

# 宽度学习的虾新鲜度检测方法

叶荣珂<sup>1</sup>, 孔庆辰<sup>1</sup>, 李道亮<sup>1,2</sup>, 陈英义<sup>1,2</sup>, 张玉泉<sup>1</sup>, 刘春红<sup>1,2\*</sup>

1. 中国农业大学信息与电气工程学院, 北京 100083  
2. 国家数字渔业创新中心, 北京 100083

**摘要** 为了提升虾新鲜度判别的准确性, 提出了一种基于宽度学习(BLS)的虾新鲜度检测方法。首先采用多元散射校正(MSC)、标准正态变量校正(SNV)和直接正交信号校正(DOSC)对不同冷藏天数虾的原始高光谱进行预处理, 再使用t分布随机邻域嵌入(t-SNE)将预处理之后的数据可视化, 可视化结果表明DOSC聚类效果最佳。然后使用随机森林(RF)、主成分分析(PCA)和二维相关光谱分析(2D-COS)对经DOSC预处理之后的光谱数据进行特征选择。最后基于选择的特征波长对虾新鲜度进行建模分析。将宽度学习(BLS)首次用于虾新鲜度建模, 同时与偏最小二乘判别(PLS-DA)和极限学习机(ELM)等经典判别模型做比较。研究结果表明RF方法最大限度地消除了光谱中的冗余信息, 而BLS与线性建模方法PLS-DA以及非线性建模方法ELM相比, 准确率更高并且判别时间更短, 因此RF-BLS组合模型获得了最佳新鲜度判别效果, 表明高光谱成像技术结合宽度学习识别虾的新鲜度是可行的, 可以为在线检测虾新鲜度系统的开发提供理论依据。

**关键词** 虾新鲜度; 直接正交信号校正; 随机森林; 宽度学习; 高光谱成像技术

**中图分类号**: O657.3 **文献标识码**: A **DOI**: 10.3964/j.issn.1000-0593(2022)01-0164-06

## 引言

虾含有丰富的锌、硒、铜等物质, 维生素B12含量高, 质地细腻易消化, 受到消费者的欢迎, 由于虾含有丰富的水分和营养物质, 随着冷藏时间的延长虾的品质会迅速劣变<sup>[1]</sup>。王素华<sup>[2]</sup>等发现对虾在5℃的保质期为4d。凌萍华<sup>[3]</sup>等研究发现南美白对虾置于4℃条件下贮藏货架期仅为4d。传统虾新鲜度评定方法多采用仪器分析和人工感官评定的方法, 这些方法可靠性能差, 花费时间长且不可重复, 不能精确、实时地反映全部产品的新鲜度<sup>[4]</sup>。

近年来, 大量研究利用高光谱成像技术对各类农产品进行新鲜度的检测。与仪器分析和感官评价的方法相比较, 高光谱成像技术具有预测精确、价格较低和无损等特点, 在食物新鲜度检测领域得到广泛应用<sup>[5]</sup>。Yu等<sup>[6]</sup>采用高光谱成像技术结合基于堆栈式自动编码器的逻辑回归模型(SAEs-LR), 对虾仁冷藏保鲜过程中的鲜度进行判别。Zhang等<sup>[7]</sup>采用近红外高光谱成像系统结合卷积神经网络(CNN)测定了黑枸杞中总酚, 总黄酮和总花色苷含量。Xin等<sup>[8]</sup>将高光谱技术和堆栈卷积自动编码器(SCAE)相结合用于检测生菜

中的重金属。然而, 上述研究都是基于深度学习算法, 而深度学习存在样本需求量大、参数多且不易设置等问题, 导致其在小样本量的检测中的应用受到了限制。宽度学习(BLS)是近年来在IEEE会议上提出的新算法, 该算法为扁平结构, 且具有泛化性能好, 迭代收敛快等优点, 使其在解决中小样本的问题上有独特的优势。目前, 宽度学习用于光谱检测领域的研究较少。经过充分的理论研究与前期调研, 我们将宽度学习引入到高光谱成像技术检测虾新鲜度的研究中。

