

基于表面增强拉曼光谱的毒品鉴别方法

杨志超^{1,2}, 蔡 竞¹, 张 辉¹, 石 璐¹

1. 浙江警察学院刑事科学技术系, 浙江 杭州 310053

2. 毒品防控技术浙江省重点实验室, 浙江 杭州 310053

摘 要 毒品的快速检测在抑制毒品的传播, 打击毒品犯罪方面有着举足轻重的作用。表面增强拉曼光谱(SERS)技术具有指纹识别、检测速度快、样品用量少、无损伤等众多优点而受到了关注, 其特点特别适合于公安机关现场快速检测执法。本文利用金纳米粒子溶胶作为增强试剂对拉曼光谱进行增强, 制作 $1 \mu\text{g} \cdot \text{mL}^{-1}$ 的苯丙胺、氯胺酮、芬太尼、海洛因、可卡因和甲基苯丙胺六种溶液, 毒品溶液、增强试剂、NaCl 溶液的体积比为 20 : 6 : 5 混合, 取 $30 \mu\text{L}$ 混合溶液滴在载玻片表面, 在空气中自然挥干后待检。每类毒品溶液各制作 5 个样本, 每个样本随机采集 10 个点的拉曼光谱数据。6 种毒品溶液拉曼光谱数据共 300 组, 随机选择 60 组拉曼数据作为训练集, 利用训练集数据对模型进行训练。其余 240 组数据作为测试集, 测试模型的分类效果。经过预实验比较, 实验选择波长为 785 nm 激光作为激发光源, 采用 $50\times$ 物镜, 激光强度为 3.0 mW, 曝光时间为 0.2 s, 扫描次数为 1 000 次, 选取 $400\sim 1\,700 \text{ cm}^{-1}$ 波段测试研究。采用 Savitzky-Golay 方法对拉曼数据进行平滑降噪, 采用 airPLS 方法进行基线校正, 完成数据的 0-1 归一化。利用主成分分析法、方差筛选法、遗传选择算法、互信息法对数据降维处理, 通过支持向量机、随机森林、人工神经网络和最近邻四种算法分别进行建模训练, 并利用测试集数据测试模型分类效果, 重复 10 遍取平均准确率。结果表明, 拉曼光谱数据经过 PCA 降维后, 选取 5 个主成分, 各分类器准确率都在 95% 以上。另外三种波段选择方法中, 遗传选择算法结合 SVM 分类器准确率较高, 遗传选择算法筛选出的 5 个拉曼波段的组合, 分类准确率已达到 95% 以上, 25 个拉曼波段组合时, 准确率达到 99%。遗传选择算法作为波段选择算法, 不仅可以降低拉曼光谱采集数据的维度, 而且可解释性更强, 有更重要的意义, 为毒品的快速检测技术提供参考。

关键词 表面增强拉曼; 毒品; 分类模型

中图分类号: O433.4 **文献标识码**: A

DOI: 10.3964/j.issn.1000-0593(2022)04-1168-05

引 言

2019 年, 我国破获毒品相关案件 6.2 万起, 缴获毒品达 49.1 吨, 抓获犯罪嫌疑人 9 万名。214.8 万名现有吸毒人员中, 滥用冰毒人员占 55.2%, 滥用海洛因占 37.5%, 滥用氯胺酮占 2.3%^[1]。毒品检测技术是发现毒品、固定证据的重要手段, 传统的毒品检测方法主要有气相色谱质谱法^[2-3]、高效液相色谱法^[4-5]、毛细管电泳^[6]等, 此类检测方法过程复杂, 耗时长, 需要专门的实验环境和实验人员。拉曼光谱技术具有“指纹”光谱、无需前处理、分析速度快等优势, 适合犯罪现场检测^[7]。通过分析样品的拉曼谱图就可以得到样品的结构信息。但是, 绝大多数分子的拉曼散射截面非常小, 拉曼光谱强度较低, 使得拉曼很难有较低的检出限。对

