DOI:10.12441/spyswjs.20250710001 [文章编号] 1673-1689(2025)12-0135-11

基于双波段高光谱与机器学习的花生中黄曲霉毒素B_1污染程度快速检测

任 璐[1] 沈书玥[1] 史婉荣[1] 芮 闯[1] 董茂锋[2] 李婷婷[1*]

(1. 上海健康医学院医学技术学院,上海 201318; 2. 上海市农业科学院农产品质量标准与检测技术研究所,上海 201403)

摘要:【目的】通过双波段高光谱与机器学习技术实现花生中黄曲霉毒素B_1(AFB_1)污染程度的快速检测。**【方法】**采用400~1 000 nm可见近红外与900~1 700 nm短波近红外双波段高光谱,结合机器学习构建花生AFB_1污染程度预警模型。通过采集4个污染梯度的花生样本,基于超高效液相色谱-串联质谱测定AFB_1的质量分数;获取双波段光谱数据后,比较4种预处理方法[一阶导数、二阶导数、多元散射校正、标准正态变换(standard normal variate,SNV)]、2种特征波长选择方法[连续投影算法、非信息变量剔除技术(uninformative variables elimination,UVE)]和2种机器学习方法[支持向量机(support vector machine,SVM)、反向传播神经网络]对分类效果的影响。**【结果】**SNV-UVE-SVM的组合模型在双波段均能较好地判断AFB_1污染程度,在900~1 700 nm波段该模型性能最优,其中特征波长主要集中在1 126~1 308 nm和1 588~1 720 nm,其训练集和测试集准确率均为100.00%。**【结论】**首次采用双波段高光谱与机器学习算法构建花生AFB_1污染程度智能预警模型,为农产品真菌毒素污染的无损检测提供了新型技术范式,对实现加工环节的实时在线监测具有重要应用价值。

关键词:黄曲霉毒素B_1;花生;近红外高光谱;机器学习

Rapid Detection of Aflatoxin B_1 Contamination Levels in Peanuts Based on Dual-band Hyperspectral Imaging and Machine Learning

REN Lu[1] *SHEN Shuyue*[1] *SHI Wanrong*[1] *RUI Chuang*[1] *DONG Maofeng*[2] *LI Tingting*[1*]

(1. The College of Medical Technology, Shanghai University of Medicine & Health Sciences, Shanghai 201318, China; 2. The Institute of Agro-food Standards and Testing Technology, Shanghai Academy of Agricultural Sciences, Shanghai 201403, China)

Abstract: [**Objective**] This study aims to achieve rapid detection of AFB_1 contamination levels in peanuts through dual-band hyperspectral imaging and machine learning. [**Method**] A peanut AFB_1 contamination level warning model was constructed *via* a dual-band hyperspectral imaging system consisting of visible near-infrared spectroscopy at 400~1 000 nm and short-wave near-infrared spectroscopy at 900~1 700 nm, combined with machine learning. Peanut samples with four contamination levels were collected, in which the mass fraction of AFB_1 was determined by ultra-high performance liquid chromatography-tandem mass spectrometry. The effects of four pretreatment methods [first derivative, second derivative, multiplicative scatter correction, and standard

基金项目:上海市东方英才计划青年项目(2023);上海市科委"青年科技英才扬帆计划"项目(21YF1418900);上海健康医学院校级科研基金项目(SSF-23-05-002)。

通信作者:李婷婷(1989—),女,博士,副教授,硕士研究生导师,主要从事食药用菌活性物质分离与作用机制研究。E-mail:litt@sumhs.edu.cn

收稿日期:2025-07-10 **修回日期:**2025-08-21

normal variate (SNV)], two characteristic wavelength selection methods [successive projections algorithm and uninformative variables elimination (UVE)], and two machine learning methods [support vector machine (SVM) and back-propagation neural network] on classification performance were compared. [**Result**] The combined model SNV-UVE-SVM effectively determined AFB_1 contamination levels in both bands, with optimal performance in the 900~1 700 nm band. The characteristic wavelengths were primarily concentrated in the ranges of 1 126~1 308 nm and 1 588~1 720 nm, achieving 100.00% accuracy in both the training and test sets. [**Conclusion**] This paper pioneered the construction of an intelligent early-warning model for AFB_1 contamination levels in peanuts by using dual-band hyperspectral imaging and machine learning algorithms, providing a novel technical paradigm for non-destructive detection of mycotoxin contamination in agricultural products. This approach holds significant engineering application value for real-time online monitoring in processing stages.

Keywords: aflatoxin B_1; peanut; near-infrared hyperspectral; machine learning

花生是一种重要的油料作物，其AFB_1污染防控直接关系食品安全。根据GB 2761—2017《食品安全国家标准 食品中真菌毒素限量》的规定，花生及其制品中AFB_1限量为20 μg/kg[1]。鉴于AFB_1的剧毒性和剂量效应特征，构建污染程度分级体系不仅符合国际监管要求，而且实现风险梯度预警与分级管控，对保障消费者健康具有重要现实意义[2-3]。

检测黄曲霉毒素的传统方法主要涵盖了感官评估、聚合酶链反应、高效液相色谱以及液相色谱-质谱联用技术等[4-7]。但这些技术方法存在操作误差显著或操作流程复杂等问题，难以实现快速检测。而酶联免疫吸附测定、胶体金免疫层析法等技术虽可实现快速检测（在2~4 h可获取检测结果），但其前处理过程复杂且存在操作误差，无法做到无损检测[8-9]。因此，开发兼具快速响应与无损特性的新型检测技术已成为突破粮油安全防控瓶颈的关键。

近红外高光谱成像技术是一种基于光学原理的无损检测技术，其融合传统成像与光谱学，能同时捕获空间与光谱信息，显著增强细微特征检测能力，已成功应用于黄曲霉毒素检测[10]。然而，光谱信息复杂且易受干扰，预测模型的选择依赖先验知识，传统化学定量方法处理困难。近年来，机器学习技术凭借其强大的数据处理能力和特征提取能力，为解决光谱信息复杂性和减少经验知识依赖提供了新的途径，为智能化、高精度黄曲霉毒素检测提供了可能[11]。

