李文峰, 夏意雯, 杨蕊, 等. 色泽测定结合机器学习预测胭脂萝卜中花色苷的含量 [J]. 食品工业科技, 2026, 47(11): 335−343. doi: 10.13386/j.issn1002-0306.2025060181

LI Wenfeng, XIA Yiwen, YANG Rui, et al. Prediction of Anthocyanin Content in Red Radish Using Colorimetry and Machine Learning[J]. Science and Technology of Food Industry, 2026, 47(11): 335−343. (in Chinese with English abstract). doi: 10.13386/j.issn1002-0306.2025060181

· 分析检测 ·

色泽测定结合机器学习预测胭脂萝卜中花色苷的含量

李文峰[1,2,3], 夏意雯[3], 杨　蕊[3], 明　建[2,*]

（1.重庆市涪陵榨菜集团股份有限公司, 重庆 408000;
2.西南大学食品科学学院, 重庆 400715;
3.长江师范学院生命科学与技术学院, 重庆 408100）

摘　要: 为了实现胭脂萝卜花色苷含量的快速测定，以提高胭脂萝卜色素生产时原材料的筛选效率。利用色差仪测定了萝卜剖面不同位置的 L^*、a^*、b^*、C^*和 H^*颜色参数。采用超高效液相色谱-质谱联用仪分析了胭脂萝卜中单体花色苷的含量。结果表明 a^*值与花色苷含量呈正相关，据此建立的 a^*值与单体花色苷和总花色苷含量的线性回归模型，对总花色苷含量预测准确率为 68.4%，对主要单体花色苷含量的预测准确率为 28.9%~52.6%。利用 8 个机器学习算法优化了基于颜色参数构建的花色苷含量预测模型。Lasso 回归模型能用于预测总花色苷含量，预测准确率达 87.5%。Decision Tree 回归模型能用于预测天竺葵素-3-(阿魏酰) 二葡萄糖苷-5-(丙二酰) 葡萄糖苷和天竺葵素-3-(咖啡酰) 二葡萄糖苷-5-葡萄糖苷含量，预测准确率均为 75%。这些结果表明色泽测定结合机器学习能快速预测胭脂萝卜的花色苷含量，可用于胭脂萝卜色素生产中原材料筛选。

关键词: 胭脂萝卜, 花色苷, 机器学习, 色泽, 含量预测

中图分类号: TS207.3　　**文献标识码:** A　　**文章编号:** 1002−0306(2026)11−0335−09

DOI: 10.13386/j.issn1002-0306.2025060181

本文网刊:

Prediction of Anthocyanin Content in Red Radish Using Colorimetry and Machine Learning

LI Wenfeng[1,2,3], XIA Yiwen[3], YANG Rui[3], MING Jian[2,*]

（1.Chongqing Fuling Zhacai Croup Co., Ltd., Chongqing 408000, China;
2.College of Food Science, Southwest University, Chongqing 400715, China;
3.School of Life Science and Biotechnology, Yangtze Normal University, Chongqing 408100, China）

Abstract: To rapidly determine anthocyanin content in red radish (*Raphanus sativus* L.) and improve raw material screening efficiency for pigment production, color parameters (L^*, a^*, b^*, C^*, and H^*) were measured at different positions on red radish using a colorimeter. Ultra-performance liquid chromatography-mass spectrometry was employed to analyze content of individual anthocyanin. A significant positive correlation was observed between a^* values and anthocyanin contents. Linear regression models based on a^* values achieved a 68.4% prediction accuracy for total anthocyanin content and from 28.9% to 52.6% for major individual anthocyanins. Eight machine learning algorithms were employed to optimize the color parameter-based prediction model of anthocyanin content. The Lasso regression model demonstrated strong predictive performance for total anthocyanin content, achieving a prediction accuracy of 87.5%. The Decision Tree regression model achieved prediction accuracies of 75% for both pelargonidin-3-(feruloyl)diglucoside-5-(malonyl)glucoside and pelargonidin-3-(caffeoyl)diglucoside-5-glucoside, indicating its effectiveness in estimating specific anthocyanin compounds. Collec-

收稿日期: 2025−06−16

基金项目: 重庆市自然科学基金面上项目: 胭脂萝卜射流干燥过程中花色苷的降解机制（cstc2019jcyj-msxmX0141）。

作者简介: 李文峰（1987−）, 男, 博士, 副教授, 研究方向: 食品化学与营养学, E-mail: lwffood@163.com。

*** 通信作者:** 明建（1972−）, 男, 博士, 教授, 研究方向: 食品化学与营养学, E-mail: mingjian1972@163.com。

tively, these findings indicate that combining colorimetric measurement with machine learning enables rapid prediction of anthocyanin content, which is conducive to raw material selection in red radish pigment production.

Key words: red radish; anthocyanins; machine learning; color; content prediction

胭脂萝卜(*Raphanus sativus* L.)又称红心萝卜,是重庆市涪陵区特色萝卜品种。胭脂萝卜单根重量约为 250~300 g,形状近似小球状或葫芦形,少数呈圆柱形,大多数个体从表皮到核心都含有大量花色苷,从而呈现为鲜红色[1]。花色苷属于多酚类化合物[2],具有良好的水溶性,不仅安全无毒[3],还具有多种生理活性,可作为天然色素应用于食品、化妆品和药品中[4−6]。前期研究发现,胭脂萝卜花色苷的主要成分是多酰化天竺葵苷,它们与花青素、花青素单糖苷和非酰化花色苷相比,拥有更强的极端 pH、热和紫外线稳定性,是极具应用潜力的天然色素[7]。然而胭脂萝卜的花色苷含量在不同个体中存在较大差异(图 1)。如果利用花色苷含量较少的胭脂萝卜生产色素,不仅徒增生产成本,还会增加色素的杂质含量。因此,非常有必要在色素提取前,快速准确地筛选出拥有较高花色苷含量的个体。



图 1 胭脂萝卜外观与剖面图

Fig.1 Images of appearance and profile of red radish

目前,针对胭脂萝卜花色苷含量的测定方法,主要包括 pH 示差法、高效液相色谱(HPLC)法、直接标准曲线测量法等[8−10]。这些方法虽然准确,但操作复杂且耗时长,难以在生产中迅速测定大规模胭脂萝卜群体的花色苷含量。研究表明果蔬呈现出的特定色泽与其花色苷含量呈强正相关[11−12]。因此,可以利用能快速测定的色泽参数构建数学模型,即时预测出花色苷的含量,实现胭脂萝卜色素生产中原材料的在线挑选。然而,目前鲜有研究文献明确报道可高效达成该目标的技术路径。机器学习作为传统数据处理过程的一种有效而经济的替代方法,具有自学习、容错能力强、预测准确率高、效率高、能够大规模快速检测等优点[13]。当前,随机森林、决策树和支持向量机等算法,已被成功用于构建食品相关的数学模型,从而解决食品领域的分类或回归问题[13]。然而,目前尚不清楚何种机器学习算法可有效构建胭脂萝卜花色苷含量的预测模型。

