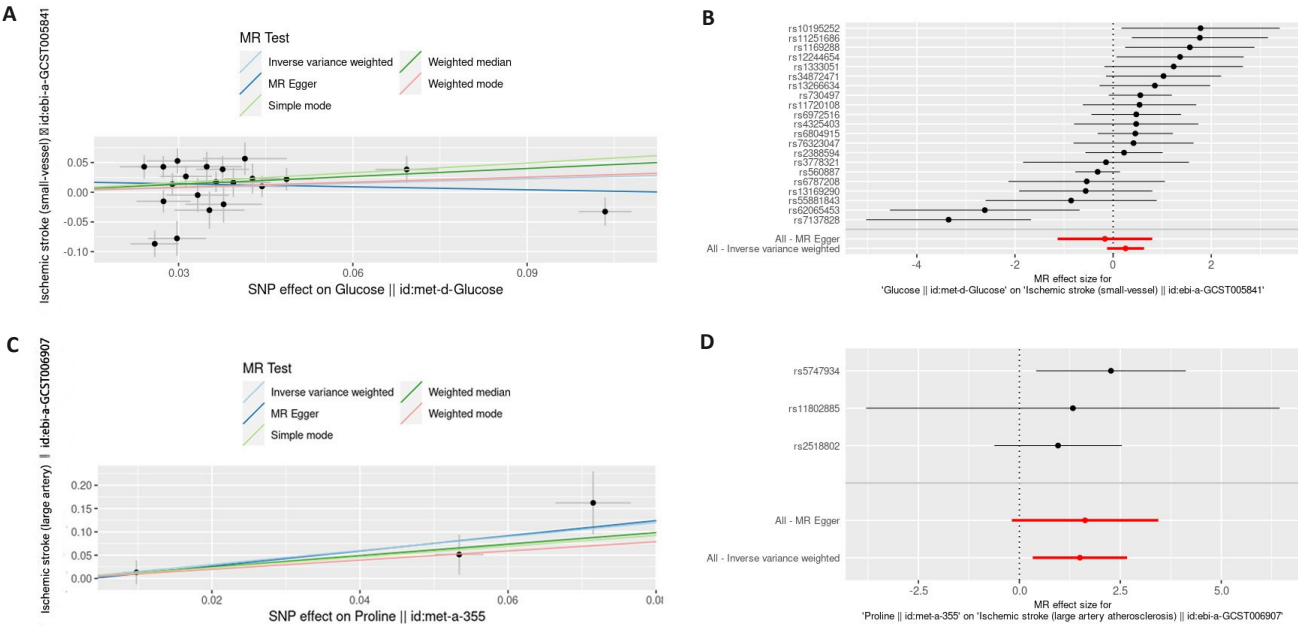


图5 孟德尔随机化揭示了葡萄糖对 SVO和脯氨酸对LAA 的影响
Fig.5 Mendelian randomization analysis of the effects of glucose on SVO and proline on LAA. A, B: Mendelian randomization analysis of glucose and SVO: regression lines and forest plots showing weighted median. C, D: Mendelian randomization analysis of proline and LAA: regression lines and forest plots showing weighted median.

