

基于飞行时间二次离子质谱的指纹遗留人个体识别的研究

吴寅¹ 田师思¹ 李文杰¹ 孙令辉¹ 宋禧林¹ 赵雅彬^{*1} 李展平^{*2,3}

¹(中国人民公安大学侦查学院, 北京 100038) ²(清华大学分析中心, 北京 100084)

³(清华大学化学系, 有机光电子与分子工程教育部重点实验室, 北京 100084)

摘要 指纹作为法庭科学领域中用于人身认定的重要物证, 蕴含了丰富的个体信息。但是, 在实际案件现场中, 提取到的多为模糊和残缺的指纹, 无法利用传统的形态比对方方法进行鉴定; 而且指纹中物质成分复杂, 也很难通过检测其中的具体物质实现遗留人的溯源分析。飞行时间二次离子质谱(TOF-SIMS)能够在不破坏检材的情况下获得指纹物质的离子信号, 同时获得对应的离子分布图像, 为利用指纹物质信息进行个体识别提供了可能。本研究利用 TOF-SIMS 采集了 5 名志愿者十指指纹的质谱信息, 初步探究了基于指纹物质离子信号进行遗留人个体识别的可行性。首先筛选出 98 种代表指纹物质的离子信号, 利用主成分分析法对筛选出的离子信号强度进行降维, 再将降维后的主成分新变量输入到不同的分类器中对比。结果表明, 使用无监督的层次聚类法的结果与实际的分类相差较大, 难以进行个体区分; 使用有监督的 Fisher 判别法对个体进行判别, 正确率可以达到 94%, 交叉验证的正确率也达到了 90%; 将数据输入到多层感知神经网络中, 经过多次训练可以得到更加理想的结果。本研究借助多种统计分析方法建立了个体分类和预测模型, 为指纹证据价值的提升提供了新渠道, 也为后期利用指纹物质信号进行个体识别的深入探究奠定了研究基础。

关键词 飞行时间二次离子质谱; 指纹物质; 个体识别; 多元统计分析

指纹因其形态具有“人各不同, 终身基本不变”的特性而被广泛用于人身的同一认定, 在刑事司法领域中发挥着重要作用。但是, 在实际案件中, 受现场环境、客体条件和遗留人行为等因素的影响, 侦查人员提取到的指纹多模糊不清或残缺不全, 以致其形态特征无法得到应用, 证据价值难以实现。为此, 研究人员展开了对残缺指纹中物质成分的检验, 以充分挖掘指纹的化学信息, 从多种信息角度进行遗留人的溯源分析。

指纹中主要包含外源性和内源性两类物质。其中, 对指纹外源性物质的检验可以反映遗留人是否沾染过特殊的物质, 如毒物毒品^[1-2]、爆炸物^[3]和避孕套润滑剂^[4-5]等, 从而与特定案件联系起来; 对内源性物质的检验可以判断遗留人是否具有吸毒^[6]和服药^[7]等情况。此外, 还能通过检测内源性物质含量的变化以及个体间差异分析遗留人的人群特征, 如年龄^[8-9]和性别^[10-11]等。研究人员多采用光谱法^[12-13]和质谱法^[14-16]对指纹物质进行检验。然而, 由于指纹物质的成分复杂且含量较少^[17-18], 光谱法的灵敏度和准确度在检测内源性物质时存在一定的局限性^[19]。灵敏度和准确度更高的气相色谱-质谱联用^[14-15]以及液相色谱-质谱联用^[16]等质谱法则无法获取指纹的图像信息, 并且还会对指纹样本造成破坏。

近年来发展的质谱成像技术(Mass spectrometry imaging, MSI)能够同时获取样本表面的物质及物质分布信息, 具有近无损、高灵敏度和多组分平行分析等优势, 受到研究人员的广泛关注^[20]。其中, 解吸电喷雾质谱(Desorption electrospray ionization-mass spectrometry, DESI-MS)^[21-22]、基质辅助激光解吸电离质谱(Matrix-assisted laser desorption ionization-mass spectrometry, MALDI-MS)^[23-24]等技术均在指纹分析领域显示出良好的应用前景。但是这些技术的空间分辨率较低, 对于残缺指纹的成像效果较差。相比而言, 飞行时间二次离子质谱(Time-of-flight secondary ion mass spectrometry, TOF-SIMS)具有更高的空间分辨和质量分辨能力^[25], 不仅能够显现出指纹的三级特征^[26-27], 还能够在无法分析出具体物质的情况下, 通过图像判断来自指纹物质的离子信号, 为利用指纹物质信息进行个体溯源提供了更多可能。目前,

2022-05-23 收稿; 2022-11-18 接受

国家重点研发计划项目(No. 2018YFA0702600)和公安部科技计划面上项目(No. 2021JSM01)资助。

* E-mail: zhaoyb04@gmail.com; zhanpingli@mail.tsinghua.edu.cn

TOF-SIMS 技术在指纹研究中的应用已有很多报道,如 Szykowskam 等利用 TOF-SIMS 检测出指纹中的苯丙胺类物质^[28],探究了转印指纹中相关物质的检验^[29]; Thandauthapani 等^[30]利用 TOF-SIMS 检测了金属客体上遗留的指纹; Muramoto 等^[31]利用 TOF-SIMS 的成像功能并根据指纹中饱和脂肪酸的扩散建立误差函数,用于判定指纹的遗留时间。但是,当前的研究多集中于指纹物质和形态信息的获取,很少进行遗留人的人群特征分析。本研究组前期利用指纹的 TOF-SIMS 离子信号进行了人群性别的区分^[32],但是对于不同遗留人之间的个体识别还没有相关的研究报道。