采用高光谱成像技术与宽度学习相结合鉴别虾新鲜度; 以白对虾为研究对象。首先将经过不同预处理后的光谱数据进行可视化, 从而判别出最佳预处理方法, 之后采用多种算法对最佳预处理的光谱数据提取特征波长, 最后将宽度学习与传统机器学习算法进行比较与分析, 最终获得对白对虾新鲜度检测的最优模型。

## 1 实验部分

### 1.1 样本

实验用白对虾购买于北京市朝阳区海鲜市场, 共420只, 平均每只重15g左右。所有样本置于4℃的恒温箱中冷

收稿日期: 2020-10-12, 修订日期: 2021-02-04

基金项目: 国家重点研发计划项目(2017YFE0122100)资助

作者简介: 叶荣珂, 1995年生, 中国农业大学信息与电气工程学院硕士研究生 e-mail: yerongke@163.com

\* 通讯作者 e-mail: sophia\_liu@cau.edu.cn

藏, 每天测定 60 个样本, 共测试 7 d。

## 1.2 仪器

Gaia Sorter 近红外高光谱仪。其主要配置为: 电控平移台、56 mm 定焦近红外镜头、4 个 200 W 溴钨灯组成的均匀光源、计算机箱、遮光罩等; 光谱范围为: 866.4~1 701.0 nm; 采集软件: SpecVIEW software (SpecView Ltd, Uckfield, UK); 曝光时间: 0.09 s; 采样间隔: 3.37 nm; 电控移动平台速度:  $0.55 \text{ cm} \cdot \text{s}^{-1}$ 。

## 1.3 光谱数据处理

### 1.3.1 光谱数据预处理

在数据采集过程中可能会由于一些干扰因素而引入噪声, 导致后续的建模相对较慢且鲁棒性低<sup>[9]</sup>。为消除噪声的影响, 使用多元散射校正(MSC)、标准正态变量校正(SNV)和直接正交信号校正(DOSC)三种方法预处理原始光谱数据。

### 1.3.2 特征波长选择

所有样本被随机分成 3 : 1 的校正集和预测集, 由于部分高光谱数据的全波段光谱信息和空间信息是冗余共线的, 这些干扰信息会影响后续建模过程中模型的鲁棒性、准确性和计算速度。因此, 分别采用随机森林(RF)<sup>[10]</sup>、主成分分析(PCA)<sup>[11]</sup>和二维相关光谱(2D-COS)<sup>[12]</sup>进行特征波长选择。

### 1.3.3 判别模型

偏最小二乘判别分析(PLS-DA)是基于偏最小二乘回归的高维线性识别多元分类模型<sup>[13]</sup>; 极限学习机(ELM)是一类基于前馈神经网络(FNN)构建的机器学习方法, 适用于监督学习和非监督学习问题<sup>[14]</sup>; 宽度学习系统(BLS)是基于随机向量函数链神经网络(RVFLNN)提出的新型替代方案<sup>[15]</sup>, 其结构如图 1 所示。BLS 的输入矩阵是由两部分组成的: 映射节点(mapped feature)和增强节点(enhancement nodes), 映射节点由光谱数据经过线性映射和激活函数变换得到, 增强节点由映射节点经过线性映射和激活函数变换得到。

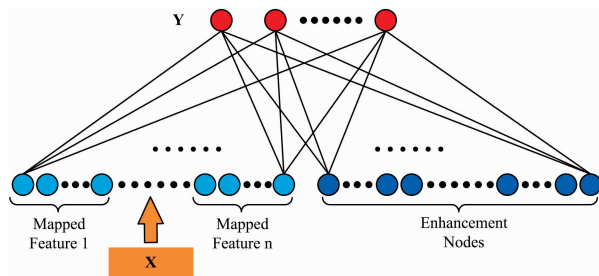


图 1 宽度学习系统结构

Fig. 1 Architecture of the broad learning system

## 2 结果与讨论

### 2.1 原始光谱分析

通过白对虾样本高光谱图像上感兴趣区域(ROI)提取光谱数据。图 2 为白对虾在不同冷藏天数的平均光谱曲线及其标准差。可以看出, 在相同的冷藏天数下, 样品分散度较小, 数据准确。不同冷藏天数的白对虾光谱曲线趋势相似, 在一些波段难以区分。但在 1 150~1 300 nm 范围内, 可以看出

