

基于可见-近红外光谱的鸡蛋胆固醇含量无损检测研究

王巧华^{1,2} 顾梦圆¹ 刘成康¹ 吴佳权¹

(1. 华中农业大学工学院, 武汉 430070; 2. 农业农村部长江中下游农业装备重点实验室, 武汉 430070)

摘要: 鸡蛋作为优质蛋白质来源在全球膳食结构中占据重要地位, 但因其含有胆固醇而引发的健康争议持续受到消费者关注。传统检测方法需破坏样本且耗时费力, 难以满足禽蛋行业大规模分级需求。为实现胆固醇含量的快速无损检测, 本研究采用可见-近红外光谱技术, 结合最小二乘支持向量机 (LSSVM)、偏最小二乘回归 (PLSR)、随机森林 (RF) 等方法构建鸡蛋胆固醇含量的回归预测模型。通过采集 135 枚新鲜鸡蛋在横放 (L)、尖端朝下 (S)、钝端朝下 (B) 3 种姿态的透射光谱, 结合多元散射校正 (MSC)、标准正态变换 (SNV) 等预处理方法, 并系统比较竞争性自适应重加权采样 (CARS)、连续投影算法 (SPA) 及非信息变量消除 (UVE) 3 种特征选择算法对模型性能的影响。结果显示: 光谱采集姿态显著影响预测精度, 其中横放 (L) 姿态采集的光谱建立模型性能最优; 经 KS 数据划分后 SNV-CARS-LSSVM 模型性能最佳, 其训练集决定系数 R^2 达 0.919 0, 测试集决定系数 R^2_p 为 0.917 4; 训练集 RMSE 为 60.868 2 mg/(100 g), 测试集 RMSE 为 41.011 9 mg/(100 g); 表明本文建立的光谱模型可实现对鸡蛋胆固醇含量的快速无损精准化预测, 为鸡蛋品质监测与消费者选择提供技术支撑。

关键词: 鸡蛋; 胆固醇; 可见-近红外光谱; 最小二乘支持向量机; 无损检测

中图分类号: TS253.7

文献标识码: A

文章编号: 1000-1298(2026)21-0400-09

OSID:



Non-destructive Detection of Egg Cholesterol Content Based on Visible – Near-infrared Spectroscopy

Wang Qiaohua^{1,2} Gu Mengyuan¹ Liu Chengkang¹ Wu Jiaquan¹

(1. College of Engineering, Huazhong Agricultural University, Wuhan 430070, China

2. Key Laboratory of Agricultural Equipment for the Middle and Lower Reaches of the Yangtze River, Wuhan 430070, China)

Abstract: As a high-quality protein source, egg plays an important role in the global dietary structure. However, the health controversy caused by their cholesterol content has continuously attracted consumers' attention. Traditional detection methods require sample destruction, which are time-consuming and labor-intensive, and thus cannot meet the needs of large-scale grading in the poultry egg industry. To achieve rapid and non-destructive detection of cholesterol content, the visible – near-infrared spectroscopy technology was employed, combined with methods such as the least squares support vector machine (LSSVM), partial least squares regression (PLSR), and random forest, to construct regression prediction models for egg cholesterol content. Transmission spectra of 135 fresh eggs were collected under three postures: horizontal placement (L), tip-down (S), and blunt-end down (B). Combined with preprocessing methods such as multiplicative scatter correction (MSC) and standard normal variate transformation (SNV), the effects of three feature selection algorithms, i. e., competitive adaptive reweighted sampling (CARS), successive projections algorithm (SPA), and uninformative variable elimination (UVE) on model performance were systematically compared. The results showed that the spectral collection posture significantly affected the prediction accuracy, among which the model built based on the spectra collected in the horizontal (L) posture exhibited the optimal performance. After KS data partitioning, the SNV – CARS – LSSVM model achieved the best performance, with the coefficient of determination (R^2) of the training set reaching 0.919 0 and that of the test set being 0.917 4; the root mean square error of RMSEC was 60.868 2 mg/(100 g), and the root mean square error of RMSEP was 41.011 9 mg/(100 g). It indicated that the established spectral model can realize rapid, non-destructive

收稿日期: 2025 – 08 – 03 修回日期: 2025 – 09 – 03

基金项目: 国家自然科学基金项目 (32372426) 和国家重点研发计划项目 (2024YFD2000901)

作者简介: 王巧华 (1970—), 女, 教授, 博士, 主要从事农产品智能化检测技术与装备研究, E-mail: wqh@mail.hzau.edu.cn

and accurate prediction of egg cholesterol content, providing technical support for egg quality monitoring and consumer choices.

Key words: egg; cholesterol; visible – near-infrared spectroscopy; LSSVM; non-destructive detection

0 引言

鸡蛋富含多种营养物质,是极具性价比的动物性蛋白食品^[1]。其具备高蛋白、低脂肪、低热量以及高吸收率和高转化率等特性,已成为居民广泛认可的优质动物蛋白来源,在城乡居民饮食结构中占据重要地位^[2]。但因其含有一定的胆固醇而引发的健康争议正持续受到消费者关注。胆固醇是人体细胞膜的重要组成成分,是合成激素和维持神经系统健康的必需物质^[3]。适量的胆固醇可以促进细胞生长和代谢,维护人体生理功能的正常运作,但过高的胆固醇摄入可能与心脑血管疾病的发生密切相关,长期摄入高胆固醇食品可能导致心血管疾病如动脉粥样硬化等^[4],此外人体若长期处于高胆固醇水平还可能增加心脏病发作和中风的风险^[5]。然而,传统胆固醇理化检测流程繁杂,不仅耗时较长,还需专门实验设施,故而亟需开发更为简便、快捷的检测技术,以保障公众健康。

目前,可见近红外光谱技术在国内外禽蛋内部品质检测领域已得到广泛应用。该技术基于可见及近红外光谱区光子与有机分子中含氢官能团的倍频和合频吸收原理,可实现对被测物质的定量与定性分析。文献[6–8]基于光谱技术实现了鸡蛋新鲜度的无损检测;文献[9–10]利用光谱技术建立了血斑蛋的在线无损检测方法;文献[11]通过可见-近红外光谱技术实现了新鲜鸡蛋蛋白质含量的无损预测;文献[12]基于可见-近红外光谱技术构建了不同品种鸡蛋贮藏期 S-卵白蛋白含量的无损检测模型;文献[13]基于 UV–VIS–NIR 建立了蛋黄胆固醇的回归预测模型,决定系数 R^2 达到 0.78。然而,在禽蛋内部品质无损检测方面的现有研究,主要聚焦于新鲜度、蛋白质、脂肪等大类物质的分析,针对胆固醇这类含量较低且对人类健康存在一定争议影响的特异性成分,其无损检测研究匮乏。

