

引用本文:曾薇,谭成森,苏元翠,等.基于中红外光谱的含钩吻蜂蜜鉴别研究[J].光通信技术,2023,47(6):77-82.

基于中红外光谱的含钩吻蜂蜜鉴别研究

曾薇¹,谭成森¹,苏元翠¹,曾蕊¹,华逸昇¹,黎远鹏^{1*},黄富荣²,赵红霞³

(1.广西师范大学 物理科学与技术学院,广西 桂林 541004;

2.暨南大学 光电工程系,广州 510632;3.广东省科学院 动物研究所,广州 510260)

摘要:为了探究一种快速、准确的含钩吻蜂蜜的检测方法,通过中红外(MIR)光谱仪采集了来自不同地区的216个无毒、含钩吻有毒的蜂蜜样品的MIR光谱图像,选取光谱图像的800~1 800 cm⁻¹波段作为主要研究波段,应用主成分分析法并结合偏最小二乘回归(PLS-DA)算法建立有无钩吻鉴别模型,比较不同数据预处理方法的预测结果。仿真结果表明:经一阶导数预处理后建立的蜂蜜有无钩吻鉴别模型的鉴别效果最佳,鉴别准确率达到90.74%。

关键词:中红外光谱;钩吻毒素;有毒蜂蜜;鉴别

中图分类号:O433 **文献标志码:**A **文章编号:**1002-5561(2023)06-0077-06

DOI:10.13921/j.cnki.issn1002-5561.2023.06.016

开放科学(资源服务)标识码(OSID):



Identification of honey containing hookkiss based on mid-infrared spectroscopy

ZENG Wei¹, TAN Chengsen¹, SU Yuancui¹,

ZENG Rui¹, HUA Yisheng¹, LI Yuanpeng^{1*}, HUANG Furong², ZHAO Hongxia³

(1. College of Physical Science and Technology, Guangxi Normal University, Guilin Guangxi 541004, China;

2. Department of Optoelectronic Engineering, Jinan University, Guangzhou 510632, China;

3. Institute of Zoology, Guangdong Academy of Sciences, Guangzhou 510260, China)

Abstract: In order to explore a rapid and accurate detection method for honey containing uncine-kisses, mid-infrared(MIR) spectral images of 216 non-toxic and uncine-toxic honey samples from different regions are collected by MIR spectrometer. The 800~1 800 cm⁻¹ band of the spectral images is selected as the main research band. Principal component analysis and partial least square regression(PLS-DA) algorithm are used to establish the identification model of hook-free and non-hook-free, and the prediction results of different data preprocessing methods are compared. The simulation results show that the identification model established after the first derivative pretreatment has the best identification effect, and the identification accuracy reaches 90.74%.

Key words: mid-infrared spectrum, Gelsemium elegans benth, poisonous honey, discrimination

0 引言

蜂蜜作为一种营养丰富的天然甜味剂在世界范围内受到消费者的广泛喜爱,在某些情况下还作为一

种有效的营养品^[1]。然而,野生植物中含有的一些有毒物质经常被蜜蜂在采蜜时引入蜂蜜产品中,导致食物中毒事件发生。其中,钩吻是我国主要有毒蜜源之一。目前,国内外鉴别钩吻的方法有高效液相色谱-串联质谱法(HPLC-MS/MS)^[2]、薄层色谱法(TLC)^[3]和气相色谱法(GC)^[4]等。这些检测方法都需要事先对检测样品进行复杂的预处理;同时,这些方法检测过程缓慢,对样品具有一定的破坏性和污染性,且费用昂贵,在检测过程中使用的药品也有可能对分析人员的身体健康造成影响^[5]。因此,需要研究一种快速、安全的检测方法来检测含钩吻的有毒蜂蜜。

中红外(MIR)光谱中的光谱吸收带表示分子键

收稿日期:2022-05-03。

作者简介:曾薇(2000—),女,大学本科生,现就读于广西师范大学物理科学与技术学院物理学(本硕分流培养班)专业,主要从事光谱及光谱成像分析及应用方面的研究工作。负责完成了省部级项目1项,参与省部级、国家级项目6项;曾荣获第七届全国大学生物理竞赛(教学赛)三等奖、第八届全国大学生物理竞赛(创新赛)三等奖。



