

文章编号: 1671-7449(2026)03-0327-08

J. Test Measur. Technol., 2026, 40(3): 327-334.

用于GC-IMS的呼气分析数据处理方法

马睿^{1,2}, 林建华^{1,2}, 慕世龙^{1,2}, 徐陈^{1,2}, 贾建¹, 何秀丽¹, 高晓光^{1*}

(1. 中国科学院空天信息创新研究院 传感器技术全国重点实验室, 北京 100190;

2. 中国科学院大学 电子电气与通信工程学院, 北京 100049)

摘要: 针对呼气分析中气相色谱-离子迁移谱(Gas Chromatography-Ion Mobility Spectrometry, GC-IMS)谱峰重叠和实际应用场景下样本有限、需持续更新等问题, 提出一种呼气分析数据处理方法。以健康志愿者饮用咖啡前后模拟不同生理状态, 利用GC-IMS直接分析呼出气体, 通过高斯二阶导数谱峰锐化算法解析GC-IMS呼气指纹图中的重叠峰, 构建蒙德里安森林(Mondrian forest, MF)增量学习模型实现分类。结果表明, 该方法成功解析了GC-IMS原始指纹图中的重叠峰, 将峰值信噪比提高至50 dB; 提取GC-IMS指纹图中解析出的谱峰位置及强度作为特征, 构建的MF增量分类模型随着样本增加, 平均分类准确率达到93.75%, 显著优于对比的随机森林和Hoeffding树模型。这种呼气分析数据处理方法提高了分类的准确性, 具有良好的实际应用前景。

关键词: 气相色谱-离子迁移谱; 呼气分析; 重叠峰解析; 蒙德里安森林

中图分类号: TP391

文献标识码: A

DOI: 10.62756/csjs.1671-7449.2026049

引用格式: 马睿, 林建华, 慕世龙, 等. 用于GC-IMS的呼气分析数据处理方法[J]. 测试技术学报, 2026, 40(3): 327-334.

MA Rui, LIN Jianhua, MU Shilong, et al. Data processing method for GC-IMS exhaled breath analysis[J]. Journal of Test and Measurement Technology, 2026, 40(3): 327-334.

Data Processing Method for GC-IMS Exhaled Breath Analysis

MA Rui^{1,2}, LIN Jianhua^{1,2}, MU Shilong^{1,2}, XU Chen^{1,2}, JIA Jian¹, HE Xiuli¹, GAO Xiaoguang^{1*}

(1. State Key Laboratory of Transducer Technology, Aerospace Information Research Institute,

Chinese Academy of Sciences, Beijing 100190, China;

2. School of Electronic, Electrical and Communication Engineering,

University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: A novel data processing method of gas chromatography-ion mobility spectrometry (GC-IMS) is proposed to address the challenges of peak overlapping in GC-IMS spectra for exhaled breath analysis and the limited and continuously updated samples in applications. Breath samples simulating different physiological states were collected from healthy volunteers before and after drinking coffee, and were analyzed by GC-IMS directly. Overlapping peaks in the GC-IMS two-dimensional spectra were resolved using a Gaussian second derivative peak sharpening algorithm, and a Mondrian forest (MF) incremental learning model was constructed for classification of the two states. The results indicate that the method

【收稿日期】2025-04-29

【基金项目】国家自然科学基金资助项目(62031022, 61871364)

【作者简介】马睿(1997—), 男, 硕士生, 主要从事呼气检测技术研究。E-mail: marui221@mails.ucas.ac.cn。

*【通信作者】高晓光(1976—), 男, 研究员, 博士, 主要从事离子迁移谱技术研究。E-mail: xggao@mail.ie.ac.cn。

successfully resolves overlapping peaks in the original GC-IMS spectra, increasing the peak signal-to-noise ratio (PSNR) to 50 dB. Features such as the positions and intensities of the resolved peaks in the GC-IMS spectra are extracted to build the MF incremental learning model, which maintain a high average classification accuracy of 93.75% as samples increase, significantly outperforming comparative models like random forest and Hoeffding trees. This exhaled breath analysis data processing method enhances classification accuracy, showing promising practical application prospects.

