

婴幼儿奶粉中多种掺假物近红外高光谱图像检测方法

赵昕¹ 马竞一¹ 陈晗¹ 姜洪喆² 褚璇³ 赵志磊^{1,4}

(1. 河北大学质量技术监督学院, 保定 071002; 2. 南京林业大学机械电子工程学院, 南京 210037;

3. 仲恺农业工程学院机电工程学院, 广州 510225; 4. 河北大学计量仪器与系统国家地方联合工程研究中心, 保定 071002)

摘要: 奶粉市场是食品掺假行为频发领域, 其中婴幼儿配方奶粉价格高, 其质量是消费者、生产企业和执法部门关注的重点。近红外高光谱成像 (Near infrared - hyperspectral imaging, NIR - HSI) 技术结合化学计量学和机器学习算法可以检测奶粉中单一掺假物含量。基于 NIR - HSI 技术研究了不同品牌婴幼儿奶粉中多掺假物 (三聚氰胺、香兰素和淀粉) 的定量预测。对基于像素点预处理后的高光谱图像划分感兴趣区域 (Region of interest, ROI), 提取 ROI 平均光谱。基于经典的过滤式特征选择算法拉普拉斯分数 (Laplacian score) (无监督) 和 ReliefF (有监督) 挑选建模关键变量, 建立偏最小二乘回归模型 (Partial least squares, PLS)。开发包含自定义选择层的一维卷积神经网络模型 (One-dimensional convolutional neural networks, 1DCNN)。自定义层根据权重系数绝对值, 可确定重要波长变量。Laplacian score - PLS 模型对预测集中奶粉、三聚氰胺、香兰素和淀粉质量分数预测结果均方根误差分别为 0.111 0%、0.057 0%、0.034 9% 和 0.348 1%。ReliefF - PLS 模型对预测集中奶粉、三聚氰胺、香兰素和淀粉预测结果均方根误差分别为 0.199 8%、0.054 0%、0.045 5% 和 0.182 3%。1DCNN 模型对预测集中奶粉、三聚氰胺、香兰素和淀粉质量分数预测结果均方根误差分别为 0.856 1%、0.091 1%、0.064 4% 和 0.294 2%。对 Laplacian score、ReliefF 和自定义选择层挑选出的前 15 个重要波长进行对比分析, 不同特征选择方法挑选的特征波长子集有所区别, 但都选择 1 210、1 474、1 524、1 680 nm 等附近波长。基于 ReliefF - PLS 模型的可视化结果表明了其良好的预测能力。

关键词: 奶粉掺假; 拉普拉斯分数算法; ReliefF 算法; 卷积神经网络; 近红外高光谱成像

中图分类号: O657.3; TS252.51 **文献标识码:** A **文章编号:** 1000-1298(2024)04-0368-08

OSID:



Feature Analysis of Detection of Multiple Adulterants Simultaneously in Infant Milk Powder Using Hyperspectral Images

ZHAO Xin¹ MA Jingyi¹ CHEN Han¹ JIANG Hongzhe² CHU Xuan³ ZHAO Zhilei^{1,4}

(1. College of Quality and Technical Supervision, Hebei University, Baoding 071002, China

2. College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing 210037, China

3. College of Mechanical and Electrical Engineering, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China

4. National & Local Joint Engineering Research Center of Metrology Instrument and System, Hebei University, Baoding 071002, China)

Abstract: Milk powder is the hardest hit area for food adulteration. Among them, infant formula milk powder is expensive and important, with quality being the focus of consumers, manufacturers, and law enforcement agencies. Near infrared - hyperspectral imaging (NIR - HSI) technology combined with chemometrics and machine learning algorithms can detect the content of single adulterant in milk powder. The quantitative prediction of multiple adulterants (melamine, vanillin and starch) in different brands of infant milk powder was studied based on NIR - HSI technology. The hyperspectral images after pixel wise pretreatment were divided into regions of interest (ROI), and the ROI average spectra were extracted. The key variables for modeling were selected based on the classic filtering feature selection algorithms, i. e. Laplacian score (unsupervised) and ReliefF (supervised). Partial least squares (PLS) regression was adopted to establish prediction models. A one-dimensional convolutional neural network (1DCNN)

收稿日期: 2023 - 09 - 12 修回日期: 2023 - 11 - 30

基金项目: 国家自然科学基金项目 (32102087)、河北省省级科技计划项目 (21344801D) 和河北省专业学位研究生教学案例建设项目 (KGJSZ2022005)

作者简介: 赵昕 (1992—), 女, 讲师, 博士, 主要从事食品药品光学无损快检技术与装置研究, E-mail: zhaoxinzj@hbu.edu.cn

通信作者: 赵志磊 (1977—), 男, 教授, 博士, 主要从事食品药品计量检测研究, E-mail: 15930721279@163.com

model with a self-defined selection layer was developed. The self-defined layer determined the important wavelength variables according to the multiplicative weight parameters learned after modeling. The root mean square errors of prediction set of Laplacian score – PLS models to predict milk powder, melamine, vanillin and starch were 0.111 0% , 0.057 0% , 0.034 9% and 0.348 1% , respectively. The root mean square errors of prediction set of ReliefF – PLS models to predict milk powder, melamine, vanillin and starch were 0.199 8% , 0.054 0% , 0.045 5% and 0.182 3% , respectively. The root mean square errors of prediction set of 1DCNN models to predict milk powder, melamine, vanillin and starch were 0.856 1% , 0.091 1% , 0.064 4% and 0.294 2% , respectively. The first 15 important wavelengths selected by Laplacian score, ReliefF and self-defined selection layer were compared and analyzed. The characteristic wavelength subsets selected by different feature selection methods were different, but the wavelengths near 1 210 nm, 1 474 nm, 1 524 nm and 1 680 nm were selected in more than one method. The visualization results based on the ReliefF – PLS model demonstrated good predictive ability.

