

# 基于光谱和色谱数据碰撞融合策略的大麻油快速识别分类

李佳瑞<sup>1</sup>, 王继芬<sup>1\*</sup>, 范琳媛<sup>1</sup>, 石学军<sup>2</sup>

<sup>1</sup>中国人民公安大学侦查学院, 北京 100038;

<sup>2</sup>北京海关缉私局司法鉴定中心, 北京 100000

**摘要** 实现药物快速无损准确性分析是药物防控视域下的关键问题。采集了 4 类共计 159 份大麻油样本的光谱和色谱数据, 利用有监督模式识别方法 Fisher 判别分析和 K 邻近分析构建不同数据集分类模型, 对比了单一模型与融合模型对分析结果的影响。结果表明, 在 4 类大麻油识别与分类过程中, 基于光谱与色谱融合数据集的分类模型分类效果优于其他数据集, 借助 K 邻近分析的融合模型分类效果最佳, 总体分类准确率达到 1。该研究实现了对不同类型大麻油快速准确的定性分析, 为准确认定缴获药物的来源、审理事实关联的上下游药物犯罪案件提供线索和证据。

**关键词** 光谱融合; 大麻油; Fisher 判别分析; K 邻近分析

中图分类号 O433.4

文献标志码 A

DOI: 10.3788/LOP202259.1630004

## Rapid Identification and Classification of Cannabis Oil Based on Data Fusion of Spectroscopy and Chromatography

Li Jiarui<sup>1</sup>, Wang Jifen<sup>1\*</sup>, Fan Linyuan<sup>1</sup>, Shi Xuejun<sup>2</sup>

<sup>1</sup>School of Investigation, People's Public Security University of China, Beijing 100038, China;

<sup>2</sup>Forensic Expertise Center of Beijing Customs Anti-Smuggling Bureau, Beijing 100000, China

**Abstract** The rapid, non-destructive, and quasi-deterministic analysis for drugs is a critical issue in the field of drug prevention and control. The experiment sorted the spectral and chromatographic data from 159 cannabis oil samples representing four types. We establish classification models for different data sets using the supervised pattern recognition methods of Fisher discriminant analysis and K-nearest neighbor analysis, and then compare the effects of single and fusion models on the analysis results. According to the results, in the process of identification and classification of four types of cannabis oil, the fusion model based on spectral and chromatographic data sets has a higher classification effect than other data sets, and the fusion model based on K-nearest neighbor analysis has achieved the best classification effect, with an overall classification accuracy of 1. This research enables the rapid and accurate qualitative analysis of different types of cannabis oil, as well as providing evidence and clues for accurately identifying the source of seized drugs and trying upstream and downstream drug crime cases related to facts.

**Key words** spectral fusion; cannabis oil; Fisher discriminant analysis; K-nearest neighbor analysis

## 1 引言

作为传统毒品的大麻, 毒性相对较小, 因价格低廉、摄取方便成为了世界上滥用的毒品。目前能够从大麻植物中提取的大麻酚类及其衍生物超过 60 种, 四氢大麻酚 (THC) 和大麻二酚 (CBD) 是其中最主要的精神活性成分, 具有法庭意义<sup>[1]</sup>。目前可将大麻植物

分为工业大麻 (THC 含量小于 0.3%)、药用大麻 (THC 含量为 0.3%~0.5%) 和毒品大麻 (THC 含量大于 0.5%)。为了避免违反相关法律, 药企们开始从许可的工业大麻入手寻找突破口, 目前含有 CBD 的食品、化妆品和保健品等新兴产品受到了广泛关注。但是 CBD 制品除了会影响其他药物的代谢动力学之外, 由于精油制品中的有效成分含量不确定, 服用剂量超

收稿日期: 2021-05-26; 修回日期: 2021-06-06; 录用日期: 2021-07-13

基金项目: 中央高校基本科研业务费专项资金 (2021JKF208)

通信作者: \*wangjifen58@126.com

过人体安全范围会带来未知风险<sup>[2-4]</sup>。更有不法分子走私大麻油等相关管制药物,试图以大麻二酚作为主要成分掩饰其中的四氢大麻酚,规避海关检查。目前除了通过行李物品夹藏等方式从空港旅检渠道走私入境,网络和物流邮寄渠道活动突出。不法分子常通过快件、空运货物等邮寄渠道走私管制药物入境。建立快速、可靠、简单、准确和经济有效的检验非法药物的方法对精准打击违法犯罪具有重要意义。

光谱检验技术提供了一种快速灵敏而又无损材料的有效方法,成为了法庭科学领域有利的分析工具,能够适用于药物成分的检验分析<sup>[5-7]</sup>。但在实际检验中,常遇到多种药物混杂的情况,光谱技术由于没有分离功能,针对各成分进行检验就存在一定的局限性。色谱与质谱联用分析技术由于具有灵敏度高、检测限低、分离效果好等优点常用于滥用药物的定性和定量分析,可为案件侦破与审判诉讼提供大量信息<sup>[8-10]</sup>。将光谱与色谱数据相结合可以对彼此的优势进行互补,从不同角度与层次反映样品间的差异,为模型的构建提供了更全面的信息。Wang 等<sup>[11]</sup>研究发现,光谱与色谱数据集融合充分利用了信息协同的优势,表现出良