Key words: gas chromatography-ion mobility spectrometry; exhaled breath analysis; overlapping peaks resolution; Mondrian forest

0 引言

呼出气体中含有多种挥发性有机物(Volatile Organic Compounds, VOCs),其种类及含量能够反映人体的生理状态。与血液、生物活检和胸透等传统诊断手段相比,呼出气体检测具有无创、无痛和操作简便等优势,适用于疾病的早期筛查与实时监测。目前呼出气体检测方面的研究主要分为两类^[1],一类是针对已知疾病生物标志物的定量检测,另一类是基于呼气中多种痕量 VOCs 的整体特征构建呼气指纹图,再利用模式识别方法进行生理状态分析,并实现疾病预警。后者在缺乏明确的疾病标志物信息情况下已实现对肺癌、慢性阻塞性肺病(Chronic Obstructive Pulmonary Disease, COPD)等多种疾病的准确判别^[2-4],显示出其在临床应用中的广阔前景。目前,获取呼气指纹图的方法主要有气相色谱-质谱联用(Gas Chromatography-Mass Spectrometry, GC-MS)^[5]、电子鼻^[6]和离子迁移谱(Ion Mobility Spectrometry, IMS)^[7]等。

气相色谱-离子迁移谱(Gas Chromatography-Ion Mobility Spectrometry, GC-IMS)技术结合了气相色谱的高效分离与离子迁移谱的灵敏检测优势,具有系统体积小、检测成本低的特点,适合构建呼气指纹图。GC-IMS可以在较短的时间内完成对呼出气体中多种痕量 VOCs 的同时检测,其谱图中特征峰的位置和强度分别表征对应组分种类与含量,能有效支持样本特征向量构建与模式识别分析。然而,由于呼出气体中存在大量结构相似的挥发性有机物,谱图中谱峰分辨率不高,严重影响谱峰识别与特征提取的准确性。如何实现谱图中重叠峰的有效解析已成为当前呼气分析研究的重要技术难题。

呼气检测的另一个重要问题是临床样本数量

有限及检测环境差异较大,这对传统模式识别方法提出了挑战。当代医学诊断中,许多模式识别方法依赖彩超或X光等病理图像进行训练与优化^[8-9],在数据规模和质量上均有较强保障。相比之下,呼气指纹图易受环境影响,基于传统机器学习的模式识别方法在样本量较小时容易产生过拟合,且在新增样本时往往需要重新训练模型,计算开销较大。

本文以饮用咖啡前后的呼出气体分别模拟正常人群与患者的呼气气体,提出一种结合重叠峰解析与增量学习的呼气检测数据处理方法,解决实际GC-IMS呼气指纹图中的重叠峰解析困难问题,同时构建增量学习分类模型解决因样本量较小导致的模型泛化能力不足问题。

1 实验方案

1.1 仪器设备

本研究采用自建的GC-IMS系统对呼出气体进行检测。该系统如图1所示,主要由呼气采样装置、气相色谱预分离装置和离子迁移谱检测装置组成。人体呼出气体由死腔气和肺泡气组成^[10],其中,只有肺泡气与人体血液进行了气体交换,肺泡气中可反映人体生理特征的VOCs浓度约为混合呼出气体的3倍,故本文选择肺泡气作为检测样本。

呼气采样装置由CO₂传感器、六通阀、抽气泵和1 mL定量环组成。在采集呼出气体时,使用CO₂传感器监测呼出气体的CO₂浓度,进而区分死腔气与肺泡气,并通过抽气泵将肺泡气采样至六通阀定量环中。为避免呼出气体中的水汽在采样管路中凝结,所有管路均维持40℃的恒温。

GC预分离装置主要由DB-5毛细色谱柱(Agilent Technologies, USA)构成,色谱柱长15 m,内径0.53 mm,涂层厚1 μm。

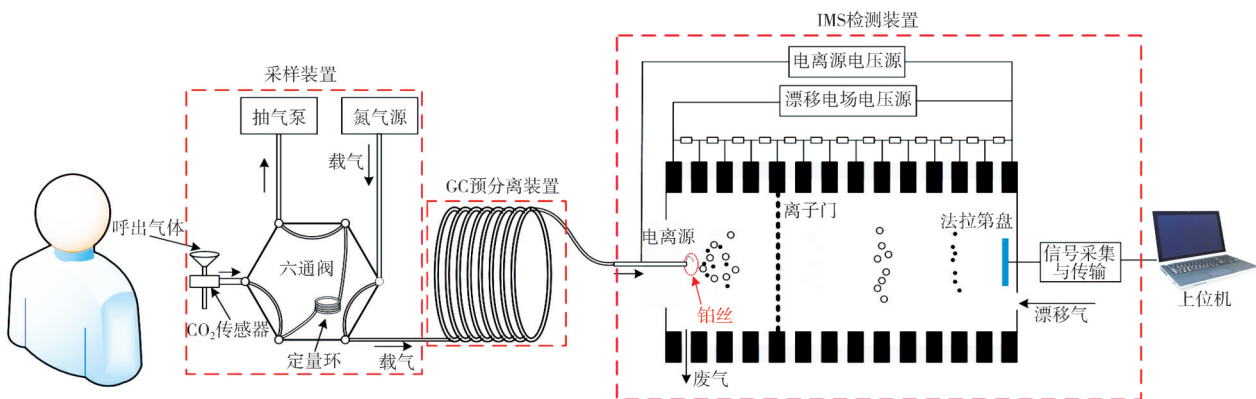


图 1 GC-IMS 呼气分析系统原理图

Fig. 1 Schematic of GC-IMS exhaled breath analysis system

IMS 检测装置由电晕放电电离源、漂移管、外部电路、信号采集与处理单元等组成^[11]。肺泡气中的不同 VOCs 经 GC 分离后在不同时刻依次进入 IMS 中, 得到随 GC 保留时间变化的一系列 IMS 谱图, 构成如图 2 所示的 GC-IMS 谱图。IMS 漂移气和 GC 载气均为高纯氮气(纯度 99.999%), GC-IMS 系统相关参数如表 1 所示。

