

高光谱技术融合纹理信息的羊肉总酚浓度快速检测

孙有瑞¹, 郭美¹, 刘贵珊^{1*}, 樊奈昀^{1*}, 张浩楠², 李月¹, 蒲芳宁², 杨世虎¹, 王昊²

1. 宁夏大学食品与葡萄酒学院, 宁夏银川 750021

2. 宁夏大学物理与电子电气工程学院, 宁夏银川 750021

摘要 为了研究可见-近红外(Vis-NIR)高光谱成像对滩羊肉中总酚浓度(TPC)快速检测的可行性, 基于光谱信息融合图像纹理特征建立 TPC 含量的预测模型, 实现滩羊肉中 TPC 含量的可视化表达。将样本集根据 3:1 的比例划分成校正集和预测集, 采用多元散射校正(MSC)、基线校准(Baseline)、去趋势(De-trending)、卷积平滑(S-G)、标准正态变量变换(SNV)、归一化(Normalize)等校正方法去除原始光谱中不良散射等干扰信息。通过竞争性自适应加权抽样(CARS)、引导软收缩(BOSS)、区间变量迭代空间收缩法(iVISA)和变量组成集群分析-迭代保留信息变量(VCPA-IRIV)提取 TPC 浓度的代表性特征光谱。采用灰度共生矩阵(GLCM)算法依次提取肉样第 1 主成分图像中纹理特征。基于特征光谱及图谱融合信息建立滩羊肉中 TPC 含量的偏最小二乘回归(PLSR)与最小二乘支持向量机(LSSVM)预测模型并进行对比分析。结果表明, (1)利用 De-trending+SNV 预处理后的光谱数据建立的 PLSR 预测模型性能较好, 其 $R_c^2=0.8749$, $R_p^2=0.7932$; (2)采用 CARS, BOSS, iVISA 和 VCPA-IRIV 分别提取出了 23, 35, 57 和 43 个特征波长, 占全光谱的 18.4%, 28%, 45.6%和 16.8%; (3)采用 BOSS 法提取的关键性波长建立的 LSSVM 模型性能较好, 其 $R_c^2=0.8513$, $R_p^2=0.7459$, RMSEC=0.1168 和 RMSEP=0.1550; (4)与基于特征波长建立的预测模型相比, BOSS-ASM-ENT-CON-LSSVM 模型为滩羊肉中 TPC 浓度的最佳图谱融合预测模型($R_c^2=0.8500$, $R_p^2=0.7709$, RMSEC=0.1160, RMSEP=0.1447); (5)利用 BOSS-PLSR 简化模型将 TPC 浓度反演到样本的高光谱图像上, 通过色彩直观化形式展现出来, 实现 TPC 含量的可视化表达。

关键词 滩羊肉; 高光谱; 总酚浓度; 图谱融合; 可视化

中图分类号: TS251 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)11-3631-06

引言

滩羊是宁夏优势特色畜种, 其肉质鲜香细嫩、不膻不腥、肌肉纹理清晰, 备受消费者喜爱^[1]。氧化是导致宰后肉品劣变的重要原因之一, 总酚浓度(total phenolic content, TPC)是肌肉内源抗氧化指标之一, 能反映宰后肉质氧化还原程度, 从而判别肉品的新鲜度。传统检测方法因费时费力, 破坏性大不能满足现代快速无损检测的需求。近年来, 高光谱成像技术因其可同时实现光谱信息分析和图像处理, 广泛应用于肉品新鲜度^[2-4]、嫩度^[5]、脂肪酸含量^[6]、食源性病原体^[7]、脂质氧化^[8]等方面, 然而, 基于高光谱成像技术结合纹理特征法检测肉品中 TPC 的研究鲜有报道。

以滩羊肉为研究对象, 采用高光谱成像结合图像纹理信息对羊肉中 TPC 进行无损检测, 通过将 TPC 反演到样本的高光谱图像以色彩直观化形式展现出来, 实现 TPC 含量的可视化表达。

1 实验部分

1.1 样本

滩羊肉样本采集于宁夏盐池鑫海牧场。宰后将背最长肌、前腿肉及后腿肉放入保温箱运至实验室, 用无菌刀剔除肉样表面多余脂肪和筋膜, 分割为 3.0 cm×2.0 cm×1.0 cm 的羊肉样本 180 个(背最长肌:前腿肉:后腿=1:1:1), 真空包装后标记, 4 °C 冰箱排酸 24 h。