Zhang等[12]构建了基于朴素贝叶斯分类器的定量模型，通过结合Savitzky-Golay二阶导数光谱预处理与正态核密度估计优化策略，其训练集与测试集判别精度均超过91%，但该模型未涉及霉变程度分级，且近红外光谱设备只能用于定点检测，不易搬运；朱昊宇等[13]采用高光谱成像系统，Savitzky-Golay平滑联合二阶导数预处理，建立了SVM分类模型，霉变样本的识别测试集准确率达95%。使用蒙特卡洛-无信息变量消除法筛选特征波长对不同程度霉变样本进行模型构建，平均测试集准确率达90%。该方法在表观显著霉变样本中识别效果较好，但对工业化场景中普遍存在的潜在霉变样本的检测灵敏度仍需提升，特别是快速便携设备开发与多级霉变量化识别技术亟待突破。

作者将双波段高光谱成像技术与机器学习模型相融合，针对外观正常但内部霉变的花生样本，建立了基于非信息变量消除算法的SVM模型，以解决传统单波段检测灵敏度不足和近红外光谱仪器不易搬运的问题，满足现场快速筛查需求。该成果为农产品加工企业提供了可集成于分选产线的实时检测解决方案，具有显著的技术转化潜力。

1 材料与方法

1.1 材料与试剂

市售带壳新鲜红皮花生：产自江苏宿迁，选取形态均匀的120粒样本进行检测；AFB_1标准品、同位素内标$^{13}C_{17}$-AFB_1：纯度≥98%，曼哈格（上海）生物科技有限公司；甲醇：色谱纯，德国默克公司；醋酸铵（色谱纯）、乙酸（色谱纯）、PTFE滤膜（0.22 μm）：上海安谱实验科技股份有限公司。

1.2　仪器与设备

高光谱成像系统（配备FX10、FX17高光谱相机）：Specim SisuCHEMA型，芬兰Specim公司；恒温培养箱：Esco Celmate型，新加坡Esco公司；超高效液相色谱仪（联用BEH C18色谱柱，2.1 mm×50 mm×1.7 μm）：Waters ACQUITY Premier型，美国Waters公司；三重四极杆质谱系统：Waters APGC/Xevo TQ-XS型，美国Waters公司。

1.3　方法

1.3.1　样品制备

将花生分为4组（n=30），置于30 ℃、相对湿度85%的培养箱进行可控霉变。在0、2、4、6 d取样，构建包含未污染、早期污染（<10 μg/kg）、临界污染（≈20 μg/kg）及超标样本（>20 μg/kg）的渐进式样本库。

1.3.2　AFB_1的检测

根据GB 5009.22—2016《食品安全国家标准 食品中黄曲霉毒素B族和G族的测定》的规定，采用超高效液相色谱-串联质谱对花生中AFB_1质量分数进行精确测定[14]。

称取1.00 mg AFB_1标准品，用乙腈溶解并定容至100 mL，质量浓度约10 μg/mL，转移至试剂瓶后于-20 ℃避光保存。取1.0 mL的1.0 μg/mL标准储备液，用乙腈定容至100 mL，密封后于-20 ℃避光保存。分别移取混合标准工作液10、50、100、200、500、800、1 000 μL至10 mL容量瓶，各加200 μL的100 ng/mL的$^{13}C_{17}$-AFB_1同位素内标工作液，用初始流动相定容，配成0.1、0.5、1.0、2.0、5.0、8.0、10.0 ng/mL的系列标准溶液。

取1 kg以上的试样，粉碎过2 mm筛，混匀缩分至100 g，称取5 g（精确至0.01 g）于50 mL离心管，加100 μL同位素内标工作液，振荡静置30 min，加20.0 mL乙腈-水（体积比84∶16），涡旋后超声振荡提取20 min，6 000 r/min离心10 min，取上清液。移取4 mL上清液，加入46 mL含体积分数1%吐温-20的PBS，混匀后将样液以1~3 mL/min过恢复至室温的免疫亲和柱，用10 mL水淋洗抽干2次，再用1 mL甲醇以1~3 mL/min洗脱并抽干2次，收集洗脱液，50 ℃氮吹至近干，加1 mL初始流动相涡旋30 s溶解，0.22 μm滤膜过滤后上机检测。

1.3.3　高光谱数据采集

使用芬兰Specim公司的SisuCHEMA高光谱成像平台，搭载FX10和FX17两款相机进行图像采集，光谱采集与处理过程见图1。样品以20 mm/s速度移动，曝光时间3 ms，镜头与样品距离30 cm，空间分辨率为640 pixel/line，波段数224，帧频50 Hz，共采集120个花生样本。使用ENVI 4.6软件分析处理高光谱图像，提取样本光谱数据。

1.3.4　数据预处理与特征波长筛选

近红外光谱分析容易受到环境因素、仪器校正、样品颗粒度等多种因素的影响。通过应用恰当的预处理技术，可以有效缓解背景噪声和常见物理量的干扰，进而提升模型的预测精度[15]。运用以下4种光谱预处理技术：一阶导数（first derivative，FD）、二阶导数（second derivative，SD）、多元散射校正（multiplicative scatter correction，MSC）以及SNV。

全光谱数据集往往包含大量数据点，这些数据点中混杂着大量与预测目标无关的干扰信息，因此对特征波长进行筛选显得尤为重要[16]。运用连续投影算法（successive projections algorithm，