为此,本研究以涪陵胭脂萝卜为研究对象,利用色差仪获取色泽参数,采用 UHPLC-QqQ-MS/MS 测得单体花色苷含量;在此基础上首先利用线性回归方程初步揭示色泽参数与花色苷含量的关系,然后系统筛选能通过色泽参数准确预测花色苷含量的机器学习算法。相关研究结果能被用于胭脂萝卜色素生产的原材料筛选,也为其他富含色素食品的研究提供了新思路。

1 材料与方法

1.1 材料与仪器

胭脂萝卜　分别购买自长江师范学院周边的 4 个市场,19 个样品新鲜无明显腐坏,个体间的形状和尺寸属性差异较大,可以让模型具有更强的适应能力(图 1);盐酸　购于四川川东化工有限公司;甲酸、甲醇　购于 Adamas试剂有限公司;盐酸为实验室常备分析纯,甲酸、甲醇为色谱纯。

TGL-16B 台式离心机　上海安亭科技仪器厂;JA5003 电子天平　上海舜宇恒平科学仪器有限公司;FE28 pH 计　梅特勒-托利多仪器(上海)有限公司;FSH2A 型分散匀浆机　常州讯生仪器有限公司;NR60CP 型色差仪　3NH 科技有限公司;1290 Ⅱ-6460C 型超高效液相色谱-电喷雾-三重四极杆质谱联用(UHPLC-QqQ-MS/MS)　安捷伦科技(中国)有限公司。

1.2 实验方法

1.2.1 胭脂萝卜色泽参数测定　首先沿胭脂萝卜生长轴切分为三块(图 1),用色差仪分别测定截面上的五点(中心以及周围四个点),记录亮度值(L^*)、红绿值(a^*)和黄蓝值(b^*)。L^*值表示样品的亮度,大于 0 的 a^*值表示红色,小于 0 的 b^*值表示蓝色。H^*值被认为是颜色的定性属性,可以将颜色定义为传统的偏红或偏绿等属性[14]。C^*值作为色彩的定量属性,其值越高,人类感知到的样品颜色强度越高[14]。依据公式(1)和(2)计算色度值(C^*)和色彩角(H^*)。利用主成分分析(principal component analysis, PCA)揭示不同样品色泽参数的整体差异。

$$C^* = [(a^*)^2 + (b^*)^2]^{0.5} \quad \text{式(1)}$$

$$H^* = \tan^{-1}(b^*/a^*) \quad \text{式(2)}$$

1.2.2 胭脂萝卜花色苷的提取与 UHPLC-QqQ-MS/MS 分析　虽然先前的研究指出,使用酸化溶剂提取胭脂萝卜花色苷,可以较好地防止花色苷降解[15]。但为避免酸化溶剂腐蚀铁质匀浆机,本文采用中性溶液提取花色苷[15]。取 100 g 胭脂萝卜与 100 mL 纯水匀浆。10 mL 匀浆与 20 mL 酸化水(pH1)混匀后在 2500 r/min 离心 10 min,取上清液过 0.22 μm 孔径超滤膜。将 5 μL 待测液用 C_{18} 柱(100 mm×2.1 mm i.d., 1.8 μm)进行色谱分离,柱温控制在 35±0.5 ℃。流动相 A 为 0.1% 甲酸-水溶液,流动相 B 为 0.1% 甲酸-甲醇溶液,梯度洗脱总流速为 0.2 mL/min。洗脱程序如下:0~0.1 min,维持 10% 流动相 B;0.1~11.5 min,

流动相 B 从 10% 增加到 40%；11.5~14.5 min，流动相 B 从 40% 增加至 95%；14.5~16 min，维持流动相 B 为 95%；16~16.5 min，流动相 B 从 95% 降低到 10%；16.5~20 min，维持流动相 B 为 10%。

质谱的电喷雾离子源（ESI）设置为正离子模式，毛细管电压 3000 V。离子源干燥气为 N_2，温度为 350 ℃，流速为 10 L/min。碰撞气为高纯 N_2（99.999%），压力为 0.15 MPa。碎裂电压和碰撞能分别为 130 V 和 30 V，分子离子和碎片离子信息参考本团队先前的报道[15]。在多反应监测（MRM）模式下采集数据，依据母离子、子离子、保留时间和离子丰度比定性，依据质谱响应最强的子离子峰面积定量。

1.2.3　构建胭脂萝卜花色苷含量预测模型　采用稀疏的偏最小二乘法（sparse partial least square，sPLS）分析色泽参数与单体花色苷含量之间的相关矩阵。利用色泽 a^* 值和线性回归方程构建模型，用于预测单体花色苷和总花色苷含量。

将包含所有色泽参数的数据集分成涵盖 30 个样品的训练集和 8 个样品的验证集。以平均绝对误差（mean absolute error，MAE）、均方误差（mean square error，MSE）、均方根误差（root-mean-square error，RMSE）、确定系数（R^2）和交叉验证结果的 RMSE（RMSECV）为依据，从线性回归（Linear）、岭回归（Ridge）、Lasso 回归、随机森林回归（Random Forest）、支持向量机回归（Support Vector）、梯度提升回归（Gradient Boosting）、决策树回归（Decision tree）和 K 最邻近回归（k-Nearest Neighbor，KNN）等多种机器学习模型中，筛选出能预测花色苷含量的最佳建模算法。为提高模型性能，建模过程使用迭代 5 折交叉验证法[16]。

1.3　数据处理

结果表示为平均值±标准偏差。利用 R 语言 agricolae 包进行基于 Tukey HSD 法的方差分析。利用 R 语言 mixOmics 包进行 PCA 和 sPLS 分析。采用 Excel 进行线性回归建模与预测。机器学习模型的建立与预测使用 Python 的 scikit-learn 库实现。

2　结果与分析

2.1　胭脂萝卜色泽数据集特征

由表 1 可知，胭脂萝卜具有较高的 L^* 值、a^* 值和 C^* 值，但 b^* 值接近 0，且大多数样品的 H^* 值小于 0。这些数据表明本研究采用的胭脂萝卜呈明显红色，与先前的研究报道[17] 和视觉观察结果一致（图 1）。然而，不同样本 a^* 值差异较大，导致据 a^* 值计算的 C^* 值和 H^* 值也存在较大不同。即便在同一市场购买的同一品种胭脂萝卜，其 a^* 值依然差异较大。说明本研究采用的胭脂萝卜呈现出不同程度红色，这可能与胭脂萝卜个体生长条件或遗传基因存在细微差别有关[4]。PCA 是最广泛用于揭露数据关系尚不清楚样本之间相似性和隐藏模式的常用工具[18]。如图 2 所示，第一主成分和第二主成分分别解释了全部变量的