好的性能。基于融合数据集构建的分类模型对地理原产地准确鉴定具有可行性,此方法对食品的质量监控和产品溯源具有重要意义。高冰等<sup>[12]</sup>基于数据融合的方法对不同植物来源的食用油进行分析,研究发现数据碰撞融合策略对不同来源的复杂样本具有较好的包容性。借助数据碰撞融合策略,构建可视化研究模型,挖掘实验数据中隐藏的信息价值,从而提升涉案药物的检验效率,强化证据认定能力。

目前针对缴获大麻油的快速无损检验鉴别鲜有报道,本研究对 4 类大麻油的红外光谱和气相色谱进行采集,通过主成分分析(PCA)提取不同波段光谱数据的特征后进行数据降维,考察并比较了基于 Fisher 判别分析和 K 邻近分析的单一模型与融合模型的识别分类结果,从而实现大麻油在内的诸多现场缴获管制类药物的快速检验与定性分析。

## 2 材料与方法

### 2.1 样本数据采集方法

如表 1 所示,基于 2021 年 2 月至 5 月期间实际案件,收集了 4 类成分不同但都含有大麻二酚和四氢大麻酚的大麻油样本,共 159 份。

表 1 不同种类大麻油样品信息

Table 1 Information of cannabis oil samples with different species

Name	Ingredient
IMMUNO BOOSTER	MCT Oil, Echinacea Extract, CBD, THC
NRG BOOSTER	MCT Oil, Ashwagandha, Maca Root, Citral-Terpene Blend, CBD, THC
RELAX	MCT Oil, Dang Quai Extract, Red Clover Extract, CBD, THC
WELLNESS	MCT Oil, Passion Flower, Myrcene-Terpene Blend, CBD, THC

#### 2.1.1 红外光谱采集

采用 Nicolet is10 型傅里叶变换红外光谱仪(Thermo Fisher Scientific 公司),ATR 附件中为金刚石 HATR 晶体,KBr(Thermo Fisher Scientific 公司)作为分束器,使用 OMNIC 光谱数据处理软件,扫描次数为 64,分辨率为  $2\text{ cm}^{-1}$ ,光谱采集范围为  $4000\sim 650\text{ cm}^{-1}$ 。图 1 为 4 类大麻油的红外光谱图,每个样品

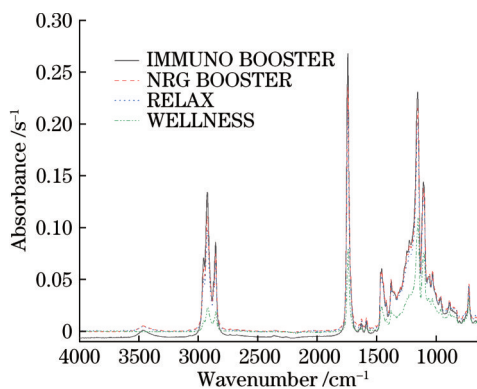


图 1 大麻油的红外光谱

Fig. 1 Infrared spectra of cannabis oil

光谱曲线采集 5 次,取均值为实验样本数据。

#### 2.1.2 气相色谱采集

7820 型气相色谱仪(美国安捷伦公司)色谱柱使用石英毛细管柱(尺寸为  $17\text{ m}\times 0.2\text{ mm}\times 0.11\text{ }\mu\text{m}$ ),进样量为  $1\text{ }\mu\text{L}$ ;进样口温度为  $250\text{ }^\circ\text{C}$ ;检测温度为  $260\text{ }^\circ\text{C}$ ;载气为氦气,恒压模式,压力为  $85\text{ kPa}$ ;分流比为  $60:1$ 。程序升温:初始温度为  $60\text{ }^\circ\text{C}$ ,以  $15\text{ }^\circ\text{C}/\text{min}$  速率升温至  $280\text{ }^\circ\text{C}$ ,保持  $27\text{ min}$ 。

### 2.2 数据处理方法

#### 2.2.1 数据预处理

以 70% 的样本共 111 份作为训练集,30% 的样本共 48 份作为测试集,构建分类模型。对采集得到的图谱信息采用基线校正、峰面积归一化<sup>[13-16]</sup>进行处理,汇总整理光谱数据集并进行标准化处理。通过 PCA 对复杂的数据结构进行简化,对冗余的数据进行降维处理的同时保留数据最具代表性的特征。构建借助不同分析方法的基于光谱数据集的单一模型与对光谱与色谱数据集分别经 PCA 降维后提取到的特征值进行串联的融合模型。

