崔希炜, 刘铭萱, 江涛, 等. 基于可见/近红外光谱的科克铁热克葡萄品质预测和地理标志识别 [J]. 食品工业科技, 2026, 47(6): 34−43. doi: 10.13386/j.issn1002-0306.2025030202
CUI Xiwei, LIU Mingxuan, JIANG Tao, et al. Quality Prediction of Kokotireke Grapes and Identification of Geographical Indication Products Based on Visible/Near-infrared Spectroscopy[J]. Science and Technology of Food Industry, 2026, 47(6): 34−43. (in Chinese with English abstract). doi: 10.13386/j.issn1002-0306.2025030202

·特邀主编专栏—地方特色食品:加工技术、感官品质、风味特性和营养健康(客座主编:易俊洁、和劲松、郭超凡)·

基于可见/近红外光谱的科克铁热克葡萄品质预测和地理标志识别

崔希炜[1], 刘铭萱[2], 江　涛[2], 刘志刚[1,*], 刘　岩[1], 郭威望[1], 于　航[2]
(1.阿米检测技术有限公司, 江苏苏州 215100;
2.江南大学, 江苏无锡 214122)

摘　要: 科克铁热克(Kokotireke)葡萄作为新疆地理标志(地标)产品,存在品质参差不齐和地标掺假的问题。本研究创新性地利用可见/近红外(Visible/near-infrared, Vis/NIR)光谱技术,同时开展科克铁热克葡萄品质预测与地标识别工作。采集地标和非地标的样本,并依次获取光谱数据并测定其可溶性固形物(Soluble Solids Content, SSC)和可滴定酸(Titratable Acidity, TA)。随后,多种光谱预处理技术结合偏最小二乘回归(Partial Least Squares Regression, PLSR)构建品质预测模型和偏最小二乘判别分析(Partial Least Squares Discriminant Analysis, PLS-DA)地标产品识别模型。结果表明,不同预处理方法对模型性能产生明显影响,其中联合预处理(First Derivative Combined Pretreatment of Savitzky-Golay and Least Squares Baseline Correction, SG-LBC-1stD)效果最佳。PLSR 模型对 SSC 和 TA 预测的决定系数(R^2)分别为 0.9361 和 0.9478,均方根误差(Root Mean Squared Error, RMSE)分别为 0.3362 和 0.0368。此外,PLS-DA 模型能有效区分地标与非地标产品,拟合效果和预测能力良好,判别准确率高达 92.8%,R^2_X 为 0.828,R^2_Y 为 0.621,Q^2 为 0.553,AUC(Area Under Curve, AUC)值为 0.980。综上,本研究为科克铁热克葡萄产业提供了一种快速、在线、无损的品质预测和地标判别方法,对保障产品品质和预防地标掺假声誉意义重大,也为其他水果的相关研究提供了参考。

关键词: 科克铁热克葡萄, 可见/近红外光谱, 品质预测, 地标产品识别, 光谱预处理, 化学计量学建模
中图分类号: TS255.35　　**文献标识码:** A　　**文章编号:** 1002−0306(2026)06−0034−10　　本文网刊:
DOI: 10.13386/j.issn1002-0306.2025030202

Quality Prediction of Kokotireke Grapes and Identification of Geographical Indication Products Based on Visible/Near-infrared Spectroscopy

CUI Xiwei[1], LIU Mingxuan[2], JIANG Tao[2], LIU Zhigang[1,*], LIU Yan[1], GUO Weiwang[1], YU Hang[2]
(1.Am Incorporation for Metrology and Testing Technology Services, Suzhou 215100, China;
2.Jiangnan University, Wuxi 214122, China)

Abstract: Kokotireke grapes, a geographical indication agricultural product in Xinjiang, exhibit the challenges in uneven quality grades and geographical indication adulteration in the market. The visible/near-infrared (Vis/NIR) spectroscopy technology was innovatively employed to perform the quality prediction of Kokotireke grapes and geographical indication

收稿日期: 2025−03−17
基金项目: 中国航天科技集团民用产业自主研发项目。
作者简介: 崔希炜(1997−),男,硕士,中级工程师,研究方向:食品品质挖掘,E-mail:cuixw@htami.com。
*** 通信作者:** 刘志刚(1978−),男,博士,高级工程师/副教授,研究方向:食品检测、智能装备,E-mail:liuzg@htami.com。

identification in this study. Kokotireke grape samples with geographical indication and non-geographical indication were collected, and the spectral data, soluble solid content (SSC) and titratable acidity (TA) were measured. Subsequently, different spectral preprocessing techniques combined with the partial least squares regression (PLSR) to construct quality prediction models and partial least squares discriminant analysis (PLS-DA) geographical indication identification models. The results showed that different preprocessing methods exhibited a significant effect on the model performance, and the SG-LBC-1stD had the best preprocessing effectiveness. Specifically, the determination coefficients (R^2) of PLSR model for SSC and TA prediction reached 0.9361 and 0.9478, respectively, and the root mean square errors (RMSE) were 0.3362 and 0.0368, respectively. Additionally, the PLS-DA model can effectively distinguish the products between geographical indication and non-geographical indication, achieving the identification accuracy of 0.928, R^2_X of 0.828, R^2_Y of 0.621, Q^2 of 0.553, and AUC of 0.980. Overall, this research provides a rapid, online, efficient and non-destructive quality prediction and geographical indication discrimination method for the Kokotireke grape industry, which is of great significance for ensuring product quality and avoiding the geographical indication adulteration and also offers the reference for related research in other fruits.

Key words: kokotireke grapes; visible/near-infrared spectroscopy; quality prediction; geographical indication product identification; spectral preprocessing; chemometric modeling

科克铁热克葡萄作为新疆地理标志(地标)水果,以其独特地理环境孕育出的高品质水果闻名,在新疆农业与果蔬产业中占据重要地位,具有极高的经济价值[1]。然而,在市场利益的驱使下,科克铁热克葡萄市场存在地标掺假现象,这主要归因于地标与非地标产品的科克铁热克葡萄,即地标与非地标产品在品质和价格上存在的显著差异。科克铁热克葡萄富含糖、酸、矿物质、维生素等丰富的营养成分,对保持人体健康发挥着重要作用。可溶性固形物(SSC)和可滴定酸(TA)是衡量水果品质的关键指标,在评估葡萄成熟度、指导加工与保鲜工作方面意义重大。因此,开发一种能够同时准确预测科克铁热克葡萄 SSC 和 TA,并且可以有效区分地标与非地标产品的方法显得尤为重要。然而,用于评估葡萄品质的传统方法,如色谱法、酸碱滴定法等破坏性技术,不仅耗费大量人力、财力,且难以满足大规模、快速检测的需求。相比之下,可见/近红外光谱技术作为一种无损、成本低廉且高效的检测手段,成为理想的选择[2]。当光与水果相互作用时,水果内部化学成分所含有的分子键例如 C-H、O-H 会产生振动,可被光谱仪所捕获形成光谱[3]。水果光谱蕴含着大量化学成分信息,借助化学计量学方法,能够建立起光谱与品质之间的关联,从而实现对水果品质的预测以及地理标志产品的识别[4]。

