

基于ERC-RF高光谱波长选择的羊肉新鲜度相关指标的预测

李丹阳, 姜新华*, 李靖, 乌丹牧其尔, 徐子洋, 鄂计祥

(内蒙古农业大学计算机与信息工程学院, 内蒙古自治区农牧业大数据研究与应用重点实验室,
内蒙古呼和浩特 010018)

摘要: 为了快速检测羊肉中与新鲜度相关的四个理化指标含量, 该研究提出一种多任务光谱特征提取算法 ERC-RF, 通过对集成回归链 (Ensemble Regression Chain, ERC) 和随机森林 (Random Forest, RF) 进行集成, 利用新鲜度指标之间的相关性, 对特征光谱赋予不同的权重从而获取光谱波段特征的重要性。实验过程中, 首先采用 SG (Savitzky-Golay) 平滑滤波和多元散射校正 (Multiplicative Scatter Correction, MSC) 结合的方法对数据进行预处理, 然后使用 RF 和 ERC-RF 模型分别选取特征波段并将特征光谱数据用于两种新鲜度指标含量预测模型: 支持向量回归器 (Support Vector Regression, SVR) 与多输出回归器 (Multi-Output Regressor, MOR) 的集成模型: SVR-MOR 和 SVR 与多输出回归链 (Multi-Output Regression Chain, MORC) 的集成模型: SVR-MORC。结果表明: ERC-RF 提取的光谱特征数据在两种预测模型上得到的确定系数 (Coefficient of Determination, R^2)、平均相对均方根误差 (average Relative Root Mean Squared Error, aRRMSE)、平均绝对误差 (Mean Absolute Error, MAE) 分别达到了 0.974 5、1.302 7、0.385 2 和 0.975 2、1.495 6、0.391 8, 相比于其他波段选择模型得到了明显的提升, 波段数也从 125 减到 31, 证明 ERC-RF 是一种有效的多任务特征波长筛选方法, 不但提高了多输出回归模型的精度, 而且降低了模型复杂度。

关键词: 高光谱; 羊肉; 新鲜度; 波长选取; 多任务

文章编号: 1673-9078(2025)07-1-11

DOI: 10.13982/j.mfst.1673-9078.2025.7.0677

Prediction of Mutton Freshness-related Indicators Based on ERC-RF Hyperspectral Wavelength Selection

LI Danyang, JIANG Xinhua*, LI Jing, WU Danmuqier, XU Ziyang, E Jixiang

(College of Computer and Information Engineering, Key Laboratory of Agricultural and Animal Husbandry Big Data Research and Application, Inner Mongolia Autonomous Region, Inner Mongolia Agricultural University, Hohhot 010018, China)

Abstract: In order to achieve rapid detection of the contents of four physicochemical indicators related to freshness of mutton, a multi-task spectral feature extraction algorithm named ERC-RF is proposed in this study. By integrating the

引文格式:

李丹阳, 姜新华, 李靖, 等. 基于ERC-RF高光谱波长选择的羊肉新鲜度相关指标的预测[J]. 现代食品科技, 2025, 41(7): 1-11.

LI Danyang, JIANG Xinhua, LI Jing, et al. Prediction of mutton freshness-related indicators based on ERC-RF hyperspectral wavelength selection [J]. Modern Food Science and Technology, 2025, 41(7): 1-11.

收稿日期: 2024-05-17

基金项目: 国家自然科学基金项目 (62061037; 31960494); 内蒙古自然科学基金项目 (2023LHMS06017); 内蒙古自治区科技攻关计划项目 (2020GG0169)

作者简介: 李丹阳 (1999-), 女, 硕士, 研究方向: 数字图像处理, E-mail: ldy19990125@163.com

通讯作者: 姜新华 (1977-), 男, 博士, 教授, 研究方向: 农业信息化与自动化, E-mail: jiangxh@imau.edu.cn

Ensemble Regression Chain (ERC) and Random Forest (RF), the correlation between freshness indicators was used to assign varying weights to the characteristic spectra to ascertain the importance of spectral band features. During the experiments, data were first preprocessed by combining the SG (Savitzky-Golay) smoothing filter and Multiplicative Scatter Correction (MSC). Subsequently, the RF and ERC-RF models were used to select characteristic bands, respectively, and the spectral data were used for two freshness indicator contents prediction models: models for integrating Support Vector Regression (SVR) with Multi-Output Regressor (MOR): SVR-MOR, and integrating SVR with Multi-Output Regression Chain (MORC): SVR-MORC. The results showed that the Coefficient of Determination (R^2), average Relative Root Mean Squared Error (aRRMSE), and Mean Absolute Error (MAE) values of the spectral feature data extracted by ERC-RF were 0.974 5, 1.302 7, 0.385 2, and 0.975 2, 1.495 6, 0.391 8, respectively in the two prediction models, which were significantly improved compared with models based on other band selection, with the number of bands being reduced from 125 to 31. These results confirm that ERC-RF is an effective multi-task feature wavelength screening method, which not only improves the accuracy of multi-output regression model, but also reduces the model complexity.

Key words: hyperspectral technology; mutton; freshness; wavelength selection; multi-task

随着消费水平的提高,羊肉的安全和质量成为了人们关注的重点,经济全球化使得食品供应链变长,其中的任何一个阶段都可能导致肉类食品发生质变^[1]。传统的肉类检测主要依靠感官和化学手段,存在主观性强、耗时耗力的特点。高光谱成像技术通过对光谱进行细致分析和识别,可以获取物体在不同波长下的光学特性信息^[2]。每种物质都有独特的光谱特征,通过分析高光谱数据,可以了解被观察物体的化学成分、物理特性和表面特征^[3]。

在当前利用高光谱对肉类理化指标进行定量分析的研究中,大多都集中在预测模型的建立和改进上,如卷积神经网络、回归链、BP神经网络和传统回归模型的改进。其中大部分是基于单个指标利用传统的特征选取算法对高光谱波段进行提取,例如连续投影算法(Successive Projections Algorithm, SPA)、RF(Random Forest, RF)、竞争性自适应重加权采样算法(Competitive Adaptive Reweighted Sampling, CARS)、变量组合总体分析算法(Variable Combination Population Analysis, VCPA)等^[4-9],这些研究中,针对多个指标进行多任务波段选取的研究较少。羊肉的新鲜度是一个复杂的多维属性,采用单一指标可能无法全面准确地评价,因此本实验旨在同时预测与羊肉新鲜度相关的四个指标,进行多任务回归,提高评价准确性。然而,在多目标任务中,各个目标变量通常存在相关性,这种相关性会影响特征选择的稳定性和可靠性。简单地删除某些预测能力较弱的特征可能会损害模型在其他目标变量上的性能。因此,在多目标任务中,特征选择需要兼顾各个目标变量的特点,寻找能够同时提高