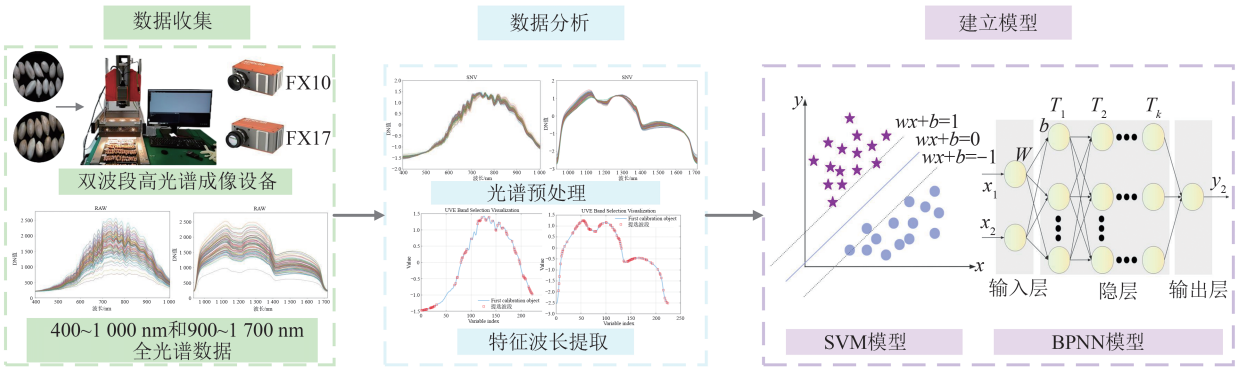


图1　花生AFB_1高光谱检测技术流程图

Fig. 1　Flowchart of hyperspectral detection technology for AFB_1 in peanuts

SPA）与 UVE，旨在筛选出关键的特征波长，并对所选波长进行建模与比较分析。

1.3.5 建模方法与模型评价标准

采用机器学习领域的SVM和反向传播神经网络（back propagation neural network，BPNN）构建了一种用于检测花生内部霉变的识别模型。在SVM模型构建中，基于径向基函数进行模型搭建，通过网格搜索法在指数空间（2^{-10}~2^{10}）内对模型参数组合进行系统优化。在模型评估过程中，按式（1）~（4）计算衡量模型分类性能的核心指标。

$$A=\frac{T_P+T_N}{T_P+F_P+T_N+F_N}\times 100\% \quad (1)$$

$$P=\frac{T_P}{T_P+F_P}\times 100\% \quad (2)$$

$$R=\frac{T_P}{T_P+F_N}\times 100\% \quad (3)$$

$$F_1=2\times\frac{R\times P}{R+P} \quad (4)$$

式中：

A——准确率，%；

P——精确率，%；

R——召回率，%；

F_1——F_1分数，%；

F_P——假阳性样本（实际不属于目标类别但被模型错误预测为该类别的样本）数量，个；

F_N——假阴性样本（实际属于目标类别但被模型错误预测为其他类别的样本）数量，个；

T_P——真阳性样本（实际属于目标类别且被模型正确预测为该类别的样本）数量，个；

T_N——真阴性样本（实际不属于目标类别且被模型正确预测为非该类别的样本）数量，个。

1.3.6 *t*-SNE 特征可视化

采用*t*-SNE算法对模型末层特征进行降维可视化，验证特征提取有效性。该算法通过高斯条件概率表征高维样本相似性，并基于*t*分布优化Kullback-Leibler散度，实现高维数据到二维空间的非线性映射。其重尾特性可同时保持数据局部聚类与全局拓扑关系，来自模型最后一层的特征向量使用*t*-SNE映射到二维空间。如果生成的数据分布表现出清晰的边界，则认为该数据是有效的聚类数据。

1.3.7 数据处理软件

使用 Matlab R2022a、Excel 2016、IBM SPSS Statistics 25、Origin 2018、Visio Pro 2024等软件进行数据分析处理和模型构建。

2 结果与讨论

2.1 不同霉变程度的AFB_1质量分数分析

试验系统测定了不同污染程度花生样本中AFB_1的质量分数，结果见表1。结果表明，对照组花生的AFB_1质量分数显著低于国家标准限量；轻度污染组的AFB_1质量分数在临界阈值下；中度污染组的AFB_1质量分数在阈值±5%；而重度污染组的AFB_1质量分数则显著超标。该试验结果有效模拟了AFB_1在花生中的动态累积过程，为食品安全风险评估提供了重要参考依据。

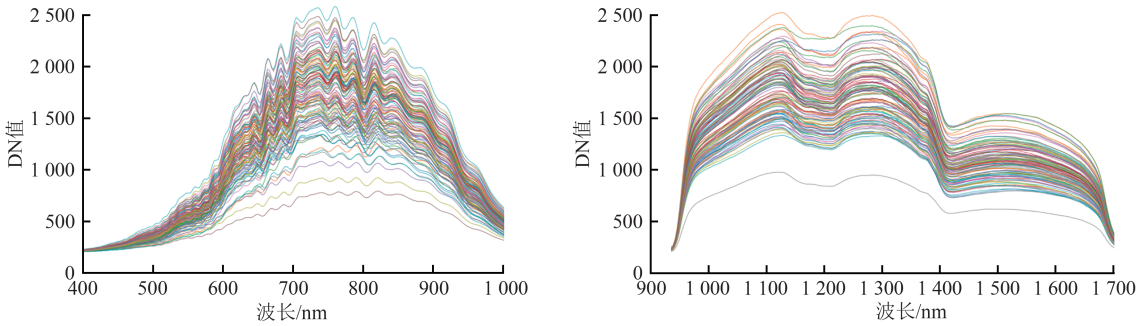
表1 不同污染程度花生中AFB_1质量分数测定结果

Table 1 Mass fractions of AFB_1 in peanuts with different contamination levels

分组	AFB_1质量分数/（μg/kg）
对照组	5.22±0.66
轻度污染组	10.88±0.74
中度污染组	21.23±0.68
重度污染组	27.07±0.87

2.2 光谱数据分析

所有花生样本的初始近红外光谱图见图2。在400~1 000 nm波长范围内的原始光谱（RAW）曲线未显现出显著的吸收峰特征，这可能与该波段光穿透深度有限、花生内部成分的特征光谱信号被掩盖有关。然而，在900~1 700 nm波段的光谱曲线在990、1 195、1 440 nm处有3个明显吸收区域。在光谱分析领域，观察到位于990 nm的吸收峰，可归因于O—H基团拉伸振动的第二泛音，该现象与淀粉分子结构中羟基振动密切相关；位于1 195 nm的吸收峰，对应于C—H基团拉伸振动的第二泛音，与淀粉分子中CH_3基团的分子结构密切相关，这是由于淀粉分子中甲基的存在；在1 440 nm的吸收峰，源于O—H基团拉伸振动的第一泛音，同样与淀粉分子中羟基振动相关。这些特定波长处的吸收峰为淀粉分子结构分析提供了关键的光谱学证据[17]。



