

基于可见-近红外光谱的鲜食葡萄成熟品质关键指标检测

刘文政¹ 周雪健¹ 平凤娇¹ 苏媛¹ 鞠延仑¹ 房玉林¹ 杨继红^{1,2}

(1. 西北农林科技大学葡萄酒学院, 陕西杨凌 712100; 2. 新疆农业大学食品科学与药学院, 乌鲁木齐 830052)

摘要: 酚类物质是评价葡萄成熟品质的重要指标, 本文利用可见-近红外光谱技术结合化学计量学定量分析方法对葡萄皮总酚、籽总酚、皮单宁和籽单宁含量开展了无损检测研究。通过手持式可见-近红外光谱仪采集巨玫瑰葡萄波长 400 ~ 1 029 nm 范围内的漫反射光谱, 采用 SPXY 算法将其划分为校正集和预测集, 结合标准正态变换 (Standard normal variate, SNV)、多元散射校正 (Multiplicative scatter correction, MSC)、一阶导数 (First derivative, 1D)、二阶导数 (Second derivative, 2D)、Savitzky-Golay 卷积平滑 (Savitzky-Golay smoothing, SG) 和 Savitzky-Golay 卷积平滑 + 一阶导数 (SG + 1D) 6 种预处理方法以及偏最小二乘回归 (Partial least squares regression, PLSR)、支持向量机回归 (Support vector machine regression, SVR) 和卷积神经网络 (Convolutional neural network, CNN) 3 种建模算法, 分别建立了基于全波段和特征波长的葡萄皮总酚、籽总酚、皮单宁和籽单宁定量预测模型并进行综合对比分析。结果表明, 对于皮总酚、籽总酚和籽单宁, 经特征波长筛选后建立的模型效果优于全波段, 而对于皮单宁, 全波段建立的模型较特征波长效果更佳; 因此, 在预测皮总酚、籽总酚、皮单宁和籽单宁含量时, 最优模型分别为 RAW-CARS-SVR、1D-CARS-SVR、RAW-CNN 和 RAW-CARS-PLSR, 校正集相关系数 (Correlation coefficient of calibration set, R_c) 分别为 0.96、0.99、0.96 和 0.91, 预测集相关系数 (Correlation coefficient of prediction set, R_p) 分别为 0.95、0.99、0.83 和 0.89, 剩余预测偏差 (Residual predictive deviation, RPD) 分别为 3.56、7.30、1.92 和 2.25。因此, 结合可见-近红外光谱和合适的回归模型, 可以实现对巨玫瑰葡萄的皮-籽总酚、皮-籽单宁含量的无损检测。

关键词: 葡萄; 可见-近红外光谱; 成熟度; 品质检测

中图分类号: S371

文献标识码: A

文章编号: 1000-1298(2024)02-0372-12

OSID:



Detection of Key Indicators of Ripening Quality in Table Grapes Based on Visible - near-infrared Spectroscopy

LIU Wenzheng¹ ZHOU Xuejian¹ PING Fengjiao¹ SU Yuan¹ JU Yanlun¹ FANG Yulin¹ YANG Jihong^{1,2}

(1. College of Enology, Northwest A&F University, Yangling, Shaanxi 712100, China)

2. College of Food Science and Pharmacy, Xinjiang Agricultural University, Urumqi 830052, China)

Abstract: Phenolic compounds play a crucial role in assessing the internal quality of grapes and hold significant importance in this regard. The capability of visible-near-infrared (Vis - NIR) spectroscopy combined with multivariate regression models was explored to detect the contents of total phenolics and tannins in grape skins and seeds. Reflectance spectra data of Muscat Kyoho grapes were collected within the wavelength range of 400 nm to 1 029 nm, and the samples were divided into correction set and prediction set by SPXY algorithm. Six commonly used preprocessing methods such as standard normal variate (SNV), multiplicative scatter correction (MSC), first derivative (1D), second derivative (2D), Savitzky-Golay smoothing (SG) and SG + 1D were applied to the spectral data, and the competitive adaptive reweighted sampling algorithm (CARS) was utilized to select informative wavelengths. The quantitative models for comprehensive analysis of total phenolics and tannins in grape skins and seeds based on full spectra and effective wavelengths were established by partial least squares regression (PLSR), support vector machine regression (SVR), and convolutional neural network (CNN). The results showed that for the total phenolics in grape skins, total phenolics and tannins in

收稿日期: 2023-12-25 修回日期: 2024-01-16

基金项目: 国家自然科学基金项目 (32201678) 和中央高校基础科研业务费专项资金项目 (2452020201)

作者简介: 刘文政 (1991—), 男, 讲师, 博士, 主要从事葡萄与葡萄酒工程和无损检测技术研究, E-mail: lwzheng@nwfau.edu.cn

通信作者: 杨继红 (1975—), 女, 副教授, 主要从事葡萄与葡萄酒工程技术研究, E-mail: yangjihong@nwsuaf.edu.cn

grape seeds, the models on the basis of effective wavelengths performed better than those with full spectra data. While for the tannins in grape skins, the models constructed with full spectra yielded better results than the feature wavelength-selected models. The optimal models for the total phenolics and tannins in grape skins and seeds were RAW - CARS - SVR, 1D - CARS - SVR, RAW - CNN and RAW - CARS - PLSR, respectively. The correlation coefficient of calibration set (R_c) were 0.96, 0.99, 0.96 and 0.91, the correlation coefficient of prediction set (R_p) were 0.95, 0.99, 0.83 and 0.89, the residual predictive deviation (RPD) were 3.56, 7.30, 1.92 and 2.25, respectively. Therefore, the developed method could realize the non-destructive detection of the contents of total phenolics and tannins in grape skins and seeds.

Key words: grape; visible - near-infrared spectroscopy; ripeness; quality detection

0 引言

葡萄(*Vitis vinifera* L.)及其相关产业在农业经济中发挥着重要的作用,其丰富的维生素含量和独特的风味深受人们喜爱。葡萄用途广泛,可鲜食、制干、制备果汁以及葡萄酒等^[1-2]。成熟度是水果品质的重要评价指标,直接影响果实产量和品质,也决定着后期果实加工、贮藏方式^[3]。酚类物质是葡萄中非常重要的一类活性物质,主要存在于葡萄籽和葡萄皮中,且其含量受成熟度的影响较大^[4-5]。因此酚类物质是判断葡萄成熟品质的重要指标^[6]。目前,葡萄成熟度判定以生理性成熟和技术性成熟为主,即通过果实颜色、质地、糖度和酸度等,虽有一定的参考价值,但不够全面,酚类物质对于成熟期果实品质影响更为显著^[7]。然而,当前酚类物质检测多采用传统理化检测方法,繁琐耗时,难以及时、准确了解葡萄品质,且破坏样品,造成样品浪费和一定程度的经济损失。因此,开展基于新型检测技术的酚类物质快速无损检测技术研究具有重要意义^[8]。

近年来,化学计量学的引入为光谱技术评估水果内部品质提供了广阔的应用前景。通过分析水果的光谱特征,可以提取与品质指标相关的信息^[9-10]。与传统的物理化学测量方法相比,该方法具有简便、快速、可靠、无损等优点。在水果品质和成熟度检测方面,国内外学者已开展大量研究。GAO 等^[11]利用高光谱成像技术(450 ~ 1 000 nm),将连续投影算法特征波长提取的光谱与用主成分分析算法降维后的图像信息融合,构建了基于高光谱图像信息融合的红地球葡萄可溶性固形物含量的偏最小二乘预测模型,研究结果表明该技术在完整评估葡萄内部品质方面具有巨大的潜力。DANIELS 等^[12]利用近红外光谱技术(800 ~ 2 500 nm)建立了 3 个葡萄品种技术成熟度和感官品质参数的偏最小二乘预测模型,为无损和无接触测定完整葡萄串内部成熟度参数奠定基础。YU 等^[13]利用自行设计的近红外检测探头,对葡萄可溶性固形物含量进行无

损检测,并利用偏最小二乘法建立了葡萄可溶性固形物含量与光谱之间的预测模型。研究结果表明,利用测试探针预测葡萄可溶性固形物(SSC)是可行的,并可为整串葡萄的无损检测提供参考。BASILE 等^[14]利用近红外光谱(833 ~ 2 500 nm)结合偏最小二乘模型和区间偏最小二乘模型建立了鲜食葡萄硬度和可溶性固形物的预测模型,得到的可溶性固形物预测模型效果较好,可以为水果行业开发高效、低成本的检测仪器提供参考。由此表明光谱技术在评估水果成熟度方面具有巨大潜力。然而,通过文献发现,目前基于可见-近红外光谱技术来系统预测葡萄皮-籽总酚、皮-籽单宁含量的研究较少。可见-近红外光谱技术因操作简单、精确度高、客观无损而成为目前极具潜力的水果品质无损检测技术^[15],因此,开展基于可见-近红外光谱的葡萄皮-籽总酚、单宁含量的预测研究尤为必要。