*通信作者:黎远鹏(1989—),男,博士,副教授,主要从事光谱、光谱成像分析及应用方面的研究工作。

曾薇,谭成森,苏元翠,等:基于中红外光谱的含钩吻蜂蜜鉴别研究

和官能团的特征。这些光谱吸收带由分子的拉伸振动、弯曲振动和摆动振动引起,整个光谱可以作为给定化合物的指纹。因此,有毒蜂蜜的指纹特征可代表其整体化学成分,是检测有毒蜂蜜的重要依据。MIR光谱技术作为食品检测工具,其检测速度极快(可在1~2 m内完成)、使用简单,不仅能由非专业人员实施,还可以在测试中搜索出大量潜在的掺假物,无需样品制备,且测试过程中不产生污染。

因此,本文提出一种基于MIR光谱的含钩吻蜂蜜鉴别方法,能对含钩吻的有毒蜂蜜实现快速、准确的检测。

1 基于MIR光谱的含钩吻蜂蜜鉴别方法

将MIR光谱技术与机器学习相结合,可以从MIR光谱中提取定性和定量信息,从而促进食品分析成分和掺假鉴别方法的发展。机器学习使用数学和算法工具来提供来自大型算法数据集的完整信息^[6]。首先,分析MIR光谱中对应识别样品中的化合物的特征基团,并为机器学习选择最佳处理算法^[7];然后,对MIR光谱进行预处理,以消除或减少测量过程中由样品性质或环境条件产生的变化。光谱预处理包括基线校正、归一化(采用多重散射校正(MSC)和标准正态变量变换(SNV))、平滑和导数预处理(采用一阶导数和二阶导数)^[8-9]。

在应用必要的光谱预处理后,结合主成分分析(PCA)和层次聚类分析(HCA)等分类工具和主成分回归(PCR)、多元线性回归(MLR)和偏最小二乘判别分析(PLS-DA)等判别模型,实现对样品的进一步分析。本文主要使用PCA算法和PLS-DA算法对蜂蜜的检测进行研究。

PCA算法是一种无监督的线性变换算法,通过确定数据的最大方差来生成称为主成分(PC)的新特征。PCA算法将高维数据集投影到一个新的子空间,其中正交轴或PC被视为最大数据方差的方向;在转换过程中,第一个PC的方差最大,随后的PC的方差逐渐减小^[10]。使用PCA算法时,通常以坐标轴PC为因子构建样品的二维或三维投影。每个PC都是原始响应的线性组合(保留了一些相关性),并且PC彼此正交。迭代计算的PC尽可能多地保留原始数据集的变化,PC₁比PC₂解释更多数据变化,PC₂比PC₃解释更多数据变化。依此类推,少数PC就可以解释大量原始数据的变化^[11]。

PLS-DA算法是一种基于偏最小二乘(PLS)算法开发的多元数据判别方法,广泛应用于化学计量学

中,在许多科学领域被广泛应用^[12]。在PLS中,需要一个与分类属性相关的矩阵Y。在这种情况下,创建一个虚拟矩阵Y,其中阳性样本为1,阴性样本为-1。为了开发分类模型,将数据集随机分成2个子集进行训练和测试。样品矩阵X和虚拟矩阵Y对训练集和测试集进行了预处理,PLS-DA模型的潜在变量数量由交叉验证中的最低预测误差标准选择,通过研究校准(RMSEC)、交叉验证(CV)和预测均方根误差(RMSEP)来评估模型质量,分类模型通过测试集进行验证。

2 实验制备与结果分析

2.1 样品的制备

本文实验中的蜂蜜样品分别来自韶关、佛山、北流等13个地区的天然蜂场,均为我国传统蜜蜂品种中华蜂所酿造的中华蜂蜜。其中,含钩吻的有毒蜂蜜样本108个,无毒蜂蜜样本108个,共216个样本。因为蜂蜜中的葡萄糖在温度低于25℃时易结晶,会对MIR光谱采集造成影响,所以需要将所有蜂蜜样品和糖浆掺假样品置于55℃水浴锅中浸泡1 h,以确保去除结晶;再将各样品置于30℃水浴锅中浸泡24 h,以确保完全去除蜂蜜样品中的气泡,并在室温下置于试管中,在试管架上贴上相应的标签,用于MIR光谱采集。

