

近红外与表面增强拉曼光谱融合技术快速检测花生油中黄曲霉毒素 B₁

吴升德^{1#}, 朱家骥^{2#}, 钱昊², 姜鑫², 许艺³, 焦天慧^{3*}

(1. 盐城市产品质量监督检验所, 盐城 224056; 2. 盐城工学院电气工程学院, 盐城 224051;
3. 集美大学海洋食品与生物工程学院, 厦门 361021)

摘要: **目的** 在近红外光谱(near infrared spectroscopy, NIR)与表面增强拉曼光谱(surface-enhanced Raman spectroscopy, SERS)特征层数据融合的基础上构建偏最小二乘回归(partial least squares regression, PLSR)模型实现花生油中黄曲霉毒素 B₁ (aflatoxin B₁, AFB₁)含量的快速检测。**方法** 首先, 分别采集待测样本的 NIR 与 SERS 光谱。其次, 将采集的 NIR 与 SERS 光谱分别进行光谱预处理。然后, 采用基于希尔伯特-施密特独立准则的变量空间迭代优化算法(Hilbert-Schmidt independence criterion based variable space iterative optimization, HSIC-VSIO)分别筛选 NIR 与 SERS 光谱的特征变量。最后, 将筛选的特征变量进行融合并构建 PLSR 模型用于定量检测花生油中 AFB₁ 含量。**结果** 与 NIR 光谱数据、SERS 光谱数据以及 NIR 与 SERS 光谱直接融合数据构建的 PLSR 模型相比, NIR 与 SERS 光谱特征层融合数据构建的 PLSR 模型具有最佳的预测性能: 校正集均方根误差(root mean squared error of calibration set, RMSEC)为 0.1569, 校正集决定系数(coefficient of determination of calibration set, R^2_c)为 0.9908, 预测集均方根误差(root mean squared error of prediction set, RMSEP)为 0.1827, 预测集决定系数(coefficient of determination of prediction set, R^2_p)为 0.9854, 性能偏差比(ratio of performance to deviation, RPD)为 8.2761。将本方法与标准方法分别检测真实含有 AFB₁ 的花生油样本, 结果表明两者的检测性能无显著性差异($P=0.84>0.05$)。**结论** 本方法可实现花生油中 AFB₁ 含量的快速、高精度定量检测, 也验证了 NIR 与 SERS 光谱融合的可行性与有效性。

关键词: 近红外光谱; 表面增强拉曼光谱; 光谱数据融合; 黄曲霉毒素 B₁

Rapid determination of aflatoxin B₁ in peanut oil by the fusion of near infrared spectroscopy and surface-enhanced Raman spectroscopy

WU Sheng-De^{1#}, ZHU Jia-Ji^{2#}, QIAN Hao², JIANG Xin², XU Yi³, JIAO Tian-Hui^{3*}

(1. Yancheng Products Quality Supervision and Inspection Institute, Yancheng 224056, China; 2. School of Electrical Engineering, Yancheng Institute of Technology, Yancheng 224051, China; 3. College of Ocean Food and Biological Engineering, Jimei University, Xiamen 361021, China)

基金项目: 国家自然科学基金项目(32302211)、江苏省市场监督管理局科技计划项目(KJ2022050)、江苏省高等学校基础科学(自然科学)面上项目(21KJD550002)、福建省自然科学基金青年项目(2022J01233199)

Fund: Supported by the National Natural Science Foundation of China (32302211), the Science and Technology Project of Jiangsu Market Supervision Administration of China (KJ2022050), the Natural Science Research of Jiangsu Higher Education Institutions of China (21KJD550002), and the Natural Science Foundation Youth Program of Fujian Province of China (2022J01233199)

#吴升德、朱家骥为共同第一作者

#WU Sheng-De and ZHU Jia-Ji are Co-first Authors

*通信作者: 焦天慧, 博士, 讲师, 主要研究方向为食品快速无损检测技术。E-mail: jth@jmu.edu.cn

*Corresponding author: JIAO Tian-Hui, Ph.D, Lecturer, College of Ocean Food and Biological Engineering, Jimei University, No.43, Yindou Road, Jimei District, Xiamen 361021, China. E-mail: jth@jmu.edu.cn

ABSTRACT: Objective To construct a partial least squares regression (PLSR) model for rapid detection of aflatoxin B₁ (AFB₁) content in peanut oil based on the fusion of near infrared spectroscopy (NIR) and surface-enhanced Raman spectroscopy (SERS) feature layer data. **Methods** First, the NIR and SERS spectra of samples were respectively collected. Second, the collected NIR and SERS spectra were processed by spectral preprocessing techniques. Then, the characteristic variable of NIR and SERS spectra were respectively selected by employing Hilbert-Schmidt independence criterion based variable space iterative optimization (HSIC-VSIO) algorithm. Finally, the selected characteristic variable were fused, and then a PLSR model was built based on the fused characteristic variable for rapid determination of AFB₁ in peanut oil. **Results** Compared with the PLSR models built by NIR spectra, SERS spectra, and the data directly fused by NIR and SERS spectra, the PLSR model built based on the fused characteristic variable which were respectively selected from NIR and SERS spectra possessed superior prediction performance: The value of root mean squared error of calibration set (RMSEC) was 0.1569, the value of coefficient of determination of calibration set (R_c^2) was 0.9908, the value of root mean squared error of prediction set (RMSEP) was 0.1827, the value of coefficient of determination of prediction set (R_p^2) was 0.9854, and the value of ratio of performance to deviation (RPD) was 8.2761. The detection performance of this method and the standard method for real peanut oil samples containing AFB₁ was not significantly different, and the results showed that there was no significant difference between 2 kinds of methods ($P=0.84>0.05$). **Conclusion** The proposed method is a promising alternative for rapid and high-precision determination of the content of AFB₁ in peanut oil. The feasibility and effectiveness of the fusion of NIR and SERS spectra are also verified.

KEY WORDS: near infrared spectroscopy; surface-enhanced Raman spectroscopy; spectral data fusion; aflatoxin B₁

0 引言

黄曲霉毒素 B₁ (aflatoxin B₁, AFB₁) 是一种典型的真菌毒素, 它是二氢呋喃氧杂萜邻酮的衍生物^[1]。AFB₁ 是目前已知的化学物质中致癌性最强的一种, 主要对肝脏功能造成严重损伤, 故 AFB₁ 是国家市场监督管理总局指定的食品安全必检指标之一^[2]。油料作物(如花生、玉米等)由于其含水率高, 在储存与加工过程中容易发生霉变, 从而受到 AFB₁ 的污染。因此, 相关部门需要加大对粮油食品中 AFB₁ 的检测力度, 防止食品安全事件的发生。

目前, 食品中真菌毒素的常规检测方法包括: (1) 仪器理化分析法, 如高效液相色谱法(high performance liquid chromatography, HPLC)和液相色谱-质谱法(liquid chromatography-mass spectrometry, LC-MS)等^[3-5]; (2) 免疫化学检测法, 如酶联免疫吸附分析(enzyme-linked immunosorbent assay, ELISA)和化学发光免疫分析(chemiluminescence immunoassay, CLIA)等^[6-8]。由于上述方法均需要复杂的样本前处理, 使其难以满足食品加工、流通过程中现场快速检测的需求。因此, 开发一种快速、高精度的 AFB₁ 检测方法, 对于保障食品安全及维护消费者权益有着重要的意义。