此外,光谱信息采集中,包含大量无关的噪声信号,为提升模型预测精度,通常采用传统的机器学习方法构建光谱预测模型,其中,常见的有偏最小二乘回归(Partial least squares regression, PLSR)、支持向量机回归(Support vector machine regression, SVR)等。目前,深度学习方法在人工智能领域引起了广泛的关注,是一系列基于人工神经网络的机器学习算法典型代表。然而,在光谱数据分析方面,仍主要侧重于传统的机器学习算法,仅有少数研究探究了深度学习解析光谱数据。深度学习具备出色的模式识别能力,适于从广泛的光谱数据库中提取有价值的信息^[16],同时,深度学习和传统机器学习在特定应用中将产生不同的效果。相关研究已将深度学习应用于水果品质的定量预测,取得了较好的结果^[17]。

本文基于可见-近红外光谱技术,结合 PLSR、SVR 和卷积神经网络(Convolutional neural network, CNN)建模方法,以巨玫瑰葡萄品种为研究对象,系统构建巨玫瑰葡萄从转色至成熟的皮-籽总酚、皮-籽单宁含量与可见-近红外光谱之间的关系模型,并

对比分析 PLSR、SVR、CNN 模型预测结果,确定各品质指标的最佳检测模型,以期为葡萄酚类物质的无损在线监测提供理论依据与技术支持。

1 材料与方法

1.1 样品采集

巨玫瑰葡萄取自曹新庄葡萄园(34°18'0" N, 108°05'23.9" E, 杨凌),随机选取两行葡萄植株作为固定采样点。采样时兼顾果穗上中下及阴阳面,利用五点取样法从每穗采集 5 粒无损伤和病害的葡萄颗粒。采样周期从葡萄转色期(2021 年 7 月 15 日)至成熟期(2021 年 10 月 15 日)共计约 90 d。根据葡萄生长特点并结合种植经验,确定每隔 3~5 d 进行 1 次采样,每次采集 280~320 粒葡萄,并按照颜色和硬度进行分组,40 粒为一组,分为 7~8 组。采样周期内共采集 24 批试验样品,共计 178 组,每个样品测定 20 颗巨玫瑰果粒的光谱信息,并将相应的 20 颗巨玫瑰果粒-40℃ 冷冻保存进行后续理化值测定。

1.2 试验方法

1.2.1 光谱采集

利用手持式光纤光谱仪(ATP3030 型,厦门奥谱天成光电有限公司)对巨玫瑰葡萄进行光谱采集。仪器参数:积分时间 1 ms,分辨率 0.5 nm,HL2000 卤钨灯光源,漫反射方式。在室温下进行试验,测量前将光谱仪预热,光谱数据由 Optosky Spectra 软件进行采集;测量时为避免环境干扰,采用标准漫反射白板(WR-D97 型,材料为聚四氟乙烯)作为背景以放置样本,由卤素灯光源、暗室、光纤探头等构成的手柄灯先与标准白板贴合,以采集白板反射光谱进行校准,再将单粒葡萄样本放在漫反射白板上并置于手柄灯暗室内,以最终获取样本反射光谱;此外,测量过程沿葡萄样本赤道以 120° 间隔测量 3 次取平均值作为最终光谱数据,其采集过程原理如图 1 所示。为消除光谱仪及外界因素影响,去除前端部分光谱波段,选择波长范围 400~1 100 nm 共 1 508 个波长点,利用 Python 3.11 分析处理并进行后续建模分析。

1.2.2 葡萄预处理

光谱测量结束后,将葡萄样品置于-40℃ 的低温条件下冷冻保存,待分离葡萄皮、籽进行总酚和单宁含量测定。在进行酚类物质测定之前,先对葡萄皮、籽进行预处理。首先,在液氮冷冻条件下将葡萄皮、籽分别磨成粉末,然后将粉末放入真空冷冻干燥器(上海叶拓科技有限公司)中冻干 24 h。随后,取出的粉末用盐酸甲醇溶液(0.1% 盐酸:60% 甲醇)



图 1 光谱采集系统

Fig.1 Reflectance spectral acquisition system

在温度 30℃、功率 40 W 下超声处理 30 min,在温度 4℃、10 000 r/min 下离心 10 min,收集上清液。重复以上提取步骤 2 次,将所有上清液混合摇匀,然后储存于-40℃ 冰箱中以保持酚类物质的稳定性^[18],用于后续酚类物质测量。

1.2.3 葡萄皮、籽酚类物质测定

总酚是葡萄皮、籽提取物中主要的酚类物质,采用 Folin-Ciocalteu 方法^[19]测定并进行适当修改。使用没食子酸标准曲线(图 2),以毫克/克(mg/g)没食子酸当量表征总酚含量。采用 SARNECKIS 等^[20]提出的甲基纤维素沉淀法测定单宁含量,并根据校准曲线(图 3),以 mg/g 的儿茶素当量表征单宁含量。

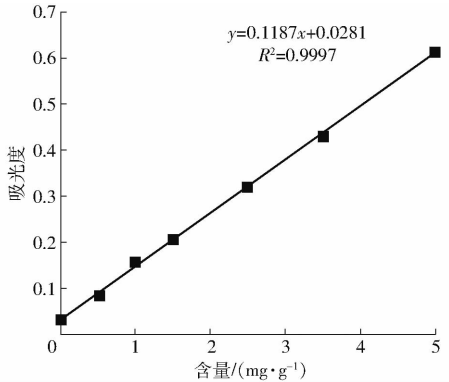


图 2 总酚标准曲线

Fig.2 Total phenols standard curve

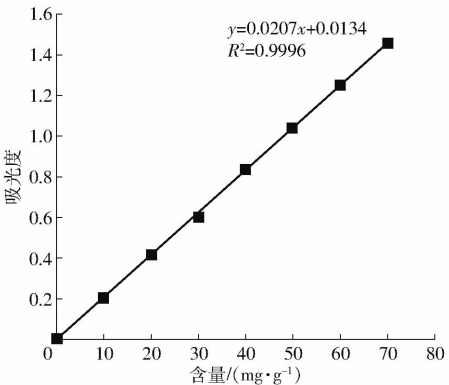


图 3 单宁标准曲线

Fig.3 Tannin standard curve

1.3 数据分析与模型构建

1.3.1 样本集划分

为了确保校正集和预测集的分布最大程度地表

征样品分布,使样品更具代表性,采用 SPXY (Sample set partitioning based on joint $x-y$ distances) 算法^[21]对样本集进行划分。该算法基于光谱空间(x)和目标组成空间(y)的欧氏距离来选择样本。首先,根据欧氏距离选择具有最大距离的两个样本作为校正集的初始样本。随后,计算剩余样本与校正集中每个样本之间的欧氏距离,逐步将具有最长和最短距离的样本添加到校正集中,反复执行这一过程,直到校正集达到所需的样本数量^[22-23]。通过使用 SPXY 算法,将所有样本按照比例 3:1 划分为校正集和预测集。

1.3.2 光谱预处理

在光谱法测量过程中,外部环境、测量条件以及仪器本身的特点对测量结果有较大的影响。所以,收集到的光谱信息非常丰富,不仅包括有价值的葡萄样品的漫反射光谱信息,还包括仪器本身的噪声。在采集过程中,从标准白板上所反射的其他光谱信息,还有周围杂散光等。为减弱或消除该影响,需对采集到的光谱数据进行预处理,从而得到有效的光谱数据,提升模型检测准确度和稳定性。采用标准正态变换 (Standard normal variate, SNV)、多元散射校正 (Multiplicative scatter correction, MSC)、一阶导数 (First derivative, 1D)、二阶导数 (Second derivative, 2D)、Savitzky-Golay 卷积平滑 (Savitzky-Golay smoothing, SG) 以及 Savitzky-Golay 卷积平滑 + 一阶导数 (SG + 1D) 6 种预处理方法对原始光谱 (Raw spectra, RAW) 数据进行处理。

1.3.3 特征波长提取

采集的原始葡萄数据通常包含大量的冗余和共线性数据,显著影响模型精度及数据处理效率^[24]。竞争自适应重加权算法 (Competitive adaptive reweighted sampling, CARS) 可以消除非信息性变量,并且增强模型预测性能^[25],筛选出偏最小二乘

模型中相关系数最大的波长,并且利用交叉验证找到均方根误差最小的子集,而子集中的变量就是被选择的波长变量。

1.3.4 模型构建

偏最小二乘回归是一种基于因子分析的多元校准方法,将主成分分析、典型相关性分析和多元线性回归的优势集于一体,可同时考虑理化值与光谱之间基本关系,为有效解决回归分析中的自变量多重共线性问题提供了很好的解决途径^[26-27]。

