

快递寄递渠道的减肥药物光谱模式识别方法比较

张傲林¹, 王继芬^{1*}, 刘松¹, 石学军², 徐晓杰², 周娣², 张震¹¹中国人民公安大学侦查学院, 北京 100038;²北京市海关缉私局司法鉴定中心, 北京 100000

摘要 通过寄递渠道传播的含有非法添加成分的减肥药物是公安机关侦办食药环犯罪的重点打击对象。为快速鉴别此类药物,本研究采用分子光谱分析技术,对含艾司唑仑、西地那非、西布曲明、氟硝西泮、唑吡坦等 5 种精神管制类药物成分的“减肥药”进行检验,获取了 145 组光谱数据。采用主成分分析提取主成分因子对数据降维。基于所提取的 20 维数据,建立 FDA 模型、KNN 模型、SVM 模型并进行对比。在模型 1 中构建 3 个 Fisher 判别函数对 5 类样品进行判别,准确度达到 100%;在模型 2 中 K 值的变化会影响分类器精度,通过对 K 值的调整能够快速对 5 类样品进行分类,准确率达到 100%;在模型 3 中选用 RBF 核函数,分别对比唑吡坦与其他 4 类减肥药物分类效果,准确率均达到 100%。通过实验中的数据集对唑吡坦不同品牌的样本进行识别和对实际案件进行分析,对公安机关侦办此类案件具有一定参考。

关键词 光谱学; 减肥药物; 精神管制类药物; 寄递; 分子光谱; 机器学习; 模式识别

中图分类号 D918.9;O657.33

文献标志码 A

DOI: 10.3788/LOP213018

Comparison of Spectral Pattern-Recognition Methods for Slimming Drugs in Express Delivery Channels

Zhang Aolin¹, Wang Jifen^{1*}, Liu Song¹, Shi Xuejun², Xu Xiaojie², Zhou Di², Zhang Zhen¹¹College of Investigation, People's Public Security University of China, Beijing 100038, China;²Forensic Expertise Center of Beijing Customs Anti-Smuggling Bureau, Beijing 100000, China

Abstract Slimming medicines containing illegal additives, often distributed through mail and other delivery channels, are primary targets for public security organizations to curb food, drug, and environmental crimes. To enable rapid identification of such drugs, this study adopted molecular spectral analysis technology to examine slimming medication containing estazolam, sildenafil, sibutramine, flurazepam, and zolpidem; 145 groups of spectral data were obtained. Principal component analysis was utilized to extract the principal component factors and reduce data dimension. Based on the extracted 20-dimensional data, the Fisher discriminant analysis (FDA), K-nearest neighbor (KNN), and support vector machine models were established for comparison. In model 1, three Fisher discriminant functions were constructed to discriminate five types of samples, and the accuracy reached 100%. In model 2, the change of K value will affect the accuracy of the classifier. Through the adjustment of K value, 5 kinds of samples can be classified quickly, with an accuracy rate of 100%. In model 3, RBF kernel function was used to compare the classification effect of zolpidem and other four kinds of slimming drugs, and the accuracy rate reached 100%. Through the dataset in the experiment, the samples of different brands of zolpidem are identified and the actual cases are analyzed, which has a certain reference for the public security organs to investigate such cases.

Key words spectroscopy; slimming drug; psychotropic substance; mail and delivery; molecular spectroscopy; machine learning; pattern recognition

1 引言

随着自贸港的建设,海外入境快递投递量激增,

“互联网+寄递”已成为日益突出的贩毒方式。含有非法添加成分的减肥药物在电商平台的消费不断提升,不法分子通过非法渠道大量进口国外处方减肥药物,

收稿日期: 2021-11-22; 修回日期: 2021-12-15; 录用日期: 2021-12-27; 网络首发日期: 2022-01-08

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

并通过快递寄递的渠道销往国内各地。公安部等三部门联合部署,在 2021 年 9 月 1 日—12 月 10 日期间开展寄递渠道禁毒百日攻坚行动,全力打击整治寄递渠道涉毒活动。

西布曲明是一种 5-羟色胺去甲肾上腺素再摄取抑制剂,临床研究表明其具有促进肥胖者体重减轻的效果^[1]。药物毒理研究表明西布曲明会使中枢神经系统过度兴奋并伴有感官异常等生理反应。艾司唑仑、氟硝西泮属于苯二氮䓬类药物,具有催眠、抗焦虑等作用,耐受性和依赖性较大。长期服用含有此类添加成分的药物会导致尿频、周身乏力等,严重时会出现精神恍惚、呼吸抑制等危害。服用西地那非后最常见的药理反应是头痛、呕吐、严重影响消化系统和呼吸道。鉴于减肥药物对人体健康产生的重大危害,打击通过快递寄递渠道非法销售的处方减肥药物已成为公安机关与物证鉴定人员的重要任务。