2.2 实验仪器与数据采集

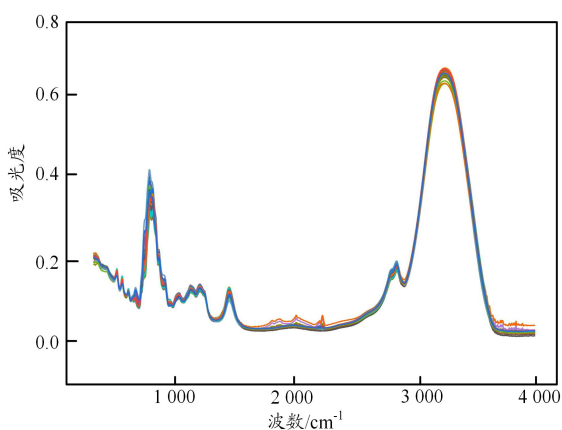
本文实验使用了用于MIR光谱分析的傅里叶MIR光谱仪(ALPHA II,德国)及其传输附件采集MIR光谱。在采集有毒蜂蜜和无毒蜂蜜的MIRS光谱时,将1 mL样品部分置于石英比色皿中,在400~4 000 cm⁻¹光谱范围内以2 cm⁻¹的波数增量记录光谱。测量各样品的光谱数据并取其平均值(实验室温度为26℃,相对湿度为40%)。

2.3 样本划分

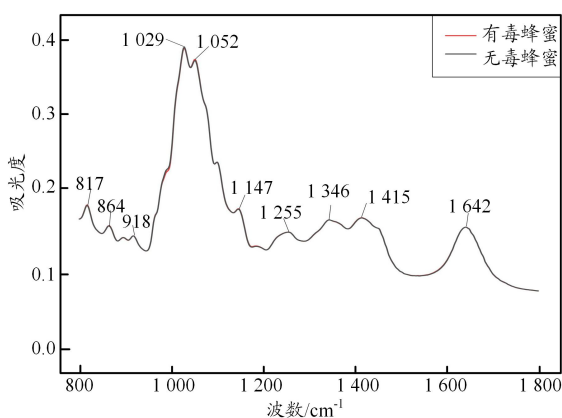
数据处理时,使用光谱-理化值共生距离(SPX Y)方法将所采集的216个含钩吻的有毒蜂蜜及无毒蜂蜜的数据分别按照3:1的比例划分训练集和预测集。SPXY方法由GALVÃO R K H等人^[13]基于凯纳德·斯通(KS)算法提出。与其它样本划分方法相比,SPXY方法在样本间距的计算中同时包含了x变量和y变量的差异,可以有效地应用于MIR定量模型的建立。