(a) 所有样品400~1 000 nm的原始近红外光谱　　(b) 所有样品900~1 700 nm的原始近红外光谱

图2　所有样品的原始近红外光谱

Fig. 2　Original near-infrared spectra of all samples

通过光谱曲线可清晰识别出不同质量分数的AFB_1样品间存在显著差异。这种差异是由于在花生霉变过程中，微生物活动及霉菌毒素产生的同时消耗原料成分，导致化学组成发生改变，进而在光谱特征上呈现变化。为降低随机噪声，突出关键信息，消除基线漂移，提高信噪比，作者采用4种预处理技术对原始光谱数据进行分析。不同预处理方法对光谱特征的优化效果存在差异，结果见图3。FD和SD通过增强高频成分和消除基线漂移来突出局部特征，但可能放大噪声，导致光谱形态与原始数据差异显著；SNV和MSC在保留光谱整体形状的同时，校正散射和基线偏移并抑制噪声，可更有效地提取化学信息。因此，SNV和MSC在优化模型性能方面更具优势，更适合作为光谱预处理的优选方法。

2.3　样本组划分

采用Kennard-Stone（KS）算法对数据集进行划分，依据AFB_1质量分数对120个花生样本进行分组，随后按照4∶1的比例将样本划分为训练集和测试集。训练集与测试集中花生样品AFB_1质量分数的统计分析概况见表2。2个子集的AFB_1质量分数均无显著差异（P>0.05），验证了样本划分的化学计量学合理性，确保模型具有可靠泛化能力。

2.4　基于全波段波长的模型

将花生样本进行分组，对照组的样本标记为“0”，轻度污染组的样本标记为“1”，中度污染组的样本标记为“2”，重度污染组的样本标记为“3”。通过应用多种预处理技术于全波段近红外光谱数据，并结合相应的分类值，构建了用于花生中AFB_1检测的SVM判别模型，试验结果见表3。

构建的模型揭示了应用不同预处理技术对模型精确度的提升作用。在400~1 000 nm波段中，SVM模型经FD、SD和MSC预处理后准确率提升幅度为6.67%~20.00%，而SNV预处理将训练集、测试集准确率分别显著提升至95.83%、100.00%；BPNN模型在MSC预处理下准确率仅提升5.56%，SNV预处理则提高11.11%。在900~1 700 nm波段中，SVM和BPNN模型对FD、SD、MSC及SNV预处理的响应均较为敏感，准确率提升了15.79%~21.04%。其中SVM模型经SNV预处理后取得最高性能，其训练集准确率为93.75%，测试集准确率为95.83%。试验表明，SNV预处理通过消除基线漂移和背景干扰，能有效抑制表面散射效应，其在双波段分析中展现出最优特征增强效果，故选定为后续建模的标准预处理方法。

2.5　基于特征波长的模型

在全波段光谱分析中存在大量与目标物质预测无关的干扰信息。通过采用恰当的波长筛选技术，可以有效剔除不相关的变量，减少变量间的干扰，降低数据维度和计算量，从而提升模型的预测精度。作者采用SPA和UVE算法，对经过SNV预处理后的光谱数据进行了特征波长的筛选，结果见图4。

图4（a）展示了在400~1 000 nm波长范围内采用SPA算法选取的41个特征波长，UVE算法选取的77个特征波长。图4（b）展示了在900~1 700 nm波长范围内采用SPA算法选取的23个特征波长，UVE算法选取的94个特征波长。