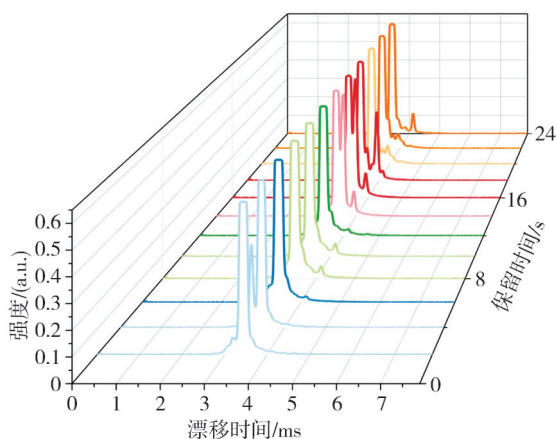


图 2 GC-IMS 谱图

Fig. 2 GC-IMS spectra

表 1 GC-IMS 呼气分析系统参数

Tab. 1 Parameters of GC-IMS exhaled breath analysis system

仪器参数	数值
GC 温度/℃	40
IMS 漂移管温度/℃	100
IMS 电场强度/(V·cm ⁻¹)	380
载气流量/(mL·min ⁻¹)	10
漂移气流量/(mL·min ⁻¹)	200

1.2 呼气指纹图采集

选取 7 名健康志愿者参与实验, 所有参与者均签署知情同意书。为模拟不同的呼气样本, 分别采集每位志愿者在饮用咖啡前后的呼出气体。

采样前受试者漱口并静坐 30 s 以清除口腔残留物及缓解运动影响。首先, 采集志愿者在未饮用咖啡时的呼出气体, 作为健康人群呼气模拟样本; 随后, 每位志愿者饮用 100 mL 预先冲泡好的蓝山风味咖啡(福建苏卡食品公司生产, 300 mL 沸水中加入 15 g 咖啡粉冲泡), 20 min 后采集其呼出气体作为患者呼气模拟样本。

在正式采样前先启动抽气泵, 用洁净空气冲洗采样管路和定量环, 并获取洁净空气的 IMS 谱图以确认系统正常。采样时, 受试者通过一次性吹嘴缓慢呼气至采样装置, 在整个呼气过程中利用 CO₂ 传感器实时监测呼出气体中 CO₂ 分压, 当达到阈值时(指示肺泡气开始排出), 采样系统自动开启抽气泵收集肺泡气到定量环。呼气结束后, 六通阀切换至进样模式, 将定量环中的肺泡气注入 GC 进行预分离后由 IMS 检测。每次检测完成后, 启动抽气泵抽取洁净空气冲洗定量环和采样管路, 以避免残留干扰。

将受试者未饮用咖啡时的呼出气体样本标记为“0 类样本”, 将饮用咖啡后的呼出气体样本标记为“1 类样本”, 分别对应标签 0 和标签 1。每组均包含 35 个呼气样本。

1.3 呼气分析数据处理方法

呼气谱图数据处理流程如图 3 所示, 包括数据预处理、特征提取和模式识别 3 个阶段。

在数据预处理阶段对原始 GC-IMS 谱图进行去噪和平滑处理, 以降低环境背景对信号的干扰。利用气相色谱分离原理, 提取 GC 载气信号的 IMS 谱图作为背景信号, 然后将该背景信号从原始谱图中以差分操作剔除, 从而消除 GC-IMS 装置参数漂移和环境干扰气体的影响。

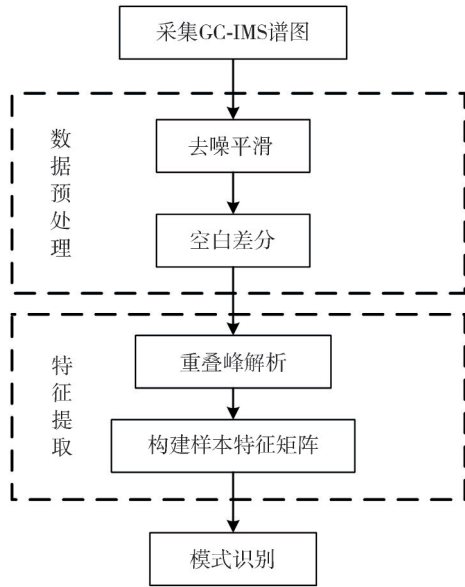


图3 呼气分析数据处理流程

Fig. 3 Data processing flow of exhaled breath analysis

$$\begin{cases} F(t) = \sum_{\tau=-m}^m Y(t) [Gauss(t-\tau) - k \times Gauss''(t-\tau)], \\ Gauss(t) = A \times e^{-\frac{(t-\mu)^2}{2\sigma^2}}, \end{cases} \quad (2)$$

