

文章编号:1004-5929(2025)01-0047-07

## 基于偏最小二乘回归法和原位拉曼光谱技术在线 检测奶粉成分的研究

王利祥<sup>1\*</sup>, 郭向伟<sup>2</sup>, 郎沁争<sup>3</sup>, 赵诵<sup>4</sup>

1. 河南护理职业学院, 公共学科部, 河南 安阳, 455000;
2. 河南理工大学, 电气工程与自动化学院, 河南 焦作, 454000;
3. 焦作工贸职业学院, 信息工程系, 河南 焦作, 454000;
4. 河南农业大学, 信息与管理科学学院, 河南 郑州, 450003)

**摘要:** 奶粉成分的在线非接触式原位光谱探测技术对乳制品生产企业非常有益, 这对于加工商实时评估生产产品的品质至关重要。拉曼光谱是一项极具商业价值的原位光谱探测技术, 本文分析了原位拉曼光谱法检测奶粉主成分(脂肪、蛋白质和乳糖等)的可行性。针对性开展了不同混合比例的脱脂奶粉和全脂奶粉的拉曼光谱实验, 研究了拉曼光谱特征谱峰与脂肪含量的统计学规律。建立了基于偏最小二乘回归法的预测模型, 并基于回归模型预测了脂肪含量, 回归模型的决定系数大于 0.9886, 均方根误差小于 0.6248。总之, 基于原位拉曼光谱技术对在线评估奶粉品质具有潜在的商业应用价值。

**关键词:** 拉曼光谱; 奶粉; 偏最小二乘回归法; 定量分析

**中图分类号:** O657.3; O433 **文献标志码:** A **doi:** 10.13883/j.issn1004-5929.202501007

## Research on online detection of milk powder components based on partial least squares regression and in-situ Raman spectroscopy technology

WANG Lixiang<sup>1\*</sup>, GUO Xiangwei<sup>2</sup>, LANG Qinzheng<sup>3</sup>, Zhao Song<sup>4</sup>

1. Henan Vocational College Of Nursing, Anyang 455000, China;
2. Henan Polytechnic University School of Electrical Engineer and Automation, Jiaozuo 454000, China;
3. Jiaozuo College of Industry and Trade, Jiaozuo 454000, China;
4. College of Information and Management Science, Henan Agricultural University, Zhengzhou 450002, China)

**Abstract:** The online non-contact in-situ spectral detection technology of milk powder composition is very beneficial to dairy production enterprises, which is very important for processors to evaluate the quality of products in real-time. Raman spectroscopy is a very commercial potential in-situ spectral detection technology. This paper analyzes the feasibility of in-situ Raman spectroscopy in detecting the main components of milk powder (fat, protein, lactose, etc.). The Raman spectrum experiments of skimmed milk powder and whole milk powder with different mixing ratios were carried out, and the statistical laws of Raman spectrum characteristic peaks and fat content were studied. The prediction model based on the partial least squares regression method was established, and the fat content was predicted based on the regression model. The determination coefficient of the regression model was more significant than 0.9886, and the root mean square error was less than 0.6248. In con-

收稿日期: 2024-04-11, 修改日期: 2024-05-20

基金项目: 国家自然科学基金(62003133)

通讯作者: 王利祥(1987-), 男, 硕士, 讲师, 主要从事计算机网络安全以及机器人方向。E-mail: wanglx87@126.com

clusion, in situ, Raman spectroscopy has potential commercial application value for the on-line evaluation of milk powder quality.