不同冷藏天数的白对虾光谱反射率不同。随着冷藏天数的延长, 反射率逐渐增大。总的来说, 第 1 天和第 7 天的平均光谱曲线差异最大, 而第 2, 3, 4, 5 天的平均光谱曲线差异很小。

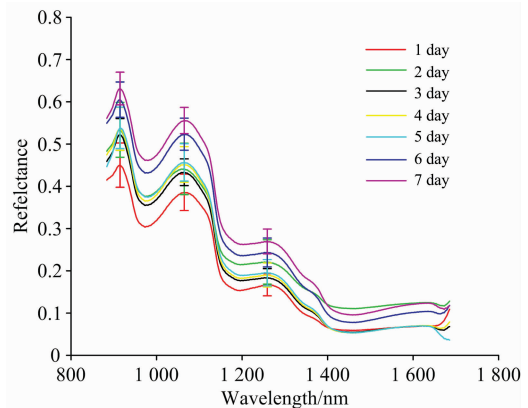


图 2 平均光谱曲线及其标准差

Fig. 2 Average spectral curves with standard deviation

### 2.2 预处理后光谱数据的可视化分析

t 分布随机邻域嵌入(t-SNE)方法可对不同组别样品之间的差异进行可视化分析<sup>[16]</sup>。t-SNE 是由 SNE 衍生出来的一种算法, 它将高维映射到低维, 并试图确保分布概率在它们之间是恒定的。在本研究中, t-SNE 用于将原始光谱和经三种方法(MSC, SNV, DOSC)预处理的光谱的可视化, 并将这些高维光谱还原到二维平面进行分析和比较。图 3 为原始光谱和经过不同预处理的光谱曲线及其相应的 t-SNE 可视化结果。其中 RAW 代表原始光谱, 从图 3(b)中可以直观地看出经 DOSC 预处理后的光谱显示出较好的聚类效果, 虽然仍有交叉重叠, 但与经过 MSC 和 SNV 处理的光谱以及原始光谱相比聚类效果更佳。因此选择经 DOSC 处理后的光谱进行特征波长的建模分析。

### 2.3 特征波长的选择

#### 2.3.1 随机森林

随机森林(RF)是由多棵决策树组成的机器学习算法, 具有分析复杂交互分类特征的能力。当随机森林算法作为高维数据的特征选择工具时, 可以得到所有变量的重要性及其排序。图 4 为使用 RF 后的波长重要性排序, 设置波长变量重要性阈值为 0.15, 高于阈值的变量被选中作为新鲜度特征, 最终通过 RF 选择了 15 个特征波长。

#### 2.3.2 主成分分析

主成分分析(PCA)可提供频谱分析中对不同数据点之间方差的解释。PC<sub>1</sub> 与 PC<sub>2</sub> 组合的 2D 得分图如图 5(a)所示, 一般来说, 同一冷藏天数的数据点倾向于聚集在一起, 随着冷藏天数的持续上升, 相应的样品趋向于沿着 PC<sub>1</sub> 轴正方向移动。图 5(b)为 PC<sub>1</sub> 和 PC<sub>2</sub> 负载线, 显著峰谷处的波长被认为是识别冷藏天数的重要信息, 应加以选择, 通过 PCA 最终提取了 7 个特征波长。

