拉曼光谱结合化学计量学方法鉴别糖浆掺假蜂蜜

寇泽坤[1,2]，陈国通[1]，李思雨[3]，杨　中[1]，欧阳玲秀[4]，龚　龑[2,*]
（1.新疆维吾尔自治区分析测试研究院，新疆 乌鲁木齐　830011；2.新疆大学纺织与服装学院，新疆 乌鲁木齐　830017；
3.新疆农业大学食品科学与药学学院，新疆 乌鲁木齐　830052；4.北京服装学院材料设计与工程学院，北京　100029）

摘　要：为区分掺加糖浆的假蜂蜜，确定其糖浆含量，提出一种以拉曼光谱技术结合化学计量学方法快速鉴别掺假蜂蜜的方法。利用拉曼光谱技术测定蜂蜜样本的光谱数据，利用主成分分析对光谱数据进行特征提取，选取累计贡献率达85%以上的主成分进行建模和预测。通过建立线性判别分析（linear discriminant analysis，LDA）和偏最小二乘判别分析（partial least squares-discriminant analysis，PLS-DA）模型，能够判别掺假蜂蜜中20%的糖浆含量差异。通过建立支持向量机（support vector machine，SVM）模型，能够判别掺假蜂蜜中5%的糖浆含量差异，且LDA、PLS-DA和SVM皆能以0.9以上的准确率区分1%糖浆含量的掺假蜂蜜样本和真蜂蜜样本。拉曼光谱技术结合化学计量学方法是一种快速无损、准确率高的掺假蜂蜜鉴别方法，其为蜂蜜及蜂蜜产品的快速鉴定提供了一种可行的思路，对维持蜂蜜市场秩序具有一定的意义。

关键词：蜂蜜；糖浆掺假；拉曼光谱技术；主成分分析；线性判别分析；偏最小二乘判别分析；支持向量机

Identification of Honey Adulterated with Syrup by Raman Spectroscopy and Chemometrics

KOU Zekun[1,2], CHEN Guotong[1], LI Siyu[3], YANG Zhong[1], OUYANG Lingxiu[4], GONG Yan[2,*]
(1. Xinjiang Uygur Autonomous Region Institute for Analysis and Testing, Ürümqi　830011, China;
2. School of Textile and Clothing, Xinjiang University, Ürümqi　830017, China;
3. College of Food Science and Pharmacy, Xinjiang Agricultural University, Ürümqi　830052, China;
4. School of Materials Design & Engineering, Beijing Institute of Fashion Technology, Beijing　100029, China)

Abstract: In order to qualitatively and quantitatively identify syrup adulteration in honey, a method for rapid identification of adulterated honey by Raman spectroscopy and chemometrics was proposed. Raman spectroscopy was used to acquire spectral data of honey samples, and principal component analysis (PCA) was used to extract features from the spectral data. Principal components with a cumulative contribution rate of more than 85% were selected for modeling and prediction. By using linear discriminant analysis (LDA) and partial least squares-discriminant analysis (PLS-DA), models to identify honey adulterated with 20% syrup were established. A support vector machine (SVM) model to identify honey adulterated with 5% syrup, and all LDA, PLS-DA and SVM models could distinguish adulterated honey samples with 1% syrup content from pure honey with an accuracy of more than 0.9. Raman spectroscopy combined with chemometrics is a fast and non-destructive method for the identification of adulterated honey with high accuracy, which is significant to maintaining the order of the honey market.

Keywords: honey; adulteration with syrup; Raman spectroscopy; principal component analysis; linear discriminant analysis; partial least squares-discriminant analysis; support vector machine
DOI:10.7506/spkx1002-6630-20230323-230
中图分类号：O657.37　　文献标志码：A　　文章编号：1002-6630（2024）01-0254-07
引文格式：
寇泽坤, 陈国通, 李思雨, 等. 拉曼光谱结合化学计量学方法鉴别糖浆掺假蜂蜜[J]. 食品科学, 2024, 45(1): 254-260. DOI:10.7506/spkx1002-6630-20230323-230.　http://www.spkx.net.cn

收稿日期：2023-03-23
基金项目：新疆维吾尔自治区科技支疆项目（2020E02122）；新疆维吾尔自治区自然科学基金项目（2022D01A113）；新疆维吾尔自治区天山创新团队计划项目（202110498）；国家自然科学基金地区科学基金项目（21964015）
第一作者简介：寇泽坤（1998—）（ORCID: 0009-0006-6389-7099），男，硕士研究生，研究方向为拉曼光谱技术与应用。E-mail: zx13791683008@163.com
*通信作者简介：龚龑（1980—）（ORCID: 0000-0003-0375-4986），男，教授，博士，研究方向为拉曼光谱技术与应用。E-mail: 2205206742@qq.com

KOU Zekun, CHEN Guotong, LI Siyu, et al. Identification of honey adulterated with syrup by Raman spectroscopy and chemometrics[J]. Food Science, 2024, 45(1): 254-260. (in Chinese with English abstract) DOI:10.7506/spkx1002-6630-20230323-230. http://www.spkx.net.cn