为此,本文采用实验室自行搭建的可见-近红外光谱采集系统,采集鸡蛋在横放(L)、尖端朝下(S)、钝端朝下(B)3种姿态下的光谱数据,为分析采集姿态对模型的影响提供基础数据。随后,对原始光谱数据进行系列预处理,包括滤波平滑(Savitzky–Golay, SG)、多元散射校正(Multiplicative scatter correction, MSC)、标准正态变换(Standard normal variate transformation, SNV)及归一化

(Normalization, NX),以降低噪声干扰;同时结合竞争性自适应重加权采样(CARS)、连续投影算法(SPA)及非信息变量消除(UVE)3种算法完成特征提取,筛选关键光谱特征。在此基础上,选用最小二乘支持向量机(LSSVM)、偏最小二乘回归(PLSR)、随机森林(RF)3种方法构建回归预测模型,通过对比不同模型的预测性能及3种采集姿态对模型精度的影响,深入测定分析鸡蛋胆固醇含量,探讨鸡蛋特异性光谱特征与胆固醇浓度的内在关联,最终建立鸡蛋胆固醇含量的品质预测模型,以期为提升禽蛋生产的经济效益与智能化水平、推动高质量家禽养殖提供理论依据。

1 材料与方法

1.1 试验材料

样本选择市面上常见的普通海兰褐鸡蛋,总计 135 枚。该鸡蛋样本被妥善保存于恒温恒湿箱(温度 5℃;相对湿度 75%)内,以保持其新鲜度和稳定性。同时确保光谱检测的所有程序也在同一天内顺利完成。这种同步性有助于减少由于时间延迟可能产生的样本变性,从而提高试验结果的可靠性。

1.2 试验试剂与仪器

试验试剂:乙醚;石油醚(30~60℃);无水乙醇;超纯水。

试验仪器:Maya2000Pro 型光纤光谱仪(海洋光学公司,美国);84–UV 系列准直透镜(广州航鑫光电科技有限公司);超声波清洗机(深圳福洋科技集团有限公司);HPLC 气相色谱仪(Thorlabs 公司,美国)。

1.3 数据采集方法

1.3.1 光谱采集

本研究采用自搭建的可见-近红外光谱检测装置(示意图如图 1 所示)进行数据采集,光信号经长度为 1 000 mm、光纤直径为 7 mm 的光纤传导至另一端后,光纤终端从鸡蛋下方中部抵接其表面,于蛋壳表面形成与光纤直径一致的 7 mm 圆形光斑;在距该光纤终端正上方 120 mm 处架设准直透镜以采集光信号,采集到的光信号经光纤传送至光纤光谱仪,转换为电信号后,再通过数据线传输至计算机终端显示。为消除暗电流及环境因素的干扰,每次试验前均需采集暗光谱并在鸡蛋样本的光谱采集过程

中予以剔除。同时,为确保光源的光谱能量分布达到稳定状态提升光谱采集的稳定性,需将光源预热 30 min 后再开展试验。试验全程于恒温室内进行,以保证光谱采集过程的稳定性。每次扫描时,从该期样本中依次抽取单枚鸡蛋置于可见-近红外光谱检测系统中,透射光信号经鸡蛋后,由上方的准直透镜接收。为考察鸡蛋不同部位的光谱特征,试验采集了 3 种姿态下的光谱数据,分别为尖端朝下(S)、钝端朝下(B)和水平横放(L),如图 1 左下角所示。透射光信号通过玻璃光纤传输至光谱仪,并在上位机软件中记录光谱仪输出的原始光谱数据。光谱仪参数设置为:积分时间 50 ms,平均扫描次数 3 次,平滑处理 3 次。后续分析采用吸光度 A 作为光谱数据处理原型,计算公式为

$$A = -\lg \frac{I - I_1}{I_0 - I_1}$$

(1)

式中 A ——吸光度
 I ——样品透射光强
 I_0 ——参考光强
 I_1 ——暗光谱光强

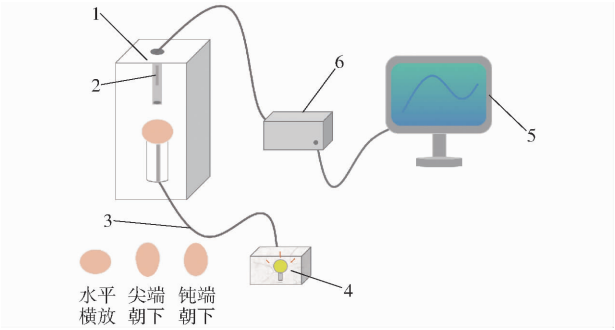


图 1 可见-近红外光谱检测装置示意图

Fig. 1 Schematic of visible - near-infrared spectrum detection device

1. 暗箱 2. 准直透镜 3. 光纤 4. 光源 5. 计算机 6. 光谱仪

1.3.2 胆固醇含量对比测定

鸡蛋胆固醇含量测定依据 GB 5009. 128—2016《食品安全国家标准 食品中胆固醇的测定》进行检测。鸡蛋样品匀质后,称取制备后的样品 0.5 g 于 250 mL 圆底烧瓶中,加入 30 mL 无水乙醇,10 mL 质量分数 60% 的氢氧化钾溶液混匀,进行皂化回流。皂化结束,用 5 mL 无水乙醇自冷凝管顶端冲洗其内部,取下圆底烧瓶,用流水冷却至室温(20℃),定量转移全部皂化液于 250 mL 分液漏斗中,用 30 mL 水分 3 次冲洗圆底烧瓶,洗液并入分液漏斗,再用 40 mL 石油醚-无水乙醚混合液(体积比 1:1)分 3 次冲洗圆底烧瓶并入分液漏斗,振荡 2 min,静置,分层。转移水相,合并 3 次有机相,每次用水 100 mL 洗涤提取液至中性。提取液浓缩至干,用无

水乙醇溶解定容至 5 mL 容量瓶中后,封样,采用气相色谱法检测、外标法定量^[14],试验过程如图 2 所示。为提高试验效率,皂化环节应用超声波振荡机和加热提高试验皂化效果,超声温度 70℃、超声功率 500 W、超声时间 30 min^[15]。为保证试验严谨性,同一样本做 3 组重复取平均值。胆固醇浓度 y (mg/(100 g)) 和气相色谱峰面积 x (pA·s) 响应关系为

$$y = 0.4038x + 12.33 \quad (R^2 = 0.999)$$

(2)