收稿日期: 2021-09-13, 修订日期: 2022-03-02

基金项目: 宁夏回族自治区科技创新领军人才项目(2020GKLRXLX05), 国家自然科学基金项目(31760435)资助

作者简介: 孙有瑞, 1996 年生, 宁夏大学食品与葡萄酒学院硕士研究生 e-mail: youruisun@163.com

郭美, 1995 年生, 宁夏大学食品与葡萄酒学院硕士研究生 e-mail: 13120270799@163.com

孙有瑞, 郭美: 并列第一作者 * 通讯作者 e-mail: liugs@nxu.edu.cn; fny0606@163.com

1.2 图像采集

可见/近红外高光谱成像系统(400~1 000 nm), 由高光谱成像光谱仪、CCD 相机、光纤卤素灯(150 W)、电控位移平台、数据采集软件五部分组成。启动高光谱系统预热 30 min, 进行黑白校正和参数设定^[6]。

1.3 总酚浓度测定

参考 Luciano^[9] 的方法并略作修改^[10]。

1.4 数据分析

1.4.1 预处理方法

采用多元散射校正(multiplicative scatter correction, MSC)、基线校准(baseline)、去趋势(De-trending)、卷积平滑(Savitzky-Golay, S-G)、标准正态变量变换(standard normal variate transformation, SNV)、归一化(normalize)和 De-trending-SNV 预处理算法减少原始光谱中噪音和背景信息的干扰, 增强光谱数据的有用信号, 提高模型预测性能。

1.4.2 特征波长选择方法

引导软收缩(bootstrapping soft shrinkage, BOSS)算法引入自助抽样和加权自助抽样来生成随机变量组合并构建子模型^[11]。竞争性自适应加权抽样(competitive adaptive reweighted sampling, CARS)算法遵循每组波长变量通过各自的权重相互竞争, 大的权重集保留, 较小的权重集去除的原理。变量组成集群分析-迭代保留信息变量(variable combination population analysis-iteratively retains informative variables, VCPA-IRIV)算法采用变量选择混合策略, 通过改进的 VCPA 缩小变量空间, 再使用 IRIV 进一步优化。区间变量迭代空间收缩法(interval variable iterative space shrinkage approach, iVISSA)算法是从整体和个体对波段进行搜索。其中整体和个体分析采用波段的权重优化, 以准确区分各个波段的位置、宽度和组合^[5]。

1.4.3 图像纹理信息的提取

灰度共生矩阵(gray-level co-occurrence matrix, GLCM)用于计算相邻像素之间不同灰度值的组合数目, 以此反映在指定方向上的梯度信息^[12], 在灰度梯度信息基础上从不同角度($\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$)提取能量(ASM)、对比度(CON)、相关性(COR)和熵(ENT)作为图像纹理特征参数的典型性。提取的四个纹理参数根据式(1)~式(8)进行计算

$$ASM = \sum_{i=0}^X \sum_{j=0}^Y p(i, j)^2 \quad (1)$$

$$ENT = - \sum_{i=0}^X \sum_{j=0}^Y p(i, j) \log p(i, j) \quad (2)$$

$$CON = \sum_{i=0}^X \sum_{j=0}^Y |i - j|^2 p(i, j) \quad (3)$$

$$COR = \sum_{i=0}^X \sum_{j=0}^Y \frac{(i - \mu_i)(j - \mu_j) p(i, j)}{\sigma_i \sigma_j} \quad (4)$$

$$\mu_i = \sum_{i=0}^X i \sum_{j=0}^Y p(i, j) \quad (5)$$

$$\mu_j = \sum_{j=0}^Y j \sum_{i=0}^X p(i, j) \quad (6)$$

$$\sigma_i = \sqrt{\sum_{i=0}^X (i - \mu_i)^2 \sum_{j=0}^Y p(i, j)} \quad (7)$$

$$\sigma_j = \sqrt{\sum_{j=0}^Y (j - \mu_j)^2 \sum_{i=0}^X p(i, j)} \quad (8)$$

式中, X 为 GLCM 的列数, Y 为 GLCM 的行号, $p(i, j)$ 为灰度共生矩阵。

1.4.4 模型建立

基于单一光谱信息和图谱融合信息建立羊肉中总酚浓度的偏最小二乘回归(partial least squares regression, PLSR)和最小二乘支持向量机(least-squares support vector machines, LSSVM)预测模型, 经对比分析确定最佳定量分析模型。模型性能主要由以下统计参数进行评估: 校正集及预测集相关系数(R_c^2, R_p^2); 校正集及预测集均方根误差(RMSEC, RMSEP)。