偏最小二乘回归(PLSR)和偏最小二乘判别分析(PLS-DA)作为经典的化学计量学方法,因其具备较强的处理高维光谱信息的能力和较高的鲁棒性,成为分析光谱数据的重要工具[5]。已有研究证实了可见/近红外光谱技术在水果品质检测和产地判别方面的可行性。例如,可见/近红外光谱技术已经实现了对苹果[6]、柑橘[7] 品质的定量分析。此外,李佳洁等[8]通过使用可见/近红外光谱技术实现了对小米的产地溯源,且 Huck-Pezzei 等[9] 针对阿尔卑斯山的三种典型产品(苹果、奶酪和肉类)的可见/近红外光谱建立判别模型,完成了阿尔卑斯山产品的鉴定工作。然而,目前尚未有研究使用可见/近红外光谱技术同时实现科克铁热克葡萄的品质预测和地标识别。

基于此,本研究选取地标产品和非地标产品的科克铁热克葡萄作为研究对象,在获取样本后立即进行光谱信息采集,同时准确测定其 SSC 和 TA,同时结合光谱预处理技术建立 PLSR 模型进行品质定量分析,并以 PLS-DA 模型进行地标产品判别。这项研究为科克铁热克葡萄的品质控制和地理标志产品保护提供了科学依据和技术支持。

1　材料与方法

1.1　材料与仪器

科克铁热克葡萄　品种为'木纳格',采集于新疆科克铁热克乡,其地理标志(地标)保护的区域范围为皮山县科克铁热克乡全乡 25 个行政村,地理坐标为东经 77°31′00″~79°38′00″,北纬 35°22′00″~39°01′00″。相同品种的非地理标志(非地标)葡萄,采集于新疆克孜勒苏柯尔克孜自治州和吐鲁番市。地标区域属暖温带大陆性干旱气候,年均日照超 2800 h,昼夜温差超 15 ℃,独特光热条件促进葡萄糖分积累与风味物质合成。此外,产区土壤以沙质土为主,透气性佳、矿物质丰富,助力根系发育。非地标产品虽为同品种,而克孜勒苏柯尔克孜自治州和吐鲁番市则是典型高温干燥环境,且光照时长、温差幅度、土壤质地等均与科克铁热克乡有所差异;氢氧化钠、1% 酚酞指示剂　上海国药集团化学试剂有限公司。

DT380 型台式可见/近红外光谱采集装置　无锡迪特莱科技有限公司;PAL-1 型糖度仪　日本岛津有限公司;GZL-1.7 型真空冷冻干燥机　奥特佳美科技(北京)有限公司;AD500S-H 型数显型均质机　广州市联合机械有限公司。

1.2　实验方法

1.2.1　样品采集　在葡萄果园内,从众多葡萄行中随机挑选出两行葡萄植株,并将其设为固定采样点,确保样本的随机性。在采样时,充分考虑果穗不同位置

和受光面的差异，对果穗的上部、中部、下部以及向阳面和背阴面均进行采样。

具体采样操作采用五点取样法，在选定的每穗葡萄上，选取五个具有代表性的位置，采集 5 粒表皮无损伤、无病害迹象的葡萄颗粒，保证样本的高质量。为保证样本的代表性，每颗葡萄藤上只采取一次，共采集科克铁热克葡萄 80 组，共 400 粒，色格孜库勒葡萄 45 组，共 225 粒。每组葡萄样本单独编号后，立刻冷链运输至江南大学协同创新中心实验室，并在室温下存放 8 h 后采集光谱数据，以降低温度对光谱数据的干扰。

1.2.2 可见/近红外光谱采集 利用台式光纤光谱仪对科克铁热克葡萄进行可见/近红外光谱数据采集。如图 1 所示，该设备主要由承托装置和光谱采集装置构成。承托装置包括水果托盘和葡萄；光谱采集装置包括光源、光谱仪和计算机。光源为两盏 50 W 的 HL2000 卤钨灯；光谱仪配备有光纤，放置在承托装置下方的暗室中，以防止环境杂散光的干扰，暗室顶部固定一个带有信号通道的光纤，形成一个直径为 10 mm 的信号窗口，光源与信号窗口之间的距离设定为 180 mm。光谱数据在 640~1048 nm 可见/近红外波段范围内采集，积分时间 1 ms，分辨率 2 nm，漫反射方式。

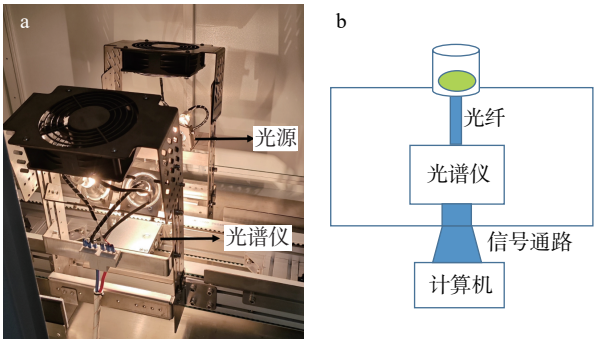


图 1 光谱采集台式设备的现场图(a)和台式设备示意图(b)
Fig.1 Spectral acquisition in on-site picture (a) and schematic diagram (b) of of the desktop device

光谱测量在一个配备空气循环系统的控制室进行(海尔科技有限公司，中国青岛)，室内温度保持在 25 ℃，以确保数据稳定性。测量前将光谱仪预热，光谱数据由计算机收集；测量时将葡萄放置在专用托盘上，测量过程沿葡萄样本赤道以 120°间隔测量三次取平均值作为每粒葡萄的光谱数据，每组五粒葡萄光谱数据取平均作为最终原始光谱数据。

1.2.3 可溶性固形物测定 采集完光谱数据的葡萄，每组选 3 粒葡萄破碎后取汁液，利用手持式糖度仪，测定糖度，每组测定 6 个平行数据，并求平均值。每组剩余的葡萄，以 1 cm 的厚度切片冻干 48 h。冻干后样品于−80 ℃ 冰箱保存，用于后续可滴定酸测定。

1.2.4 可滴定酸测定 采用酸碱滴定法测量酸度。取冻干后的样品 0.5 g，加蒸馏水复溶，利用高压均质机制成均一溶液。加入 2 滴 1% 酚酞指示剂，用 0.1 mol/L 氢氧化钠溶液进行滴定。滴定至溶液初显粉色并在 0.5 min 内不褪色时为终点，记录氢氧化钠滴定液的用量，重复三次。并以蒸馏水作为空白对照，进行滴定。

根据氢氧化钠溶液消耗量，计算可滴定酸含量，以质量分数(%)表示。计算公式：

$$可滴定酸(\%)=\frac{c\times(V_1-V_0)\times f}{m\times 2.5}\times 100 \qquad 式（1）$$

式中：c 表示氢氧化钠滴定液浓度，mol/L；V_1 表示滴定时消耗的氢氧化钠溶液体积，mL；V_0 表示滴定蒸馏水消耗的氢氧化钠溶液体积，mL；m 表示样品质量，g；f 表示折算系数，葡萄为酒石酸，取 f=0.075 g/mmol；2.5 表示冻干葡萄复水比。

1.2.5 样本集划分

1.2.5.1 品质预测模型 将得到的 125 条光谱按照比例 4∶1 分层随机划分，4 份作为模型训练集(样本集)，1 份作为预测集检验模型的性能。

1.2.5.2 地标产品识别模型 将地标产品的 80 条和非地标产品的 45 条光谱数据，分别按照 4∶1 比例划分为训练集和预测集。

1.2.6 光谱预处理 在光谱法测量过程中，环境、测量条件以及仪器本身的特点，对测量结果都有较大影响[1]。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它光谱信息，还有周围杂散光等[10]。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。本研究系统采用了多种光谱预处理技术对原始光谱(RAW)数据进行处理，以提升数据质量与分析性能。