## 2.2.2 建模原理

### 1) Fisher判别分析

Fisher判别分析(FDA)的目的是找到一条对训练点进行投影后可以对不同类数据进行有效区分的直线。不同类型数据间的距离越大且该类数据内部距离越小,分类效果越好<sup>[17-18]</sup>。以二分类为例, $N$ 个样本组成训练集  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ , 其中每个样本  $\mathbf{x} = (x_1, x_2, \dots, x_d)^\top$  是一个  $d$  维向量; 第一类样本为  $\mathbf{X}_1 = \{x_1^1, \dots, x_{N_1}^1\}$ , 第二类样本为  $\mathbf{X}_2 = \{x_1^2, \dots, x_{N_2}^2\}$ ; 投影方向  $\mathbf{w} = (w_1, w_2, \dots, w_d)^\top$ , 同样为  $d$  维向量。在投影下, 样本转换为

$$y_i = \mathbf{w}^\top \mathbf{X}_i, \quad i = 1, 2, \dots, N. \quad (1)$$

投影前, 原样本空间中类均值向量  $\mathbf{m}_i$  为

$$\mathbf{m}_i = \frac{1}{N_i} \sum_{x_j \in X_i} \mathbf{x}_j, \quad i = 1, 2. \quad (2)$$

定义各类型的内部离散度矩阵  $\mathbf{S}_i$  为

$$\mathbf{S}_i = \sum_{x_j \in X_i} (\mathbf{x}_j - \mathbf{m}_i)(\mathbf{x}_j - \mathbf{m}_i)^\top, \quad i = 1, 2. \quad (3)$$

总类内离散度矩阵  $\mathbf{S}_w$  为

$$\mathbf{S}_w = \mathbf{S}_1 + \mathbf{S}_2. \quad (4)$$

各类型之间的离散度矩阵  $\mathbf{S}_b$  为

$$\mathbf{S}_b = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^\top. \quad (5)$$

以判别标准 Rayleigh 商  $\max J_F(\mathbf{w}) = \frac{\mathbf{w}^\top \mathbf{S}_b \mathbf{w}}{\mathbf{w}^\top \mathbf{S}_w \mathbf{w}}$  确定

最佳投影方向为  $\mathbf{w}^* = \mathbf{S}_w^{-1}(\mathbf{m}_1 - \mathbf{m}_2)$ , 最终通过判断  $g(\mathbf{x}) = \mathbf{w}^{*\top} \mathbf{x} + \mathbf{w}_0$  的值是否大于或小于 0 来进行分类。

其中阈值  $\mathbf{w}_0 = -\frac{1}{2}(\tilde{\mathbf{m}}_1 + \tilde{\mathbf{m}}_2)$ 。

### 2) K 邻近分析

K 邻近分析(KNN)参考样本所有已知的类型, 计算未知样本与已知样本之间的位置, 选择与未知样本距离最近的  $K$  个已知样本, 最终未知样本归属于这  $K$  个样本中占比较大的类型<sup>[19-21]</sup>。需要一个距离函数以计算两个样本之间的距离, 以计算二维空间中的  $A(x_1, y_1)$ 、 $B(x_2, y_2)$  两点之间的距离为例。欧几里得距离表达式为

$$D_{\text{Euclidean}} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}, \quad (6)$$

选用适用于连续变量的欧氏距离作为距离的度量, 简单来说 KNN 通过最邻近的一个或几个样本的类型判断未知样本的类型, 此类分析方法中样本间距离的计算具有重要的意义。

## 2.3 模型评价标准

通过计算灵敏度、特异性、精密度和准确率等相关参数对模型进行评价。参数值越接近 1, 分类效果越好<sup>[22]</sup>。参数的表达式分别为

$$P_{\text{sensitivity}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}}, \quad (7)$$

$$P_{\text{specificity}} = \frac{N_{\text{TN}}}{N_{\text{TN}} + N_{\text{FP}}}, \quad (8)$$

$$P_{\text{precision}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}}, \quad (9)$$

$$P_{\text{accuracy}} = \frac{N_{\text{TP}} + N_{\text{TN}}}{N_{\text{TP}} + N_{\text{FP}} + N_{\text{TN}} + N_{\text{FN}}}, \quad (10)$$

式中:  $N_{\text{TP}}$  为真阳性样本数;  $N_{\text{TN}}$  为真阴性样本数;  $N_{\text{FP}}$  为假阳性样本数;  $N_{\text{FN}}$  为假阴性样本数。

## 3 结果与分析

### 3.1 大麻油光谱分析

图 2 为各样本的红外光谱图。各样本在波数为 1500~650  $\text{cm}^{-1}$  时峰形有所差异, 在波数 1200  $\text{cm}^{-1}$  附近有较强尖峰, 主要是酚类 C—O 单键伸缩振动引起的, 其余范围内有多个中强峰, 主要是萜烯类不饱和烃基=C—H 弯曲振动引起的; 波数 3000~2750  $\text{cm}^{-1}$  范围内有两个尖峰, 主要是醛类羰基 C—H 单键伸缩振动引起的; 在 1750  $\text{cm}^{-1}$  附件出现明显的特征峰, 主要是醛酸类 C=O 双键伸缩振动引起的; 总体出峰位置相似, 所含物质成分结构相似, 但峰强度有所不同, 各类物质中相同成分含量有所差异。总体峰形较为混杂, 难以依靠图谱直观地进行区分。