Key words: milk powder adulteration; Laplacian score algorithm; ReliefF algorithm; convolutional neural network; near infrared hyperspectral imaging

0 引言

奶粉的质量安全问题一直是消费者关注的焦点。然而由于商业利益驱使,且掺伪手段容易且较难检出,使得奶粉成为食品掺假行为频发领域,并已成为全球性食品安全问题^[1]。

目前,奶粉掺假检测采用随机抽样、实验室送检方式,流程长效率低。多数借助于化学方法,主要包括高效液相色谱法^[2]、气相色谱法^[3]、质谱法^[4]、毛细管电泳法^[5]和荧光光谱法^[6]等。然而,这些常规方法往往需要专业、繁琐的前处理步骤和复杂、昂贵的分析仪器,导致其单次样品检测时间长,难以满足现场快速实时评价的检测需求。近年来,光谱技术快速发展,其无损、快速和绿色等检测优点,为乳粉掺杂检测提供了一种高效的新手段。其中,高光谱成像技术基于含氢基团振动吸收和二维空间成像的原理,能通过光谱和图像,获取内部成分和外部物理特征信息,在肉类^[7]、农产品(如水稻^[8]、玉米^[9]、大米^[10])和调味品(如香料^[11])等的掺假检测领域均取得了系列研究成果。

对于奶粉掺假检测,文献[12]通过高光谱成像结合 PLS(偏最小二乘)算法,在 990 ~ 1 700 nm 光谱范围内检测奶粉中掺杂的三聚氰胺,基于可视化 PLS 结果图和阈值法能够检测混合样本中低至 0.02% 掺假质量分数水平的三聚氰胺颗粒。文献[13]通过近红外高光谱成像技术和多元曲线分辨率(Multivariate curve resolution, MCR)方法对奶粉一元(乳清粉、淀粉、尿素或三聚氰胺)、二元(淀粉和尿素)和三元(淀粉、尿素和乳清粉)等不同复杂情况下的掺假检测进行了研究。所提出的方法不需要大量的先验信息,预测的掺假物质量分数的绝对误差均低于 5%。以上研究表明高光谱成像技术结合适当的分析算法,可以有效检测奶粉中低含量

和多元掺假情况。

高光谱图像数据的共线性、冗余性和高维性,以及近红外谱区吸收峰重叠等特点,使其数据解析和特征挖掘成为一项艰巨的任务。不同于以往的建模算法,深度学习算法基于端到端的建模思想,从输入数据中自动提取抽象特征并学习,预测结果可达到较理想的精确水平。文献[14]利用近红外光谱结合基于注意力机制的集成卷积神经网络(Convolutional neural networks, CNN)建模方法,检测不同品牌婴幼儿配方奶粉中单一掺杂的水解皮革蛋白和三聚氰胺,与经典回归方法(PLS、核偏最小二乘(Kernel partial least squares, K-PLS)和 CNN)相比,所提出的集成方法表现出更优越的回归性能。

尽管已有以上较多针对掺假奶粉的检测研究,但大多主要讨论单一掺假情况,且奶粉样本的品牌类型较单一,而实际掺假情况往往比较复杂。本文选取 4 种品牌的婴幼儿配方奶粉,向其中同时掺入三聚氰胺、香兰素和淀粉 3 种掺假物,分别建立基于特征选择算法的经典机器学习模型和深度学习卷积神经网络模型。利用两种经典过滤式特征选择算法:无监督的拉普拉斯分数(Laplacian score)和有监督的 ReliefF,建立 PLS 模型。构建能进行解释性分析的添加自定义选择层的 1DCNN 模型^[15],对不同掺假物进行定量预测。讨论 3 种方法确定的最优波长并对最优模型的预测结果进行可视化分析。

1 实验

1.1 样品制备

从河北省保定市某大型超市购买 4 个品牌的 1 段(0 ~ 6 个月)婴幼儿配方奶粉:A(飞鹤)、B(完达山)、C(贝因美)、D(雅培)。香兰素和三聚氰胺购自北京易秀博谷生物科技有限公司,纯度 99.5%

以上。按照表 1 所示,将奶粉与不同质量分数的三聚氰胺、香兰素和淀粉充分混合,制备多掺假样品。

表 1 掺假样品中各掺杂物的混合质量分数

Tab.1 Mass fractions of different adulterants in mixed samples %

奶粉品牌	质量分数			
	奶粉	三聚氰胺	香兰素	淀粉
A	94.5	5.0	0.5	0
A	94.5	0	0.5	5.0
B	85.0	5.0	0	10.0
C	79.0	5.0	1.0	15.0
D	62.0	10.0	3.0	25.0

将每种混合物放入 3 个塑料方形培养皿(100 mm × 100 mm × 15 mm)中,粉末层厚度约占培养皿高度的 1/2。按照相同方式制备纯奶粉、三聚氰胺、香兰素和淀粉样本。

为了验证文中所建方法对外部验证样品的预测能力,从网上购买同样 4 个品牌的不同批次奶粉样品。三聚氰胺、香兰素购自上海麦克林生化科技有限公司,淀粉购自国药集团化学试剂有限公司。同样按照表 1 所示质量分数制备混合物样品作为外部验证集。此外,为了进一步分析模型对不同品牌奶粉掺假的预测能力,如表 2 所示,设计了新的外部验证样品。新外部验证样本制备所用的奶粉和掺假物与外部验证样品所用相同,每种混合物放于一个塑料方形培养皿进行后续的数据采集。