针对减肥药物的检验,目前主要采用高效液相色谱法(LC)及气相色谱-质谱联用(GC-MS)等分析方法。洪灯等^[2]采用高效液相色谱-四极杆/静电场轨道阱高分辨质谱法,通过比较不同流动相下的定性效果,对 6 类西布曲明及衍生物进行分析。孙婷婷等^[3]借助分子印迹整体柱在线固相萃取-液相色谱-串联质谱技术完成了对数组保健品样品中西地那非的定量分析。但此类方法检验周期长、操作难度大,案件中涉案样品数量较大时捉襟见肘,检验人员亟须一种快速、无损、准确的新型检测方法。红外光谱法利用物质分子对红外光的吸收所产生的红外吸收光谱对物质的组成和结构进行分析测定,作为一种常用的无损检验技术,其具有分析速度快、样品用量少,操作简单等优点,一直以来在各检测领域有着十分广泛的应用。实际工作中,单纯依靠肉眼观察谱图差异进行鉴别分析,不仅主观因素影响大而且效率低。而红外光谱分析与机器学习结合能够达到快速准确鉴别的目的。例如,李佳瑞等^[4]在对苯二氮䓬类和吩噻嗪类镇静安眠药样品的研究中,选用红外光谱数据融合对数据进行建模处理,总体分类准确率达 92.7%。侯伟等^[5]使用随机森林模型对安眠镇静类药物光谱数据进行建模分析,对 3 种安眠药的分类识别率可达 100%。基于此,本研究借助红外光谱分析,建立了 3 种机器学习模型,并对比这 3 类模型的分类准确率。

2 基本原理

1.1 实验样本与仪器

本研究中的实验样本为采集自己知含唑吡坦、西地那非、西布曲明、氟硝西泮、艾司唑仑等 5 种非法添加成分的 145 份减肥药物样本,并分别编号 1、2、3、4、5 类,其中 1、2、4、5 类分别有 30 个样本,3 类有 25 个样本,通过研磨法粉碎药物样本,便于使用红外光谱进行检测。实验中使用的设备是傅里叶红外光谱仪(Nicolet is20 型,

美国 Thermo Fisher Scientific 公司),具体参数如表 1 所示。

表 1 实验设备及参数

Facility	Parameter
Nicolet is20 Fourier transform infrared spectrometer	Spectral resolution: 2 cm^{-1}
	Spectral repeatability: better than 0.02 nm
	Number of scans: 32
	Measuring range: $4000\text{--}400\text{ cm}^{-1}$
	Dynamic adjustment up to 130000 times/s
	Signal-to-noise ratio: 50:1

1.2 数据前处理

实验室环境下,由于温度、湿度及仪器自身因素影响,光谱背景信息冗杂,常存在基线摆动等问题。本研究中通过 Ominic 采用基线自动预处理消除光谱背景信息。在数学上通过主成分分析(PCA)探究不同维度数据的模型准确率。PCA 是一种在损失很少信息的前提下把多个指标转化为几个综合指标的多元统计方法,转化生成的综合指标称之为成分,其中每个成分都是原始变量的线性组合,且各个成分之间互不相关。Denman 等^[6]尝试应用质谱法对油墨进行分析鉴别,并利用 PCA 处理相关质谱图特征数据,成功地鉴别 41 对(占 91%,共 45 对)墨水样本。

本实验通过相关系数矩阵计算主成分得分发现,在 20 维后累计方差贡献率达 100%,20 维数据时冗余信息较少,有效信息完整,准确率最高,如图 1 所示。

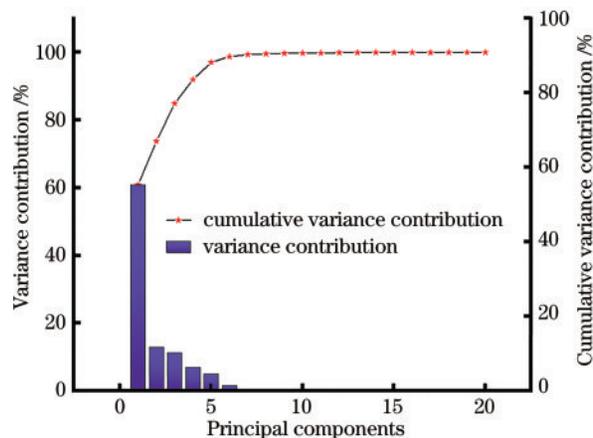


图 1 主成分的累计方差贡献率

Fig. 1 Cumulative variance contribution rate of principal component

1.3 实验建模

1.3.1 FDA 模型

Fisher discriminant analysis(FDA)判别是一种有效的分类方法^[9-13],其原理是将多维数据投影到某个方向上,拟定已知类别为标准,将待定类别之间尽可能分开,类内尽可能聚合,然后选择合适的判别规则对未知样品进行分类判别。

假设整个数据样本集一共可以分成 K 类, 那么有

$$s_w = \sum_{k=1}^K s_k, \quad (1)$$

式中: s_w 为散度矩阵; s_k 为原始矩阵。两个总体的 Fisher 判别函数为

$$c(Y) = C_1 Y + C_2 Y_2 + \dots + C_p Y_p = C' Y', \quad (2)$$

式中: C_1, \dots, C_p 是组间离差系数。多总体的 Fisher 判别函数为

$$E(C'Y) = E(C'Y|G_i) = C'E(Y|G_i) = C' \mu_i, \quad i = 1, 2, \dots, k, \quad (3)$$