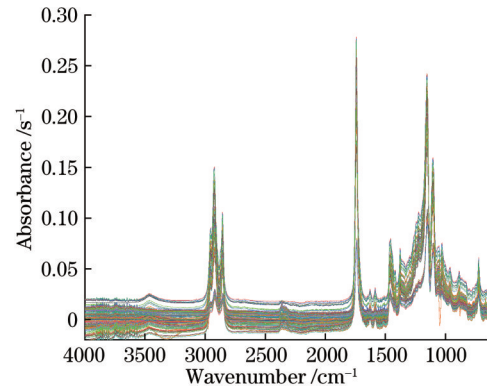


图 2 全部样本的红外光谱

Fig. 2 Infrared spectra of all samples

### 3.2 Fisher判别分析分类模型

分别构建基于 4 类麻油光谱数据集的单一模型与基于光谱和色谱数据集的融合模型, 比较它们借助 FDA 的分类结果。考察了灵敏度、特异性、精密度和准确率等相关参数, 由表 2 可知, 基于光谱数据集的单一模型中各类灵敏度与特异性均不低于 0.669, 精密度最高达 0.853, 总体分类准确率可达 0.818。基于光谱与色谱数据集的融合模型中各类灵敏度与特异性均不低于 0.951, 精密度最高达 0.992, 比起单一模型, 融合模型各参数值更接近 1, 总体分类准确率达 0.987, 分类结果更优。

图 3 为 FDA 的混淆矩阵, 图 3(a) 为基于光谱数据集的单一模型, 图 3(b) 为基于光谱与色谱数据的融合

表 2 FDA 模型参数结果  
Table 2 Parameter results of FDA model

Type	Data set	Sensitivity	Specificity	Precision	Accuracy
FTIR	IMMUNO BOOSTER	0.809	0.912	0.853	0.818
	NRG BOOSTER	0.809	0.912	0.853	
	RELAX	0.856	0.669	0.729	
	WELLNESS	0.833	0.862	0.836	
FTIR+GC	IMMUNO BOOSTER	0.983	1.000	0.992	0.987
	NRG BOOSTER	0.983	1.000	0.992	
	RELAX	0.983	1.000	0.992	
	WELLNESS	1.000	0.951	0.974	

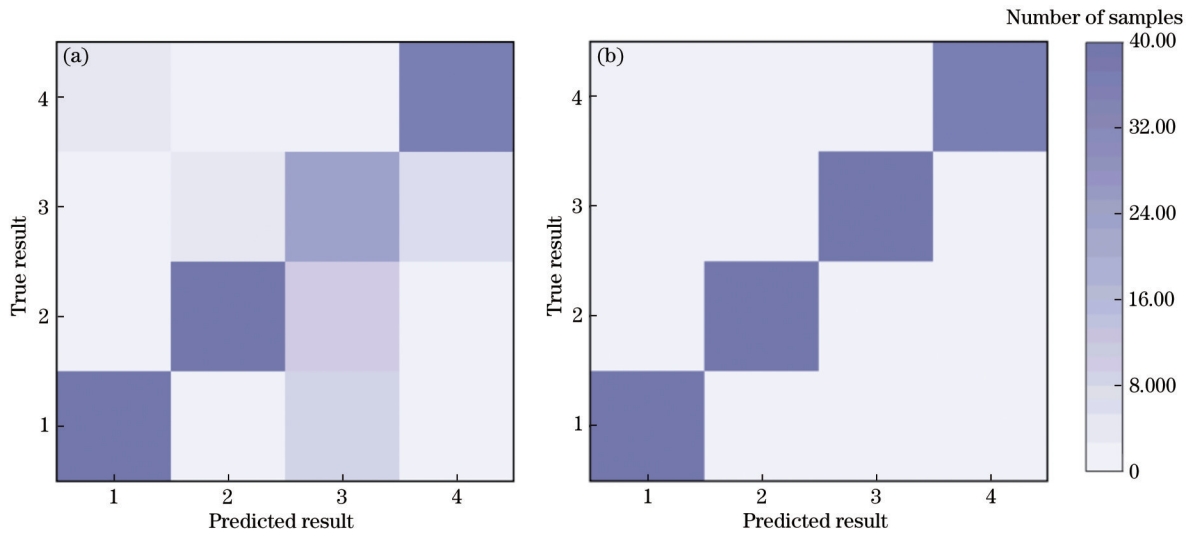


图 3 FDA 混淆矩阵。(a)单一模型；(b)融合模型

Fig. 3 FDA confusion matrix. (a) Single model; (b) fusion model

模型。坐标 1 代表大麻油 IMMUNO BOOSTER 类，2 代表大麻油 NRG BOOSTER 类，3 代表大麻油 RELAX 类，4 代表大麻油 WELLNESS 类。颜色越深，证明该类样本数越多。图 3(a)中，除真实分类结果与预测分类结果相符的对角线外，真实分类结果与预测分类结果不相符的部分颜色也较为明显，仍然有样本真实分类结果与预测结果不符。图 3(b)中几乎所有样本都可以得到准确分类，相比单一模型，融合模型分类效果更好。分析认为，融合模型结合了光谱与色谱数据的优势，对单一数据集进行了补充，对特征进行了进一步统计分析，使相似成分能够得到有效区分，为模型的构建提供了更全面的信息。