采用多种光谱预处理技术以提升数据质量与分析性能。具体采用了：一阶导数(1stD)[11] 预处理、Savitzky-Golay 卷积平滑(SG)[12] 预处理、多元散射校正(MSC)[13] 预处理、局部基线校正(LBC)[14] 预处理、最大值归一化(Max Nor)[15] 预处理。各预处理方法的协同应用构建了多层次的数据优化体系，为后续建模提供了高保真、低噪声的光谱特征输入。

1.2.7 建模

1.2.7.1 品质预测模型 采用 Unscrambler X10.4 版本软件构建偏最小二乘回归(PLSR)模型进行 SSC 与 TA 品质预测。Unscrambler X 软件是一款开源的数据分析软件，广泛用于光谱预处理和模型构建。使用五折交叉验证方式评价模型性能，具体过程为：将训练集拆分为五份子集，以其中一份作为验证、另外四份为训练，进行建模，共进行五轮验证，其目的是最小化均方根误差，选出模型最佳成分数，最后以预测集作为外部验证集评估模型的泛化能力，其输出指标包括训练集指标 RMSEC、R^2_C 以及预测集 RMSEP、R^2_P。

1.2.7.2　地标产品识别模型　采用 SIMCA 14.1 版本软件(MKS Data Analytics Solutions，瑞典)建立偏最小二乘判别分析(PLS-DA)模型进行分析，对地标产品进行判别分析。SIMCA 版本软件是也一款开源的数据分析软件，广泛用于构建各种化学计量学判别分析模型。为评估模型的预测能力与可靠性，判断是否存在过拟合现象，采用七折交叉验证与置换验证方法。其中，置换验证进行 200 次，具体操作为对因变量的类别标签随机置换 200 次，每次置换后重新构建 PLS-DA 模型并计算性能指标。此外，绘制受试者工作特征曲线(ROC)，并计算其曲线下面积(AUC)。开展混淆矩阵(Confusion Matrix)检验以计算地标识别准确率。通过计算自变量拟合指数(R^2_X)、因变量拟合指数(R^2_Y)、模型预测指数(Q^2)以及变量重要性投影(VIP)值，对模型进行全面评价。

1.3　数据处理

利用 Microsoft Excel 2021 软件对原始光谱数据进行整理与计算，对实验获得的重复测量数据分别计算其平均值和标准差。利用 SPSS 12.0 采用 t 检验对品质数据进行显著性差异分析，P<0.05 表示差异显著。所有图由 Origin Pro 2021 绘制。

2　结果与分析

2.1　样本 SSC 和 TA 分析

图 2 展示了所有样本中 SSC 和 TA 的频率分布，表 1 为不同数据集中 SSC 和 TA 含量的具体情况。由图 2A、图 2B 和表 1 可知，所有样本的 SSC 和 TA 含量呈正态分布，且分布区间为 13.80%~20.80% 和 0.52%~1.42%，平均值分别为 17.68% 和 0.90%，标准差分别为 1.34% 和 0.17%，表明样本的分布范围较广，采样具有代表性。吴虹璋等[16] 利用可见/近红外光谱成功建立了葡萄 SSC 含量的定量分析，其中红提葡萄、巨峰葡萄和辽峰葡萄的糖度含量分布区间分别为 11.6%~14.8%、13.4%~16.4% 和 15.6%~18.1%，与本实验科克铁热克葡萄糖度分布相似。同一产地的科克铁热克葡萄 SSC 和 TA 存在差异，可能是由于遗传变异所致[17]，而不同产地的科克铁热克葡萄品质差异，可能是由于不同地区的土壤成分、光照条件和栽培技术等因素会影响同品种果实的形态、大小、颜色、味道和成分[18]。另外，总集、训练集和预测集中的各参数(最大值、最小值、平均值和标准偏差)结果相近，表明数据集划分具备一致性和可靠性，有助于提高预测模型的稳定性[19]。图 2C 和图 2D 显示，地标产品与非地标产品的 SSC 含量具有显著性差异(P<0.05)，而 TA 含量没有显著性差异(P>0.05)。高升等[20] 基于可见/近红外漫反射光谱，成功设计了便携式生长期红提葡萄多品质参数无损检测仪，SSC 和 TA 差异与本研究结果基本一致。

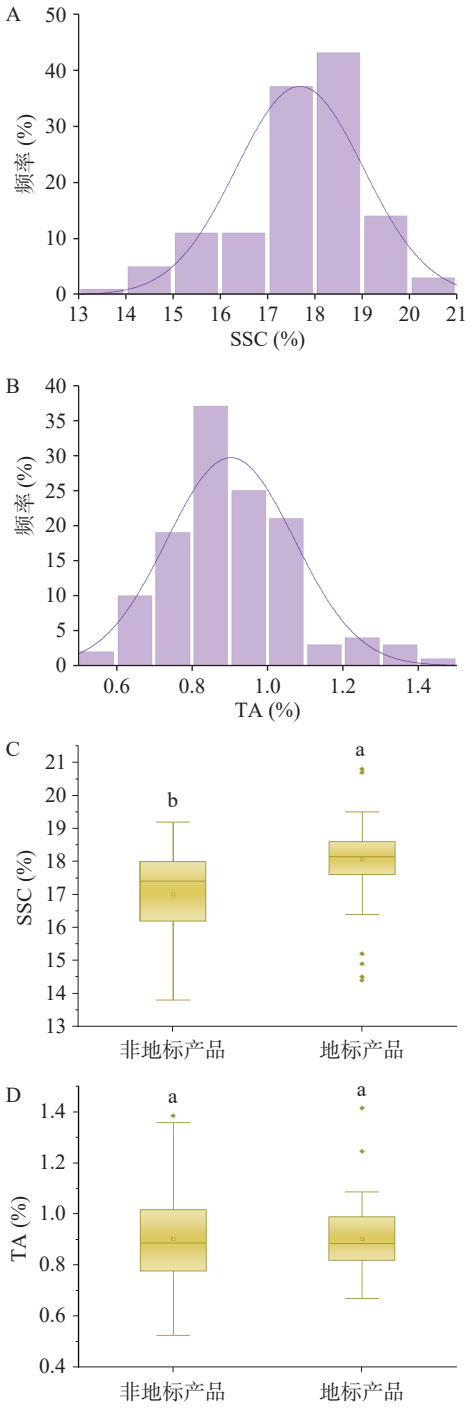


图 2　不同样本集中的 SSC(A 和 C)和 TA(B 和 D)的频率分布

Fig.2　Frequency distribution diagrams of SSC (A and C) and TA (B and D) in different sample sets

注：不同字母表示差异显著(P<0.05)。

2.2　科克铁热克葡萄的光谱特征

图 3 展示了科克铁热克葡萄在经过不同预处理后的可见/近红外光谱图，为深入了解科克铁热克葡萄的光谱特征提供了丰富信息。由图 3 可知，不同的光谱虽有重叠但整体趋势有较高一致性。葡萄的主要成分包括水分、糖类、有机酸、色素和酚类物质，光谱图中蕴含着丰富的葡萄成分信息[21]。水分的 O-H 键伸缩振动以倍频与合频形式在近红外区产生吸收；糖类的 C-H、O-H 键，有机酸的 C=O 键、O-