近年来, 以近红外光谱(near infrared spectroscopy, NIR)和拉曼光谱(Raman spectroscopy)为代表的光谱快速、无损检测技术异军突起。由于 NIR 和拉曼光谱响应速度快、一般不需要预处理且易实现现场快速检测等优势, 已在食

品质量与安全检测领域得到了广泛的应用。例如, FERNÁNDEZ-IBÁÑEZ 等^[9]采用傅里叶变换近红外光谱仪分别预测玉米和小麦中的黄曲霉毒素, 预测相关性在 0.80~0.85 之间, 检出限可达 20 μg/kg。DURMS 等^[10]利用 NIR 结合模式识别方法, 定性鉴别了黄曲霉毒素污染的无花果, 识别率可达 100%。虽然 NIR 技术在食品中真菌毒素的快速、无损检测方面具有良好的应用潜力, 但由于食品中真菌毒素含量低、分子量小, 导致其检测精度难以达到国际和国家规定的检出限。表面增强拉曼光谱(surface-enhanced Raman spectroscopy, SERS)是拉曼光谱的延伸, 既拥有强大的分子识别能力, 又具有灵敏度高的特点, 是检测食品中痕量有害物质的理想手段^[11]。例如, KO 等^[12]制备了一种二氧化硅包覆的中空金纳米颗粒作为 SERS 基底, 再利用磁珠分离和免疫手段构建了纳米探针, 实现了食品中 AFB₁ 的快速检测, 检出限达到了 0.1 ng/mL, 检测时间小于 30 min。FANG 等^[13]通过合成一种金包覆镍磁性纳米颗粒作为 SERS 基底, 借助竞争性免疫手段构建了用于检测玉米中 AFB₁ 的探针, 检出限达到了 0.05 fg/mL, 加标回收率为 87.4%~117.7%。芦然等^[14]合成了 SERS 适配体传感器用于检测植物蛋白肉中的 AFB₁, 该传感器由巯基适配体修饰的金包四氧化三铁磁性纳米粒子作为捕获底物与巯基适配体互补链修饰的金包二氧化硅纳米粒子作为信号探针组成。通过适配体与 AFB₁ 的特异性结合可以导致捕获底物释放信号探针, 从而实现对 AFB₁ 的检测, 检出限达到了 0.37 pg/mL。将该方法用于实际样品中 AFB₁ 的分析,

加标回收率为 98.0%~110.0%，相对标准偏差为 3.9%~5.7%。

目前，在食品真菌毒素的光谱快速、无损检测应用中仍采用 NIR 或 SERS 单一技术手段。从理论角度来看，NIR 反映的是电偶极矩变化引起的振动，SERS 反映的是分子极化引起的振动，两种光谱信息在分子信息表达上具有互补性^[15]。因此，有必要将两种光谱信息进行融合，实现信息互补，以提高检测精度。

鉴于此，本研究以花生油中 AFB₁ 为检测指标，分别采集其 NIR 和 SERS 光谱，然后采用基于希尔伯特-施密特独立准则的变量空间迭代优化算法(Hilbert-Schmidt independence criterion based variable space iterative optimization, HSIC-VSIO)分别筛选 NIR 与 SERS 光谱的特征变量，将筛选的特征变量进行融合并在此基础上构建多元校正模型，实现花生油中 AFB₁ 的快速、高精度定量检测，为食品中真菌毒素的检测提供新的研究思路，对保障食品安全、提高政府监管效率、维护公共安全具有重要意义。

1 材料与方法

1.1 材料与试剂

鲁花牌花生油购于当地永辉超市。

甲醇(分析纯，国药集团化学试剂有限公司)；AFB₁ 标准品(纯度 99%，天津阿尔塔科技有限公司)；SERS 增强基底芯片 Q-SERS(江阴韵翔光电技术有限公司)；实验用水均为超纯水。

1.2 仪器与设备

FA1004E 分析天平(精度 0.0001 g，常州幸运电子设备有限公司)；RMS1000 手持式拉曼光谱仪、TS9214 近红外光谱仪(上海如海光电科技有限公司)；HY-2 涡旋混合器(上海仪电科学仪器股份有限公司)；PS-20A 超声波振荡器(深圳得康科技有限公司)。

1.3 实验方法

1.3.1 样本制备

将 AFB₁ 标准品溶于甲醇中制备母液，经稀释后获得质量浓度梯度为 1.0×10^{-5} 、 1.0×10^{-4} 、 1.0×10^{-3} 、 1.0×10^{-2} 和 1.0×10^{-1} $\mu\text{g/mL}$ 的 AFB₁ 标准溶液，所得溶液密封保存于 4℃ 下以备用。称取 5 g 花生油于 50 mL 离心管中，然后加入 1.0×10^{-5} $\mu\text{g/mL}$ 的 AFB₁ 标准溶液，随后用甲醇定容至 25 mL 并涡旋混匀，置于超声振荡器下 20 min，取混合溶液作为待测样本^[16]。其他浓度的待测样本均按此方法制备。

1.3.2 光谱数据采集

SERS 光谱的采集：将待测样本滴入 Q-SERS 检测区域，采用 RMS1000 手持式拉曼光谱仪采集其 SERS 光谱。手持式拉曼光谱的激光功率设置为 50 mW，积分时间设置为 2 s，扫描次数设置为 3 次。NIR 光谱的采集：将待测样本倒入石英比色皿中，采用 TS9214 近红外光谱仪采集待

测样本的 NIR 光谱。近红外光谱仪积分时间设置为 3 s，扫描次数设置为 5 次。对于每个浓度的待测样本均采集 10 条 SERS 光谱和 NIR 光谱，总共获得了 50 条 SERS 光谱和 50 条 NIR 光谱。

将已获得的光谱数据集划分为校正集与预测集：(1)无论是 NIR 还是 SERS 光谱数据，对于每种浓度的待测样本，从采集的 10 条光谱中随机挑选 6 条光谱划入校正集，故 NIR 或 SERS 校正集中光谱的数量均为 36；(2)无论是 NIR 还是 SERS 光谱数据，对于每种浓度的待测样本，将剩余的 4 条光谱划入预测集，故 NIR 或 SERS 预测集中光谱的数量均为 24。

1.4 光谱数据预处理

无论是 NIR 光谱还是 SERS 光谱的采集都将一定程度地带入干扰信息，如仪器噪声、基线漂移和光散射等。因此，对光谱数据进行预处理对于后续的定位或定量分析准确性的提高是有帮助的。另一个需要注意的细节是，由于 NIR 光谱数据与 SERS 光谱数据在光谱强度上差异较大，使得二者并不处在同一个量级。因此，为了实现光谱信息的融合，需对二者的光谱数据进行归一化操作，即将二者的光谱强度等比缩放到一个量级^[17]。本研究中，光谱数据预处理策略主要包含以下 4 个步骤^[18-21]：(1)采用自适应迭代重加权惩罚最小二乘(adaptive iteratively reweighted penalized least squares, AIRPLS)算法进行基线校正；(2)采用多元散射校正(multiplicative scatter correction, MSC)算法进行光散射校正；(3)采用卷积平滑(Savitzky-Golay, S-G)算法进行光谱信号的平滑处理；(4)采用 Min-Max 归一化算法将二者的光谱强度缩放到 0 与 1 之间。