整体预测精度的核心特征子集。

针对多任务特征选择问题,Qin等^[10]提出了一种自适应图学习和标签信息增强(Adaptive Graph Learning and Label Information Enhancement, AGLE)的多标签特征选择方法,它通过自适应图谱学习和标签信息增强来选择最具判别性的特征子集。Mohsen等提出了一种基于蚁群优化(Ant Colony Optimization, ACO)的新型多标签相关-冗余特征选择方法,它通过引入两个无监督和有监督的启发函数,在特征空间中进行搜索^[11]。虽然这两种方法都能够平衡特征的相关性和冗余性,但是它们无法充分利用标签之间的相关性信息。针对上述问题,本文提出了一种基于集成回归链(Ensemble Regression Chain, ERC)和RF集成的多任务特征选择算法ERC-RF,ERC通过构建目标变量之间的关系链,将多个相关任务联系起来进行建模;RF是一种强大的基于树模型的特征选择方法,这两种算法的集成,可以充分利用目标变量间的相关性,提高特征选择的稳定性。ERC-RF通过构建一个基于相关性的得分函数来对特征重要性进行加权,这可以更好地捕捉目标间的潜在联系,从而提高特征选择的性能。

1 材料与方法

1.1 实验材料

1.1.1 材料与试剂

材料:羊肉背部肌肉样本购于内蒙古呼和浩特市惠民区食品有限公司,将羊肉样品均匀地分

为 6 cm×6 cm×1 cm 的肉块,并用保鲜膜非接触式包装测试样品,存放 14 d。

试剂:浓盐酸,深圳市安泽欣科技有限公司;硼酸,北京百灵威科技有限公司;氧化镁悬液(10 g/L),国药集团化学试剂有限公司;灭菌生理盐水,山东齐都药业有限公司。

1.1.2 仪器与设备

HS-2x 磁力搅拌棒,美国 Scilogex;KDN-520 凯氏定氮仪,青岛路博伟业环保科技有限公司;MIK-PH6.5PH 计,杭州美控自动化技术有限公司;SPX-L50 恒温培养箱,天津市泰斯特仪器有限公司;光谱范围为 400~1 025 nm 波段数为 125 的 DL-604M-OEM 高光谱成像仪,美国 Hyperspec VNIR N-Series;150 W 光卤素灯,佛山市永裕照明有限公司;扫描平台:SMA905 光纤准直器,深圳市鑫锐光技术有限公司;装有 SpecView 数据采集软件的计算机。

1.2 数据采集

1.2.1 样品处理

从经过酸处理的羊肉背部肌肉的 5 个不同部分取样进行实验。在本研究中,样品被立即运送至医用安全箱(0~4 ℃)并送至实验室。新鲜羊肉样品被均匀分割成 6 cm×6 cm×1 cm 的肉块,并用保鲜膜对试验样品进行非接触式包装。包装好的肉样都有编号,放在 4 ℃ 的冰箱中保存,不得挤压。理化指标值和高光谱数据的测量持续了 14 d,共准备了 5 组测试样品。其中,四组用于实验室理化指标(TVB-N、pH 值)和微生物指标(TBC 和 MPN)的测量,羊肉样品数量为 4(指标)×5(背区)×14(天)=280。其中一组用于高光谱测量,羊肉样本数为 5(背区)×14(天)=70。制备的实验样品涵盖羊肉新鲜度的三个等级:新鲜、次新鲜和变质。每次采集前,将样品表面的保鲜膜撕掉,将羊肉样品放在室内 20 min,使表面水分挥发,并于每天下午 4 点测量 TVB-N、pH 值、TBC、MPN 和高光谱图像。

1.2.2 羊肉新鲜度理化指标含量测定

羊肉储存时,其蛋白质、脂类和碳水化合物会被酶和细菌分解。这就会产生含氮化合物,如 NH₃ 和 R-NH₂,从而使肉变碱^[12]。羊肉中挥发性碱基氮(的含量随羊肉腐败程度的变化而变化,是中国现行肉类安全法规规定的主要理化指标,可定期指示羊肉的新鲜程度。羊肉新鲜度的参考指标之一

是测量 pH 值,它有助于评估肉的新鲜度和屠宰前动物的健康状况。新鲜羊肉在保存初期含有多种细菌,其菌落总数符合国家规定。随着时间的推移,细菌会大量繁殖并散发出难闻的气味^[13]。根据中国国家食品安全检测条例,菌落总数表示微生物污染和羊肉变质的程度^[14]。这是一个需要检查的指标。肉类腐烂是微生物大肠菌群造成的,国家法规规定,食品生产商必须在产品出厂前检测其是否含有大肠菌群。因此,大肠菌群的估计数量是本研究用来衡量羊肉新鲜度的指标之一。本研究中,羊肉中的挥发性碱基氮(The Volatile Salt Base Nitrogen, TVB-N)、pH 值、菌落总数(Total Bacteria Count, TBC)和大肠菌群近似数(Coliforms Most Probable Number, MPN)分别采用微量凯氏定氮法^[15]、非平均样品测定法^[16]、《食品安全国家标准食品微生物检验细菌总数测定》^[17]中的单位质量细菌总数标准值和《食品安全国家标准食品微生物检验大肠菌群计数》^[18]中的单位质量大肠菌群近似数进行测定。

1.2.3 高光谱数据采集

高光谱采集平台如图 1 所示:

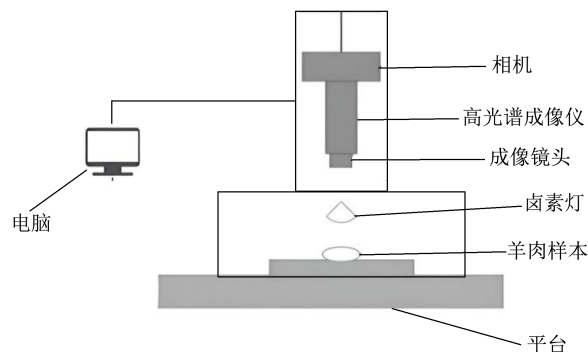


图 1 高光谱采集系统

Fig.1 Hyperspectral acquisition system

在采集高光谱图像之前,应将光谱仪提前打开预热 30 min,以此来避免噪声的影响。设置像元混合次数 6 次,通过调节高光谱仪的曝光时间使光谱亮度值曲线的峰值保持在 8 500 左右^[19]。将黑白条纹聚焦板放置在距离光谱仪透镜 30 cm 的位置,调整光谱仪透镜的焦距,直到界面上显示出黑白清晰的光谱图像。由于暗电流噪声和光源强度在每个波段的分布不均匀,原始高光谱图像需要通过黑白参考图像进行校正。采用系统自带软件 SpectraView 对高光谱图像进行黑白校正。黑白校正公式如下:

$$I = \frac{I_R - I_B}{I_W - I_B} \quad (1)$$

式中：
 I ——经过黑白校正后获得的最终实验图像；
 I_R ——实验中收集的原始光谱图像；
 I_w ——白色参考图像是通过白板校正获得的；
 I_B ——黑色图像是通过关灯获得的，相机镜头完全被不透明的盖子覆盖。