于类似于毒品之类的痕量物质检测, 因为样品浓度低, 常规拉曼光谱无法被检测出来。表面增强拉曼散射(surface-enhanced Raman spectroscopy, SERS)利用金、银等贵金属颗粒制作的溶胶与备件样品混合后, 拉曼信号显著增强几个数量级, 表面增强拉曼技术因其灵敏度高、操作简单方便等优势受到检测技术人员的普遍关注, 并逐步实现特定应用场景的痕量检测^[8-11]。

2011 年, Farquharson 等^[12]使用固相萃取和 SERS 技术检测唾液中的药物, 该方法成功检测了唾液中 1 ppm 的苯丙胺, 地西洋, 美沙酮和哌啶, 可卡因的检测灵敏度高达 50 ppb。董荣录等^[13]在模拟交易或吸毒现场使用印刷的纸质基材检测毒品。同时, 使用金纳米棒来检测人类尿液中的药物。在支持向量机(SVM)的帮助下, 该方法成功地检测了真实吸毒者尿液样本中的药物, 准确率达 92% 以上。2018 年,

收稿日期: 2021-05-20, 修订日期: 2021-07-28

基金项目: 国家重点研发计划项目(2018YFC0807400), 浙江警察学院项目(2020XJY015)资助

作者简介: 杨志超, 1985 年生, 浙江警察学院刑事科学技术系讲师

e-mail: yangzhichao@zjpcxy.cn

Haddad 等^[14]使用银纳米粒子浸渍的纸质基底作为衬底,用于检测痕量芬太尼或是作为海洛因的掺杂物。同时,与每种物质相关的特征峰的强度比符合 Langmuir 等温线校正模型,可以用于海洛因混合物中芬太尼的定量分析。此外,用这些纸质 SERS 基底有助于从表面擦拭回收芬太尼,证明这是一种非常适用于犯罪现场调查的检测技术。2020 年,颜文杰等^[15]分别获取了海洛因、甲基苯丙胺与其他物质的 90 组毒品混合物光谱数据,建立基于支持向量机与多层感知器神经网络的融合分类模型。结果表明,基于高斯核函数、线性核函数、多项式核函数的 SVM 模型能够实现对不同质量分数海洛因混合品样本 97.8%, 97.8% 和 95.6% 的准确分类,多层感知器能够对甲基苯丙胺混合品样本实现 96.5% 的准确分类。

既往研究都是基于拉曼光谱的全光谱数据,由于数据量较大,一方面影响运算速度,另一方面由于冗余波段信息影响,可能造成分类准确率降低。通过光谱数据降维可以实现数据的压缩,更加节约运算资源。本文利用主成分分析法、方差阈值法、遗传选择算法和互信息法四种降维算法,将六种毒品拉曼光谱数据降维,利用最近邻(KNN)、支持向量机(SVM)、人工神经网络(ANN)和随机森林(RF)四种分类器对降维后的拉曼数据进行分类,讨论不同的降维方法及分类器对准确率的影响。

1 实验部分

1.1 样品

实验所用苯丙胺 (Amphetamine, AM)、海洛因 (Heroin)、可卡因 (Cocaine)、氯胺酮 (Ketamine, KET)、甲基苯丙胺 (Methamphetamine, METH)、芬太尼 (Fentanyl) 均由上海物证鉴定中心提供。制作 $1 \mu\text{g} \cdot \text{mL}^{-1}$ 六类毒品的水溶液,并加入纳米金增强试剂和 5% NaCl 溶液,毒品溶液、金溶胶增强试剂、NaCl 溶液的体积比为 20 : 6 : 5。取 $30 \mu\text{L}$ 混合溶液滴在载玻片表面,在空气中自然挥干后检测。每类毒品溶液各制作 5 个样本,每个样本采集 10 个拉曼光谱。6 种毒品溶液拉曼光谱数据共 300 组,随机选择 60 组拉曼数据作为训练集,其余 240 组数据作为测试集。

1.2 仪器与试剂

实验采用美国 Thermo Fisher 公司生产的 DXR2xi 显微激光拉曼成像光谱仪,拉曼光谱仪具有超低暗噪声,单光子信号探测器等优势。计算机环境为 Intel (R) Core (TM) i5-5200U CPU @ 2.2GHz, RAM: 12.0 GB, 64 位操作系统。实验采用柠檬酸钠还原法制备的金纳米颗粒 (Au NPs) 作为拉曼增强试剂,采购自厦门普识纳米科技公司,金纳米颗粒尺寸在 50 nm 左右,其电子显微图像如图 1。