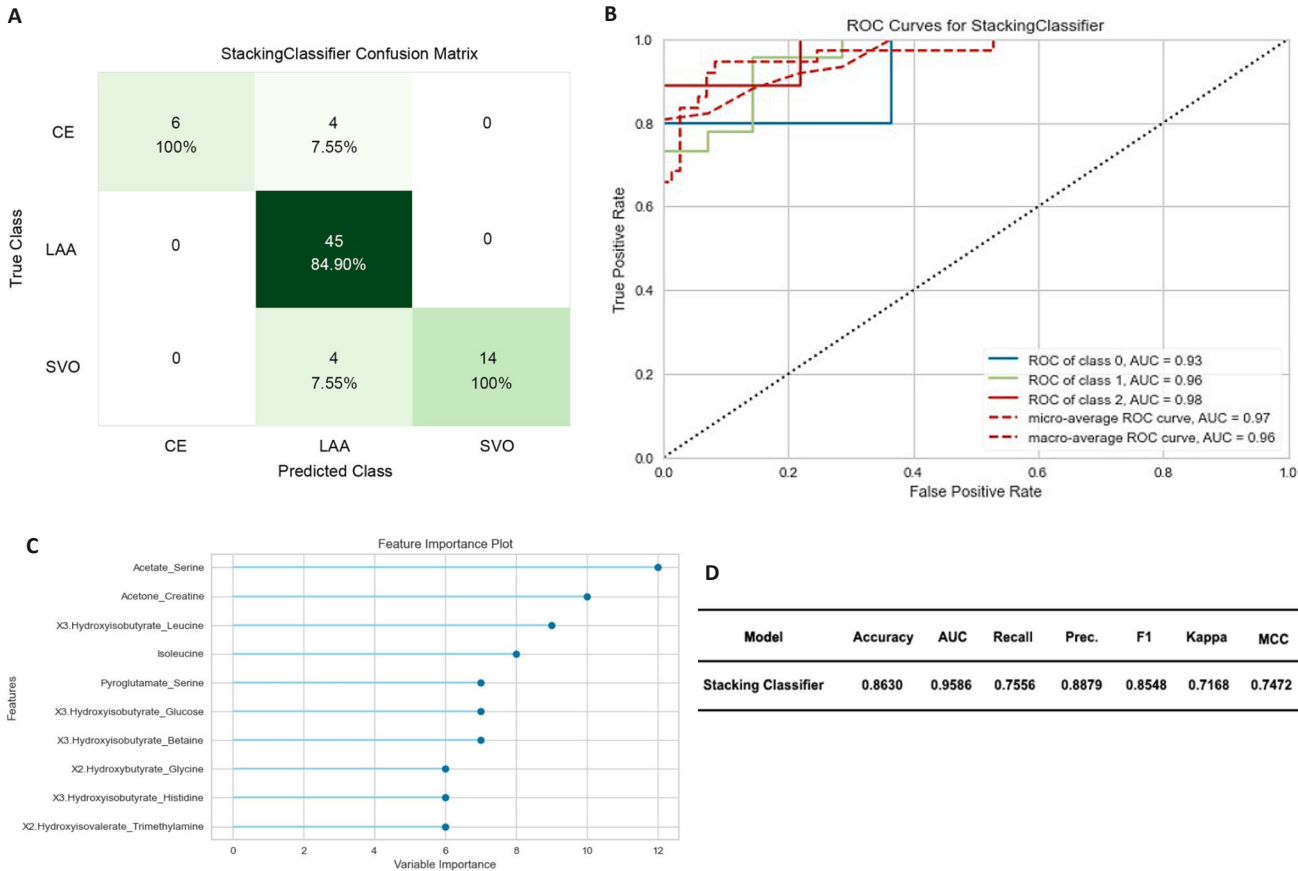


图6 血浆代谢物机器学习(PMML)模型及其表现
Fig. 6 The Plasma Metabolite Machine Learning (PMML) model and its predictive performance. A: Confusion matrix of the Stacking Classifier model. B: ROC curve of the Stacking Classifier model. C: Top 10 factors ranked based on their impact on the Stacking Classifier. D: Parameters indicating the prediction accuracy of the Stacking Classifier model.

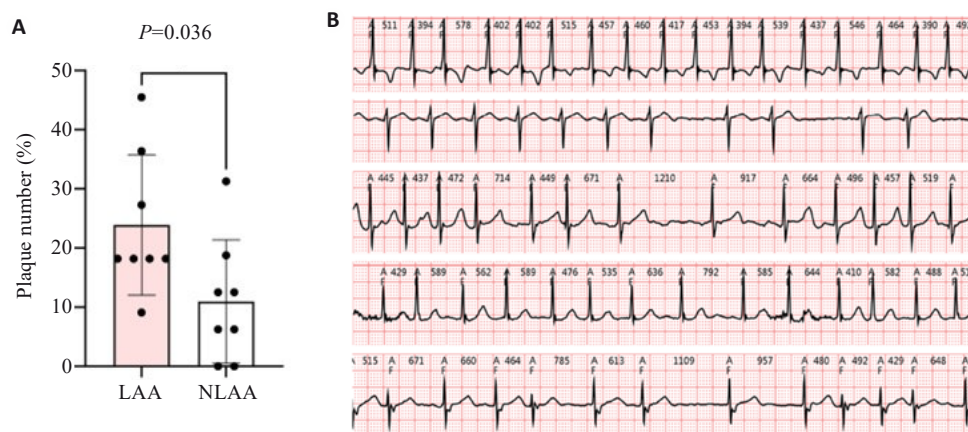


图7 ML模型用于预测SUE患者的表现

Fig.7 Validation of the PMML Model in SUE patients. A: A significantly higher plaque burden was observed in patients predicted to have LAA than in the non-LAA group. B: Five patients predicted to have CE had paroxysmal atrial fibrillation episodes captured by long-range ECG.