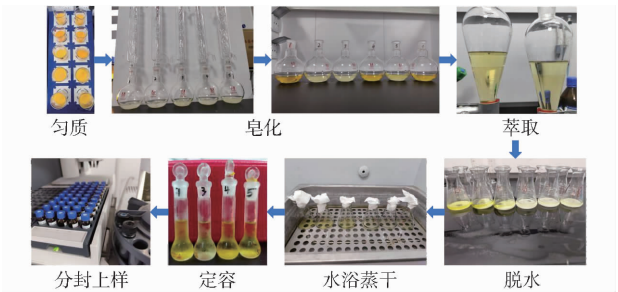


图 2 鸡蛋胆固醇物质提取与浓度测定

Fig. 2 Extraction and concentration determination of cholesterol in eggs

1.4 光谱数据分析方法

1.4.1 样本集划分方法

研究采用基于联合 $X - Y$ 距离的样本集划分方法(Sample set partitioning based on joint $X - Y$ distance,SPXY)和 Kennard - Stone 算法(Kennard - Stone algorithm,KS)样本划分方法。

1.4.2 光谱预处理方法

由于仪器和样品背景等因素的影响,可见-近红外光谱可能会出现偏移和漂移现象。因此在建模前需对光谱进行预处理。本试验中,采用 SG、MSC、SNV、NX 等方法对光谱进行预处理。

1.4.3 特征波长提取算法

本研究采用 3 种特征波长选择算法优化模型性能:CARS^[16]、SPA^[17]和 UVE^[18]。

1.4.4 模型建立与性能评估方法

研究采用 LSSVM、PLSR、RF 等方法构建鸡蛋胆固醇含量的回归预测模型。LSSVM 通过将数据映射到高维特征空间,并利用最小二乘准则求解等式约束优化问题,有效构建线性决策边界以实现回归预测^[16]。PLSR 通过最优化预测变量和响应变量间的协方差结构,有效地从数据中提取关键特征,并据此构建预测模型^[17-18]。RF 通过集成大量决策树并进行随机特征选择与样本采样,有效降低模型方差并提升泛化能力,最终通过多棵树结果平均机制进行回归预测^[19-20]。

模型性能使用决定系数(R_c^2 、 R_p^2)、均方根误差(RMSEC、RMSEP)评估。决定系数越接近 1,均方根误差越小,则模型性能更稳定、准确率越高。

2 结果与分析

2.1 光谱分析

2.1.1 原始光谱曲线分析

图 3 为鸡蛋在水平横放(L)、尖端朝下(S)及钝端朝下(B)3 种采集姿态下的原始光谱图。分析结果表明,3 种采集姿态下的光谱曲线整体趋势基本一致,尖端朝下(S)与钝端朝下(B)两种姿态所获取的平均光谱近乎重合;而横放姿态(L)采集的光谱吸光度在全波段范围内均显著低于尖端朝下和钝端朝下两种采集姿态,该差异可能与光线在鸡蛋内部的传输路径及散射特性变化相关。

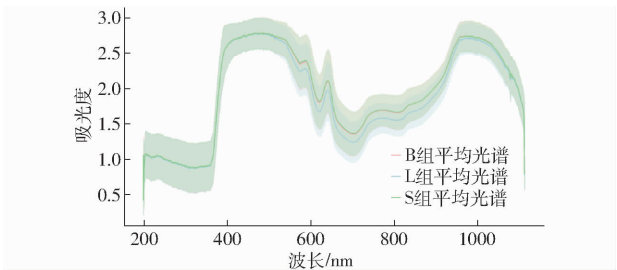


图 3 不同采集姿态下平均光谱及其光谱带
Fig. 3 Average spectral and spectral bands under different collection postures

观察光谱曲线可以明显发现,在波长为 580.86、587.14、620.73、640、700、806.86、978.1、1 084 nm 波段附近有明显的吸收峰。在 580.86、587.14 nm 波段附近可见近红外光谱吸收峰主要是电子跃迁与氧化产物吸收,以及脂质氧化产物的共轭双键吸收^[21-22];赵杰文等^[23]指出,鸡蛋的透射光谱在 500~900 nm 范围内具有显著的波峰和波谷,这些差异可能与蛋白质、脂肪含量和含水率有关;脂肪、胆固醇同属于脂质类物质。620.73、640 nm 波段附近对应蛋黄内色素选择性吸收与散射和 C—H 伸缩振动(第 3 倍频)以及水分子 O—H 的合频振动^[24];700、806.86 nm 主要是米氏散射与 O—H 振动偏移^[25];978.1 nm 反映的是 C—H/C=O 振动与倍频效应^[26];1 084 nm 附近反映的是水分中 O—H 键的倍频振动^[27]。

2.1.2 不同预处理光谱和鸡蛋胆固醇含量相关性分析

在光谱数据处理中,预处理方法的选择影响回归模型的预测精度。对原始光谱进行 SG、MSC、SNV、NX 变换,获取不同光谱变换后的光谱变量与鸡蛋胆固醇含量的相关性,胆固醇含量和光谱变量的相关系数图以及最大相关系数和敏感波段如图 4 所示。

由图 4 可见,经 MSC 和 SNV 变换后的光谱与胆固醇含量的相关性走势相近。其中,MSC 变换后的光谱在 520~670 nm、685~777 nm 及 850~

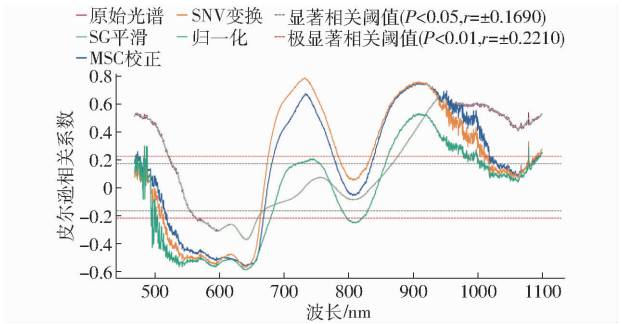


图 4 不同预处理在不同波长下光谱变量和胆固醇含量相关性
Fig. 4 Correlation between spectral variables in different wavelength bands and cholesterol content under different pretreatments