2 结果与讨论

2.1 光谱曲线分析

光谱反射率是肉样化学成分、颜色、质地、温度等性质的综合反映, Vis-NIR 区域的光谱吸收峰与 C—H, O—H 和 N—H 分子键组成的有机化合物的基本振动组合有关^[13]。如图 1 所示, 430, 580, 760 和 970 nm 为肉样的主要吸收带, 430 和 580 nm 左右的吸收带与肉样中色素蛋白有关^[14]; 760 和 970 nm 左右的吸收峰由 O—H 基团的第二泛音和第三泛音引起^[15], 对应于肉样中的水分。

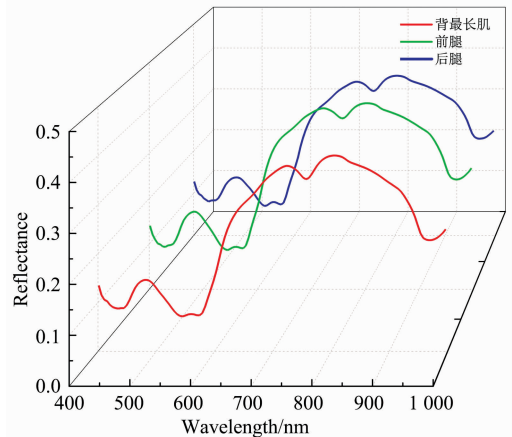


图 1 滩羊肉样本的光谱曲线

Fig. 1 Spectral curves of Tan mutton samples

2.2 光谱预处理

不同预处理方法下 TPC 含量的全波长模型性能如图 2 所示, 与原始光谱模型效果相比, Normalize 和 S-G 处理后的光谱模型性能较低, 其模型性能变差可能是因为预处理消除了与 TPC 含量相关的光谱信息。经 De-trending+SNV 处理后的光谱模型效果较原始光谱模型性能明显提高, 其 $R_p^2 = 0.793 2$, $R_c^2 = 0.874 9$, RMSEC = 0.106 9, RMSEP = 0.137 3, 表明 De-trending+SNV 预处理方法降低了原始光谱信息中仪器漂移散射和信号噪声等干扰信息。因此, 后续数据分析采用 De-trending+SNV 预处理后的光谱数据。

2.3 特征波长提取

采用 CARS, BOSS, iVISSA 和 VCPA-IRIV 对全波段光

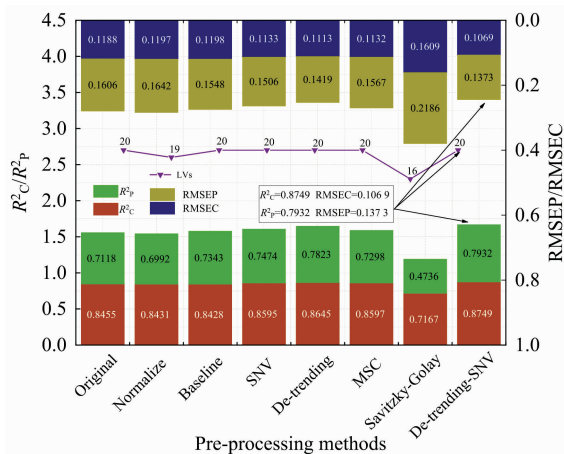


图 2 不同预处理方法下 TPC 含量的全波段模型的性能
Fig. 2 Performances of full wavelength models based on different pre-processing methods

谱进行降维处理, 结果分析如图 3 所示。BOSS 和 CARS 算法提取特征波长时, 对于不太重要的特征, 分配较小的权重, 而不是直接消除, 如图 3(a)和(b)所示。CARS 算法最终选取了 23 个有效波长, 占全光谱的 18.4%。BOSS 算法最终选取了 35 个有效波长, 占全光谱的 28%。VCPA-IRIV 算法采样过程中 RMSECV 分布如图 3(c)所示, 当采样次数为 43 时, RMSECV 下降到 0.162 4, 共选取出 21 个有效波长, 占全光谱的 16.8%。如图 3(d)所示, iVISSA 方法选择的是一些位置、组合和宽度都经过优化的波长区间, 最终获得了 57 个光谱变量, 占全光谱的 45.6%。