本研究利用 TOF-SIMS 初步探索了利用指纹物质进行个体识别的可行性。通过对来自不同个体的指纹进行检测,筛选出共有的指纹物质的离子信号;然后利用主成分分析法(Principal component analysis, PCA)对数据进行降维,以减少重叠信息;最后使用层次聚类法(Hierarchical clustering)、Fisher 判别法(Fisher discriminant)以及多层感知机神经网络(Multi-layer perception neural networks, MLP)对降维获得的新变量进行分析,并对比分类结果。最终建立了个体的分类和预测模型,为利用指纹物质信息进行人群溯源分析提供了新思路。

1 实验部分

1.1 仪器与材料

TOF-SIMS5 飞行时间二次离子质谱仪(德国 ION-TOF 公司);硅片(5 mm×5 mm,浙江立晶硅材料公司)。

1.2 仪器测试参数

TOF-SIMS 参数设置:一次离子源选用 Bi₃⁺(液态金属离子枪),能量为 30 keV,电流为 0.8 pA,扫描面积为 500 μm × 500 μm,扫描时间为 500 s,扫描像素为 256 × 256 pixel,质量范围为 0~2000 Da (一次离子束脉冲周期 150 μs),二次离子模式为正离子模式。

1.3 样品制备过程

本研究募集了 5 名志愿者(3 名女性,2 名男性),分别收集每名志愿者 10 根手指的指纹样本,所采集的指纹仅供研究使用。在统一捺印指纹样本前,志愿者先将手洗净,待水自然晾干后,确保志愿者不触摸任何物品和部位,等待 30 min,保证指纹中均为自然分泌物质;然后,将每根手指的指尖部位轻放在硅片上约 3 s 后抬起,获得指纹样本;最后,用导电胶将硅片粘在样品托上,将样品托放入真空中待测。本研究通过了相关伦理委员会的审批。

1.4 检测和数据筛选方法

分别对 5 名志愿者共 50 枚指纹进行检测,检测选取的扫描面积为 500 μm×500 μm,指纹纹线的宽度约为 200~400 μm,犁沟的宽度约为 100~200 μm^[33],操作过程中调整测试位置,保证每次检测的区域至少包含一段完整形态的指纹纹线。

2 结果与讨论

2.1 数据的相关性分析与降维

检测完成后,根据出峰位置的离子分布图像是否能够呈现指纹形态筛选出可能来自于指纹物质的离子信号(如图 1),最后将所选离子的信号强度导出。

2.1.1 相关性分析

按照 1.4 节所述方法,根据 5 名志愿者指纹质谱信息中共有的信号峰的离子分布图像筛选出了 98 种离子信号(表 1)。由于 TOF-SIMS 一次离子束能量较大,大分子量的物质容易裂解成若干碎片离子,各离子信号之间存在较强的相关性^[34]。为减小系统误差,首先对各离子信号进行相关性分析,验证是否存在信息上的重叠。

本研究使用 Pearson 相关系数(Pearson correlation coefficient)表示两个离子信号之间的相关程度^[35]。将两个变量的协方差除以标准差之积计算出 Pearson 相关系数,表 2 展示了前 10 个离子信号之间的 Pearson 相关系数矩阵。Pearson 相关系数介于-1~1 之间,绝对值越接近 1 表明相关性越强^[36],*m/z* 45.98 和 61.97 的相关系数达到了 0.953,*m/z* 61.95 和 62.96 的相关系数达到了 0.986,均属于极度相关。此结果

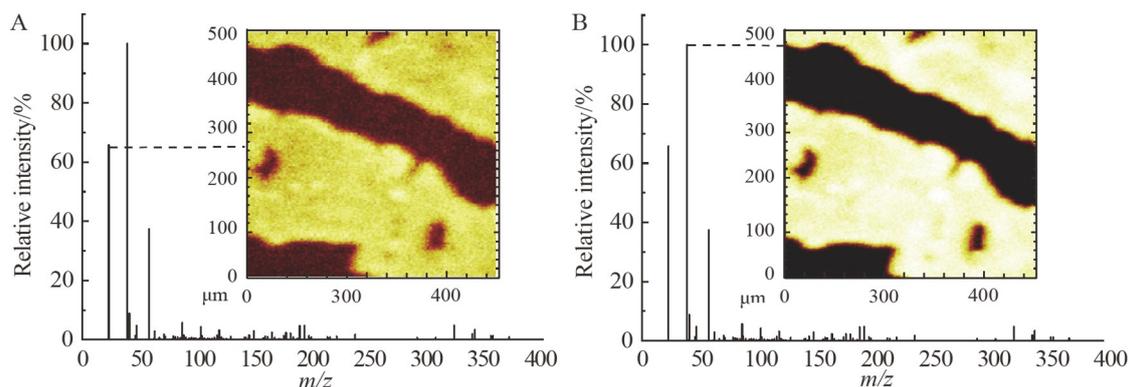


图1 指纹飞行时间二次离子质谱中(A)钠离子和(B)钾离子的分布图像

Fig.1 Images of distribution of (A) Na^+ and (B) K^+ within time-of-flight secondary ion mass spectrometry of fingerprints