1 050 nm 波段与胆固醇含量的相关性达到 $P < 0.01$ 的极显著水平,在 920.7 nm 处,光谱吸光度与胆固醇含量的相关性最高,达 0.7456。该波段与胆固醇分子中 C—H 键的 3 阶倍频振动密切相关——脂质(包括胆固醇)的亚甲基(CH_2)伸缩振动在 920 nm^[28]附近存在特征吸收,而胆固醇的 C—H 键因具有高度非谐性,在近红外区会产生强吸收,尤其是甲基(CH_3)和亚甲基(CH_2)的合频振动^[29-31]。SNV 变换后的光谱在 515~664 nm、677~784 nm 及 850~1 050 nm 波段与胆固醇浓度的相关性达到 $P < 0.01$ 的极显著水平,在 733.4 nm 处,光谱吸光度与胆固醇含量的相关性最高,达 0.7813。该波段被认为对应 C—H 弯曲振动与伸缩振动的合频模式^[32],接近脂肪族 C—H 键的合频振动区域(如 CH_2 弯曲 + C—H 伸缩),可能表明胆固醇的脂肪链结构在此处贡献显著^[31]。SG 变换后的光谱与原始光谱基本重合,其与胆固醇含量的相关性曲线在 470~522 nm、565~660 nm 及 877~1 100 nm 波段附近达到 $P < 0.01$ 的极显著水平,在 948 nm 处相关性最高,为 0.6469。该波段与 N—H 键的 2 阶伸缩振动相关^[29],可能反映了胆固醇分子中羟基(O—H)的倍频吸收。NX 变换后的光谱与胆固醇含量的相关性曲线在 472.3~679.9 nm、798~822 nm 及 861~978 nm 波段达到 $P < 0.01$ 的极显著水平,在 642.6 nm 处相关性最高,为 0.5913。该波段对胆固醇的特异性较低,可能受到蛋黄中其他成分(如卵磷脂)的干扰^[33]。综上可知,经 MSC 和 SNV 变换后的光谱吸光度与胆固醇含量的相关性略优于其他变换方法。

2.2 不同样本集划分的样本集数据

试验样本共计 135 枚,按照 3:1 比例进行训练集和测试集划分。统计得出训练集和测试集的最大值、最小值、平均值和标准差如图 5 所示。由图 5 可

知训练集样品的数值分布覆盖了测试集样品的数值分布。表明,训练样品集与测试样品集 的样品划分是合理的。

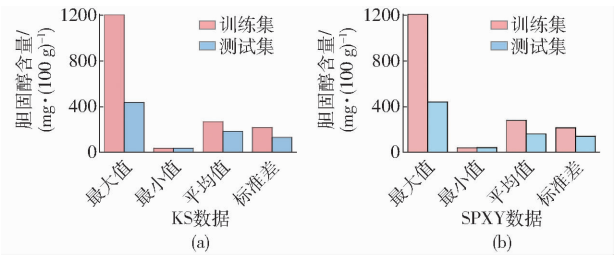


图 5 不同划分方式的数据分布
Fig. 5 Data distribution under different division methods

2.3 特征波长提取结果分析与可视化

2.3.1 CARS 算法提取特征波长

使用 CARS 算法对采集到的鸡蛋光谱数据进行处理。如图 6a 所示,随着迭代次数的增加,大量无用的光谱信息被逐步剔除。图 6b 显示,在交叉验证均方根误差(RMSECV)降至最低点(118 次迭代)之前,主要是消除无用的光谱信息;在超过最低点之后,随着有效光谱信息的剔除,RMSECV 开始上升。图 6c 中,当 RMSECV 达到最低点时,对应的迭代次数为 118 次,表明此时光谱中的有效信息量达到最大,从而确定了最优特征波长集合。筛选后的特征波长包括 389.93、394.53、389.01、396.37、485.5、500、657.7、669.7、752.25、753.13、750.49、752.69、974.1、971.6、974.3、1 023.2、1 034.5、1 067.5、1 082.1、1 092.8 nm 共计 20 个,CARS 筛选的特征波长涵盖了 2.1.1 节中可见-近红外光谱分析的部分特征光谱区域。

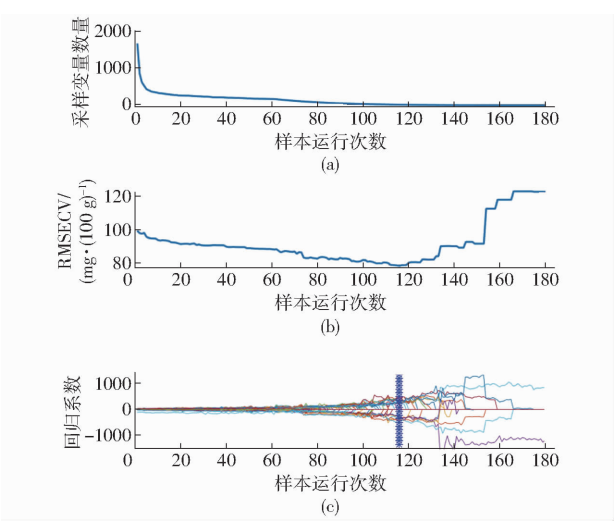


图 6 波长变量数目、RMSECV 和回归系数变化曲线
Fig. 6 Number of wavelength variables, RMSECV trend and regression coefficient

2.3.2 SPA 算法提取特征波长

利用 SPA 算法对鸡蛋光谱的特征变量进行筛

选。如图 7a 所示,当均方根误差(RMSE)降低到合理的数值范围内时,将对应拐点的变量选为特征变量,此时大量无用的光谱信息被剔除,剩余的光谱信息将被作为特征光谱用于后续分析。该拐点对应的波长个数为 19,RMSE 为 37.276 6 mg/(100 g),选取的特征变量数目远少于原始变量,显著降低了数据的维度。图 7b 为选取的特征波长点,经过 SPA 筛选后,选择了 396.4、411.5、422.1、439、459.1、467、492.3、495.3、510.9、531.3、675.5、746.5、789.9、948.8、990.8、1 032.4、1 049.6、1 079.2、1 098.7 nm 共计 19 个特征波长,虽然未完全与 2.1.1 节可见-近红外光谱的典型特征区域重合,但仍具有一定的代表性。

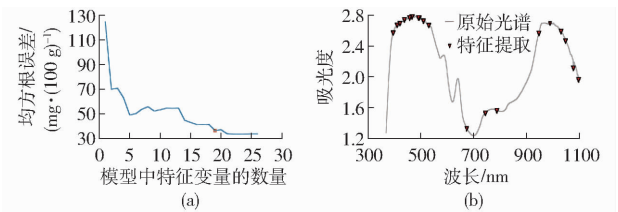


图 7 SPA 算法提取特征波长 RMSE 趋势图和 SPA 特征波长图

Fig. 7 RMSE trend chart of characteristic wavelengths extracted by SPA algorithm and characteristic wavelength chart