1.5 算法原理

1.5.1 偏最小二乘回归

偏最小二乘回归(partial least squares regression, PLSR)是一种化学计量学线性回归建模方法，其能够处理变量数远大于样本数的回归问题^[22]。假设光谱矩阵为 X ，响应矩阵为 Y ，PLSR 寻求最大化 X 与 Y 之间的协方差，在此基础上提取隐变量，提取的隐变量不仅解释了 X 的可变性，而且也最大程度地与 Y 相关。PLSR 对光谱矩阵 X 与响应矩阵 Y 的分解模型如公式(1)、(2)：

$$X = TP^T + E_X \quad (1)$$

$$Y = UQ^T + E_Y \quad (2)$$

其中， T 和 U 分别为 X 和 Y 的得分矩阵； P 和 Q 分别为 X 和 Y 的载荷矩阵； E_X 和 E_Y 分别为 X 和 Y 的 PLSR 拟合残差矩阵。

T 与 U 之间存在线性关系，即 $U=bT$ ， b 为常数。将矩阵 X 用得分矩阵 T 来表示，即将 X 投影到子空间 T ，可得公式(3)：

$$T = XW \quad (3)$$

其中， W 表示投影方向，因此， X 与 Y 在原始空间的线性关

系可以表示为公式(4):

$$Y = bXR + E_Y \quad (4)$$

其中, $R=WQ^T$ 为回归系数矩阵; 矩阵 X^TYY^TX 最大特征值对应的特征向量可用于计算 W 。

1.5.2 光谱数据直接融合策略

光谱数据直接融合是一种最直接和最简单的融合策略, 将来自于不同光谱源的光谱数据按一定的顺序串接成一个矩阵, 然后在此基础上构建定性或定量模型(本研究中采用 PLSR 构建定量模型), 以达到提高模型预测性能的目的^[23]。需要注意的是, 原始光谱数据先进行预处理后再融合。光谱数据直接融合策略示意图如图 1 所示。

1.5.3 光谱数据特征层融合策略

光谱数据特征层融合是基于原始光谱数据的特征信息层面进行的融合。将来自于不同光谱源的光谱数据经光谱预处理后筛选特征变量, 并将筛选的特征变量按一定的顺序串接成一个矩阵, 然后在此基础上构建定性或定量模型(本研究中采用 PLSR 构建定量模型)^[24]。由于原始光谱数据中含有大量的冗余变量, 故特征层融合的性能要优于直接融合。光谱数据特征层融合策略示意图如图 2 所示。

1.5.4 基于 HSIC 的变量空间迭代优化算法

本研究中, 光谱特征变量的筛选由 HSIC-VSIO 实现, 该算法主要包括以下内容:

(1) 加权二值矩阵采样策略

加权二值矩阵采样(weighted binary matrix sampling, WBMS)策略是由二值矩阵采样(binary matrix sampling, BMS)策略演变而来。在 WBMS 中, 根据变量的权重, 为每一列分配不同数量的“0”和“1”, 即权重越大的变量被选中的概率就越大。

(2) 希尔伯特-施密特独立准则

HSIC 是在再生核希尔伯特空间(reproducing kernel Hilbert space, RKHS)基础上提出来的。对于随机变量 X 和 Y , 设 F 是一个具有非线性特征映射的 RKHS($\Phi: X \rightarrow F$), 对于 $x \in X$, 则 $\Phi(x) \in F$ 。类似的, 设 G 也是一个具有非线性特征映射的 RKHS($\Psi: Y \rightarrow G$), 对于 $y \in Y$, 则 $\Psi(y) \in G$ 。 F 与 G 之间的交叉协方差算子可定义为公式(5):

$$C_{xy} = E_{x,y}[(\Phi(x) - \mu_x) \otimes (\Psi(y) - \mu_y)] \quad (5)$$

其中, $\mu_x = E_x[\Phi(x)]$; $\mu_y = E_y[\Psi(y)]$; $E_{x,y}$ 、 E_x 和 E_y 分别是对联合概率密度函数 P_{xy} 、边缘概率密度函数 P_x 和边缘概率密度函数 P_y 的期望; $\otimes$ 代表张量积的运算符。

HSIC 的理论值可表示为交叉协方差算子的希尔伯特-施密特范数的平方, 如公式(6):

$$\begin{aligned} \text{HSIC}(X, Y) = \|C_{xy}^2\| = & E_{x,x',y,y'}[k(x,x')l(y,y')] + \\ & E_{x,x'}[k(x,x')]E_{y,y'}[l(y,y')] - 2E_{(x,y)}[E_x[k(x,x')] \\ & E_y[l(y,y')]] \end{aligned} \quad (6)$$

其中, $k(x,x') = \langle \Phi(x), \Phi(x') \rangle$; k 是一个核函数($k: X \times X \rightarrow R$); $l(y,y') = \langle \Psi(y), \Psi(y') \rangle$; l 也是一个核函数($l: Y \times Y \rightarrow R$); $E_{x,x',y,y'}$ 表示数据对 (x,y) 和 (x',y') 的联合期望。

给定 N 个独立观测值, $Z = \{(x_1, y_1) \dots (x_N, y_N)\} \subset (X \times Y)$, HSIC 的经验值为公式(7)^[25]:

$$\text{HSIC}(Z, F, G) = \frac{1}{(N-1)^2} \text{tr}(KHLH) \quad (7)$$

其中, $K, H, L \in R^{N \times N}$; $K_{ij} = k(x_i, x_j)$; $L_{ij} = l(y_i, y_j)$; $H_{ij} = I - N^{-1}ee^T$ (e 是全 1 向量); tr 表示矩阵 $KHLH$ 的迹。

如果两个随机变量是相互独立的, 他们的 HSIC 值为零, 故 HSIC 可用于独立性测试。



图 1 光谱数据直接融合策略示意图

Fig.1 Schematic diagram of spectral data direct fusion strategy

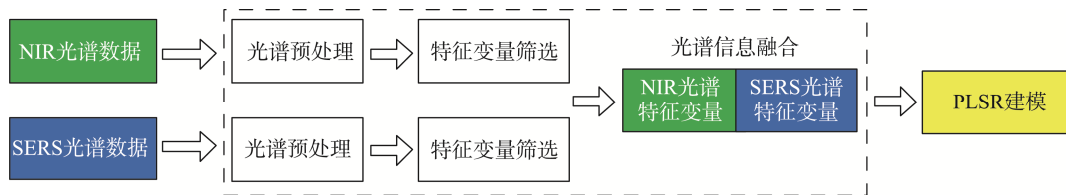


图 2 光谱数据特征层融合策略示意图

Fig.2 Schematic diagram of spectral data characteristic layer fusion strategy

(3) 基于 HSIC 评估波长变量的重要性

假设有光谱数据矩阵 X 和响应向量 y , 首先构建全光谱多元校正模型, 公式(8):

$$y = f_{\text{full}}(x_1, \dots, x_p) + \varepsilon_{\text{full}} \quad (8)$$

其中, f_{full} 表示由 PLSR 构建的全光谱多元校正模型; $\varepsilon_{\text{full}}$ 表示预测残差。

将第 m 个波长剔除, 构建多元校正模型, 如公式(9):

$$y = \hat{f}_m(x_1, \dots, x_{m-1}, x_{m+1}, \dots, x_p) + \varepsilon_m \quad (9)$$