表 2 新外部验证集混合掺假样品的奶粉品牌和各物质质量分数

Tab.2 Milk powder brands and mass fractions of different adulterants of mixed adulterated samples in new external validation set %

奶粉品牌	质量分数			
	奶粉	三聚氰胺	香兰素	淀粉
B	94.5	5.0	0.5	0
C	94.5	0	0.5	5.0
C	85.0	5.0	0	10.0
D	79.0	5.0	1.0	15.0
A	62.0	10.0	3.0	25.0

1.2 高光谱图像获取

使用推扫式高光谱成像系统(SPECIM SisuCHEMA 型, Spectral Imaging Ltd. 公司, 芬兰)(图 1)采集样品的高光谱图像。基于 InGaAs 检测器的线扫描高光谱相机(SPECIM FX10 型, Spectral Imaging Ltd. 公司, 芬兰)采集的光谱范围为 935.61 ~ 1 720.23 nm,光谱分辨率为 3.45 nm,图像分辨率约为 0.26 mm/像素。

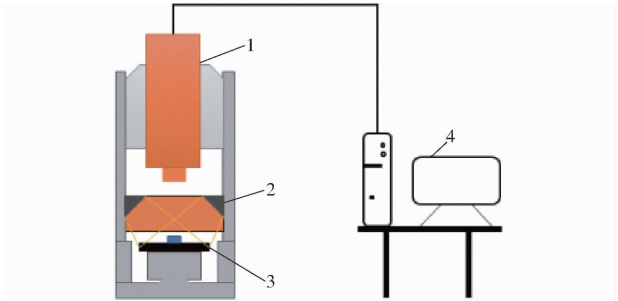


图 1 高光谱成像仪结构示意图

Fig.1 Schematic of hyperspectral imaging system

1. 高光谱成像仪 2. 光源 3. 载物台 4. 计算机

为了减少暗电流和环境光变化等因素对高光谱图像的影响,需要对原始高光谱图像进行黑白校正。校正公式为

$$R = \frac{l_0 - B}{W - B}$$

式中 R ——校正图像反射率
 l_0 ——原始图像反射强度
 B ——黑参考图像反射强度
 W ——白参考图像反射强度

为了去除每幅图像中培养皿和背景等不相关像素,选择 3 个培养皿中的粉末像素作为感兴趣区域(ROI),并从校正图像中裁剪出来。每个裁剪子图像尺寸为 360(像素) × 360(线) × 224(波段)。总共获得 15 个子图像(5 个含量梯度 × 3 个重复)。粉末样品反射率在光谱范围 936 ~ 1 157 nm 内几乎饱和(接近 100%)。此外,1 700 nm 之后光谱范围内的数据通常信噪比较低。因此,研究仅对 1 158 ~ 1 702 nm 的 155 个波段数据进行分析 and 处理。

1.3 数据处理方法

1.3.1 预处理

为了消除探测器灵敏度差异、样本不规则表面散射等干扰对光谱数据的不利影响,将亮度校正、吸收光谱转换和标准正态变换(Standard normal variate, SNV)等预处理算法应用于高光谱图像中的每个像素点光谱,具体说明详见文献[16]。

随后,每个子图像被均等分成 9 个 ROI,并计算 ROI 平均光谱。每种粉末混合物和纯粉末样品,提取 27 个 ROI 平均光谱(9 个 ROI × 3 个平行),共获得 135 条掺假样本光谱(27 条光谱 × 5 个掺假比例)。

1.3.2 偏最小二乘回归

偏最小二乘是一种经典的定量建模方法,它通过将大量自变量减少为一组新的不相关变量(称为潜在变量)来避免多重共线^[17]。

1.3.3 卷积神经网络

CNN 通过正向和反向传播更新神经网络的权

值,以最小化预测值与真实标签之间的差异。本文设计了一种卷积神经网络模型,包括 1 层输入层、1 层自定义层、2 层卷积层、2 层激活层、2 层池化层、1 层全连接层和 1 层输出层。网络学习率和迭代次数分别为 0.001 和 700。

1.3.4 特征波长选择

特征选择 (Feature selection, FS)^[18] 是解决高维数据维度灾难问题的主要降维方法。过滤式特征选择算法具有独立于训练模型、结构简单、速度快的优点。本文采用了两种经典的过滤式特征选择方法,分别是无监督的拉普拉斯分数 (Laplacian score) 和有监督的 ReliefF。Laplacian score 基于局部保持投影和拉普拉斯特征映射,结合局部结构信息和方差,根据特征的局部保持能力来评估特征得分^[19]。ReliefF 基于特征对近距离样本的区分能力度量特征与类别标签的相关性,基于此计算特征的重要性^[20]。

1.3.5 预测结果可视化

从粉末混合物中随机选择尺寸为 150 像素 × 150 像素的 ROI 用于可视化。将感兴趣区域中每个像素点视作一个样本提取光谱构建二维 (2-D) 矩阵输入最优模型,获得可视化预测结果。

2 结果与讨论

2.1 纯样品的光谱特征

图 2 为纯样本奶粉和 3 种掺假物:三聚氰胺、香兰素和淀粉的平均光谱。4 种物质的光谱曲线趋势总体差异明显,为以后的分类预测提供了基础。掺假物中淀粉与奶粉的光谱曲线趋势相近,但奶粉在 1 210 nm 处的吸收峰更高,该吸收峰与蛋白质中 C—H 键伸缩振动的第二泛音有关^[21]。香兰素、奶粉和淀粉均在 1 577 nm 处出现吸收峰,其中香兰素在该处的吸收较强,奶粉和淀粉该处吸收较弱。三聚氰胺在 1 524 nm 和 1 492 nm 处存