蜂蜜作为一种营养丰富、用途广泛的天然食品，不仅是一种营养品，而且在医用、药用方面发挥着独到的作用[1]。近年来，各种低制作成本的假蜂蜜悄然出现并流入市场，造成局部市场蜂蜜产品鱼龙混杂。蜂蜜造假的方式多种多样，常见的有直接用糖浆和色素等制成假蜂蜜、真假蜂蜜混合制成掺假蜜、除去未成熟蜂蜜中的多余水分制成浓缩蜜和给蜜蜂喂养白糖等。其中真蜂蜜掺加假蜂蜜或糖浆的掺假方式较容易实现，制出的掺假蜜形貌与味道变化不大，有些理化指标还获得了加强[2]，也因此品尝和肉眼观察等方法不能直接准确地分辨出真假[3]。

因为天然蜂蜜的主要成分是果糖、葡萄糖和水分，故以其为主要成分的糖浆成为了掺假物的首选[4]。使用果糖、葡萄糖和蜂蜜香精等制成味道、形貌和理化指标都和真蜂蜜相近的糖浆，并掺加到真蜂蜜中可以模拟掺假蜂蜜的制作过程。

GB 14963—2011《蜂蜜卫生标准》规定了蜂蜜不同指标对应的检测手段，诸如液相色谱、紫外光谱和液相色谱-质谱联用等。基于上述检测手段不同程度存在前处理复杂或无法脱离实验室场景等缺点，提出利用拉曼光谱技术结合化学计量学方法对掺加蜂蜜进行现场快速鉴别的方法[5-6]，并对数据分析模型进行评价。

1 材料与方法

1.1 材料与试剂

伊犁百信黑蜂蜂蜜、伊犁天山黑蜂结晶白蜜、唐布拉黑蜂薰衣草原蜜、尼勒克山花蜜。

果糖（分析纯） 上海蓝季生物公司；葡萄糖（分析纯） 天津市盛奥化学试剂有限公司；蜂蜜香精 可菲生物科技有限公司。

1.2 仪器与设备

SSR-3000拉曼光谱检测仪 南京简智仪器设备有限公司；AL204-IC电子天平 上海梅特勒-托利多仪器有限公司；SHA-CA水浴恒温振荡器 北京市光明医疗仪器有限公司；MS3basic微量振荡器 德国IKA集团；石英进样瓶 安捷伦科技（中国）有限公司。

1.3 方法

1.3.1 样本制作及采样

1.3.1.1 真蜂蜜采样

将4 种蜂蜜参照进行GB 14963—2011的相关检验，证实样品均属于符合国标要求的天然蜂蜜。

蜂蜜样本存放于4 ℃环境下保存，测试之前取出，并于50 ℃水浴1 h以上直至蜂蜜样本中的所有结晶物溶于蜂蜜中后，将样本取出并于25 ℃环境下放置至恢复常温且气泡消失，待用。

样品采集时，取1 g蜂蜜样本置于离心管中，加入1 mL超纯水，涡旋振荡至溶液均匀。采集时每个样本加入石英进样瓶上机，结果取平均光谱进行分析，假蜂蜜和掺假蜂蜜样品采集亦如此。

如图1所示，对4 种蜂蜜样本进行拉曼光谱采集后，发现4 种蜂蜜的谱图形状相近，特征峰位置几乎相同，强度上有所差异，为了避免真蜂蜜样本间的差异影响鉴别结果，选择信号强度较好的山花蜜作为掺假鉴别的对象。

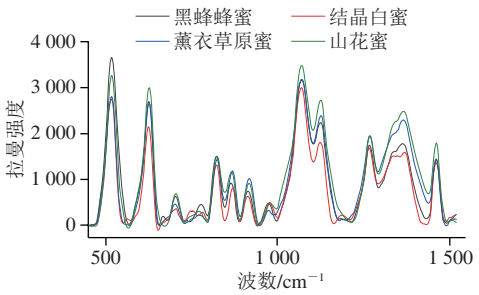

图 1 4 种真蜂蜜样本的拉曼光谱图

Fig. 1 Raman spectra of four pure honey samples

1.3.1.2 糖浆制作

糖浆样本采用果糖、葡萄糖和蔗糖进行制作，按照GB 14963—2011的相关检验要求蔗糖质量分数不超过5%，葡萄糖和果糖质量分数不低于60%。制作时将果糖、葡萄糖和蔗糖和适量超纯水一并加入并加热至180 ℃，待所有糖完全溶解且保证溶液冷却后的黏度、形貌和色泽等较为接近真蜂蜜时加入少量香精，搅拌并冷却待用。

1.3.1.3 掺假蜂蜜制作

掺假蜂蜜样本的制备通过山花蜜掺加糖浆样本进行制作，并且引入掺假度R_f的量化指标：

$$R_f/\% = \frac{m_f}{m_r} \times 100 \quad (1)$$

式中：m_f和m_r分别为糖浆质量和真蜂蜜质量。

本实验将设置1%、5%、10%、15%、20%、30%、40%、50%共8 个掺假度梯度，每个掺假度10 个样本，共80 个假蜂蜜样本，与山花蜜纯品蜂蜜进行比较判别[7]，并在此将上述种类的掺假样本依次命名为F1、F5、F10、F15、F20、F30、F40、F50，真山花蜜样本则命名为R，方便进行数据分析。