表 1　不同胭脂萝卜样本的色泽数据

Table 1　Color data of different red radish samples

样品编号	L^*	a^*	b^*	C^*	H^*
B-1-A-1	32.32±2.72	16.64±2.13	−0.31±0.9	16.66±2.15	−0.91±2.88
B-1-B-1	33.09±2.73	17.01±1.6	−1.17±0.84	17.07±1.6	−3.97±2.78
B-2-A-1	42.47±10.59	9.05±3.68	0.91±1.51	9.31±3.37	10.31±15.44
B-2-B-1	38.72±7.43	14.03±2.83	−0.8±1.12	14.1±2.82	−2.86±5.1
B-3-A-1	36.03±4.78	12.7±3.44	−2.47±2.87	13.25±3.32	−10.79±13.15
B-3-B-1	37.13±6.18	14.26±1.89	−2.4±1.39	14.51±1.97	−9.33±5.11
B-4-A-1	32.27±2.32	15.35±2.85	−1.55±1.4	15.49±2.81	−4.93±6.76
B-4-B-1	32.88±2.02	16.01±2.43	−2.06±0.82	16.15±2.46	−7.24±2.73
B-5-A-1	35.96±2.9	18.72±1.95	−2.12±0.72	18.85±1.98	−6.45±1.95
B-5-B-1	33.15±5.11	17.96±2.18	−1.91±1.18	18.09±2.22	−5.92±3.59
B-6-A-1	38.12±8.07	13.77±2.21	−0.87±0.57	13.81±2.21	−3.59±2.32
B-6-B-1	34.17±5.51	14.94±1.62	−1.42±0.85	15.02±1.65	−5.33±3.01
M-1-A-1	34.26±2.69	16.17±2.94	−0.72±1.16	16.22±2.93	−2.71±4.12
M-1-B-1	31.52±1.41	17.77±1.53	0.34±0.82	17.79±1.51	1.25±2.85
M-2-A-1	30.01±2.19	14.82±1.07	0.37±0.84	14.85±1.07	1.46±3.26
M-2-B-1	30.08±1.13	14.61±1.64	0.35±0.68	14.63±1.62	1.49±2.85
M-3-A-1	28.59±1.32	12.47±1.55	−0.64±0.83	12.5±1.58	−2.6±3.68
M-3-B-1	29.3±0.93	13.63±1.53	−1.11±0.61	13.69±1.54	−4.56±2.51
M-4-A-1	36.92±6.21	11.22±4.47	−1.12±1.61	11.36±4.53	−4.58±7.66
M-4-B-1	38.42±4.67	8.8±4.2	−0.49±1.34	8.91±4.19	−0.1±11.13
M-5-A-1	32.79±2	18.63±1.3	−1.09±0.56	18.67±1.31	−3.34±1.6
M-5-B-1	35.13±3.45	17.94±2.37	−1.64±0.82	18.03±2.38	−5.23±2.5
S-1-A-1	36.41±6.1	11.25±3.63	−0.16±1.91	11.44±3.47	2.1±13.25
S-1-B-1	38.71±5.17	11.21±3.44	−0.57±1.02	11.26±3.47	−1.97±4.86
S-2-A-1	38.33±3.89	8.04±4.58	−0.63±1.51	8.17±4.61	0.2±12.55
S-2-B-1	38.48±4.97	7.99±4.64	−0.49±1.69	8.17±4.61	3.56±16.87
S-3-A-1	32.59±3.36	10.8±3.99	−2.2±1.58	11.07±4.16	−10.89±5.24
S-3-B-1	32.35±2.87	10.23±4.38	−1.73±1.17	10.4±4.48	−8.29±5.43
Y-1-A-1	30.72±1.66	12.57±1.74	−0.97±0.67	12.62±1.76	−4.11±2.96
Y-1-B-1	30.28±1.56	11.94±1.75	−0.68±0.65	11.98±1.75	−3.28±3.03
Y-2-A-1	30.35±2.04	15.1±1.59	−1.41±0.74	15.17±1.63	−5.24±2.54
Y-2-B-1	30.1±1.69	14.39±1.81	−1.86±0.67	14.52±1.82	−7.36±2.51
Y-3-A-1	28.5±1.62	14.47±2.39	−0.86±0.55	14.51±2.4	−3.24±2.03
Y-3-B-1	29.05±1.54	13.34±3.26	−0.73±1.17	13.39±3.32	−2.29±4.3
Y-4-A-1	28.19±1.06	14.42±1.15	−1.13±1.19	14.5±1.24	−4.23±4.28
Y-4-B-1	29.61±1.24	16.94±1.26	−0.98±0.87	16.99±1.24	−3.34±3.13
Y-5-A-1	35.02±2.8	17.27±1.35	−2.08±0.45	17.4±1.32	−6.92±1.75
Y-5-B-1	35.01±5.54	16.32±2.08	−1.69±1.32	16.45±2.15	−5.68±4.24

注：样品编号中第一个字母表示样品的来源，B表示百汇广场超市，M表示马鞍街道农贸市场，S表示蔬果园超市，Y表示亿万家超市；第二位的数字表示第几个样品；第三位的字母表示一个胭脂萝卜从中间剖开后的A面与B面。