表1 筛选出的指纹物质的离子信号

Table 1 Selected ion signals of compounds of fingerprint

序号 No.	离子信号 Ion signal (m/z)										
1	22.99	18	78.93	35	95.91	52	116.89	69	157.88	86	212.83
2	38.96	19	78.95	36	96.92	53	116.91	70	158.91	87	214.83
3	39.96	20	79.92	37	98.91	54	118.93	71	160.90	88	220.87
4	40.96	21	79.94	38	100.93	55	119.92	72	164.91	89	236.84
5	45.98	22	80.95	39	101.92	56	120.93	73	166.91	90	290.80
6	46.99	23	82.93	40	101.95	57	121.92	74	170.84	91	306.76
7	58.07	24	84.91	41	102.92	58	122.92	75	172.84	92	306.83
8	61.95	25	84.93	42	102.96	59	124.91	76	174.88	93	322.80
9	61.97	26	84.96	43	103.93	60	125.89	77	176.87	94	338.78
10	62.96	27	86.96	44	103.95	61	125.91	78	180.87	95	340.78
11	62.98	28	87.95	45	105.93	62	128.95	79	182.87	96	354.75
12	66.96	29	87.97	46	108.94	63	140.90	80	186.82	97	356.75
13	70.98	30	89.95	47	109.91	64	141.88	81	188.81	98	370.73
14	71.98	31	92.97	48	109.94	65	143.88	82	192.84		
15	77.92	32	93.92	49	110.90	66	144.92	83	196.86		
16	77.94	33	93.94	50	112.89	67	148.92	84	197.86		
17	77.97	34	94.92	51	114.89	68	156.88	85	198.86		

说明部分离子之间的确存在信息重叠,确认了数据降维的必要性。

2.1.2 主成分分析

验证了存在信息重叠后,本研究采用PCA对数据进行降维分析。PCA是将原始变量重新组合成几个新的综合变量,这些新变量尽可能多地反映原来变量的信息,并且互不相关,可以有效消除数据的多重共线性问题。基于主成分的特征值大于1,即包含了1个以上原始变量的信息,共选取了6个主成分作为新变量,累计方差贡献率达到94.90%,起到了较好的降维作用,同时也极大地保留了原始数据的信息。

PCA本身也可以作为一种无监督聚类分析方法,保留第一主成分和第二主成分绘制出一幅二维的PCA得分(Score)散点图(图2)。其中,Donor 1、2、3为女性志愿者,Donor 4、5为男性志愿者。由图2可见,5名志愿者样本信息总体重叠严重,难以区分,但个别志愿者之间还是存在差异,如Donor 1和Donor 4之间的信息虽然存在部分重叠,但可以看出其组内距离小于组间距离;Donor 1和Donor 2则是完全直接区分开的。虽然并非每个个体之间都有较好的区分度,但是后续还可以基于PCA后保存的

表2 Pearson相关系数矩阵

Table 2 Pearson correlation coefficient matrix

离子信号 Ion signal	相关系数 Correlation coefficient									
	<i>m/z</i> 22.99	<i>m/z</i> 38.96	<i>m/z</i> 39.96	<i>m/z</i> 40.96	<i>m/z</i> 45.98	<i>m/z</i> 46.99	<i>m/z</i> 58.07	<i>m/z</i> 61.95	<i>m/z</i> 61.97	<i>m/z</i> 62.96
<i>m/z</i> 22.99	1.000	0.868	0.388	0.839	0.781	0.660	0.197	0.821	0.887	0.826
<i>m/z</i> 38.96	0.868	1.000	0.323	0.910	0.525	0.550	-0.109	0.699	0.673	0.683
<i>m/z</i> 39.96	0.388	0.323	1.000	0.618	0.661	0.589	0.100	0.775	0.601	0.791
<i>m/z</i> 40.96	0.839	0.910	0.618	1.000	0.670	0.713	-0.106	0.848	0.783	0.843
<i>m/z</i> 45.98	0.781	0.525	0.661	0.670	1.000	0.798	0.412	0.825	0.953	0.872
<i>m/z</i> 46.99	0.660	0.550	0.589	0.713	0.798	1.000	0.027	0.748	0.805	0.745
<i>m/z</i> 58.07	0.197	-0.109	0.100	-0.106	0.412	0.027	1.000	0.055	0.307	0.155
<i>m/z</i> 61.95	0.821	0.699	0.775	0.848	0.825	0.748	0.055	1.000	0.865	0.986
<i>m/z</i> 61.97	0.887	0.673	0.601	0.783	0.953	0.805	0.307	0.865	1.000	0.906
<i>m/z</i> 62.96	0.826	0.683	0.791	0.843	0.872	0.745	0.155	0.986	0.906	1.000