表 1　所有样本中 SSC、TA 值分布

Table 1　Distribution of SSC and TA values in all samples

数据集	数量	SSC(%)				TA(%)			
		最小值	最大值	平均值	标准偏差	最小值	最大值	平均值	标准偏差
总集	125	13.80	20.80	17.68	1.34	0.52	1.42	0.90	0.17
训练集	93	13.80	20.80	17.70	1.35	0.52	1.42	0.90	0.17
预测集	32	14.50	20.80	17.64	1.33	0.65	1.34	0.91	0.16
地标产品	80	14.4	20.8	18.07	1.23	0.67	1.42	0.90	0.15
非地标产品	45	13.8	19.2	17.00	1.27	0.52	1.39	0.90	0.18

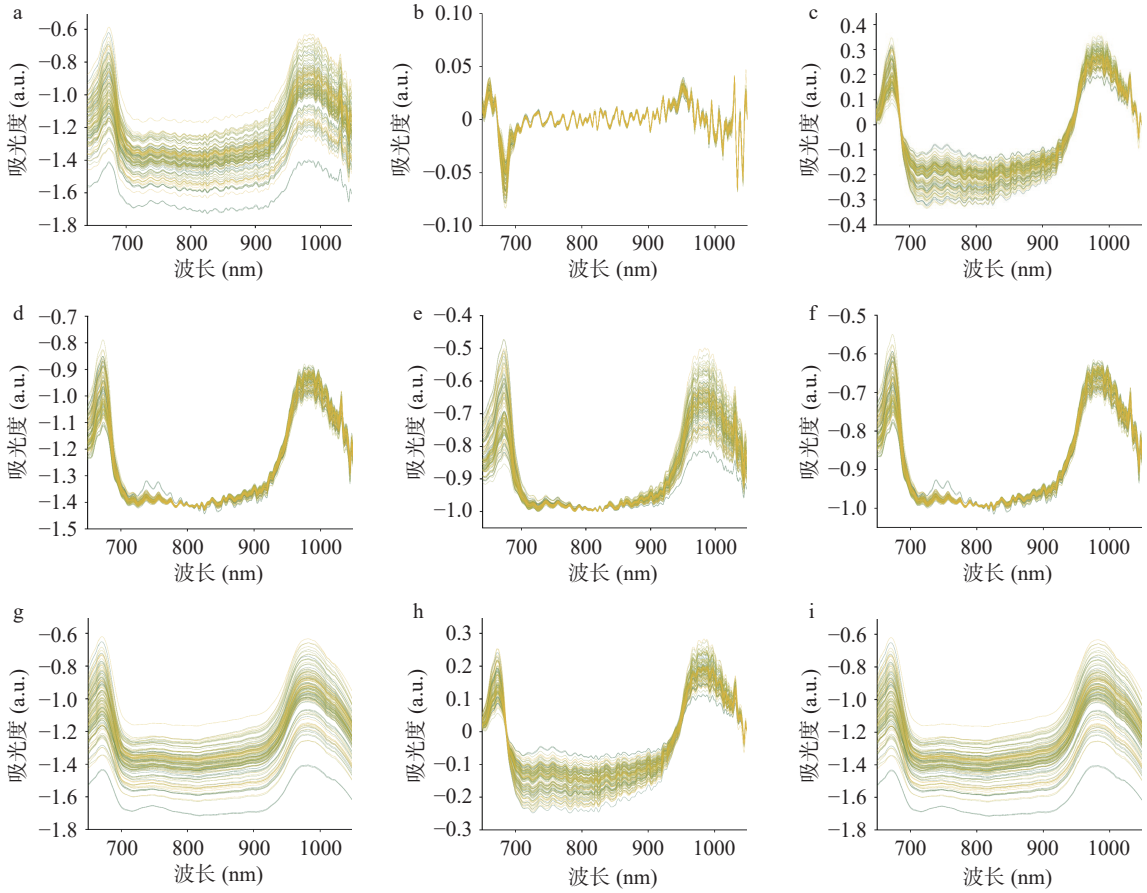


图 3　所有葡萄的原始光谱图(a)以及预处理后(b~i)的光谱图

Fig.3　Raw spectral diagrams (a) of all grapes and the spectral diagrams after preprocessing (b~i)

注: b: 1stD; c: LBC; d: MSC; e: Max Nor; f: Max Nor-MSC; g: SG; h: Max Nor-LBC; i: SG-LBC-1stD。

H 键和 C-H 键, 蛋白质的 N-H 键、C=O 键, 以及酚类物质的 C-H 键、C=C 键, 它们的振动在近红外区均有相应吸收[21−22]。在采集到的光谱图中(图 3a), 有两个明显的吸收峰, 分别在 675 和 970 nm 处。675 nm 吸收峰可能与 N-H 键伸缩振动倍频或合频相关, 该吸收峰蕴含叶绿素的信息[23]。而 970 nm 吸收峰通常源于水分子 O-H 键伸缩振动的一阶倍频, 对研究葡萄水分状态与糖酸含量具有重要意义[24]。

然而, 其他成分的散射信号和仪器噪声可能掩盖掉了其他特征吸收峰, 为了进一步解释光谱数据中的潜在吸收峰, 采用了多种光谱预处理方法(图 3b~图 3i)。经过 LBC、MSC、Max Normalization 和 1stD 预处理后, 在 675 和 970 nm 处出现了更明显的吸收峰。此外, 经过 1stD 预处理后的光谱 690 nm 处出现明显波谷, 但在 700~900 nm 波段范围内却出现了杂峰, 这可能归因于这种预处理方式将仪器噪声放大, 扰乱了主要特征峰识别[25]。因此, 需要提前处理背景噪声, 选用 SG 处理和 LBC 处理, 动态识别特征峰外的背景信号区域, 实现复杂基线的自适应修正[26]。SG-LBC-1stD 联合预处理后的光谱图, 在扣除背景噪声的同时, 增强了特征峰的强度, 可以看到在 760 nm 处还出现了一个潜在的吸收峰, 这一般认为是 O-H 键伸缩振动与弯曲振动的合频吸收峰, 这反映了葡萄中水、糖和酸这几个主要成分的信息。在其它果蔬中也有类似的研究结果, 例如苹果[27]、西瓜[28]、香蕉[29]、西红柿[30] 和腌黄瓜[31]。

2.3　不同预处理方法对科克铁热克葡萄品质预测模型性能的影响

针对不同预处理方法, 分别构建了 PLSR 预测

表 2　不同预处理方法的 PLSR 预测模型预测集结果

Table 2　Results of the prediction set of the PLSR prediction model for different pretreatment methods