提前将羊肉样品从冰箱中取出，静置 20 min，使表面颜色从暗红色恢复为亮红色，然后测量光谱数据。最后，将待测样品放置在采集平台上，设置光谱仪的扫描方向、次数和移动速度等参数，设置完成后，开始采集高光谱图像。

1.3 实验方法

1.3.1 数据预处理

利用 ENVI 软件从测得的高光谱图像中取感兴趣区域 (Region of Interest, ROI)，在这项研究中，选取羊肉的肌肉像素用作 ROI，从每个高光谱图像取 20 个大小为 100 ± 10 像素的矩形区域^[19]，每个 ROI 的平均光谱用于后续建模和分析。

1.3.2 特征波段选择

高光谱数据庞大，可能会存在许多与羊肉的四个新鲜度理化指标含量无关的信息，因此采用适当的方法对特征波段进行选择尤为重要。本实验从多任务回归的问题转换法和算法适应法两个角度^[20]得到启发，采用两种不同的方法进行多任务特征提取，以此来测试多任务特征提取的性能。首先，多任务回归的问题转换法指的是多目标回归问题被转换为多个单目标回归问题，在本实验的多任务特征选择中，分别以 TVB-N、pH、TBC 和 MPN 为目标采用 RF 模型，计算 125 个特征波段的重要性，依次选出重要性更高的敏感波段，进而将每个指标选择的波段进行合并去重作为多任务回归模型的输入；而多任务回归的算法适应法指的是仅构造一个模型来输出所有的目标值来完成预测任务，在本实验的多任务特征选择中，采用新提出的方法 ERC-RF 计算 125 个波段相对于四个指标的综合重要性，选取综合重要性高的波段作为多任务回归模型的输入。

RF 进行特征选择的基本原理是使用随机森林模型来评估特征的重要性，随机森林由多个决策树组成，利用森林的概念，可以解决单一决策树结果片面和不准确的问题。构建随机森林的过程是随机的，每棵决策树彼此不相关，因此可以降低过拟合的可能性^[21]。本实验首先使用网格搜索来确定随机森林的理想参数值。然后，对模型进行训练，确定

特征重要性得分，得到每个波段的重要性后，将采用相应的筛选标准，筛选出重要性高的变量。特征重要性得分的计算步骤如下：①随机将全光谱数据放回并采样，形成袋内数据，未采样数据形成袋外数据 (OBB) 作为测试集；②利用袋内数据建立随机森林模型，并利用网格搜索优化参数，然后利用袋外数据计算模型的 OBB 误差 (OBB_{e1})；③随机改变 OBB 数据所有样本中变量 D 的值，重新计算其 OBB 误差 (OBB_{e2})；④假设随机森林中有 K 棵决策树，变量 D 的重要度按下式 (2) 计算：

$$VI = \frac{\sum OBB_{e2} - OBB_{e1}}{K}$$

(2)

ERC-RF 进行多任务特征选择的核心原理是使用随机森林作为基础模型计算特征重要性，又利用集成回归链 (Ensemble Regression Chains, ERC) 的原理对各个特征进行加权来进行多任务特征波段选择。在介绍 ERC-RF 算法之前，首先对集成回归链 ERC 进行简单的介绍。

在处理多目标回归问题时，传统方法通常是每个目标构建独立的预测模型，如图 2 所示。然而，这种方法未能充分考虑目标变量之间的内在相关性。受分类器链思想的启发，有研究者提出了回归器链的概念，如图 3 所示。回归链方法本质上是一种序列建模技术，其核心思想在于每个模型尝试预测目标变量的一个分量，而该分量是前一个模型的预测结果和输入特征的函数。这种方法的优点在于其能够捕获数据中的复杂依赖关系，有效地处理了多目标预测问题。但是这种方式存在一个明显的缺点，就是回归链的性能对链顺序极为敏感。

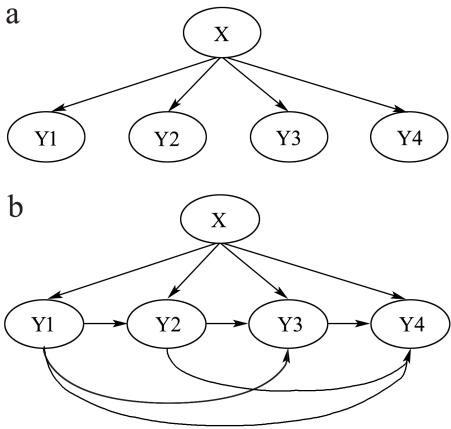


图 2 处理多目标回归问题的两种传统方法
Fig.2 Two traditional methods for dealing with multi-objective regression problems