式中: G_i 为总体; C' 代表组间差; Y 代组内差。本实验模型 1 中分别将 5 种减肥药物的 20 维光谱数据以 Fisher 判别函数投影到 XYZ 方向上, 建立 FDA1、FDA2、FDA3 判别轴, 可以对维度较高的光谱数据进行降维运算, 同时可通过训练模型对未知的减肥药物进行判别, 但多维矩阵的运算导致时间复杂度较高, 样本量大时处理较慢。

1.3.2 KNN 模型

K-nearest neighbors (KNN) 是一种较为简单的机器学习算法^[14-18], 通过欧氏距离计算交叉验证样本与数据样本集的最邻近距离从而对样本进行分类。根据算法性能评估可知, 随 K 值的变化, 模型的拟合性会随之变化, K 值与准确率有关, 在度量计算上根据数据维度不同, 常见有欧氏距离和马氏距离:

$$d_E(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x}_1 - \mathbf{y}_1)^2 + (\mathbf{x}_2 - \mathbf{y}_2)^2 + \dots + (\mathbf{x}_n - \mathbf{y}_n)^2}, \quad (4)$$

$$d_M = \sum_{k=1}^n |\mathbf{x}_{1k} - \mathbf{x}_{2k}|, \quad (5)$$

式中: n 表示特征数; $\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{y}_1, \dots, \mathbf{y}_n$ 表示向量特征; d_E 和 d_M 分别表示欧氏距离和马氏距离下的函数。在模型 2 中, 由于 5 种药物的数据维度仅达 20 维, 冗余

信息上较少, 欧氏距离具有较高的精度且计算复杂度低, 相较于 FDA 模型和支持向量机 (SVM) 模型能够更便捷解决 5 类减肥药物的多分类问题。

1.3.3 SVM 模型

SVM 是一种属于监督学习的机器学习算法^[19-22], 在不同核函数 (Kernel) 下过度数据维度, 通过约束条件 $\mathbf{y}_i(\boldsymbol{\omega} \mathbf{x}_i + \mathbf{b}) - 1 \geq 0$ 确定最小分类间隔 $\frac{\|\boldsymbol{\omega}\|^2}{2}$, 可以

通过 Lagrange 方法解决, 令 $L(\boldsymbol{\omega}, \mathbf{b}, \mathbf{a}) = \frac{1}{2} \|\boldsymbol{\omega}\|^2 - \sum_{i=1}^m [\mathbf{y}_i(\boldsymbol{\omega} \mathbf{x}_i + \mathbf{b}) - 1]$, 其中 \mathbf{a} 为每个样本的拉格朗日乘

子。由此得到决策函数 $f(\mathbf{x}) = \text{sgn} [\sum_{i=1}^m \mathbf{a}_i \mathbf{y}_i(\mathbf{x} + \mathbf{x}_i) + \mathbf{b}]$ 。常见核函数有以下 4 种: 1) Linear Kernel, $K(x, x_i) = x \cdot x_i$; 2) Polynomial Kernel, $K(x, x_i) = [(x, x_i) + 1]^q$; 3) RBF/Gaussian Kernel, $K(x, x_i) = \exp \left[-\frac{(x - x_i)^2}{\sigma^2} \right]$; 4) Sigmoid Kernel, $K(x, x_i) = \tanh [v(x, x_i) + c]$ 。

在本实验模型 3 中, 选用图像特征辨识度较低的唑吡坦样本为分类标准, 对比其他 4 种减肥药物在径向基函数 (RBF) 核函数支持下与唑吡坦的分类准确率, 在此分类问题上, SVM 模型具有较好的鲁棒性。

3 分析与讨论

2.1 谱图分析

5 组减肥药 (分别含唑吡坦、艾司唑仑、氟硝西泮、西布曲明、西地那非) 的 145 个样本红外谱图如图 2 所示。从图中可以看出, 每组样本的峰形、峰的走向和出峰位置基本一致, 如唑吡坦在波数为 900 cm^{-1} 左右、 1000