2.4 纹理信息提取

对肉样光谱图像进行主成分分析, 前 3 个主成分图(累计贡献率 > 99%)如图 4 所示。对表达最多的 PC1 从 $\theta=0^\circ, 45^\circ, 90^\circ$ 和 135° 角度下进行对比度、能量、熵和相关性纹理信息提取。对肉样 ROI 图像的纹理变量进行计算, 得到了一个包含 180×4 (样本 \times 变量) 的纹理矩阵。

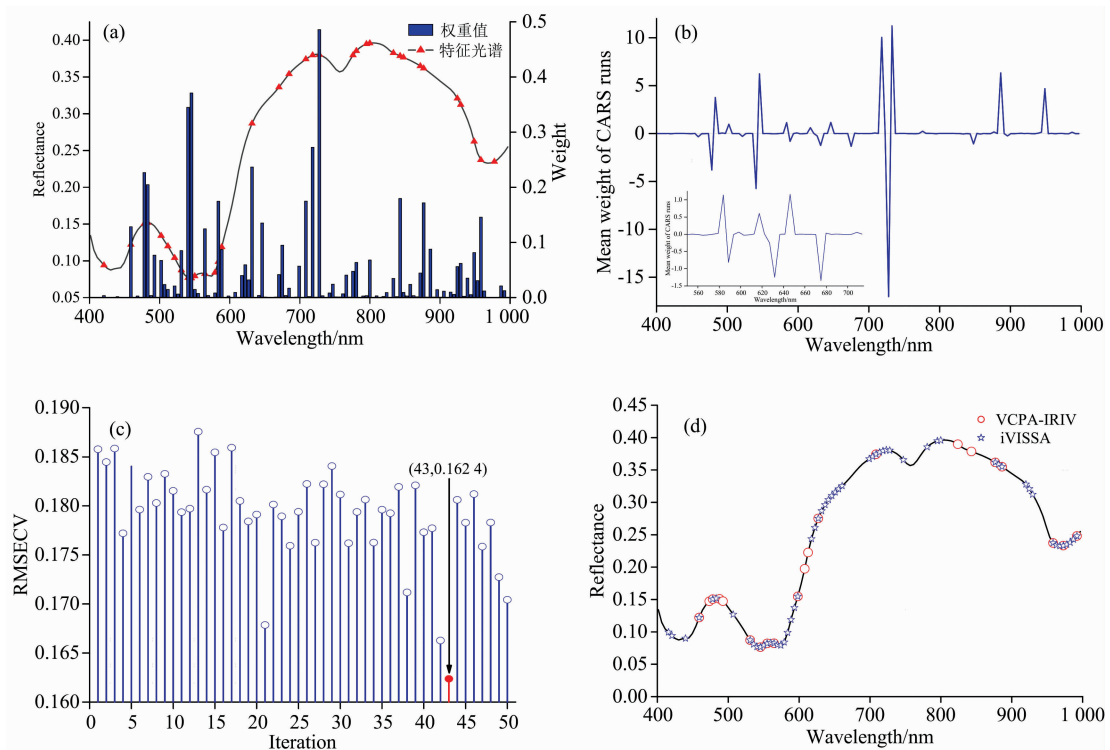


图 3 特征波长的选择

(a): BOSS 算法权重值变化和提取特征波长分布图; (b): CARS 算法平均权重值变化曲线; (c): VCPA-IRIV 算法的 RMSECV 变化图; (d): VCPA-IRIV 和 iVISSA 提取特征波分布图

Fig. 3 Selection of the characteristic wavelengths

(a): Change curve of mean weight value and the distribution map based on BOSS algorithm;
(b): Change curve of mean weight value by CARS algorithm;
(c): Changes map of RMSECV by VCPA-IRIV algorithm;
(d): Distribution maps based on the characteristic wavelengths extracted by VCPA-IRIV and iVISSA