支持向量机^[28]是一种基于统计学习理论和结构风险最小化的机器学习方法,以其特殊的泛化能力和在多变量问题中的可靠性而被广泛关注,尤其适用于小样本问题。SVR 在光谱数据中已被证明是一种可靠且高效的方法^[29]。为了降低计算复杂度,采用高斯径向基函数 (RBF) 作为核函数。关键参数包括惩罚因子 (C) 和核参数 (g),通过 RandomizedSearchCV 方法和 5 折交叉验证确定,取值范围均设为 $10^{-4} \sim 10^3$ 。

卷积神经网络是一种利用卷积层的人工神经网络的变体,通过局部连接和权重共享来减少参数数量^[30]。典型的 CNN 包括输入层、卷积层、池化层、扁平层和全连接层。本文 CNN 回归框架如图 4 所示,包括 1 个输入层、5 个卷积层、5 个池化层、1 个扁平层、3 个全连接层和 1 个输出层,其中卷积层卷积核数量分别为 192、128、64、32、16,卷积核尺寸均为 3,池化层尺寸和步长均设置为 2,修正线性单元 (Rectified linear unit, ReLU) 选为卷积层和全连接层的激活函数。为了防止过拟合,在扁平层之后添加 1 个随机失活层,失活概率为 0.5。初始学习率设置为 0.000 1,迭代次数为 500,并选择 Adam 作为优化器^[31]。为了加快网络的收敛速度,批量大小设置为 20,并使用均方误差损失函数来进一步降低过拟合的风险^[32]。

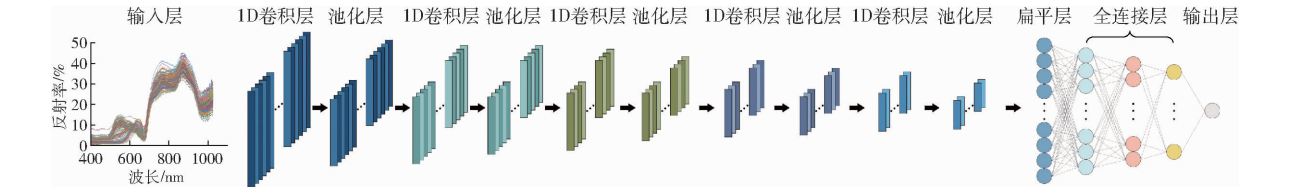


图 4 CNN 模型架构示意图
Fig. 4 Architecture of CNN model

1.3.5 模型评价

模型准确性和稳定性由校正集均方根误差 (Root mean square error of calibration set, RMSEC)、预测集均方根误差 (Root mean square error of prediction set, RMSEP)、校正集相关系数 (Correlation coefficient of calibration set, R_c)、预测集相关系数

(Correlation coefficient of prediction set, R_p) 和剩余预测偏差 (Residual predictive deviation, RPD) 进行评估^[33]。其中,相关系数 (R_c 、 R_p) 表示一元多项式回归方程的拟合度,均方根误差 (RMSEC、RMSEP) 则是用来平衡观测值与真实值之间的偏差;相关系数越接近于 1,均方根误差越小,模型精准度和稳定性

越高^[34]。RPD 是模型预测精度的重要指标,在模型评价中,当 $RPD \geq 3$ 时,表明模型具有极好预测能力;当 $2 \leq RPD < 3$ 时,表明模型具有一定的预测能力;当 $1.5 \leq RPD < 2$ 时,表明模型预测能力不足,但可用于定量分析;当 $RPD < 1.5$ 时,表明模型预测能力较差^[35-36]。

2 结果与讨论

2.1 葡萄成熟期光谱特征

图 5 为 178 组不同成熟品质葡萄的漫反射光谱。由图 5 可知,在原始光谱中葡萄样品的可见-近红外光谱虽存在重叠和交叉的情况,但整体上变化趋势基本一致。在光谱范围 500 ~ 720 nm 内,光谱曲线存在一定的差异性,原因可能在于葡萄成熟品质不同致使其果皮颜色存在差异,进而影响光谱的吸收^[10]。光谱曲线在波长 720 nm 附近存在明显的吸收峰,分析其原因可能与胡萝卜素^[37]、叶绿素^[38]的吸收有关。在光谱范围 850 ~ 920 nm 内,光谱曲线出现吸收带,其原因可能与酚类^[39]和花青素、纤维素和蔗糖^[12]等的吸收有关^[40]。此外,光谱曲线在 1 000 ~ 1 050 nm 范围内出现吸收带,可能与水和维生素等成分有关^[41-42]。

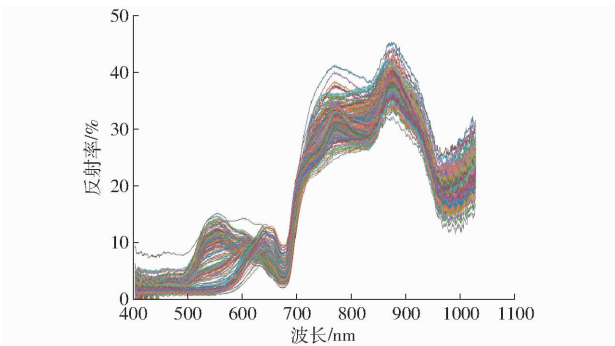


图 5 成熟期葡萄光谱曲线

Fig. 5 Spectral curves of grapes in ripening period

2.2 葡萄成熟期理化指标

巨玫瑰葡萄样本酚类物质含量频数分布如图 6 所示,可以看出皮总酚、籽总酚、皮单宁和籽单宁的含量呈正态分布,且分布区间分别为 37 ~ 96 mg/g、53 ~ 101 mg/g、14 ~ 76 mg/g、31 ~ 80 mg/g,表明样本的分布范围较广,采样具有代表性。利用 SPXY 算法按照比例 3:1 将葡萄样本划分为 134 个校正集样本,44 个预测集样本。由表 1 可知,样本集划分后,各指标的校正集和预测集的酚类物质含量范围相近,说明通过 SPXY 划分的数据集分布较为合理,有助于提升预测模型稳定性。

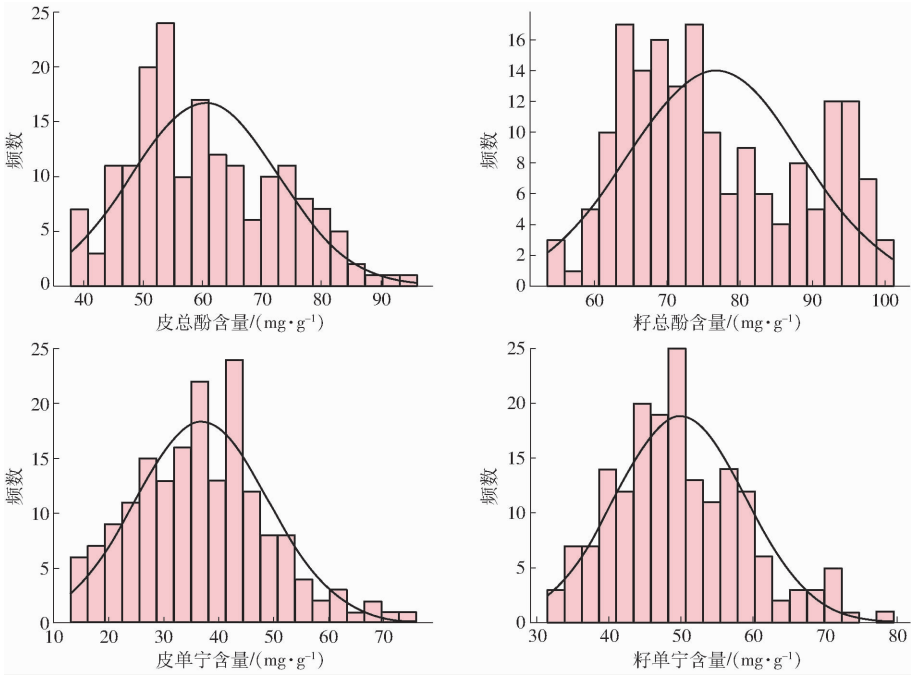


图 6 巨玫瑰葡萄酚类物质含量频数分布

Fig. 6 Frequency distributions of phenolic substances content in Muscat Kyoho grapes

2.3 不同预处理下全波长预测模型构建

采用 PLSR、SVR 和 CNN 3 种机器学习方法对原始光谱和经过不同预处理的光谱数据建立了酚类物质的预测模型。模型评价指标如图 7 (图中 RMSEC、RMSEP 单位为 mg/g,下同) 所示。由图 7

可知,经过 1D 和 2D 预处理后的光谱数据建立的皮总酚模型相较于其他预测模型均有着较高的 R_p 、RPD 和较低的 RMSEP,因此这两种预处理后建立的皮总酚模型的性能较差。而原始光谱所建立的 3 个皮总酚预测模型均具有较高的 R_p 和 RPD,同时具有