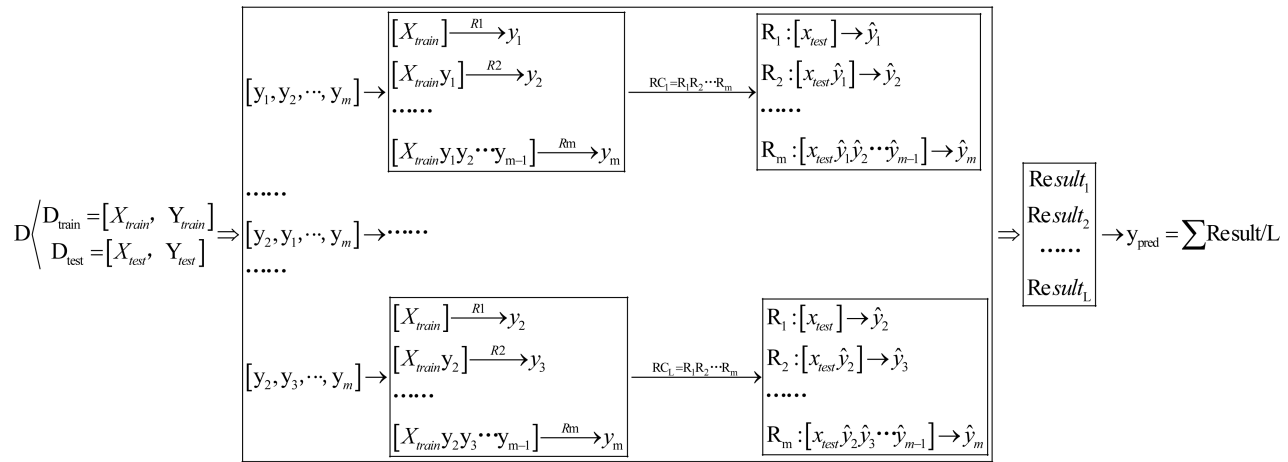


图3 集成回归链流程图

Fig.3 Flowchart of the ensemble regression chain

目前解决该问题的技术主要有使用进化算法来优化链结构或利用集成回归链的方法，即独立训练 n 个随机链，然后取 n 个估计的平均值作为最终结果^[21]。集成回归链的计算流程如图3所示，本实验的 ERC-RF 算法利用了集成回归链的原理通过考虑目标之间的相关性来对特征重要性进行加权。

在进行多任务特征选择时，如果只关注单个目标变量的特征重要性，很容易忽略一些对相关性较低的目标变量同样重要的特征。在 ERC-RF 算法中，通过计算四个新鲜度理化指标的相关性矩阵，给予相关性较低的指标更高的权重，可以确保算法选择的特征能够兼顾所有目标变量，而不会偏向于某一个目标变量。首先给定多目标数据集 $X \in R^{n \times d}$, $Y \in R^{n \times m}$ ，其中 n 表示样本数， d 表示特征数， m 表示目标数。(1) 计算目标相关性矩阵 $C \in R^{m \times m}$ ，其中 $C_{ij} = \text{corr}(y_i, y_j)$ ，其中各个目标之间的相关性通过皮尔相关系数进行计算；(2) 考虑所有可能的标签排列顺序 $\Pi = \pi | \pi = (\pi_1, \pi_2, \pi_3, \pi_4)$ ，针对每个排列顺序 $\pi \in \Pi$ ，计算相关性得分 $S_\pi = \prod_{i=1}^{m-1} \prod_{j=i+1}^m (1 - |C_{\pi_i, \pi_j}|)$ ；(3) 构建随机森林的集成回归链：初始化 T 个随机森林回归器 $f_1^{(1)}, f_1^{(1)}, \dots, f_m^{(1)}, \dots, f_1^{(T)}, f_2^{(T)}, \dots, f_m^{(T)}$ ，对于每个排列顺序，依次训练 m 个回归器，第 j 个回归器的输入是 X 拼接前 $j-1$ 个目标得到的，将所有回归器的特征重要性加权求平均得到当前排列下的特征重要性 $\text{impor tan } ce^{(\pi)} = \frac{1}{T \cdot M} \sum_{t=1}^T \sum_{j=1}^m \text{impor tan } ce^{(t,j)}$ ；(4) 将所有排列顺序下的特征重要性加权求平均，得到最终的特征重要性 $\text{impor tan } ce \in R^d$ ， impor tan

$ce = \frac{1}{|\pi|} \sum_{\pi \in \Pi} s_\pi \cdot \text{impor tan } ce^{(\pi)}$ ，根据最终计算的特征重要性，选取特征重要性高的变量作为最终的结果。

为了验证 ERC-RF 特征选择方法的有效性，本研究将采用两种不同的多任务回归建模技术，即多输出回归链 (Mult-Output Regression Chain, MORC) 和多输出回归器 (Multi-Output Regressor, MOR)，并将它们分别与支持向量回归 (Support Vector Regression, SVR) 模型集成，对经过 ERC-RF 特征选择的光谱数据进行性能评估。多输出回归链采用级联的建模方式，通过逐步预测各个输出变量并将其作为其他输出变量的输入特征，能够充分利用目标变量间的依赖关系。而多输出回归器则将所有输出变量视为一个整体，使用单一的模型同时预测所有目标变量，未考虑变量相关性。将基于 ERC-RF 特征选择的数据分别输入这两种多输出回归模型，旨在探究 ERC-RF 方法在不同多任务建模策略下的适用性和鲁棒性，避免陷入局部最优现象。

1.3.3 评价指标

多输出回归模型的泛化性能评估与传统单输出回归模型存在一定差异。在多输出回归问题中，每个样本可以同时与多个输出变量相关联，这使得模型性能评估更加复杂^[22]。本研究采用以下指标来分析实验结果：确定系数 (Coefficient of Determination, R^2)、平均相对均方根误差 (average Relative Root Mean Squared Error, aRRMSE)、平均绝对误差 (Mean Absolute Error, MAE)，这些指标不仅能够评估模型的整体预测性能，还可以反映模型对各个输出变量的拟合情况。

2 结果与讨论

2.1 光谱曲线分析

将三个不同新鲜度样本的光谱数据进行平均，得到三种新鲜度羊肉样本的光谱曲线，如图 4 所示。从图中可以看出，因为不同新鲜度理化指标值的变化，不同新鲜度的光谱图像有很大的区别，在第 14 天的时候，由于水分的减少和脂肪的氧化，导致反射率增大。从图 4 观察到了多个吸收峰，分别位于 580、750、760 和 800 nm 附近，其中 580、760、750 nm 分别是 TVB-N、pH 值、菌落总数和大肠菌群近似数变化的主要原因，光谱曲线在三个吸收峰处的差别，代表不同新鲜度的羊肉样本在四项理化指标含量上的差异。

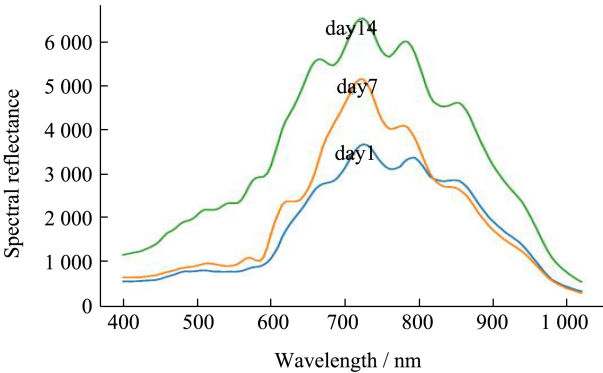


图 4 三个不同新鲜度羊肉样本的光谱曲线
Fig.4 Hyperspectral profiles of three different freshness samples of mutton