预处理方法	SSC					TA				
	R_c^2	RMSEC	R_P^2	RMSEP	最佳因素数	R_c^2	$RMSE_C$	R_P^2	RMSEP	最佳因素数
SG	0.7181	0.7084	0.9102	0.3985	1	0.8913	0.0559	0.8621	0.0598	2
1stD	0.7632	0.6491	0.5757	0.8662	4	0.5545	0.1132	0.6674	0.0929	2
Max Nor	0.7896	0.6119	0.8031	0.5901	5	0.6974	0.0933	0.8121	0.0670	3
MSC	0.6807	0.7539	0.7280	0.6935	5	0.4193	0.1227	0.4193	0.1227	6
LBC	0.5762	0.8686	0.6232	0.8162	4	0.4703	0.1234	0.6273	0.0983	2
Max Nor-LBC	0.7174	0.7092	0.7825	0.6201	4	0.8370	0.0685	0.8386	0.0647	2
Max Nor-MSC	0.6807	0.7539	0.7280	0.6935	5	0.7592	0.0832	0.4193	0.1227	6
SG-LBC-1stD	0.9076	0.4056	0.9361	0.3362	2	0.9433	0.0399	0.9478	0.0368	9

模型，表 2 展示了各模型对 SSC 和 TA 含量的预测性能。图 4 是本研究筛选出的最优预处理方式(SG-LBC-1stD)对训练集和测试集的预测值与观测值的散点图。

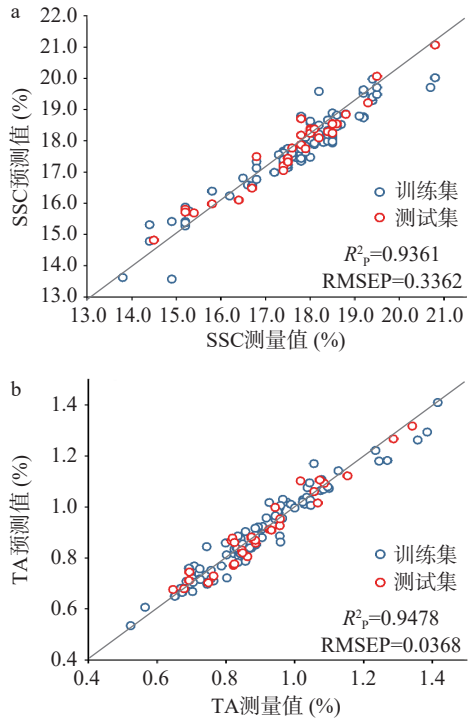


图 4　SG-LBC-1stD 预处理后 SSC(a)和 TA(b)训练集和测试集的散点图

Fig.4　Scatter plots of predicted values versus observed values in SSC (a) and TA (b) for the training set and the test set after SG-LBC-1stD preprocessing

在 SSC 预测方面，SG-LBC-1stD 预处理的模型 R^2 为 0.9361，RMSEP 为 0.3362，最佳因素数为 2。较高的 R_P^2 表明该模型对 SSC 观测值的拟合程度较好，而较低的 RMSEP 意味着预测值与实际值的偏差较小，最佳因素数则反映了模型复杂度与解释能力的平衡[32]。对比其他预处理方法，如 SG、LBC 和 1stD，其 R_P^2 分别为 0.9102、0.6232 和 0.5757，RMSEP 分别为 0.3985、0.8162 和 0.8662，因素数为 1、4 和 4，可以看出 SG-LBC-1stD 的组合预处理在拟合度和精度上均更优。SG 预处理后的模型与 SG-LBC-1stD 预处理后，R_P^2、RMSEP 有显著性差异，且因素更小，说明通过联合预处理后增强了模型的拟合能力并且能够从光谱数据中提取更多的信息[33]。另外，除了 SG-LBC-1stD，其他预处理方式均降低了预测模型的性能。结合图 3 可知，这些预处理方式增强了杂峰强度，因素数更大，提高了模型复杂程度引入了冗余信息。而采用了 MSC 和 LBC 预处理后(图 3c、图 3d)，观察到光谱数据出现不期望的改变，这可能会导致重要波长或波段的丢失。对于 TA 预测，各预处理方法也呈现出不同的结果。SG-LBC-1stD 的模型 R_P^2 为 0.9478，RMSEP 为 0.0368，最佳因素数为 9。SG 预处理的 R_P^2 为 0.8621，RMSEP 为 0.0598，最佳因素数为 2。通过对比这些参数，发现 SG-LBC-1stD 预处理 R_P^2 明显高于 SG 预处理，说明前者构建的模型在 TA 预测上对观测数据的解释能力更强，并且因素数高，明显增强了预测模型的拟合程度，且泛化能力较强。

对光谱应用的每种预处理方法都有其独特的优点和局限性，这会对预测模型的性能产生显著影响。在以往的研究中，对于可见/近红外光谱分析，通常采用平滑和标准化相结合的方法。如 SNV 和 SG 及 MSC 预处理分别提高了 PLSR 模型对柑橘[34]和梨[35]内部品质预测的准确性。

综合对 SSC 和 TA 的预测模型，发现 SG-LBC-1stD 预处理后的模型最优。另外，对于 SSC 预测模型，采用 SG 处理后的光谱与 SSC 之间呈现了很强的线性关系，这可能是由于 SG 预处理减少了光谱噪声。Jiang 等[36]基于可见/近红外光谱，建立机器学习模型实现了对柑橘 SSC 和 TA 含量的在线监测，相志勇等[37]的研究也有类似的结果。PLSR 模型能够提取与 SSC 相关的高维光谱信息，对多重共线性表现出良好的鲁棒性，同时还具备抵抗异常值干扰的能力。正因如此，PLSR 模型成为分析光谱与水果品质的得力工具。例如，在测定菠萝[38]的 SSC 时，采用基于 MSC 光谱的 PLSR 模型，得到的决定系数 R^2 为 0.88，均方根误差 RMSE 为 1.04。而在预测番茄[39]的 SSC 时，使用基于原始光谱的 PLSR 模型，所得决定系数 R^2 为 0.860、均方根误差 RMS 为 0.201。这些研究已证实，无论何种水果类型以及光谱预处理方式，PLSR 模型均可优先用于分析可见/近

红外光谱，以准确预测 SSC，这一观点在本文的研究中通过对比分析得到了进一步验证。其他采用 PLSR 模型对 TA 的预测，相对预测性能较差，如梁子兆等[40]运用 SNV 处理方法的 PLSR 模型，对 TA 的预测效果 R^2 仅为 0.842。但本文的 SG-LBC-1stD 模型对 TA的预测效果较好，可能是因为这种预处理方式与水果和光谱仪的适配度最佳[41−42]。

2.4 不同预处理方法对科克铁热克葡萄地理标志产品识别模型性能的影响

PLS-DA 是一种有监督的判别分析统计方法，它通过主成分将自变量(X)与因变量(Y)的相关性最大化，从而建立分类预测模型。其核心作用在于通过降维简化高维数据结构，有效区分不同类别样本(如地标产品与非地标产品)。图 5 为使用不同预处理光谱进行 PLS-DA 模型建模后，进行地标与非地标产品的判别分析结果，SG-LBC-1stD 预处理后的光谱判别效果最佳。

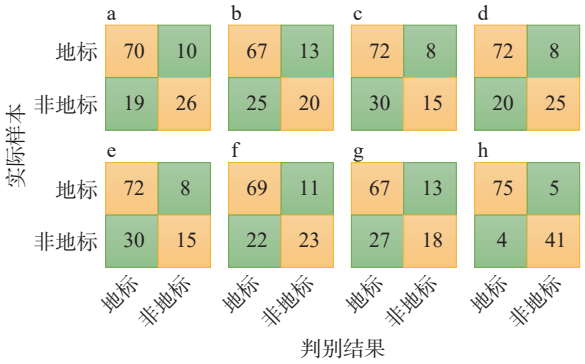


图 5 不同预处理后光谱建立 PLS-DA 模型的混淆矩阵

Fig.5 Missclassification tables of the PLS-DA model established with spectra after different preprocessing