表 1 利用 SPXY 算法划分葡萄样品样本集的数据统计

品质指标	校正集			预测集		
	最大值	最小值	平均值	最大值	最小值	平均值
	mg/g					
皮总酚含量	95.96	37.89	60.63	83.46	38.43	59.91
籽总酚含量	101.02	53.73	76.51	98.87	53.56	76.85
皮单宁含量	75.79	13.05	36.79	61.10	15.56	37.44
籽单宁含量	79.43	31.48	50.36	71.12	32.93	48.51

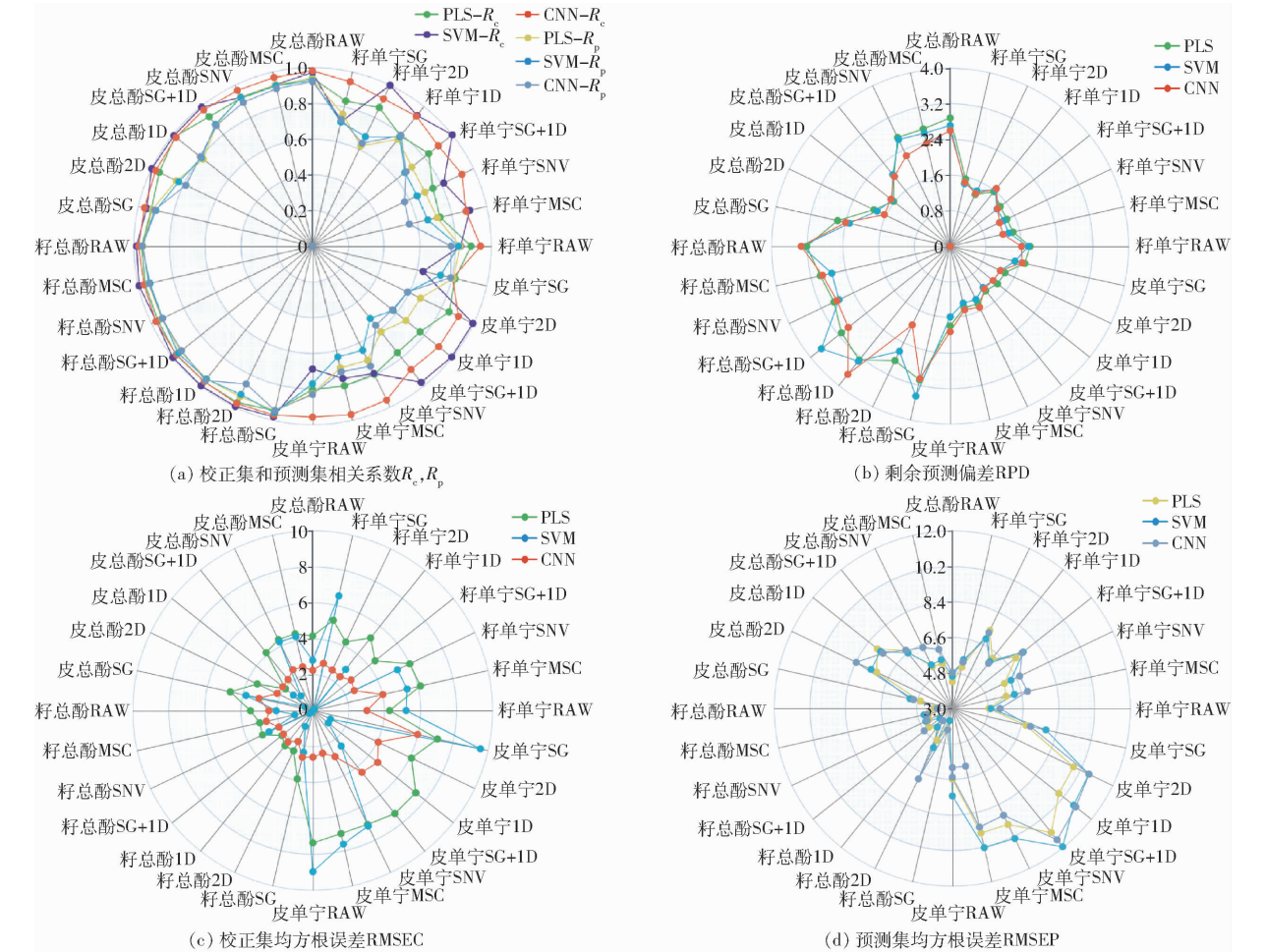


图 7 不同预处理后全波段 PLS、SVR、CNN 模型的 R 、RMSE 及 RPD

Fig. 7 R , RMSE and RPD values of PLSR, SVR and CNN models by using different preprocessing methods

预测集的 R_p 和 RPD 减小, RMSEP 增大, 表明过多的预处理可能破坏了与皮单宁相关的原始光谱信息^[44-45]。而原光谱建立的皮、籽单宁的预测模型效果较好, 其中皮单宁的 CNN 模型效果最佳, R_p 、RMSEP 和 RPD 分别为 0.830 3、6.487 6 mg/g 和 1.917 5, 籽单宁的 PLSR 模型效果最佳, R_p 、RMSEP 和 RPD 分别为 0.825 9、4.847 8 mg/g 和 2.174 3。因此, 对于巨玫瑰葡萄的籽总酚含量, 原始光谱经过 1D 预处理后模型的预测精度明显提高, 选取原始光谱经过 1D 预处理进行特征波长提取和建模。而对于巨玫瑰葡萄皮总酚、皮单宁和

较低的 RMSEP。其中, PLSR 模型表现最佳, 其 R_p 、RMSEP 和 RPD 分别为 0.936 4、4.353 1 mg/g 和 2.883 2。在籽总酚所建立的模型中, 经过 2D 预处理后的模型性能较差, 可能是由于 2D 预处理在辨别重叠峰的同时放大了局部误差^[43]。而经过 1D 预处理后模型具有较高的 R_p 和 RPD, 同时具有较低的 RMSEP。其中, CNN 建立的模型效果最优, 其 R_p 、RMSEP 和 RPD 分别为 0.961 4、3.786 1 mg/g 和 3.677 5。原光谱经不同预处理后的皮、籽单宁模型

籽单宁含量, 经过预处理后所建模型的预测精度反而降低, 后续利用原始光谱进行特征波长提取和建模分析。

2.4 特征波长提取

采集到的光谱数据包含 1 508 个波长点, 利用 CARS 算法分别提取葡萄皮总酚、籽总酚、皮单宁、籽单宁经过最佳预处理后的光谱数据的特征波长。以籽总酚特征波长选取为例, 阐述 CARS 算法对特征波长的选取过程, 如图 8 所示。本研究设定蒙特卡罗采样次数为 100, 通过 10 倍交叉验证方法确定变量的显著性。由图 8b 可知, 交互验证均方根误差

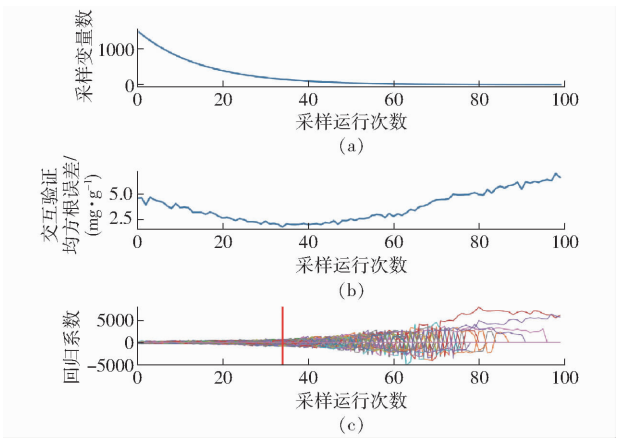


图 8 基于 CARS 算法的巨玫瑰籽总酚特征波长选取
Fig.8 Characteristic wavelength selection charts
of total phenolic content in Muscat Kyoho grapes
extraction based on CARS algorithm

(RMSECV)随着取样运行次数增加先降低后增加;当 RMSECV 逐渐降低时,表明光谱数据中无用信息被剔除,而当 RMSECV 逐渐增加时,表明光谱数据中有用信息开始被剔除。因此,选择取样运行次数 100 次所对应的最小 RMSECV 作为最优结果,结合图 8b 和图 8c,当 RMSECV 达到最小值时,各变量的回归系数位于图 8c 中竖直线位置,采样次数为 34 次。由此,基于 CARS 算法选取葡萄的皮总酚、籽总酚、皮单宁、籽单宁波长变量数分别为 50、155、91 和 50,所选波长点分别占总波长的 3.3%、10.3%、6.0% 和 3.3%,数据减少量均在 89% 以上。由 CARS 算法选取的波长如图 9 所示。