2.4 原始光谱分析

首先,对含钩吻的有毒蜂蜜和无毒蜂蜜的原始光谱进行分析,原始光谱图如图1所示。从图1(a)可以看出,在光谱采集400~4 000 cm⁻¹的MIR波段中,



(a) 所有样品原始光谱图



(b) 平均光谱图

图1 原始光谱图

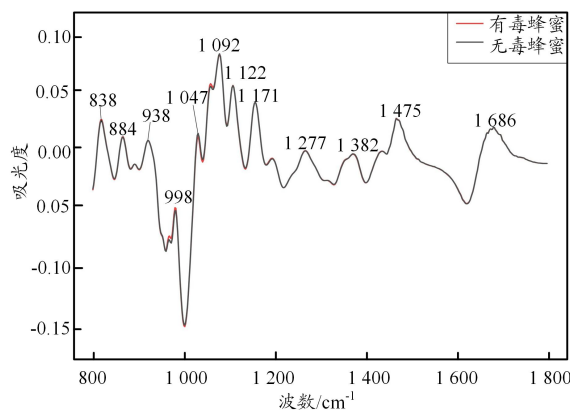
3 400 cm^{-1} 附近的峰是蜂蜜中含有的水分子在 MIR 区域的主要吸收峰,这个大型的吸收峰包含了 H_2O 的反对称和对称伸缩振动、 H_2O 变形振动的一级倍频,而 2 950 cm^{-1} 处的峰则归属为 O-H 伸缩振动和弯曲振动。蜂蜜的 MIR 光谱主要是由蜂蜜中水和糖类物质的吸收产生的,而 3 400 cm^{-1} 处附近的峰的吸光度较高,这种高吸光度主要来源于水的扰动。由于高吸光度会产生高噪音,因此本文选取了吸光度较低的 800~1 800 cm^{-1} 波段作为主要研究波段。从图 1(b)可以看出,该波段最强的峰集中在 900~1 200 cm^{-1} 波段区域,该区域也被称为糖类物质的指纹区域^[14]。其中,在 1 052 cm^{-1} 处归属于葡萄糖中 C-O-H 伸缩振动,在 1 415 cm^{-1} 和 817 cm^{-1} 处归属于果糖,1 029 cm^{-1} 处归属于蜂蜜中 O-H 伸缩振动。在较低能量下,由 C-H 和 O-H 弯曲振动引起的能带也可用于区分和量化目的。另外,出现在 1 147 cm^{-1} 、1 255 cm^{-1} 和 1 346 cm^{-1} 处的峰则归含有 C-C-H、C-O-H 和 O-C-H 基团的弯曲振动。而在 1 642 cm^{-1} 处,由于蜂蜜中有机酸的吸收(醛羰基 C=O 伸缩振动),因

此在此处为一个低位独立峰;此外,含钩吻有毒蜂蜜中的葫曼藤碱因其在 1 029 cm^{-1} 处的 $=\text{CH}_2$ 扭曲振动有别于钩吻碱、钩吻素子和钩吻绿碱的碳碳双键扭曲振动,可作为其主要鉴别区;同时,钩吻素子的脒类化合物 C=N 伸缩振动、钩吻碱和钩吻绿碱的酰胺类 C=O 伸缩振动以及葫曼藤碱的碳碳双键伸缩振动也是平均光谱呈现为低位独立峰的重要原因。因为含有苯环,钩吻碱、钩吻素子和钩吻绿碱可在 817 cm^{-1} 和 1 255 cm^{-1} 处贡献芳香族芳环 CH 面的内、外弯曲扭动,相似的环醚与叔胺官能团存在,也让这 3 种化学成分在 918 cm^{-1} 和 1 147 cm^{-1} 处都有(六元环)环醚 -O- 的伸缩振动和叔胺的 C-N 伸缩振动。但由于钩吻素子中 N 的分布差异,故在 1 417 cm^{-1} 处,不同钩吻碱、钩吻绿碱以及葫曼藤碱产生了 C-N 的伸缩振动。

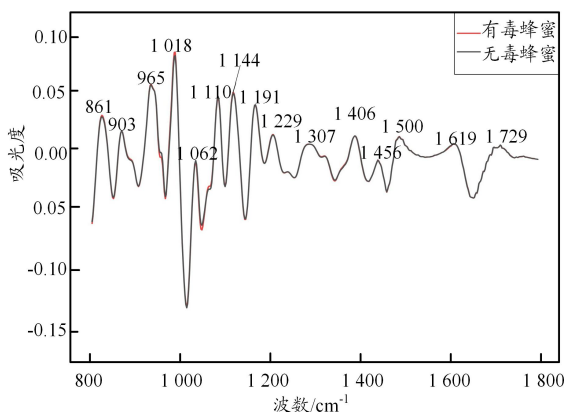
由于原始光谱图中含钩吻的有毒蜂蜜与无毒蜂蜜光谱差异不十分明显,故本文需要进一步处理原始光谱,以求找到更明显的差异^[15]。

2.5 预处理光谱分析

MIR 光谱中包含了蜂蜜样品的特性官能团的信息,但在提取光谱的过程中因受到了噪声等外界因素的干扰,影响模型判断的正确率,所以本文将提取在 800~1 800 cm^{-1} 波段的蜂蜜样品的光谱进行一阶导数、二阶导数、MSC、SNV 这 4 种不同的预处理,其光谱图如图 2 所示。从图 2(a)可以看出,除在 998 cm^{-1} 处的峰值出现相对明显的差异外,含钩吻的有毒蜂蜜和无毒蜂蜜的光谱几乎重叠,此处峰位对应葡萄糖中的 CH_2OH (糖键)的指纹振动也可归因于有毒蜂蜜中的葫曼藤碱的 $=\text{CH}_2$ 扭曲振动或者归因于钩吻碱、钩吻素子和钩吻绿碱的碳碳双键扭曲振动。从图 2(b)可以看出,经二阶导数转换后出现更明显的峰,其中最大的正峰在 1 018 cm^{-1} 处,2 种蜂蜜的二阶导数平均光谱的