#### 2.3.3 二维相关光谱

二维相关光谱(2D-COS)是一种常用的数学分析形式。

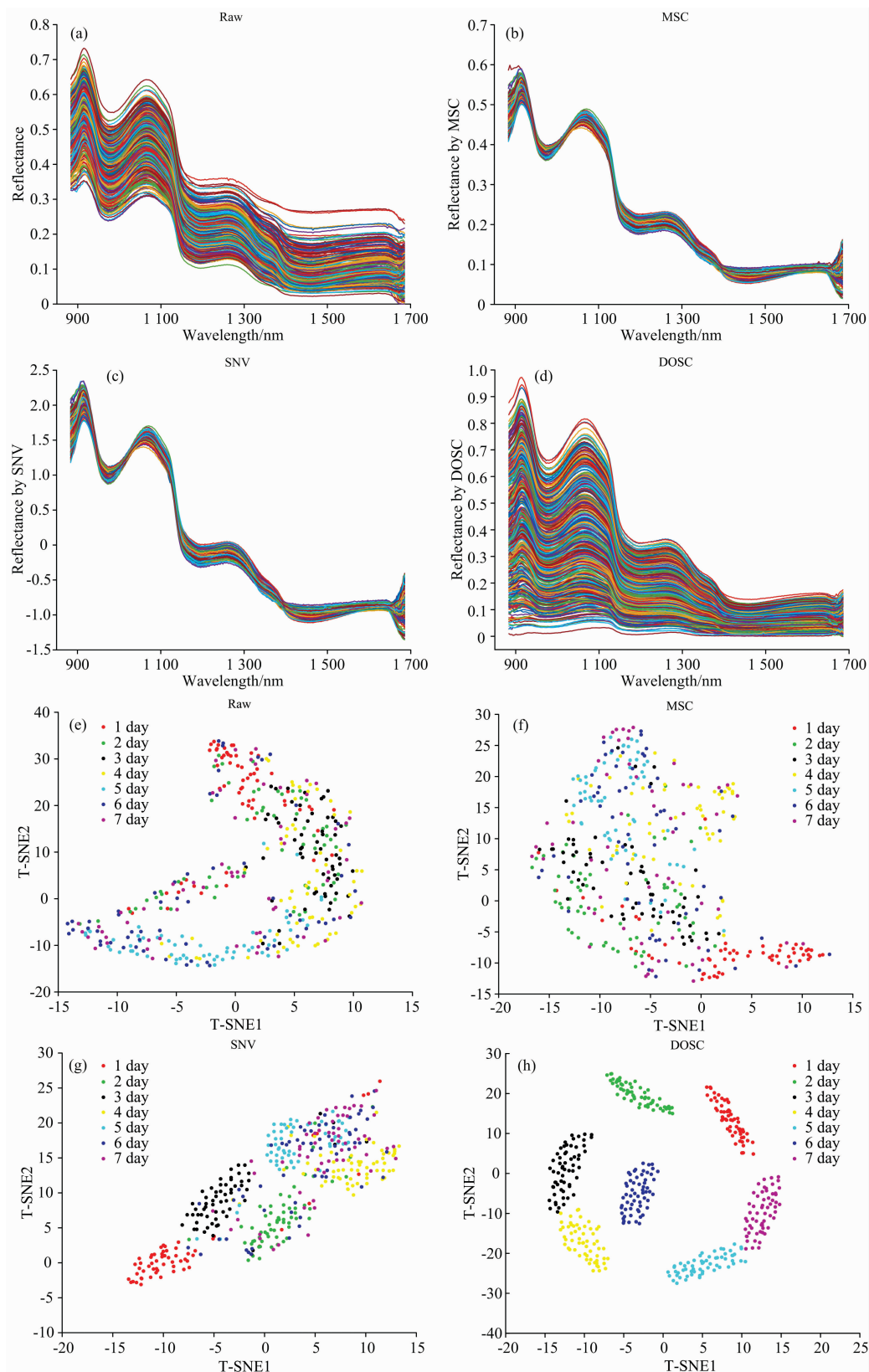


图 3 可视化分析

(a), (b), (c), (d): 不同预处理方法的光谱曲线; (e), (f), (g), (h): t-SNE 可视化

Fig. 3 Visualization analysis

(a), (b), (c), (d): Spectral curves by using different preprocessing methods; (e), (f), (g), (h): Visualization using t-SNE