注：a: 1stD; b: LBC; c: MSC; d: Max Nor; e: Max Nor-MSC; f: SG; g: Max Nor-LBC; h: SG-LBC-1stD。

对 SG-LBC-1stD 预处理后的光谱的 PLS-DA 建模进行分析，如表 3 所示，R^2_X 为 0.828，R^2_Y 为 0.621，Q^2 为 0.553，均大于 0.5，模型拟合结果良好。图 6a 为混淆矩阵的结果，可知地标产品与非地标产品的样本有明显分离效果。禹洁等[43]基于稳定同位素与矿物元素利用 PLS-DA 模型对蜂蜜进行产地溯源，实现了产地之间良好的分离效果。

表 3 不同预处理方法的 PLS-DA 预测模型预测集结果

Table 3 Results of the prediction set of the PLS-DA prediction model with different pretreatment methods

预处理方法	R^2_X	R^2_Y	Q^2	识别准确率	AUC值
SG	0.994	0.271	0.231	0.736	0.826
1stD	0.522	0.254	0.229	0.768	0.818
Max Nor	0.619	0.205	0.189	0.776	0.794
MSC	0.287	0.148	0.094	0.696	0.740
LBC	0.735	0.197	0.179	0.696	0.781
Max Nor-LBC	0.795	0.206	0.192	0.680	0.786
Max Nor-MSC	0.287	0.148	0.096	0.696	0.741
SG-LBC-1stD	0.828	0.621	0.553	0.928	0.980

用已构建的 PLS-DA 模型进行 200 次置换检验，如图 6b 所示，最右侧的 R^2 值和 Q^2 值均大于 0.5，且均高于左侧的 R^2 值和 Q^2 值，并且 Q^2 回归线和纵轴的截距小于 0，说明构建的 PLS-DA 模型未过拟合，具有较好的预测能力。因此，SG-LBC-1stD 模型可以用于地标产品与非地标产品的判定识别，最终该模型对产品判别的平均准确率高达 92.8%(表 3)。

根据 PLS-DA 生成的 VIP 值来筛选差异特征波

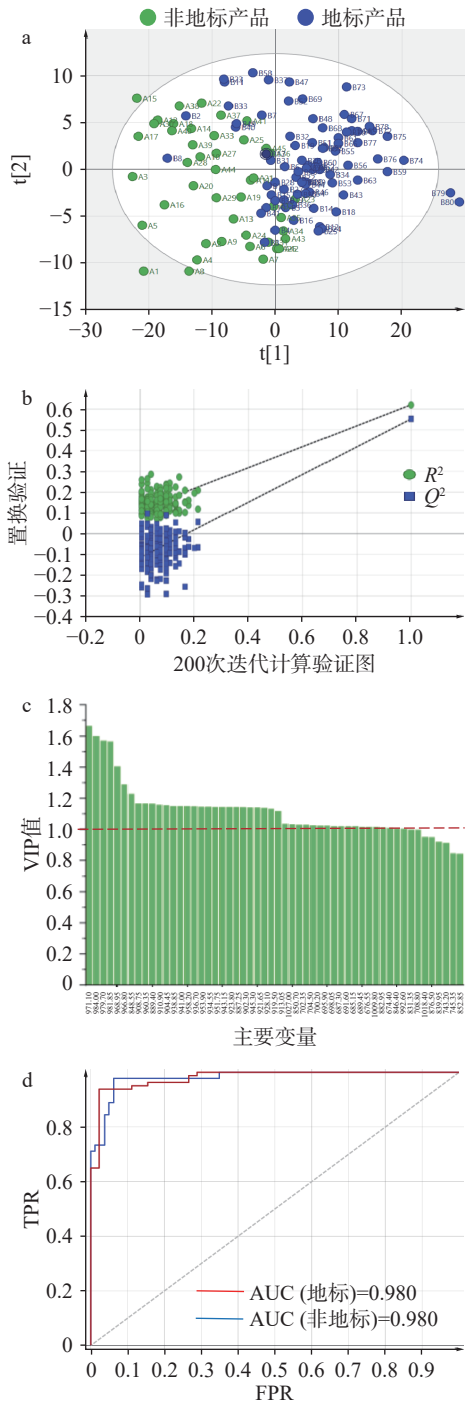


图 6 地标产品与非地标产品的 SG-LBC-1stD 预处理光谱 PLS-DA 得分图(a)、模型置换检验结果(b)、VIP 图(c)与 ROC 曲线图(d)

Fig.6 SG-LBC-1stD pretreatment spectral PLS-DA score plots (a), model permutation test results (b), VIP plot (c), and ROC curve (d) for products with and without geographical indication

长（VIP>1），VIP 值越大，说明该差异成分对样品组间的分类影响强度越大。在本研究中，VIP 值最大的波长为 971.1 nm，其他特征波长主要集中在 900~1000 nm，还有部分特征波长范围集中在 680~700 nm，这一结果也分别对应了与糖酸相关和颜色相关的特征波长（图 6c）。

此外，由模型所得的 ROC 曲线可知（图 6d），地标与非地标产品均具有较高的 AUC（0.980），表明本研究有效实现了科克铁热克葡萄的产地溯源。在先前的研究中，使用传统元素分析法实现农产品的地理溯源已经成为一种公认可靠的方法。例如，陈秋生[44]基于元素指纹图谱对玫瑰香葡萄进行产地溯源，5 种特征元素：K、Ca、Cu、Na 和 Mn，而后建立 PLS-DA 模型，其产地判别率达到 100%。显然，元素分析法具有更高的识别准确率。然而，相比于传统元素分析法，本研究具有三个明显优势：相比于元素分析法，可见/近红外光谱技术具有无损检测优势，实际应用中只需将光谱仪探头靠近或接触样品表面，即可快速获取光谱数据，不会对样品造成任何物理损伤，尤其适合像葡萄这类需要保持外观和完整性用于销售的农产品；本研究使用的台式可见/近红外光谱设备可大大提高在线检测效率，具备工业化和商业水平，可大规模应用以满足市场监管和生产企业检测需求；可见/近红外光谱技术所需的仪器维护成本相对较低，且单个产品检测成本低，无需使用大量的化学试剂，是一种环境友好型方法。本研究的结果与之相似，Li 等[45]利用可见/近红外光谱技术对鸭蛋进行产地分类，建立随机森林和支持向量机模型的最终检测准确率分别为 90% 和 91.1%。这项研究仅用于产地分类任务，而本研究具备多任务分析潜力，可同时实现地标判别分析和品质回归预测。