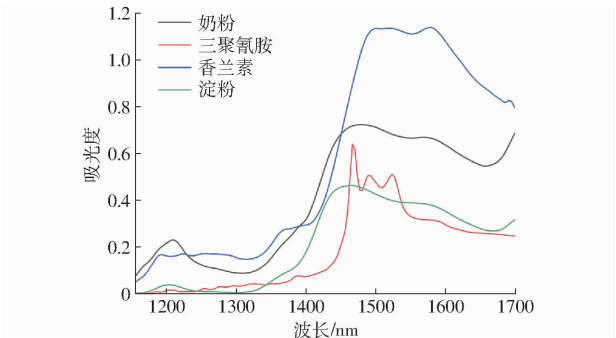


图 2 奶粉、三聚氰胺、香兰素和淀粉纯样品的平均光谱

Fig.2 Average spectra of pure samples of milk, melamine, vanillin and starch

在的 2 个较强吸收峰,分别与 N—H 对称和反对称拉伸振动的第一泛音有关,与文献[22]研究结果一致。三聚氰胺中芳香胺结构的存在,使得三聚氰胺和其他物质在 1 467 nm 左右的光谱处存在最显著的光谱差异^[23]。

2.2 PLS 掺假模型建立

2.2.1 全波段 PLS 模型

将每种混合类型的 27 条光谱按照比例 2:1 随机划分为校正集和预测集,共得到校正集 90 条光谱 (18 条 × 5 种),预测集 45 条光谱 (9 条 × 5 种)。采用五折法建立交叉验证集。通过交叉验证均方根误差 (RMSEV),确定最优潜变量因子数。PLS 全波长模型预测结果如表 3 所示。可以看出,奶粉、三聚氰胺、香兰素和淀粉预测模型的校正集、交叉验证集和预测集决定系数 R^2 均大于 0.960 0,表明建立的 PLS 模型能够有效预测混合物中各物质含量。根据交叉验证集均方根误差,模型对三聚氰胺的预测效果最优,淀粉次之,其次是奶粉,香兰素最差。

表 3 PLS 全波长模型预测结果
Tab.3 Prediction results of PLS models based on full wavelengths

物质	潜变量因子数	校正集		交叉验证集		预测集	
		R^2_c	均方根误差/%	R^2_{cr}	均方根误差/%	R^2_p	均方根误差/%
奶粉	7	0.999 9	0.141 0	0.999 8	0.155 3	0.999 8	0.160 5
三聚氰胺	7	0.999 9	0.038 0	0.999 8	0.041 7	0.999 8	0.047 0
香兰素	7	0.999 3	0.028 3	0.963 2	0.213 2	0.999 1	0.031 4
淀粉	6	0.999 8	0.120 8	0.999 8	0.135 3	0.999 8	0.133 3

2.2.2 基于特征波长的 PLS 模型

利用两种特征波长筛选方法 Laplacian score 和 ReliefF 优选 15 个特征波长,并基于此设计多光谱 PLS 模型,结果如表 4 所示。所有模型决定系数均大于 0.960 0,预测能力略低于全波长模型,但仍能较理想地预测奶粉、三聚氰胺、香兰素和淀粉。在奶粉、三聚氰胺和淀粉预测中,ReliefF - PLS 模型效果更好,在香兰素预测中,Laplacian score - PLS 模型效果更好。

与全波段 PLS 模型相比,采用筛选特征波长的方法会大大减小建模波长数量,进而缩减模型运算时间,提高检测效率。Laplacian score 和 ReliefF 方法分别在奶粉、三聚氰胺、香兰素和淀粉检测模型中筛选出的波长在平均光谱上的分布如图 3 所示。由于 Laplacian score 为非监督算法,故对不同物质筛选出的波长基本相同,微小差异是由建立不同物质的检测模型时随机划分的不同训练集导致。所选最优波长点大多分布在 1 450 ~ 1 550 nm 之间特征吸

表 4 基于特征波长的多光谱 PLS 模型预测结果

Tab.4 Prediction results of multispectral PLS models based on characteristic wavelengths									
物质	方法	特征波长	潜变量	校正集		交叉验证集		预测集	
		个数	因子数	R_c^2	均方根误差/%	R_{cr}^2	均方根误差/%	R_p^2	均方根误差/%
奶粉	Laplacian score	15	7	0.999 1	0.358 1	0.998 7	0.449 1	0.998 5	0.111 0
三聚氰胺		15	9	0.999 8	0.042 0	0.999 8	0.048 4	0.999 7	0.057 0
香兰素		15	9	0.999 2	0.030 4	0.963 3	0.213 6	0.998 9	0.034 9
淀粉		15	8	0.998 9	0.283 4	0.998 5	0.331 6	0.998 4	0.348 1
奶粉	RelieFF	15	9	0.999 8	0.159 7	0.999 8	0.188 0	0.999 7	0.199 8
三聚氰胺		15	11	0.999 9	0.036 8	0.999 8	0.044 1	0.999 7	0.054 0
香兰素		15	8	0.998 8	0.035 8	0.961 9	0.215 1	0.998 2	0.045 5
淀粉		15	12	0.999 8	0.107 5	0.999 8	0.130 1	0.999 6	0.182 3