1.3 拉曼光谱数据的获取与校正

利用拉曼光谱仪获取血痕的拉曼光谱,实验考查了不同的激发波长、物镜倍数、激光强度、曝光时间、扫描次数等采集参数,综合比较对样本的破坏、荧光干扰、拉曼信号信噪比、实验效率等方面。实验选择波长为 785 nm 激光作为激发光源,采用 50×物镜,激光强度为 0.7 mW,曝光时间

为 0.2 s,扫描次数为 1 000 次。实验采用迭代自适应加权惩罚最小二乘法校正基线,使用 S-G 平滑滤波实现平滑处理,选取 $400 \sim 1700 \text{ cm}^{-1}$ 波段测试研究。

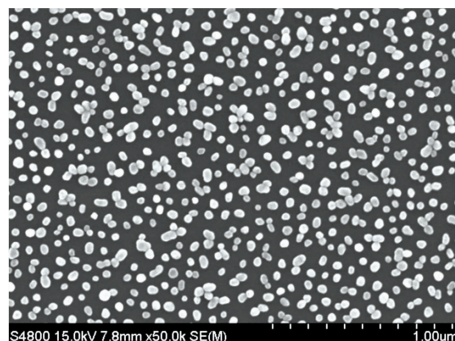
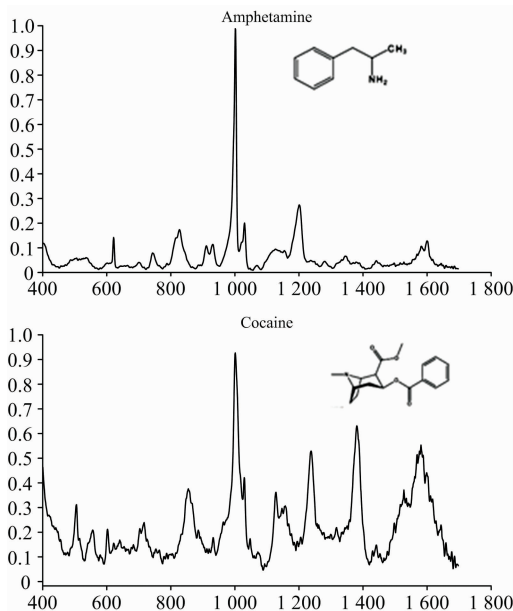


图 1 金纳米颗粒的电子显微图像

Fig. 1 Electron microscopic image of Au NPs

1.4 毒品的拉曼光谱特征

六种毒品的平均拉曼光谱如图 2 所示。苯丙胺拉曼 1024 cm^{-1} 位置归属为苯环呼吸振动, 835 cm^{-1} 的特征峰归属为 NH 键摇摆振动, 1201 cm^{-1} 处较弱的特征峰归属为 CH 摇摆振动。可卡因拉曼光谱 1707 cm^{-1} 特征峰归属于苯甲酸酯, 1591 cm^{-1} 特征峰归属于芳香环, 1024 cm^{-1} 归属为苯环呼吸振动。芬太尼拉曼光谱 810 cm^{-1} 特征峰归属于芳香环, 998 cm^{-1} 特征峰归属于苯环上的 C=C 伸缩振动, 1030 cm^{-1} 为最强的特征峰归属于 C=C 对称性伸缩振动, 1152 cm^{-1} 特征峰归属于 CH 反对称变形振动。海洛因拉曼光谱特征峰 1605 cm^{-1} 特征峰均归属于 C=C 对称性伸缩振动, 1023 和 1271 cm^{-1} 特征峰共同归属于 C—O—R 非对称性伸缩振动。氯胺酮拉曼光谱特征峰 455 cm^{-1} 归属为 C—C 变形振动, 1447 cm^{-1} 为 CH_3 和 CH_2 变形振动。甲基苯丙胺拉曼光谱特征峰位于 1024 cm^{-1} 位置归属为苯环呼吸振动, 835 cm^{-1} 位置的特征峰归属为 NH 键摇摆振动, CH_3 反对称变形振动在拉曼光谱上位于 1455 cm^{-1} 位置。