3　结论

基于可见/近红外光谱技术结合化学计量学模型，本研究实现了对科克铁热克葡萄品质预测和地标产品识别的双重任务。首先，明确了科克铁热克葡萄 SSC 和 TA 的分布特征，以及地标和非地标样本的品质差异，为后续建模奠定坚实的数据基础。光谱分析过程中，不仅识别出原始光谱中关键吸收峰的物质关联，还通过多种预处理技术对比，确定了 SG-LBC-1stD 预处理在优化光谱数据、提升模型性能方面的显著优势。其中，PLSR 模型对科克铁热克葡萄的 SSC 和 TA 实现了精准预测，其中 SSC 预测的 R^2 为 0.9361 和 RMSEP 为 0.3362，TA 预测的 R^2 为 0.9478 和 RMSEP 为 0.0368，展示出高拟合度和低误差的特性。此外，PLS-DA 模型区分地标与非地标产品时表现卓越，通过 VIP 值筛选出的特征波长为产地判别提供了关键依据，且模型具备良好的稳健性和预测能力，准确率高达 92.8%。该研究成果为科克铁热克葡萄品质把控和地理标志产品保护提供了切实可行的技术方案，有助于规范市场秩序、提升产品竞争力。同时，本研究突破了传统检测方法的局限，为水果品质检测和产地溯源领域提供了创新性的研究范式，有望推动可见/近红外光谱技术在农业领域的广泛应用和深度发展。但是，本研究在模型普适性方面仍有提升空间，后续可考虑纳入更多产地、不同年份的样本，进一步优化模型，增强其在复杂环境下的适应性和可靠性，更好地服务于实际生产。

© 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] 李杰, 魏佳, 张政, 等. 二氧化硫(SO_2)熏蒸改善木纳格葡萄的采后品质[J]. 现代食品科技, 2020, 36(2): 114−121,151. [LI Jie, WEI Jia, ZHANG Zheng, et al. Improving on postharvest quality of munag grape berries fumigated by sulfur dioxide(SO_2)[J]. Modern Food Science and Technology, 2020, 36(2): 114−121,151.]

[2] 郭志明, 桑伟兴, 杨忱, 等. 近红外光谱及成像在果品无损检测中的应用[J]. 包装与食品机械, 2024, 42(5): 1−14. [GUO Zhiming, SANG Weixing, YANG Chen, et al. Application of nondestructive detection of fruit by near-infrared spectroscopy and imaging[J]. Packaging and Food Machinery, 2024, 42(5): 1−14.]

[3] DING Gang, JIN Ke, CHEN Xiaoya, et al. Non-destructive prediction of ready-to-eat kiwifruit firmness based on Fourier transform near-infrared spectroscopy[J]. Postharvest Biology and Technology, 2024, 212: 112908.

[4] MARÍN-SAN ROMÁN S, FERNÁNDEZ-NOVALES J, CEBRIÁN-TARANCÓN C, et al. Application of near-infrared spectroscopy for the estimation of volatile compounds in *Tempranillo blanco* grape berries during ripening[J]. Journal of the Science of Food and Agriculture, 2023, 103(13): 6317−6329.

[5] ZHENG Runyu, JIA Yuyao, ULLAGADDI C, et al. Optimizing feature selection with gradient boosting machines in PLS regression for predicting moisture and protein in multi-country corn kernels *via* NIR spectroscopy[J]. Food Chemistry, 2024, 456: 140062.

[6] 张宇, 张重阳, 段鑫鑫, 等. 基于线性判别分析和机器学习的可见-近红外光谱苹果损伤分级[J]. 食品科学, 2024, 45(22): 255−261. [ZHANG Yu, ZHANG Chongyang, DUAN Xinxin, et al. Visual/near-infrared spectroscopy combined with linear discriminant analysis and machine learning for classification of apple damage[J]. Food Science, 2024, 45(22): 255−261.]

[7] 刘豪, 龚中良, 文韬, 等. 基于可见/近红外光谱的柑橘糖度在线检测分选系统的设计与试验[J]. 湖南农业大学学报(自然科学版), 2023, 49(4): 497−502. [LIU Hao, GONG Zhongliang, WEN Tao, et al. Design and experiment of an online detection and sorting system for *Citrus* sugar content based on visible/near infrared spectroscopy[J]. Journal of Hunan Agricultural University (Natural Sciences), 2023, 49(4): 497−502.]

[8] 李佳洁, 吴建虎, 张海波. 利用可见/近红外光谱技术对小米产地进行溯源研究[J]. 食品安全质量检测学报, 2017, 8(8): 3037−3043. [LI Jiajie, WU Jianhu, ZHANG Haibo. Discrimination of the origin of millet by visual/near infrared reflectance spectroscopy[J]. Journal of Food Safety and Quality, 2017, 8(8): 3037−3043.]

[9] HUCK-PEZZEI V A, SEITZ I, KARER R, et al. Alps food authentication, typicality and intrinsic quality by near infrared spec-