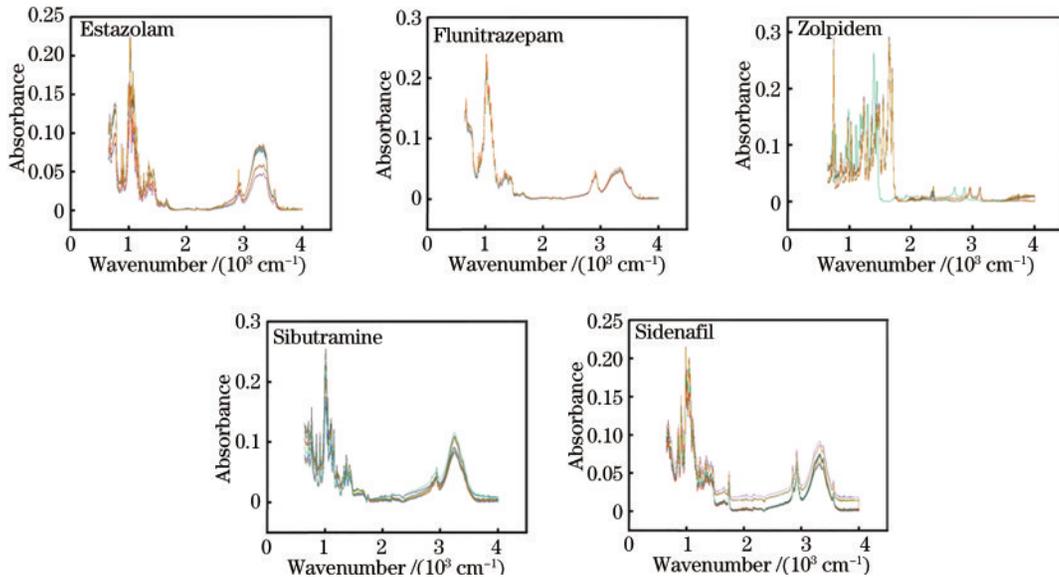


图 2 145 组样本红外光谱
Fig. 2 Infrared spectrograms of 145 samples

cm^{-1} 左右、 2900 cm^{-1} 左右均有中间高两侧低的多个尖峰,在波数为 1400 cm^{-1} 左右均有两个尖峰,且相对于右峰左峰较宽,在波数为 1000 cm^{-1} 左右均有多个高强度尖峰,在波数为 670 cm^{-1} 左右均有一个尖峰。部分样本在峰的个数及相对峰高有所区别,在波数为 3400 cm^{-1} 处有一弱峰,由于红外光谱对不同样本形成波形图有较大差异,波数不确定,使用人工观察法直接对比红外谱图实现对样本的准确区分难度较大。

不同类型的“减肥药”及“减肥茶”在组分和配比上会存在一定差异,当样本数量较多时,借助红外谱图直接分析不仅主观误差较大而且耗时耗力,此外成分的混杂使得谱图之间交叉混淆现象较多^[23-24],无法客观对样品进行准确有效的区分检验。基于此,本实验借助 PCA 降维提取样本特征峰光谱数据,最终选择 30 维有效信息对比 PCA-FDA、PCA-KNN、PCA-

SVM 等 3 种机器学习模型分类准确率。

2.2 化学光谱模式识别

2.2.1 Fisher 判别分析

Fisher 判别能够通过投影转化的思维对数据进行降维处理,卫辰洁等^[7]借助 PCA-FDA 模型实现了对 12 个品牌 44 个汽车灯罩样本的 100% 分类,并指出通过 PCA 对光谱信息进行提取后,能大量排除冗余信息,从而提高 FDA 判别的精度。本实验以 5 组药品样本(唑吡坦-1 类、西地那非-2 类、氟硝西洋-3 类、西布曲明-4 类、艾司唑仑-5 类)为分类依据,构建判别预测模型,实现对 145 个样本的分类工作。选择各类别先验概率,按照各类样本数量进行计算,使用合并组内协方差矩阵进行分析,得到了各样本的判别函数摘要,如表 2 所示。

表 2 5 组药物样本判别函数摘要

Table 2 Summary of discriminant functions of 5 groups of drug samples

Function	Function test	Correlation	Total	Wilks' Lambda	Sig	Eigenvalue
FDA1	1 to 4	0.998	2263.451	0	0.000	286.651
FDA2	2 to 4	0.996	1482.129	0	0.000	124.386
FDA3	3 to 4	0.99	815.397	0.003	0.000	50.449
FDA4	4	0.928	271.596	0.14	0.913	6.157

Fisher 模型构建了 4 个判别函数,由表 2 可知,函数 1、函数 2、函数 3、函数 4 的累计方差贡献率达 100%,表明其能很好地解释样本的基本特征信息。相关性均大于 92%,表明不同分组与函数 1、函数 2、函数 3、函数 4 的相关性较强。函数检验中,Wilks' Lambda 系数表明函数 1、2、3 对分类模型影响的显著性高于函数 4,能很好解释各样本的分类情况。综上所述,选择函数 1、函数 2、函数 3 作为判别函数,构建判别预测模型,得到了 145 个样本的判别分布图(图 3)和判别结果(表 3)。函数 1、函数 2、函数 3 的判别式分别为

$$\text{FDA1}(X) = -7.306X_1 + 1.976X_2 - 1.897X_3 + 13.096X_4 - 6.39X_5 - 1.57X_6 - 1.995X_7, \quad (5)$$

$$\text{FDA2}(X) = 2.429X_1 - 7.747X_2 + 4.055X_3 - 0.176X_4 - 5.676X_5 + 1.979X_6 + 1.841X_7, \quad (6)$$

$$\text{FDA3}(X) = -2.816X_1 - 3.495X_2 + 1.732X_3 + 1.552X_4 + 4.921X_5 + 0.506X_6 - 0.030X_7. \quad (7)$$