其中, $\hat{f}_m$ 表示由 PLSR 构建的多元校正模型, ε_m 表示预测残差, 其包含了噪声及被剔除波长变量 x_m 的信息。

因此, 可以通过计算预测残差 ε 与输出 y 之间的统计独立性差别来获得波长重要性的合理度量。波长变量 x_m 的重要性(WI_m)可由式(10)进行估算^[26]。

$$WI_m = \frac{\text{HSIC}(y, \varepsilon_m)}{\text{HSIC}(y, \varepsilon_{\text{full}})} \quad m = 1, 2, \dots, p \quad (10)$$

(4) HSIC-VSIO 具体实施步骤

步骤 1: 估算每个波长变量的重要性, 并存储在向量 $T = [WI_1, WI_2, \dots, WI_p]$, 然后对每个波长变量的重要性进行归一化处理, 如公式(11):

$$WI'_i = \frac{WI_i}{\sum_{i=1}^p WI_i} \quad i = 1, 2, \dots, p \quad (11)$$

其中, WI'_i 表示第 i 个变量的重要性归一化后的值, 并存储在向量 $T' = [WI'_1, WI'_2, \dots, WI'_p]$ ($0 \leq WI'_i < 1$)。

步骤 2: 基于 WBMS(其中, 二值矩阵的尺寸为 $M \times p$, 每个波长变量的权重被初始化为 0.5), 则可得 M 个由 PLSR 构建的多元校正模型, 并计算得到每个模型的 5 折交叉验证均方根误差(root mean squared error of cross-validation, RMSECV), 从所有模型中挑选出具有较小 RMSECV 值的模型(占有模型的比例为 σ), 并将最小的 RMSECV 值记为 $\min\text{RMSECV}$ 。

步骤 3: 计算每个波长变量在挑选出的具有较小 RMSECV 值的模型中出现的频率, 如公式(12):

$$f_i = \frac{N_i}{M \times \sigma} \quad i = 1, 2, \dots, p \quad (12)$$

其中, N_i 表示第 i 个波长变量在挑选出的具有较小 RMSECV 值的模型(数量为 $M \times \sigma$)中出现的次数, 则 f_i 表示第 i 个波长变量在挑选出的具有较小 RMSECV 值的模型(数量为 $M \times \sigma$)中出现的频率, 并存储在向量 $Q = [f_1, f_2, \dots, f_p]$ 。更新每个波长变量的权重 $W = 0.5 \times T' + 0.5 \times Q$ 。

步骤 4: 基于更新后的波长变量的权重 W 及 WBMS 策略, 由 PLSR 构建 M 个新的多元校正模型, 与步骤 2 中类似, 从所有模型中挑选出具有较小 RMSECV 值的模型, 并记录下最小的 RMSECV 值。

步骤 5: 如果 $\min\text{RMSECV}_I < \min\text{RMSECV}_{I-1}$ (I 为迭代

次数), 重复执行步骤 3~5, 否则, 迭代结束并输出最优的波长变量。

根据参考文献[27~28]中的结论, HSIC-VSIO 主要参数设置为: 1) WBMS 中二值矩阵的行的数量 $M=1000$; 2) 从所有模型中挑选出具有较小 RMSECV 值的模型的比例 $\sigma=10\%$ 。在本研究 2.2 节中对参数设置的合理性进行了验证。

1.6 模型评价指标

为了评估各模型的性能, 一般采用校正集均方根误差(root mean squared error of calibration set, RMSEC)、预测集均方根误差(root mean squared error of prediction set, RMSEP)、校正集决定系数(coefficient of determination of calibration set, R_c^2)、预测集决定系数(coefficient of determination of prediction set, R_p^2)和性能偏差比(ratio of performance to deviation, RPD)等模型评价指标, 如公式(13)、(14)、(15):

$$\text{RMSE} = \sqrt{\frac{1}{L} \sum_{i=1}^L (y_i - \hat{y}_i)^2} \quad (13)$$

$$R^2 = 1 - \frac{\sum_{i=1}^L (y_i - \hat{y}_i)^2}{\sum_{i=1}^L (y_i - \bar{y})^2} \quad (14)$$

$$\text{RPD} = \frac{1}{\sqrt{1 - R_p^2}} \quad (15)$$

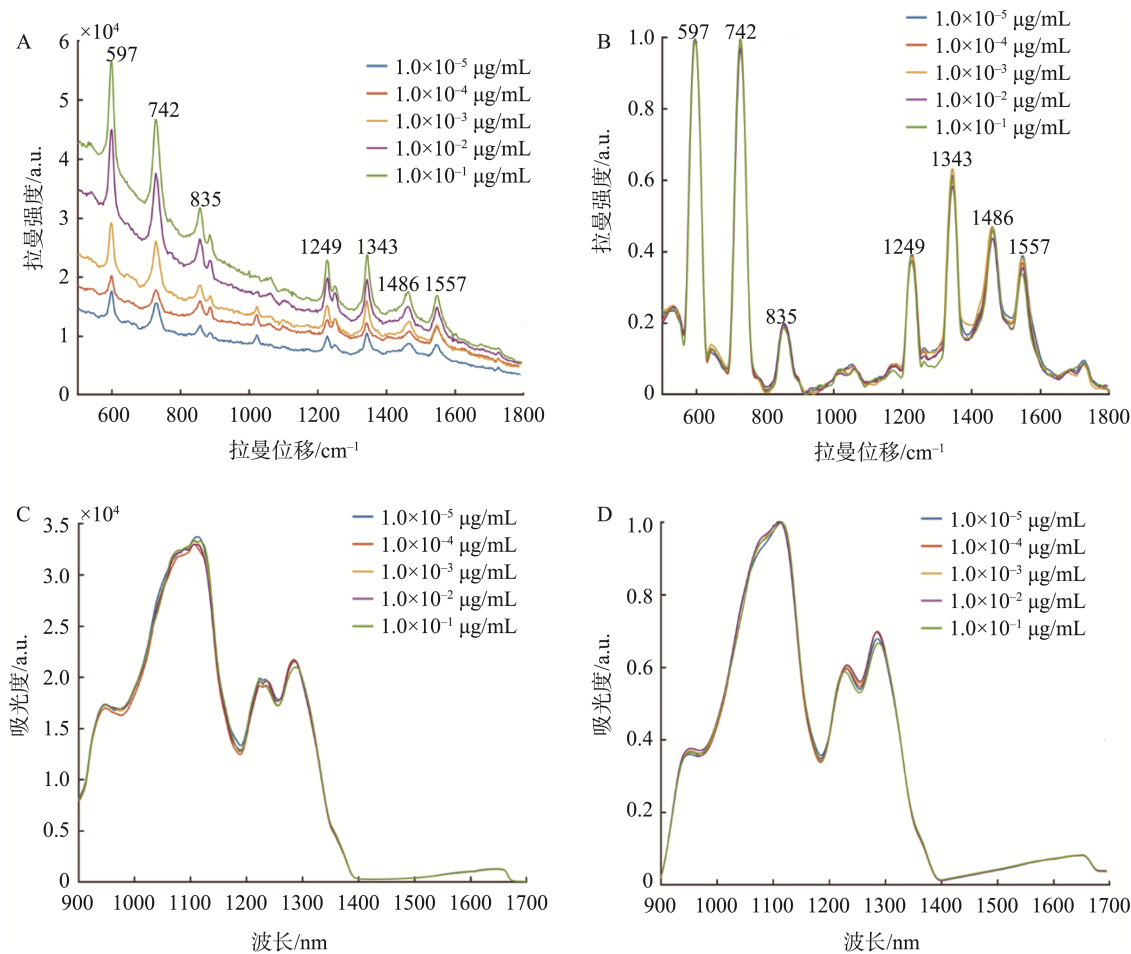
其中, y_i 表示真实值; $\hat{y}_i$ 表示预测值; $\bar{y}$ 表示真实值的平均值。