(a) 一阶导数平均光谱图



(b) 二阶导数平均光谱图

图2 蜂蜜预处理光谱图

最大差异也出现在此处,与原始光谱中 $1\,029\text{ cm}^{-1}$ 处无毒蜂蜜中 O-H 伸缩振动和有毒蜂蜜中的蒽曼藤碱因的主要鉴别区对应。由图 2 可知,相比于原始光谱图,经过一阶导数和二阶导数预处理后的光谱的峰更多,且峰位明显,同时更加光滑。另外,由于 MSC、SNV 这 2 种预处理方式的核心是归一化,得到的光谱图与原始平均光谱图一致,故本文不对其进行分析。

2.6 PCA 算法

由于 MIR 光谱高度相关,在一定程度上造成数据冗余,而 PCA 算法具有数据简化的能力,能最大程度地保留原始数据的信息。因此,本文将其应用于 $800\sim 1\,800\text{ cm}^{-1}$ 波段的有毒和无毒蜂蜜(共 216 个)样品的 MIR 光谱之中,通过 PCA 算法降低数据维度、减少变量、简化数据,以便对 MIR 光谱数据进行分析,其中前 6 个 PC 的得分情况如表 1 所示。可以看出,前 3 个 PC 的累积贡献率达到了 97.17%,能够代表 216 个蜂蜜样品的基本信息。因此,本文选择前 3 个 PC 来建立含钩吻的有毒蜂蜜和无毒蜂蜜鉴别三维散点图,如图 3 所示。可以看出,含钩吻的有毒蜂蜜和无毒蜂蜜在前 3 个 PC 的区域重叠率和样本分布粘度均较高。由此可见,基于 PCA 算法的有毒、无毒蜂蜜鉴别方法体现了有毒、无毒蜂蜜中一些相同的组成成分(如葡萄糖、

表 1 前 6 个 PC 得分表

类别	差异贡献率/%	累计贡献率/%
PC ₁	86.80	86.80
PC ₂	6.68	93.48
PC ₃	3.69	97.17
PC ₄	2.02	99.19
PC ₅	0.38	99.57
PC ₆	0.17	99.74

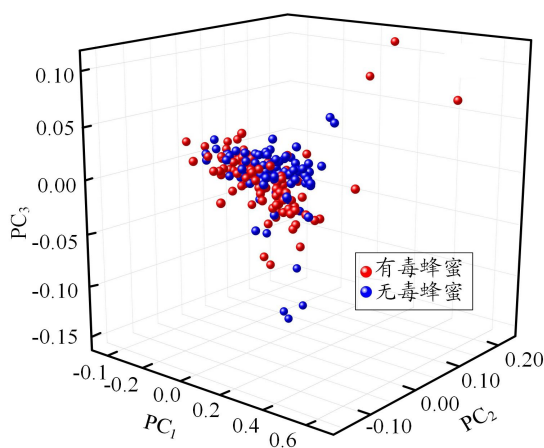


图3 采用 PCA 算法的前 3 个 PC 散点图

果糖、结合水等物质),反映了有毒蜂蜜和无毒蜂蜜主要成分的共同特性。但 PCA 算法对于蜂蜜中是否含有钩吻等有毒物质的区分度有限,需要采用 PLS-DA 算法进行进一步的判别分析。

2.7 PLS-DA 算法模型

为了寻找更好的鉴别有毒蜂蜜和无毒蜂蜜的方法,本文采用经典的线性算法——PLS-DA 算法建立含钩吻的有毒蜂蜜和无毒蜂蜜的判别模型。在建模前,考虑到在采集样品过程中产生的噪音对检测结果的干扰,本文对 MIR 光谱数据进行预处理来提高光谱数据的信噪比以期获得更好的鉴别结果。为尽可能消除噪音的影响,将原始光谱分别进行一阶导数、二阶导数、MSC、SNV 这 4 种不同的光谱预处理方式,有利于建立更加准确的 PLS-DA 算法模型。