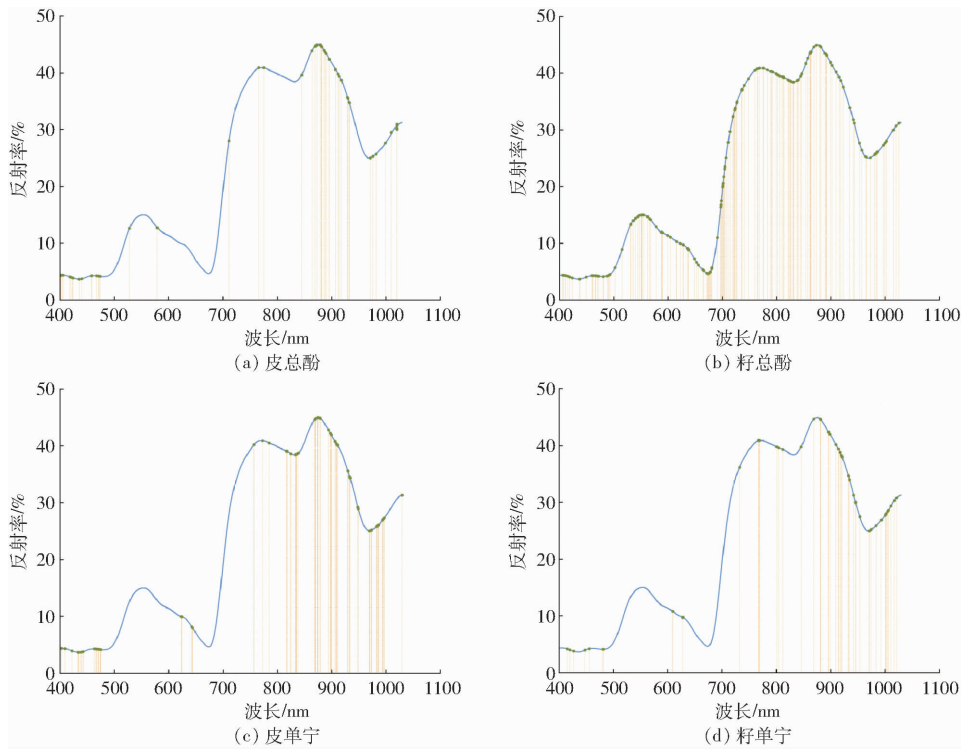


图 9 CARS 选择的不同酚类物质特征波长的位置

Fig.9 Location of variables selected with CARS for phenolic content detection

2.5 基于特征波长的光谱预测模型构建

采用 CARS 方法辨识关键光谱区域,可最大限度地降低输入变量数量,有助于建立更为稳定、简洁的模型。在对皮-籽总酚、皮-籽单宁进行特征波长筛选后,使用 PLSR、SVR 和 CNN 对光谱数据进行建模。图 10 为所建立 PLSR 模型的回归系数,其中绿色线、红色线和紫色线分别表示对品质指标影响较高-中等-较低的波长。由图 10 可知,皮总酚(图 10a)、籽总酚(图 10b)和籽单宁(图 10d)的绿色和红色线条较多,表明 CARS 算法很好地识别了对应指标的特征波长。而对于皮单宁(图 10c),紫色线条较多,显示 CARS 算法在与该指标相关的特征波长识别能力方面有待提升。

图 11 为 CNN 模型校正集和预测集的损失值。可知,4 个模型损失值在开始阶段迅速下降,之后下降速度减缓,并且发生一定的波动。经过 200 次迭代后,校正集和预测集损失值逐渐趋于稳定。籽总酚校正集和预测集损失值最终趋近于 0,而皮单宁损失值仍然较高。

皮-籽总酚、皮-籽单宁经特征波长筛选后建立的 PLSR、SVR 和 CNN 模型评价指标见图 12。可以看出,皮、籽总酚总体建模效果均优于皮、籽单宁,且皮总酚、籽总酚和籽单宁特征波长筛选后所建立的 3 个模型预测精度较全波段均有提高,而皮单宁特征波长筛选后建立的 3 个模型较全波段均有降低,这与图 10、11 相一致,原因可能是 CARS 未能较好

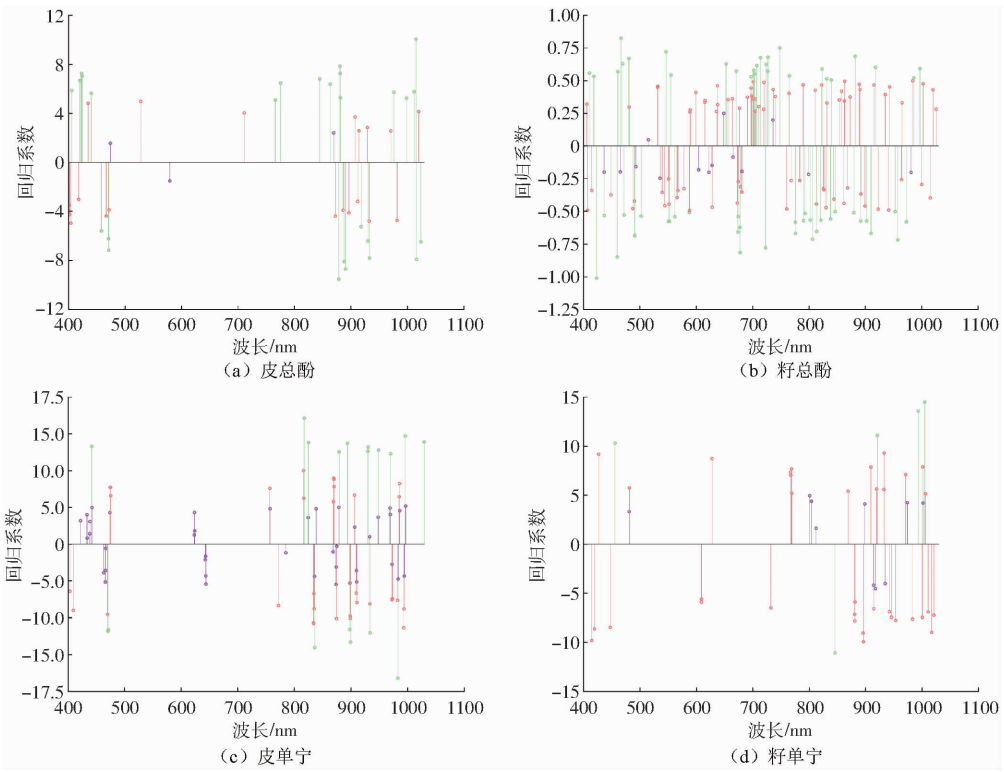


图 10 PLSR 模型回归系数
Fig. 10 Partial least squares regression coefficients

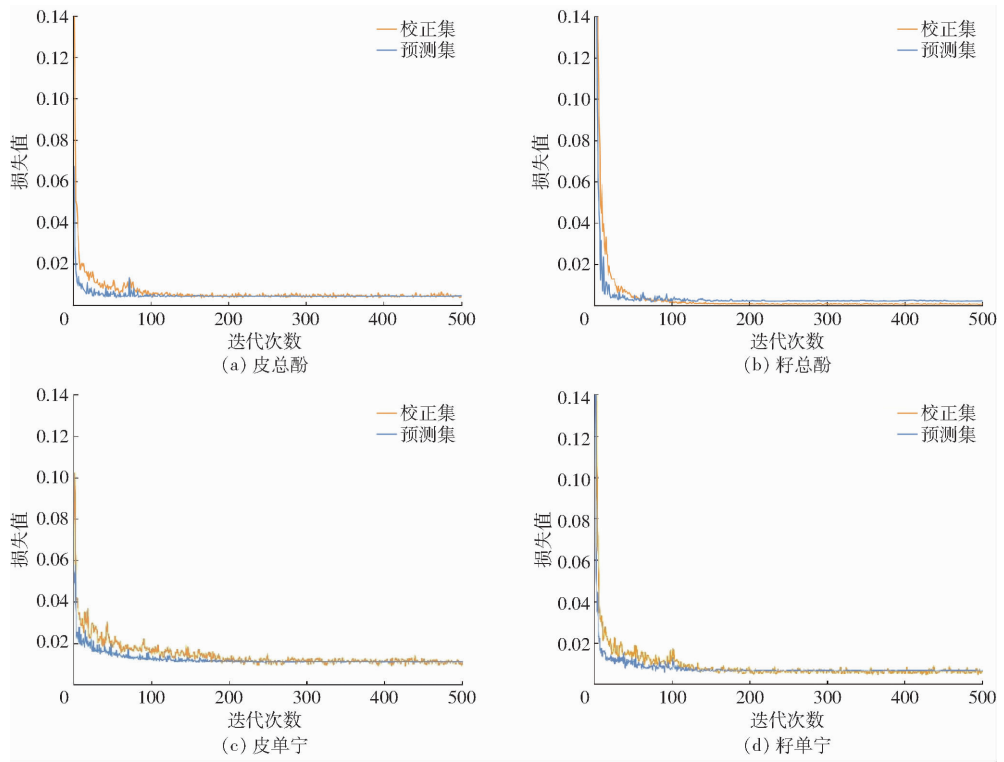


图 11 CNN 模型校正集和预测集损失值变化曲线
Fig. 11 Loss curves of CNN models by calibration set and prediction set