式中: $F(t)$ 为平滑处理后的IMS信号; $Gauss(t)$ 为高斯锐化算子; A 、 μ 和 σ 分别为峰高、峰位偏差和标准差。为使IMS信号卷积平滑处理前后峰面积保持不变, 高斯锐化算子 $Gauss(t)$ 满足

$$\int_{-m}^m Gauss(t) dt = 1, \quad (3)$$

式中: m 为平滑卷积的半窗宽。

在峰值拟合阶段, 首先基于锐化后的信号提取各个IMS特征峰的峰高 A 、峰位置 μ 和标准差 σ 。在后续的优化迭代过程中, 调整峰高与标准差以最小化模型拟合信号与原始IMS信号之间的均方误差, 从而保证拟合结果的准确性与可靠性。

$$\min \sum_t \left[Y(t) - \sum_{i=1}^n A_i e^{-\frac{(t-\mu_i)^2}{2\sigma_i^2}} \right]^2, \quad (4)$$

式中: n 为单幅IMS谱图提取到的所有峰数量。

1.3.2 模式识别

针对呼气指纹图的模式识别任务, 本文选取3种具有代表性的分类模型进行对比分析: 传统批量学习模型随机森林(Random Forest, RF)、数据流增量学习模型Hoeffding树分类器(Hoeffding Tree Classifier, HTC)以及增量集成学习模型蒙德里安森林(Mondrian Forest, MF)。

RF是一种典型的机器学习方法, 由多个决策

1.3.1 GC-IMS谱图特征提取

为了提取GC-IMS谱图中的谱峰位置及强度信息作为分类识别的特征, 采用基于二阶导数的高斯锐化方法实现IMS重叠峰解析。该方法可分为谱峰锐化与峰值拟合两个阶段。

在谱峰锐化阶段, 首先利用二阶导数锐化增大重叠峰分离度。

$$S(t) = Y(t) - k \times Y''(t), \quad (1)$$

式中: $S(t)$ 为谱峰锐化后的IMS一维信号; $Y(t)$ 为IMS原始信号; $Y''(t)$ 为IMS信号二阶导数; t 为漂移时间; k 为锐化权重因子。

为了避免谱峰锐化过程中高频噪声放大形成伪峰, 进一步引入高斯卷积对锐化信号进行平滑处理。

树构成, 模型训练依赖完整的数据集, 采用批量学习方式完成建模过程^[12]。HTC属于在线学习范式, 其构建基于Hoeffding边界理论, 可在数据流环境中动态更新模型结构, 适用于样本规模较大的场景^[13]。上述两类方法均依赖人工设定超参数及模型结构, 模型性能在很大程度上受制于使用者的经验水平与参数调整能力。

作为一种新的增量学习方法, MF结合了随机森林与增量学习的优势, 以蒙德里安过程为理论基础, 实现了模型的在线高效更新^[14]。MF由若干棵独立的蒙德里安树(Mondrian Tree, MT)组成, 每棵树利用一种随机空间划分过程进行构建。这种划分过程具备自洽性, 使得在完整数据集上采样得到的树在其任意子集上的限制依然服从相同的分布。

图4为MT的构建过程, 其中, 左图为对应的树结构及时间序列, 右图为输入特征空间的划分示意图。MT是通过不断地在特征空间中随机选取维度并设置划分阈值, 将数据区域逐步细分而构建完成的。每次划分只发生在当前叶节点所对应的区域内, 保证模型结构具有局部自适应性。图4中特征空间由两个特征变量组成, 分别为横轴的 x_1 和纵轴的 x_2 。MT首先在时间序列 $t=0.42$ 时进行第一次划分, 从 $x_1=0.33$ 处将特征空间分为左右两个区域(B_{left} 和 B_{right}), 分别对应树结构中