1.3.2 拉曼光谱工作条件

扫描范围180～3 500 cm^{-1}，激光光源采用785 nm，积分时间2 000 ms，激光功率250 mW，平均次数3 次。

1.3.3 样本集划分

在建立线性判别分析（linear discriminant analysis，LDA）和偏最小二乘判别分析（partial least squares-discriminant analysis，PLS-DA）和支持向量机（sport vector machine，SVM）模型时，训练集和测试集的划分比例往往会对训练结果产生影响，比例失衡会影响模型的性能，甚至造成过拟合或欠拟合等现象，通常训练集与测试集的比例在6∶4与8∶2之间较为适宜[8]。采用留出法[9]划分训练集和测试集，即将数据集划分为两个互斥的子集。经验证，7∶3的划分比例在训练集上的准确率优于其他比例且和交叉验证的总体准确率相差不超过0.1，故将比例统一确定为7∶3，各类样本的样品集均包含35 个训练集和15 个测试集，9 类共315 个训练集和135 个测试集，总计450 个样本。

1.3.4 数据预处理

1.3.4.1 背景扣除与平滑处理

在拉曼光谱测试中，除去容器和环境等的影响产生的背景光，实验样品本身也会出现荧光现象，如来自羰基、硝基和乙烯基等常见荧光发色团的荧光背景。荧光的产生会降低拉曼光谱的信噪比，掩盖光谱中的重要信息，所以需要采取一定的方法扣除荧光产生的基线漂移。基线扣除的常见方法有对样本做表面增强等预处理手段和通过airPLS等化学计量学方法进行扣除等[10]。本实验中采取软件自带的扣除基线算法进行处理。

在拉曼光谱检测中，因为激发激光光强的漂移、CCD检测器热稳定噪声、样品放置位置与方向等多方面因素的影响，拉曼信号可能会有比较大的噪声，在波形上表现为剧烈波动的锯齿状或毛刺状信号。一般来讲，信号平滑处理的方式有Savitzky-Golsy平滑（SG平滑）、相邻平均（adjacent averaging，AAV）法和小波变换（wavelet transform，WT）去噪等。

采取SG平滑的方法进行处理。SG平滑滤波是一种移动窗口的加权平均算法，在滤除信号中的噪声的同时保证波形的形状和宽度不发生改变，处理后的拉曼光谱与原数据波形较为接近，且提高了信号信噪比。光谱在波长i处经SG平滑处理后的数值为：

$$x_{i,\mathrm{SG}}=\frac{\sum_{j=-m}^{m}c_j\,x_{i+j}}{N} \tag{2}$$

式中：m为波长一侧的平滑窗口数，$2m+1$则为总平滑窗口数；$N=\Sigma_{j=-m}^{m}c_j$为归一化指数；c_j为多项式拟合得到的平滑系数；x_{i+j}为光谱在波长$i+j$处的数值。在本实验中，选择15窗口点数与二阶多项式进行平滑处理。

1.3.4.2 均值中心化处理

均值中心化是将样品光谱数据集的每一个元素减去该元素所在列的均值的处理方法，经过均值中心化处理后的第i行第j列的元素如下：

$$X'_{ij}=X_{ij}-X_j \tag{3}$$

式中：X_{ij}为原始数据矩阵X的第i行第j列的元素；X_j为X第j列的n 个样本的平均值；X'_{ij}为均值中心化处理之后的第i行第j列的元素。经过均值中心化处理的数据矩阵具有每一列的均值都等于零的性质，样品光谱之间的差异性会被放大，模型的稳定性和预测能力会得到一定程度的提高。原始光谱数据经过背景扣除、平滑处理和均值中心化后方可作为输入数据。

1.3.5 建模方法

1.3.5.1 主成分分析（principal component analysis，PCA）降维

PCA是一种获得原始数据矩阵中主要信息的无监督的线性变换算法，它通过降低数据维度简化复杂的数据集[11]，通过确定数据的方差产生新的特征，称为PC。产生的第1个PC具有最高的方差，随后的PC方差递减[12]。

用贡献率对每个PC代表原变量的能力进行量化。累计贡献率越高，代表PC综合原变量的程度越高，一般选取累计贡献率达85%以上的前几个PC作为PCA的结果[13]。

1.3.5.2 LDA

LDA是基于类别之间的马氏距离最大的判别思想[14]，使变换后类间距离最大、类内距离最小，以寻找对分类最有帮助的特征向量[15]。LDA算法的核心使选择某个投影方向，使得投影后样本类间具有尽可能大的离散度而样本类内具有尽可能小的离散度[16]。

1.3.5.3 PLS-DA

PLS-DA是一种有监督的多变量统计分析方法，采用经典的偏最小二乘回归模型[17]，其将变量数据与分类信息划分为两组数据集，将降维分析与组类别相结合，从而度样本进行区分[18]。