基于 4 种预处理方法的 PLS-DA 算法模型预测结果如表 2 所示。可以看出,原始 MIR 光谱数据经一阶导数预处理后建立的模型的最佳 PC 数为 9,判别准确率达到 90.74%,明显高于其它预处理方法,灵敏度和特异性也达到了 92.84%和 88.89%,是鉴别有毒蜂蜜和无毒蜂蜜的最佳 PLS-DA 模型,对含钩吻的有毒蜂蜜具有良好的判别效果。这与在原始光谱图 1(a)中所

表 2 基于 4 种不同预处理方法的 PLS-DA 算法模型预测结果

预处理方法	最佳 PC 数	灵敏度/%	特异性/%	准确率/%
原始光谱	10	70.73	80.25	74.07
一阶导数	9	92.84	88.89	90.74
二阶导数	8	77.78	90.12	85.19
MSC	10	75.31	86.42	81.48
SNV	10	71.60	83.95	77.78

选取吸光度较低的 $800\sim 1\,800\text{ cm}^{-1}$ 波段光谱所体现的较多峰值的出现相吻合,且一阶导数建立的复杂度最低。因此,经一阶导数预处理后建立的有毒蜂蜜鉴别效果最佳。

经过一阶导数处理后的前 3 个 PC 的散点图如图 4 所示。可以看出,含钩吻的有毒蜂蜜和无毒蜂蜜在前 3 个 PC 的区域重叠率和样本分布粘合度明显降低,说明 PLS-DA 算法模型对鉴别含有钩吻的有毒蜂蜜和无毒蜂蜜的效果更好。

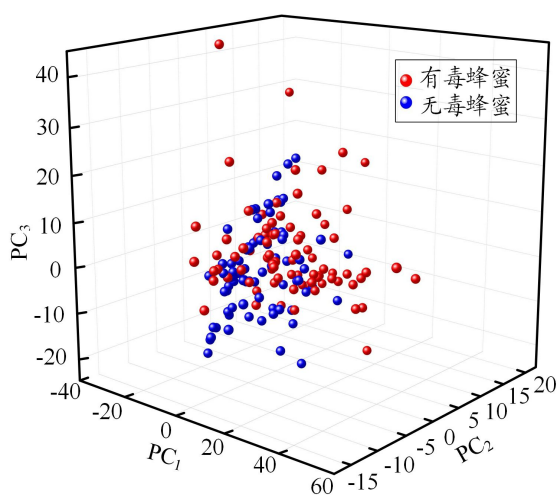


图 4 PLS-DA 算法前 3 个 PC 散点图

采用一阶导数预处理方法建立的 PLS-DA 算法模型对预测集中的 54 个蜂蜜样品进行预测,其结果如表 3 所示。可以看出,预测集中的 27 个无毒蜂蜜样品均被准确识别,而 27 个有毒蜂蜜样品中有 4 个样品被错判为无毒蜂蜜,模型整体判别准确率达到了 90.74%,其中特异性为 100%,灵敏度为 81.48%。因此,选择最佳的预处理方法有助于提高模型鉴别的准确性和可靠性。

表 3 一阶导数的 PLS-DA 算法预测模型

蜂蜜类型	实际集	预测集	灵敏度/%	特异性/%	准确率/%
有毒蜂蜜	27	23	81.48	100.00	90.74
无毒蜂蜜	27	31	81.48	100.00	90.74

3 结束语

本文采集了 216 个来自不同产地的含钩吻的有毒、无毒蜂蜜样品光谱,发现含钩吻的有毒蜂蜜与无毒蜂蜜的 MIR 平均光谱基本重合,无明显差异,不能通过普通的光谱分析得出其中的差异。因此,本文建立 MIR 与 PLS-DA、PCA 等算法相结合的有毒蜂蜜检