**Keywords:** Raman spectroscopy; Milk powder; Partial Least squares regression; Quantitative analysis

## 0 引言

由于人民生活质量的不断提高和饮食习惯的改变,大众对乳制品的需要也迅猛增长,这导致生产厂商不断提高产品生产效率,而乳制品成分的实时分析技术是保障产品不同批次一致性和品质控制监管的重要途径。

光谱技术是现代光学计量中的重要分支,通过对物质散射、吸收、反射等光谱的分析可以反演物质的化学组成、含量等信息,且光谱技术具有非接触、多组分、高灵敏及实时连续探测等优势,其中光散射的经典技术—拉曼光谱是目前科学家应用最广的、最具商业潜力的光散射技术之一。在乳制品行业中,拉曼光谱已被用于奶制品的检测<sup>[1]</sup>、表征判别<sup>[2]</sup>、掺杂物分类<sup>[3]</sup>等。由于奶源产地、加工工艺等因素会引起生牛奶成分的差异,这会给生产厂商的产品加工和品质控制监管带来挑战。近些年发表的研究成果中报道了几项使用拉曼光谱定性定量分析牛奶成分的研究。陈珂等<sup>[4]</sup>报道了一种基于目标物诱导的等离子激元耦合以及表面增强拉曼光谱、利用未修饰的金纳米颗粒检测牛奶中三聚氰胺的纳米生物传感方法,实现牛奶中三聚氰胺 8 nmol/L 的检测限。王琳等<sup>[5]</sup>提出一种现场快检的表面增强拉曼光谱检测三聚氰胺的分析方法,可在 5 分钟内实现牛奶中三聚氰胺 1  $\mu\text{g}/\text{mL}$  的检测。王海燕等<sup>[6]</sup>利用拉曼光谱结合模式识别算法开展了牛初乳奶片和乳酸奶片的判别研究,建立了优化的智能判别模型。Mendes 等<sup>[7]</sup>研究了牛奶的拉曼光谱和红外光谱,并结合气相色谱技术进一步分析了光谱特征信息代表的化学信息,建立了 PLSR 回归模型,其中模型校准的决定系数为 0.97,均方根误差为 0.44。Li 等<sup>[8]</sup>开发了一种单变量回归模型,预测了牛奶中乳糖的含量,检测限达到 0.019 mol/L,回归模型决定系数为 0.9992。Huang 等<sup>[9]</sup>使用表明增强拉曼光谱和蛋白质对磷钼酸的影响量化牛奶中的蛋白,报道的检测限为 1.5  $\mu\text{g}/\text{mL}$ 。Mazurek 等<sup>[10]</sup>采用傅里叶变换拉曼光谱和红外光谱分析了牛奶样品的脂肪、蛋白质等物质,并建立了 PLSR 回归模型,发现红外光谱结果优于拉曼光谱。

以上基于拉曼光谱技术开展的乳制品定性定量分析虽取得了较大进展,但所使用的拉曼光谱具体实施方案很难实现在线实时分析。实时分析技术需结合实验室光谱分析和在线检测的模拟,因此本文创新性的提出一种基于便携式原位拉曼光谱技术结合偏最小二乘回归法在线实时定量分析奶粉所含蛋白质、脂肪和碳水化合物的方法,使用便携式近红外拉曼光谱技术详细分析了脱脂奶粉和全脂奶粉的光谱特性,并建立三种营养物质的 PLSR 回归模型,验证该方案对于奶粉在线检测的应用潜力。

## 1 实验与算法

### 1.1 样品制备

脱脂奶粉和全脂奶粉的品牌均为河套,生产厂名为内蒙古金河套乳业有限公司,两种奶粉均不含植脂末、防腐剂、香精和色素等添加物。使用英衡电子天平秤(量程 500 g,精度 0.001 g)配比不同质量分数的混合奶粉,其中全脂奶粉的质量分数分别为 0.00、0.05、0.10、0.15、0.20、0.25、0.30、0.35、0.40、0.45、0.50、0.55、0.60、0.65、0.70、0.75、0.80、0.85、0.90、0.95 和 1.00。为避免人为引入过多的误差,每种混合比例需单独制备样品且每个浓度各单独制备 15 个样品。