2.3.3 UVE 算法提取特征波长

应用 UVE 算法时,将阈值设定为 0.99,主成分数设定为 20。如图 8 所示,UVE 提取特征波长的过程详细展示了筛选步骤。图中左侧的黄线代表鸡蛋光谱矩阵的稳定值,而右侧的红色曲线则表示加入随机噪声后的光谱矩阵变量。蓝色虚线表示随机噪声的最大阈值(±9.804 7),位于这两条虚线之间的光谱信息被视为无效信息并予以剔除,而其余信息则保留为特征波长。经过 UVE 筛选,共筛选出 52 个特征波长,筛选后的特征波长包括 505.5、601.9、730.7、741.3、811.2、852.8、974.3 nm 等,部分涵盖了 2.1.1 节可见近红外光谱分析中关键波段。

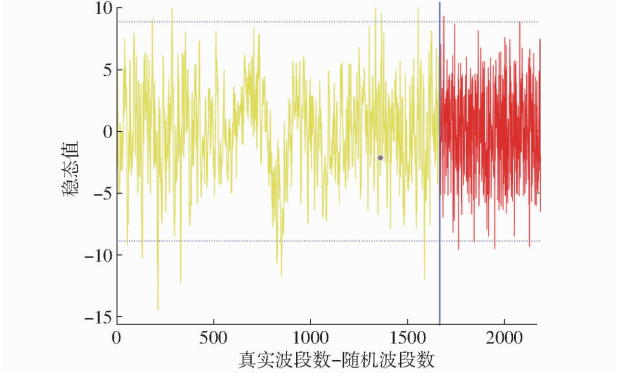


图 8 UVE 算法选择特征波长
Fig. 8 UVE algorithm selected characteristic wavelengths

2.4 回归模型的建立与预测

2.4.1 LSSVM 模型

为确定鸡蛋内部胆固醇含量的最优预测模型，本研究采用水平横放(L)、尖端朝下(S)、钝端朝下(B)3种采集姿态；MSC、SNV、NX、SG共4种算法进行原始光谱预处理，对不同预处理的光谱数据分别采用CARS、SPA、UVE共3种特征提取结合LSSVM算法进行品质预测，检测结果如图9所示。

图9展示了KS、SPXY两种不同样本划分策略下，光谱预处理与特征提取算法组合对模型性能的影响。结果表明，基于KS划分方法构建的模型在训练集和测试集的拟合优度与预测精度方面，均显著优于SPXY划分方法。具体而言，SPXY划分所得模型的训练集决定系数 R_c^2 普遍低于测试集决定系数 R_p^2 ，该现象表明SPXY数据划分方式效果不佳，不利于实际预测应用。因此，本研究后续分析采用KS方法进行样本划分以确保模型泛化能力。在光谱预处理方法方面，观察图9可知，对比发现：除SG平滑外，SNV、MSC、NX等3种预处理方法，对3个不同采集姿态下获取的光谱数据均展现出积极的回归效果提升。其中， R_p^2 与 R_c^2 的提升程度因光谱采集姿态和预处理方法而异。

表1呈现了在KS划分方式下，不同预处理、采集姿态及特征提取算法组合对应模型 R^2 与RMSE。由表中数据可知，在采用相同预处理和特征提取算法前提下，横放姿态采集的光谱数据构建模型回归效果最优。SG预处理对模型性能未产生任何提升作用。为更清晰展示结果，图10单独呈现了3种采集姿态下不同算法组合模型的 R_p^2 分布（已剔除SG预处理对应结果）。由图10可知，采用L姿态采集的光谱数据集所构建模型的 R_p^2 整体高于其余两种采集姿态。横置姿态下，蛋黄因重力作用均匀分布于赤道平面，形成稳定的光学检测界面，光照区域在蛋壳表面分布相对更广且更均匀，减少了因曲率突变（如竖直放置时蛋的尖端或钝端）导致的光散射异常或阴影区。均匀的光照有利于获取更能代表整体蛋内容物（特别是蛋黄、胆固醇主要富集区）的光谱信息。相反，在竖直姿态下，光源可能更集中于尖端或钝端的较小区域，光照穿透不均有效信息量减少，且可能受蛋壳厚度局部差异和气室位置影响更大^[34]。蛋黄作为胆固醇主要载体，在横放姿态下，近红外光束穿过鸡蛋的路径更可能横贯蛋黄中心区域或其附近，获取的光谱信号包含更丰富的蛋黄成分信息。