图 4 为借助 Fisher 判别分析的基于光谱和色谱融合数据集分类模型所得到的 4 类大麻油的空间分布。不同类型样品在不同判别轴上分布差异性大，其中 WELLNESS 类与 NRG BOOSTER 类判别空间聚敛度较为集中；NRG BOOSTER 类和 IMMUNO BOOSTER 类与 RELAX 类空间位置相距较远，彼此之间差异性较大；WELLNESS 类与 NRG BOOSTER 类空间位置相距较近，两类大麻油相似度较高。结合 FDA1、FDA2 与 FDA3 判别轴，Fisher 判别分析对所

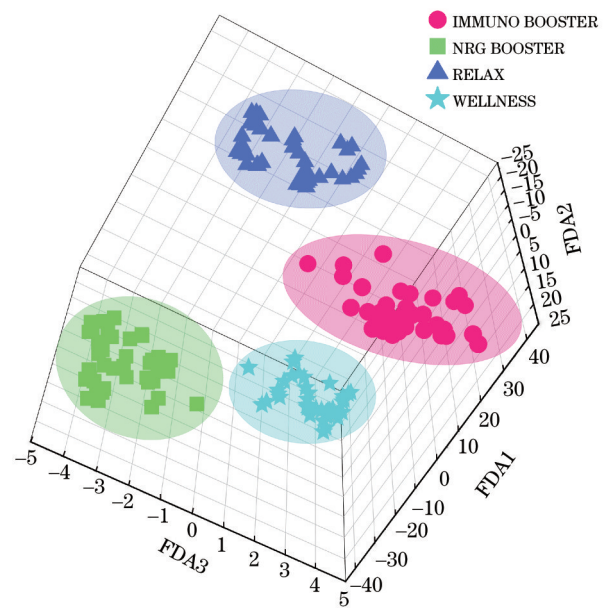


图 4 大麻油的空间分布

Fig. 4 Spatial classification detail of cannabis oil

有大麻油样品的分类准确率达到 0.987。Fisher 判别分析计算较为繁琐，没有考虑各类样本的出现概率，无

法估计错误率和后验概率,虽然最大程度地保留了分类信息,但依旧可能会造成信息损失,从而影响精度<sup>[23-25]</sup>。

### 3.3 K 邻近分析分类模型

借助 KNN 对经 PCA 降维后的光谱数据集进行分析,提取特征值进行降维处理,选择适用于建模的数据维度,除去冗余变量以提高算法的准确度。如图 5 所示,纵坐标为模型中提取的特征值,横坐标特征值重要性反映了此特征值对整个数据集的代表程度,特征 2 对整个数据集的分类准确率贡献性最强。

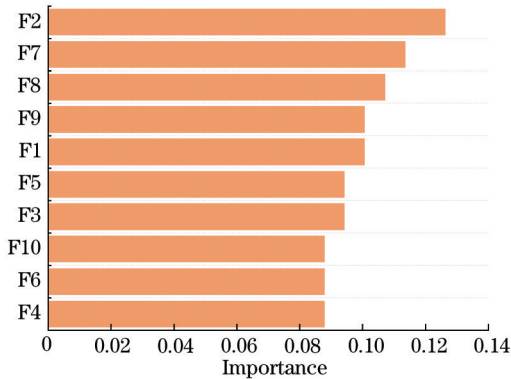


图 5 特征值重要性

Fig. 5 Eigenvalue importance

KNN 中,对 K 值的选择直接影响分类的结果。若 K 值过大,距离较远的不同类别的训练点就可能成为邻点而影响预测结果,使得偏差变大,准确率降低,模型容易欠拟合。图 6 为该模型 K 值的选择错误率。K=1 时选择错误率为最小值 0.1838,因此选择 1 为最优 K 值,在此条件下构建基于 4 类大麻油光谱数据集的 KNN 分类模型,分析发现,总体识别率达 0.842。在此模型基础上进一步考察基于 4 类大麻油光谱与色谱融合数据集的分类准确率。

使用 KNN 对光谱和色谱数据集融合模型进行分析,对比单一模型与融合模型对 4 类大麻油分类预测

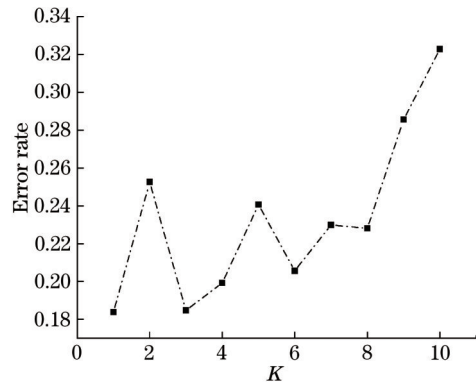


图 6 K 值选择错误率

Fig. 6 K value selection error rate

准确率。由表 3 数据能清晰地看到,基于光谱数据集的单一模型中各类灵敏度与特异性均不低于 0.758,精密度最高达到 0.871,总体分类准确率可达到 0.842。基于光谱与色谱数据集的融合模型中各类灵敏度、特异性、精密度和准确率等参数均可达到 1,分类结果理想。