### 1.2 仪器设备

实验仪器为上海如海光电科技有限公司 830 拉曼光谱仪,其中激励源为稳谱激光器(Laser830-0.08-500-FC-3A),中心波长为(830 $\pm$ 0.5) nm,谱线宽度小于 0.08 nm,激光功率可调(0~500 mW);光谱采集模块为光纤光谱仪(XS11639-790-1050-25-1A),光谱采集范围为 260~2950  $\text{cm}^{-1}$ ,信噪比为 380:1,光学分辨率为 0.4 nm。激励和接收组件为 830 拉曼探头(RPB830-1.5-FS),探头耦合效率 $\geq 80\%$ ,工作距离 7.5 mm。拉曼光谱实验中聚焦在样品表面的激光功率为 50 mW,单条光谱的积分时间为 500 ms,单个样品随机探测 10 条拉曼光谱,即每个样品有 10 个光谱数据。仪器及工作原理如图 1 所示。

### 1.3 数据处理算法

#### 1.3.1 光谱校正

在进行拉曼光谱分析时,需要对原始光谱数

表 1 混合奶粉成分表

Table 1 Composition of mixed milk powder

编号	全脂奶粉质量分数	蛋白质质量分数( $\times 10^{-2}$ )	脂肪质量分数( $\times 10^{-2}$ )	碳水化合物质量分数( $\times 10^{-2}$ )
1	0.00	32.60	1.30	54.50
2	0.05	32.16	2.59	54.03
3	0.10	31.72	3.87	53.55
4	0.15	31.28	5.16	53.08
5	0.20	30.84	6.44	52.60
6	0.25	30.40	7.73	52.13
7	0.30	29.96	9.01	51.65
8	0.35	29.52	10.30	51.18
9	0.40	29.08	11.58	50.70
10	0.45	28.64	12.87	50.23
11	0.50	28.20	14.15	49.75
12	0.55	27.76	15.44	49.28
13	0.60	27.32	16.72	48.80
14	0.65	26.88	18.01	48.33
15	0.70	26.44	19.29	47.85
16	0.75	26.00	20.58	47.38
17	0.80	25.56	21.86	46.90
18	0.85	25.12	23.15	46.43
19	0.90	24.68	24.43	45.95
20	0.95	24.24	25.72	45.48
21	1.00	23.80	27.00	45.00



图 1 实验系统光路结构示意图

Fig. 1 Schematic diagram of the optical path structure of the experimental system

据进行基线去除和降噪处理,从而获得准确的振动光谱信息。基线去除包括消除背景信号和杂散光的过程,背景信号可能源于环境光、样本自身的非拉曼散射等。去除基线可以提高拉曼光谱谱峰的对比度,便于后期特征振动峰的识别和分析。降噪处理是降低拉曼光谱信号中噪声的过程,噪声可能源于环境干扰、光学系统和探测器等因素,噪声直接干扰了拉曼光谱谱峰的强度和谱形,降低噪声干扰是提高拉曼光谱信噪比的重要举措。总之,基线去除和降噪处理是拉曼光谱校正必不可少的预处理过程,光谱校正可以提高拉曼光谱谱峰质量、增强特征峰对比度及定性定量分析能力,保障拉曼光谱分析的可靠性和准确性。本文

的光谱校正采用小波算法实现基线去除和降噪处理<sup>[11]</sup>,通过小波分解提取原始信号中高、低频信息,通过高、低通滤波器完成原始信号的重构。

### 1.3.2 回归模型

偏最小二乘回归法(Partial least squares regression, PLSR)是一种经典的监督学习方法,在线性多元统计方法中非常具有实用性,在现代光谱分析中的化学计量学方法中具有重要的应用价值,擅长处理高维变量和多重共线性问题。PLSR可最大限度的提高待分析物质浓度与光谱信息的线性关系,减少不确定因素的干扰<sup>[12]</sup>。PLSR的目标是找到估计函数的参数,并确保总平方误差最小<sup>[13]</sup>。本文使用的PLSR模型是基于Matlab R2019b数学软件开发的。定量模型预测性能采用决定系数(Coefficient of determination,  $R^2$ )和均方根误差(Root mean square error, RMSR)两个核心参数作为评价指标<sup>[14]</sup>。