的左右两个子节点。随后,在时间顺序 $t=0.5$ 时,右区域在 $x_2=0.5$ 处进一步划分。每一次划分

均伴随时间标记,该时间序列沿树结构的分支路径记录,完整反映了 MT 的构建过程。

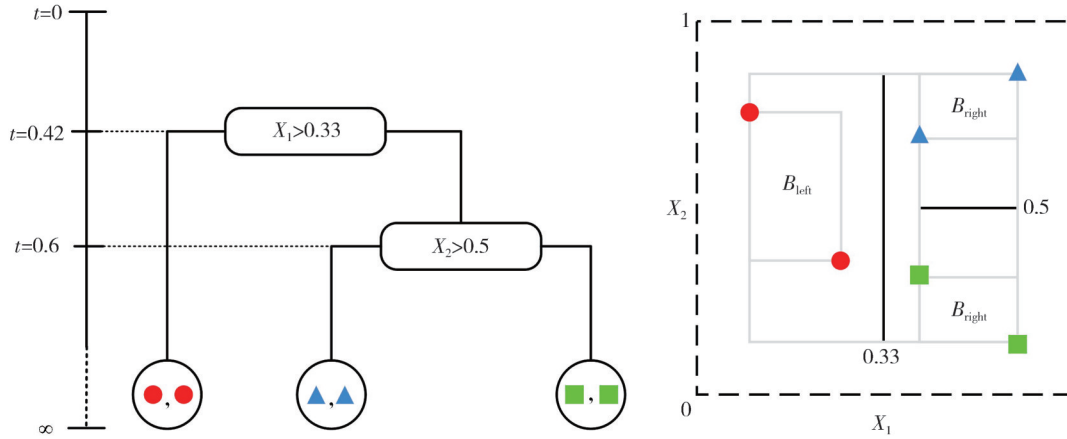


图 4 蒙德里安树构建过程

Fig. 4 Construction process of Mondrian tree

在完成空间划分后, MT 通过叶节点内的类别分布实现分类判别。模型依据输入样本的特征向量沿划分路径向下递归遍历,直至到达对应的叶节点。该叶节点基于训练阶段聚集的样本类别信息输出预测结果,通常采用最大类概率作为分类判定标准。每棵 MT 独立进行分类判别,最终综合森林中所有树的分类结果形成整体的最终决策。

为了解释模型分类决策的内在机制,利用 SHAP(Shapley Additive Explanations)方法^[15]进行特征重要性分析。SHAP 方法通过计算每个特征对模型预测输出的贡献,量化每个特征的重要性。

2 实验结果与分析

2.1 IMS 谱图高斯锐化

图 5(a) 所示为 IMS 呼气谱图中存在的多个不同强度的特征峰。1#峰受邻近主峰影响,常规方法难以精确识别。同样,2#峰也与相邻峰重叠,影响了峰形特征的准确表征。这种谱峰重叠现象在呼气分析中十分常见,主要原因是呼气成分复杂并且 IMS 仪器分辨率有限。

为量化峰分离效果,采用色谱学中的分离度作为衡量指标^[16]。

$$R_s = 2 \frac{|t_1 - t_2|}{w_1 + w_2}, \quad (5)$$

式中: t_1 和 t_2 分别为相邻峰漂移时间; w_1 和 w_2 为对应两峰的底峰宽(底峰宽 $w=4\sigma$)。当 $R_s > 0.5$ 时,两峰之间出现明显峰谷, $R_s > 1$ 表明两峰基本分离。

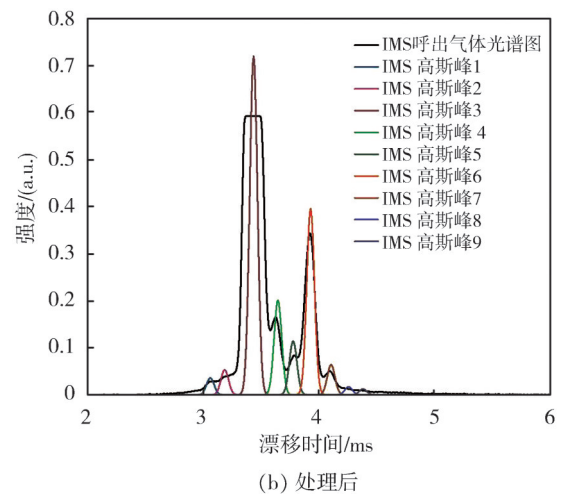
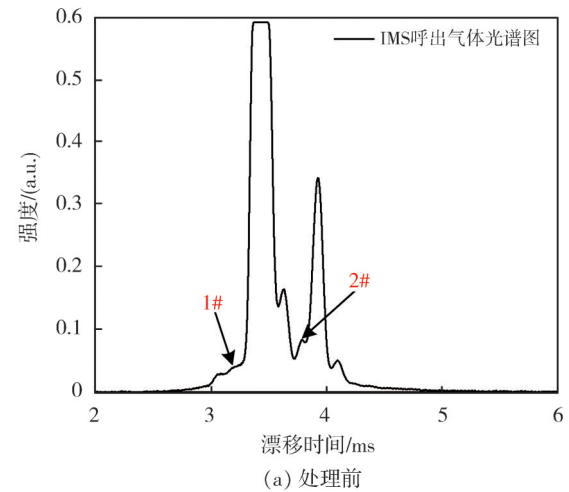