地识别出与皮单宁相关的特征波段^[46]。对于皮总酚,经过特征波长筛选后所建立的预测模型 R_p 均在 0.94 以上,RPD 均大于 2.5,其中 SVR 所建立的模型效果最佳,RPD 可达 3.38,模型预测效果较好;籽总酚 3 个模型的建模效果均优于其他 3 个指标,原

因可能在于 CARS 筛选籽总酚的特征波段时保留了较多的波长,且从图 10b 可知,这些波长与籽总酚相关性较高,从而建模效果较好,其中 3 个模型的 R_p 均大于 0.98,RPD 均高于 5.5,综合对比 3 个模型的各个评价指标,SVR 所建模型的 R_c 、 R_p 和 RPD 均

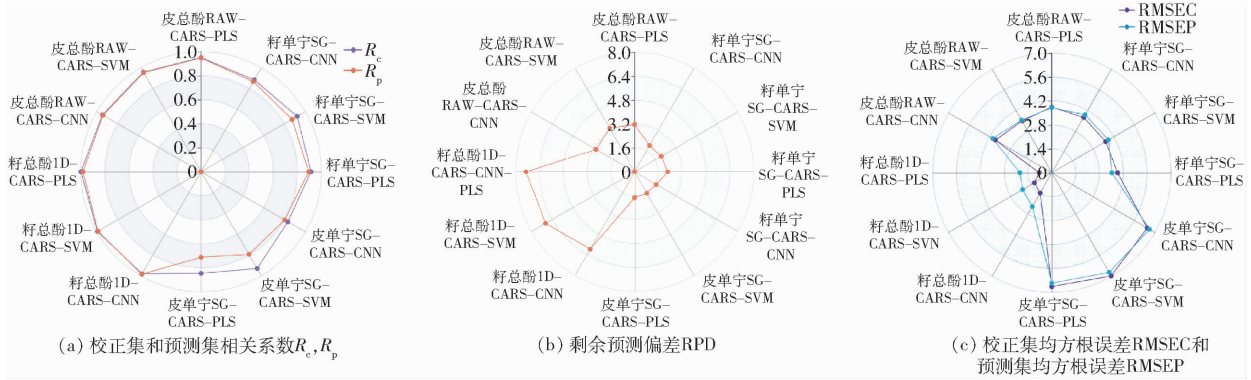


图 12 基于特征波长建立的 PLSR、SVR 和 CNN 模型的 R 、RMSE 及 RPD

Fig. 12 R , RMSE and RPD values of PLSR, SVR and CNN models based on selected spectra

高于其他模型, RMSEP 和 RMSEC 均低于其他模型。由此可知, 经特征波长筛选后建立的 SVR 模型是籽总酚的最优模型; 而经特征波长筛选后建立的皮单宁的预测模型 R_p 和 RPD 均低于基于全波段的预测模型, 且 RMSEP 和 RMSEC 均较高, 则基于全波段建立的 CNN 模型为皮单宁最佳预测模型; 籽单宁经过特征波长筛选后建立的 3 个模型 RPD 均大于 2, 效果均优于全波段所建立的模型, 其中 PLSR 模型的 R_c 可达 0.91, R_p 可达 0.89, RPD 为 2.25, 比 SVR 和 CNN 所建立的模型效果好, 则经特征波长筛选后建立的 SVR 模型是籽单宁的最佳模型。

2.6 最优模型确立

综合图 7、12 的结果, 对所构建的 4 个指标的预测模型进行分析。结果显示, 原始光谱经特征波长筛选后建立的 SVR 模型为皮总酚最佳模型, 其 R_c 、

R_p 和 RPD 分别为 0.96、0.95 和 3.56。对于籽总酚, 经过 1D 预处理后建立的 SVR 模型表现最佳, 其 R_c 、 R_p 和 RPD 分别为 0.99、0.99 和 7.30。而针对皮单宁的预测模型 R_p 和 RPD 均低于基于全波段的预测模型, 且 RMSEP 和 RMSEC 均较高, 则基于全波段建立的 CNN 模型为皮单宁最佳预测模型; 籽单宁经过特征波长筛选后建立的 3 个模型 RPD 均大于 2, 效果均优于全波段所建立的模型, 其中 PLSR 模型是最优模型, 其 R_c 、 R_p 和 RPD 分别为 0.91、0.83 和 2.25。葡萄样品皮总酚、籽总酚、皮单宁和籽单宁最佳模型的校正集和预测集样本的预测值与参考测定值之间的关系散点图见图 13。

3 讨论

已有研究探究了利用可见-近红外光谱预测葡萄品质指标的可行性, 且利用传统机器学习构建葡萄糖酸指标预测模型已成为较为常见的研究方

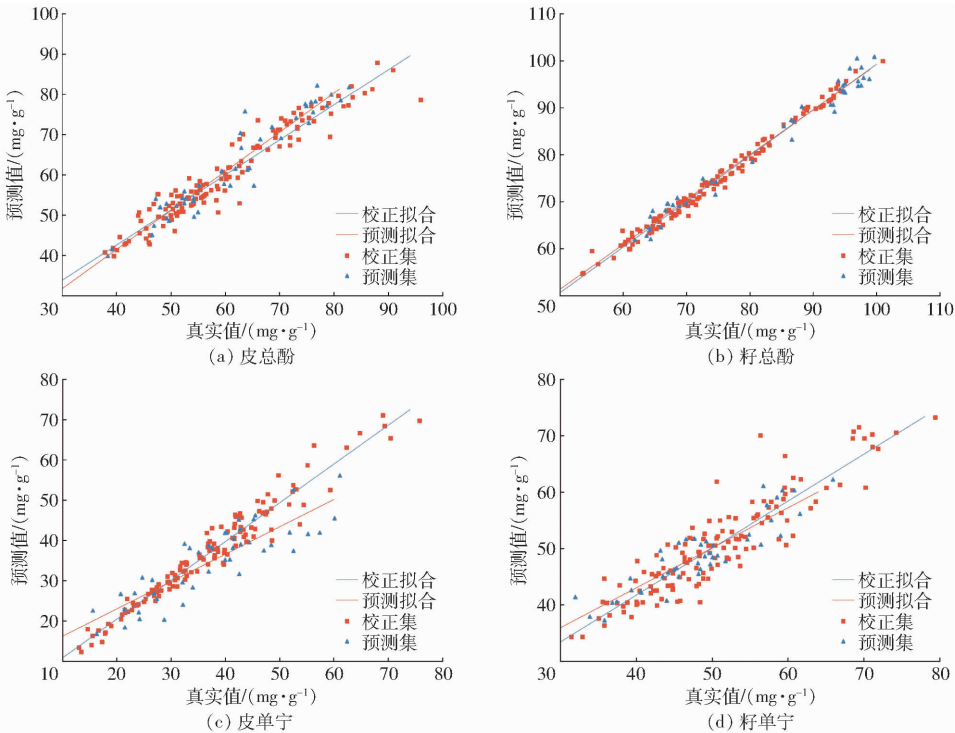


图 13 葡萄样品可见/近红外光谱皮-籽总酚与单宁的最优模型

Fig. 13 The best models of total phenols and tannins in grape samples

法^[3, 47]。然而,酚类物质对葡萄的品质、颜色和口感至关重要,且与成熟度密切相关^[48]。传统的酚类物质测定方法需大量实验、时间和化学试剂,故探寻一种简单快速测定葡萄酚类物质的方法尤为必要。因此,本文采用可见-近红外光谱对葡萄皮-籽总酚和皮-籽单宁含量进行预测,有助于根据采摘需求有选择性地收获葡萄,达到所需的葡萄酚类成熟度,从而确保最佳的质量品质。除了传统的预测方法和人工选择特征光谱区域在识别样本特征方面的有效性,还研究了利用深度学习方法预测葡萄酚类物质的可行性,并将全光谱与基于CARS选择的特征波长的预测效果进行了比较。采用的CNN深度神经网络学习方法,具有通过多个卷积层自主学习和提取特征的能力,已有的光谱研究中,该方法展现出了较传统机器学习方法更出色的预测能力^[49-50]。尽管CNN模型表现良好,但有关皮总酚、籽总酚和籽单宁的预测模型效果略逊于传统的机器学习;可能原因在于葡萄成熟过程中,酚类物质变化更加显著,故线性模型拟合效果较非线性模型更好,因此,采用线性算法建立模型可以提高预测效果^[51]。对于皮-籽总酚和皮-籽单宁的最优预测模型, R_p 均大于0.8,其中皮总酚和籽总酚的 R_p 均超过0.95,表明模型具有较好的预测能力。然而,考虑到本研究的局限性,在构建CNN模型时减少了网络深度,主要原因在于光谱数据仅提供一维信息,使用过深的网络易导致过拟合,而深层的深度学习模型通常更适用于更大的数据集。未来研究中,可以进一步扩大葡萄样本的收集,包括来自不同来源