## 2 结果与讨论

### 2.1 奶粉的典型拉曼光谱

图2中蓝色曲线为全脂奶粉的原始光谱数据,生物样本中包含了较大的基线(荧光等),全脂奶粉的拉曼光谱中同样出现明显的荧光信号,因此必须抑制荧光背景对拉曼光谱的影响<sup>[15]</sup>。图2中红色曲线为通过小波算法重建的全脂奶粉原始光谱数据的基线,黑色曲线为全脂奶粉去基线和降噪后的拉曼光谱。

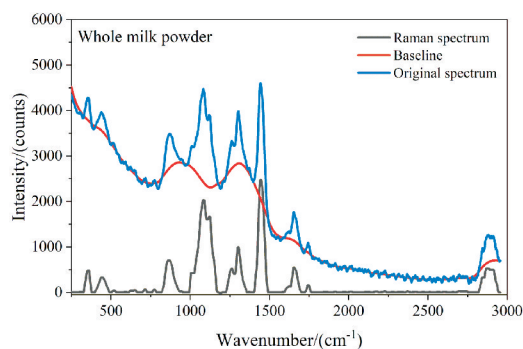


图 2 全脂奶粉的原始拉曼光谱机器校正后的谱图。(黑色的标注说明:校正后的拉曼谱)

Fig. 2 The original Raman spectra of whole milk powder were corrected. (black curve: corrected Raman spectrum)

利用光谱技术结合化学计量学方法开展定性定量分析时,必须要对数据进行归一化处理,数据归一化的过程可以消除激光能量抖动和研究人

人为操作引入的光谱整体强度的浮动<sup>[16]</sup>。图 3 为 21 种混合奶粉的归一化拉曼光谱,通过对比可以发现不同混合浓度奶粉的拉曼光谱的谱峰差异主要集中在 354、448、867、1004、1080、1121、1303、

1448、1659、1743 及 2815~2947  $\text{cm}^{-1}$ 附近,这些谱峰强度的差异反映了不同奶粉成分含量的差异,因此有必要对拉曼光谱的谱峰进行振动的归属判定,从而明确各个特征谱峰代表的成分信息。