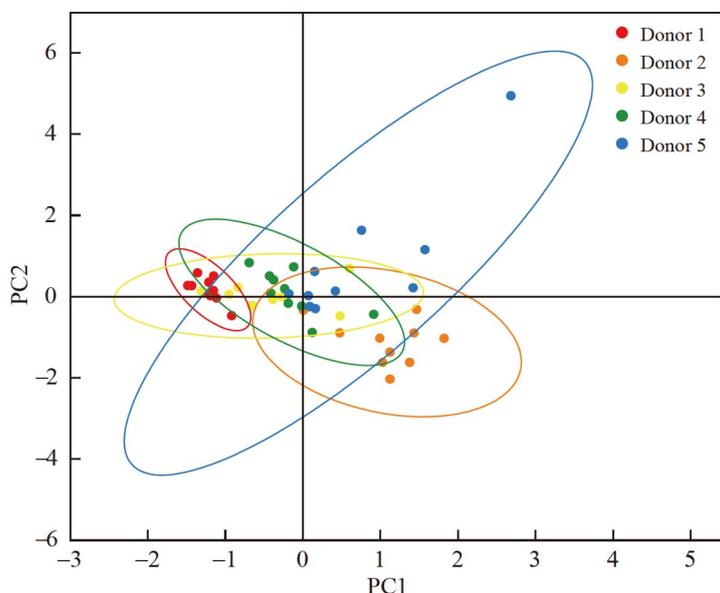


图2 主成分分析(PCA)得分散点图

Fig.2 Principal component analysis (PCA) scatter score plot

6 个新变量做进一步分类。

2.2 数据的分类与预测

2.2.1 层次聚类法

首先通过无监督的层次聚类法利用降维后的新变量进行个体之间分类效果的探索。选择欧几里得距离作为类内距离的测量方法, Ward 法作为类间距离的测量方法,对数据进行聚类分析,并绘制了聚类热图(图 3),横轴代表 5 名志愿者的 50 枚指纹样本,纵轴代表降维后的 6 个主成分变量,中间色块的颜色对应变量的数值,上方的谱系图则表示样本间聚类的远近关系。由图 3 可见,各样本之间颜色的相似度高,大多数为红色,仅根据颜色差异无法直接判断出样本的组别。从下方的聚类结果也可以看出,有较多不是来自同一个体的样本被分在一起。此结果说明个体间的差异较小,使用层次聚类法不能将不同人的指纹准确地区分开,也无法有效构建分类模型达到个体识别的目的。

2.2.2 Fisher判别法

由于聚类的效果并不理想,本研究选择了有监督的 Fisher 判别法对样本个体进行分类。Fisher 判别通过线性函数将高维空间的点向低维空间投影,并寻求组间差异的最大化。本研究共提取了 4 个判别函

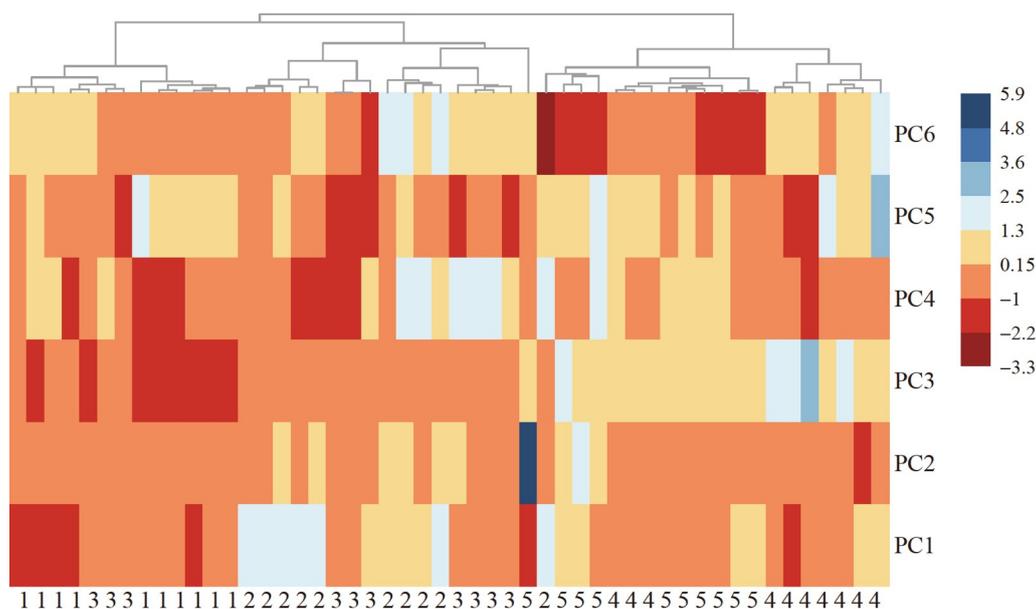


图3 层次聚类热图

Fig.3 Heatmap of hierarchical clustering

数,并根据第一和第二判别函数绘制出一幅联合分布图(图4),两个函数包含了91%的总信息量,其中, Donor 1、2、3为女性志愿者, Donor 4、5为男性志愿者。可以看出,在第一个判别维度上, Donor 1、2、3和 Donor 4、5的投影之间不存在重合,说明可以将女性和男性志愿者较好地地区分开; Donor 4和 Donor 5的投影略有重合,也具有一定的区分度, Donor 1、2、3之间则重合严重,无法进行区分。但是,根据在第二个判别维度上的投影, Donor 1、2、3之间也可以得到较好的区分效果。