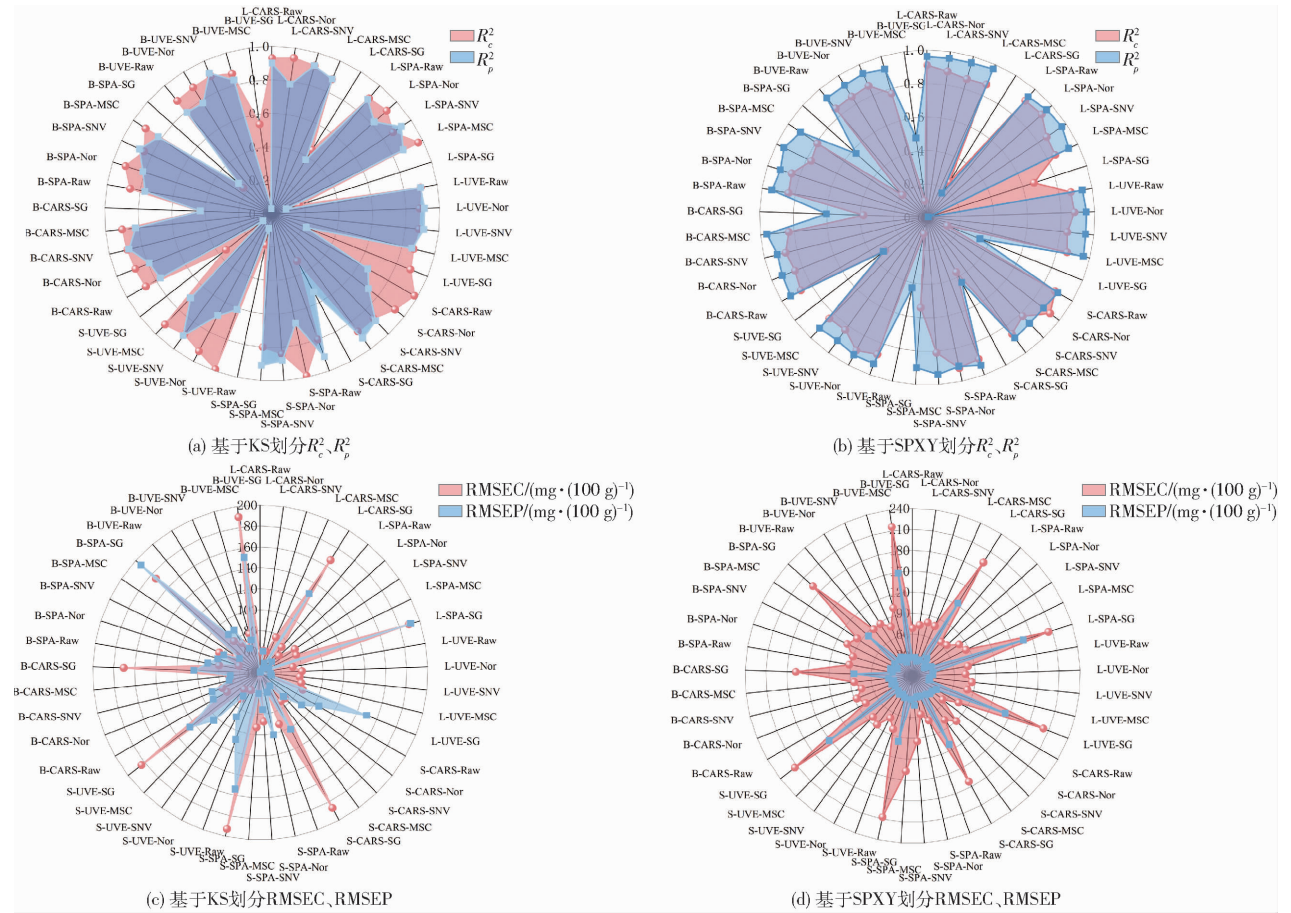


图9 不同姿态、预处理结合特征选择后 LSSVM 模型的 R^2 、RMSE

Fig. 9 R^2 and RMSE of LSSVM model after combining feature selection with different postures and preprocessing

表 1 不同采集姿态算法组合下 LSSVM 模型 R^2 和 RMSE

Tab. 1 R^2 and RMSE of LSSVM models under different algorithm and postures combinations

预处理 方法	评价指标	CARS			SPA			UVE		
		L	S	B	L	S	B	L	S	B
Raw	R_c^2	0. 928 1	0. 984 0	0. 871 6	0. 896 3	0. 803 7	0. 858 6	0. 901 6	0. 989 7	0. 876 2
	R_p^2	0. 902 9	0. 660 2	0. 766 9	0. 885 0	0. 915 8	0. 770 6	0. 908 5	0. 611 1	0. 787 2
	RMSEC/($\text{mg} \cdot (100 \text{ g})^{-1}$)	59. 778 0	27. 366 1	76. 696 4	71. 833 4	92. 628 7	80. 150 9	71. 558 8	22. 542 1	79. 286 1
	RMSEP/($\text{mg} \cdot (100 \text{ g})^{-1}$)	40. 431 9	104. 714 7	91. 639 7	44. 838 0	60. 117 4	90. 833 9	40. 206 4	108. 340 3	87. 373 7
NX	R_c^2	0. 938 7	0. 929 8	0. 877 5	0. 922 4	0. 992 6	0. 922 1	0. 881 7	0. 936 7	0. 886 6
	R_p^2	0. 782 5	0. 729 8	0. 791 2	0. 823 1	0. 672 8	0. 812 2	0. 915 6	0. 694 7	0. 787 2
	RMSEC/($\text{mg} \cdot (100 \text{ g})^{-1}$)	61. 252 9	35. 667 5	79. 349 2	64. 158 4	21. 078 1	61. 705 0	79. 919 0	59. 048 9	73. 347 3
	RMSEP/($\text{mg} \cdot (100 \text{ g})^{-1}$)	60. 478 4	90. 314 5	89. 606 6	55. 082 3	101. 302 8	82. 701 3	37. 633 0	88. 311 4	87. 521 9
SNV	R_c^2	0. 919 0	0. 874 0	0. 888 7	0. 872 0	0. 835 7	0. 854 9	0. 881 5	0. 890 4	0. 896 4
	R_p^2	0. 917 4	0. 892 0	0. 882 4	0. 936 1	0. 877 2	0. 882 7	0. 912 2	0. 906 4	0. 923 8
	RMSEC/($\text{mg} \cdot (100 \text{ g})^{-1}$)	60. 868 2	74. 180 3	69. 765 5	79. 942 5	87. 072 3	83. 802 8	77. 816	71. 647 9	68. 842 8
	RMSEP/($\text{mg} \cdot (100 \text{ g})^{-1}$)	41. 011 9	72. 007 1	70. 850 6	40. 114 7	75. 497 4	78. 574 1	50. 094 9	67. 889 3	65. 135 4
MSC	R_c^2	0. 877 7	0. 868 9	0. 901 1	0. 870 7	0. 803 7	0. 905 4	0. 865 9	0. 918 7	0. 865 4
	R_p^2	0. 875 2	0. 918 4	0. 819 0	0. 876 7	0. 915 8	0. 817 6	0. 862 4	0. 703 5	0. 832 8
	RMSEC/($\text{mg} \cdot (100 \text{ g})^{-1}$)	76. 906 2	73. 867 3	70. 309 6	78. 420 5	92. 628 7	68. 139 7	80. 469 6	62. 306 1	78. 449 0
	RMSEP/($\text{mg} \cdot (100 \text{ g})^{-1}$)	49. 526 2	55. 764 8	68. 680 4	49. 227 1	60. 117 4	63. 975 8	51. 399 2	103. 846 3	72. 135 4
SG	R_c^2	0. 448 2	0. 316 3	0. 433 4	0. 170 5	0. 196 2	0. 232 2	0. 895 6	0. 351 8	0. 544 5
	R_p^2	0. 385 1	0. 539 3	0. 434 9	0. 087 4	0. 090 2	0. 277 5	0. 224 3	0. 076 8	0. 039 3
	RMSEC/($\text{mg} \cdot (100 \text{ g})^{-1}$)	166. 688 0	186. 704 7	170. 491 0	189. 271 7	192. 837 9	174. 107 8	83. 806 7	183. 846 6	189. 914 7
	RMSEP/($\text{mg} \cdot (100 \text{ g})^{-1}$)	128. 639 0	101. 936 7	103. 328 6	191. 097 4	154. 139 0	193. 116 7	149. 428 0	124. 903 1	150. 779 7

结合表 1 和图 10,通过对比最终选择横放姿态采集的光谱数据,基于 SNV 数据预处理 CARS 算法构建 SNV - CARS - LSSVM 模型用于预测鸡蛋内部胆固醇浓度。鸡蛋胆固醇含量的预测值与测定值之间的关系散点图如图 11 所示。 R_c^2 为 0.919 0、 R_p^2 为 0.917 4;RMSEC 为 60.868 2 $\text{mg}/(100 \text{ g})$,RMSEP 为 41.011 9 $\text{mg}/(100 \text{ g})$ 。训练集 RPD 为 2.61,测试集 RPD 为 3.22。