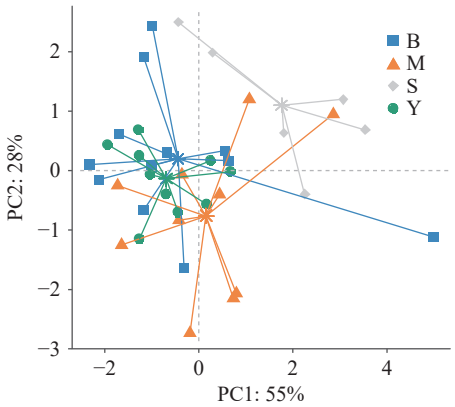


图 2　不同胭脂萝卜样本色泽数据的主成分分析得分图

Fig.2　Score plots of PCA for color data of different red radish samples

注：B 表示百汇广场超市，M 表示马鞍街道农贸市场，S 表示蔬果园超市，Y 表示亿万家超市。

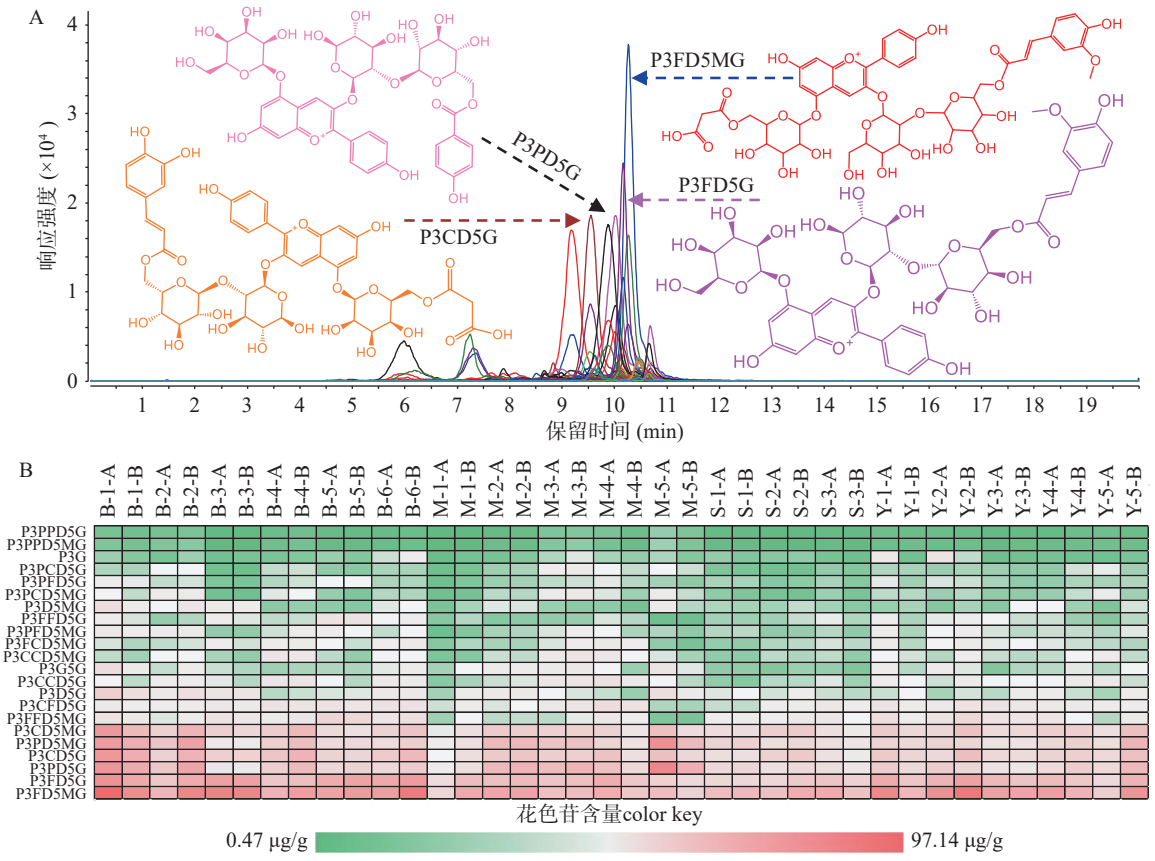
55% 和 28%。此外，购买自“蔬果园”超市的胭脂萝卜与购买自其他市场的萝卜聚类在不同位置，说明“蔬果园”超市胭脂萝卜的色泽与其他市场胭脂萝卜的色泽具有较大差异(图 2)。购买自百汇广场超市、马鞍街道农贸市场和亿万家超市的胭脂萝卜聚集在左下方，表明它们的色泽较为接近。但即便如此，购买自同一市场的胭脂萝卜并未聚类在同一点，而是散落于坐标系各象限，说明同一市场胭脂萝卜色泽也具有明显差异。因此，从表 1 和图 2 可知，本研究所选样品的色泽呈红色，但个体差异较大，确保了样品色泽多样性，可以用于构建高适应性模型。

2.2 胭脂萝卜花色苷单体的组成特征及含量

本研究利用 UPHLC-QqQ-MS/MS 从胭脂萝卜中鉴定出 22 种花色苷，包括 3 种非酰化天竺葵苷和 19 种酰化天竺葵苷(图 3A、B)。花色苷湿基含量介于 0.47 μg/g 到 97.14 μg/g(图 3B)，与本团队先前的检测结果(0.05~27.74 μg/g)基本一致[19]。天竺葵素-3-(咖啡酰)二葡萄糖苷-5-(丙二酰)葡萄糖苷(P3CD5MG)、天竺葵素-3-(对香豆酰)二葡萄糖苷-5-(丙二酰)葡萄糖苷(P3PD5MG)、天竺葵素-3-(咖啡酰)二葡萄糖苷-5-葡萄糖苷(P3CD5G)、天竺葵素-3-(对香豆酰)二葡萄糖苷-5-葡萄糖苷(P3PD5G)、天竺葵素-3-(咖啡酰)二葡萄糖苷-5-葡萄糖苷(P3FD5G)和天竺葵素-3-(阿魏酰)二葡萄糖苷-5-(丙二酰)葡萄糖苷(P3FD5MG)的含量显著(P<0.05)高于其他花色苷(图 3B、C)。其中，P3FD5MG 平均含量最高可达到 49.73±19.66 μg/g，是胭脂萝卜最主要花色苷。这与之前的研究结果一致[14,17,19]。然而，不同胭脂萝卜的 P3FD5MG 含量差异巨大，最高含量甚至是最低含量的 5.14 倍(图 3B、C)。此外，单体花色苷 PCA 得分图显示，不同个体分布在不同象限，表明本研究中胭脂萝卜的花色苷含量具有明显个体差异(图 3D)。许多研究要求平行样品间不能存在显著差异，这可能存在离群值。但这种离散程度较大的数据集为建模奠定了基础[20]。此外，该数据亦表明胭脂萝卜群体中存在花色苷含量显著偏低的个体，此类材料不适合用于天然色素生产。因此，建立一种快速检测胭脂萝卜花色苷含量的方法，对于高效率筛选适宜用于色素生产的优质材料具有重要意义。

2.3 色泽与花色苷含量的相关性

在色泽参数与花色苷相关性矩阵图中(图 4)，红色圈越大，正相关性越强；蓝色圈越大，负相关性越强。虽然大多数花色苷含量与 a^*值或 C^*值具有正相关性，但相关系数值大约仅有 0.5。此外，a^*值呈现出了与 C^*值非常类似的相关系数。这是因为 C^*值是 a^*值和 b^*值的函数，且胭脂萝卜 b^*值接近 0，导致 C^*值主要受 a^*值影响[15]。这些结果表明 a^*值和 C^*值与花色苷含量之间具有较强正相关性，证实了花色苷是赋予胭脂萝卜红色的重要物质，可以尝试用色泽参数构建预测花色苷含量的模型。这一发现与李培嗣等[21]发现山楂果肉花色苷含量与果肉 a^*值呈强正相关的结果一致。此外，P3PD5G、P3PD5MG、P3G5G 和总花色苷含量与 a^*值和 C^*值之间具有更强的正相关性，说明花色苷分子构型对色泽有较大



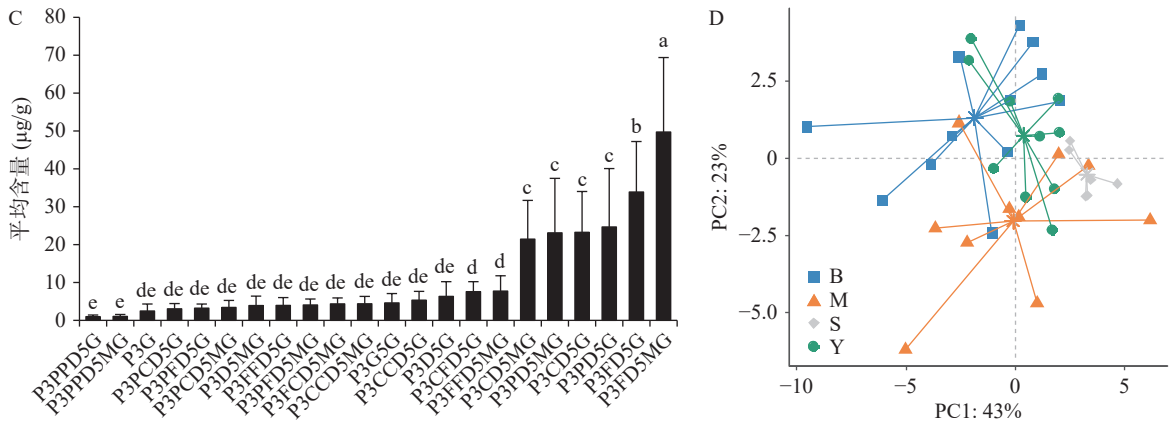


图 3　UHPLC-QqQ-MS/MS 测定胭脂萝卜的单体花色苷含量

Fig.3　Content of individual anthocyanins in red radish was determined by UHPLC-QqQ-MS/MS