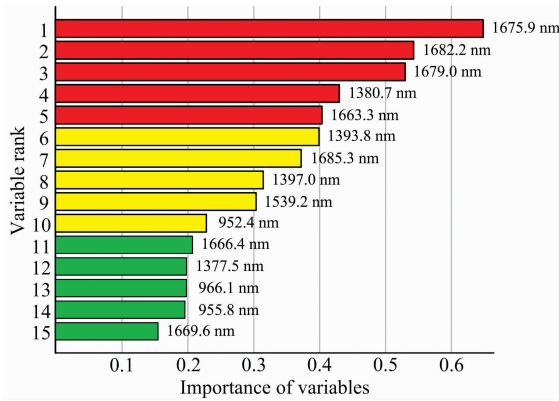


图 4 使用随机森林的波长变量重要性分析  
Fig. 4 Analysis of the importance of wavelength variables by RF

所获得的不同冷藏天数 2D-COS 分析如图 6 所示。在图 6(a) 同步等高线图的对角线观察到 914.6 和 1 058.0 nm 这两个主要的自相关峰，另一个弱的自相关峰 1 249.1 nm 也在图 6(b) 相应的自相关峰强度曲线图可以清楚地看到。这些频段的强度随冷藏天数的不同而有很大差异，因此，通过这 3 个波长可有效识别冷藏天数。最终通过 2D-COS 选择了 3 个特征波长。

2.4 基于特征波长的建模分析

为了进一步选择最佳的模型组合，将特征波长光谱数据与判别模型相结合，以获得最佳的模型组合。表 1 为不同特征选择算法结合判别模型建模分析的结果。PLS-DA 参数为潜在变量个数。ELM 参数为隐含层层数。BLS 参数为正归化参数、增强节点的收缩尺度。具体分析来说：(1) 不同特征选择算法(RF, PCA, 2D-COS)与相同的判别模型中，RF 算法与判别模型相结合的校正集和预测集的精度最高(校正集和预

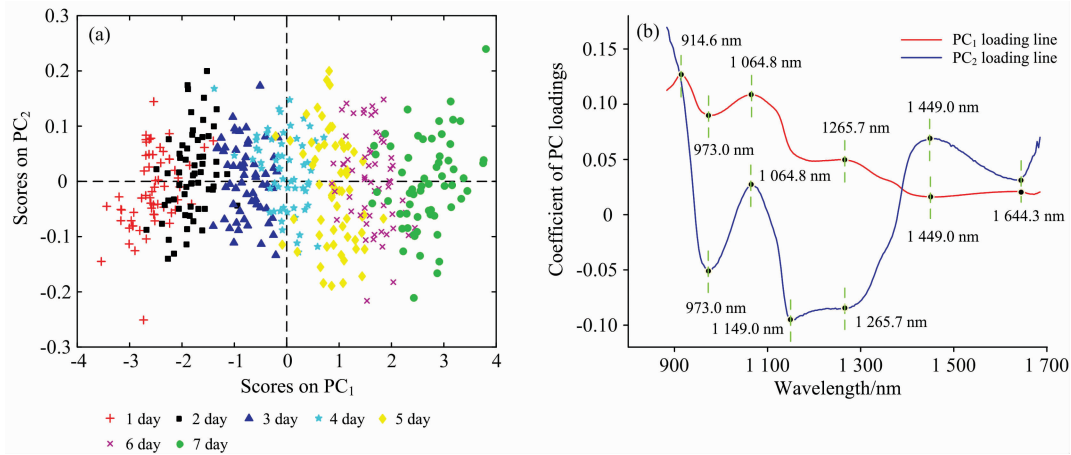


图 5 有效 PC 得分和负载分析  
(a): PC<sub>1</sub> 与 PC<sub>2</sub> 的 PCA 得分图; (b): PC<sub>1</sub> 和 PC<sub>2</sub> 负载线上的选择出来的波长

Fig. 5 Analysis of effective PC scores and loading

(a): PCA score plot of PC<sub>1</sub> vs. PC<sub>2</sub>; (b): Wavelength selection on PC<sub>1</sub> and PC<sub>2</sub> loading lines