当 L 表示校正集中样本的数量时, 则式(13)对应为 RMSEC, 式(14)对应为 R_c^2 ; 当 L 表示预测集中样本的数量时, 则式(13)对应为 RMSEP, 式(14)对应为 R_p^2 。本研究中的算法及运算均由 Matlab R2020a 实现。

2 结果与分析

2.1 光谱数据分析结果

含有不同浓度 AFB₁ 的 5 条代表性的花生油待测样本的 SERS 光谱如图 3A 所示。图 3A 中主要的 SERS 特征谱带及其归属为^[29]: 597 cm⁻¹ (C-O 伸缩振动)、742 cm⁻¹ (C-H 面外弯曲振动)、835 cm⁻¹ (C-H 伸缩振动)、1249 cm⁻¹ (C-H 面内弯曲振动)、1343 cm⁻¹ (CH₃ 变形振动)、1486 cm⁻¹ (C=C 伸缩振动)和 1557 cm⁻¹ (C-C 伸缩振动)。这里需要说明的是, 由于 SERS 光谱区域(500~1800 cm⁻¹)信噪比高且包含了主要的特征谱带, 故本研究中将此区域用于 AFB₁ 的定量分析。含有不同质量浓度 AFB₁ 的 5 条代表性的花生油待测样本的 NIR 光谱如图 3C 所示。图 3C 中 NIR 特征谱带及其归属为^[30]: 930~970 nm (CH₂ 与 CH₃ 一阶倍频伸缩振动)、1090~1130 nm (C-H 伸缩振动)、1210~1240 nm (CH₂ 二阶倍频伸缩振动)和 1270~1300 nm (C=O 二阶倍频伸缩振动、C=O 合频振动及 N-H 伸缩振动)。AFB₁ 与 NIR 特征谱带有密切关系, 这是由于花生油中的蛋白质、碳水化合物以



注: A. 5 条代表性的原始 SERS 光谱; B. 预处理后的 5 条代表性的 SERS 光谱; C. 5 条代表性的原始 NIR 光谱; D. 预处理后的 5 条代表性的 NIR 光谱。

图 3 含有不同质量浓度 AFB₁ 的花生油待测样本的 SERS 与 NIR 光谱
Fig.3 SERS and NIR spectra of peanut oil samples containing different mass concentrations of AFB₁

及脂肪酸易受到 AFB₁ 的影响, 从而影响分子的振动。无论是 NIR 还是 SERS 光谱, 在光谱采集过程中带入干扰信息往往是无法避免的, 故需要对光谱数据进行预处理。经 AIRPLS 基线校正、MSC 光散射校正、S-G 平滑以及 Min-Max 归一化处理之后的 SERS 与 NIR 光谱分别如图 3B 与 3D 所示, 与原始光谱(图 3A 与 3C)对比发现, 预处理后的 SERS 和 NIR 光谱的基线漂移得到了抑制, 光谱信号更加平滑, 为后续的定量分析起到了积极的作用。

2.2 HSIC-VSIO 算法参数设置合理性验证

对 HSIC-VSIO 算法参数设置合理性进行验证: 在设置不同的参数情况下, 分别对 NIR 和 SERS 光谱数据筛选特征变量, 并将每次筛选的特征变量进行融合建立 PLSR 模型, 记录 RMSEC、 R_c^2 、RMSEP、 R_p^2 和 RPD 值进行对比分析。

(1) WBMS 中二值矩阵的行的数量 M

首先, 将 σ 的值分别设置为 10%; 然后, 将 M 的值分

别设置为 1000、1500、2000 和 2500 进行对比分析。由表 1 中的运行结果可知, 模型的性能受 M 的影响并不大。但是, 如果 M 的值越大, 模型的计算量将显著增大, 综合考虑模型精度与计算量, 将 M 设置为 1000 是合理的。

表 1 参数 M 对模型预测性能的影响
Table 1 Effects of parameter M on model's prediction performances

参数 M	校正集		预测集		
	RMSEC	R_c^2	RMSEP	R_p^2	RPD
1000	0.1569	0.9908	0.1827	0.9854	8.2761
1500	0.1571	0.9901	0.1831	0.9849	8.1380
2000	0.1570	0.9905	0.1829	0.9851	8.1920
2500	0.1572	0.9899	0.1832	0.9845	8.0321

(2) 从所有模型中挑选出具有较小 RMSECV 值的模型的比例 σ

首先, 将 M 的值设置为 1000; 然后, 将 σ 的值分别设

置为 10%、20%、30%和 40%进行对比分析。由表 2 中的运行结果可知,当 $\sigma=10\%$ 时,模型的性能最优。具体表现为, RMSEC 和 RMSEP 值较小, R_c^2 、 R_p^2 和 RPD 值较大,故将 σ 设置为 10%是合理的。

表 2 参数 σ 对模型预测性能的影响
Table 2 Effects of parameter σ on model's prediction performances

参数 σ	校正集		预测集		
	RMSEC	R_c^2	RMSEP	R_p^2	RPD
10%	0.1569	0.9908	0.1827	0.9854	8.2761
20%	0.1642	0.9886	0.1925	0.9830	7.6699
30%	0.1735	0.9869	0.2011	0.9747	6.2869
40%	0.1778	0.9861	0.2032	0.9738	6.1782

2.3 各方法检测结果

将 NIR 光谱数据、SERS 光谱数据、NIR 与 SERS 光谱直接融合数据以及 NIR 与 SERS 光谱特征层融合数据分别构建 PLSR 多元校正模型检测花生油中 AFB₁ 的含量。PLSR 建模过程中,最佳隐变量数(latent variables, LVs)由 5 折交互验证产生的 RMSECV 值所确定。各方法的检测结果如表 3 所示。