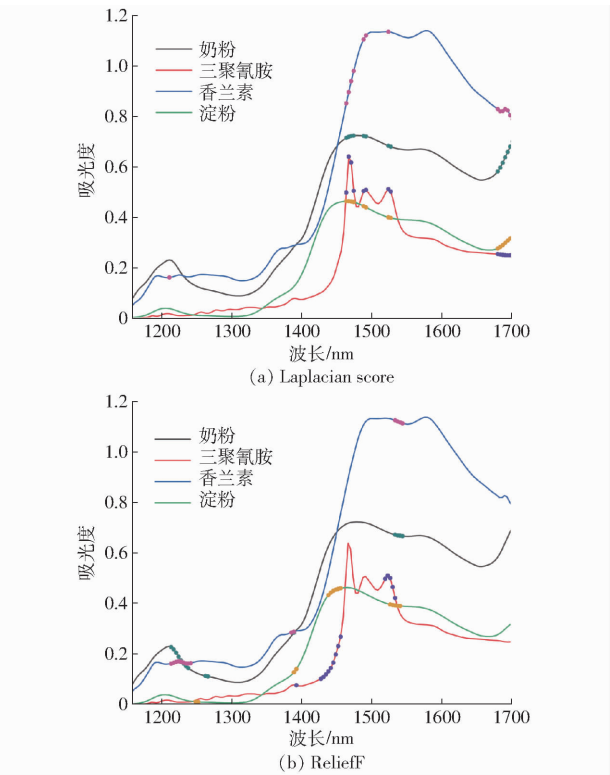


图 3 特征波长在平均光谱图上的分布
Fig.3 Distributions of characteristic wavelengths on average spectra

收峰附近,以及 1 680 ~ 1 700 nm 之间。Laplacian score 筛选了 1 467、1 492、1 524 nm 共 3 个与三聚氰胺最密切相关的吸收特征(见 2.1 节)。RelieFF 筛选出的特征波长集中分布于 1 213 ~ 1 241 nm、1 386 ~ 1 393 nm、1 428 ~ 1 456 nm 以及 1 527 ~ 1 541 nm。其中,波长 1 224 nm 与香兰素中的 C—H 键有关^[24],波长 1 393、1 446、1 450 nm 与淀粉中的 O—H 键的拉伸振动有关^[25-27]。不同物质模型所对应的特征波长分布较接近。Laplacian score 和 RelieFF 方法均挑选 2.1 节中讨论过的特征吸收峰 1 213 nm 和 1 524 nm 附近波长。

2.3 CNN 掺假模型建立

将每种混合类型的 27 条光谱按照比例 2:1:1

随机划分为训练集、验证集和预测集,共得到训练集 70 条光谱(14 条×5 种),验证集 35 条光谱(7 条×5 种),以及预测集 30 条光谱(6 条×5 种)。奶粉、三聚氰胺、香兰素和淀粉基于全波段建立的 CNN 模型的预测结果如表 5 所示。所有模型决定系数均在 0.995 0 以上,预测效果略低于全波长 PLS 模型和多光谱 PLS 模型,具有较好的预测性能。根据自定义选择层的权重系数绝对值,优选前 15 个特征波长,并标注在平均光谱图上(图 4)。

表 5 CNN 模型的预测结果 Tab.5 Prediction results of CNN model						
物质	训练集		验证集		预测集	
	R_c^2	均方根 误差/%	R_{cr}^2	均方根 误差/%	R_p^2	均方根 误差/%
奶粉	0.996 3	0.786 0	0.996 8	0.753 4	0.995 8	0.856 1
三聚氰胺	0.999 9	0.071 4	0.999 6	0.106 8	0.999 8	0.091 1
香兰素	0.998 5	0.072 5	0.999 0	0.075 4	0.998 1	0.064 4
淀粉	0.999 4	0.248 2	0.999 4	0.246 6	0.999 4	0.294 2

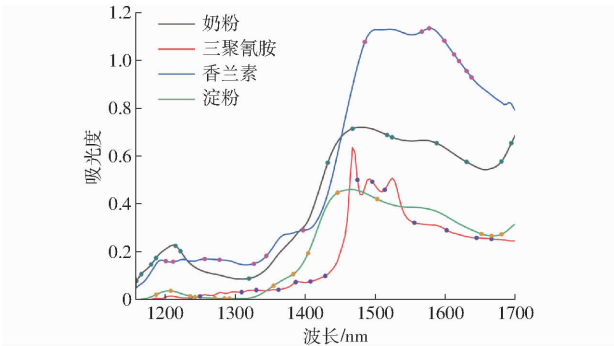


图 4 CNN 模型自定义层挑选的特征波长在平均光谱图上的分布
Fig.4 Distribution of characteristic wavelengths selected by self-defined layer in CNN model on average spectra

对比图 3 和图 4 可得 CNN 模型对应筛选出的 15 个波长点的分布较 Laplacian score 和 RelieFF 方法更分散。对于奶粉预测模型,Laplacian score、RelieFF 和自定义层均选了 1 524 nm 作为特征变量。对于三聚氰胺预测模型,Laplacian score 和自定义层

均选了 1 474 nm,文献[28]表明该波长与 N—H 键的第二泛音有关。对于香兰素预测模型,Laplacian score 和自定义层均选了 1 210 nm(C—H 振动)^[29]。对于淀粉预测模型,Laplacian score 和自定义层均选了 1 680 nm,ReliefF 和自定义层均选了 1 446 nm,文献[30]在马铃薯淀粉的掺假检测研究中同样筛选了 1 446 nm 为特征波长。

2.4 可视化

因 ReliefF-PLS 模型建模所需变量数少,且效果较优,故将包含空间信息像素点的光谱输入 ReliefF-PLS 模型,对其预测结果进行可视化。具体地,从 SNV 预处理后的子图像中随机选取 150 像素×150 像素尺寸的 ROI,提取 ROI 中每个像素点的光谱数据。以每个像素点作为一个样本点,构建尺寸为 22 500×155 的二维数据矩阵。提取 ReliefF 筛选的特征波长下数据输入构建好的相应多光谱 PLS 模型,得到每个像素点对应的物质