2.2 数据预处理

表 1 预处理方法结果对比

Table 1 Comparison of results of pre-processing methods

预处理方法	R^2	aRRMSE	MAE
None	0.788 0	1.922 3	2.325 7
D1	0.771 0	1.932 1	2.302 7
MSC	0.942 3	1.291 9	1.542 8
SNV	0.794 4	1.839 9	2.278 7
SG	0.942 2	1.272 3	0.985 5
MA	0.938 3	1.329 0	1.541 6
Z-Score	0.769 9	1.976 6	2.314 4
SG-MSC	0.943 1	1.297 1	1.575 4
SG-SNV	0.777 9	1.980 3	2.319 7
MA-MSC	0.930 7	1.442 0	1.543 8
MA-SNV	0.768 6	1.920 4	2.263 1

未经过预处理的原始光谱曲线存在噪声和基线漂移现象，在对原始数据进行建模和分析之前，应使用预处理方法对其进行处理。

本研究从基线校正、散射校正、平滑处理以及尺度缩放四个维度出发，探讨了这些预处理方法在偏最小二乘回归模型（Partial Least Squares Regression, PLSR）中对多输出回归性能的影响，其中基线校正采用一阶导（first-order Derivative, D1）方法，散射校正采用多元散射校正（Multiplicative Scatter Correction, MSC）和标准正态变量变换（Standard Normal Variate, SNV），平滑处理采用 SG 平滑和移动平均（Moving Average, MA），尺度缩放采用 Z-Score 方法。本实验采用多输出回归的性能度量指标： R^2 、aRRMSE、MAE 来表示预处理方法的效果，实验结果如表 1 所示。从表中可知，当采用散射校正和平滑处理两种预处理方式对原始数据进行处理时，可以获得较好的效果， R^2 相较于原始光谱提升了 15.51%，aRRMSE 和 MAE 分别降低了 0.625 2、0.750 3。因此，将这两种预处理方式进行组合^[23]，测试不同组合方式的效果，从表 1 可以看出，当采用 SG-MSC 对原始数据进行处理时，获得了最好的效果，将其作为本实验的预处理方法。

2.3 最佳波段数的确定

2.3.1 RF 特征选择

本实验分别以 TVB-N、pH、TBC 和 MPN 为目标，使用 RF 算法结合网格搜索确定了 125 个特征波段分别相对于四个指标的重要性，计算结果如图 5 所示。为了确定最佳的重要性阈值，分别以 0.009、0.01、0.012、0.013、0.014、0.014 86、0.015、0.015 2、0.016、0.018 为阈值，选择了 59、51、41、37、35、32、31、30、27、24 个波段，以此来测试不同阈值对多输出回归模型的性能影响。将选择的特征波段数据依次代入 SVR-MOR 和 SVR-MORC 两种不同的多输出回归模型中，来测试 RF 特征选择方法对于不同模型的适用性，结果如表 2 所示。从表中可以看出，当选择 0.015 作为阈值时，选择的 31 个波段在两个多输出回归模型都获得了最佳的性能， R^2 、aRRMSE、MAE 分别为 0.972 0、1.713 1、0.422 9 和 0.972 7、1.713 0、0.422 1。此时选择的波段为 438.864 1、443.665 7、448.467 2、453.268 7、458.070 3、467.673 3、501.284 1、510.887 2、539.696 4、544.497 9、582.910 2、587.711 7、606.917 9、645.330 2、

654.933 3、659.734 8、664.536 3、722.154 8、726.956 3、731.757 9、736.559 4、741.360 9、746.162 5、750.964、755.765 6、760.567 1、765.368 6、770.170 2、933.422 4、943.025 5、947.827 nm。

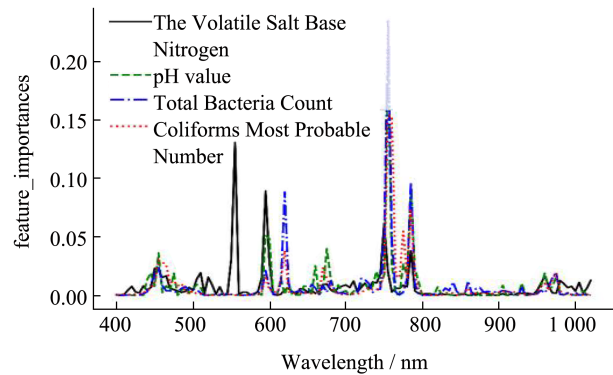


图 5 采用 RF 计算的各个波段相对于四个理化指标的重要性曲线

Fig.5 Importance curves of each band relative to the four physical and chemical indicators calculated using RF

2.3.2 ERC-RF特征选择

本实验采用 ERC-RF 方法计算了 125 个波段相对于四个理化指标的综合重要性，各个波段的重要性如图 6 所示，当光谱特征变量和因变量相关性越强，特征重要性就越大，那么优选该变量为特征变量^[24]。为了确定最佳波段数，实验选择不同的波段数 25、28、30、31、32、35、40、50、59、68、74 依次代入 SVR-MOR 和 SVR-MORC 两种不同的多

输出回归模型中，来测试的特征选择方法对于不同模型的适用性，结果如表 3 所示。从表中可以看出，当选择的特征波段数量为 31 时两个多输出回归模型都获得了最佳的性能， R^2 、aRRMSE、MAE 分别为 0.974 5、1.302 7、0.385 2 和 0.975 2、1.495 6、0.391 8。此时选择的波段为 448.467 2、453.268 7、458.070 3、462.871 8、467.673 3、472.474 9、477.276 4、482.078、486.879 5、544.497 9、549.299 4、582.910 2、587.711 7、592.513 3、616.521、621.322 5、702.948 6、707.750 2、712.551 7、717.353 3、722.154 8、726.956 3、731.757 9、736.559 4、741.360 9、746.162 5、750.964、755.765 6、760.567 1、765.368 6、770.170 2 nm。