由表 3 可知,基于 NIR 光谱数据的定量检测结果如下: LVs=10, RMSEC=0.2812, $R_c^2=0.9533$, RMSEP=0.3447, $R_p^2=0.9211$, RPD=3.5601,花生油中 AFB₁ 含量 PLSR 预测值与真实值之间的关系如图 4A 所示。基于 SERS 光谱数据定量检测结果如下: LVs=8, RMSEC=0.2105, $R_c^2=0.9726$, RMSEP=0.2349, $R_p^2=0.9689$, RPD=5.6705,花生油中 AFB₁ 含量 PLSR 预测值与真实值之间的关系如图 4B 所示。基于 NIR 与 SERS 光谱直接融合数据定量检测结果如下: LVs=10, RMSEC=0.1923, $R_c^2=0.9836$, RMSEP=0.2117, $R_p^2=0.9703$, RPD=5.8026,花生油中 AFB₁ 含量 PLSR 预测值与真实值之间的关系如图 4C 所示。基于 NIR 与 SERS 光谱特征层融合数据定量检测结果如下: LVs=9, RMSEC=0.1569,

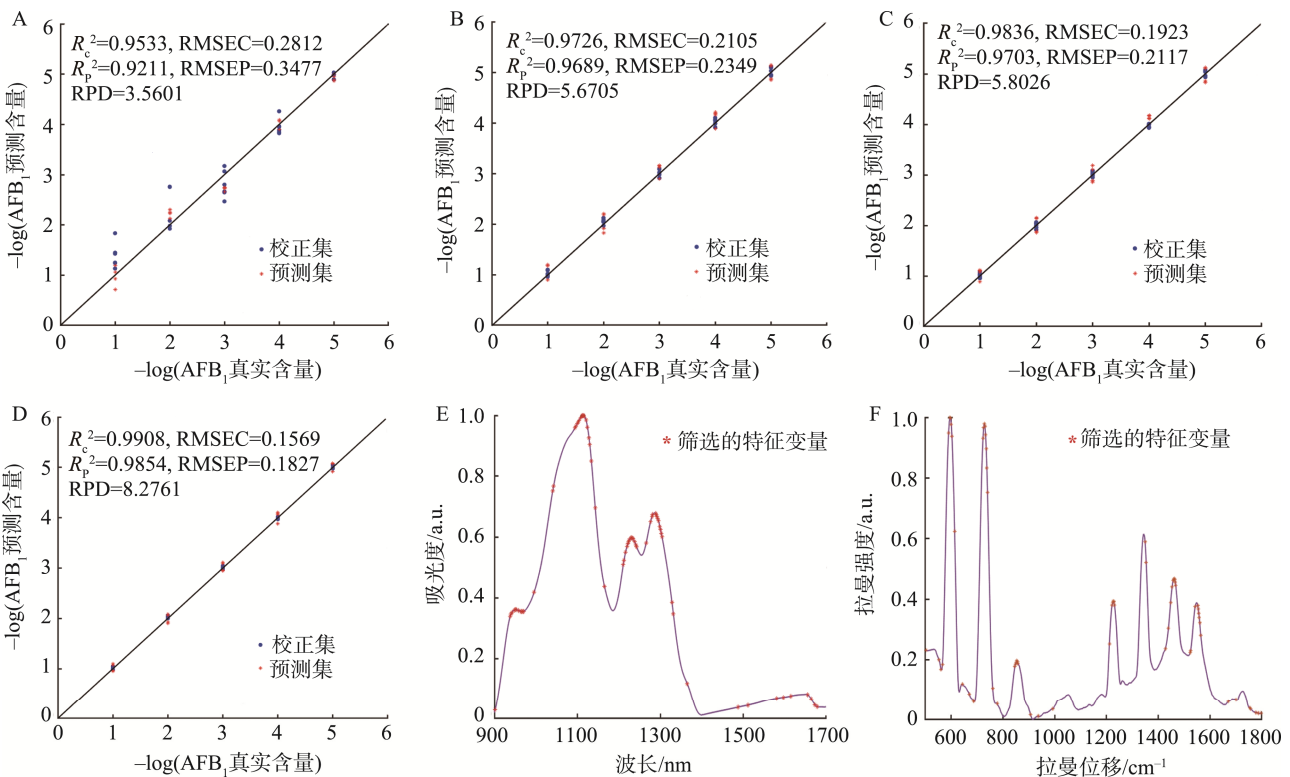
$R_c^2=0.9908$, RMSEP=0.1827, $R_p^2=0.9854$, RPD=8.2761,花生油中 AFB₁ 含量 PLSR 预测值与真实值之间的关系如图 4D 所示。由 HSIC-VSIO 筛选的 NIR 光谱特征变量如图 4E 所示,其中部分特征变量覆盖了 NIR 特征谱带 930~970、1090~1130、1210~1240 和 1270~1300 nm。由 HSIC-VSIO 筛选的 SERS 光谱特征变量如图 4F 所示,其中部分特征变量覆盖了 SERS 特征谱带 597、742、835、1249、1486 和 1557 cm⁻¹。

2.4 各方法检测结果对比分析

各方法所建 PLSR 模型评价指标的变化趋势如图 5 所示,显然,由 NIR 光谱数据构建的 PLSR 模型预测性能最差,主要在于花生油中 AFB₁ 含量低,分子量小,内部含氢基团振动在近红外区域吸收的能量低,对应的光谱信号弱,影响了其检测精度。相较于 NIR 光谱数据构建的 PLSR 模型,由 SERS 光谱数据、NIR 与 SERS 光谱直接融合数据以及 NIR 与 SERS 光谱特征层融合数据所构建的 PLSR 模型的预测性能均获得了提高。以 NIR 光谱数据构建的 PLSR 模型的预测性能作为基准, SERS 光谱数据、NIR 与 SERS 光谱直接融合数据以及 NIR 与 SERS 光谱特征层融合数据所构建的 PLSR 模型的 RMSEC 分别降低了 25.14%、31.61%和 44.20%; R_c^2 分别提高了 2.02%、3.18%和 3.93%; RMSEP 分别降低了 31.85%、38.58%和 47.01%; R_p^2 分别提高了 5.19%、5.34%和 6.98%; RPD 分别提高了 59.28%、62.99%和 132.47%。综上所述,由 SERS 光谱数据构建的 PLSR 模型的预测性能明显提高,主要在于 SERS 技术通过增强基底 Q-SERS 获得拉曼增强效应使得花生油中痕量 AFB₁ 的信号获得了放大,从而提高了其检测精度。相较于采用 NIR 或 SERS 光谱单一检测技术,将 NIR 光谱与 SERS 光谱直接融合后,实现了光谱信息的互补,有助于检测精度的进一步提高。然而,光谱直接融合数据中包含大量的冗余甚至干扰变量,采用 HSIC-VSIO 分别对 NIR 与 SERS 光谱筛选特征变量,然后将筛选得到的特征变量进行融合并构建 PLSR 模型,其检测精度获得了较大的提高。

表 3 各方法定量检测花生油中 AFB₁ 含量的结果
Table 3 Quantitative determination results of AFB₁ content in peanut oil by various methods

方法	LVs	校正集		预测集		
		RMSEC	R_c^2	RMSEP	R_p^2	RPD
NIR	10	0.2812	0.9533	0.3447	0.9211	3.5601
SERS	8	0.2105	0.9726	0.2349	0.9689	5.6705
NIR 与 SERS 直接融合	10	0.1923	0.9836	0.2117	0.9703	5.8026
NIR 与 SERS 特征层融合	9	0.1569	0.9908	0.1827	0.9854	8.2761