和不同年份的样本,以改进模型的完整性,此外,通过对多年收集的样本进行微调和训练,可以增强模型的稳定性和通用性,进而扩展CNN模型的深度,提高模型的鲁棒性和预测准确性,促进深度学习在该领域的应用。

4 结论

- (1)采用6种方法对原始光谱进行预处理,并结合PLSR、SVR和CNN模型结果发现,皮总酚、皮单宁和籽单宁基于原始光谱无需任何预处理即可获得较好的建模效果;而对于籽总酚,原始光谱经1D预处理模型预测精度明显提高。
- (2)利用CARS算法对经过最佳预处理后的光谱进行特征波长选择,可得皮总酚、籽总酚、皮单宁、籽单宁特征波长数量分别为50、155、91和50,所选波长数分别占总波长的3.3%、10.3%、6.0%和3.3%,数据减少量均在89%以上。
- (3)综合对比分析基于全波段和特征波长的PLSR、SVR和CNN皮-籽总酚、皮-籽单宁含量预测模型发现,与全波段相比,特征波长预测模型大幅降低计算量,且皮总酚、籽总酚、籽单宁的预测精度有所提高,而皮单宁的模型预测精度略有下降。由此可得皮总酚、籽总酚、皮单宁、籽单宁最优模型分别为RAW-CARS-SVR、1D-CARS-SVR、RAW-CNN和RAW-CARS-PLSR,其预测集相关系数 R_p 分别为0.95、0.99、0.83和0.89,预测集均方根误差RMSEP分别为3.56、1.87、6.49、3.53 mg/g;剩余预测偏差RPD分别为3.56、7.30、1.92和2.25。

参 考 文 献

[1] 贾璐,叶中华. 基于注意力机制和特征融合的葡萄病害识别模型[J]. 农业机械学报, 2023, 54(7): 223-233.
JIA Lu, YE Zhonghua. Grape disease recognition model based on attention mechanism and feature fusion[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(7): 223-233. (in Chinese)

[2] 张宏鸣,张国良,朱珊珊,等. 基于U-Net的葡萄种植区遥感识别方法[J]. 农业机械学报, 2022, 53(4): 173-182.
ZHANG Hongming, ZHANG Guoliang, ZHU Shanna, et al. Remote sensing recognition method of grape planting regions based on U-Net[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(4): 173-182. (in Chinese)

[3] 高升,王巧华,施行,等. 便携式红提葡萄多品质可见/近红外检测仪设计与试验[J]. 农业机械学报, 2021, 52(2): 308-319.
GAO Sheng, WANG Qiaohua, SHI Hang, et al. Design and test of portable red globe grape extraction multi-quality visible/near infrared detector[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(2): 308-319. (in Chinese)

[4] XIAO Hui, LI Ang, LI Meiyu, et al. Quality assessment and discrimination of intact white and red grapes from *Vitis vinifera* L. at five ripening stages by visible and near-infrared spectroscopy[J]. Scientia Horticulturae, 2018, 233: 99-107.

[5] ZHOU Xuejian, LIU Wenzheng, LI Kai, et al. Discrimination of maturity stages of Cabernet Sauvignon wine grapes using visible-near-infrared spectroscopy[J]. Foods, 2023, 12(23): 4371.

[6] BACA-BOCANEGRA B, HERNÁNDEZ-HIERRO J M, NOGALES-BUENO J, et al. Feasibility study on the use of a portable micro near infrared spectroscopy device for the “in vineyard” screening of extractable polyphenols in red grape skins[J]. Talanta, 2019, 192: 353-359.

[7] JEDIYI H, NAAMANI K, AIT ELKOCH A, et al. First study on technological maturity and phenols composition during the ripeness of five *Vitis vinifera* L grape varieties in Morocco[J]. Scientia Horticulturae, 2019, 246: 390-397.

[8] MOHD A M, HASHIM N. Non-destructive methods for detection of food quality[M]. America: Academ Press, 2022: 645-667.

[9] DUCKENA L, ALKSNIS R, ERDBERGA I, et al. Non-destructive quality evaluation of 80 tomato varieties using Vis-NIR

- spectroscopy[J]. *Foods*, 2023, 12(10): 1990.
- [10] ZHANG Xiao, ZHOU Chenxin, SUN Qi, et al. Prediction of solid soluble content of green plum based on improved CatBoost[J]. *Agriculture-Basel*, 2023, 13(6): 1122.
- [11] GAO Sheng, XU Jianhua. Hyperspectral image information fusion-based detection of soluble solids content in red globe grapes[J]. *Computers and Electronics in Agriculture*, 2022, 196: 106822.
- [12] DANIELS A J, POBLETE-ECHEVERRIA C, OPARA U L, et al. Measuring internal maturity parameters contactless on intact table grape bunches using NIR spectroscopy[J]. *Frontiers in Plant Science*, 2019, 10: 1517.
- [13] YU Jing, WANG Hui, SUN Xiangyu, et al. Parameter optimization in soluble solid content prediction of entire bunches of grape based on near infrared spectroscopic technique[J]. *Journal of Food Measurement and Characterization*, 2017, 11(4): 1676 – 1680.
- [14] BASILE T, MARSICO A D, PERNIOLA R. NIR analysis of intact grape berries: chemical and physical properties prediction using multivariate analysis[J]. *Foods*, 2021, 10(1):113.
- [15] 郭志明, 王郡艺, 宋焯, 等. 手持式可见近红外苹果品质无损检测系统设计与试验[J]. *农业工程学报*, 2021, 37(22): 271 – 277.
- GUO Zhiming, WANG Junyi, SONG Ye, et al. Design and experiment of the handheld visible-near infrared nondestructive detecting system for apple quality[J]. *Transactions of the CSAE*, 2021, 37(22): 271 – 277. (in Chinese)
- [16] ZHENG Yinjie, CAO Yuchen, YANG Jie, et al. Enhancing model robustness through different optimization methods and 1-D CNN to eliminate the variations in size and detection position for apple SSC determination[J]. *Postharvest Biology and Technology*, 2023, 205: 112513.
- [17] LI Siyi, SONG Qiming, LIU Yongjie, et al. Hyperspectral imaging-based detection of soluble solids content of loquat from a small sample[J]. *Postharvest Biology and Technology*, 2023, 204:112454.
- [18] ESPARZA I, CIMMINELLI M J, MOLERO J A, et al. Stability of phenolic compounds in grape stem extracts[J]. *Antioxidants*, 2020, 9(8): 720.
- [19] CHENG Xianghan, MA Tingting, WANG Panpan, et al. Foliar nitrogen application from veraison to preharvest improved flavonoids, fatty acids and aliphatic volatiles composition in grapes and wines[J]. *Food Research International*, 2020, 137: 109566.
- [20] SARNECKIS C J, DAMBERGS R G, JONES P, et al. Quantification of condensed tannins by precipitation with methyl cellulose: development and validation of an optimised tool for grape and wine analysis[J]. *Australian Journal of Grape and Wine Research*, 2006, 12(1): 39 – 49.
- [21] 李佳盟, 王楠, 李震, 等. 基于特征波长优化的便携式作物叶绿素检测仪研究[J]. *农业机械学报*, 2023, 54(增刊2): 270 – 277.
- LI Jiameng, WANG Nan, LI Zhen, et al. Development of handheld chlorophyll detector based on characteristic wavelengths optimization[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(Supp. 2): 270 – 277. (in Chinese)
- [22] SHAO Yuanyuan, SHI Yukang, QIN Yongdong, et al. A new quantitative index for the assessment of tomato quality using Vis – NIR hyperspectral imaging[J]. *Food Chemistry*, 2022, 386: 132864.
- [23] TIAN Han, ZHANG Linna, LI Ming, et al. Weighted SPXY method for calibration set selection for composition analysis based on near-infrared spectroscopy[J]. *Infrared Physics & Technology*, 2018, 95: 88 – 92.
- [24] WANG Jingang, TIAN Tian, WANG Haijiang, et al. Estimating cotton leaf nitrogen by combining the bands sensitive to nitrogen concentration and oxidase activities using hyperspectral imaging[J]. *Computers and Electronics in Agriculture*, 2021, 189: 106390.
- [25] LIANG Long, WEI Lulu, FANG Guigan, et al. Prediction of holocellulose and lignin content of pulp wood feedstock using near infrared spectroscopy and variable selection[J]. *Spectrochimica Acta Part A—Molecular and Biomolecular Spectroscopy*, 2020, 225: 117515.
- [26] 杨增玲, 黄圆萍, 沈广辉, 等. 基于在线近红外光谱的堆肥全过程关键参数快速检测[J]. *农业机械学报*, 2019, 50(5): 356 – 361, 384.
- YANG Zengling, HUANG Yuanping, SHEN Guanghui, et al. Rapidly detection of key parameters in whole composting process based on online near infrared spectroscopy[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2019, 50(5): 356 – 361, 384. (in Chinese)
- [27] 赵娟, 沈懋生, 浦育歌, 等. 基于近红外光谱与多品质指标的苹果出库评价模型研究[J]. *农业机械学报*, 2023, 54(2): 386 – 395.
- ZHAO Juan, SHEN Maosheng, PU Yuge, et al. Out-of-warehouse evaluation and prediction model of apple based on near-infrared spectroscopy combined with multiple quality indexes[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(2): 386 – 395. (in Chinese)
- [28] YU Shui, HUAN Kewei, LIU Xiaoxi. Application of quantitative non-destructive determination of protein in wheat based on pretreatment combined with parallel convolutional neural network[J]. *Infrared Physics & Technology*, 2023, 135: 104958.
- [29] 边明博, 马彦鹏, 樊意广, 等. 融合无人机多源传感器的马铃薯叶绿素含量估算[J]. *农业机械学报*, 2023, 54(8): 240 – 248.
- BIAN Mingbo, MA Yanpeng, FAN Yiguang, et al. Estimation of potato chlorophyll content based on UAV multi-source sensor[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(8): 240 – 248. (in Chinese)
- [30] YANG J, XU J F, ZHANG X L, et al. Deep learning for vibrational spectral analysis: recent progress and a practical guide