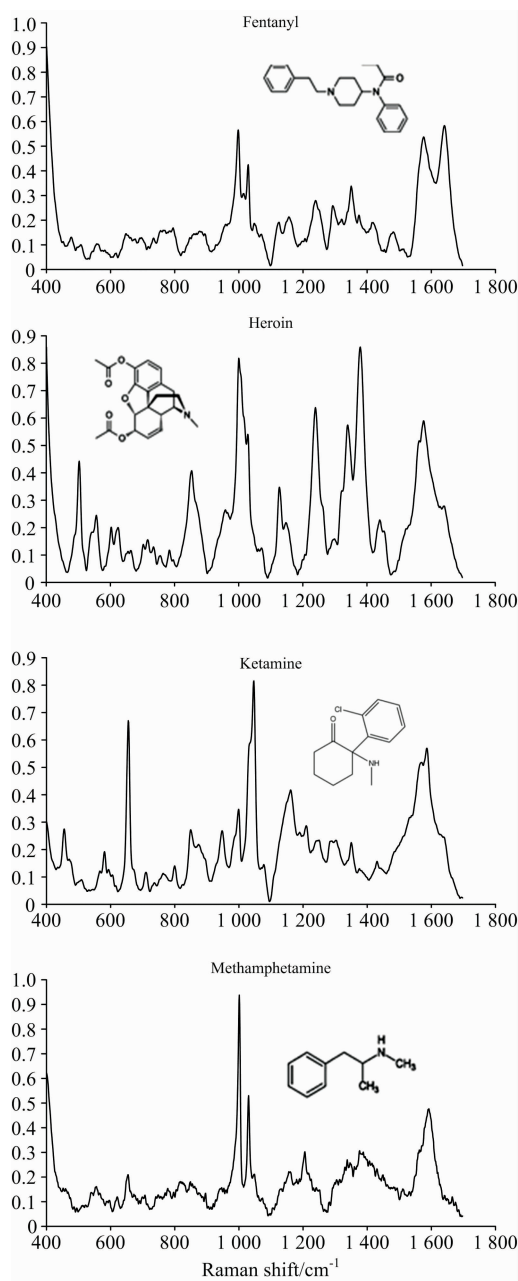


图 2 六类毒品溶液的拉曼光谱曲线

Fig. 2 Raman spectral curves of six drug solutions

2 结果与讨论

拉曼光谱数据特征较多,且大部分特征是冗余信息。这些冗余信息不仅对分类没有帮助,而且会浪费计算资源,降低分类效率,所以需要拉曼光谱数据进行降维。降维算法主要分为两种,一种是基于数学变换的方法,比如主成分分析。另一种是基于波段选择的方法,其特点在于保留了原来波段的物理特性,可解释性强。比如方差阈值法、遗传选择算法和互信息法。方差阈值法(Variance Threshold),是一种通过特征的方差值过滤方差的方法,计算每一个特征的方差,选择方差值最大的前 N 个波段,形成波段子集。遗传选

择算法(genetic algorithm, GA),模拟了生物种群的迭代进化原理,从一组随机的波段组合开始,通过交叉和变异过程,逐步迭代出最适合的波段组合。本实验中遗传选择算法主要参数:变异概率 2%,迭代次数 100 次,种群个体数为 200。互信息(mutual information, MI),通过计算每个波段的强度值与类别标签之间的互信息值,互信息值表示了两者的相关程度,选择互信息值最大的前 N 个波段,形成波段组合。

实验采用主成分分析法、方差阈值法、遗传选择算法和互信息法四种降维算法,将六种毒品拉曼光谱数据降维至 30 个特征以内。利用最近邻、支持向量机、神经网络和随机森林四种分类器对降维后的拉曼数据完成分类,讨论不同的降维方法及分类器的准确率。