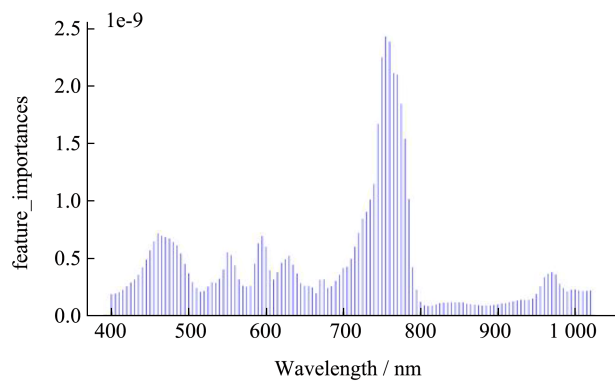


图 6 采用 ERC-RF 计算的各个波段的重要性

Fig.6 Importance of individual bands calculated using ERC-RF

表 2 RF特征选择不同阈值范围内的特征波长预测结果

Table 2 Prediction results of feature wavelengths in different threshold ranges for RF feature selection							
阈值	波段数	SVR-MOR			SVR-MORC		
		R^2	aRRMSE	MAE	R^2	aRRMSE	MAE
0.009	59	0.960 7	2.569 9	0.528 1	0.957 2	2.570 1	0.534 1
0.01	51	0.961 1	2.542 3	0.519 7	0.958 2	2.542 6	0.525 5
0.012	41	0.959 5	2.582 9	0.521 9	0.956 8	2.583 2	0.526 9
0.013	37	0.954 5	2.704 1	0.625 8	0.948 3	2.704 5	0.632 1
0.014	35	0.952 9	2.744 4	0.643 9	0.967 4	2.038 8	0.485 7
0.014 86	32	0.963 1	2.041 7	0.484 7	0.969 3	1.954 5	0.457 3
0.015	31	0.972 0	1.713 1	0.422 9	0.972 7	1.713 0	0.422 1
0.015 2	30	0.971 1	1.496 2	0.394	0.972 6	1.512 6	0.394 7
0.016	27	0.970 2	1.871 0	0.441 9	0.971 8	1.802 4	0.435 4
0.018	24	0.969 7	1.911	0.447 7	0.971 3	1.836 8	0.440 1

表 3 ERC-RF选择的特征波长预测结果
Table 3 Predicted results for different numbers of feature wavelength selected by ERC-RF

波段数	SVR-MOR			SVR-MORC		
	R^2	aRRMSE	MAE	R^2	aRRMSE	MAE
25	0.942 1	2.684 9	0.605 7	0.955 3	2.173 7	0.520 1
28	0.959 8	2.093 3	0.492 8	0.963 1	1.962 8	0.470 4
30	0.971 2	1.837 2	0.441 3	0.972 0	1.713 1	0.422 9
31	0.974 5	1.302 7	0.385 2	0.975 2	1.495 6	0.391 8
32	0.968 3	1.614 8	0.426 1	0.968 8	1.614 7	0.425 9
35	0.965 3	1.674 5	0.439 1	0.939 8	2.847	0.592
40	0.960 5	2.582 8	0.540 8	0.955 3	2.583 1	0.548
50	0.953 5	2.616 4	0.639 7	0.954 9	2.517 1	0.548 6
59	0.958 4	2.591 4	0.568 8	0.953 8	2.591 7	0.575 4
68	0.959 1	2.563 8	0.544 7	0.955 3	2.564 1	0.549 9
74	0.957 9	2.581 7	0.543 2	0.954 6	2.58	0.545 7

2.4 不同波段选择方法的比较

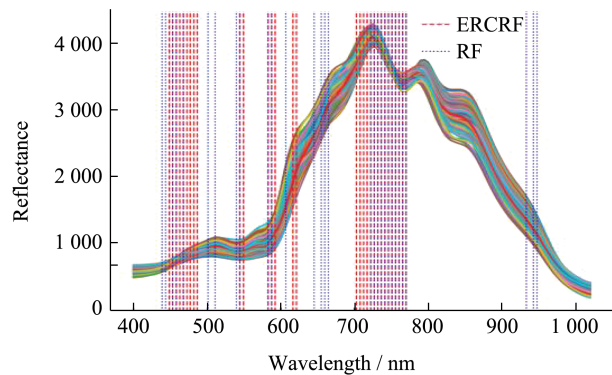


图 7 ERC-RF 和 RF 提取的最优特征波段的分布
Fig.7 Distribution of optimal eigenbands for ERC-RF and RF extraction

RF 和 ERC-RF 两种特征选择选择算法选择的特征波段的分布如图 7 所示，将 ERC-RF 与 RF 算法筛选的波长结果进行对比可知，二者相同的波长为 448.467 2、453.268 7、458.070 3、467.673 3、544.497 9、582.910 2、587.711 7、722.154 8、726.956 3、731.757 9、736.559 4、741.360 9、746.162 5、750.964、755.765 6、760.567 1、765.368 6、770.170 2 nm，主要集中在 440~590 nm、700~780 nm 之间，经前人研究，550 nm 和 580 nm 附近是引起 TVB-N 值变化的血红蛋白的吸收波长^[5]；760 nm 附近的变化主要是由 O-H 拉伸第三泛音或肌红蛋白氧化产生的吸收带引起的，它是 pH 变化的原因^[25]；微生物腐败是由蛋白质降解与水分子变化引起的，755 nm 处与水分子的 O-H 键有关^[26]，表明提取的特征波长可能与样品

的四项理化指标有较高的相关性。除了这些共同波段，RF 提取的特征波段还提取了这些分布之外的一些波段，而 ERC-RF 经过改进之后，提取的波段比较集中的分布在 440~620 nm 和 700~780 nm 之间，主要是因为 970 nm 处与水分子的 O-H 伸缩振动二级倍频有关，与 755 nm 和 760 nm 附近的波段信息类似，并且这部分的波段判别能力可能相对较弱，予以舍弃。综上所述，选用 ERC-RF 进行特征波段选择时能够减少全光谱的数据冗余，提高回归性能和准确率。

2.5 不同波段选择方法的建模

为了验证 ERC-RF 特征选择方法的有效性，分别将全光谱、SVR-GA、SVR-HHO、SVR-ACO、SPA、CARS、RF 和 ERC-RF 选择的特征波段作为输入代入 SVR-MOR 和 SVR-MORC 两种不同的多输出回归模型中进行测试，SVR-GA、SVR-HHO、SVR-ACO 分别指的是遗传算法（Genetic Algorithm, GA）、哈里斯鹰优化算法（Harris Hawks Optimization, HHO）和蚁群算法（Ant Colony Optimization, ACO）三种智能优化算法与 SVR 结合组成的波段选择算法^[23,24]，又为了避免这三种算法在 SVR 模型中的局限性，本实验采用传统方法 SPA、CARS 选取特征波段进行对比，其中 SVR-GA、SVR-HHO、SVR-ACO、SPA、CARS 都是以单个化学指标 TVB-N、pH、TBC 和 MPN 为目标进行特征提取最后进行合并去重，各个算法筛选的特征波长数和预测结果如表 4 所示。