1.3.5.4 SVM

SVM的主要思想是寻找某个超平面，使得它能够尽可能多地将两类数据点正确分开，并且使分开的两类数据点离分类面的距离最远。对于重合区域比较大，线性分类难度比较高的SVM分类问题，即通过引进输入空间到另一个高维空间的变换，将原输入空间的训练集转化为高维空间中新的训练集，并使其在高维空间线性可分，或利用核函数进一步计算并构造分类函数[19]，此时涉及到核函数的选取问题，选择合适的核函数可以使映射到特征空间的样品点类间混合程度降低，使得数据集类间线性可分的程度更高[20]。常用的核函数有线性内核函数、多项式核函数、径向基核函数等[21]。

1.3.5.5 交叉验证

在PCA-LDA、PLS-DA和SVM交叉验证时，循环方法往往有五折、十折和留一法。在执行交叉验证循环时，选取每一种方法进行操作，将准确率最高的方法作为模型适用的方法，PCA-LDA和PLS-DA为留一法，SVM为十折法最佳。

1.3.6 掺假度鉴别实验与真假鉴别实验

实验主要分为假蜂蜜掺假度鉴别和真假蜂蜜鉴别两方面进行。掺假度鉴别即为20%梯度（F10、F30、F50）和10%梯度（F10、F20、F30、F40、F50）和5%梯度（F5、F10、F15）掺假蜂蜜建立分类模型，若某模型可以在3 次或以上的交叉验证中对上述分类的掺假蜂蜜保持0.9以上的总体准确率，则认定该模型可以达到对应该梯度的掺假度辨别能力。若某模型无法在上述5 类掺假度蜂蜜分类的多次训练中一直保持0.9以上的总体准确率，则认定其无法达到掺假度鉴别要求。在进行模型训练与评价时，先从10%梯度的掺假度鉴别开始。成功则继续进行5%梯度鉴别，失败则进行20%梯度。

在真假蜂蜜对比中，利用真样本R分别与F10、F5、F1等掺假度梯度的样本各自配对后进行PCA降维及后续的一系列建模分析，同样认定可以在3 次或以上的交叉验证中对R与某掺假样本具有0.9以上的总体准确率的模型为有效区分模型，具备区分真蜂蜜和该掺假度假蜂蜜的能力。

PCA降维方法与PC选取原则通用，均为将预处理后的光谱数据进行PCA，选取累计贡献率达85%以上的前 n 个PC作为光谱数据的降维结果。

1.4 数据计算

使用准确率、灵敏度、特异性、F1-Score、Macro-F1 Score和GScore对模型性能进行评价[22]。准确率是指在分类模型中，已建立的模型在通过测试集测试时，被正确判别的样本占总样本数的比例，灵敏度是指正样本被正确分类的百分率，特异性是指负样本被正确分类的百分率。F1-Score常用以度量二分类问题的模型特征识别能力，对于多种分类的模型则需要其他评价方法，此处使用Macro-F1 Score和GScore[23-24]两种度量方式，均为F1-Score向多分类集中相关性度量问题的推广，上述评价指标计算公式如下：

$$准确率=\frac{n_c}{n_t} \tag{4}$$

$$灵敏度=\frac{TP}{TP+FN} \tag{5}$$

$$特异性=\frac{FP}{FP+TN} \tag{6}$$

$$F1\text{-}Score=\frac{2PR}{P+R} \tag{7}$$

$$P=\frac{TP}{TP+FP};\ R=\frac{TP}{TP+FN} \tag{8}$$

$$Macro\text{-}F1\ Score=\frac{2P_{Ma}R_{Ma}}{P_{Ma}+R_{Ma}} \tag{9}$$

$$P_{Ma}=\sum_{i=1}^{n}\frac{1}{n}P_i;\ R_{Ma}=\sum_{i=1}^{n}\frac{1}{n}R_i \tag{10}$$

$$GScore=\frac{\sum_{j=1}^{l}(\bar{x}_i^{(j)}-\bar{x}_i)^2}{\sum_{j=1}^{l}\frac{1}{m_j-1}\sum_{k=1}^{m_j}(x_{k,i}^{(j)}-\bar{x}_i^{(j)})^2} \tag{11}$$

式中：n_c为测试集数据被模型正确分类的样本个数；n_t为测试集的所有样本个数；TP、FN、FP和TN分别为真正样本、假负样本、假正样本和真负样本的数量；P为查准率；R为查全率计算同灵敏度；P_{Ma}和R_{Ma}分别为P和R对应所有类别的均值；$\bar{x}_i^{(j)}$为第j类数据集上的第i个特征的均值；$\bar{x}_i$为整个数据集上第i个特征的均值；l为总样本个数（$l\geq2$）；m为某类样本个数；$x_{k,i}^{(j)}$为第j类数据集的第k个样本的第i个特征值。

2 结果与分析

2.1 背景扣除与SG平滑处理结果

采取拉曼光谱处理算法AutoBaseline进行背景扣除。采取SG平滑算法进行平滑处理，平滑处理时选择的窗口点数是对平滑效果有决定性作用的参数。采取相关系数和残差均方根评价对上述处理的结果进行选优，确定SG平滑点数为15。原始光谱与经背景扣除、SG平滑处理后光谱的对比图见图2，可以观察到幅度小而杂乱的信号波动被基本除去，特征峰的相对强度和峰宽等重要信息得以保留。背景扣除后，原光谱曲线落入下方，光谱整体的强度区间缩小。