注：(A)样品花色苷的 UHPLC-QqQ-MS/MS 色谱图；(B)所有样品花色苷的含量，图中不同颜色代表不同含量；(C)所有样品中单体花色苷含量平均值，不同字母标注表示 P<0.05；(D)花色苷含量的 PCA 得分图。P3PPD5G：天竺葵素-3-(对香豆酰)(对香豆酰)二葡萄糖苷-5-葡萄糖苷；P3PPD5MG：天竺葵素-3-(对香豆酰)(对香豆酰)二葡萄糖苷-5-(丙二酰)葡萄糖苷；P3G：天竺葵素-3-葡萄糖苷；P3PCD5G：天竺葵素-3-(对香豆酰)(咖啡酰)二葡萄糖苷-5-葡萄糖苷；P3PFD5G：天竺葵素-3-(对香豆酰)(咖啡酰)二葡萄糖苷-5-葡萄糖苷；P3PCD5MG：天竺葵素-3-(对香豆酰)(咖啡酰)二葡萄糖苷-5-(丙二酰)葡萄糖苷；P3D5MG：天竺葵素-3-二葡萄糖苷-5-(丙二酰)葡萄糖苷；P3FFD5G：天竺葵素-3-(阿魏酰)(阿魏酰)二葡萄糖苷-5-葡萄糖苷；P3PFD5MG：天竺葵素-3-(对香豆酰)(阿魏酰)二葡萄糖苷-5-(丙二酰)葡萄糖苷；P3FCD5MG：天竺葵素-3-(阿魏酰)(阿魏酰)二葡萄糖苷-5-(丙二酰)葡萄糖苷；P3CCD5MG：天竺葵素-3-(咖啡酰)(咖啡酰)二葡萄糖苷-5-(丙二酰)葡萄糖苷；P3G5G：天竺葵素-3-葡萄糖苷-5-葡萄糖苷；P3CCD5G：天竺葵素-3-(咖啡酰)(咖啡酰)二葡萄糖苷-5-葡萄糖苷；P3D5G：天竺葵素-3-二葡萄糖苷-5-葡萄糖苷；P3CFD5G：天竺葵素-3-(咖啡酰)(阿魏酰)二葡萄糖苷-5-葡萄糖苷；P3FFD5MG：天竺葵素-3-(阿魏酰)(阿魏酰)二葡萄糖苷-5-(丙二酰)葡萄糖苷；P3CD5MG：天竺葵素-3-(咖啡酰)二葡萄糖苷-5-(丙二酰)葡萄糖苷；P3PD5MG：天竺葵素-3-(对香豆酰)二葡萄糖苷-5-(丙二酰)葡萄糖苷；P3CD5G：天竺葵素-3-(咖啡酰)二葡萄糖苷-5-葡萄糖苷；P3PD5G：天竺葵素-3-(对香豆酰) 二葡萄糖苷-5-葡萄糖苷；P3FD5G：天竺葵素-3-(咖啡酰)二葡萄糖苷-5-葡萄糖苷；P3FD5MG：天竺葵素-3-(阿魏酰)二葡萄糖苷-5-(丙二酰)葡萄糖苷。

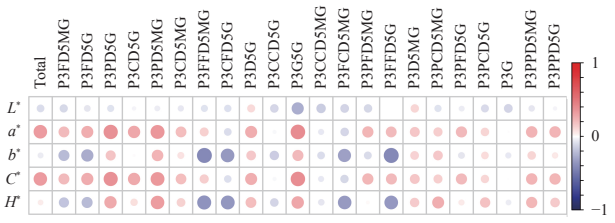


图 4　颜色参数与花色苷单体含量的相关性矩阵

Fig.4　Correlation plot of color parameters with anthocyanin monomer content

注：颜色标尺表示皮尔森相关系数，计算方法为 sPLS。

影响。先前研究表明，天竺葵素分子的 C3 和 C5 位糖基化修饰能显著提升红色强度[22]，酰化糖基化修饰能通过共色效应进一步增强红色[23]。虽然 P3CD5MG、P3CD5G、P3FD5G 和 P3FD5MG 也是胭脂萝卜中主要花色苷(图 3)，但它们与色泽的皮尔森相关系数值不高(图 4)。因此，本文推测，这些花色苷分子可能与胭脂萝卜中酚酸、金属离子等成分发生了共色或络合，从而导致了颜色偏离、强度变化和色调改变[24−25]。

2.4　线性回归模型的建立与验证

鉴于 a^*值和 C^*值都与花色苷含量呈正相关性，本研究首先尝试利用线性回归方法建立预测模型。通过直线性方程(y=ax+b)，将所有样本 a^*值或 C^*值与总花色苷含量和 3 种主要花色苷含量(P3FD5MG、P3FD5G、P3PD5G)进行线性拟合。a^*值和 C^*值构建模型的 R^2 分别为 0.043~0.142 和 0.045~0.138，基本一致。然而，C^*值依据 a^*和 b^*值计算获得，当面临大量样品在线检测时，需要更多算力维持系统快速响应。因此，在后文主要评估直接利用 a^*值构建模型的预测能力。从图 5 可以看出，直线性模型对 26 个样本的总花色苷含量预测误差小于 25.0%，占总样本量的 68.4%。然而，直线性模型针对单体花色苷含量的预测准确率远小于对总花色苷含量的预测准确率。具体而言，直线性模型对 P3PD5G、P3FD5G 和 P3FD5MG 含量的预测准确性高于 75% 的样品个数分别占 28.9%、52.6% 和 44.7%(图 5)。这可能是每种花色苷都为胭脂萝卜贡献了部分红色。此外，直线性模型对单个花色苷含量的预测准确性差异可能与花色苷化学结构的光谱响应异质性有关。例如，酰化修饰(如丙二酰)可能通过花色苷分子内共色或电荷转移效应[26]，导致可见光吸收峰偏移[27]，从而降低 a^*值对特定异构体的预测能力。这表明基于单一色泽参数构建的简单线性回归模型无法准确预测花色苷含量。邵雅芳等[28]在利用颜色参数判定黑米花色苷含量研究中，强调单一参数(如仅用 a^*值)无法覆盖不同结构花色苷，需要更多色泽参数协同建模，才能有效提高模型预测精度。总之，仅利用直线性模型和 a^*值不能准确预测胭脂萝卜花色苷含量。