由图4可知，在400~1 000 nm波段内的特征波长分布主要集中在397~465 nm、715~839 nm、

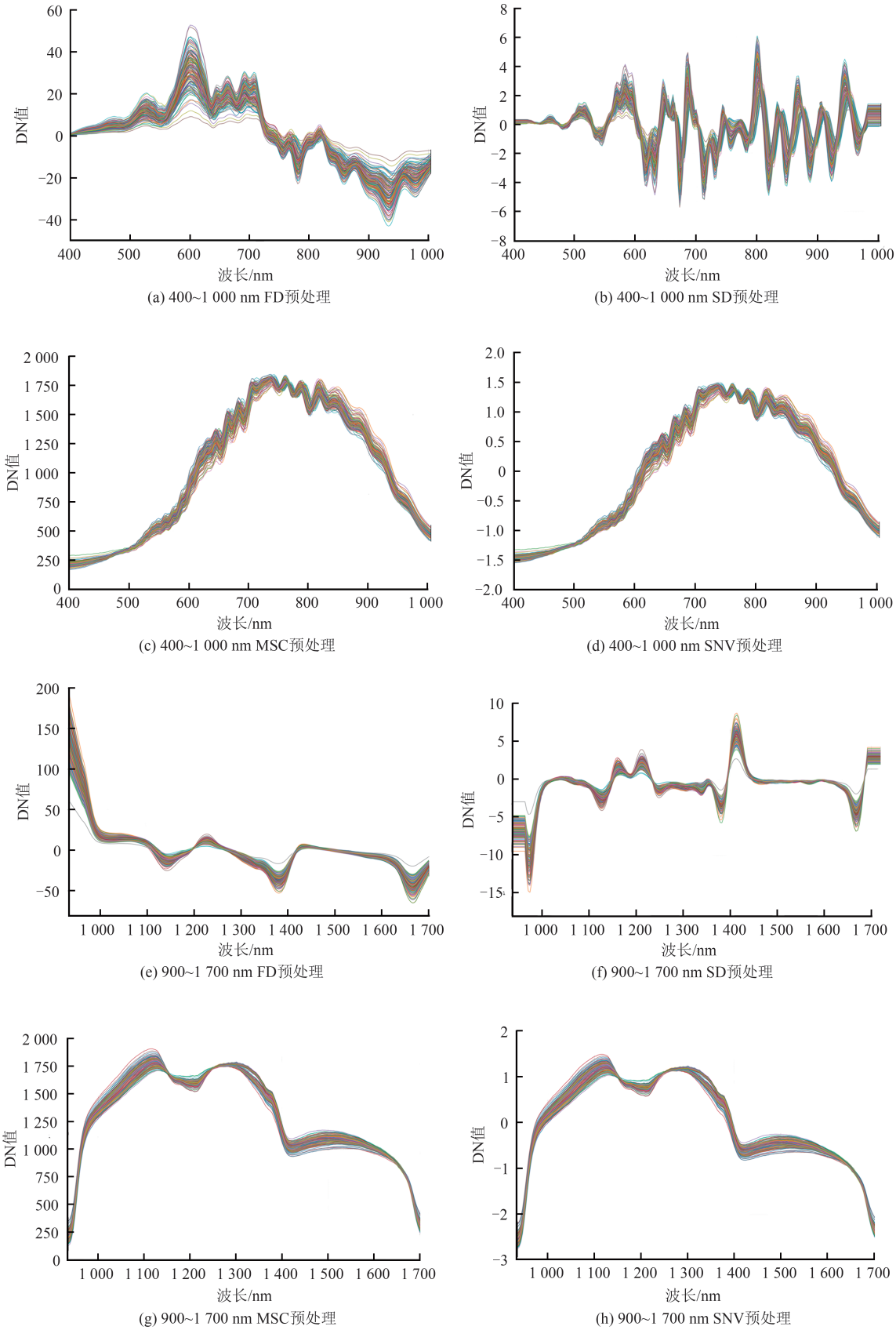


图 3　经预处理后所有样品的近红外光谱

Fig. 3　Near-infrared spectra of all samples after pretreatment

表 2　花生样品 AFB_1 在训练集和测试集中的测量

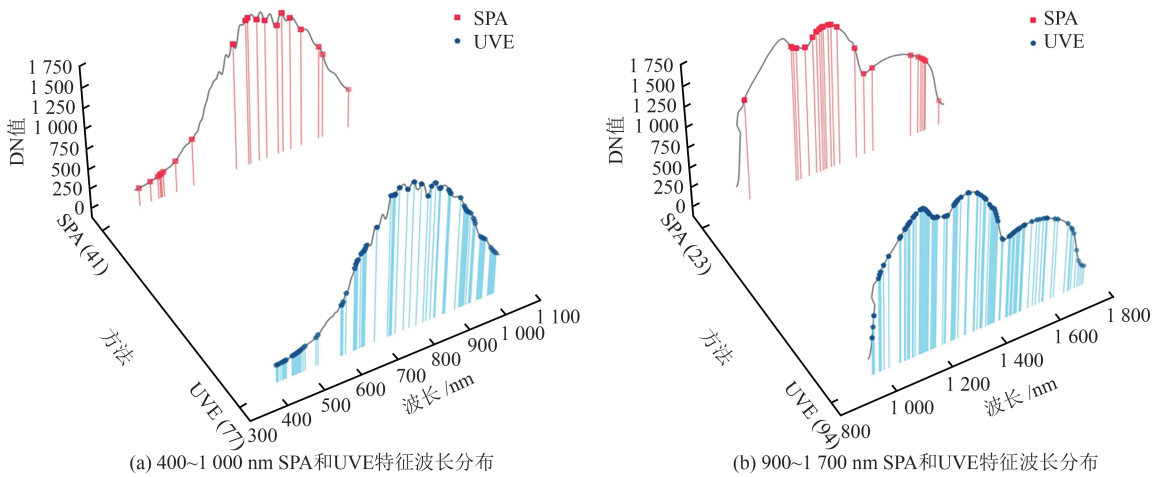
Table 2　Measurement of AFB_1 in peanut samples in the training set and test set

波长范围/nm	分组	样本数量/个					AFB_1 质量分数/(μg/kg)
		对照组	轻度污染	中度污染	重度污染	总计	
400~1 000	训练集	24	24	22	26	96	16.23±0.20
	测试集	6	6	8	4	24	16.75±0.13
900~1 700	训练集	24	23	23	26	96	16.14±0.19
	测试集	6	7	7	4	24	16.67±0.18

表 3　不同预处理方法下全波段 SVM 模型和 BPNN 模型的性能

Table 3　Performance of full-waveband SVM models and BPNN models under different pretreatment methods

波长范围/nm	训练模型	预处理方法	训练集				测试集			
			准确率/%	精确率/%	召回率/%	F1 分数/%	准确率/%	精确率/%	召回率/%	F1 分数/%
400~1 000	SVM	RAW	93.75	93.47	93.86	93.58	62.50	66.67	62.20	63.33
		FD	90.63	90.48	91.07	90.65	75.00	75.71	77.38	75.96
		SD	94.79	94.77	94.27	94.36	66.67	75.00	67.08	68.75
		MSC	90.63	90.16	90.24	90.12	75.00	77.59	77.50	76.94
		SNV	95.83	95.70	95.92	95.74	100.00	100.00	100.00	100.00
	BPNN	RAW	82.29	82.37	81.80	82.08	75.00	74.84	76.04	75.43
		FD	81.25	83.40	80.70	82.02	70.83	72.50	73.57	73.03
		SD	89.58	92.22	86.54	89.29	54.17	61.67	72.92	66.82
		MSC	89.58	89.57	88.92	89.24	79.17	82.08	82.59	82.34
		SNV	92.71	93.36	92.81	93.08	83.33	82.29	79.76	81.01
900~1 700	SVM	RAW	86.46	86.04	86.00	85.89	79.17	82.14	83.75	85.89
		FD	93.75	93.39	93.85	93.33	91.67	88.54	88.54	88.54
		SD	100.00	100.00	100.00	100.00	91.67	92.86	91.88	91.38
		MSC	100.00	100.00	100.00	100.00	91.67	93.75	92.26	92.23
		SNV	93.75	93.48	94.64	93.42	95.83	96.43	96.43	96.15
	BPNN	RAW	89.58	89.49	89.13	89.31	79.17	86.25	82.14	84.15
		FD	98.96	98.91	99.07	98.99	83.33	82.64	82.64	82.64
		SD	93.75	94.17	93.60	93.89	87.50	89.58	89.29	89.43
		MSC	92.71	92.42	92.37	92.40	83.33	84.52	86.46	85.48
		SNV	95.83	95.94	95.70	95.82	87.50	86.01	86.61	86.31



(a) 400~1 000 nm SPA和UVE特征波长分布　　(b) 900~1 700 nm SPA和UVE特征波长分布

图 4　所有样品经 SNV 预处理后的特征波长分布图

Fig. 4　Distribution of characteristic wavelengths for all samples after SNV pretreatment