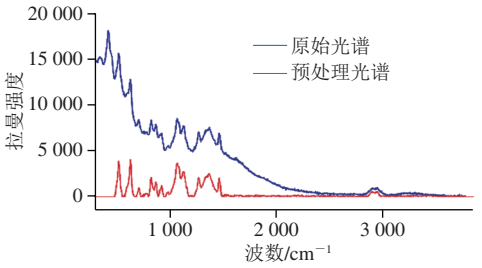

图 2 原始光谱与经背景扣除、SG平滑处理光谱对比图

Fig. 2 Comparison of original spectra before and after background subtraction and Savitzky-Golsy smoothing

2.2 蜂蜜拉曼光谱分析

天然蜂蜜与掺假蜂蜜的拉曼光谱对比图见图3。可以观察到两种谱图在形状上大致相似[25]，但部分特征峰的形状存在差异，同时光谱重叠比较严重，特征峰的位置较为接近，肉眼判别区分谱图的方式难以实现。

同时通过观察天然蜂蜜光谱，可知实验所用山花蜜在422、520、627、705、819、864、916、1 071、1 123、1 265、1 361 cm^{-1}和1 461 cm^{-1}处存在峰。其中，705 cm^{-1}对应—CO—和CCO键的伸缩振动、OCO键的弯曲振动；864 cm^{-1}对应CH(12)；819 cm^{-1}对应C(1)H；916 cm^{-1}与C(1)—H和COH相关；1 071 cm^{-1}与碳水化合物中的C—H和蛋白质和氨基酸中的C—N基团有关；1 123 cm^{-1}与糖中的C—O和氨基酸中的C—N基团有关；1 265 cm^{-1}用于定量C(6)—OH和C—OH；1 461 cm^{-1}则与C—H和—COO—基团有关[2]。

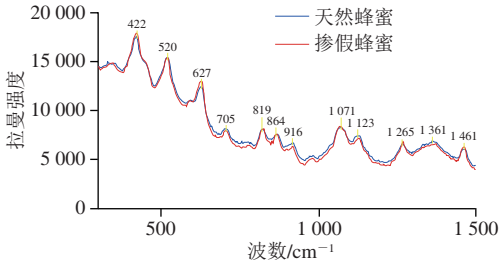

图 3 天然蜂蜜与掺假蜂蜜光谱对比图

Fig. 3 Spectral comparison between natural honey and adulterated honey

2.3 PCA降维和特征提取

进行10%掺加度判别的模型训练时，将对应的5 类样本的光谱数据输入PCA模型进行降维和特征提取。F10、F20、F30、F40和F50共250 条光谱数据经PCA降维后，前3 个PC累计贡献率达63.64%，前7 个PC累计贡献率达85.14%。于是，选择前7 个PC作为上述5 类掺假蜂蜜建立分类模型使用的光谱数据的PCA降维结果[26]。由图4可以观察到F10和F50的得分点分布范围可分程度比较高，但F20、F30和F40的PC1和PC得分点彼此之间入侵和重合的现象比较严重，肉眼观察和线性划分均无法准确地对5 个掺假度进行聚类分析，故需要将上述前7 个PC作为LDA输入数据，构建PCA-LDA模型进行进一步的判别[27]。

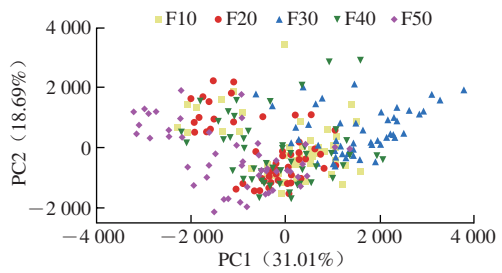

图 4 PC1和PC2得分情况

Fig. 4 Scores of the first and second principal components

2.4 PCA-LDA、PLS-DA建模与模型性能

2.4.1 PCA-LDA模型10%梯度鉴别

采用PCA与LDA相结合的方法，将PCA特征提取结果输入LDA模型进行判别[28]。在10%梯度的LDA模型掺假度鉴别中，F40样本发生了较为严重的错判现象，35 个测试集样本共错判13 个，其中判为F20类5 个、F30类6 个、F50类2 个，单类GScore低至0.715 7。其他4 个样品的错判数都在6 个及以下，交叉验证的总体准确率都达到了0.92以上，但由于受F40影响，模型总体准确率为0.84，Macro-F1 Score为0.8373，未达到0.9的标准。

2.4.2 PLS-DA模型10%梯度鉴别

在10%梯度的PLS-DA模型掺假度鉴别中。F10和F50作为掺假度区间的最小值和最大值，只存在相邻一个样本的错判。而F20、F30和F40错判现象比较严重，GScore都低于0.8。PLS-DA模型的总体准确率为0.8057，Macro-F1 Score为0.806 4，未达到0.9的标准。故PLS-DA模型也未达成10%梯度的掺假度鉴别目标。