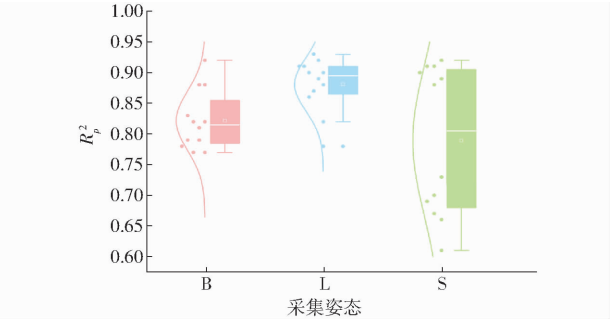


图 10 不同采集姿态下 R_p^2 分布

Fig. 10 Distribution of R_p^2 under different collection postures

2.4.2 不同模型比较

为进一步验证 LSSVM 模型性能,选择偏最小二乘回归(PLSR)、随机森林(RF)两种经典算法方法作为对比。采用水平横放(L)、尖端朝下(S)、钝端朝下(B)3种采集姿态;原始数据与4种预处理算法;CARS、SPA和UVE共3种波长选择降维算法进

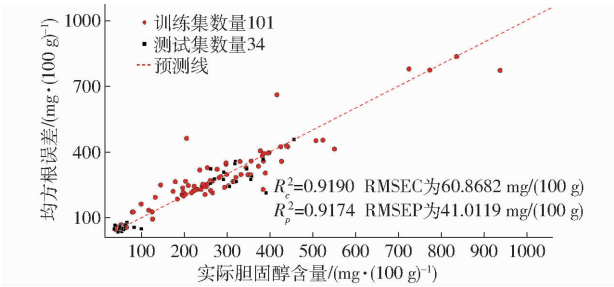


图 11 预测散点图

Fig. 11 Prediction scatter plot

行组合,从 90 种组合算法中得到每类模型综合最优预测结果与 LSSVM 方法中最优预测模型进行对比,结果如表 2 所示,LSSVM 模型效果最优,此外,不同模型都是以水平横放姿态采集的光谱建立模型效果最优,印证 2.4.1 节所述光谱采集姿态显著影响预测精度,其中水平横放(L)姿态采集的光谱建立模型性能最优。

3 结论

(1)构建了基于 SNV - CARS - LSSVM 的鸡蛋胆固醇含量无损检测模型。SNV 预处理与 CARS 特征提取协同优化;SNV 有效校正蛋壳曲率光散射的干扰,CARS 算法筛选出与鸡蛋胆固醇关联密切的稳健特征波段,优化后的模型在胆固醇含量的预测能力中 R^2 达 0.917 4,实现了对鸡蛋胆固醇含量的高效精确预测。

表 2 不同模型最优组合的 R^2 和 RMSE

Tab.2 R^2 and RMSE of optimal combination for different models

模型名称	训练集		测试集	
	R_c^2	RMSEC/(mg·(100 g) ⁻¹)	R_p^2	RMSEP/(mg·(100 g) ⁻¹)
L-SNV-CARS-LSSVM	0.919 0	60.868 2	0.917 4	41.011 9
L-NX-CARS-PLSR	0.847 7	77.438 4	0.858 6	78.561 2
L-MS-C-SPA-RF	0.891 7	70.249 8	0.884 3	43.988 7

(2) 试验结果表明,3 种模型中,均是鸡蛋横置位(L)的光谱采集方案对鸡蛋胆固醇含量的预测效能最优。该结果证实样本横置策略可有效优化胆固醇定量分析,为基于光谱的禽蛋品质无损检测提供了高效的样本姿态优化方案。

参 考 文 献

[1] 周龙珠,李萍漾,赵丽媛,等. 畜产品 $\omega-3$ 多不饱和脂肪酸生物营养强化研究应用进展[J]. 中国科学:生命科学, 2025, 55(3): 463-478.
Zhou Longzhu, Li Pingyang, Zhao Liyuan, et al. Advances in the application of biofortification research on omega-3 polyunsaturated fatty acids in livestock products[J]. Scientia Sinica(Vitae), 2025, 55(3): 463-478. (in Chinese)

[2] 国家统计局. 中国统计年鉴 2023[M]. 北京:中国统计出版社, 2023.

[3] 谭金鹏,张婉婷,徐成,等. 基于无监督学习方法的细胞膜内单分子扩散运动分析:胆固醇对模型膜和活细胞膜流动性的不同影响[J]. 物理学报,2024,73(18):284-294.
Tan Jinpeng, Zhang Wanting, Xu Cheng, et al. Analysis of single-molecule diffusion movement in cell membrane based on unsupervised learning methods:different effects of cholesterol on flowability of model membrane and living cell membrane[J]. Acta Physica Sinica,2024,73(18):284-294. (in Chinese)

[4] 周筱丹,董晓芳,佟建明. 食用鸡蛋与心血管疾病关系的研究进展[J]. 动物营养学报,2012,24(1):55-61.
Zhou Xiaodan, Dong Xiaofang, Tong Jianming. The relationship between egg consumption and cardiovascular diseases[J]. Chinese Journal of Animal Nutrition,2012,24(1):55-61. (in Chinese)

[5] 童彤摘译. 美国:草莓可降低胆固醇且有益心脏健康[J]. 中国果业信息,2024,41(11):60.

[6] Fu D, Li Q, Chen Y, et al. Assessment of integrated freshness index of different varieties of eggs using the visible and near-infrared spectroscopy[J]. International Journal of Food Properties, 2023, 26(1): 155-166.

[7] Dong X, Zhang B, Dong J, et al. Egg freshness prediction using a comprehensive analysis based on visible near infrared spectroscopy[J]. Spectroscopy Letters, 2020, 53(7): 512-522.

[8] Dong X, Dong J, Li Y, et al. Maintaining the predictive abilities of egg freshness models on new variety based on VIS-NIR spectroscopy technique[J]. Computers and Electronics in Agriculture, 2019, 156: 669-676.

[9] Chen M, Zhang L, Xu H. On-line detection of blood spot introduced into brown-shell eggs using visible absorbance spectroscopy[J]. Biosyst En, 2015(131): 95-101.