2.1 毒品拉曼光谱主成分分析

主成分分析的前 3 主成分空间分布如图 3 所示。从拉曼光谱数据的前 3 主成分在空间的分布可知,六类毒品样本的分离性较好。前 K 个主成分特征的方差贡献率如图 4 所示,可知前 5 个主成分的方差贡献率已达 80%,取前 5 个主成分即应得到较好的分类准确率表现。

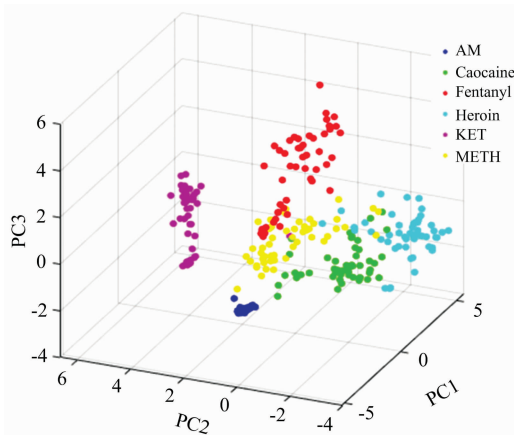


图 3 六类毒品溶液的前三个主成分分布图

Fig. 3 Principal components analysis score plot of samples

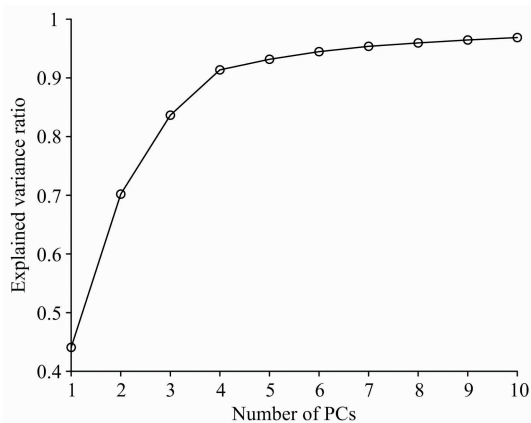


图 4 主成分特征方差贡献率

Fig. 4 Principal component feature explained variance contribution ratio

2.2 降维与建模结果

利用 4 种降维方法压缩特征数量后, 分别采用 KNN, SVM, ANN, RF 分类器重复 10 遍, 其准确率表现如图 5 所示。其中, 拉曼光谱数据经过 PCA 方法降维后, 各分类器的准确率均较高。在选取的主成分为 5 时, 各分类器准确率都在 95% 以上。这主要是由于 PCA 算法是一种基于数学变换的降维算法, 变换后的每一个主成分特征包含了所有波段的组合信息, 所以 PCA 降维后的分类效果好。在三种波段选择

方法中, 遗传选择算法得到的波段组合准确率相对较高。结合 SVM 分类器, 遗传选择算法筛选出的 5 个拉曼波段的组合, 分类准确率已达到 95% 以上, 在选择的波段数达到 30 时, 准确率达到 99.5%。在取 25 个以上的波段时, 方差阈值法准确率达到 95% 以上。互信息法确定的波段子集的分类准确率较低, 特别是在波段数量大于 15 后, 准确率还有下降。可能是由于, 在波段数量大于 15 后, 互信息法选择了无价值的冗余波段, 致使准确率下降。