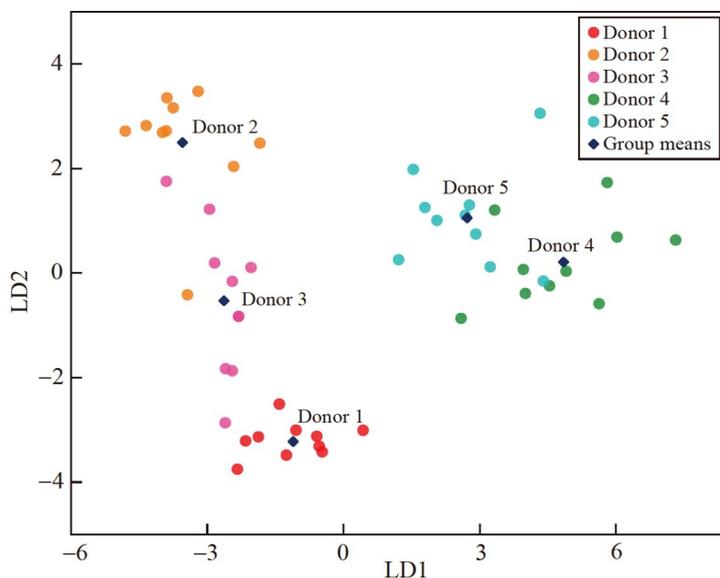


图4 Fisher判别散点图

Fig.4 Fisher discriminant scatter plot

通过4个判别函数计算出各个点的判别得分,根据判别公式判定每枚指纹来自的个体,并将所有案例的判别结果输出(表3)。由表3可知,利用Fisher判别可正确地对94.0%的原始样本进行分类,并且对每个个体分组判别的正确率均达到了90%。随后在总样本中剔除一份样本建立判别函数,再通过对剔除的样本判别分组进行交叉验证,获得的结果可正确地对90.0%的验证组样本进行分类。此实验结果表明,使用Fisher判别对个体分类的效果比较理想,并且具有较好的预测能力。

表3 Fisher判别分类结果

Table 3 Results of Fisher discriminant

个体 Individual	原始分类组 Initial classification group			交叉验证组 Cross-validation group		
	真阳性数 True positive	假阳性数 False positive	正确率 Accuracy/%	真阳性数 True positive	假阳性数 False positive	正确率 Accuracy/%
Donor1	10	1	100	10	1	100
Donor2	9	0	90	8	1	80
Donor3	9	1	90	8	1	80
Donor4	9	0	90	9	0	90
Donor5	10	1	100	10	2	100

2.2.3 多层感知机神经网络

MLP 是一种非线性分类方法,相比于线性分类方法其性能更优越。因此,进一步采用 MLP 对数据进行分析,以探究能否进一步提高模型的准确率,从而达到更加理想的效果。

MLP 模型由输入层、输出层、一个或多个隐藏层构成。隐藏层中可以包含多个神经元,隐藏层的层数和神经元个数直接决定了神经网络的复杂程度,而复杂度越高,模型的性能越好,但复杂度过高则会导致过拟合的问题^[37]。因此,首先需对模型的参数进行设置,达到模型的最优化。由于本实验样本量较小,所以只选用最简单的 3 层结构,即只有 1 个隐藏层。然后分别设定 1~6 个神经元,将每种条件的模型各训练 10 次,记录下每次训练后的正确率(图 5)。如图 5 所示,当神经元个数为 5 时,模型的预测正确率基本都能保持在 90%以上,相比其它条件下的模型正确率更高且更加稳定;继续增加节点数,正确率并未上升,甚至还出现了过度拟合的倾向,所以最终隐藏层神经元个数设为 5。此外,隐藏层的激活函数设为双曲正切,输出层的激活函数设为 softmax,训练集和验证集的比例设为 7:3。