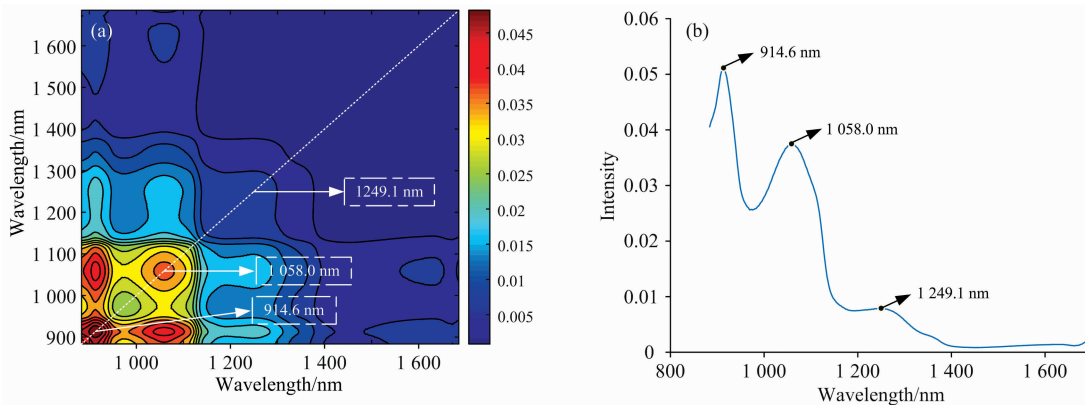


图 6 不同冷藏天数样品的 2D-COS 光谱

(a): 同步等高线图; (b): 自相关峰强度曲线

Fig. 6 The 2D-COS spectrum of samples with different days of refrigeration

(a): Synchronous contourmap plot; (b): Autocorrelation peak intensity curve

表 1 基于特征波长的建模结果分析  
Table 1 Analysis of modeling results based on feature wavelengths

Model	Parameter	Accuracy of calibration set	Accuracy of prediction set
RF-PLS-DA	13	91.75%(289/315)	90.48%(95/105)
RF-ELM	34	95.56%(301/315)	95.24%(100/105)
RF-BLS	( $2^{-30}$ , 0.9)	98.41%(310/315)	97.14%(102/105)
PCA-PLS-DA	5	90.48%(285/315)	89.52%(94/105)
PCA-ELM	21	93.97%(296/315)	91.43%(96/105)
PCA-BLS	( $2^{-20}$ , 0.9)	96.19%(303/315)	94.29%(99/105)
2D-COS-PLS-DA	2	85.10%(269/315)	81.90%(86/105)
2D-COS-ELM	14	89.21%(281/315)	87.62%(92/105)
2D-COS-BLS	( $2^{-20}$ , 0.7)	91.43%(288/315)	89.52%(94/105)

测集准确率都超过 90%)，表明 RF 可以最大程度地消除冗余并保留有用信息。(2)特征选择算法结合了 3 个判别模型(PLS-DA, ELM, BLS)，其中，BLS 模型结合特征选择算法可实现最佳分类精度(校正集和预测集准确率都超过 89%)。综上所述，将 RF-BLS 作为新鲜度最佳判别模型(校正集和预测集准确率分别为 98.41%和 97.14%)。

### 3 结 论

将宽度学习引入对白对虾新鲜度的检测研究。采用 3 种预处理方法(MSC, SNV, DOSC)，3 种特征选择算法(RF,

PCA, 2D-COS)，2 种经典机器学习算法(PLS-DA, ELM)和宽度学习模型(BLS)建立了虾新鲜度的检测方法。经过实验与结果分析，DOSC 聚类效果最佳，RF 算法在 3 种特征选择算法中消除冗余信息性能最好，而 BLS 与线性机器学习算法(PLS-DA)以及非线性机器学习算法(ELM)相比获得了更高的分类精度，因此 RF-BLS 模型效果最好(校正集和预测集准确率分别为 98.41%和 97.14%)，验证了宽度学习用于光谱检测领域的可行性。实验结果表明高光谱成像技术结合宽度学习识别虾的新鲜度是可行的，方法可以为在线检测虾新鲜度系统的开发提供理论依据。