2.4.3 PCA-LDA和PLS-DA模型20%梯度鉴别

利用PCA-LDA模型进行20%梯度的掺假度鉴别，即F10、F20和F30的分类判别。在20%梯度的掺假度鉴别中，PCA-LDA模型GScore都在0.94以上，总体准确率为0.961 9，Macro-F1 Score为0.961 8。

PLS-DA模型GScore都在0.95以上，总体准确率0.971 4，Macro-F1 Score为0.971 4，达到了总体准确率的最低要求。表1为20%梯度PCA-LDA与PLS-DA模型交叉验证性能评价。

表 1 20%梯度PCA-LDA与PLS-DA梯度模型性能评价

Table 1 Performance evaluation of PCA-LDA and PLS-DA models in discriminating honey samples adulterated at 20% gradient

类别	PCA-LDA			PLS-DA		
	F10	F30	F50	F10	F30	F50
灵敏度	0.885 7	1	1	0.942 9	0.971 4	1
特异性	1	0.942 9	1	0.985 7	0.971 4	1
准确率	0.961 9	0.961 9	1	0.971 4	0.971 4	1
Macro-F1 Score	0.939 4	0.945 9	1	0.956 5	0.957 7	1
GScore	0.941 1	0.947 3	1	0.956 6	0.957 8	1

2.4.4 PCA-LDA和PLS-DA模型真假鉴别

在真蜂蜜与掺假5%蜂蜜的鉴别中，PCA-LDA和PLS-DA模型的总体准确率达到了1，不存在错判现象，故进行掺假1%蜂蜜的鉴别。

如表2所示，在真蜂蜜与掺假1%蜂蜜的鉴别中，PCA-LDA存在一定的错判现象，但总体准确率达到0.9，Macro-F1 Score为0.899 8。而PLS-DA总体准确率为0.971 4，Macro-F1 Score为0.971 4，达到了总体准确率的最低要求。其中PLS-DA的准确率达到了0.97以上，GScore均到达了0.97以上，相比PCA-LDA具有更高的判别准确率。

表 2 PCA-LDA与PLS-DA真假鉴别模型性能评价

Table 2 Performance evaluation of PCA-LDA and PLS-DA models in discriminating pure from adulterated honey

类别	PCA-LDA		PLS-DA	
	R	F1	R	F1
灵敏度	0.857 1	0.942 9	1	0.942 9
特异性	0.942 9	0.857 1	0.942 9	1
准确率	0.900 0	0.900 0	0.971 4	0.971 4
Macro-F1 Score	0.895 5	0.904 1	0.972 2	0.970 6
GScore	0.896 4	0.904 9	0.972 6	0.971 0

2.5 SVM建模与模型性能

SVM模型训练的步骤是，若SVM模型使用线性核函数即可通过调整参数在10%梯度的掺假度鉴别中达到0.9以上总体准确率，则可以进行5%梯度的鉴别，若总体准确率还可维持0.9以上，则进行1%梯度的鉴别，过程中总体准确率低于0.9时，再使用其他核函数并将进行调参，直至总体准确率达标。

SVM模型的输入数据选择预处理数据进行建模，通过调试核函数参数优化SVM模型性能，观察模型训练集和测试集中的准确率变化情况，若某核函数模型在训练集上准确率较高，在测试集上的准确率却很低，则认为该核函数模型出现过拟合。选择准确率较高且未出现过拟合现象的核函数进行预测[29]。核函数在线性核函数、多项式核函数和径向基核函数中选用[30]，在调整参数过程中选优[31]。

2.5.1 SVM模型5%梯度鉴别

SVM模型在10%梯度的掺假度鉴别中总体准确率达到了1，不存在错判现象，故进行掺假5%梯度蜂蜜的鉴别。线性核函数SVM在F5和F15判别中不存在错判现象，有3 个F10样本错判为F5，F15单类评价参数均为1，如表3所示，F5的F1-Score和GScore到达0.9以上，F10的F1-Score为0.888 9，GScore为0.894 4，SVM模型总体准确率为0.93，Macro-F1 Score为0.932 7，达到了总体准确率的最低要求，最优c值为0.312 5，$\log_2 c=-1.678$，图5为随线性核SVM准确率随c值变化的曲线。

表 3 SVM梯度鉴别模型性能评价

Table 3 Performance evaluation of SVM model in discriminating honey samples adulterated with different proportions of syrup

类别	F5	F10	F15
灵敏度	1	0.800 0	1
特异性	0.800 0	1	1
准确率	0.933 3	0.933 3	1
F1-Score	0.909 1	0.888 9	1
GScore	0.912 9	0.894 4	1

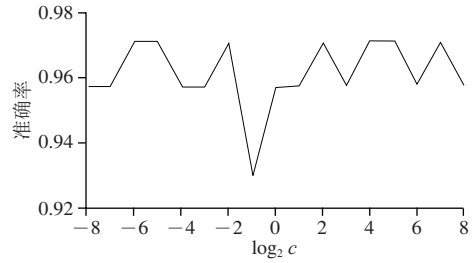

图 5 线性核SVM的$\log_2 c$值与准确率关系图

Fig. 5 Relationship between $\log_2 c$ value and accuracy of linear kernel SVM

2.5.2 SVM模型真假鉴别

SVM模型在真蜂蜜与5%掺假度蜂蜜的鉴别中总体准确率达到了1，不存在错判现象，故进行真蜂蜜和掺假度1%蜂蜜的鉴别。

在应用线性核时，R不存在错判现象，有1 个F1样本错判为R。R和F1的GScore和F1-Score均达到了0.96以上，总体准确率0.966 7（表4）。Macro-F1 Score为0.966 6。在此SVM模型中，使用的是线性核函数，最优c值为0.156 25，$\log_2 c=-2.678$。