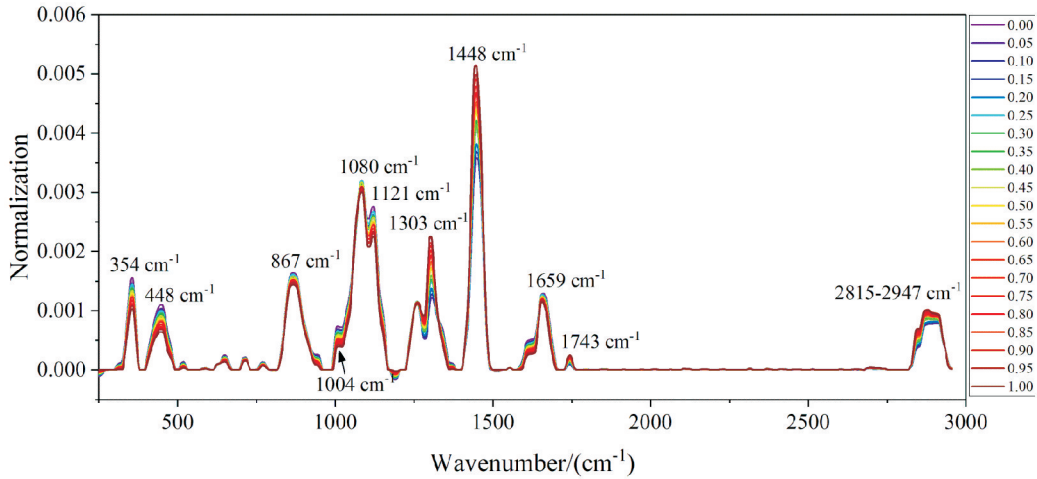


图 3 21 种混合奶粉的归一化拉曼光谱

Fig. 3 Normalized Raman spectra of 21 kinds of mixed milk powder

图 4 为对 21 种混合奶粉拉曼光谱谱峰振动模式归属,结果表明脱脂奶粉和全脂奶粉的拉曼光谱差异归因于脂肪和牛奶蛋白,脂肪以  $\nu(\text{C-H})$  和  $\nu(\text{C=O})$  模式表示,1004  $\text{cm}^{-1}$  散射峰归属于苯丙氨酸环,与蛋白质的含量有关,全脂奶粉中  $\delta$

$(\text{CH}_2)_{\text{twisting}}$  模式相对较强。总之,奶粉中蛋白质和脂肪分子的拉曼光谱谱峰的特征十分明显,这些拉曼光谱差异为定性定量分析提供了可能性。表 2 为对典型拉曼光谱谱峰振动的归属分类。表 2 总结了奶粉拉曼光谱主要特征峰的分子振动模式。

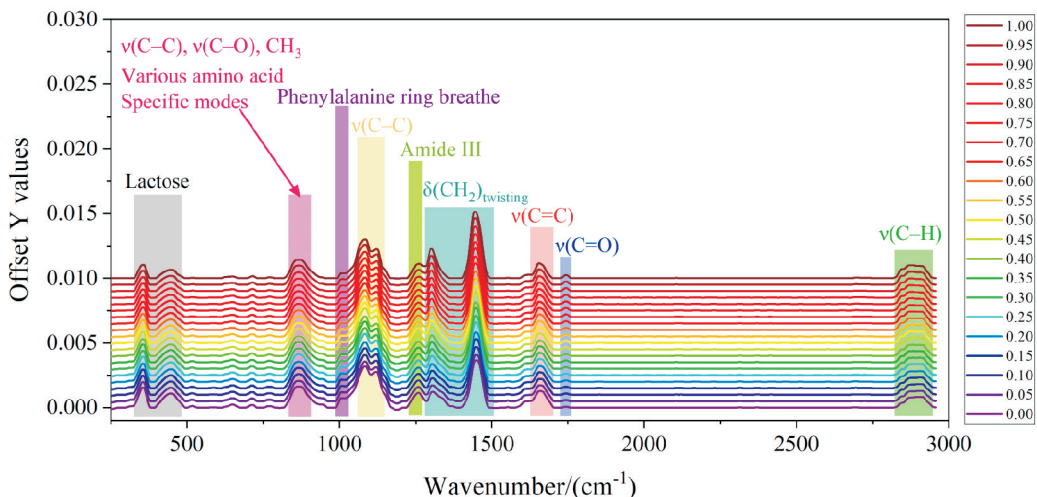


图 4 不同混合比例的脱脂奶粉和全脂奶粉的拉曼光谱

Fig. 4 Raman spectra of skimmed milk powder and whole milk powder with different mixing ratios