图 5 高斯锐化处理前后的 IMS 谱图

Fig. 5 IMS spectra comparison before and after Gaussian sharpening

图 5(b) 为基于二阶导数的高斯锐化方法处理后的结果。经过高斯锐化处理,原本复杂的呼气谱图被解析为 9 个独立的高斯峰(IMS 高斯峰 1~9),并以不同颜色进行区分。

表 2 列出了图 5(b) 中识别出的高斯峰之间相邻峰的分离度。高斯峰 1~高斯峰 9 共 8 组相邻峰的 R_s 均大于 0.9,表明该高斯锐化方法能够显著改善重叠峰的可分辨性。

表 2 高斯锐化法重叠峰解析结果

Tab. 2 Resolution after Gaussian sharpening of overlapping peaks

相邻高斯峰峰对	μ_1/ms	μ_2/ms	R_s
高斯峰 1-2	3.065	3.191	0.95
高斯峰 2-3	3.191	3.441	1.90
高斯峰 3-4	3.441	3.651	1.60
高斯峰 4-5	3.651	3.783	1.00
高斯峰 5-6	3.783	3.934	1.15
高斯峰 6-7	3.934	4.112	1.35
高斯峰 7-8	4.112	4.263	1.15
高斯峰 8-9	4.263	4.382	0.95

2.2 GC-IMS 指纹图特征提取

将上述高斯锐化算法应用于 GC-IMS 指纹图,评估其对指纹图特征提取的改善情况。图 6 为处理前后的 GC-IMS 呼气指纹图。

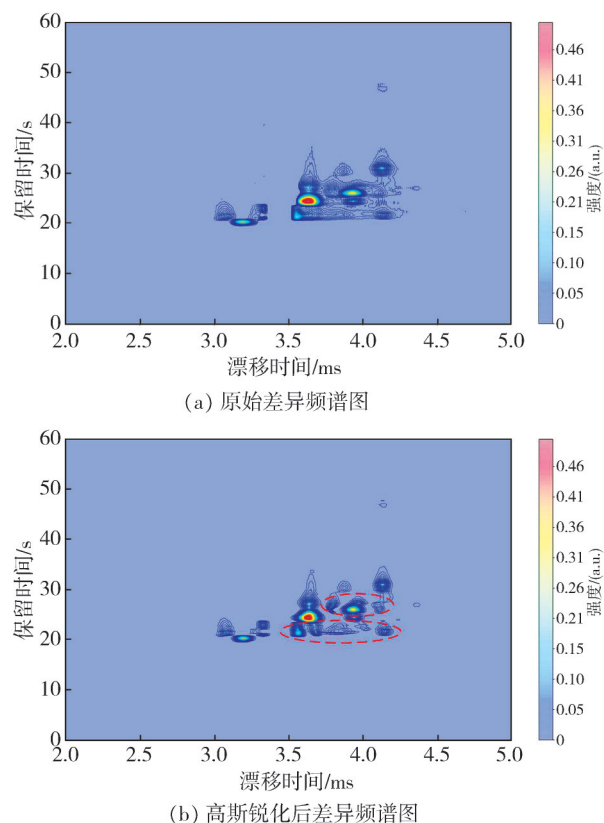


图 6 高斯锐化处理前后 GC-IMS 指纹图

Fig. 6 GC-IMS fingerprints before and after Gaussian sharpening

由图 6 可以明显看出,经过高斯锐化处理,原先粘连在一起的谱峰分离效果良好。为定量评价指纹图的重构质量,采用峰值信噪比(Peak Signal to Noise Ratio, PSNR)^[17]对处理前后的指纹图进行评估。PSNR 是评估图像质量的常用指标,其定义为处理前后信号之间的均方误差。通常认为 PSNR 值超过 30 dB 时图像失真较小,而本文高斯锐化处理后的指纹图 PSNR 达到 50 dB,说明指纹图在经过处理后失真极低。同时,处理后的指纹图中提取到更多的特征峰。高斯锐化算法不仅提高了指纹图整体信噪水平,还抑制了基线漂移及拖尾干扰,显著提高了重叠峰的解析效果,从而提高了呼气指纹图特征提取的准确性。

2.3 模式识别

2.3.1 模型性能评估

图 7 给出了 MF、HTC 和 RF 3 种模型随训练样本数量增加时训练时间的变化趋势。所有模型的训练时间均随样本量增加而增长,但 RF 的训练耗时始终明显高于 MF 和 HTC,且增长速率最快。这是由于 RF 在处理增量数据时每次都都需要使用全部样本重新训练整个模型,计算开销极大,难以满足实时增量学习的需求。HTC 的训练时间最低且随样本数增加变化最小。这得益于 HTC 仅构建单棵决策树并采用增量更新机制,有效降低了模型更新的计算成本,显示出其在临床快速分析中的效率优势。