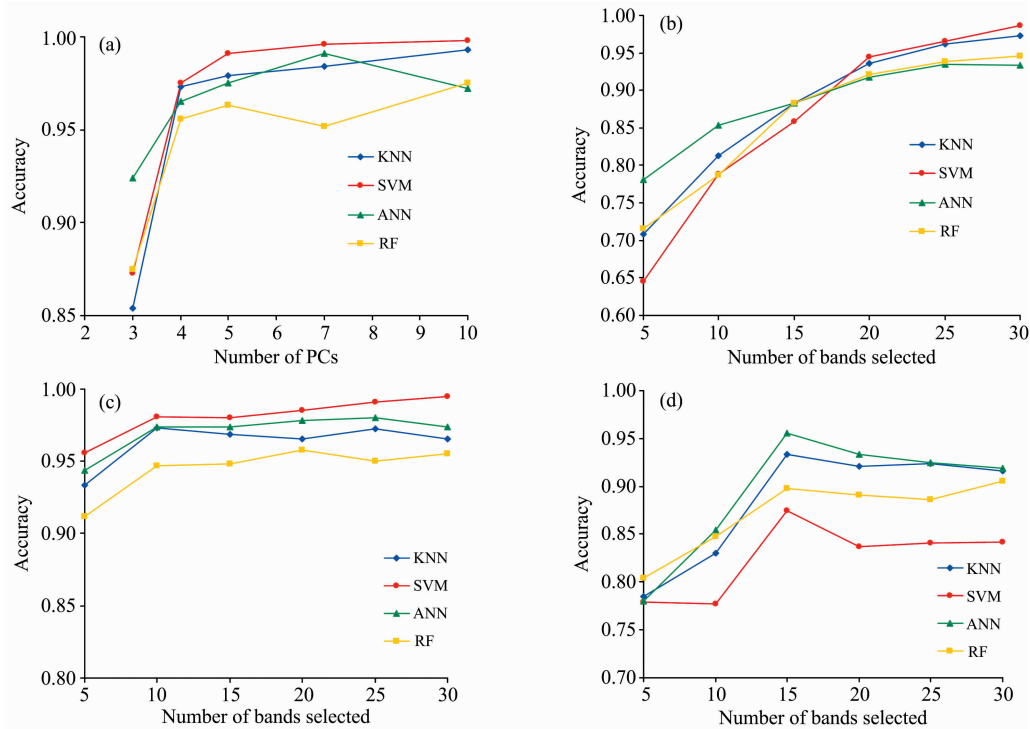


图 5 不同的降维方法在测试集中的准确率比较

(a): PCA 算法; (b): 方差阈值法; (c): 遗传选择算法; (d): 互信息法

Fig. 5 Comparison of accuracy of different dimensionality reduction methods in testing set

(a): PCA; (b): Variance Threshold; (c): GA; (d): Mutual Information

3 结论

在表面拉曼光谱技术区分毒品种类方面, 将拉曼光谱数据降维后再进行分类, 依然保持了较高的分类准确率, 降维方法在毒品拉曼光谱分类上体现出有效性。在主成分分析法、方差阈值法、遗传选择算法和互信息法四种降维算法

中, 基于数学变换的主成分分析降维方法效果最好, 在降维至 5 个特征时, 各分类算法的准确率依然能达到 95% 以上。其他三种波段选择算法中, 遗传选择算法筛选特征的效果最好, 在波段数为 20, 25, 30 时, SVM 算法的分类准确率分别达到 98.5%, 99.1%, 99.5%。遗传选择算法作为波段选择算法, 不仅可以降低拉曼光谱采集数据的维度, 而且可解释性更强, 有更重要的意义。

References

- [1] Drug Situation in China (2019)(2019 年中国毒品形势报告). China Police Daily(人民公安报), 2020-06-25(002).
- [2] ZHANG Yi-chen, CHEN Xue-guo, ZHAO Dan(张一辰, 陈学国, 赵丹). Journal of Instrumental Analysis(分析测试学报), 2021, 40(4): 560.
- [3] Seymour Craig, Shaner Rebecca L, Feyereisen Melanie C, et al. Journal of Analytical Toxicology, 2019, 43(4): 266.
- [4] Sara Odoardi, Valeria Valentini, et al. Microchemical Journal, 2017, 133: 302.
- [5] Fabio Vaiano, Busardo Francescop, et al. Journal of Pharmaceutical and Biomedical Analysis, 2016, 129: 441.
- [6] LI Xiao-juan, LIU Xin, YANG Li(李晓娟, 刘心, 杨丽). Journal of Molecular Science(分子科学学报), 2016, 32(4): 334.