2.5 模型建立

2.5.1 特征光谱模型建立

分别采用 PLSR 及 LSSVM 方法对特征波长的降维能力

进行评估, 建模效果如表 1 所示。由表 1 可知, CARS 和 VCPA-IRIV 方法提取的特征波长建立的预测模型性能较差。基于 BOSS 算法优选的特征波长建立的 LSSVM 模型性能较

好, 具有较高的相关系数和较低的均方根误差, 其 $R_c^2 = 0.8513$, $R_p^2 = 0.7459$, $RMSEC = 0.1168$ 和 $RMSEP = 0.1550$ 。

2.5.2 图谱融合模型建立

根据 2.5.1 的建模结果, 将 BOSS 方法提取的特征波长与肉样纹理特征进行融合建立 TPC 含量的预测模型如表 2 所示, PLSR 和 LSSVM 模型采用四种数据融合建立, 包括一个纹理变量(COR)、两个纹理变量(ASM 和 CON)、三个

纹理变量(ASM, ENT 和 CON)和四个纹理变量(COR, ASM, ENT 和 CON)与 BOSS 算法融合。BOSS-ASM-ENT-CON-LSSVM 图谱融合模型稳健性较好, 其 $R_c^2 = 0.8500$, $R_p^2 = 0.7709$, $RMSEC = 0.1160$, $RMSEP = 0.1447$, 是滩羊肉中 TPC 浓度的最佳定量分析模型, 图谱融合模型较单一光谱模型产生了令人满意的预测效果, 说明提取的纹理特征里携带了部分与肉样中 TPC 含量相关的有效信息。

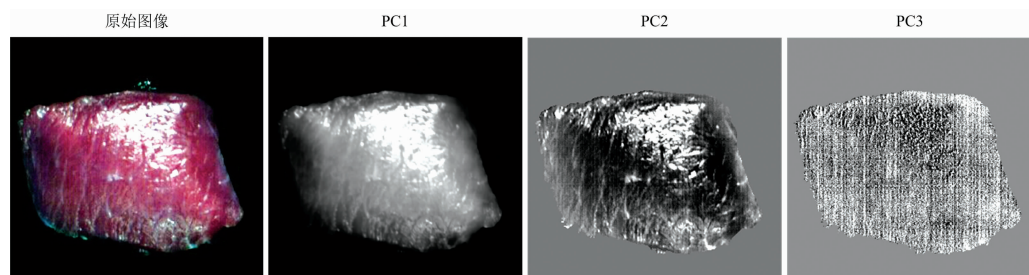


图 4 羊肉样本的前 3 个主成分图像

Fig. 4 The images of first three PCs of mutton samples

表 1 特征波长建模效果

Table 1 Model performances based on different feature-wavelength methods

Modeling methods	Selection method	LVs	Calibration set		Prediction set	
			R_c^2	RMSEC	R_p^2	RMSEP
PLSR	CARS	16	0.7634	0.1475	0.6578	0.1755
	BOSS	20	0.8252	0.1263	0.7386	0.1541
	iVISSA	20	0.8526	0.1160	0.7357	0.1564
	VCPA-IRIV	19	0.7370	0.1549	0.6096	0.1868
LSSVM	CARS	/	0.7906	0.1384	0.7002	0.1649
	BOSS	/	0.8513	0.1168	0.7459	0.1550
	iVISSA	/	0.8968	0.0975	0.7142	0.1626
	VCPA-IRIV	/	0.7400	0.1544	0.5861	0.1924

表 2 图谱融合建模效果

Table 2 Model performance based on image and spectroscopy fusion

Modeling methods	Selection method	LVs	Calibration set		Prediction set	
			R_c^2	RMSEC	R_p^2	RMSEP
PLSR	BOSS-COR	20	0.8256	0.1262	0.7184	0.1598
	BOSS-ASM-CON	20	0.8127	0.1310	0.7558	0.1527
	BOSS-ASM-ENT-CON	20	0.8085	0.1314	0.7614	0.1510
	BOSS-COR-ASM-ENT-CON	20	0.8074	0.1328	0.7424	0.1578
LSSVM	BOSS-COR	/	0.8581	0.1140	0.7512	0.1505
	BOSS-ASM-CON	/	0.8544	0.1154	0.7477	0.1522
	BOSS-ASM-ENT-CON	/	0.8500	0.1160	0.7709	0.1447
	BOSS-COR-ASM-ENT-CON	/	0.8572	0.1143	0.7578	0.1485