分辨率MRI和其他昂贵且耗时的检查。近来,许多研究旨在识别新的生物标志物并开发用于识别和诊断SUE的临床预测模型。例如,一项研究开发了一种ML分类器,使用患者人口统计学、合并症、生命体征、实验室结果和超声心动图结果来预测SUE的病因,在预测心源性栓塞方面实现了0.85的AUC和44%的准确率^[25];另一项研究利用ML技术对电子病历进行自然语言处理,不同模型拟合之间的Kappa值在0.4~0.8,SUE队列的Kappa值为0.4752^[35]。相较于上述研究,本模型的AUC和准确率优于上述模型,Kappa值为0.6~0.8,表明模型具有优越的一致性。流行病学数据表明,不明原因的栓塞,特别是源自颈动脉或主动脉弓中不稳定性动脉粥样硬化斑块的动脉栓塞,是导致SUE的主要潜在因素。此外,斑块形态的计数结果与本ML模型区分疑似LAA患者和非LAA患者的结果显示出较好的一致性。在所有纳入的IS患者中,CE约占这些病例的20%,同时,在本模型对SUE的预测中,CE病例的比例为18.5%,与真实世界的数据非常吻合。此外,在样本量有限的后续研究中,5名预计患有CE的患者,其后续的长程心电图结果与阵发性心房颤动(PAF)一致,这一发现表明PAF可能是他们发生IS的基础,支持了该研究的最初的预测。这些结果可能有助于临床医生选择适当的二级预防策略。

本研究存在的局限性:首先,样本量有限、亚型分布不均衡作为本研究的核心局限性,期待后续的多中心大样本研究,以进一步验证该模型的普遍性。其次,仅纳入中国患者,限制了本研究结果对其他种族的适用性。因此,未来的研究应该纳入更多样化的人群,在进一步缩小采血窗的同时更严格控制混杂因素,进一步提升结果的可靠性,并将其扩展到临床实践之中。

Declaration of interests: The authors declare no competing interests.

参考文献:

- [1] Wang HD, Liddell CA, Coates MM, et al. Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013[J]. Lancet, 2014, 384(9947): 957-79.
- [2] Campbell BCV, De Silva DA, MacLeod MR, et al. Ischaemic stroke[J]. Nat Rev Dis Primers, 2019, 5: 70.
- [3] Adams HP Jr, Bendixen BH, Kappelle LJ, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment[J]. Stroke, 1993, 24(1): 35-41.
- [4] Powers WJ, Rabinstein AA, Ackerson T, et al. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: a guideline for healthcare professionals from the American heart association/American stroke association[J]. Stroke, 2019, 50(12): e344-418.
- [5] Kamarova M, Baig S, Patel H, et al. Antiplatelet use in ischemic stroke[J]. Ann Pharmacother, 2022, 56(10): 1159-73.
- [6] Diener HC, Hankey GJ. Primary and secondary prevention of ischemic stroke and cerebral hemorrhage JACC focus seminar[J]. J Am Coll Cardiol, 2020, 75(15): 1804-18.
- [7] Benz AP, Hohnloser SH, Eikelboom JW, et al. Outcomes of patients with atrial fibrillation and ischemic stroke while on oral anticoagulation[J]. Eur Heart J, 2023, 44(20): 1807-14.
- [8] Kolmos M, Christoffersen L, Kruuse C. Recurrent ischemic stroke-a systematic review and meta-analysis[J]. J Stroke Cerebrovasc Dis, 2021, 30(8): 105935.
- [9] Ritter MA, Kochhäuser S, Duning T, et al. Occult atrial fibrillation in cryptogenic stroke: detection by 7-day electrocardiogram versus implantable cardiac monitors[J]. Stroke, 2013, 44(5): 1449-52.
- [10] Seet RCS, Friedman PA, Rabinstein AA. Prolonged rhythm monitoring for the detection of occult paroxysmal atrial fibrillation