## 2.2 奶粉成分的定量回归模型

PLSR 是一种基于主成分分析的多元线性回归方法,因此主成分的因子数直接影响了定量模

型的预测性能。将 3150 条光谱构成的所有数据集按照 3 : 1 : 1 的比例划分训练集和测试集,其中采用“留一法”进行模型的交叉验证,在进行定量分

表 2 奶粉拉曼光谱的谱带归属

Table 2 Band assignments of milk powder Raman spectra

拉曼位移 Raman shift/( $\text{cm}^{-1}$ )	归属 Assignments	参考文献 References
354	Lactose	[17][18]
448	Lactose	[17][18]
867	$\nu(\text{C}-\text{C})$ , $\nu(\text{C}-\text{O})$ , $\text{CH}_3$ , rock, various amino acid specific modes	[17][18]
1004	Phenylalanine ring breathe	[17][18]
1080	$\nu(\text{C}-\text{C})$	[17][18]
1118	$\nu(\text{C}-\text{C})$	[17][18]
1121	$\nu(\text{C}-\text{C})$	[17][18]
1260	Amide III	[17][18]
1303	$\delta(\text{CH}_2)$ twisting	[17][18]
1448	$\delta(\text{CH}_2)$ twisting	[17][18]
1659	$\nu(\text{C}=\text{C})$	[17][18]
1743	$\nu(\text{C}=\text{O})$	[17][18]
2815~2947	$\nu(\text{C}-\text{H})$	[17][18]

析时,采用全光谱作为定量模型的输入。影响 PLSR 回归模型预测效果的核心参数是主成分因子数,本文计算了前 10 个主成分因子的方差贡献,如图 5 所示。当主成分因子数在达到 10 个时,PLSR 模型的累计方差贡献已超过 98%。

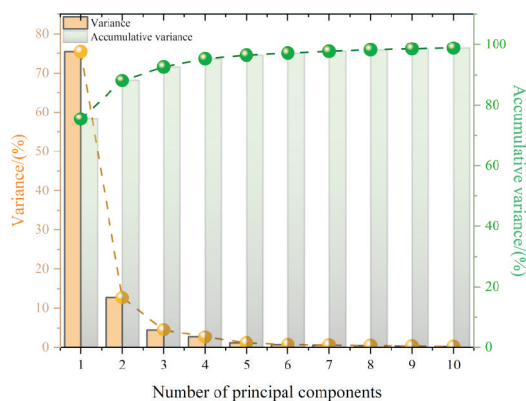


图 5 主成分因子的累计方差贡献

Fig. 5 Accumulative variance contribution of principal component factors

综合以上分析,使用前 10 个主成分因子构建 PLSR 定量回归模型,本文对比了前 10 个主成分因子对 PLSR 定量回归模型训练集的影响,并采用 RMSE 作为评价指标,如图 6 所示。结果表明当主成分因子达到 10 时,PLSR 定量回归模型预测效果达到最优。

PLSR 定量回归模型预测效果如图 7 所示。采用  $R^2$  和 RMSE 评价 PLSR 定量回归模型的性能,其中训练集、验证集和测试集的  $R^2$  分别为

0.9942、0.9912 以及 0.9886, RMSE 分别为 0.5924、0.5981 及 0.6248。

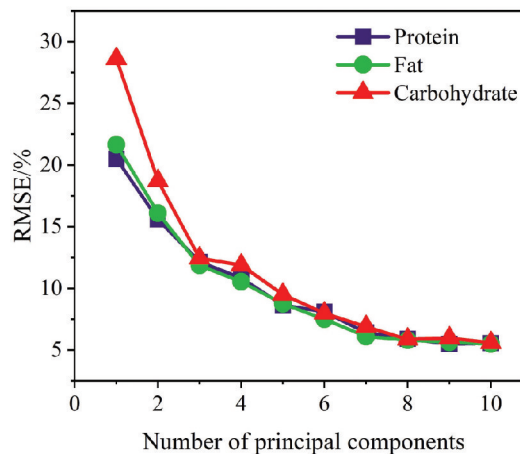


图 6 训练集均方根误差随主成分因子变化

Fig. 6 Variation of RMSE of training set with principal component factors

就 PLSR 模型的预测效果而言,  $R^2$  和 RMSE 均达到了较为理想的效果。真实值和预测值的差别主要由三方面因素造成,一是人工配比的混合模拟样本存在人为不可控的误差引入,即模拟样本本身在配比的过程中就不可避免的存在一定的精准度;二是便携式光谱仪器在面向现场原位快检的需求,其光谱稳定性也是干扰模型预测的重要因素;三是 PLSR 模型自身存在的缺陷,例如预测能力受限、模型复杂性高、多重共线性问题和数据解释性差等。