经 PCA 和 FDA 两次降维处理后,光谱数据冗余

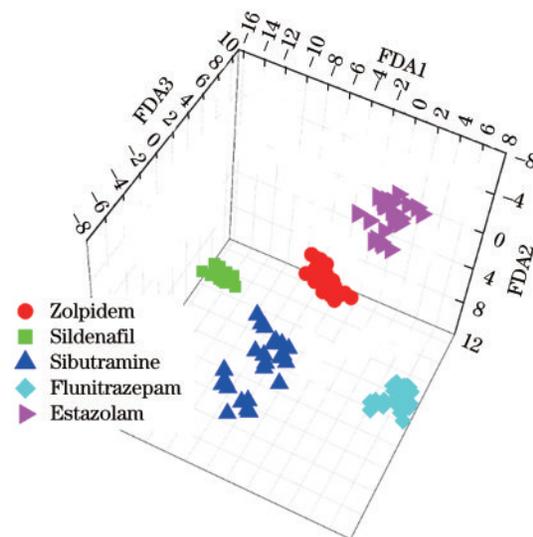


图 3 Fisher 判别样本识别空间分布

Fig. 3 Spatial distribution of Fisher discriminant samples

信息大量消除,对于模型进行判别具有一定帮助,实现

表 3 5 组药物分类准确率

Table 3 Accuracy of drug classification in 5 groups

Type	Zolpidem	Estazolam	Flunitrazepam	Sibutramine	Sildenafil	Accuracy / %
Zolpidem	30					100
Estazolam		30				100
Flunitrazepam			30			100
Sibutramine				25		100
Sildenafil					30	100

了对 5 类药品的 100% 分类。由图 3 可知, 唑吡坦、艾司唑仑样本在函数 FDA1 上区分明显, 西地那非样本在函数 FDA2 上区分明显, 西布曲明、氟硝西洋样本在函数 FDA3 上区分明显。唑吡坦、西地那非、氟硝西洋样本在模型上聚敛程度高, 西布曲明、艾司唑仑样本聚敛程度低。

2.2.2 KNN 分类

欧氏度量下的 KNN 模型可以通过计算距离解决多分类问题, 李佳瑞等^[8]借助 PCA-KNN 对不同类型的大麻油构建分类模型发现, 在欧氏度量下的 KNN 模型对于数据维度较低的光谱分类准确率较高, 由于光谱数据在 [0, 1] 变化且通过 PCA 降维后会大量减少计算的复杂度, 所以保证了 100% 分类准确。基于此, 本实验模型 2 采用欧氏度量, 符合上述使用环境。在以 70% 样本数据为训练集、30% 样本数据为测试集进行交叉验证的方法下, 调整 K 值使模型达到最高准确率。基于 20 维的成分, 总体分类准确率随 K 值增加整体呈现上升趋势, 具体如图 4 所示: 其中在 K=1 时仅对单个临近进行预测, 模型复杂度高, 决策边界崎岖, 准确率最低; 在 K=5 时总体分类准确率达 100%, 训练效果较好, 验证准确; 在 K 值大于等于 7 以后准确率稳定在 100%, 但模型计算复杂度提高。因此 K=5 时最佳, 能够对 5 组减肥药物样品进行较好的区分, 具体如表 4 所示。本实验中数据量达 145 组, 使用 KNN 进行聚类具有使用意义, 通过降维和调整 K 值的方式使得准确率可达 100%。

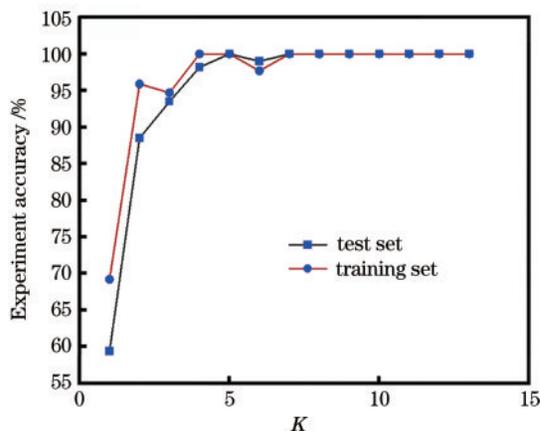


图 4 KNN 模型下样本分类准确率

Fig. 4 Classification accuracy of samples under KNN model

2.2.3 SVM 分类

由于 SVM 的二分类问题效果较好, 且唑吡坦样本的光谱特征明显, 故在实验中选择 1 类 (唑吡坦) 为参照样本集, 选取 2 类 (西地那非)、3 类 (西布曲明)、4 类 (氟硝西洋)、5 类 (艾司唑仑) 为对照分类样本集。惩罚因子、伽马值等参数环境一致, 对比 4 种药物的分类准确率。结果表明, RBF 核函数下的 SVM 对于样本分类效果较好, 唑吡坦与其他 4 类减肥药物的分类

表 4 KNN 模型下样本分类准确率

Table 4 Classification accuracy of samples under KNN model units: %

K	Zolpidem	Estazolam	Flunirazepm	Sibutramine	Sildenafil
1	59.4	59.4	59.4	59.4	59.4
2	88.5	88.5	88.5	88.5	88.5
3	93.5	93.5	93.5	93.5	93.5
4	98.2	98.2	98.2	98.2	98.2
5	100	100	100	100	100
6	99	99	99	99	99
7	100	100	100	100	100
8	100	100	100	100	100
9	100	100	100	100	100
10	100	100	100	100	100
11	100	100	100	100	100
12	100	100	100	100	100
13	100	100	100	100	100

准确率均达 100%, 如图 5 所示。对于维度较高的样本光谱数据, 运用 PCA-SVM 模型降维后进行 SVM 分类识别, 对减肥药物实现了准确的分类。局部支持向量可解释最终结果, 以核函数代替高维空间的非线性映射, 消除了光谱数据大量的冗余信息, 对比其他算法具有较好的“鲁棒”性。