- in ischemic stroke of unknown cause[J]. *Circulation*, 2011, 124(4): 477-86.
- [11] Mendelson SJ, Prabhakaran S. Diagnosis and management of transient ischemic attack and acute ischemic stroke: a review[J]. *JAMA*, 2021, 325(11): 1088-98.
- [12] Li WT, Shao CY, Li C, et al. Metabolomics: a useful tool for ischemic stroke research[J]. *J Pharm Anal*, 2023, 13(9): 968-83.
- [13] German JB, Hammock BD, Watkins SM. Metabolomics: building on a century of biochemistry to guide human health[J]. *Metabolomics*, 2005, 1(1): 3-9.
- [14] Jové M, Mauri-Capdevila G, Suárez I, et al. Metabolomics predicts stroke recurrence after transient ischemic attack[J]. *Neurology*, 2015, 84(1): 36-45.
- [15] Sun JX, Xiao CX, Zhang JY, et al. Advances in metabolomics of biomarkers for ischemic stroke: From bench to clinic[J]. *Neural Regen Res*, 2026, 21(8): 3588-97.
- [16] Sun DK, Tiedt S, Yu B, et al. A prospective study of serum metabolites and risk of ischemic stroke[J]. *Neurology*, 2019, 92(16): e1890-8.
- [17] Isenegger J, Meier N, Lämmle B, et al. D-dimers predict stroke subtype when assessed early[J]. *Cerebrovasc Dis*, 2009, 29(1): 82-6.
- [18] Gąsiorek P, Sakowicz A, Banach M, et al. Arterial stiffness and indices of left ventricular diastolic dysfunction in patients with embolic stroke of undetermined etiology[J]. *Dis Markers*, 2019, 2019: 9636197.
- [19] Guo MN, Hao XY, Tian J, et al. Human blood metabolites and lacunar stroke: a Mendelian randomization study[J]. *Int J Stroke*, 2023, 18(1): 109-16.
- [20] Kumar P, Swarnkar P, Misra S, et al. Lipoprotein (a) level as a risk factor for stroke and its subtype: a systematic review and meta-analysis[J]. *Sci Rep*, 2021, 11: 15660.
- [21] Arnold M, Schweizer J, Nakas CT, et al. Lipoprotein(a) is associated with large artery atherosclerosis stroke aetiology and stroke recurrence among patients below the age of 60 years: results from the BIOSIGNAL study[J]. *Eur Heart J*, 2021, 42(22): 2186-96.
- [22] Baranovicova E, Kalenska D, Kaplan P, et al. Blood and brain metabolites after cerebral ischemia[J]. *Int J Mol Sci*, 2023, 24(24): 17302.
- [23] Heo J, Yoon JG, Park H, et al. Machine learning-based model for prediction of outcomes in acute stroke[J]. *Stroke*, 2019, 50(5): 1263-5.
- [24] Wang JN, Gong XX, Chen HF, et al. Causative classification of ischemic stroke by the machine learning algorithm random forests[J]. *Front Aging Neurosci*, 2022, 14: 788637.
- [25] Kamel H, Navi BB, Parikh NS, et al. Machine learning prediction of stroke mechanism in embolic strokes of undetermined source[J]. *Stroke*, 2020, 51(9): e203-10.
- [26] Pereira S, Kildegaard HF, Andersen MR. Impact of CHO metabolism on cell growth and protein production: an overview of toxic and inhibiting metabolites and nutrients[J]. *Biotechnol J*, 2018, 13(3): 1700499.
- [27] Kolb H, Kempf K, Röhling M, et al. Ketone bodies: from enemy to friend and guardian angel[J]. *BMC Med*, 2021, 19(1): 313.
- [28] Balestrino M, Sarocchi M, Adriano E, et al. Potential of creatine or phosphocreatine supplementation in cerebrovascular disease and in ischemic heart disease[J]. *Amino Acids*, 2016, 48(8): 1955-67.
- [29] You R, McNeil JJ, O'Malley HM, et al. Risk factors for lacunar infarction syndromes[J]. *Neurology*, 1995, 45(8): 1483-7.
- [30] Horowitz DR, Tuhim S, Weinberger JM, et al. Mechanisms in lacunar infarction[J]. *Stroke*, 1992, 23(3): 325-7.
- [31] Caplan LR. Lacunar infarction and small vessel disease: pathology and pathophysiology[J]. *J Stroke*, 2015, 17(1): 2-6.
- [32] Kim MS, Krawczyk SA, Doridot L, et al. ChREBP regulates fructose-induced glucose production independently of insulin signaling[J]. *J Clin Investig*, 2016, 126(11): 4372-86.
- [33] Goulart VAM, Sena MM, Mendes TO, et al. Amino acid biosignature in plasma among ischemic stroke subtypes[J]. *BioMed Res Int*, 2019, 2019(1): 8480468.
- [34] Hermann DM, Matter CM. Tissue plasminogen activator – induced reperfusion injury after stroke revisited[J]. *Circulation*, 2007, 116(4): 363-5.
- [35] Garg R, Oh E, Naidech A, et al. Automating ischemic stroke subtype classification using machine learning and natural language processing[J]. *J Stroke Cerebrovasc Dis*, 2019, 28(7): 2045-51.

(编辑:经 媛)