预测质量分数一维向量结果(22 500×1)。将该向量重构为 150×150 矩阵数据,并制图,得到不同物质质量分数的可视化结果,如图 5 所示。图 5 中不同颜色变化表示模型对相应物质预测质量分数变化。图 5a 显示模型对质量分数 62%、79%、85% 和 94.5% 奶粉的可视化图由蓝到浅绿到黄绿变化明显,质量分数 62% 奶粉结果图中蓝色像素点多,表明预测质量分数主要集中在大约 65% 附近,质量分数 94.5% 奶粉结果图中主要为黄绿色,表明预测质量分数结果集中在大约 95% 附近。图 5b 显示掺有质量分数 0、5% 和 10% 三聚氰胺的样本的可视化图由黄到红变化明显,图 5c 显示掺有质量分数 0、0.5%、1% 和 3% 香兰素的样本的可视化图由浅蓝到绿到黄变化明显,图 5d 显示掺有质量分数 0、5%、10%、15% 和 25% 淀粉样本的可视化图由紫到红再到黄变化明显。以上表明模型具有良好的预测性能。

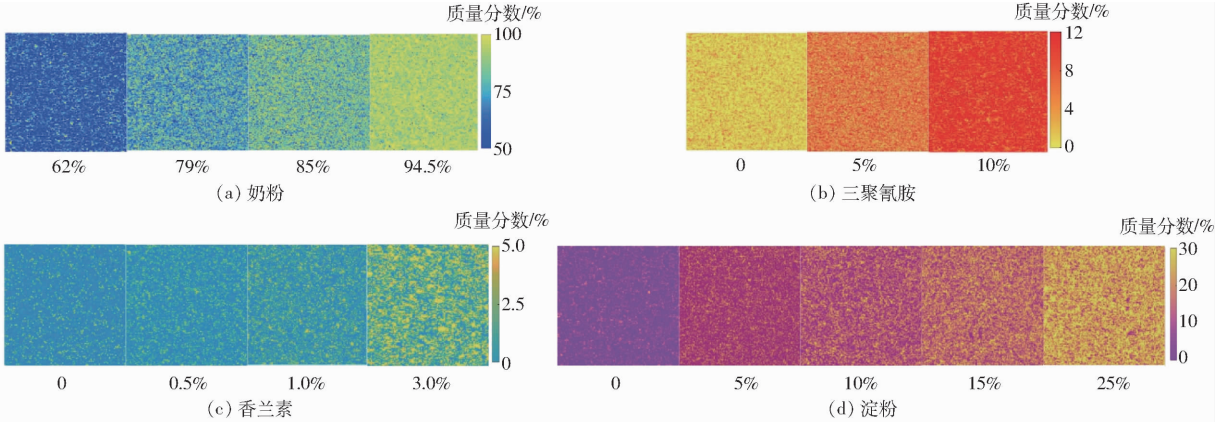


图 5 基于 ReliefF-PLS 模型的 4 种物质含量预测的可视化图

Fig. 5 Visualization of predicted contents of four substances based on ReliefF-PLS models

2.5 模型验证

对采集的外部验证集和新外部验证集的高光谱图像数据应用相同的预处理,然后提取 ROI 平均光谱数据,获得外部验证集和新外部验证集各 45 条光谱(9 条×5 种)。提取 ReliefF 筛选的特征波长下的数据输入之前构建好的多光谱 PLS 模型。模型对预测集、外部验证集和新外部验证集的预测结果如表 6 所示。

尽管相比预测集,模型对外部验证集中的奶粉和淀粉的预测精度有所降低(R^2 由原来的 0.999 7 和 0.999 6 降为 0.937 4 和 0.654 1;均方根误差由原来的 0.199 8% 和 0.182 3% 增长为 4.380 8% 和 6.059 0%),但奶粉仍获得较理想的预测结果,淀粉的预测表现一般。精度降低可能是由外部验证集中奶粉、淀粉与建模样本集中奶粉、淀粉产品的批次等因素引起的样本差异导致的。模型对外部验证集中

表 6 ReliefF-PLS 多光谱模型对不同数据集的预测结果

物质	预测集		外部验证集		新外部验证集	
	R_p^2	均方根 误差/%	R_w^2	均方根 误差/%	R_{nw}^2	均方根 误差/%
奶粉	0.999 7	0.199 8	0.937 4	4.380 8	0.795 3	6.981 6
三聚氰胺	0.999 7	0.054 0	0.271 2	3.763 0	0.075 7	4.625 8
香兰素	0.998 2	0.045 5	0.220 1	3.942 8	0.268 8	4.012 1
淀粉	0.999 6	0.182 3	0.654 1	6.059 0	0.795 2	6.076 8

的三聚氰胺和香兰素预测结果差, R^2 只有 0.2 左右。原因可能是两者的掺假质量分数较低,批次等样品差异严重干扰了不同质量分数导致的光谱变化信息,从而影响了模型预测效果。