图 7 为 KNN 的混淆矩阵,图 7(a)为基于光谱数据集的单一模型,图 7(b)为基于光谱与色谱数据的融合模型。坐标 1 代表大麻油 IMMUNO BOOSTER 类,2 代表大麻油 NRG BOOSTER 类,3 代表大麻油 RELAX 类,4 代表大麻油 WELLNESS 类。图 7(a)中除真实分类结果与预测分类结果相符的对角线外,真实分类结果与预测分类结果不相符的部分也有颜色分布;图 7(b)中所有样本都可以得到准确分类,融合模型比起单一模型分类结果更优。与图 3 相比较可以看出,KNN 比 FDA 的分类准确率更高,具有更强的分类能力。KNN 原理简单,易于实现。Abu 等<sup>[26]</sup>使用不同距离量度,基于实际数据集进行测试,以准确性、精确度和召回率来评估 KNN 的性能。实验结果表明,KNN 的核心主要取决于训练集与测试集之间的距离或相似度。Nurdiawan 等<sup>[27]</sup>通过比较不同算法对潮湿土壤进行分类的准确率发现,由于 KNN 主要通过附近

表 3 KNN 模型的参数结果

Table 3 Parameter results of KNN model

Type	Parameter	Training set				Test set			
		IMMUNO BOOSTER	NRG BOOSTER	RELAX	WELLNESS	IMMUNO BOOSTER	NRG BOOSTER	RELAX	WELLNESS
FTIR	Sensitivity	0.822	0.807	0.822	0.811	0.840	0.824	0.849	0.849
	Specificity	0.758	0.947	0.764	0.839	0.854	0.932	0.789	0.789
	Precision	0.786	0.864	0.790	0.824	0.847	0.871	0.816	0.816
	Accuracy	0.731	0.957	0.741	0.846	0.857	0.941	0.769	0.842
FTIR+GC	Sensitivity	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Specificity	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Precision	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Accuracy	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

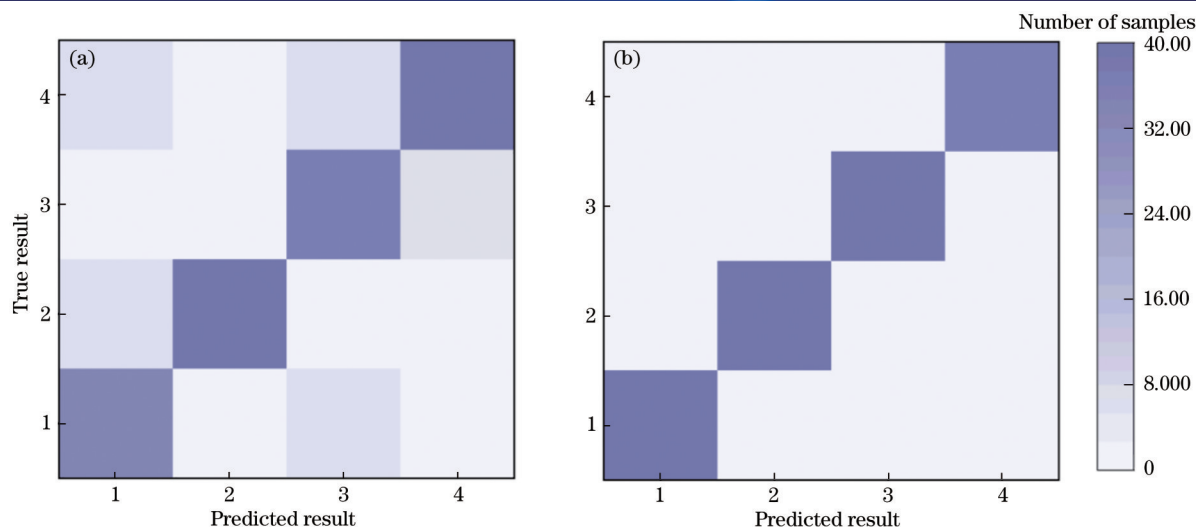


图7 KNN混淆矩阵。(a)单一模型;(b)融合模型

Fig. 7 KNN confusion matrix. (a) Single model; (b) fusion model

样本点而不是判别类别来确定所属的类别,所以更加适用于类域交叉或重叠较多的数据集。与FDA相比,KNN对数据没有假设、对异常点不敏感、分类准确率更高。分析认为KNN因分类原理,对分布均匀的样本数据集来说较其他方法更为适合<sup>[28-30]</sup>。以上实验表明,借助KNN的基于光谱和色谱融合的数据集模型为最佳分类模型。

## 4 结 论

借助Fisher判别分析和K邻近分析对不同类型大麻油构建分类模型,实现对其快速、无损的有效区分。在本实验中,对基于光谱数据集的单一模型和基于光谱与色谱数据集的融合模型进行分类效果进行比较,借助KNN的融合模型分类准确率可达到100%。光谱与色谱融合数据集和化学计量学相结合可作为大麻油分类评价的可靠、稳定、快速、准确、低成本的分析方法。成分不同的药物光谱与色谱数据不同,可以以此为依据实现对海关现场缴获的管制药物的分类鉴别,提升公安机关维护药物安全的能力。随着互联网技术的发展和销售渠道的扩大,管制药物走私情况日益严峻。在下一步的工作中,可以进一步筛选含有违禁成分的药物并尝试对不同来源地的同类药物进行区分,提供准确可靠的数据支撑。数据碰撞融合技术以特有的方式和优点在药物特征分析及来源推断方面发挥着独特的作用,为管制类药物的研究提供了一种快速而又无损的有效方法,有望成为法庭科学领域有利的分析工具。

## 参 考 文 献

[1] 王继芬. 大麻毒品滥用与检验[M]. 北京: 中国人民公安大学出版社, 2009.  
Wang J F. Cannabis drug abuse and testing[M]. Beijing:

People's Public Security University of China Publishing House, 2009.