由表 4 可知，7 种不同的波长变量算法在提取特征波长后构建的模型效果都要高于全光谱，经过 ERC-RF 进行特征波段提取之后，SVR-MOR 和 SVR-MORC 都获得了最好的预测结果。三种智能优化算法组合成的波段选择算法选择的冗余特征波段数量较多，仅减少为原来的 60% 左右，但是相比于传统特征选择算法，预测精度得到了提升；SPA、CARS 这两种传统的特征选择方法最大程度的减少了特征波段数量，波段数量仅为原来的 20% 左右，但是模型预测准确度较低；RF 和 ERC-RF 两种算法都选择了 31 个波段，波长数量仅为原始光谱的 24.8%，且预测精度进一步得到了提升，且 ERC-RF 相较于原始的 RF 特征提取算法，准确度提高了 0.25% 左右，主要原因可能是 ERC-RF 考虑了各个目标之间的相关性，按照相应的权重对各个波段的重要性进行赋值，在此背景下，可能会更适用于多输出回归模型，并且在考虑目标相关性的 SVR-MORC 模型中表现更好。

表 4 不同波段选取方法建模结果
Table 4 Modeling results of different band selection methods

特征选择方法	波段数	SVR-MOR			SVR-MORC		
		R^2	aRRMSE	MAE	R^2	aRRMSE	MAE
None	125	0.942 2	2.682 2	0.599 6	0.965 4	2.344 8	0.481 4
SVR-GA	89	0.962 4	2.501 3	0.494 6	0.961 6	2.501 4	0.496 2
SVR-HHO	76	0.961 6	2.501 9	0.496 3	0.960 2	2.213 2	0.519 6
SVR-ACO	70	0.965 1	2.463 2	0.466 7	0.968 0	1.996 1	0.475 2
SPA	18	0.957 4	2.416 7	0.548 1	0.960 7	2.201 7	0.517 7
CARS	23	0.969 7	1.911 0	0.447 7	0.971 3	1.836 8	0.440 1
RF	31	0.972 0	1.713 1	0.422 9	0.972 7	1.713 0	0.422 1
ERC-RF	31	0.974 5	1.302 7	0.385 2	0.975 2	1.495 6	0.391 8

2.6 单个理化指标的预测效果

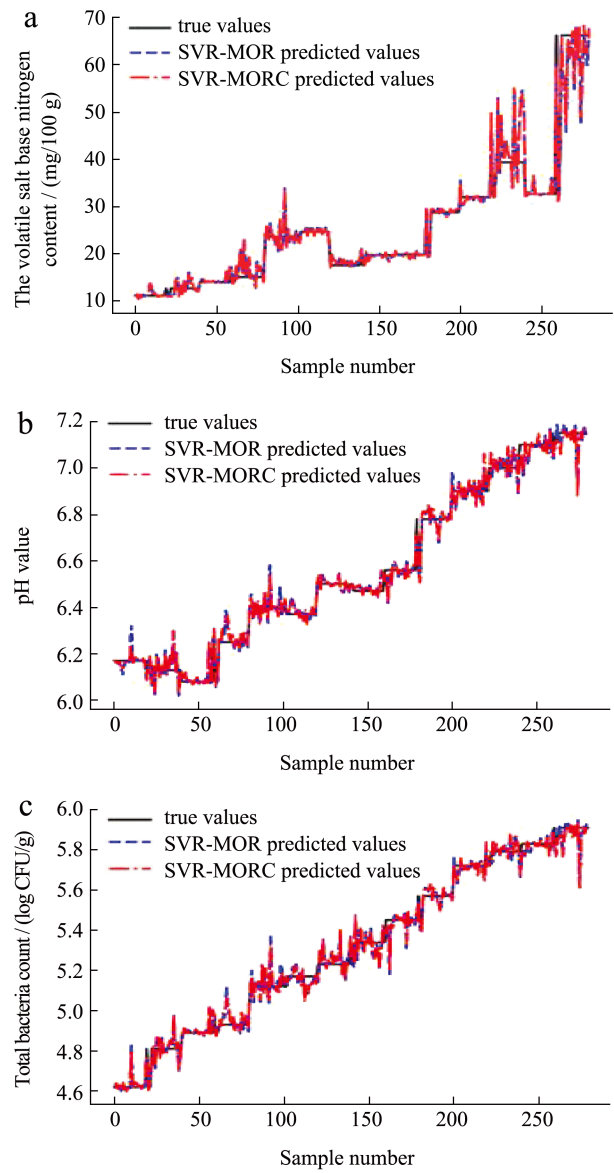
为了进一步验证 ERC-RF 的有效性，除了测试多输出回归效果，还需要测试其对单个理化指标的预测效果。针对单个理化指标，本实验采用 R^2 、RMSE、MAE 三个评价指标来表示回归效果，同样将 ERC-RF 特征选择之后的波段数据代入 SVR-MOR 和 SVR-MORC 两种多输出回归模型中，分别测试每个理化指标的回归效果。四个理化指标的测试结果如表 5 所示。分析表 5 可知，采用 ERC-RF 特征提取的波段数据在四个理化指标都取得了不错的效果，但针对 TVB-N 的预测，两个模型的预测效果都不如其他三个模型，究其原因，可能是因为 TVB-N 含量在第 14 天出现了陡增，导致 TVB-N 的预测效果不如其他三项指标^[27,28]。

表 5 单个理化指标的预测效果

理化指标	SVR-MOR			SVR-MORC		
	R^2	RMSE	MAE	R^2	RMSE	MAE
TVB-N	0.940 7	3.331 6	1.246 0	0.946 1	3.175 6	0.489 3
pH	0.983 1	0.045 6	0.030 1	0.984 0	0.044 4	0.030 3
TBC	0.980 7	0.055 4	0.036 6	0.981 4	0.054 3	0.037 1
MPN	0.988 5	0.123 0	0.076 1	0.989 5	0.117 8	0.073 9

采用 ERC-RF 方法筛选的特征波长建立多输出回归模型，四个理化指标的真实值与预测值的拟合图如图 8 所示。从图 8 可以直观地看出，经过 ERC-RF 特征提取之后模型的拟合效果表现较好。综上所述，尽管在 TVB-N 指标的预测中效果略逊于其他指标，但整体来看，高光谱结合 ERC-RF 算法筛选特征波长并建立预测模型，具有较高的预测

准确度，可以实现羊肉新鲜度化学指标含量的定量分析，在理化指标回归预测中展示了较高的应用价值和潜力。



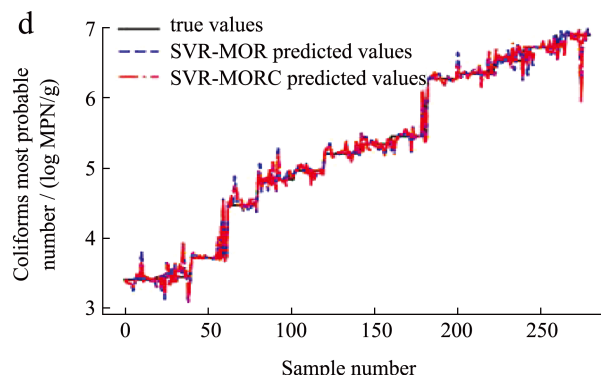


图 8 四项理化指标真实值与预测值关系图

Fig.8 Relationship between the true and predicted values of the four physical and chemical indicators