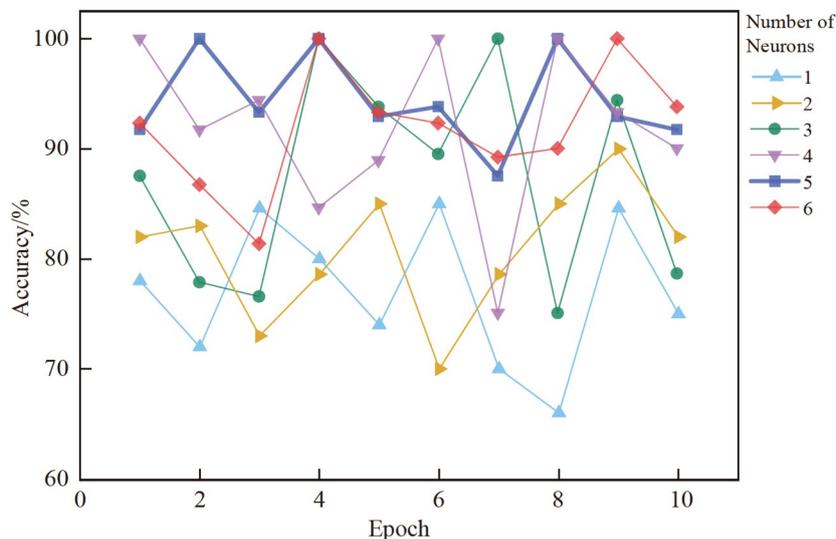


图5 不同神经元数的模型多次训练的准确率

Fig.5 Accuracy of models with different neurons after multiple trainings

由最终优化后的模型训练结果可知,采用非线性的多层感知机神经网络对人群分类的效果同样较好,多次训练的平均正确率达到 94.4%,并且高于线性判别的结果。虽然模型的预测效果比较理想,但是其中一次训练的正确率仅为 87.5%,说明此模型的预测可能并不稳定。为此,本研究基于本次训练建立了每个个体的受试者工作特征曲线(Receiver operating characteristic curve, ROC 曲线)(图 6)。ROC 曲线体现了模型预测结果的真假阳性比例,ROC 曲线下面积(Area under roc curve, AUC)越接近 1,模型的性能越好。如图 6 所示,每名志愿者对应曲线线下面积都大于 0.99,说明此模型具有较好的预测能力。预

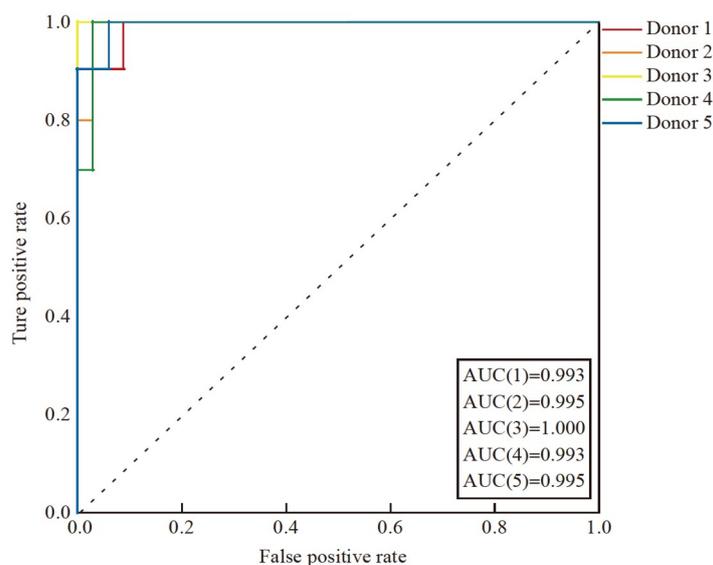


图6 基于5名志愿者判别结果的受试者工作特征曲线

Fig.6 Receiver operating characteristic (ROC) curve based on the discriminant of 5 donors

测结果的不稳定可能是由于每个分类中的样本量较小导致的,如果能增加训练的样本,可能会得到更好的分类和预测结果。

2.3 数据的优化

本研究中,指纹物质各离子信号的强度本质上与手指自然分泌物质的含量一致。实验通过控制分泌相同的时间,然后根据每个人分泌能力的不同进行个体区分。实际上,分泌能力除了与个人体质有关外,还容易受到情绪、外界温度和运动等因素的影响。某些因素在实验中难以控制,并且对结果的影响程度不得而知,这表明同一个人不同条件下捺印的指纹也可能存在显著差异,这将导致模型对一些外部样本的预测不够准确。

理论上,无论指纹物质的总量怎么变化,每个人所分泌的各种物质之间的比例应该是相对稳定的。实验中的每一枚指纹检测完后皆可得到该指纹所有离子强度的总和,用各信号强度与对应总离子强度的比值代替原始数据,然后采取相同方法进行数据处理。先用PCA法进行降维,再将降维后的新变量直接导入到Fisher判别当中进行分类,结果见表4,对样本分类的正确率达到了100%,交叉验证的正确率也达到了96%,均高于之前的结果,也证明了优化后的模型数据更加稳定。

表4 数据优化后的判别结果

Table 4 Results of discriminant after optimization

个体 Individual	原始分类组 Initial classification group			交叉验证组 Cross-validation group		
	真阳性数 True positive	假阳性数 False positive	正确率 Accuracy/%	真阳性数 True positive	假阳性数 False positive	正确率 Accuracy/%
	Donor1	10	0	100	10	0
Donor2	10	0	100	8	0	80
Donor3	10	0	100	10	2	100
Donor4	10	0	100	10	0	100
Donor5	10	0	100	10	0	100

3 结论

本研究基于TOF-SIMS质谱成像技术初步探究了利用指纹物质信号进行个体识别的可行性,借助多种数理统计方法建立了个体的分类和预测模型。结果显示,通过线性分类器Fisher判别法对个体进行分