- [7] Penido C A F D, Pacheco M T T, Lednev I K, et al. *Journal of Raman Spectroscopy*, 2016, 47(1) : 28.
- [8] Mirsafavi Rustin, Moskovits Martin, Meinhart Carl. *Analyst*, 2020, 145(9): 3440.
- [9] Fedick Patrick W, Pu F, Morato N M, et al. *Journal of the American Society for Mass Spectrometry*, 2020, 31(3): 735.
- [10] Wang Ling, Deriu C, Wu W, et al. *Journal of Raman Spectroscopy*, 2019, 50(10): 1405.
- [11] Shende Chetan, Brouillette Carl, Farquharson Stuart. *Analyst*, 2019, 144(18) : 5449.
- [12] Frank Inscore, Chetan Shende, Atanu Sengupta, et al. *Applied Spectroscopy*, 2011, 65(9): 1004.
- [13] DONG Rong-lu, LI Shao-fei, LIN Dong-yue(董荣录, 李绍飞, 林东岳). *Scientia Sinica Chimica(中国科学: 化学)*, 2021, 51(3): 294.
- [14] Haddad Abed, Comanescu Mircea Alex, Green Omar, et al. *Analytical Chemistry*, 2018, 90(21): 12678.
- [15] YAN Wen-jie, LU Wen-hui, WANG Ji-fen(颜文杰, 卢雯慧, 王继芬). *Laser & Optoelectronics Progress(激光与光电子学进展)*, 2021, 58(14): 1404003.

Drug Classification Method Based on Surface-Enhanced Raman Spectroscopy

YANG Zhi-chao^{1, 2}, CAI Jing¹, ZHANG Hui¹, SHI Lu¹

1. Zhejiang Police College, Hangzhou 310053, China

2. Key Laboratory of Drug Prevention and Control Technology of Zhejiang Province, Hangzhou 310053, China

Abstract Rapid detection of drugs plays an important role in restraining the spread of drugs and cracking down on drug crimes. Surface Enhanced Raman Spectroscopy (SERS) technology has many advantages such as fingerprint identification, fast detection speed, less sample consumption, no damage, high sensitivity and so on, which has attracted much attention. Its characteristics are especially suitable for the rapid detection and law enforcement of public security organs on the spot. This paper used gold nanoparticle sol as the enhancement reagent to enhance the Raman spectrum. Six solutions of amphetamine, ketamine, fentanyl, heroin, cocaine and methamphetamine were prepared by $1 \mu\text{g} \cdot \text{mL}^{-1}$. The volume ratio of drug solution, enhancement reagent and NaCl solution was 20 : 6 : 5, and $30 \mu\text{L}$ of the mixed solution was dropped on the surface of the slide. Let dry in the air and wait for inspection. Five samples were made for each drug solution, and Raman spectral data of 10 points were randomly collected for each sample. 300 groups of Raman spectral data of 6 drug solutions were collected, and 60 groups of Raman data were randomly selected as the training set. The model was trained by using the training set data. The remaining 240 groups of data were used as test sets to test the classification effect of the model. After pre-experiment comparison, 785 nm laser was selected as the excitation light source in the experiment, $50\times$ objective lens was used, the laser intensity was 3.0 mW, the exposure time was 0.2 seconds, and the scanning times were 1 000 times. The bands from 400 to $1\,700 \text{ cm}^{-1}$ were selected for test and research. Savitzky-Golay method was used for smoothing and de-noising Raman data, and the airPLS method was used for baseline correction to complete 0–1 normalization of data. Principal component analysis (PCA), variance screening, genetic selection algorithm and mutual information method were used to process the dimensioning of the data. Modeling training was carried out by the four support vector machine algorithms, random forest, artificial neural network and nearest neighbor respectively. The classification effect of the model was tested by the test set data, and the average accuracy was obtained by repeating 10 times. The results show that the accuracy of all classifiers is more than 95% when the principal component is 5, after the dimension reduction of Raman spectral data by the PCA method. Among the three bands selection methods, the combination of genetic selection algorithm and SVM classifier has higher accuracy. The classification accuracy of the combination of 5 Raman bands screened by the genetic selection algorithm has reached more than 95%, and the accuracy of the combination of 25 Raman bands has reached 99%. As a band selection algorithm, genetic selection algorithm can reduce the dimension of Raman spectral data collection and have stronger interpretation and more important significance, which provides a reference for the rapid detection technology of drugs.

Keywords SERS; Drug classification; Classification model

(Received May 20, 2021; accepted Jul. 28, 2021)