### References

- [1] WANG Wei(王伟). Food Research and Development(食品研究与开发), 2017, 38(22): 176.
- [2] WANG Su-hua, CHEN Ji-ming, ZHU Hai, et al(王素华, 陈积明, 朱海, 等). Science and Technology of Food Industry(食品工业科技), 2011, 32(7): 386.
- [3] LING Ping-hua, XIE Jing, ZHAO Hai-peng, et al(凌萍华, 谢晶, 赵海鹏, 等). Jiangsu Journal of Agricultural Sciences(江苏农业学报), 2010, 26(4): 828.
- [4] WANG Wei, CHAI Chun-xiang, LU Xiao-xiang(王伟, 柴春祥, 鲁晓翔). Food & Machinery(食品与机械), 2013, 29(4): 233.
- [5] Baek I, Kusumaningrum D, Kandpal L, et al. Sensors, 2019, 19(2): 271.
- [6] Yu X J, Tang L, Wu X F, et al. Food Analytical Methods, 2018, 11(3): 768.
- [7] Zhang C, Wu W Y, Zhou L, et al. Food Chemistry, 2020, 319: 126536.
- [8] Zhou X, Sun J, Tian Y, et al. Food Chemistry, 2020, 321: 126503.
- [9] Cheng J H, Dai Q, Sun D W, et al. Journal of Food Engineering, 2015, 161: 33.
- [10] HUANG Chong, XU Zhao-xin, ZHANG Chen-chen, et al(黄翀, 许照鑫, 张晨晨, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2020, 36(9): 177.
- [11] Zhang L, Sun H, Rao Z H, et al. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2020, 229: 117973.
- [12] Jiang H Z, Cheng F N, Shi M H. Foods, 2020, 9(2): 154.
- [13] Bai Z Z, Hu X J, Tian J P, et al. Food Chemistry, 2020, 331: 127290.
- [14] Ye R K, Chen Y Y, Guo Y C, et al. Applied Sciences, 2020, 10(16): 5498.
- [15] Chen C L, Liu Z L, et al. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(1): 10.
- [16] Zhang L, Rao Z H, Ji H Y. Spectroscopy Letters, 2020, 53(3): 207.

# Shrimp Freshness Detection Method Based on Broad Learning System

YE Rong-ke<sup>1</sup>, KONG Qing-chen<sup>1</sup>, LI Dao-liang<sup>1, 2</sup>, CHEN Ying-yi<sup>1, 2</sup>, ZHANG Yu-quan<sup>1</sup>, LIU Chun-hong<sup>1, 2\*</sup>

1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

2. National Innovation Center for Digital Fishery, China Agricultural University, Beijing 100083, China

**Abstract** To improve the accuracy of shrimp freshness discrimination, we proposed a shrimp freshness detection method based on a broad learning system in this paper. In this study, firstly, multivariate scatter correction (MSC), standard normal variate (SNV), and direct orthogonal signal correction (DOSC) were used to preprocess the raw hyperspectral data of shrimp with different days of refrigeration. And secondly, t-distributed stochastic neighbor embedding (t-SNE) was used to visualize the data after preprocessing, and the visualization results showed that the DOSC-processed data had the best clustering effect. Then, the spectral data after DOSC preprocessing were used for feature extraction using random forest (RF), principal component analysis (PCA), and two-dimensional correction spectroscopy (2D-COS). Finally, the shrimp freshness was modeled based on the characteristic wavelength, and the broad learning system (BLS) was used in shrimp freshness modeling for the first time in this paper and compared with the classical discriminant models such as partial least squares discrimination analysis (PLS-DA) and extreme learning machine (ELM). The results indicated that the RF method minimized the redundant information in the spectra, while the BLS had high accuracy and shorter discrimination time than the linear modeling method PLS-DA and the nonlinear modeling method ELM, and thus the combined RF-BLS model obtained the best freshness discrimination performance. The experimental results indicated that the hyperspectral imaging technology combined with broad learning system to identify shrimp freshness is feasible. The method proposed in this paper can provide a theoretical basis for developing an online shrimp freshness detection system.

**Keywords** Freshness of shrimp; Direct orthogonal signal correction; Random forest; Broad learning system; Hyperspectral imaging technology

(Received Oct. 12, 2020; accepted Feb. 4, 2021)

\* Corresponding author