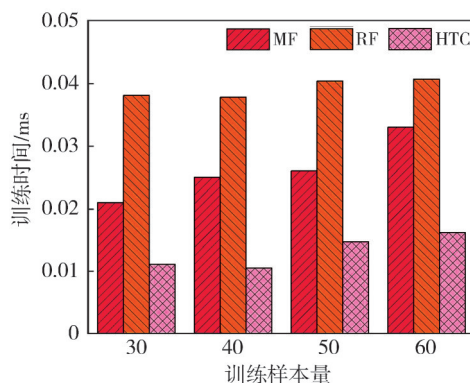


图 7 不同模型训练时间

Fig. 7 Training time of different models

图 8 为 3 种模型随训练样本数量增加时分类准确率变化情况。结果表明,在 30 个样本的起始阶段,MF 凭借其在线更新机制和自适应划分策略,能够迅速建立决策边界,测试准确率接近

100%。随着样本量逐渐增加,数据分布开始覆盖更多边界情形和罕见模式,MF需在既有树结构上局部增添结点以吸收新信息,早期过于理想化的划分因而得到修正,准确率回落并最终稳定在90%。模型的平均准确率保持在93.75%,反映了模型由小样本拟合最优向样本增长稳健后模型性能泛化的自然收敛。

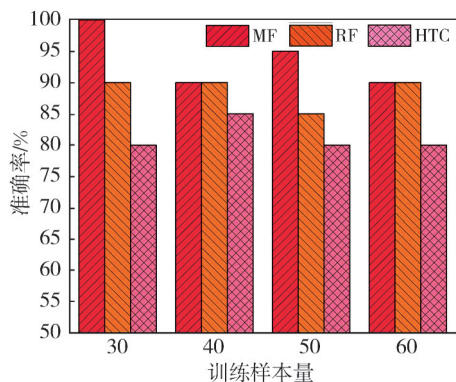


图8 不同模型分类准确率

Fig. 8 Classification accuracy of different models

进一步比较可见,RF在样本量扩大后才能充分发挥集成优势,最终准确率与MF持平,但由于每轮需重新训练整个森林,在线更新开销显著高于MF。HTC虽训练耗时最短,但在全程准确率上始终落后。3种模型对数据划分均具有良好的稳定性,通过三折交叉验证并对训练-测试划分进行多次随机重采样,各模型在所有实验中的准确率方差均小于10%。

综合训练效率与准确率结果可知,MF模型在保持高分类准确性的同时具备良好的增量训练效率,尤其在初始训练样本很少的情况下展现了卓越的泛化性能,并且随着样本规模扩大仍能维持优异表现。MF模型适合作为实际GC-IMS呼气检测中的增量学习分类器。

2.3.2 特征重要性分析

图9显示了MF模型中各特征对分类结果影响程度的SHAP值分布,其中峰28、峰1、峰10和峰14这4个特征对MF模型分类结果影响显著。

当峰28的特征值较高时,模型更倾向于将样本判定为“1”类;而峰14越大,样本越倾向于被识别为“0”类。这些峰对应的呼气组分在分类中起到关键作用,可能蕴含重要的生理意义,可以为后续深入探究呼气指纹与疾病表型关联及分析疾病潜在标志物提供线索。

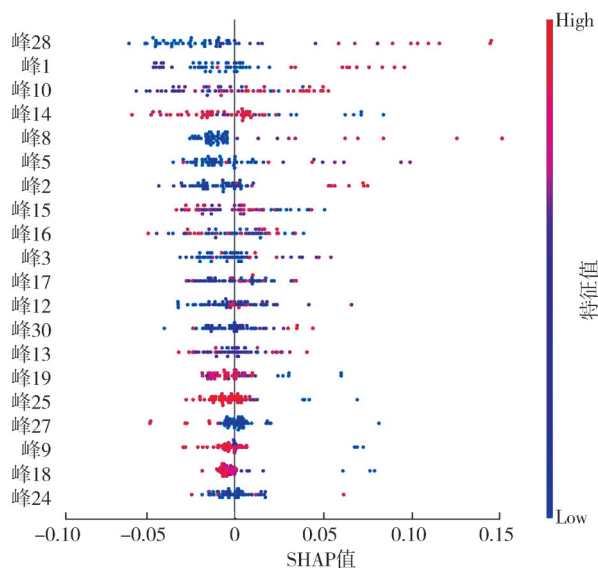


图9 MF模型的SHAP分析

Fig. 9 SHAP analysis of the MF model

3 结 论

针对GC-IMS呼气检测中指纹图重叠导致的特征提取困难,以及实际应用场景下样本有限、需支持持续更新等问题,提出一种结合高斯二阶导数锐化与增量学习模型的呼气分析数据处理方法。特征提取阶段,引入基于二阶导数的高斯锐化算法,在处理单一重叠峰方面取得良好效果,有效提升了GC-IMS指纹图中重叠峰的分离度与指纹图重构质量。MF的增量学习分类模型在小样本环境下能够高效适应新样本的动态引入,避免传统随机森林在新增样本处理时所需的高计算成本。在面向临床应用的场景中,MF模型的在线更新能力和较低计算开销,使其具备很好的实际应用潜力。