925~1 003 nm波段范围内；在900~1 700 nm波段内的特征波长分布主要集中在1 126~1 308 nm、1 588~1 720 nm波段范围内。该分布规律与霉变过程中脂肪分解、蛋白质变性及淀粉降解引起的光谱响应变化高度吻合。UVE算法筛选的波段覆盖全光谱范围，印证了霉变对花生多组分协同影响的本质特征。模型建模结果的特征波长见表4。

表4 不同特征波长下SVM模型和BPNN模型的性能

Table 4 Performance of SVM models and BPNN models at different feature wavelengths

波长范围/nm	训练模型	提取特征波长方法	提取特征波长数量/个	训练集				测试集			
				准确率/%	精确率/%	召回率/%	F1分数/%	准确率/%	精确率/%	召回率/%	F1分数/%
400~1 000	SVM	全波长	224	95.83	95.70	95.92	95.74	100.00	100.00	100.00	100.00
		SPA	41	95.83	95.70	95.92	95.73	95.83	93.75	96.88	94.76
		UVE	77	92.71	92.48	92.87	92.50	100.00	100.00	100.00	100.00
	BPNN	全波长	224	92.71	93.36	92.81	93.08	83.33	82.29	79.76	81.01
		SPA	41	73.96	73.78	73.83	73.14	75.00	74.11	76.34	72.80
		UVE	77	87.50	87.23	87.59	87.23	91.67	91.67	91.88	90.56
900~1 700	SVM	全波长	224	93.75	93.48	94.64	93.42	95.83	96.43	96.43	96.15
		SPA	23	87.50	87.09	88.02	87.03	91.67	92.85	91.43	91.13
		UVE	94	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	BPNN	全波长	224	95.83	95.94	95.70	95.82	87.50	86.01	86.61	86.31
		SPA	23	93.75	93.66	93.72	93.61	83.33	81.85	86.67	83.36
		UVE	94	96.87	96.78	96.91	96.82	83.33	80.36	80.95	80.22

在400~1 000 nm波段中，SNV-UVE-SVM模型训练集准确率为92.71%，测试集准确率为100.00%，性能与全波段模型相当；在900~1 700 nm波段中，测试集准确率达100.00%，较全波段提升4.35%，成为最优模型。相较于SPA算法，UVE算法通过剔除冗余光谱信息并保留与AFB_1相关的C—H键及酰胺键特征波长，显著提升了模型泛化能力。虽然全波段模型平均准确率高出3.13%，但SNV-UVE-SVM组合通过减少60%~90%特征维度，在模型复杂度与精度间取得最优平衡。结合SVM算法对高维光谱数据的优异处理性能，本结果为开发高效便携的AFB_1检测设备提供理论依据。

2.6 对模型进行全面比较和验证

系统构建了SVM与BPNN的定性分析模型体系，旨在实现花生样本中AFB_1污染水平的精准判别。试验结果表明，不同光谱预处理方法对模型性能具有显著影响。具体而言，SNV预处理技术展现出卓越的模型优化能力，其通过有效消除光散射效应与基线漂移，显著提升了光谱数据的信噪比。

在特征波长选择方面，SPA虽可实现变量空间的降维，但试验表明该策略导致模型分类准确率下降。相比之下，UVE通过构建蒙特卡洛扰动矩阵，实现了光谱变量的智能筛选，在保留大部分有效信息量的同时，成功剔除冗余变量，使模型在测试集上的泛化误差降低。

模型架构对比分析表明，SVM模型在AFB_1检测任务中表现出显著优势。其核心机理在于通过核函数将原始光谱数据映射至高维特征空间，构建最优分类超平面，实现最大间隔分类。而BPNN模型受限于高维数据下的过拟合现象，且存在收敛速度慢等固有缺陷。

跨波段对比结果表明，基于SNV-UVE-SVM的集成模型在双波段（400~1 000 nm及900~1 700 nm）均展现出卓越性能。特别在900~1 700 nm短波近红外波段，模型在训练集和测试集均获得100.00%的准确率，特征波长主要富集于1 126~1 308 nm（C—H键三级倍频）及1 588~1 720 nm（C═O键伸缩振动）等特征官能团响应区域。图5~6的混淆矩阵可视化结果表明，该模型对4个污染梯度均实现零误判，Kappa系数达1.0，展现出完美的分类一致性。

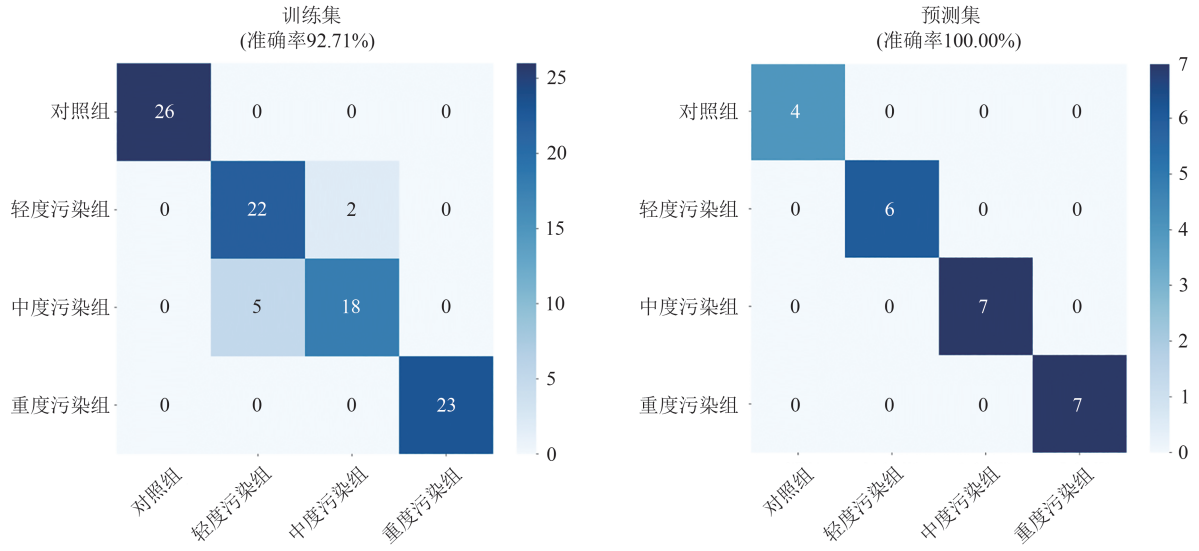


图 5　基于 400~1 000 nm 波长范围 SNV-UVE-SVM 模型的混淆矩阵图

Fig. 5　Confusion matrix diagram of the SNV-UVE-SVM model based on the wavelength range of 400~1 000 nm