表 4 SVM真假鉴别模型性能评价

Table 4 Performance evaluation of SVM model in discriminating pure from adulterated honey

类别	R	F1
灵敏度	1	0.933 3
特异性	0.933 3	1
准确率	0.966 7	0.966 7
F1-Score	0.967 7	0.965 5
GScore	0.968 2	0.966 1

在应用径向基核函数时，均不存在错判现象，各评价指标均为1，最优c值为16，最优g值为4 096，$\log_2 c=12$。SVM准确率与$\log_2 c$、$\log_2 g$值的关系见图6。

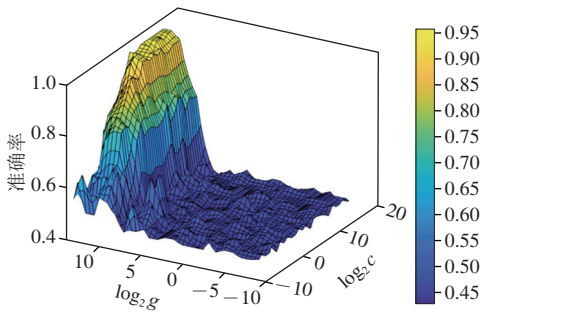

图 6 径向基核SVM的准确率与$\log_2 c$、$\log_2 g$值的关系图

Fig. 6 Relationship between $\log_2 c$ and $\log_2 g$ values and accuracy of radial basis function SVM

3 结 论

综上所述，在蜂蜜掺假度鉴别中，PCA-LDA和PLS-DA模型可以以0.9以上的准确率做到20%掺假度区别的蜂

蜜鉴别，线性核函数SVM则可以达到5%精度的掺假度区别；在真假蜂蜜鉴别中，上述3 个模型均可以做到1%掺假度蜂蜜和真蜂蜜的区分，其中PCA-LDA、PLS-DA和线性核SVM总体准确率在0.9以上，径向基SVM总体准确率为1。

参考文献：

[1] GUTTENTAG A, KRISHNAKUMAR K, COKCETIN N, et al. Inhibition of dermatophyte fungi by australian Jarrah honey[J]. Pathogens, 2021, 10(2): 194. DOI:10.3390/pathogens10020194.

[2] PARADKAR M M, IRUDAYARAJ J. Discrimination and classification of beet and cane inverts in honey by FT-Raman spectroscopy[J]. Food Chemistry, 2002, 76(2): 231-239. DOI:10.1016/S0308-8146(01)00292-8.

[3] SAHLAN M, KARWITA S, GOZAN M, et al. Identification and classification of honey's authenticity by attenuated total reflectance Fourier-transform infrared spectroscopy and chemometric method[J]. Veterinary World, 2019, 12(8): 1304-1310. DOI:10.14202/vetworld.2019.1304-1310.

[4] 郑优, 王欣, 毛锐. 蜂蜜常见的掺假类型及真伪鉴别方法的研究进展[J]. 食品与发酵科技, 2018, 54(6): 75-82; 104. DOI:10.3969/j.issn.1674-506X.2018.06-015.

[5] OROIAN M, ROPCIUC S, PADURET S. Honey adulteration detection using Raman spectroscopy[J]. Food Analytical Methods, 2018, 11(4): 959-968. DOI:10.1007/s12161-017-1072-2.

[6] 相倩倩, 张云权, 王小花, 等. 化学计量学方法在蜂蜜鉴伪中的应用研究进展[J]. 江苏农业科学, 2020, 48(8): 32-40. DOI:10.15889/j.issn.1002-1302.2020.08.006.

[7] 杨心浩. 基于红外光谱分析蜂王浆品质及鉴别麦卢卡蜂蜜掺假的方法研究[D]. 广州: 暨南大学, 2020. DOI:10.27167/d.cnki.gjinu.2020.000827.

[8] 展晓日, 朱向荣, 史新元, 等. SPXY样本划分法及蒙特卡罗交叉验证结合近红外光谱用于橘叶中橙皮苷的含量测定[J]. 光谱学与光谱分析, 2009, 29(4): 964-968. DOI:10.3969/j.issn.1002-6819.2014.06.030.

[9] PU H, SUN D W, MA J, et al. Hierarchical variable selection for predicting chemical constituents in lamb meats using hyperspectral imaging[J]. Journal of Food Engineering, 2014, 143: 44-52. DOI:10.1016/j.jfoodeng.2014.06.025.

[10] 李水芳, 张欣, 李姣娟, 等. 拉曼光谱法无损检测蜂蜜中的果糖和葡萄糖含量[J]. 农业工程学报, 2014, 30(6): 249-255. DOI:10.3969/j.issn.1002-6819.2014.06.030.