- [J]. *Analytica Chimica Acta*, 2019, 1081: 6–17.
- [31] 冯喆, 李卫豪, 崔笛. 基于高光谱成像和深度学习的山核桃内源性异物检测[J]. *农业机械学报*, 2021, 52(增刊): 466–471.
- FENG Zhe, LI Weihao, CUI Di. Inspection of endogenous foreign body in Chinese hickorynut based on hyperspectral imaging and deep learning[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(Supp.): 466–471. (in Chinese)
- [32] ZENG Shuochong, ZHANG Zongyi, CHENG Xiaodong, et al. Prediction of soluble solids content using near-infrared spectra and optical properties of intact apple and pulp applying PLSR and CNN[J]. *Spectrochimica Acta Part A—Molecular and Biomolecular Spectroscopy*, 2024, 304: 123402.
- [33] 刘彩云, 李慧颖, 张倩玮, 等. 基于紫外可见光谱的红葡萄酒关键质量指标快速监测[J]. *农业机械学报*, 2023, 54(6): 401–409.
- LIU Caiyun, LI Huiying, ZHANG Qianwei, et al. Rapid monitoring of key quality indicators of red wines based on UV–Vis spectroscopy[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(6): 401–409. (in Chinese)
- [34] EOM Y H, CHUNG Y, PARK M, et al. Deep learning-based prediction method on performance change of air source heat pump system under frosting conditions[J]. *Energy*, 2021, 228: 120542.
- [35] HUANG Zhuangrong, SHA Sha, RONG Zhengqin, et al. Feasibility study of near infrared spectroscopy with variable selection for non-destructive determination of quality parameters in shell-intact cottonseed[J]. *Industrial Crops and Products*, 2013, 43: 654–660.
- [36] 陈远哲, 王巧华, 高升, 等. 皮蛋凝胶品质含水率和弹性的高光谱预测及其可视化[J]. *食品科学*, 2022, 43(2): 324–331.
- CHEN Yuanzhe, WANG Qiaohua, GAO Sheng, et al. Hyperspectral imaging for prediction and visualization of water content and springiness as indicators of the gel quality of preserved eggs[J]. *Food Science*, 2022, 43(2): 324–331. (in Chinese)
- [37] SHUTOVA V V T E, CHURIN A A. IR and Raman spectroscopy in the study of carotenoids of *Cladophora rivularis* algae[J]. *Biophysics*, 2016, 61(4): 601–605.
- [38] FLYNN K C F A, ADMAS S. Performance of chlorophyll prediction indices for *Eragrostis tef* at Sentinel–2 MSI and Landsat–8 OLI spectral resolutions[J]. *Precision Agriculture*, 2020, 21(5): 1057–1071.
- [39] BROWN A, ANDERSON D, RACICOT K, et al. Evaluation of phenolic phytochemical enriched commercial plant extracts on the *In Vitro* inhibition of α -Glucosidase[J]. *Frontiers in Nutrition*, 2017, 4: 56.
- [40] HERNÁNDEZ-SÁNCHEZ N, GÓMEZ-DEL-CAMPO M. From NIR spectra to singular wavelengths for the estimation of the oil and water contents in olive fruits[J]. *Grasas Y Aceites*, 2018, 69(4): 278.
- [41] FATCHURRAHMAN D, NOSRATI M, AMODIO M L, et al. Comparison performance of visible-NIR and near-infrared hyperspectral imaging for prediction of nutritional quality of goji berry (*Lycium barbarum* L.)[J]. *Foods*, 2021, 10(7): 1676.
- [42] BADARÓ A T, MORIMITSU F L, FERREIRA A R, et al. Identification of fiber added to semolina by near infrared (NIR) spectral techniques[J]. *Food Chemistry*, 2019, 289: 195–203.
- [43] 张梦宇, 郝敏, 田海清, 等. 基于高光谱成像技术的青贮玉米饲料 pH 值无损检测[J]. *农业工程学报*, 2023, 39(4): 239–247.
- ZHANG Mengyu, HAO Min, TIAN Haiqing, et al. Nondestructive detection of the pH value of silage maize feeds based on hyperspectral images[J]. *Transactions of the CSAE*, 2023, 39(4): 239–247. (in Chinese)
- [44] 高升, 王巧华, 李庆旭, 等. 基于近红外光谱的红提维生素 C 含量、糖度及总酸含量无损检测方法[J]. *分析化学*, 2019, 47(6): 941–949.
- GAO Sheng, WANG Qiaohua, LI Qingxu, et al. Non-destructive detection of vitamin C, sugar content and total acidity of red globe grape based on near-infrared spectroscopy[J]. *Chinese Journal of Analytical Chemistry*, 2019, 47(6): 941–949. (in Chinese)
- [45] LU Bing, LIU Nihong, LI Huiling, et al. Quantitative determination and characteristic wavelength selection of available nitrogen in coco-peat by NIR spectroscopy[J]. *Soil & Tillage Research*, 2019, 191: 266–274.
- [46] 李冠稳, 高小红, 肖能文, 等. 基于 sCARS–RF 算法的高光谱估算土壤有机质含量[J]. *发光学报*, 2019, 40(8): 1030–1039.
- LI Guanwen, GAO Xiaohong, XIAO Nengwen, et al. Estimation soil organic matter contents with hyperspectra based on sCARS and RF algorithms[J]. *Chinese Journal of Luminescence*, 2019, 40(8): 1030–1039. (in Chinese)
- [47] COSTA D D, MESA N F O, FREIRE M S, et al. Development of predictive models for quality and maturation stage attributes of wine grapes using vis-nir reflectance spectroscopy[J]. *Postharvest Biology and Technology*, 2019, 150: 166–178.
- [48] ARMSTRONG C E J, GILMORE A M, BOSS P K, et al. Machine learning for classifying and predicting grape maturity indices using absorbance and fluorescence spectra[J]. *Food Chemistry*, 2023, 403: 134321.
- [49] LIN P, LI X L, CHEN Y M, et al. A deep convolutional neural network architecture for boosting image discrimination accuracy of rice species[J]. *Food and Bioprocess Technology*, 2018, 11(4): 765–773.
- [50] WU Na, ZHANG Chu, BAI Xiulin, et al. Discrimination of *Chrysanthemum* varieties using hyperspectral imaging combined with a deep convolutional neural network[J]. *Molecules*, 2018, 23(11): 2831.
- [51] 陈玥瑶, 夏静静, 韦芸, 等. 近红外光谱法无损检测平谷产大桃品质方法研究[J]. *分析化学*, 2023, 51(3): 454–462.
- CHEN Yueyao, XIA Jingjing, WEI Yun, et al. Research on nondestructive quality test of Pinggu peach by near-infrared spectroscopy[J]. *Chinese Journal of Analytical Chemistry*, 2023, 51(3): 454–462. (in Chinese)