注: 花生油中 AFB₁ 含量 PLSR 预测值与真实值之间的关系: A. 基于 NIR 光谱数据; B. 基于 SERS 光谱数据; C. 基于 NIR 与 SERS 光谱直接融合数据; D. 基于 NIR 与 SERS 光谱特征层融合数据; E. HSIC-VSIO 筛选的 NIR 光谱特征变量; F. HSIC-VSIO 筛选的 SERS 光谱特征变量。

图 4 花生油中 AFB₁ 含量 PLSR 预测值与真实值之间的关系及 HSIC-VSIO 筛选的光谱特征变量

Fig.4 Relationship between PLSR predicted and true values of AFB₁ content in peanut oil and spectral characteristic variables screened by HSIC-VSIO

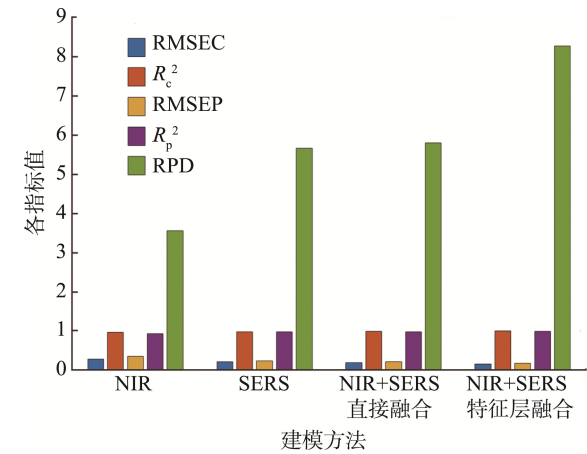


图 5 各方法所建 PLSR 模型评价指标变化趋势

Fig.5 Trends in evaluation indicators of PLSR models built by various methods

2.5 真实样本检测分析结果

从青岛普瑞邦生物工程有限公司购买一批含有 AFB₁ 的花生油样本(AFB₁ 含量范围为: $1.0 \times 10^{-5} \sim 1.0 \times 10^{-3} \mu\text{g/mL}$)。每个样本分别采用 NIR 与 SERS 光谱特征层融合数据构建的 PLSR 模型(以下简称光谱特征融合方法)以及标准方法(HPLC)

检测 AFB₁ 含量, 检测结果如表 4 所示。将两种方法的检测结果做双侧配对 t 检验, 结果表明两者无显著性差异 ($P=0.84>0.05$)。根据检出限的计算公式 $3S_0/K$ (S_0 为多个空白样本响应值标准差, K 为校正曲线的斜率), 可估算得到光谱特征融合方法对 AFB₁ 含量的检出限为 $5.27 \times 10^{-6} \mu\text{g/mL}$ 。欧盟与中国设置的花生油中 AFB₁ 最大残留限量分别为 $2.0 \mu\text{g/kg}$ 和 $20 \mu\text{g/kg}$ ^[31]。为了与上述标准进行对比, 可将溶液(花生油+AFB₁)密度设为 1 g/mL , 从而实现将 $5.27 \times 10^{-6} \mu\text{g/mL}$ 粗略地转换为 $5.27 \times 10^{-3} \mu\text{g/kg}$ 。故本研究提出的光谱特征融合方法可满足对花生油中 AFB₁ 含量是否超标的定量检测。

表 4 光谱特征融合方法与标准方法检测真实花生油样本中 AFB₁ 含量的结果($n=5$)

Table 4 Results of spectral feature fusion method and standard method for detecting AFB₁ content in real peanut oil samples ($n=5$)

样本	AFB ₁ 含量检测结果/($\mu\text{g/mL}$)	
	光谱特征融合方法	HPLC
样本 1	$0.94 \times 10^{-5} \pm 1.28 \times 10^{-7}$	$0.97 \times 10^{-5} \pm 0.82 \times 10^{-7}$
样本 2	$0.92 \times 10^{-5} \pm 1.57 \times 10^{-7}$	$0.98 \times 10^{-5} \pm 0.89 \times 10^{-7}$
样本 3	$0.95 \times 10^{-4} \pm 1.72 \times 10^{-6}$	$0.99 \times 10^{-4} \pm 0.85 \times 10^{-6}$
样本 4	$0.95 \times 10^{-3} \pm 1.61 \times 10^{-5}$	$0.97 \times 10^{-3} \pm 0.76 \times 10^{-5}$
样本 5	$0.93 \times 10^{-3} \pm 1.45 \times 10^{-5}$	$0.97 \times 10^{-3} \pm 0.80 \times 10^{-5}$

3 结 论

本研究提出了一种基于 NIR 与 SERS 光谱特征层融合数据构建 PLSR 模型实现花生油中 AFB₁ 快速、高精度检测的方法。与 NIR 光谱数据、SERS 光谱数据以及 NIR 与 SERS 光谱直接融合数据构建的 PLSR 模型相比, NIR 与 SERS 光谱特征层融合数据构建的 PLSR 模型具有最佳的预测性能: RMSEC=0.1569, R_c^2 =0.9908, RMSEP=0.1827, R_p^2 =0.9854, RPD=8.2761。同时, 将本研究方法与标准方法分别检测真实的花生油样本中 AFB₁ 含量, 结果表明两者的检测性能无显著性差异($P=0.84>0.05$), 本研究方法的检出限可换算为 5.27×10^{-3} $\mu\text{g/kg}$, 远远低于欧盟与中国设置的花生油中 AFB₁ 最大残留限量 2.0 $\mu\text{g/kg}$ 和 20 $\mu\text{g/kg}$ 。综上, 实验结果表明本研究方法可实现花生油中 AFB₁ 含量的快速、高精度定量检测, 验证了 NIR 与 SERS 光谱融合的可行性与有效性, 尤其是经特征变量筛选后, NIR 与 SERS 光谱数据在特征层的融合能够最大限度地提高模型的检测精度。