[10] 徐惠荣,徐文豪,陈华瑞,等. 基于光谱技术的褐壳血斑蛋鉴别方法研究[J]. 农业机械学报,2014,45(2): 194-198.
Xu Huirong, Xu Wenhao, Chen Huarui, et al. Detection of blood spots in brown eggs based on spectroscopic techniques[J]. Transactions of the Chinese Society for Agricultural Machinery, 2014, 45 (2): 194-198. (in Chinese)

[11] 吴建虎,黄钧. 可见/近红外光谱技术无损检测新鲜鸡蛋蛋白质含量的研究[J]. 现代食品科技,2015,31(5):285-290.
Wu Jianhu, Huang Jun. Non-destructive detection of protein content in fresh eggs by visible/near-infrared reflectance spectroscopy[J]. Modern Food Science and Technology,2015,31(5):285-290. (in Chinese)

[12] 付丹丹,王巧华,高升,等. 不同品种鸡蛋贮期 S-卵白蛋白含量分析及其可见/近红外光谱无损检测模型研究[J]. 分析化学,2020,48(2):289-297.
Fu Dandan, Wang Qiaohua, Gao Sheng, et al. Analysis of S-ovalbumin content of different varieties of eggs during storage and its nondestructive testing model by visible-near infrared spectroscopy[J]. Chinese Journal of Analytical Chemistry, 2020, 48(2):289-297. (in Chinese)

[13] Puertas G, Vazquez M. Cholesterol determination in egg yolk by UV-VIS-NIR spectroscopy[J]. Food Control, 2019, 100: 262-268.

[14] GB 5009.128—2016 食品安全国家标准 食品中胆固醇的测定[S].

[15] 李静. 禽蛋胆固醇检测方法的研究及应用[D]. 武汉:华中农业大学,2013.
Li Jing. Study and application on determination methods of cholesterol in eggs[D]. Wuhan:Huazhong Agricultural University, 2013. (in Chinese)

[16] Yao K, Sun J, Chen C, et al. Visualization research of egg freshness based on hyperspectral imaging and binary competitive adaptive reweighted sampling[J]. Infrared Physics & Technology, 2022, 127: 104414.

[17] Qin Y, Song K, Zhang N, et al. Robust NIR quantitative model using MIC – SPA variable selection and GA – ELM[J]. Infrared Physics & Technology, 2023, 128: 104534.

[18] Wang Z, Chen J, Fan Y, et al. Evaluating photosynthetic pigment contents of maize using UVE – PLS based on continuous wavelet transform[J]. Computers and Electronics in Agriculture, 2020, 169: 105160.

[19] Breiman L. Random forests[J]. Machine Learning,2001,45(1):5 – 32.

[20] 蒋蔚, 郭成波, 寇家华, 等. 基于改进随机森林算法的 RFID 室内定位方法[J]. 计算机科学,2025,52(增刊 1):819 – 825.
Jiang Wei, Guo Chengbo, Kou Jiahua, et al. RFID indoor positioning method based on improved random forest algorithm[J]. Computer Science,2025,52(S1):819 – 825. (in Chinese)

[21] Brant A W, Norris K H, Gilbert C. A spectrophotometric method for detecting blood in white-shell eggs[J]. Poultry Science, 1953, 32(2): 357 – 363.

[22] Bamelis F R, Tona K, De Baerdemaeker J G, et al. Detection of early embryonic development in chicken eggs using visible light transmission[J]. British Poultry Science,2002,43(2): 204 – 212.

[23] 赵杰文, 毕夏坤, 林颖, 等. 鸡蛋新鲜度的可见-近红外透射光谱快速识别[J]. 激光与光电子学进展,2013,50(5): 213 – 220.
Zhao Jiewen, Bi Xiakun, Lin Hao, et al. Visible – near-infrared transmission spectra for rapid analysis of the freshness of eggs [J]. Laser & Optoelectronics Progress, 2013, 50(5): 213 – 220. (in Chinese)

[24] Kuroki S, Kanoo T, Itoh H, et al. Nondestructive measurement of yolk viscosity in lightly heated chicken shell eggs[J]. Journal of Food Engineering, 2017, 205: 18 – 24.

[25] Yao K, Sun J, Cheng J, et al. Development of simplified models for non-destructive hyperspectral imaging monitoring of S-ovalbumin content in eggs during storage[J]. Foods, 2022, 11(14): 2024.

[26] Ghaderi M, Mireei S A, Masoumi A, et al. Fertility detection of unincubated chicken eggs by hyperspectral transmission imaging in the Vis – SWNIR region[J]. Scientific Reports, 2024, 14(1): 1289.

[27] Hoffman L C, Ni D, Dayananda B, et al. Unscrambling the provenance of eggs by combining chemometrics and near-infrared reflectance spectroscopy[J]. Sensors, 2022, 22(13): 4988.

[28] Ye R, Chen Y, Guo Y, et al. NIR hyperspectral imaging technology combined with multivariate methods to identify shrimp freshness[J]. Applied Sciences, 2020, 10(16): 5498.

[29] 谢安国, 纪思媛, 李月玲, 等. 基于遗传算法和深度神经网络的近红外高光谱检测猪肉新鲜度[J]. 食品工业科技, 2024,247(17):345 – 351.
Xie Anguo, Ji Siyuan, Li Yueling, et al. Detection of pork freshness using NIR hyperspectral imaging based on genetic algorithm and deep neural network[J]. Science and Technology of Food Industry,2024,247(17):345 – 351. (in Chinese)

[30] Tan L, Fu Y, Kang S, et al. Broadband NIR-emitting Te cluster-doped glass for smart light source towards night-vision and NIR spectroscopy applications[J]. Photonics Research, 2022, 10(5): 1187 – 1193.

[31] Abdelkader M F. Novel algorithms and instrumentation for vibrational spectroscopic methods of analysis[M]. Norfolk:Old Dominion University, 2013.

[32] 仇逊超. 机器视觉和近红外光谱对红松籽品质检测方法的研究[D]. 哈尔滨:东北林业大学,2017.
Qiu Xunchao. Study on quality detection method of Korean pine seed using machine vision and near infrared spectrum[D]. Harbin:Northeast Forestry University,2017. (in Chinese)

[33] 何仙玉, 黄祎, 教小磐. 超声辅助提取蛋黄卵磷脂工艺优化及其理化特性研究[J]. 食品工业科技,2019,40(3):185 – 190,206.
He Xianyu, Huang Yi, Jiao Xiaopan. Ultrasonic-assisted extraction of lecithin from egg yolk and its physicochemical properties [J]. Science and Technology of Food Industry,2019,40(3):185 – 190,206. (in Chinese)

[34] 孙俊, 刘彬, 毛罕平, 等. 基于介电特性与蛋黄指数回归模型的鸡蛋新鲜度无损检测[J]. 农业工程学报,2016, 32(21):290 – 295.
Sun Jun, Liu Bin, Mao Hanping, et al. Non-destructive examination for freshness of eggs based on dielectric properties and yolk index regression model[J]. Transactions of the CSAE,2016,32(21):290 – 295. (in Chinese)