2.5　机器学习模型的建立与验证

为进一步提高预测精度，本研究利用 8 种机器

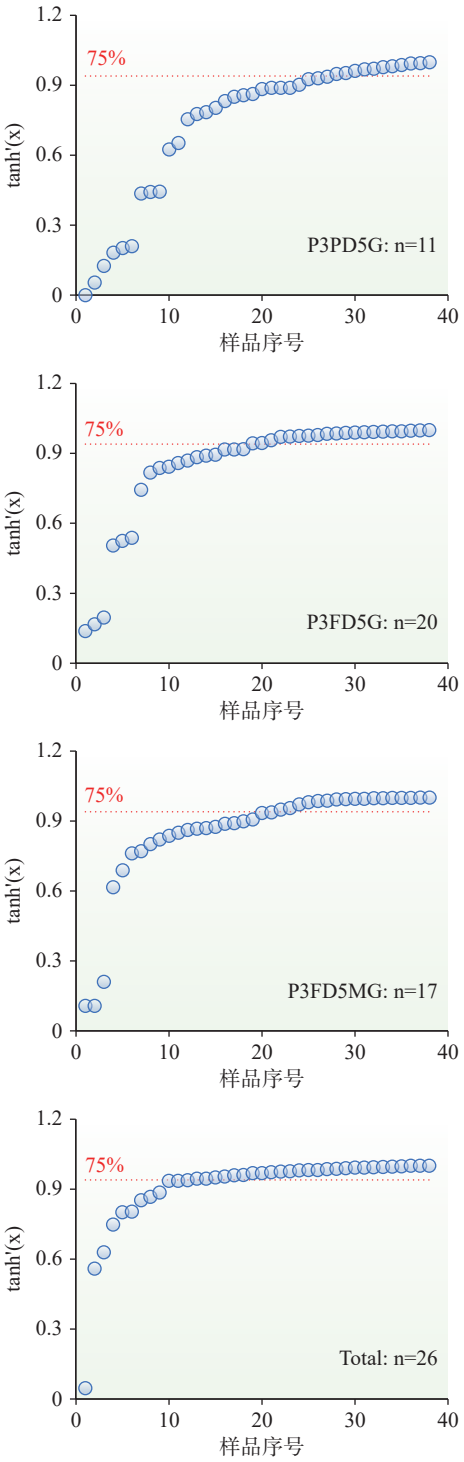


图 5　利用线性回归方程构建的模型和萝卜色泽参数预测花色苷含量

Fig.5　Model constructed using a linear regression equation and the color parameters of radish employed to predict the anthocyanin content

注: n 为模型预测花色苷含量准确率高于 75% 的样品数量。

学习算法, 针对训练集构建了预测模型。如表 2 所示, 8 种机器学习模型在预测总花色苷和单体花色苷含量时最大 R^2 值均高于前文所述的简单线性模型 R^2 值, 机器学习模型预测准确度更高。部分模型的 R^2 为负值, 这说明该模型的预测准确性甚至低于简单的平均值计算结果, 应该剔除这些模型。此外, 在

表 2　8 种机器学习模型预测胭脂萝卜花色苷含量的性能

Table 2　Capabilities of eight machine learning models in predicting the anthocyanin content of red radish

回归模型	MAE	MSE	RMSE	R^2	RMSE (Cross-Validation)
花色苷总含量					
Linear	34.56	2231	47.23	0.3089	101.45
Ridge	35.60	2190	46.80	0.3216	86.73
Lasso	36.10	2213	47.04	0.3145	87.21
Random Forest	44.60	3762	61.33	−0.1654	91.98
Support Vector	39.90	2963	54.44	0.0819	78.57
Gradient Boosting	39.05	2365	48.63	0.2672	99.37
Decision Tree	33.62	1767	42.04	0.4524	73.09
k-Nearest Neighbor	61.53	5217	72.23	−0.6163	119.84
P3FD5MG					
Linear	13.35	300	17.32	0.0138	24.95
Ridge	12.98	288	16.96	0.0545	20.79
Lasso	12.76	295	17.17	0.0311	21.20
Random Forest	13.78	257	16.02	0.1565	23.26
Support Vector	13.47	267	16.33	0.1233	20.34
Gradient Boosting	11.92	275	16.57	0.0976	22.25
Decision Tree	9.63	200	14.15	0.3416	21.79
k-Nearest Neighbor	13.60	265	16.28	0.1293	32.53
P3FD5G					
Linear	10.23	150.88	12.28	0.0192	16.69
Ridge	10.66	147.40	12.14	0.0418	14.29
Lasso	11.96	180.11	13.42	−0.1708	13.81
Random Forest	7.21	76.82	8.76	0.5006	15.12
Support Vector	10.63	178.94	13.38	−0.1633	13.61
Gradient Boosting	8.51	104.31	10.21	0.3219	16.59
Decision Tree	4.73	54.20	7.36	0.6476	15.73
k-Nearest Neighbor	8.14	146.65	12.11	0.0467	22.98
P3PD5G					
Linear	8.50	128	11.32	−0.5720	18.34
Ridge	9.90	115	10.72	−0.4093	14.91
Lasso	9.95	120	10.94	−0.4700	15.25
Random Forest	11.93	231	15.20	−1.8367	19.34
Support Vector	6.38	54	7.34	0.3384	14.72
Gradient Boosting	13.02	185	13.59	−1.2671	15.72
Decision Tree	17.04	527	22.95	−5.4648	17.84
k-Nearest Neighbor	9.78	111	10.54	−0.3640	16.10

模型筛选时 MAE、MSE、RMSE 和交叉验证 RMSE 值更低的模型可能预测效果更好, 特别是交叉验证 RMSE 值较低的模型。例如预测总花色苷含量的决策树回归、支持向量机回归、岭回归和 Lasso 回归模型; 预测 P3FD5MG 含量的岭回归、支持向量机回归、Lasso 回归和决策树回归模型; 预测 P3FD5G 含量的岭回归、梯度提升、随机森林回归和决策树回归模型; 预测 P3PD5G 含量的支持向量机回归模型。然而, 仅依据验证集的拟合与交叉验证结果, 有时筛选出的模型在应用时的预测能力并不理想。因此, 需要进一步在验证集上验证模型的预测能力。

通常预测准确率设定 75% 以上作为目标, 可以确保模型在大多数情况下具有足够的可靠性和泛化能力, 能满足许多实际应用的需求[29]。例如利用深度学习模型预测豆浆原料加工潜力的准确率为 78%[30]。从图 6 可以看出, 模型的验证集预测精度与训练集结果不一致, 再次说明模型筛选中验证集的重