相比预测集,模型对新外部验证集中的奶粉和淀粉的预测精度降低,对三聚氰胺和香兰素的预测结果差。造成以上结果的原因除了不同批次因素

外,还包括不同奶粉品牌之间的成分含量差异引起的光谱数据的干扰变动。然而,批次和品牌因素的综合影响是复杂的,体现在奶粉和三聚氰胺的 R_{xw}^2 明显小于 R_w^2 ,而香兰素的 R_{xw}^2 与 R_w^2 的结果相近,淀粉的 R_{xw}^2 却优于 R_w^2 。尽管如此,奶粉、三聚氰胺、香兰素和淀粉的新外部验证集均方根误差均大于外部验证集的均方根误差,这也说明不能单独依据 R^2 评估模型的预测能力。模型对新外部验证集中奶粉和淀粉的预测结果表明,模型具有预测新品牌底物掺假样本的潜力,但预测精度仍需采用校正方法或其他措施进一步提高。

3 结束语

通过高光谱成像对婴幼儿奶粉中的多种掺假物(三聚氰胺、香兰素和淀粉)进行了定量检测分析。对预处理后的高光谱图像提取 ROI 及其平均光谱,然后在全波段建立 PLS 模型。采用 Laplacian score(无监督)和 ReliefF(有监督)两种特征波长筛选方

法,并结合 PLS 算法建立多光谱预测模型。基于全波长数据训练包含自定义波长选择层的一维卷积神经网络模型。对 Laplacian score、ReliefF 和自定义选择层挑选的前 15 个重要波长进行对比分析。结果表明,与 1DCNN 模型相比,全波长 PLS 模型效果更佳,其预测集中奶粉、三聚氰胺、香兰素和淀粉质量分数预测结果的均方根误差分别为 0.173 6%、0.047 0%、0.031 4% 和 0.133 3%;在基于特征波长的多光谱模型中,ReliefF-PLS 表现出更好的预测效果,其预测集中奶粉、三聚氰胺、香兰素和淀粉质量分数预测结果的均方根误差分别为 0.199 8%、0.054 0%、0.045 5% 和 0.182 3%。最后基于 ReliefF-PLS 模型预测结果进行可视化分析,同一物质不同浓度的可视化图颜色变化明显。采用两种外部验证集对模型进行了评估,模型对具有大质量分数的奶粉和淀粉预测结果较好。研究表明,高光谱成像技术可实现对于奶粉中同时掺入三聚氰胺、香兰素和淀粉 3 种掺假物含量的快速无损检测。

参 考 文 献

[1] NAGRAIK R, SHARMA A, KUMAR D, et al. Milk adulterant detection: conventional and biosensor based approaches: a review[J]. Sensing and Bio-Sensing Research,2021,33:100433.

[2] 季美泉,费晓庆,丁涛,等. 高效液相色谱法鉴别蜂蜜中掺入蜂王浆来源的蛋白质[J]. 食品安全质量检测学报,2020,11(1):153-157.

JI Meiquan, FEI Xiaoqing, DING Tao, et al. Identification of proteins from honey adulterated with royal jelly by high performance liquid chromatography[J]. Journal of Food Safety & Quality, 2020,11(1): 153-157. (in Chinese)

[3] 孙超仁,王凤玲,王玉玮,等. 基于特征脂肪酸变化对花生油掺伪快速鉴别方法研究[J]. 食品与发酵工业,2023,49(3):296-300.

SUN Chaoren, WANG Fengling, WANG Yuwei, et al. Rapid identification of adulteration in peanut oil based on the changes of characteristic fatty acid[J]. Food and Fermentation Industries, 2023,49(3):296-300. (in Chinese)

[4] 侯颖烨,王志元,谢建军,等. 元素分析-稳定同位素质谱法结合化学计量学鉴别橄榄油掺假[J]. 中国油脂,2023,48(6):73-78.

HOU Yingye, WANG Zhiyuan, XIE Jianjun, et al. Olive oil adulteration identification using elemental analysis-stable isotope ratio mass spectrometer coupled with chemometrics[J]. China Oils and Fats, 2023,48(6):73-78. (in Chinese)

[5] TRIMBOLI F, COSTANZO N, LOPREIATO V, et al. Detection of buffalo milk adulteration with cow milk by capillary electrophoresis analysis[J]. Journal of Dairy Science, 2019, 102(7): 5962-5970.

[6] 胡珂青,刘欣悦,白金泽,等. 基于同步荧光光谱的杜仲籽油掺假检测研究[J]. 中国油脂,2019,44(7):103-106.

HU Keqing, LIU Xinyue, BAI Jinze, et al. Adulteration detection of *Eucommia ulmoides* seed oil using synchronous fluorescence spectroscopy[J]. China Oils and Fats, 2019,44(7): 103-106. (in Chinese)

[7] 赵茂程,吴泽本,汪希伟,等. 基于光谱成像的猪肉新鲜度空间分布预测评价方法[J]. 农业机械学报,2022,53(3):412-422.

ZHAO Maocheng, WU Zeben, WANG Xiwei, et al. Evaluation of spectral imaging-based spatial predictions of freshness spatial distribution over pork[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022,53(3): 412-422. (in Chinese)

[8] 曹益飞,徐焕良,吴玉强,等. 基于时序高光谱和多任务学习的水稻病害早期预测研究[J]. 农业机械学报,2022,53(11):288-298.

CAO Yifei, XU Huanliang, WU Yuqiang, et al. Early forecasting of rice diseases based on time series hyperspectral imaging and multi-task learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022,53(11): 288-298. (in Chinese)

[9] 王春雷,陈婉芝,卢彩云,等. 基于高光谱成像的玉米收获后根茬行分割方法[J]. 农业机械学报,2020,51(增刊2):421-426.

WANG Chunlei, CHEN Wanzhi, LU Caiyun, et al. Segmentation method of maize stubble row based on hyperspectral imaging[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020,51(Supp.2):421-426. (in Chinese)

[10] GUO L B, YU Y X, YU H Y, et al. Rapid quantitative analysis of adulterated rice with partial least squares regression using hyperspectral imaging system[J]. Journal of the Science of Food and Agriculture, 2019, 99(12): 5558-5564.