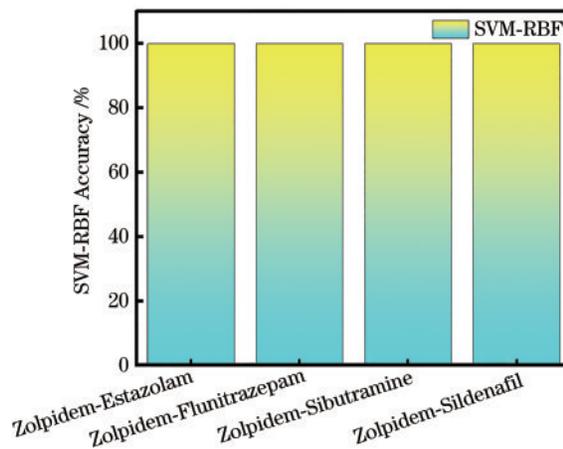


图 5 SVM 模型分类准确率

Fig. 5 Classification accuracy of SVM model

2.3 品牌分类与案例分析

含 1 类非法添加成分的减肥药物样本分别为曲姿秀减肥胶囊、纤魅脱脂胶囊、曲姿秀减肥药、赘克丽尔胶囊、纤体稳定素 III 等 8 个品牌减肥药, 将唑吡坦 8 种品牌样本按 C1~C8 依次标注, 具体如表 5 所示。模型 3 以含 1 类为例, 对 1 类、2 类、3 类、4 类、5 类等 5 种精神管控类药物的减肥药物样本进行 100% 的准确区分。

基于此, 在模型 3-SVM 的分类基础上, 针对唑吡坦样本, 以品牌为单位, 继续构建模型 1-FDA 判别分

表 5 唑吡坦 8 种品牌编号
Table 5 Number of 8 brands of Zolpidem

Drug	Number
Miaozi Slimming capsules	C1
Xianmei capsules	C2
Xianfeixianwei tablet	C3
Quzixiu Slimming capsules	C4
Zhuikelier capsules	C5
Xianti pill III	C6
Meizhixuan Pill	C7
Xiuzhixiu fu capsules	C8

类模型,对品牌进行判别分类,得到了该类样本品牌间判别结果(表 6)。由表 6 可知,8 类样本的总体分类准确率为 83.34%,其中曲姿秀减肥胶囊、曲姿秀减肥药、赘克丽尔胶囊、纤体稳定素 III、美之选纤美素这 5 种品牌的唑吡坦药品样本实现 100% 的准确区分,纤魅脱脂胶囊等 5 类减肥药品牌的唑吡坦样本分类有误,模型对该 5 类样本的区分能力较弱,区分效果一般。

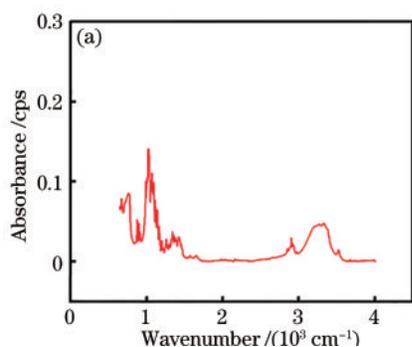


表 6 唑吡坦 8 种品牌分类准确率
Table 6 Classification accuracy of 8 brands of Zolpidem

Brand	C1	C2	C3	C4	C5	C6	C7	C8	Accuracy
C1	3								100%
C2		1				1			50%
C3			1			1			50%
C4				10					100%
C5					1				100%
C6						7			100%
C7							2		100%
C8						1		2	66.70%

2021 年某月某日,北京市海关缉私局在一起案件中查获通过快递寄递渠道进口的减肥药物 2 份,将样本命名为 A、B,以上样本需经过分子光谱无损检验确定后再进行气相色谱-质谱定量分析,A、B 样本的衰减全反射-傅里叶变换红外光谱谱图如图 6 所示。由图 6 可知,A 和 B 的红外光谱图具有较高的相似度,仅依靠谱图分析难以判别其种类,需借助机器学习的相关方法进行分析 and 认定。

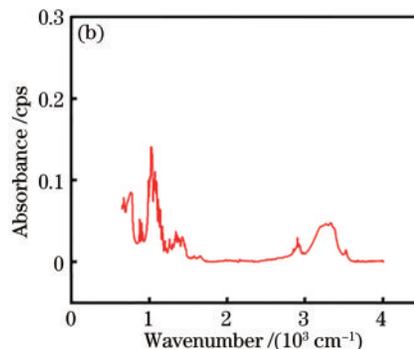


图 6 A、B 样本红外光谱图。(a) A 样本; (b) B 样本

Fig. 6 A, B infrared spectrum. (a) A sample; (b) B sample

以上述 145 份减肥药物样本为训练集,A、B 两组未知减肥药物为测试集,建立基于分子水平光谱下的 SVM 模型,判定结果为 A 含有艾司唑仑、B 含有唑吡坦,后经气相色谱-质谱检测证实。因此,在公安物证领域实际应用中,SVM 模型对于减肥药物的光谱模式识别具有良好效果。