利益冲突声明/Conflict of Interests

所有作者声明不存在利益冲突。

All authors declare no relevant conflict of interests.

参考文献

- [1] GANEEV A A, GUBAL A R, LUKYANOV G N, et al. Analysis of exhaled air for early-stage diagnosis of lung cancer: opportunities and challenges[J]. Russian Chemical Reviews, 2018, 87(9): 904-921.
- [2] DENT A G, SUTEDJA T G, ZIMMERMAN P V. Exhaled breath analysis for lung cancer[J]. Journal of Thoracic Disease, 2013, 5(Suppl 5): S540-S550.

- [3] CHRISTIANSEN A, DAVIDSEN J R, TITLESTAD I, et al. A systematic review of breath analysis and detection of volatile organic compounds in COPD[J]. *Journal of Breath Research*, 2016, 10(3): 034002.
- [4] SUBALI A D, WIYONO L, YUSUF M, et al. The potential of volatile organic compounds-based breath analysis for COVID-19 screening: a systematic review & meta-analysis [J]. *Diagnostic Microbiology and Infectious Disease*, 2022, 102(2): 115589.
- [5] VELDE S V D, NEVENS F, HEE P V, et al. GC-MS analysis of breath odor compounds in liver patients [J]. *Journal of Chromatography B*, 2008, 875(2): 344-348.
- [6] KONONOV A, KOROTETSKY B, JAHATSPANIAN I, et al. Online breath analysis using metal oxide semiconductor sensors (electronic nose) for diagnosis of lung cancer[J]. *Journal of Breath Research*, 2020, 14(1): 016004.
- [7] RUSZKIEWICZ D M, SANDERS D, OBRIEN R, et al. Diagnosis of COVID-19 by analysis of breath with gas chromatography-ion mobility spectrometry-a feasibility study[J]. *EClinicalMedicine*, 2020, 29-30: 100609.
- [8] XING G, FAN J, ZHENG Y, et al. Multi-scale input mirror network for tuberculosis detection in CXR image[J]. *Journal of Measurement Science and Instrumentation*, 2025, 16(1): 1-10.
- [9] JUMA A R, CHEN Jing, PENG Junlan, et al. EAC-Net: Ensemble adversarial co-training neural network for handling missing modalities in MRI images for brain tumor segmentation[J]. *Journal of Measurement Science and Instrumentation*, 2025, 16(1): 11-25.
- [10] LAWAL O, AHMED W M, NIJSEN T M E, et al. Exhaled breath analysis: a review of 'breath-taking' methods for off-line analysis [J]. *Metabolomics*, 2017, 13(10): 1-16.
- [11] 吴良杰, 贾建, 何秀丽, 等. 基于气相色谱-离子迁移谱的呼气分析系统研制[J]. *中国医疗器械信息*, 2024, 30(5): 45-48.
- WU Liangjie, JIA Jian, HE Xiuli, et al. Development of a breath analysis system based on gas chromatography-ion mobility spectrometry [J]. *China Medical Device Information*, 2024, 30(5): 45-48. (in Chinese)
- [12] PARMAR A, KATARIYA R, PATEL V. A review on random forest: an ensemble classifier[C]//International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI), 2018: 758-763.
- [13] DOMINGOS P, HULTEN G. Mining high-speed data streams [C]//Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2000: 71-80.
- [14] LAKSHMINARAYANAN B, ROY D M, TEH Y W. Mondrian forests for large-scale regression when uncertainty matters[C]//Artificial Intelligence and Statistics, PMLR, 2016: 1478-1487.
- [15] MARCILIO W E, ELER D M. From explanations to feature selection: assessing SHAP values as feature selection mechanism[C]//2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), 2020: 340-347.
- [16] DEWULF J O, VAN LANGENHOVE H, Wittmann G. Analysis of volatile organic compounds using gas chromatography[J]. *Trends in Analytical Chemistry*, 2002, 21(9-10): 637-646.
- [17] KAUR P, SINGH J. A study on the effect of Gaussian noise on PSNR value for digital images[J]. *International Journal of Computer and Electrical Engineering*, 2011, 3(2): 319-321.