### 3 结论

乳制品加工过程中不同批次产品的品质实时监控是目前生产企业最为关注的技术难点,本文提出的一种基于原位拉曼光谱结合 PLSR 算法实现了奶粉蛋白质、脂肪和碳水化合物等主要成分的定量分析,研究结果表明该方案对于增强乳制品生产过程的实时控制具有重要的实际应用价值。原位拉曼光谱是一种极具商业潜力的无损、非接触式、高效的物质成分检测技术,非常适用于监测奶粉生产过程中各种营养成分的动态变化,展示了其作为在线检测技术的工业潜力。总之,本研究工作证实了原位拉曼光谱可有效应用于量化全脂奶粉和脱脂奶粉的成分,其具备了在线检测蛋白质、脂肪和碳水化合物浓度的工业应用潜力。

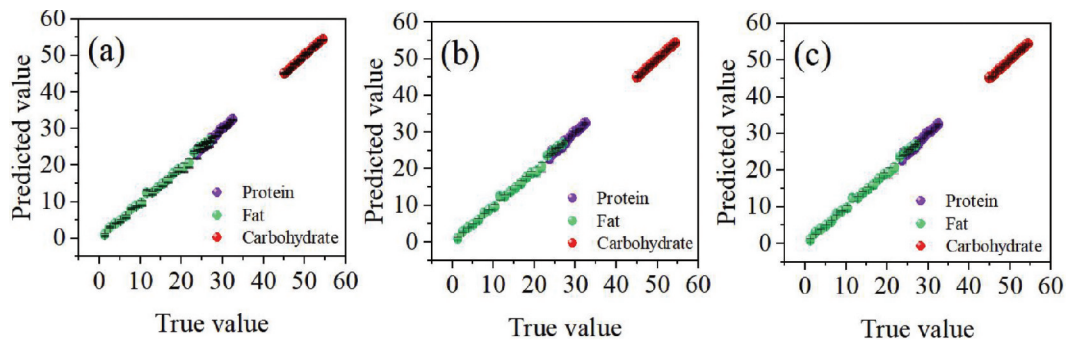


图 7 PLSR 定量回归模型预测效果。(a) 训练集预测效果,(b) 验证集预测效果,(c) 测试集预测效果

Fig. 7 Prediction effect of PLSR quantitative regression model. (a) The prediction effect of the training set, (b) the prediction effect of the verification set, and (c) the prediction effect of the test set