测模型,利用 MIR 光谱分析方法提取含有钩吻的有毒蜂蜜和正常无毒蜂蜜的光谱数据。实验结果表明:采用 PLS-DA 算法在 $800\sim 1800\text{ cm}^{-1}$ 波段光谱的 MIR 波段建立有无钩吻毒素判别模型结果,经一阶导数预处理后建立 PLS-DA 算法模型的判别效果最佳,准确率可以达到 90.74%,灵敏度和特异性分别为 92.84%和 88.89%,对含有钩吻的有毒、无毒蜂蜜可以有效地进行鉴别。下一步将在此基础上建立有毒蜂蜜样本库,增加样本来源的广泛性,包括不同地区、不同蜂蜜品种的蜂蜜和不同类型的蜜源植物等,增强 PLS-DA 算法模型对于不同类型有毒蜂蜜的鉴别能力。

参考文献:

- [1] SOARES S, AMARAL J S, OLIVEIRA M B P P, et al. A comprehensive review on the main honey authentication issues: production and origin [J]. Comprehensive Reviews in Food Science and Food Safety, 2017, 16 (5): 1072-1100.
- [2] MATUSZEWSKI B K, CONSTANZER M L, CHAVEZ-ENG C M. Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS[J]. Analytical Chemistry, 2003, 75 (13): 3019-3030.
- [3] MARSTON A, KISSLING J, HOSTETTMANN K. A rapid TLC bioautographic method for the detection of acetylcholinesterase and butyrylcholinesterase inhibitors in plants [J]. Phytochemical Analysis: An International Journal of Plant Chemical and Biochemical Techniques, 2002, 13 (1): 51-54.
- [4] GRAHAM B, MAYOL-BRACERO O L, GUYON P, et al. Water-soluble organic compounds in biomass burning aerosols over Amazonia 1. characterization by NMR and GC-MS[J]. Journal of Geophysical Research: Atmospheres, 2002, 107(D20): LBA 14-1-LBA 14-16.
- [5] WADOOD S A, BOLI G, XIAOWEN Z, et al. Recent development in the application of analytical techniques for the traceability and authenticity of food of plant origin[J]. Microchemical Journal, 2020, 152: 3-39.
- [6] LEARDI R. Chemometric methods in food authentication [J]. Modern techniques for food authentication, 2018, 70: 687-729.
- [7] OLIVERI P, MALEGORI C, CASALE M. 2-Chemometrics: multivariate analysis of chemical data [J]. Chemical Analysis of Food (Second Edition), 2020, 16: 33-76.
- [8] LUNA A S, DE GOIS J S. Application of chemometric methods coupled with vibrational spectroscopy for the discrimination of plant cultivars and to predict physicochemical properties using R[J]. Comprehensive Analytical Chemistry, 2018, 80: 165-194.
- [9] BONA E, MARÇO P H, VALDERRAMA P. Chemometrics applied to food control[J]. Food control and biosecurity, 2018, 10: 105-133.
- [10] ANOWAR F, SADAoui S, SELIM B. Conceptual and empirical comparison of dimensionality reduction algorithms (pca, kpca, lda, mds, svd, lle, isomap, le, ica, t-sne)[J]. Computer Science Review, 2021, 40: 2-4.
- [11] GRANATO D, SANTOS J S, ESCHER G B, et al. Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for mul-

曾薇,谭成森,苏元翠,等:基于中红外光谱的含钩吻蜂蜜鉴别研究

tivariate association between bioactive compounds and functional properties in foods: a critical perspective[J]. Trends in Food Science & Technology, 2018, 72: 83–90.

[12] WOLD S, SJÖSTRÖM M, ERIKSSON L. PLS-regression: a basic tool of chemometrics [J]. Chemometrics and Intelligent Laboratory Systems, 2001, 58(2): 109–130.

[13] GALVÃO R K H, ARAÚJO M C U, MARTINS M D M, et al. An application of subagging for the improvement of prediction accuracy of multi-

variate calibration models [J]. Chemometrics and Intelligent laboratory systems, 2006, 81(1): 60–67.

[14] SANTOS M I, ARAUJO-ANDRADE C, TYMCZYSZYN E E, et al. Determination of amorphous/rubbery states in freeze-dried prebiotic sugars using a combined approach of near-infrared spectroscopy and multivariate analysis[J]. Food Research International, 2014, 64: 514–519.

[15] 翁诗甫,徐怡庄. 傅里叶变换红外光谱分析[M]. (第3版)北京:化学工业出版社,2016.