[11] 王伟,赵昕,褚璇,等. 基于可见/近红外高光谱的八角茴香与莽草无损鉴别[J]. 农业机械学报,2019,50(11):373-379.

- WANG Wei, ZHAO Xin, CHU Xuan, et al. Nondestructive identification of star anise and shikimmi by visible/near infrared hyperspectral images[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(11): 373 – 379. (in Chinese)
- [12] LIM J, KIM G, MO C, et al. Detection of melamine in milk powders using near-infrared hyperspectral imaging combined with regression coefficient of partial least square regression model[J]. Talanta, 2016, 151: 183 – 191.
- [13] FORCHETTI D A P, POPPI R J. Use of NIR hyperspectral imaging and multivariate curve resolution (MCR) for detection and quantification of adulterants in milk powder[J]. LWT—Food Science and Technology, 2017, 76: 337 – 343.
- [14] LIU Y S, ZHOU S B, HAN W, et al. Detection of adulteration in infant formula based on ensemble convolutional neural network and near-infrared spectroscopy[J]. Foods, 2021, 10(4): 785.
- [15] ZHOU L, TAN L H, ZHANG C, et al. A portable NIR-system for mixture powdery food analysis using deep learning[J]. LWT, 2022, 153: 112456.
- [16] ZHAO X, LI C H, ZHAO Z L, et al. Generic models for rapid detection of vanillin and melamine adulterated in infant formulas from diverse brands based on near-infrared hyperspectral imaging[J]. Infrared Physics & Technology, 2021, 116: 103745.
- [17] LENG T, LI F, CHEN Y, et al. Fast quantification of total volatile basic nitrogen (TVB-N) content in beef and pork by near-infrared spectroscopy: comparison of SVR and PLS model[J]. Meat Science, 2021, 180: 108559.
- [18] ROFFO G, MELZI S, CASTELLANI U, et al. Infinite latent feature selection: a probabilistic latent graph-based ranking approach[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 1398 – 1406.
- [19] HE X F, CAI D, NIYOGI P. Laplacian score for feature selection[C]//WEISS Y, SCHÖLKOPF B, PLATT J. Advances in Neural Information Processing Systems 18: Proceedings of the 2005 Conference. Nineteenth Annual Conference on Neural Information Processing Systems (NIPS 2005). Cambridge: MIT Press, 2005: 507 – 514.
- [20] 项颂阳, 许章华, 张艺伟, 等. 高光谱图像分类的 ReliefF – RFE 特征选择算法构建与应用[J]. 光谱学与光谱分析, 2022, 42(10): 3283 – 3290.
- XIANG Songyang, XU Zhanghua, ZHANG Yiwei, et al. Construction and application of ReliefF – RFE feature selection algorithm for hyperspectral image classification[J]. Spectroscopy and Spectral Analysis, 2022, 42(10): 3283 – 3290. (in Chinese)
- [21] KANG R X, WANG X, ZHAO M, et al. A comparison of benchtop and micro NIR spectrometers for infant milk formula powder storage time discrimination and particle size prediction using chemometrics and denoising methods[J]. Journal of Food Engineering, 2022, 329: 111087.
- [22] FU X P, KIM M S, CHAO K L, et al. Detection of melamine in milk powders based on NIR hyperspectral imaging and spectral similarity analyses[J]. Journal of Food Engineering, 2014, 124: 97 – 104.
- [23] MAUSER L J, CHERNYSHOVA A A, HIATT A, et al. Melamine detection in infant formula powder using near- and mid-infrared spectroscopy[J]. Journal of Agricultural and Food Chemistry, 2009, 57(10): 3974 – 3980.
- [24] HOFFMAN L, INGLE P, KROLE A H, et al. Discrimination of lamb (*Ovis aries*), emu (*Dromaius novaehollandiae*), camel (*Camelus dromedarius*) and beef (*Bos taurus*) binary mixtures using a portable near infrared instrument combined with chemometrics[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2023, 294: 122506.
- [25] BABAE B, KHANMOHAMMADI M, GARMARUDI A B, et al. Effect of peeling and point of spectral recording on sucrose determination in sugar beet root using near infrared spectroscopy[J]. Infrared Physics & Technology, 2019, 103: 103065.
- [26] DELANEY J K, CONOVER D M, DOOLEY K A, et al. Integrated X-ray fluorescence and diffuse visible-to-near-infrared reflectance scanner for standoff elemental and molecular spectroscopic imaging of paints and works on paper[J]. Heritage Science, 2018, 6: 31.
- [27] ZHAO Y R, YU K Q, LI X, et al. Detection of fungus infection on petals of rapeseed (*Brassica napus* L.) using NIR hyperspectral imaging[J]. Scientific Reports, 2016, 6(1): 38878.
- [28] LABORDE A, PUIG-CASTELLVÍ F, BOUVERESSE D J R, et al. Detection of chocolate powder adulteration with peanut using near-infrared hyperspectral imaging and multivariate curve resolution[J]. Food Control, 2021, 119: 107454.
- [29] SANTOS I A, CONCEIÇÃO D G, VIANA M B, et al. NIR and MIR spectroscopy for quick detection of the adulteration of cocoa content in chocolates[J]. Food Chemistry, 2021, 349: 129095.
- [30] ZHAO X, WANG W, CHU X, et al. Rapid and nondestructive quantification of cassava starch adulterants in potato starch by using hyperspectral imaging[C]//2018 ASABE Annual International Meeting. American Society of Agricultural and Biological Engineers, 2018.