3 结论

本研究通过采集 14 天内的羊肉高光谱图像,采用 SG-MSC 预处理方法对 400~1025 nm 波段的光谱进行预处理,采用新提出的 ERC-RF 和一些经典的特征选择算法进行波段选择,同时建立 SVR-MOR 和 SVR-MORC 两种不同类别的多输出回归模型。结果显示,经过 ERC-RF 特征波段选择之后建立的 SVR-MOR 和 SVR-MORC 模型的 R^2 、aRRMSE、MAE 分别为 0.974 5、1.302 7、0.385 2 和 0.975 2、1.495 6、0.391 8,效果优于其他波段选择算法。实验结果表明,采用 ERC-RF 进行特征波长选择,可以提高多输出回归模型的准确性,减少模型的运行时间。由此证明,利用高光谱图像检测技术可以很好地实现羊肉新鲜度的精确检测,为羊肉新鲜度的无损检测提供了新的解决方案。本实验的 ERC-RF 算法虽然考虑到了各个目标变量之间的关系,但是还未考虑到各个特征之间的关系,这将作为此项研究后续改进的重点。

参考文献

- [1] QAMAR Z, MOHAMMAD A, NUR F K M, et al. Current analytical methods for porcine identification in meat and meat products [J]. Food Chemistry, 2020, 324: 126664.
- [2] 董鑫鑫,杨方威,于航,等.基于拉曼光谱技术的猪瘦肉新鲜度快速无损检测方法研究[J].光谱学与光谱分析, 2023,43(2):484-488.
- [3] 白雪冰,马殿坤,张梦杰,等.红肉质量的高光谱无损检测研究进展分析[J].光谱学与光谱分析, 2022,42(7):1993-1998.
- [4] LIU C C, CHU Z J, WENG S Z, et al. Fusion of electronic nose and hyperspectral imaging for mutton freshness detection using input-modified convolution neural network [J]. Food Chemistry, 2022, 385: 132651.
- [5] 赵亭亭,王克俭,司永胜,等.基于AW-OPS高光谱波长选择方法的羊肉新鲜度检测[J].光谱学与光谱分析,2023, 43(3):830-837.
- [6] ZUO J W, PENG Y K, LI Y Y, et al. Nondestructive detection of nutritional parameters of pork based on NIR hyperspectral imaging technique [J]. Meat Science, 2023, 202: 109204.
- [7] XIE C Q, WANG C Y, ZHAO M Y. Detection of the 5-hydroxymethylfurfural content in roasted coffee using machine learning based on near-infrared spectroscopy [J]. Food Chemistry, 2023, 422: 136199.
- [8] 陈蕊,王雪,王子文.基于随机森林特征重要性和区间偏最小二乘法的近红外光谱波长筛选方法[J].光谱学与光谱分析,2023,43(4):1043-1050.
- [9] 罗琪,庾先国,张贵宇,等.近红外光谱技术结合iPLS-SPA波段筛选在黄水酒精度预测模型中的应用[J].现代食品科技,2023,39(4):311-317.
- [10] QIN Z, CHEN H M, MI Y, et al. Multi-label feature selection with adaptive graph learning and label information enhancement [J]. Knowledge-Based Systems, 2024, 285: 111363.
- [11] MOHSEN P, MOHAMMAD B D, HOSSEIN N. MLACO: A multi-label feature selection algorithm based on ant colony optimization [J]. Knowledge-Based Systems, 2020, 192: 105285.
- [12] YANG D, HE D, LU A, et al. Detection of the freshness state of cooked beef during storage using hyperspectral imaging [J]. Applied Spectroscopy, 2017, 71(10): 2286-2301.
- [13] ZHOU Y B, AI Q J, ZHANG D Q. Analysis of bacterial phase changes on the surface of chilled mutton during storage at 4 °C [J]. Food Science, 2015, 36(6): 242-245.
- [14] GB/T5009.44-2003,肉和肉制品卫生标准的分析方法[S].
- [15] 于见亮,李开雄,卢士玲,等.应用KDY-9820凯氏定氮仪测定羊肉中挥发性盐基氮品[J].食品研究与开发,2008, 29(2):141-143.
- [16] GB/T9695.5-2008,肉与肉制品pH测定[S].
- [17] GB 4789.2-2022,食品安全国家标准 食品微生物学检验 菌落总数测定[S].
- [18] GB 4789.3-2010,食品安全国家标准 食品微生物学检验 大肠菌群计数[S].

- [19] 徐子洋.基于高光谱成像技术的多指标冷鲜羊肉新鲜度无损检测[D].呼和浩特:内蒙古农业大学,2023.
- [20] 张睿.基于多类别相关性分析的多目标回归方法研究[D].重庆:重庆邮电大学,2022.
- [21] XIANG L, YANG X, HU A, et al. Condition monitoring and anomaly detection of wind turbine based on cascaded and bidirectional deep learning networks [J]. *Journal of Network and Computer Applications*, 2022, 305: 17925.
- [22] SPYROMITROS E, TSOUMAKAS G, GROVES W. Multi-target regression via input space expansion: treating targets as inputs [J]. *Machine Learning*, 2016, 104(1): 55-98.
- [23] 黄清,薛河儒,刘江平,等.基于蚁群-遗传算法的光谱选择方法与应用[J].*光谱学与光谱分析*,2022,42(7):2262-2268.
- [24] 王仲雨,高美凤.基于改进鲸鱼优化算法的近红外光谱波长变量选择方法及其应用[J].*分析测试学报*,2023,42(1):37-44.
- [25] 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.
- [26] HE H J, SUN D W. Hyperspectral imaging technology for rapid detection of various microbial contaminants in agricultural and food products [J]. *Trends in Food Science*, 2015, 46(1): 99-109.
- [27] 姜新华,薛河儒,郅晓晶,等.高光谱图像与稀疏核典型相关分析冷鲜羊肉新鲜度无损检测[J].*光谱学与光谱分析*,2018,38(8):2498-2504.
- [28] ZHUANG Q B, PENG Y K, YANG D K, et al. Detection of frozen pork freshness by fluorescence hyperspectral image [J]. *Journal of Food Engineering*, 2022, 316: 110840.