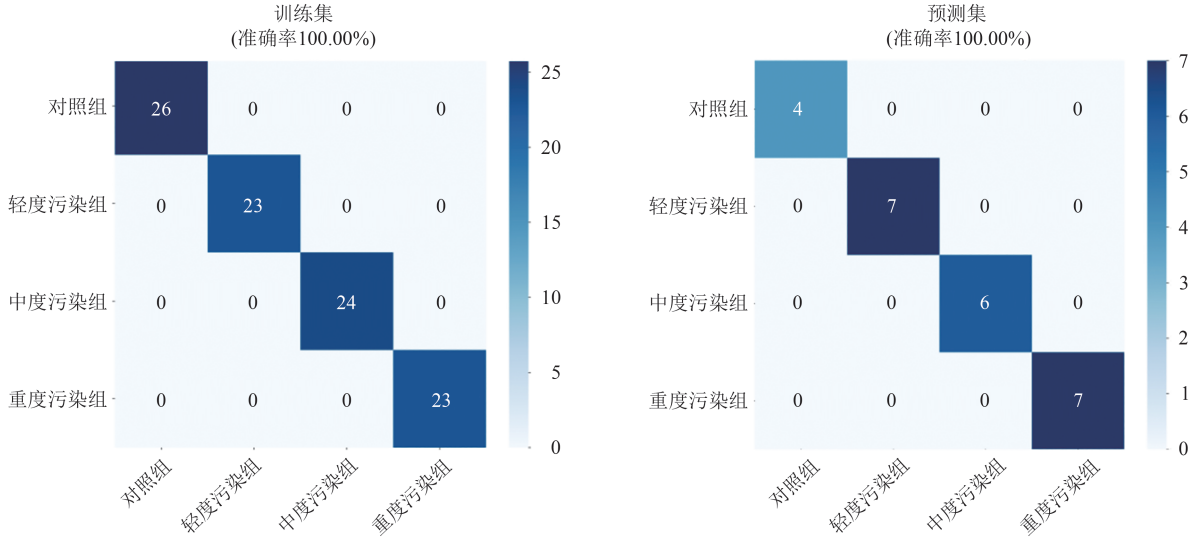


图 6　基于 900~1 700 nm 波长范围 SNV-UVE-SVM 模型的混淆矩阵图

Fig. 6　Confusion matrix diagram of the SNV-UVE-SVM model based on the wavelength range of 900~1 700 nm

根据训练集和测试集的预测结果，该模型在双波段均能非常有效地实现对花生中 AFB_1 污染的定性分析。在所有评估模型中，900~1 700 nm 波长范围内的 SNV-UVE-SVM 模型展现了最优性能，可被认定为花生中 AFB_1 污染定性分析的首选方案。

2.7　*t*-SNE 特征可视化

为了更直观地评估 SNV-UVE-SVM 模型预测不同 AFB_1 污染程度花生的分类性能，采用 *t*-SNE 算法对模型训练前后最后一层输出的高维特征进行可视化分析。*t*-SNE 缩小高维特征映射到二维平面绘制二维散点图，不同类别的样本以差异化色彩进行标识，直观地展示了特征的分布，为评估模型分类效能提供可视化依据。由图 7(a)可知，训练前重度污染组样本形成较为紧密的聚类簇，而其他 3 个类别的样本则呈现相互交叠的聚集态势，表明预训练数据无法对每个污染程度进行直接分类。由图 7(b)可知，经过训练后，相同污染程度的花生样本特征呈现高度聚集特性，不同污染程度的样本簇之间形成明显的边

界与间隔距离，仅存在少量轻度污染组样本被误分至对照组。该结果充分表明，SNV-UVE-SVM模型能够有效识别同一污染程度花生样本的特征模式，并且具备区分不同污染程度花生样本特征的能力，从而实现精准分类。综上所述，基于*t*-SNE可视化的实证分析，有力验证了SNV-UVE-SVM模型在不同AFB_1污染程度花生分类任务中的有效性与可靠性，为该模型在农产品质量安全检测领域的实际应用提供了坚实的理论依据与技术支撑。

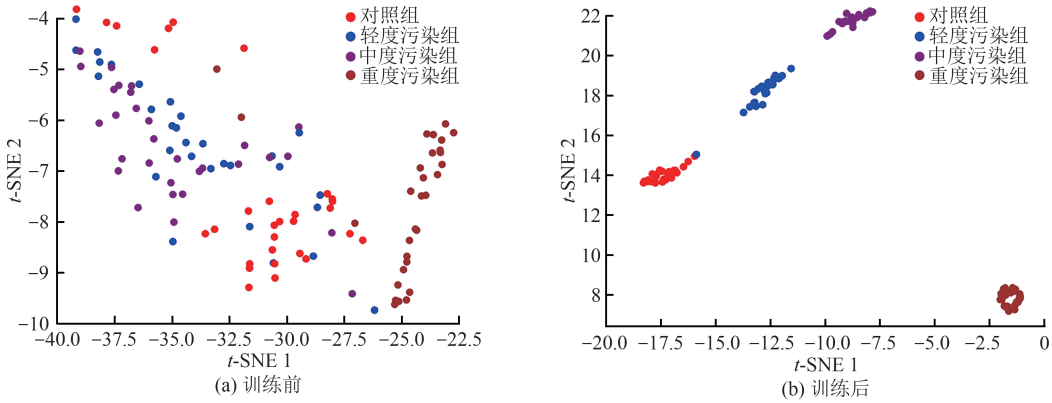


图 7　训练前后的高维特征*t*-SNE可视化结果

Fig. 7　Visualization results of high-dimensional features by *t*-SNE before and after training