[11] CORVUCCI F, NOBILI L, MELUCCI D, et al. The discrimination of honey origin using melissopalynology and Raman spectroscopy techniques coupled with multivariate analysis[J]. Food Chemistry, 2015, 169: 297-304. DOI:10.1016/j.foodchem.2014.07.122.

[12] 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: 100378. DOI:10.1016/j.cosrev.2021.100378.

[13] 王帅星, 黄茜, 王晓笋, 等. WPT、PCA与SVM结合的滚动轴承故障程度诊断[J]. 机械设计与制造, 2022(4): 5-9. DOI:10.19356/j.cnki.1001-3997.20211111.024.

[14] ZHU F, GAO J, YANG J, et al. Neighborhood linear discriminant analysis[J]. Pattern Recognition, 2022, 123: 108422. DOI:10.1016/j.patcog.2021.108422.

[15] 黄忠民, 曾万聃, 吴敏, 等. 基于PCA和LDA的食源性致病菌拉曼光谱分类识别[J]. 激光杂志, 2022, 43(9): 55-59. DOI:10.14016/j.cnki.jgzz.2022.09.055.

[16] LEI T, LIN X H, SUN D W. Rapid classification of commercial Cheddar cheeses from different brands using PLSDA, LDA and SPA-LDA models built by hyperspectral data[J]. Journal of Food Measurement and Characterization, 2019, 13(4): 3119-3129. DOI:10.1007/s11694-019-00234-0.

[17] 黄富荣, 宋晗, 郭鎏, 等. 近红外光谱结合化学计量学的常见中国蜂蜜掺杂糖浆鉴别[J]. 光谱学与光谱分析, 2019, 39(11): 3560-3565. DOI:CNKI:SUN:GUAN.0.2019-11-047.

[18] 阿基业. 代谢组学数据处理方法: 主成分分析[J]. 中国临床药理学与治疗学, 2010, 15(5): 481-489. DOI:CNKI:SUN:YLZL.0.2010-05-002.

[19] 李雨珊. 基于高光谱数据的甜瓜嫁接愈合预测[D]. 武汉: 华中农业大学, 2022. DOI:10.27158/d.cnki.ghznu.2022.001434.

[20] 吴桂芳, 何勇. 应用可见/近红外光谱进行纺织纤维鉴别的研究[J]. 光谱学与光谱分析, 2010, 30(2): 331-335. DOI:10.3964/j.issn.1000-0593(2010)02-0331-05.

[21] PHILLIPS T, ABDULLA W. Developing a new ensemble approach with multi-class SVMs for Manuka honey quality classification[J]. Applied Soft Computing, 2021, 111: 107710. DOI:10.1016/j.asoc.2021.107710.

[22] FOLLI G S, SANTOS L P, SANTOS F D, et al. Food analysis by portable NIR spectrometer[J]. Food Chemistry Advances, 2022, 1: 100074. DOI:10.1016/j.focha.2022.100074.

[23] XIE J, WANG C. Using support vector machines with a novel hybrid feature selection method for diagnosis of erythemato-squamous diseases[J]. Expert Systems with Applications, 2011, 38(5): 5809-5815. DOI:10.1016/j.eswa.2010.10.050.

[24] 吉新媛. 基于极限学习机的集成特征选择算法研究与应用[D]. 西安: 陕西师范大学, 2019. DOI:10.27292/d.cnki.gsxfu.2019.001645.

[25] 阮真, 朱鹏飞, 张磊, 等. 基于单细胞拉曼技术鉴定非结核分枝杆菌的方法研究[J]. 光谱学与光谱分析, 2021, 41(11): 3468-3473. DOI:10.3964/j.issn.1000-0593(2021)11-3468-06.

[26] SHAO Y, SHI Y, XUAN G, et al. Hyperspectral imaging for non-destructive detection of honey adulteration[J]. Vibrational Spectroscopy, 2022, 118: 103340. DOI:10.1016/j.vibspec.2022.103340.

[27] 谭航彬, 姜丽, 金尚忠, 等. 基于拉曼光谱的鸡蛋新鲜度检测及分类方法[J]. 中国计量大学学报, 2022, 33(2): 181-188; 203.

[28] 张宝萍, 宁甜, 张富荣, 等. 基于多变量光谱数据分析方法的乳腺癌血清拉曼光谱特征研究[J]. 光谱学与光谱分析, 2023, 43(2): 426-434.

[29] 周启超, 刘剑, 刘丽, 等. 基于SVM的通风系统故障诊断惩罚系数与核函数系数优化研究[J]. 中国安全生产科学技术, 2019, 15(4): 45-51. DOI:10.11731/j.issn.1673-193x.2019.04.007.

[30] 郭志明. 基于近红外光谱及成像的苹果品质无损检测方法和装置研究[D]. 北京: 中国农业大学, 2015.

[31] 吴迪, 曹芳, 冯水娟, 等. 基于支持向量机算法的红外光谱技术在奶粉蛋白质含量快速检测中的应用[J]. 光谱学与光谱分析, 2008(5): 1071-1075.