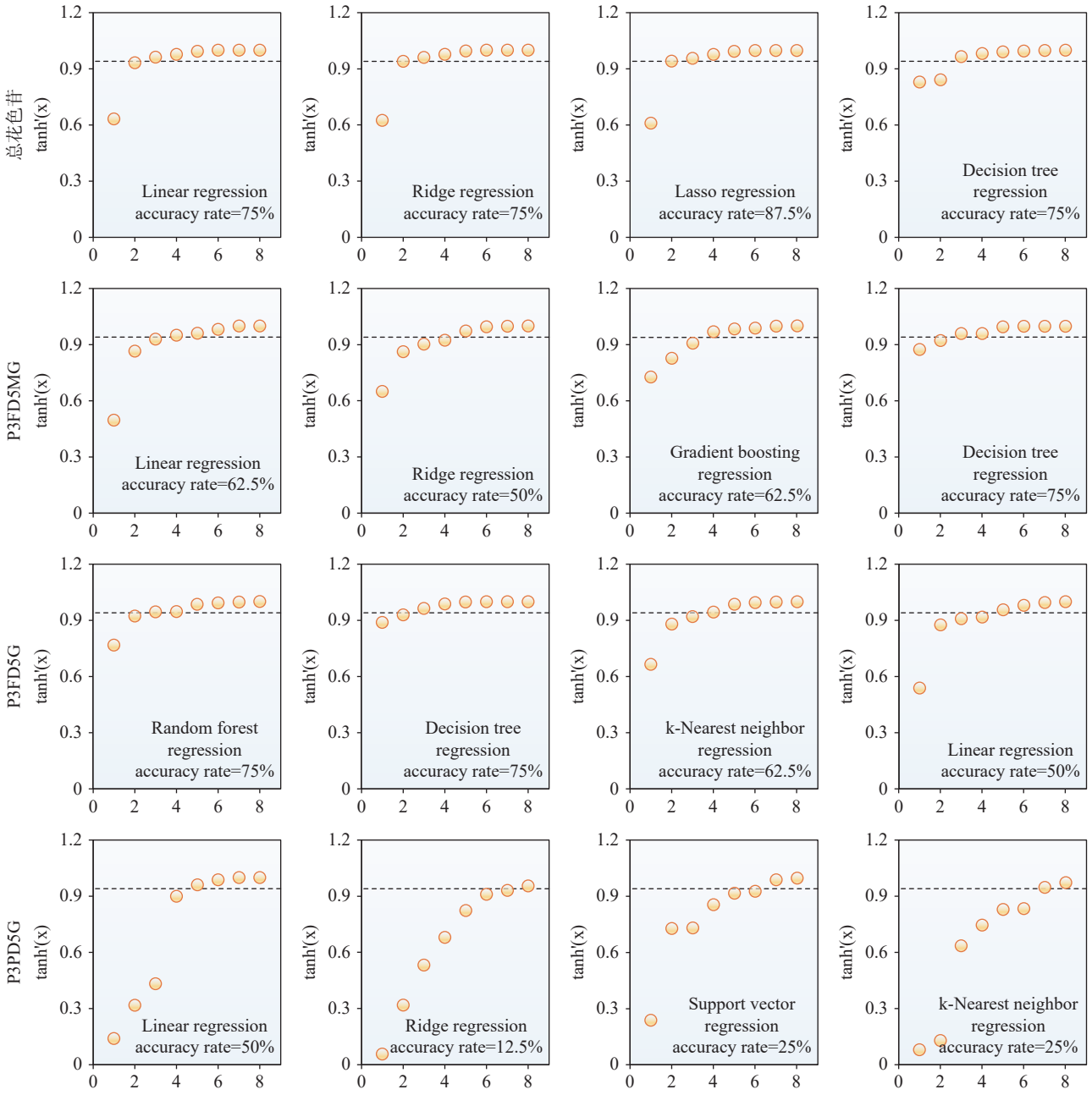


图 6　在验证集上对花色苷含量预测效果最好的 4 种机器学习模型

Fig.6　Four machine learning models with the best prediction effect on anthocyanin content on the validation set

要性。线性回归、岭回归、Lasso 回归和决策树回归等机器学习模型对总花色苷的预测准确率均超过了 75%，甚至 Lasso 回归机器学习模型的预测准确率达到了 87.5%，是预测总花色苷含量的最佳模型（图 6）。同样，决策树回归模型对于 P3FD5MG 和 P3FD5G 含量的预测准确率最高达 75.0%。虽然随机森林模型在预测 P3FD5G 含量时准确率也能达到 75.0%，但在准确率低于 75.0% 的样品中，决策树回归模型的预测结果更接近真实值。然而，虽然支持向量机回归模型对 P3PD5G 含量的 R^2 为正，但该模型对验证集数据的预测准确度仅为 25%。整体上，决策树回归模型在预测总花色苷含量、P3FD5MG 含量和 P3FD5G 含量时准确率均达到 75.0%。这一发现与余媛媛等[31]的研究结果一致，决策树回归模型更直观且不受变量间共线性影响。这些结果也说明，机器学习模型能够较准确地预测花色苷含量，不仅克服了单一色泽参数（a^*值）预测花色苷含量的局限性，还通过整合多色泽参数增强了预测能力和泛化能力。

3　结论

本研究通过分析胭脂萝卜色泽参数与花色苷含量相关性，发现胭脂萝卜 a^*值和 C^*值均与花色苷含量呈正相关关系，但仅利用 a^*值构建的线性回归模型在预测花色苷含量时最高准确度仅有 68.4%。利用 Lasso 回归构建的机器学习模型预测总花色苷含量的准确率达 87.5%。此外，基于决策树回归的机器学习模型对总花色苷、P3FD5MG 和 P3FD5G 含量的预测准确率均达 75%。因此，利用色泽参数构

建的机器学习模型能用于预测胭脂萝卜花色苷含量，从而实现胭脂萝卜色素生产原料的快速智能化筛选。

© The Author(s) 2026. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by-nc-nd/4.0/).

参考文献

[1] 董楠, 宋会歌, 刘嘉, 等. 咖啡酸对胭脂萝卜红色素辅色作用及稳定性的影响[J]. 食品科学, 2011, 32(7): 61−64. [DONG Nan, SONG Huige, LIU Jia, et al. Effect of caffeic acid on co-pigmentation and stability of kermes radish red pigment[J]. Food Science, 2011, 32(7): 61−64.]

[2] 王二雷, 黄佳莹, 段海章, 等. 花色苷稳态化技术研究进展及应用前景[J]. 食品工业科技, 2024, 45(18): 394−403. [WANG Erlei, HUANG Jiaying, DUAN Haizhang, et al. Progress on the stabilization technology of anthocyanins and the application prospects[J]. Science and Technology of Food Industry, 2024, 45(18): 394−403.]

[3] 黄方, 宁楠, 倪春蕾, 等. 不同多酚对蓝莓花色苷单体辅色效果的研究[J]. 食品工业科技, 2024, 45(24): 75−81. [HUANG Fang, NING Nan, NI Chunlei, et al. Effects of different polyphenols on the co-pigmentation of blueberry anthocyanin monomers[J]. Science and Technology of Food Industry, 2024, 45(24): 75−81.]

[4] 芮春梅. 胭脂萝卜红色素形成的动态分析与提取方法及稳定性研究[D]. 重庆: 西南大学, 2007. [RUI Chunmei. Study on the red pigment extracted from carmine radish[D]. Chongqing: Southwest University, 2007.]

[5] 董全, 李洪军, 尚永彪, 等. 食用天然色素: 萝卜红研究进展[J]. 保鲜与加工, 2004, 4(6): 9−11. [DONG Quan, LI Hongjun, SHANG Yongbiao, et al. Research and application of red radish pigment as natural food colorants[J]. Storage and Process, 2004, 4(6): 9−11.]

[6] 熊玥, 孟余燕, 张伦, 等. 萝卜红色素的研究进展及未来展望[J]. 现代食品, 2018(19): 130−133,136. [XIONG Yue, MENG Yuyan, ZHANG Lun, et al. The nature and research progress of radish red pigment[J]. Modern Food, 2018(19): 130−133,136.]

[7] LI Wenfeng, GU Mengyuan, GONG Pengling, et al. Glycosides changed the stability and antioxidant activity of pelargonidin[J]. LWT, 2021, 147: 111581.

[8] PARK N I, XU H, LI X H, et al. Anthocyanin accumulation and expression of anthocyanin biosynthetic genes in radish (*Raphanus sativus*)[J]. Journal of Agricultural and Food Chemistry, 2011, 59(11): 6034−6039.