3　结论

采用近红外高光谱成像技术与机器学习相结合，建立了对不同霉变程度花生的分类模型。在此过程中，制备了AFB_1质量分数梯度分布的4组共120份花生样品，并在双波段构建了全波段与特征波长模型。结果表明，SNV预处理方法在提升预测精度方面表现最为突出。同时，基于联合使用UVE的特征波长筛选方法，在不降低预测精度的前提下显著减少了数据处理量。最终，构建了基于SVM和BPNN的花生内部AFB_1定性鉴别模型。结果显示，在400~1 000 nm和900~1 700 nm波长范围内，SNV-UVE-SVM模型对花生AFB_1污染的鉴别效果最佳，其训练集准确率均超过90.00%，测试集准确率均为100.00%。在900~1 700 nm波长范围内，SNV-UVE-SVM模型展现了最优性能，其训练集和测试集准确率均为100.00%。作者通过双波段融合及流程优化显著提升了识别精度，为花生霉变分级检测提供了高效方案，验证了近红外高光谱技术的可靠性，其设备可满足现场快速筛查需求，为农产品质量安全智慧监管提供了可推广的技术范式。但该方案仍存在局限性，如样本库仅含单一品种、产地及贮存条件下的AFB_1质量分数梯度样本，以及模型对复杂实际场景的适应性有待验证。未来需扩充样本多样性（涵盖不同花生品种、产地、贮存条件等），增强模型对实际检测环境因素的抗干扰能力，推动技术向实际生产应用转化。

参考文献

[1] 中华人民共和国国家卫生和计划生育委员会，国家食品药品监督管理总局. 食品安全国家标准 食品中真菌毒素限量:GB 2761—2017[S]. 北京:中国标准出版社,2017.

[2] 尹玉云，陶健，焦强，等. 用于黄曲霉毒素B_1快速检测的胶体金试剂盒质量评价[J]. 食品与生物技术学报，2024,43(5):130-137.
YIN Y Y, TAO J, JIAO Q, et al. Quality evaluation of colloidal gold kits for rapid detection of aflatoxin B_1[J]. Journal of Food Science and Biotechnology, 2024, 43(5): 130-137.(in Chinese)

[3] 韩仲志，刘杰. 高光谱亚像元分解预测花生中的黄曲霉毒素B_1[J]. 中国食品学报,2020,20(3):244-250.
HAN Z Z, LIU J. Detecting aflatoxin B_1 in peanuts by hyperspectral subpixel decomposition [J]. Journal of Chinese Institute of Food Science and Technology, 2020, 20(3):244-250.(in Chinese)

[4] CHAHARAEIN M, SADEGHI E, MOHAMMADI R, et al. The effect of β-glucan and inulin on the reduction of aflatoxin B_1 level and assessment of textural and sensory properties in chicken sausages[J]. Current Research in

Food Science，2021，4：765-772.

[5] DALIRI A，SHAMS-GHAHFAROKHI M，RAZZAGHI-ABYANEH M. Detection of Aflatoxin B_1-producing *Aspergillus flavus* strains from pistachio orchards soil in Iran by multiplex polymerase chain reaction method [J]. Current Medical Mycology，2023，9(3)：1-7.

[6] WANG Y Z，HOU C，DAI Y Q，et al. Determination of aflatoxin B_1 by novel nanofiber-packed solid-phase extraction coupled with a high performance liquid chromatography-fluorescence detector[J]. Analytical Methods，2023，15(4)：472-481.

[7] SAILAJA O，MANORANJANI M，KRISHNAVENI G. Simultaneous estimation of aflatoxins (B_1，B_2，G_1 and G_2) by liquidchromatography coupled with mass spectrometry (LC-MS) in corn samples[J]. Asian Journal of Chemistry，2021，33(3)：521-526.

[8] 赵昕，郑树亮，牛晓颖，等. 基于高光谱成像技术和近红外光谱技术的金冠苹果货架期判别及其品质分析[J]. 食品工业科技，2025，46(11)：302-312.
ZHAO X，ZHENG S L，NIU X Y，et al. Shelf life identification and quality analysis of golden delicious apples based on hyperspectral imaging and near infrared spectroscopy [J]. Science and Technology of Food Industry，2025，46(11)：302-312.(in Chinese)

[9] LI X F，LIU L Q，SONG S S，et al. Colloidal gold immunochromatographic assay for the detection of total aflatoxins in cereals[J]. Food Chemistry，2025，472：142877.

[10] LANDGREBE D. Hyperspectral image data analysis[J]. IEEE Signal Processing Magazine，2002，19(1)：17-28.

[11] 朱宏飞. 基于深度学习的花生黄曲霉毒素高光谱检测关键技术研究[D]. 天津：天津工业大学，2023：1-84.

[12] ZHANG S，LI Z X，AN J，et al. Identification of aflatoxin B_1 in peanut using near-infrared spectroscopy combined with naive Bayes classifier[J]. Spectroscopy Letters，2021，54(5)：340-351.

[13] 朱昊宇，王俊杰，杨一，等. 基于近红外高光谱图像的花生内部霉变快速判别方法研究[J]. 食品安全质量检测学报，2024，15(1)：85-91.
ZHU H Y，WANG J J，YANG Y，et al. Research on rapid discrimination for internal mold detection in peanuts based on near-infrared hyperspectral image[J]. Journal of Food Safety & Quality，2024，15(1)：85-91.(in Chinese)

[14] 中华人民共和国国家卫生和计划生育委员会，国家食品药品监督管理总局. 食品安全国家标准 食品中黄曲霉毒素B族和G族的测定：GB 5009.22—2016[S]. 北京：中国标准出版社，2017.

[15] BIAN X H. Spectral dimensionality reduction methods [M]. Singapore：Springer Nature Singapore，2022：209-236.

[16] LI Z X，TANG X Y，SHEN Z X，et al. Comprehensive comparison of multiple quantitative near-infrared spectroscopy models for *Aspergillus flavus* contamination detection in peanut[J]. Journal of the Science of Food and Agriculture，2019，99(13)：5671-5679.

[17] DACHOUPAKAN S C，PUTTHANG R，SIRISOMBOON P. Application of near infrared spectroscopy to detect aflatoxigenic fungal contamination in rice [J]. Food Control，2013，33(1)：207-214.

（责任编辑：李春丽）