2.6 TPC 含量可视化分析

PLSR 算法结构简单, 运算效率高, 适用范围较广。因此, 选用特征变量下的最佳 PLSR 预测模型对肉样中 TPC 含量进行预测。由 2.5.1 中建模效果可知, BOSS-PLSR 预测模型为肉样中 TPC 含量的最佳简化模型。基于 BOSS-PLSR 模型将 TPC 含量定量反演到样本的掩膜图像上, 从而实现

TPC 含量在样本图像上的动态可视化表达(图 5)。图中黄色区域代表较高的 TPC 含量, 蓝色区域代表 TPC 含量较低。从可视化图中颜色的分布可以看出肉样中 TPC 含量的分布较均匀, 肉样的边缘部分呈现较低浓度值可能是由于肉样厚度不均匀导致。可视化分布图能够提供一种直观的方式去表征 TPC 值的动态变化过程。

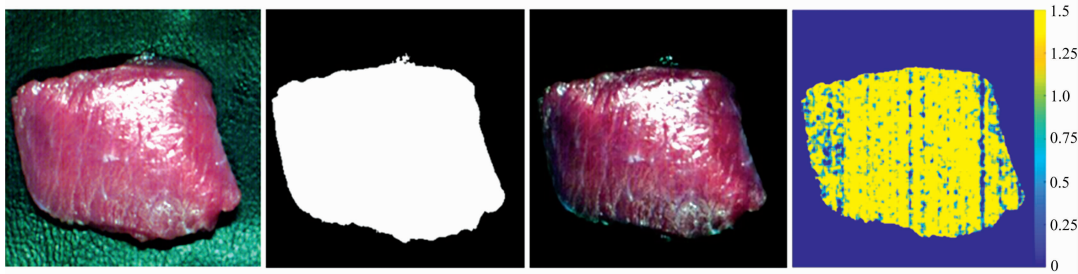


图 5 TPC 含量可视化分布图

Fig. 5 Visualizaion maps of TPC content distributions

3 结 论

利用可见-近红外高光谱技术结合图像纹理信息对滩羊肉总酚浓度含量进行预测, 主要结论如下:

(1)利用 De-trending-SNV 预处理后的光谱建立 PLSR 模型, 其 $R_p^2=0.7932$, $R_c^2=0.8749$, 较原始光谱分别提高了 0.0814 和 0.0294。

(2)CARS, BOSS, iVISSA 和 VCPA-IRIV 分别提取出了 23, 35, 57 和 43 个特征波长, 占全光谱的 18.4%, 28%, 45.6% 和 16.8%。

(3)特征波长模型中, BOSS-LSSVM 模型预测性能最佳, 其 $R_c^2=0.8513$, $R_p^2=0.7459$, RMSEC=0.1168 和 RMSEP=0.1550。

(4)与基于特征波长建立的模型相比, 图谱融合模型中, PLSR 和 LSSVM 模型的 R_p^2 值分别提高了 0.0172 和 0.0250。BOSS-ASM-ENT-CON-LSSVM 模型为最佳预测模型($R_c^2=0.8500$, $R_p^2=0.7709$, RMSEC=0.1160, RMSEP=0.1447)。