[9] 翦祎, 韩舜愈, 张波, 等. 单一 pH 法、pH 示差法和差减法快速测定干红葡萄酒中总花色苷含量的比较[J]. 食品工业科技, 2012, 33(23): 323−325,423. [JIAN Yi, HAN Shunyu, ZHANG Bo, et al. Comparison of single pH method, pH-differential method and substraction method for determining content of anthocyanins from red wine[J]. Science and Technology of Food Industry, 2012, 33(23): 323−325,423.]

[10] 刘恩芬, 孙琳, 陈子晔, 等. 黑枸杞与桑葚花色苷纯化与结构分析[J]. 食品工业科技, 2024, 45(13): 257−264. [LIU Enfen, SUN Lin, CHEN Ziye, et al. Purification and structural analysis of anthocyanins from black wolfberry and mulberry[J]. Science and Technology of Food Industry, 2024, 45(13): 257−264.]

[11] 庞文倩, 刘春菊, 李大婧, 等. 热压加工过程中紫玉米花色苷的热降解动力学及色泽变化[J]. 江苏农业学报, 2023, 39(7): 1583−1592. [PANG Wenqian, LIU Chunju, LI Dajing, et al. Thermal degradation kinetics of anthocyanins and color change in purple corn during hot pressing process[J]. Jiangsu Journal of Agricultural Sciences, 2023, 39(7): 1583−1592.]

[12] BALDASSI C, LEE C, DOSSETT M, et al. High-throughput color determination of red raspberry puree and correlation of color parameters with total anthocyanins[J]. Plant Methods, 2024, 20(1): 78.

[13] KHAN M I H, SABLANI S S, NAYAK R, et al. Machine learning-based modeling in food processing applications: State of the art[J]. Comprehensive Reviews in Food Science and Food Safety, 2022, 21(2): 1409−1438.

[14] LI Wenfeng, LIANG Chan, BAO Fangtian, et al. Chemometric analysis illuminates the relationship among browning, polyphenol degradation, Maillard reaction and flavor variation of 5 jujube fruits during air-impingement jet drying[J]. Food Chemistry, 2024, 22: 101425.

[15] LI Wenfeng, GONG Pengling, XU Mengyi, et al. Isolation and characterization of the anthocyanins derived from red radishes (*Raphanus sativus* L.) and the protective ability of *β*-lactoglobulin against heat-induced oxidation[J]. Journal of Food Science, 2022, 87(4): 1586−1600.

[16] SHAHANI N M, ZHENG X G, GUO X W, et al. Machine learning-based intelligent prediction of elastic modulus of rocks at thar coalfield[J]. Sustainability, 2022, 14(6): 3689.

[17] LI Wenfeng, ZHANG Yaxi, DENG Hanlu, et al. *In vitro* and *in vivo* bioaccessibility, antioxidant activity, and color of red radish anthocyanins as influenced by different drying methods[J]. Food Chemistry, 2023, 18: 100633.

[18] 李文峰, 陶雯, 陈露红, 等. 可溶性大豆多糖改善左旋肉碱诱导的小鼠小肠首过代谢[J]. 食品科学, 2019, 40(13): 187−194. [LI Wenfeng, TAO Wen, CHEN Luhong, et al. Soluble soybean polysaccharides ameliorate L-carnitine-induced first-pass metabolism in the small intestine of mice[J]. Food Science, 2019, 40(13): 187−194.]

[19] LI Wenfeng, PANG Xinxin, XIAO Jinfeng, et al. Degradation kinetics of pelargonidin-3-(p-coumaroyl)diglucoside-5-(malonyl)glucoside and pelargonidin-3-(feruloyl)diglucoside-5-(malonyl)glucoside in red radish during air-impingement jet drying[J]. LWT, 2020, 127: 109390.

[20] 刘耘汀. 基于机器学习的肠道菌群疾病诊断及个体识别建模研究[D]. 兰州: 兰州大学, 2023. [LIU Yunting. Disease diagnosis and individual identification modeling based on machine learning and gut microbiota[D]. Lanzhou: Lanzhou University, 2023.]

[21] 李培嗃, 孙馨宇, 王键, 等. 山楂种质资源果实颜色与花青苷含量的关系[J]. 植物遗传资源学报, 2024, 25(1): 72−83. [LI Peihao, SUN Xinyu, WANG Jian, et al. Relationship of fruit color and anthocyanin content of hawthorn germplasm resources[J]. Journal of Plant Genetic Resources, 2024, 25(1): 72−83.]

[22] CAI Dongbao, LI Xusheng, CHEN Jiali, et al. A comprehensive review on innovative and advanced stabilization approaches of anthocyanin by modifying structure and controlling environmental factors[J]. Food Chemistry, 2022, 366: 130611.

［23］ JOKIOJA J, YANG B R, LINDERBORG K M. Acylated anthocyanins: A review on their bioavailability and effects on postprandial carbohydrate metabolism and inflammation［J］. Comprehensive Reviews in Food Science and Food Safety, 2021, 20(6): 5570−5615.

［24］ CAO Y, XIA Q L, ANIYA, et al. Copigmentation effect of flavonols on anthocyanins in black mulberry juice and their interaction mechanism investigation［J］. Food Chemistry, 2023, 399: 133927.

［25］ DAI H C, FORBES A, GUO X, et al. Prediction of anthocyanin color stability against iron co-pigmentation by surface-enhanced Raman spectroscopy［J］. Foods, 2022, 11(21): 3436.

［26］ XUE Hongkun, ZHAO Jianduo, WANG Yu, et al. Factors affecting the stability of anthocyanins and strategies for improving their stability: A review［J］. Food Chemistry: X, 2024, 24: 101883.

［27］ GIUSTI M M, WROLSTAD R E. Characterization and measurement of anthocyanins by UV-visible spectroscopy［J］. Current Protocols in Food Analytical Chemistry, 2001, 00(1): F1.2. 1-F1.2. 13.

［28］ 中国水稻研究所. 利用颜色参数指标判定黑米花色苷含量的方法: 201710608018.3［P］. 2017-10-13. ［China National Rice Research Institute. Method for determining anthocyanin content in black rice by color parameter index: 201710608018.3［P］. 2017-10-13.］

［29］ 李海霞, 宋丹蕾, 孔佳宁, 等. 传统机器学习模型的超参数优化技术评估［J］. 计算机科学, 2024, 51(8): 242−255. ［LI Haixia, SONG Danlei, KONG Jianing, et al. Evaluation of hyperparameter optimization techniques for traditional machine learning models［J］. Computer Science, 2024, 51(8): 242−255.］

［30］ ZHU Guoyin, HUANG Xin, PENG Xingyun, et al. A deep learning-based quantitative prediction model for the processing potentials of soybeans as soymilk raw materials［J］. Food Chemistry, 2024, 453: 139671.

［31］ 余媛媛, 张正东, 屈波, 等. 决策树模型在老年髋部骨折术后病人 3 年内死亡风险预测中的应用［J］. 护理研究, 2023, 37(20): 3618−3624. ［YU Yuanyuan, ZHANG Zhengdong, QU Bo, et al. Application of decision tree model in predicting death risk within 3 years after hip fracture in elderly patients［J］. Chinese Nursing Research, 2023, 37(20): 3618−3624.］