参考文献

- [1] WU ZH, SUN DW, PU HB, *et al.* Ti₃C₂Tx MXenes loaded with Au nanoparticle dimers as a surface-enhanced Raman scattering aptasensor for AFB₁ detection [J]. Food Chem, 2022, 372: 131293.
- [2] YAN MM, LI HD, LI M, *et al.* Advance in surface-enhanced Raman scattering-based aptasensors for food safety detection [J]. J Agric Food Chem, 2021, 69(47): 14049–14064.
- [3] KESKIN E, EYUPOGLU OE. Determination of mycotoxins by HPLC, LC-MS/MS and health risk assessment of the mycotoxins in bee products of Turkey [J]. Food Chem, 2023, 400: 134086.
- [4] MURAD M, AHMAD KS, IRAM S, *et al.* Mycotoxins in *Zea mays*, their quantification and HPLC analysis of physico-biological detoxification [J]. Nat Prod Res, 2022, 36(22): 5798–5802.
- [5] LU Q, RUAN HN, SUN XQ, *et al.* Contamination status and health risk assessment of 31 mycotoxins in six edible and medicinal plants using a novel green defatting and depigmenting pretreatment coupled with LC-MS/MS [J]. LWT-Food Sci Technol, 2022, 161: 113401.
- [6] SUO ZG, NIU XY, WEI M, *et al.* Latest strategies for rapid and point of care detection of mycotoxins in food: A review [J]. Anal Chim Acta, 2023, 1246: 340888.
- [7] XING KY, PENG J, SHAN S, *et al.* Green enzyme-linked immunosorbent assay based on the single-stranded binding protein-assisted aptamer for the detection of mycotoxins [J]. Anal Chem, 2020, 92(12): 8422–8426.
- [8] ZONG C, JIANG F, WANG XY, *et al.* Imaging sensor array coupled with dual-signal amplification strategy for ultrasensitive chemiluminescence immunoassay of multiple mycotoxins [J]. Biosens Bioelectron, 2021, 177: 112998.
- [9] FERNÁNDEZ-IBÁÑEZ V, SOLDADO A, MARTÍNEZ-FERNÁNDEZ A, *et al.* Application of near infrared spectroscopy for rapid detection of aflatoxin B₁ in maize and barley as analytical quality assessment [J]. Food Chem, 2009, 113(2): 629–634.
- [10] DURMUS E, GUNES A, KALKAN H, *et al.* Detection of aflatoxin and surface mould contaminated figs by using Fourier transform near-infrared reflectance spectroscopy [J]. J Sci Food Agric, 2017, 97(1): 317–323.
- [11] LANGER J, ABERASTURI DJD, AIZPURUA J, *et al.* Present and future of surface-enhanced Raman scattering [J]. ACS Nano, 2020, 14(1): 28–117.
- [12] KO J, LEE C, CHOO J. Highly sensitive SERS-based immunoassay of aflatoxin B₁ using silica-encapsulated hollow gold nanoparticles [J]. J Hazard Mater, 2015, 285: 11–17.
- [13] FANG CW, WEI C, XU MM, *et al.* Ni@Au nanoparticles for surface enhanced Raman spectroscopy based ultrasensitive magnetic immunoassay on aflatoxin B₁ [J]. RSC Adv, 2016, 6: 61325–61333.
- [14] 芦然, 陈瑞鹏, 任舒悦, 等. 基于表面增强拉曼光谱适配体传感器检测植物蛋白肉中的黄曲霉毒素 B₁ [J]. 食品安全质量检测学报, 2023, 14(14): 1–9.
- LU R, CHEN RP, REN SY, *et al.* Detection of aflatoxin B₁ in plant protein meat using surface enhanced Raman spectroscopy aptamer sensor [J]. J Food Saf Qual, 2023, 14(14): 1–9.
- [15] MÁRQUEZ C, LÓPEZ MI, RUISÁNCHEZ I, *et al.* FT-Raman and NIR spectroscopy data fusion strategy for multivariate qualitative analysis of food fraud [J]. Talanta, 2016, 161: 80–86.
- [16] 杨明秀. 基于表面增强拉曼光谱技术的花生油中 AFB₁ 检测方法研究[D]. 镇江: 江苏大学, 2018.
- YANG MX. Study on detection of AFB₁ in peanut oil based on surface-enhanced Raman spectroscopy technology [D]. Zhenjiang: Jiangsu University, 2018.
- [17] ZHAO Q, YU Y, HAO N, *et al.* Data fusion of laser-induced breakdown spectroscopy and near-infrared spectroscopy to quantitatively detect heavy metals in lily [J]. Microchem J, 2023, 190: 108670.
- [18] ZHANG ZM, CHEN S, LIANG YZ. Baseline correction using adaptive iteratively reweighted penalized least squares [J]. Analyst, 2010, 135: 1138–1146.
- [19] SILALAH DD, MIDI H, ARASAN J, *et al.* Robust generalized multiplicative scatter correction algorithm on pretreatment of near infrared spectral data [J]. Vib Spectr, 2018, 97: 55–65.
- [20] ZHANG J, MOUAZEN AM. Fractional-order Savitzky-Golay filter for pre-treatment of on-line vis-NIR spectra to predict phosphorus in soil [J]. Infrared Phys Technol, 2023, 131: 104720.
- [21] MISHRA P, BIANCOLILLO A, ROGER JM, *et al.* New data preprocessing trends based on ensemble of multiple preprocessing techniques [J]. TrAC-Trend Anal Chem, 2020, 132: 116045.
- [22] WOLD S, SJOSTROM M, ERIKSSON L. PLS-regression: A basic tool of chemometrics [J]. Chem Int Lab, 2001, 58(2): 109–130.

- [23] BORRÀS E, FERRÉ J, BOQUÉ R, *et al.* Data fusion methodologies for food and beverage authentication and quality assessment—A review [J]. *Anal Chim Acta*, 2015, 891: 1–14.
- [24] WANG Q, XIAO JT, LI Y, *et al.* Mid-level data fusion of Raman spectroscopy and laser-induced breakdown spectroscopy: Improving ores identification accuracy [J]. *Anal Chim Acta*, 2023, 1240: 340772.
- [25] WANG TH, DAI XL, LIU YZ. Learning with Hilbert-Schmidt independence criterion: A review and new perspectives [J]. *Knowl-Based Syst*, 2021, 234: 107567.
- [26] ZHANG XM, WADA T, FUJIWARA K, *et al.* Regression and independence based variable importance measure [J]. *Comput Chem Eng*, 2020, 135: 106757.
- [27] ZHU JJ, JIANG X, WANG QJ, *et al.* A novel hybrid variable selection strategy with application to molecular spectroscopic analysis [J]. *Chem Int Lab*, 2023, 236: 104795.
- [28] ZHU JJ, AHMAD W, JIAO TH, *et al.* Interval combination iterative optimization approach coupled with SIMPLS (ICIOA-SIMPLS) for quantitative analysis of surface-enhanced Raman scattering (SERS) spectra [J]. *Anal Chim Acta*, 2020, 1105: 45–55.
- [29] JIAO TH, AHMAD W, ZHU JJ, *et al.* Aggregation triggered aflatoxin B₁ determination in foodstuff employing 5-aminotetramethylrhodamine decorated gold-silver core-shell nanoparticles in surface enhanced Raman scattering [J]. *Sens Actuats B*, 2021, 331: 129424.
- [30] YAO WQ, LIU RS, XU ZC, *et al.* Rapid determination of aflatoxin B₁ contamination in peanut oil by Fourier transform near-infrared spectroscopy [J]. *J Spectr*, 2022, 1: 9223424.
- [31] 尚艳娥, 杨卫民. CAC、欧盟、美国与中国粮食中真菌毒素限量标准的差异分析[J]. *食品科学技术学报*, 2019, 37(1): 10–15.
- SHANG YE, YANG WM. Variation analysis of cereals mycotoxin limit standards of CAC, EU, USA, and China [J]. *J Food Sci Technol*, 2019, 37(1): 10–15.

(责任编辑: 于梦娇 郑 丽)

作者简介



吴升德, 博士, 高级工程师, 主要研究方向为产品质量检验检测。
E-mail: shengdewu@163.com



朱家骥, 博士, 讲师, 主要研究方向为食品质量与安全快速无损检测技术。
E-mail: zhujiaji@yeah.net



焦天慧, 博士, 讲师, 主要研究方向为食品快速无损检测技术。
E-mail: jth@jmu.edu.cn