## 参考文献

- [1] 罗炳嘉,吴文超,徐祥,等.基于拉曼光谱和 k-近邻算法的乳酪制品识别优化研究[J].光谱学与光谱分析,2023,43(S1):147-148. (Luo B J, Wu W C, Xu X, et al. Optimization of cheese product identification based on Raman spectroscopy and k-nearest neighbor algorithm[J]. Spectroscopy and Spectral Analysis, 2023, 43(S1):147-148.)
- [2] 顾颖,赵姝彤,熊蓝萍,等.基于多维拉曼光谱的乳粉表征及判别分析[J].粮食科技与经济,2022,47(3):97-101. (Gu Y, Zhao S T, Xiong L P, et al. Characterization and discriminant analysis of milk powder based on multidimensional Raman spectroscopy[J]. Grain Science and Economy, 2022, 47(3): 97-101.)
- [3] 邵帅斌,刘美含,石宇晴,等.基于卷积神经网络的乳粉掺杂物拉曼光谱分类方法[J].食品科学,2022,43(14):296-301. (Shao S B, Liu M H, Shi Y Q, et al. Raman spectroscopic classification of adulterants in milk powder samples using convolutional neural network[J]. Food Science, 2022, 43(14): 296-301.)
- [4] 陈珂,刘凤娟,易姿,等.目标物诱导的等离子体激光耦合以及表面增强拉曼光谱用于牛奶中三聚氰胺的检测[J].化学传感器,2013,33(4):29-37. (Chen K, Liu F J, Yi Z, et al. Target-controlled plasmonic coupling and surface enhanced Raman scattering transduction for simple, rapid and highly sensitive detection of melamine in milk[J]. Chemical Sensors, 2013, 33(4):29-37.)
- [5] 王琳,王雪,田静秒,等.表面增强拉曼光谱检测牛奶中三聚氰胺[J].食品研究与开发,2016,37(20):164-166. (Wang L, Wang X, Tian J M, et al. Surface enhanced Raman spectroscopy (SERS) assay method for melamine in milk[J]. Food Research and Development, 2016, 37(20):164-166.)
- [6] 王海燕,桂冬冬,沙敏,等.拉曼光谱结合模式识别算法用以牛奶制品智能判别与参数优化[J].中国奶牛,2018,(2):55-60. (Wang H Y, Gui D D, Sha M, et al. Application of Raman spectroscopy combined with pattern recognition algorithm for intelligent identification of dairy products and parameter optimization[J]. China Dairy Cattle, 2018, (2):55-60.)
- [7] Mendes T O,Junqueira G M,Porto B L, et al. Vibrational spectroscopy for milk fat quantification: line shape analysis of the Raman and infrared spectra[J]. Journal of Raman Spectroscopy, 2016, 47(6):692-698.
- [8] Li M,Chen J,Xu J, et al. Determination of lactose in milk by Raman spectroscopy[J]. Analytical Letters,2015,48(8):1333-1340.
- [9] Huang Y Z, Gu J M, Xiang G, et al. Detection of total protein in milk using phosphomolybdic acid-mediated surface-enhanced Raman spectroscopy [J]. Journal of Raman Spectroscopy, 2016, 47(3): 277-282.
- [10] Mazurek S, Szostak R, Czaja T, et al. Analysis of milk by FT-Raman spectroscopy[J]. Talanta, 2015, 138:285-289.
- [11] Galloway C M,Ru E L,Etchegoin P G. An Iterative Algorithm for Background Removal in Spectroscopy by Wavelet Transforms[J]. Applied Spectroscopy, 2009,63(12):1370-1376.
- [12] Ni Y N, Huang C F, Kokot S. Simultaneous determination of iron and aluminium by differential kinetic spectrophotometric method and chemometrics[J]. Analytica Chimica Acta, 2007, 599(2):209-218.
- [13] Bayen A M,Siauw T,Chapter 13-Least Squares Regression, Editor(s): Alexandre M. Bayen, Timmy Siauw, An Introduction to MATLAB <sup>®</sup> Programming and Numerical Methods for Engineers, Aca-

- demic Press, 2015;201-210.
- [14] Yasin N, Naqvi S M D, Akhter M S. Simultaneous spectrophotometric determination of Co (II) and Co (III) in acidic medium with partial least squares regression and artificial neural networks[J]. *Heliyon*, 2024,10(4):26373.
- [15] Wei D, Chen S, Liu Q. Review of fluorescence suppression techniques in Raman spectroscopy[J]. *Applied Spectroscopy Reviews*,2015, 50(5):387-406.
- [16] Lazic V, Fantoni R, Falzone S, et al. Stratigraphic characterization of ancient Roman frescos by laser induced breakdown spectroscopy and importance of a proper choice of the normalizing lines[J]. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 2020, 168:105853.
- [17] McGoverin C M, Clark A S S, Holroyd S E, et al. Raman spectroscopic quantification of milk powder-constituents[J]. *Analytica Chimica Acta*, 2010,673(1):26-32.
- [18] Khan H M H, McCarthy U, Esmonde-White K, et al. Potential of Raman spectroscopy for in-line-measurement of raw milk composition[J]. *Food-Control*, 2023, 152:109862.