(5)利用简化模型将 TPC 浓度反演到样本的高光谱图像上, 通过色彩直观化形式展现出来, 实现 TPC 含量的可视化表达。

References

- [1] YUAN Yu-dong, YOU Li-qin, LUO Rui-ming, et al(苑昱东, 尤丽琴, 罗瑞明, 等). Food Science(食品科学), 2019, 40(18): 195.
- [2] Zhao Zhifang, Yu Hanyue, Zhang Siyu, et al. Optik, 2020, 212: 164737.
- [3] REN Ying-chun, LIU Gui-shan, ZHANG Jing-jing, et al(任迎春, 刘贵珊, 张晶晶, 等). Chinese Journal of Luminescence(发光学报), 2019, 40(3): 396.
- [4] Baek Insuck, Lee Hoonsoo, Cho Byoung-kwan, et al. Food Control, 2021, 124: 107854.
- [5] YU Yang, ZHANG Yu, TIAN Hai-qing, et al(于洋, 张珏, 田海清, 等). Journal of Agricultural Science and Technology(中国农业科技导报), 2021, 23(12): 101.
- [6] YANG Xiao-yu, DING Jia-xing, FANG Meng-meng, et al(杨晓玉, 丁佳兴, 房盟盟, 等). Food and Machinery(食品与机械), 2017, 33(11): 131.
- [7] Bonah E, Huang X Y, Aheto J H, et al. Infrared Physics & Technology, 2020, 107: 103327.
- [8] Cheng Weiwei, Klavs Martin Sørensend, Søren Balling Engelsen, et al. Journal of Food Engineering, 2019, 263: 311.
- [9] Peter C Wootton-Beard, Aisling Moran, Lisa Ryan. Food Research International, 2011, 44(1): 217.
- [10] Chen C, Han L, Yu Q L, et al. Canadian Journal of Animal Science, 2015, 95(2): 1.
- [11] Deng B C, Yun Y H, Cao D S, et al. Analytica Chimica Acta, 2016, 908: 63.
- [12] LU Bing, SUN Jun, YANG Ning, et al(芦兵, 孙俊, 杨宁, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2019, 39(8): 2515.
- [13] Pan L Q, Zhu Q B, Lu R F, et al. Food Chemistry, 2015, 167: 264.
- [14] Zheng X C, Li Y Y, Wei W S, et al. Meat Science, 2019, 149: 55.
- [15] Douglas B, Gamal E, Sun D W, et al. Meat Science, 2012, 90: 259.

Fusion of Visible Near-Infrared (VNIR) Hyperspectral Imaging and Texture Feature for Prediction of Total Phenolics Content in Tan Mutton

SUN You-rui¹, GUO Mei¹, LIU Gui-shan^{1*}, FAN Nai-yun^{1*}, ZHANG Hao-nan², LI Yue¹, PU Fang-ning², YANG Shi-hu¹, WANG Hao²

1. College of Food and Wine, Ningxia University, Yinchuan 750021, China

2. School of Physics and Electronic-Electrical Engineering, Ningxia University, Yinchuan 750021, China

Abstract The visible near-infrared (Vis-NIR) hyperspectral imaging technology was used to rapidly detect Tan mutton's total phenol concentration (TPC) content. The prediction mode and visualization of TPC content in Tan mutton were built and realized based on spectral information in combination with texture features. Firstly, the calibration set and prediction set were divided by 3 : 1, and then multiplicative scatter correction (MSC), Baseline, De-trending, savitzky-golay (S-G), and Standard normal variate transformation (SNV), and Normalize were used for model optimization. Secondly, feature bands were obtained by competitive adaptive reweighted sampling (CARS), bootstrapping soft shrinkage (BOSS), interval variable iterative space shrinkage approach (iVISSA) and variable combination population analysis coupled with iteratively retained informative variables (VCPA-IRIV), respectively. Then textural feature variables for the first principal component image were extracted by gray-level co-occurrence matrix (GLCM), respectively. Finally, partial least squares regression (PLSR) and least-squares support vector machines (LSSVM) models were built and optimized to predict TPC content. The results showed that: (1) The PLSR model yielded promising results after De-trending-SNV preprocessing, and R_p^2 and R_c^2 were 0.793 2 and 0.874 9; (2) The 23, 35, 57 and 43 characteristic bands based on the original spectral were extracted by CARS, BOSS, iVISSA and VCPA-IRIV methods, respectively, accounting for 18.4%, 28%, 45.6% and 16.8% of the total bands; (3) The simplified BOSS-LSSVM model yielded good results in assessing TPC content (R_c^2 vs. $R_p^2=0.851 3$ vs. 0.745 9, RMSEC vs. RMSEP=0.116 8 vs. 0.155 0); (4) Compared with predictive models based on characteristic wavelengths, the simply model BOSS-ASM-ENT-CON-LSSVM despited good results ($R_c^2=0.850 0$, $R_p^2=0.770 9$, RMSEC=0.116 0, RMSEP=0.144 7); (5) The simplified BOSS-PLSR model was displayed on the sample image in the form of pseudo-color to realize the visualization expression of TPC content.

Keywords Tan mutton; Hyperspectral imaging; Total phenol concentration; Fusion of spectra and texture feature; Visualization

(Received Sep. 13, 2021; accepted Mar. 2, 2022)

SUN You-rui and GUO Mei: joint first authors

* Corresponding authors