类,总体可达到 90%以上的准确率,而使用非线性的 MLP 能得到更高的准确率,但是结果并不稳定。这可能是由于本研究的取样量较少,参与训练的样本量不够充足,导致模型对外部样本预测的准确率降低。在后续研究中,需要在不同情况下对遗留人的指纹重复多次取样以提高模型的稳定性;在选取的样本人数大于本研究的情况时,需重新调整模型参数,进一步复杂化模型。本研究为充分利用指纹化学信息提供了新思路,同时也为后期深入探究利用指纹物质进行遗留人的个体识别奠定了方法和理论基础。

References

- [1] COSTA C, WEBB R, PALITSIN V, ISMAIL M, DE PUIT M, ATKINSON S, BAILEY M J. *Clin. Chem.*, 2017, 63(11): 1745-1752.
- [2] CZERWINSKA J, JANG M, COSTA C, PARKIN M C, GEORGE C, KICMAN A T, BAILEY M J, DARGAN P I, ABBATE V. *Analyst*, 2020, 145(8): 3038-3048.
- [3] FORBES T P, SISCO E. *Anal. Chem.*, 2014, 86(15): 7788-7797.
- [4] BRADSHAW R, WOLSTENHOLME R, BLACKLEDGE R D, CLENCH M R, FERGUSON L S, FRANCESE S. *Rapid Commun. Mass Spectrom.*, 2011, 25(3): 415-422.
- [5] BRADSHAW R, WOLSTENHOLME R, FERGUSON L S, SAMMON C, MADER K, CLAUDE E, BLACKLEDGE R D, CLENCH M R, FRANCESE S. *Analyst*, 2013, 138(9): 2546-2557.
- [6] ISMAIL M, STEVENSON D, COSTA C, WEBB R, DE PUIT M, BAILEY M. *Clin. Chem.*, 2018, 64(6): 909-917.
- [7] KUWAYAMA K, TSUJIKAWA K, MIYAGUCHI H, KANAMORI T, IWATA Y T, INOUE H. *Anal. Bioanal. Chem.*, 2013, 405(12): 3945-3952.
- [8] HEMMILA A, MCGILL J, RITTER D. *J. Forensic Sci.*, 2008, 53(2): 369-376.
- [9] ANTOINE K M, MORTAZAVI S, MILLER A D, MILLER L M. *J. Forensic Sci.*, 2010, 55(2): 513-518.
- [10] ASANO K G, BAYNE C K, HORSMAN K M, BUCHANAN M V. *J. Forensic Sci.*, 2002, 47(4): 805-807.
- [11] FERGUSON L S, WULFERT F, WOLSTENHOLME R, FONVILLE J M, CLENCH M R, CAROLAN V A, FRANCESE S. *Analyst*, 2012, 137(20): 4686-4692.
- [12] WOLFBEIS O. *Angew. Chem. Int. Ed.*, 2009, 48(13): 2268-2269.
- [13] WANG J, WEI T, LI X, ZHANG B, WANG J, HUANG C, YUAN Q. *Angew. Chem. Int. Ed.*, 2014, 53(6): 1616-1620.
- [14] CROXTON R S, BARON M G, BUTLER D, KENT T, SEARS V G. *J. Forensic Sci.*, 2006, 51(6): 1329-1333.
- [15] GIROD A, WEYERMANN C. *Forensic Sci. Int.*, 2014, 238: 68-82.
- [16] VAN HELMOND W, DE PUIT M. *Methods Mol. Biol.*, 2019, 2030: 439-450.
- [17] CROXTON R S, BARON M G, BUTLER D, KENT T, SEARS V G. *Forensic Sci. Int.*, 2010, 199(1-3): 93-102.
- [18] MICHALSKI S, SHALER R, DORMAN F L. *J. Forensic Sci.*, 2013, 58: S215-S220.
- [19] VAN DAM A, VAN BEEK F T, AALDERS M C G, VAN LEEUWEN T G, LAMBRECHTS S A G. *Sci. Justice*, 2016, 56(2): 143-154.
- [20] IFA D R, MANICKE N E, DILL A L, COOKS R G. *Science*, 2008, 321(5890): 805.
- [21] BAILEY M J, BRADSHAW R, FRANCESE S, SALTER T L, COSTA C, ISMAIL M, P. WEBB R, BOSMAN I, WOLFF K, DE PUIT M. *Analyst*, 2015, 140(18): 6254-6259.
- [22] SUN Q, LUO Y, SUN N, ZHANG Q, WANG Y, YANG X. *Forensic Sci. Int.*, 2022, 336: 111321.
- [23] FRANCESE S, BRADSHAW R, FERGUSON L S, WOLSTENHOLME R, CLENCH M R, BLEAY S. *Analyst*, 2013, 138(15): 4215-4228.
- [24] LAUZON N, DUFRESNE M, CHAUHAN V, CHAURAND P. *J. Am. Soc. Mass Spectrom.*, 2015, 26(6): 878-886.
- [25] LI Zhan-Ping. *Bull. Mineral., Petrol. Geochem.*, 2020, 39(6): 1173-1190.
李展平. *矿物岩石地球化学通报*, 2020, 39(6): 1173-1190.
- [26] CAI L, XIA M C, WANG Z, ZHAO Y B, LI Z, ZHANG S, ZHANG X. *Anal. Chem.*, 2017, 89(16): 8372-8376.
- [27] LI Wen-Jie, SUN Ling-Hui, YOU Wei, WANG Li-Xue, ZHAO Ya-Bin, LI Zhan-Ping. *Chin. J. Anal. Chem.*, 2020, 48(11): 96-103.
李文杰, 孙令辉, 游伟, 王利雪, 赵雅彬, 李展平. *分析化学*, 2020, 48(11): 96-103.
- [28] SZYNKOWSKAM I, CZERSKI K, ROGOWSKI J, PARYJCZAK T, PARCZEWSKI A. *Forensic Sci. Int.*, 2009, 184(1-3): E24-E26.
- [29] SZYNKOWSKA-JÓŹWIK M I, MAĆKIEWICZ E, ROGOWSKI J, GAJEK M, PAWLACZYK A, DE PUIT M, PARCZEWSKI A. *Materials*, 2021, 14(21): 6243-6253.