troscopy[J]. Food Research International, 2014, 62: 984−990.
[10] DU Zhenjiao, TIAN Wenfei, TILLEY M, et al. Quantitative assessment of wheat quality using near-infrared spectroscopy: A comprehensive review[J]. Comprehensive Reviews in Food Science and Food Safety, 2022, 21(3): 2956−3009.
[11] YANG Xinya, ZHUANG Xiaoqi, SHEN Rongjing, et al. *In situ* rapid evaluation method of quality of peach kernels based on near infrared spectroscopy[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2024, 313: 124108.
[12] 柳薇, 邱熙文, 蒋立文, 等. 基于近红外光谱技术和变量筛选-偏最小二乘判别分析方法的铁皮石斛产地无损溯源[J]. 分析测试学报, 2025, 44(2): 246−252. [LIU Wei, QIU Xiwen, JIANG Liwen, et al. Origin traceability of *Dendrobium candidum* with near infrared spectroscopy and variable selection-partial least squares discriminant analysis[J]. Journal of Instrumental Analysis, 2025, 44(2): 246−252.]
[13] 刘文政, 周雪健, 平凤娇, 等. 基于可见-近红外光谱的鲜食葡萄成熟品质关键指标检测[J]. 农业机械学报, 2024, 55(2): 372−383. [LIU Wenzheng, ZHOU Xuejian, PING Fengjiao, et al. Detection of key indicators of ripening quality in table grapes based on visible-near-infrared spectroscopy[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(2): 372−383.]
[14] JIANG Tao, DING Jianjun, YUAN Shaofeng, et al. Benchtop Vis-NIR spectroscopy meets machine learning for multi-task analysis in Hongmeiren *Citrus*: Geographical origin identification and antioxidant component quantification[J]. Food Chemistry, 2025, 489: 145007.
[15] SHI Shijie, ZHANG Wenhui, MA Yingying, et al. Near-infrared spectroscopy combined with effective variable selection algorithm for rapid detection of rice taste quality[J]. Biosystems Engineering, 2024, 237: 214−219.
[16] 吴虹璋, 蔡红星, 任玉, 等. 基于可见/近红外光谱对葡萄可溶性固形物无损检测研究[J]. 光散射学报, 2024, 36(1): 44−51. [WU Hongzhang, CAI Hongxing, REN Yu, et al. Non-destructive detection of soluble solid content based on visible-near infrared spectroscopy[J]. The Journal of Light Scattering, 2024, 36(1): 44−51.]
[17] LIU Zhongjie, WANG Nan, SU Ying, et al. Grapevine pangenome facilitates trait genetics and genomic breeding[J]. Nature Genetics, 2024, 56(12): 2804−2814.
[18] 克热曼, 赛米, 岳朝阳, 等. 不同立地条件下‘木纳格’葡萄果实品质和矿质元素含量[J]. 北方园艺, 2016(18): 9−13. [KE R M, SAI M, YUE C Y, et al. The fruit quality and mineral element content of 'Munag' grapes under different site conditions[J]. Northern Horticulture, 2016(18): 9−13.]
[19] MARMOR Y N, BASHKANSKY E, MARMOR Y N, et al. Reliability of partitioning metric space data[J]. Mathematics, 2024, 12(4): 603.
[20] 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 Xing, 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.]
[21] 张静, 徐阳, 姜彦武, 等. 近红外光谱技术在葡萄及其制品品质检测中的应用研究进展[J]. 光谱学与光谱分析, 2021, 41(12): 3653−3659. [ZHANG Jing, XU Yang, JIANG Yanwu, et al. Recent advances in application of near-infrared spectroscopy for quality detections of grapes and grape products[J]. Spectroscopy and Spectral Analysis, 2021, 41(12): 3653−3659.]
[22] COZZOLINO D, DAMBERGS R G, JANIK L, et al. Analysis of grapes and wine by near infrared spectroscopy[J]. Journal of Near Infrared Spectroscopy, 2006, 14(5): 279−289.
[23] FERRER-GALLEGO R, HERNÁNDEZ-HIERRO J M, RIVAS-GONZALO J C, et al. Evaluation of sensory parameters of grapes using near infrared spectroscopy[J]. Journal of Food Engineering, 2013, 118(3): 333−339.
[24] BEGHI R, GIOVENZANA V, BRANCADORO L, et al. Rapid evaluation of grape phytosanitary status directly at the check point station entering the winery by using visible/near infrared spectroscopy[J]. Journal of Food Engineering, 2017, 204: 46−54.
[25] HU J Q, CHEN G J, XUE C L, et al. RSPSSL: A novel high-fidelity Raman spectral preprocessing scheme to enhance biomedical applications and chemical resolution visualization[J]. Light: Science & Applications, 2024, 13: 52.
[26] KIM S Y, HONG S J, KIM E, et al. Application of ensemble neural-network method to integrated sugar content prediction model for *Citrus* fruit using Vis/NIR spectroscopy[J]. Journal of Food Engineering, 2023, 338: 111254.
[27] 杜冉. 基于近红外透射光谱技术的苹果内部品质在线检测的研究[D]. 淄博: 山东理工大学, 2008. [DU Ran. Study on online detection of internal quality in apples based on near infrared transmission spectrometric techniques[D]. Zibo: Shandong University of Technology, 2008.]
[28] 田海清, 应义斌, 陆辉山, 等. 基于可见/近红外光谱的乙烯利催熟西瓜与正常成熟西瓜分类试验研究[J]. 光谱学与光谱分析, 2009, 29(4): 940−944. [TIAN Haiqing, YING Yibin, LU Huishan, et al. Study on classification of ethylene treated and non-ethylene treated watermelons by visible/near infrared spectroscopy[J]. Spectroscopy and Spectral Analysis, 2009, 29(4): 940−944.]
[29] ZUDE M. Non-destructive prediction of banana fruit quality using VIS/NIR spectroscopy[J]. Fruits, 2003, 58(3): 135−142.
[30] ZHAO Mingrui, CANG Hao, CHEN Huixin, et al. Determination of quality and maturity of processing tomatoes using near-infrared hyperspectral imaging with interpretable machine learning methods[J]. LWT, 2023, 183: 114861.
[31] KAVDIR I, LU R, ARIANA D, et al. Visible and near-infrared spectroscopy for nondestructive quality assessment of pickling cucumbers[J]. Postharvest Biology and Technology, 2007, 44(2): 165−174.
[32] GAO Sheng, XU Jianhua. Design and test of portable comprehensive quality non-destructive detector for grape bunches based on spectrum[J]. Journal of Future Foods, 2022, 2(3): 275−283.
[33] 张恒, 梁太波, 江鸿, 等. 基于高光谱成像的区域尺度植烟土壤有机质含量和 pH 值估测[J]. 烟草科技, 2024, 57(5): 48−57,63. [ZHANG Heng, LIANG Taibo, JIANG Hong, et al. Estimation of organic matter content and pH value of regional-scale tobacco-growing soils using hyperspectral imaging[J]. Tobacco Science & Technology, 2024, 57(5): 48−57,63.]
[34] CHEN Yuan, GAO Ruizhen, XIANG Jiaxing, et al. Progress of near infrared spectroscopy in tea research[J]. Food Chemistry, 2025, 487: 144691.
[35] LI Yongjie, JIN Guoqiang, JIANG Xin, et al. Non-destructive determination of soluble solids content using a multi-region combination model in hybrid *Citrus*[J]. Infrared Physics & Technology, 2020, 104: 103138.
[36] JIANG Tao, ZUO Weidan, DING Jianjun, et al. Machine learning driven benchtop Vis/NIR spectroscopy for online detection

of hybrid *Citrus* quality[J]. Food Research International, 2025, 201: 115617.
[37] 相志勇, 苗玉彬. 基于 CNN 和近红外光谱的蜜柑 SSC 预测模型研究[J]. 中国农机化学报, 2024, 45(11): 139−144. [XIANG Zhiyong, MIAO Yubin. Study on prediction model of *Citrus* SSC based on CNN and near infrared spectroscopy[J]. Journal of Chinese Agricultural Mechanization, 2024, 45(11): 139−144.]
[38] SEMYALO D, KWON O, WAKHOLI C, et al. Nondestructive online measurement of pineapple maturity and soluble solids content using visible and near-infrared spectral analysis[J]. Postharvest Biology and Technology, 2024, 209: 112706.
[39] FENG L, ZHANG M, ADHIKARI B, et al. Nondestructive detection of postharvest quality of cherry tomatoes using a portable NIR spectrometer and chemometric algorithms[J]. Food Analytical Methods, 2019, 12: 914−925.
[40] 梁子兆, 李欣, 刘朴, 等. 基于可见-近红外光谱的多品种猕猴桃贮藏品质的多指标综合预测模型研究[J]. 食品工业科技, 2025, 46(13): 282−291. [LIANG Zizhao, LI Xin, LIU Pu, et al. Comprehensive multi-indicator prediction model for storage quality of multi-cultivar kiwifruit based on visible-near infrared spectroscopy[J]. Science and Technology of Food Industry, 2025, 46(13): 282−291.]
[41] ZHANG Xiaolei, YANG Jie. Advanced chemometrics toward robust spectral analysis for fruit quality evaluation[J]. Trends in Food Science & Technology, 2024, 150: 104612.
[42] 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. Nondestructive 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.]
[43] 禹洁, 周佳, 黄韡, 等. 基于稳定同位素与矿物元素的蜂蜜产地溯源判别模型的构建[J]. 食品安全质量检测学报, 2024, 15(18): 84−93. [YU Jie, ZHOU Jia, HUANG Wei, et al. Construction of honey origin traceability discriminant model based on stable isotopes and mineral elements[J]. Journal of Food Safety & Quality, 2024, 15(18): 84−93.]
[44] 陈秋生. 基于元素指纹图谱的茶淀玫瑰香葡萄产地溯源技术研究[D]. 天津: 天津市农业质量标准与检测技术研究所, 2015. [CHEN Q S. Research on the origin traceability technology of chadian rose-scented grapes based on elemental fingerprint spectra[D]. Tianjing: Tianjin Institute of Agricultural Quality Standards and Testing Technology, 2015.]
[45] LI Qingxu, ZHOU Wanhuai, WANG Qiaohua, et al. Research on online nondestructive detection technology of duck egg origin based on visible/near-infrared spectroscopy[J]. Foods, 2023, 12(9): 1900.