- [2] Tiwari J, Bajpai K, Gupta G, et al. Tetrahydrocannabinol: a drug of interest[J]. *Panminerva Medica*, 2018, 60(4): 228-230.
- [3] Samanta D. Cannabidiol: a review of clinical efficacy and safety in epilepsy[J]. *Pediatric Neurology*, 2019, 96: 24-29.
- [4] Huestis M A, Solimini R, Pichini S, et al. Cannabidiol adverse effects and toxicity[J]. *Current Neuropharmacology*, 2019, 17(10): 974-989.
- [5] 韦卓, 李稳稳, 林敏, 等. 基于Dropout深度信念网络的棉和涤纶含量的近红外光谱检测[J]. *光学学报*, 2021, 41(17): 1730002.  
Wei Z, Li W W, Lin M, et al. Near-infrared spectroscopy detection of cotton/polyester content based on Dropout deep belief network[J]. *Acta Optica Sinica*, 2021, 41(17): 1730002.
- [6] 肖航, 杨振发, 张雷, 等. 温度对水泥生料近红外光谱检测的影响及补偿方法[J]. *中国激光*, 2020, 47(1): 0111001.  
Xiao H, Yang Z F, Zhang L, et al. Effect of temperature on near-infrared spectrum detection of cement raw meal and compensation method[J]. *Chinese Journal of Lasers*, 2020, 47(1): 0111001.
- [7] 侯伟, 张蕾萍, 王继芬, 等. 人体毛发中常见毒品及其代谢物的超高效液相色谱串联质谱法检验和统计分析[J]. *分析试验室*, 2020, 39(11): 1264-1270.  
Hou W, Zhang L P, Wang J F, et al. Determination and statistics of common drugs and their metabolites in human hair by ultra-performance liquid chromatography coupled with mass spectrometry[J]. *Chinese Journal of Analysis Laboratory*, 2020, 39(11): 1264-1270.
- [8] 罗耀, 张建莹, 黄昌雄, 等. 液相色谱-串联质谱法测定固体及液体药物中27种新型毒品芬太尼类物质[J]. *分析测试学报*, 2020, 39(4): 427-433.  
Luo Y, Zhang J Y, Huang C X, et al. Determination of