- [30] THANDAUTHAPANI T D, REEVE A J, LONG A S, TURNER I J, SHARP J S. *Sci. Justice*, 2018, 58(6): 405-414.
- [31] MURAMOTO S, SISCO E. *Anal. Chem.*, 2015, 87(16): 8035-8038.
- [32] LI Wen-Jie, SUN Ling-Hui, YOU Wei, TIAN Shi-Si, WANG Li-Xue, ZHAO Ya-Bin, LI Zhan-Ping. *Chin. J. Anal. Chem.*, 2022, 50(1): 112-118.
李文杰, 孙令辉, 游伟, 田师思, 王利雪, 赵雅彬, 李展平. *分析化学*, 2022, 50(1): 112-118.
- [33] HINDER S J, WATTS J F. *Surf. Interface Anal.*, 2010, 42(6-7): 826-829.
- [34] MCDONNELL L A, HEEREN R M A. *Mass Spectrom. Rev.*, 2007, 26(4): 606-643.
- [35] PEARSON K. *Science*, 1909, 30(757): 23-25.
- [36] CARRASCO J L, JOVER L. *Statist. Med.*, 2005, 24(24): 4021-4034.
- [37] ZHANG Wen-Tong, DONG Wei. *SPSS Statistical Analysis of Advanced Tutorial (3rd edition)*. Beijing: Higher Education Press, 2018: 239-244, 344-353.
张文彤, 董伟. *SPSS统计分析高级教程(第三版)*. 北京: 高等教育出版社, 2018: 239-244, 344-353.

Personal Identification of Fingerprint Donors Based on Time-of-Flight Secondary Ion Mass Spectrometry

WU Yin¹, TIAN Shi-Si¹, LI Wen-Jie¹, SUN Ling-Hui¹, SONG Xi-Lin¹, ZHAO Ya-Bin^{*1},
LI Zhan-Ping^{*2,3}

¹(Academy of Forensic Science, People's Public Security University of China, Beijing 100038, China)

²(Analysis Center, Tsinghua University, Beijing 100084, China)

³(Key Laboratory of Organic Optoelectronics & Molecular Engineering of Ministry of Education,
Department of Chemistry, Tsinghua University, Beijing 100084, China)

Abstract Fingerprints, as one of the most important evidences, contain numerous personal information and are used for individual identification in the field of forensic science. However, in real case scenes, traditional way of comparison and matching with patterns may not be effective, as the detected fingerprints are mostly fuzzy or incomplete. Additionally, the compositions of fingerprint substance are so complicated that it is quite difficult to characterize donors according to some specific compounds. Time-of-flight secondary ion mass spectrometry (TOF-SIMS) can obtain the mass spectra of fingerprint substances and corresponding ion distribution images without destroying the patterns, which provides the possibility of identifying individuals based on fingerprint substances. In this study, the mass spectra of ten fingerprints from 5 volunteers were collected using TOF-SIMS to preliminarily explore the feasibility of identifying individuals based on ion signals of fingerprint substance. Firstly, 98 kinds of TOF-SIMS ion signals representing different fingerprint substances were screened out followed by the process of principal component analysis to reduce the dimension of intensity of ion signals. Afterwards, the newly saved variables after dimension reduction were input into different classifiers to compare the effects. The results showed that the effect of unsupervised hierarchical clustering was quite different from the actual classification, while supervised Fisher discriminant method could correctly discriminate 94% of the individuals and the accuracy of cross-validation also reached 90%. Meanwhile, more ideal results would be achieved when the data was input into the multi-layer perceptron neural network after multiple trainings. The individual classification and prediction models were established via various statistical analysis methods, which provided a new channel for improving the evidence value of fingerprints and laid the research foundation for further exploration of employing fingerprint substances to identify individuals.

Keywords Time-of-flight secondary ion mass spectrometry; Fingerprint substance; Personal identification; Multivariate statistical analysis

(Received 2022-05-23; accepted 2022-11-18)

Supported by the National Key R&D Program of China (No. 2018YFA0702600) and the General Science and Technology Program of the Ministry of Public Security (No. 2021JSM01).