4 结 论

本实验利用分子水平下红外光谱获取 145 组减肥药物的光谱数据,通过 PCA 对光谱数据进行降维处理获得 20 维有效信息,对比模型 1-FDA、模型 2-KNN、模型 3-SVM 等 3 种机器学习模型对艾司唑仑、氟硝西洋、唑吡坦、西布曲明、西地那非等 5 类样品的光谱识别准确率,在模型 1 中选择 FDA1、FDA2、FDA3 作为判别轴,对 5 组样本实现 100% 准确率的区分。在模型 2 中分类准确率较低,仅为 59.4%,通过调整 K 值可以增加准确率,但随着 K 值提高计算时间复杂度会提高。

在模型 3 中 RBF 核函数的 SVM 模型分类准确率达 100%,结合 FDA 判别能够对品牌实现总体准确率为 83.3% 的正确分类,且对实际案件分析准确。实验结果表明:FDA 模型可以降低高维度光谱数据,经 PCA 处理相较其余两个模型运算复杂度低,处理问题较快;KNN 模型对于多种减肥药物分类问题具有优势,可以快速分辨出 5 类不同的样本;SVM 模型对于两种减肥药物样本的区分问题较好。

参 考 文 献

- [1] 张鑫鑫,李霖,邵雅楠,等. 离子迁移谱法同时测定减肥类保健品中 5 种非法添加物的含量[J]. 理化检验-化学分册, 2018, 54(12): 1419-1424.
Zhang X X, Li L, Gao Y N, et al. Determination of five kinds of illegal additives in weight loss health products by ion mobility spectrometry[J]. Physical Testing and Chemical Analysis (Part B: Chemical Analysis), 2018, 54(12): 1419-1424.

- [2] 洪灯, 谢文, 侯建波, 等. 高效液相色谱-四极杆/静电场轨道阱高分辨质谱法快速筛查保健食品中的西布曲明及其 5 种衍生物[J]. 色谱, 2019, 37(11): 1173-1178.
Hong D, Xie W, Hou J B, et al. Rapid screening of sibutramine and five derivatives in health food by high performance liquid chromatography-quadrupole/electrostatic orbitrap high-resolution mass spectrometry[J]. Chinese Journal of Chromatography, 2019, 37(11): 1173-1178.
- [3] 孙婷婷, 徐云慧, 张雪梅, 等. 分子印迹整体柱在线固相萃取-液相色谱-串联质谱法测定保健食品中西地那非[J]. 理化检验-化学分册, 2019, 55(11): 1247-1253.
Sun T T, Xu Y H, Zhang X M, et al. LC-MS/MS determination of sildenafil in health food with on-line solid phase extraction of molecularly imprinted monolithic column[J]. Physical Testing and Chemical Analysis (Part B: Chemical Analysis), 2019, 55(11): 1247-1253.
- [4] 李佳瑞, 王继芬, 刘津彤. 公共安全视域下精神药物光谱学无损分类[J]. 分析测试学报, 2022, 41(2): 276-282.
Li J R, Wang J F, Liu J T. Non-destructive classification of psychotropic drugs based on spectroscopy from the perspective of public security[J]. Journal of Instrumental Analysis, 2022, 41(2): 276-282.
- [5] 侯伟, 王继芬, 张蕾萍, 等. 药品安全视域下安眠镇静类药物光谱模式识别研究[J]. 激光与光电子学进展, 2022, 59(5): 0530004.
Hou W, Wang J F, Zhang L P, et al. Spectral pattern recognition of sedative-hypnotic drugs in the perspective of drug safety[J]. Laser & Optoelectronics Progress, 2022, 59(5): 0530004.
- [6] Denman J A, Skinner W M, Kirkbride K P, et al. Organic and inorganic discrimination of ballpoint pen inks by ToF-SIMS and multivariate statistics[J]. Applied Surface Science, 2010, 256(7): 2155-2163.
- [7] 卫辰洁, 王继芬, 曾啸虎. 红外光谱数据融合结合化学计量学无损检测汽车灯罩[J]. 分析测试学报, 2021, 40(7): 1043-1048.
Wei C J, Wang J F, Zeng X H. Nondestructive detection of automobile lampshades by infrared spectrum data fusion combined with chemometrics[J]. Journal of Instrumental Analysis, 2021, 40(7): 1043-1048.
- [8] 李佳瑞, 王继芬, 范琳媛, 等. 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[J]. 激光与光电子学进展, 2022, 59(16): 1630004.
Li J R, Wang J F, Fan L Y, et al. Fast recognition and classification of hemp oil based on collision fusion strategy of spectral and chromatographic data[J]. Laser & Optoelectronics Progress, 2022, 59(16): 1630004.
- [9] 孙一健, 王继芬. 太赫兹时域光谱技术在食品、药品和环境领域中的应用研究进展[J]. 激光与光电子学进展, 2022, 59(16): 1600001.
Sun Y J, Wang J F. Application of terahertz time-domain spectroscopy in food, medicine and environment[J]. Laser & Optoelectronics Progress, 2022, 59(16): 1600001.
- [10] Argyri A A, Panagou E Z, Tarantilis P A, et al. Rapid qualitative and quantitative detection of beef fillets spoilage based on Fourier transform infrared spectroscopy data and artificial neural networks[J]. Sensors and Actuators B: Chemical, 2010, 145(1): 146-154.
- [11] Oravec M, Beganović A, Gál L, et al. Forensic classification of black inkjet prints using Fourier transform near-infrared spectroscopy and Linear Discriminant Analysis[J]. Forensic Science International, 2019, 299: 128-134.
- [12] Muehlethaler C, Massonnet G, Esseiva P. Discrimination and classification of FTIR spectra of red, blue and green spray paints using a multivariate statistical approach[J]. Forensic Science International, 2014, 244: 170-178.
- [13] 杨崇山, 董春旺, 江用文, 等. 基于高光谱的工夫红茶发酵品质程度判别方法[J]. 光谱学与光谱分析, 2021(4): 1320-1328.
Yang C S, Dong C W, Jiang Y W, et al. A method for judging the fermentation quality of congou based on hyperspectral[J]. Spectroscopy and Spectral Analysis, 2021(4): 1320-1328.
- [14] 何欣龙, 王继芬, 刘腾飞, 等. 傅里叶红外光谱法结合化学计量学方法区分鉴别塑钢窗[J]. 理化检验-化学分册, 2018, 54(11): 1318-1323.
He X L, Wang J F, Liu T F, et al. Discrimination and classification of plastic steel windows by FTIR in combination with chemometrics[J]. Physical Testing and Chemical Analysis (Part B: Chemical Analysis), 2018, 54(11): 1318-1323.
- [15] 刘燕德, 徐海, 孙旭东, 等. 西红柿成熟度的近红外漫透射光谱无损检测[J]. 激光技术, 2019, 43(1): 25-29.
Liu Y D, Xu H, Sun X D, et al. Non-destructive measurement of tomato maturity by near-infrared diffuse transmission spectroscopy[J]. Laser Technology, 2019, 43(1): 25-29.
- [16] 陈海秀, 胡祯林, 张江洲. 基于近红外光谱技术的易燃液体快速鉴别[J]. 激光杂志, 2017, 38(11): 46-49.
Chen H X, Hu Z L, Zhang J Z. Rapid identification of flammable liquids based on near infrared spectroscopy[J]. Laser Journal, 2017, 38(11): 46-49.
- [17] 孙宗保, 辛新, 邹小波, 等. 傅里叶变换红外光谱结合化学计量学方法对白酒基酒快速定性和定量分析[J]. 光谱学与光谱分析, 2017, 37(9): 2756-2762.
Sun Z B, Xin X, Zou X B, et al. Rapid qualitative and quantitative analysis of base liquor using FTIR combined with chemometrics[J]. Spectroscopy and Spectral Analysis, 2017, 37(9): 2756-2762.
- [18] 吴喆, 张霖, 金航, 等. 红外光谱结合化学计量学对不同采收期滇重楼的定性定量分析[J]. 光谱学与光谱分析, 2017, 37(6): 1754-1758.
Wu Z, Zhang J, Jin H, et al. Qualitative and quantitative analysis of Paris polyphylla var. yunnanensis in different harvest times with infrared spectroscopy combined with chemometrics[J]. Spectroscopy and Spectral Analysis, 2017, 37(6): 1754-1758.
- [19] 严爱花, 李贤良, 郗存显, 等. 液相色谱-串联质谱法同时检测中成药及保健食品中非法添加的 22 种苯二氮卓类药物[J]. 分析化学, 2013, 41(4): 509-516.
Yan A H, Li X L, Xi C X, et al. Simultaneous determination of 22 benzodiazepines in Chinese patent drugs and health-care foods by liquid chromatography-tandem mass spectrometry[J]. Chinese Journal of Analytical