- 27 novel fentanyl analogs and metabolites in solid and liquid drugs by liquid chromatography-tandem mass spectrometry[J]. *Journal of Instrumental Analysis*, 2020, 39(4): 427-433.
- [9] 白利文, 王瑞花, 于忠山, 等. 超高效液相色谱-串联质谱法筛查血液中 29 种常见毒品[J]. *化学试剂*, 2020, 42(3): 280-284.  
Bai L W, Wang R H, Yu Z S, et al. Ultra-high performance liquid chromatography-tandem mass spectrometry screening of 29 common drugs in the blood [J]. *Chemical Reagents*, 2020, 42(3): 280-284.
- [10] Brettell T A, Lum B J. Analysis of drugs of abuse by gas chromatography-mass spectrometry (GC-MS)[J]. *Methods in Molecular Biology*, 2018, 1810: 29-42.
- [11] Wang Q Q, Huang H Y, Wang Y Z. Geographical authentication of *Macropyria Cocos* by a data fusion method combining ultra-fast liquid chromatography and Fourier transform infrared spectroscopy[J]. *Molecules*, 2019, 24(7): 1320.
- [12] 高冰, 吴鹏飞, 许晓栋, 等. 基于色谱和光谱数据融合的不同植物源食用油判别方法与模型[J]. *分析测试学报*, 2020, 39(11): 1398-1403.  
Gao B, Wu P F, Xu X D, et al. Discriminant analysis on edible oils of botanical origins based on data fusion of gas chromatography and near infrared spectroscopy[J]. *Journal of Instrumental Analysis*, 2020, 39(11): 1398-1403.
- [13] Nakamura M. Analyses of benzodiazepines and their metabolites in various biological matrices by LC-MS (/MS) [J]. *Biomedical Chromatography*, 2011, 25(12): 1283-1307.
- [14] He X L, Wang J F. Rapid and nondestructive forensic identification of tire particles by attenuated total reflectance-Fourier transform infrared spectroscopy and chemometrics[J]. *Analytical Letters*, 2020, 53(5): 714-734.
- [15] 何欣龙, 王继芬, 李青山, 等. 基于多层感知器-Fisher 判别分析的车用保险杠红外光谱鉴别[J]. *中国测试*, 2019, 45(5): 74-78, 92.  
He X L, Wang J F, Li Q S, et al. Identification of vehicle bumper debris based on multi-layer perception-Fisher discriminant and infrared spectroscopy[J]. *China Measurement & Test*, 2019, 45(5): 74-78, 92.
- [16] 秦建强, 孔祥玉, 孙喜荣. 数据标准化对 Sevcik 分形维数算法的性能影响[J]. *仪器仪表学报*, 2016, 37(7): 1485-1491.  
Qin J Q, Kong X Y, Sun X R. Influence of different data normalizing methods on the performance of Sevcik fractal dimension algorithm[J]. *Chinese Journal of Scientific Instrument*, 2016, 37(7): 1485-1491.
- [17] 褚璇, 王伟, 张录达, 等. 高光谱最优波长选择及 Fisher 判别分析法判别玉米颗粒表面黄曲霉毒素[J]. *光谱学与光谱分析*, 2014, 34(7): 1811-1815.  
Chu X, Wang W, Zhang L D, et al. Hyperspectral optimum wavelengths and Fisher discrimination analysis to distinguish different concentrations of aflatoxin on corn kernel surface[J]. *Spectroscopy and Spectral Analysis*, 2014, 34(7): 1811-1815.
- [18] Zou M Y, Sun C H, Liang S, et al. Fisher discriminant analysis for classification of autism spectrum disorders based on folate-related metabolism markers[J]. *The Journal of Nutritional Biochemistry*, 2019, 64: 25-31.
- [19] 李景奎, 张义民. 基于 K 邻近算法的连续体结构拓扑优化设计[J]. *组合机床与自动化加工技术*, 2012(1): 5-7.  
Li J K, Zhang Y M. Topology optimization of continuum structure based on K-nearest neighbor algorithm[J]. *Modular Machine Tool & Automatic Manufacturing Technique*, 2012(1): 5-7.
- [20] 古锟山, 王继芬. 基于光谱融合的油漆分类方法比较[J]. *激光与光电子学进展*, 2021, 58(22): 2230002.  
Gu K S, Wang J F. Comparison of paint classification methods based on spectral fusion[J]. *Laser & Optoelectronics Progress*, 2021, 58(22): 2230002.
- [21] Salvador-Meneses J, Ruiz-Chavez Z, Garcia-Rodriguez J. Compressed  $k$  NN: K-nearest neighbors with data compression[J]. *Entropy*, 2019, 21(3): 234.
- [22] 于叶霞, 李鹏, 王元忠. 基于光谱和色谱数据融合策略的青叶胆及近似种的鉴别研究[J]. *光谱学与光谱分析*, 2020, 40(8): 2440-2446.  
Yu Y X, Li L, Wang Y Z. Study on differentiation of *Swertia Leducii* and its closely relative species based on data fusion of spectra and chromatography[J]. *Spectroscopy and Spectral Analysis*, 2020, 40(8): 2440-2446.
- [23] 何亚, 王继芬. 基于特征波段-Fisher-K 近邻的木器漆拉曼光谱的快速无损鉴别[J]. *激光与光电子学进展*, 2020, 57(1): 013001.  
He Y, Wang J F. Rapid nondestructive identification of wood lacquer using Raman spectroscopy based on characteristic-band-Fisher-K nearest neighbor[J]. *Laser & Optoelectronics Progress*, 2020, 57(1): 013001.
- [24] 赵丽娜. Fisher 判别法的研究及应用[D]. 哈尔滨: 东北林业大学, 2013: 47-53.  
Zhao L N. Research and improvement of Fisher discriminant analysis method[D]. Harbin: Northeast Forestry University, 2013: 47-53.
- [25] 李建军, 丁正生, 张海燕. 常用判别分类方法分析[J]. *西安科技大学学报*, 2007, 27(1): 138-142.  
Li J J, Ding Z S, Zhang H Y. The analysis of a few discriminances[J]. *Journal of Xi'an University of Science and Technology*, 2007, 27(1): 138-142.
- [26] Abu A H A, Hassanat A B A, Lasassmeh O, et al. Effects of distance measure choice on K-nearest neighbor classifier performance: a review[J]. *Big Data*, 2019, 7(4): 221-248.
- [27] Nurdiawan O, Kurnia D A, Solihudin D, et al. Comparison of the K-Nearest neighbor algorithm and the decision tree on moisture classification[J]. *IOP Conference Series: Materials Science and Engineering*, 2021, 1088(1): 012031.
- [28] 卫辰浩, 王继芬, 秦歌, 等. 基于光谱分类模型的保险杠物证无损研究[J]. *中国塑料*, 2020, 34(11): 52-58.  
Wei C J, Wang J F, Qin G, et al. Nondestructive study of bumper evidence based on spectral classification model [J]. *China Plastics*, 2020, 34(11): 52-58.
- [29] 颜文杰, 陈俊明, 宋亚军, 等. 基于 DT-KNN-FDA 建模

的车漆光谱无损鉴别[J]. 激光技术, 2021, 45(2): 182-185.

Yan W J, Chen J M, Song Y J, et al. Research on non-destructive identification about vehicle paints by DT-KNN-FDA[J]. Laser Technology, 2021, 45(2): 182-185.

[30] 何欣龙, 陈利波, 王继芬, 等. 基于K近邻算法的塑钢

窗拉曼光谱分析[J]. 激光与光电子学进展, 2018, 55(5): 053001.

He X L, Chen L B, Wang J F, et al. Raman spectroscopy analysis of plastic steel window based on K nearest neighbors algorithm[J]. Laser & Optoelectronics Progress, 2018, 55(5): 053001.