- Chemistry, 2013, 41(4): 509-516.
- [20] 古锬山, 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] 侯伟, 王继芬, 何欣龙. 基于多元建模的甲基苯丙胺及其常见添加剂混合物光谱分类识别[J]. 激光与光电子学进展, 2021, 58(3): 0330003.
Hou W, Wang J F, He X L. Spectral classification and identification of methamphetamine and its common additives based on multivariate modeling[J]. Laser & Optoelectronics Progress, 2021, 58(3): 0330003.
- [22] 颜文杰, 卢雯慧, 王继芬. 基于 SVM-MLP 融合模型的毒品混合物光谱识别研究[J]. 激光与光电子学进展, 2021, 58(14): 1404003.
Yan W J, Lu W H, Wang J F. Research on spectral recognition of drug mixture based on SVM-MLP fusion model[J]. Laser & Optoelectronics Progress, 2021, 58(14): 1404003.
- [23] 邓高琼, 陈亨业, 刘瑞, 等. 气相色谱-质谱联用技术在食药检测中的应用与发展[J]. 化学试剂, 2021, 43(5): 555-562.
Deng G Q, Chen H Y, Liu R, et al. Application and development of gas chromatography-mass spectrometry technology in food and Chinese herbal medicine testing [J]. Chemical Reagents, 2021, 43(5): 555-562.
- [24] 陈明曦, 李艳梅, 孙婷, 等. 液相色谱-质谱联用技术在中药质量研究中的应用[J]. 中国药剂学杂志, 2021, 19(1): 12-17.
Chen M X, Li Y M, Sun T, et al. Application of liquid chromatography/mass spectrometry in quality research of traditional Chinese medicine[J]. Chinese Journal of Pharmaceutics, 2021, 19(1): 12-17.