

基于近红外光谱与深度学习的紫菜水分无损检测

张雯雯, 胡亚东, 孙文珂, 潘明轩, 沈照鹏, 周兴虎, 钱津, 王鹏

Non-destructive Detection of Water Content in *Porphyra* Based on Near-infrared Spectroscopy and Deep Learning

ZHANG Wenwen, HU Yadong, SUN Wenke, PAN Mingxuan, SHEN Zhaopeng, ZHOU Xinghu, QIAN Jin, and WANG Peng

在线阅读 View online: <https://doi.org/10.13386/j.issn1002-0306.2023100153>

您可能感兴趣的其他文章

Articles you may be interested in

基于近红外光谱的条斑紫菜菌落总数快速检测技术

Rapid Detection of Total Bacterial Count of *Porphyra yezoensis* Based on Near Infrared Spectroscopy

食品工业科技. 2022, 43(16): 322–328 <https://doi.org/10.13386/j.issn1002-0306.2021110257>

基于近红外光谱的红提内部品质无损检测研究

Non-destructive Detection of the Internal Quality of Red Globe Grapes Based on Near Infrared Spectroscopy

食品工业科技. 2022, 43(22): 7–14 <https://doi.org/10.13386/j.issn1002-0306.2022030285>

基于主成分分析和人工神经网络的近红外光谱大豆产地识别

Soybean Origin Identification Based by Near-Infrared Spectrum Based on Principal Component Analysis and Artificial Neural Network Model

食品工业科技. 2021, 42(9): 270–274 <https://doi.org/10.13386/j.issn1002-0306.2020060271>

近红外光谱法测定面粉的水分、脂肪、碳水化合物和蛋白质含量

Determination of Moisture, Fat, Carbohydrates and Protein Contents in Flour by Near Infrared Spectroscopy

食品工业科技. 2020, 41(12): 256–263 <https://doi.org/10.13386/j.issn1002-0306.2020.12.042>

基于近红外光谱技术快速检测椰汁品质

Rapid Detection of Coconut Juice Quality Based on the Near Infrared Spectroscopy

食品工业科技. 2019, 40(12): 235–240 <https://doi.org/10.13386/j.issn1002-0306.2019.12.038>

近红外光谱法检测婴幼儿营养米粉蛋白质含量

Determination of Protein Contents in Infant Nutritious Rice Flour by Near Infrared Spectroscopy

食品工业科技. 2019, 40(6): 237–240,251 <https://doi.org/10.13386/j.issn1002-0306.2019.06.039>



关注微信公众号，获得更多资讯信息

张雯雯,胡亚东,孙文珂,等.基于近红外光谱与深度学习的紫菜水分无损检测[J].食品工业科技,2024,45(21):190–197. doi: 10.13386/j.issn1002-0306.2023100153

ZHANG Wenwen, HU Yadong, SUN Wenke, et al. Non-destructive Detection of Water Content in *Porphyra* Based on Near-infrared Spectroscopy and Deep Learning[J]. Science and Technology of Food Industry, 2024, 45(21): 190–197. (in Chinese with English abstract). doi: 10.13386/j.issn1002-0306.2023100153

· 分析检测 ·

基于近红外光谱与深度学习的紫菜水分无损检测

张雯雯¹,胡亚东²,孙文珂³,潘明轩²,沈照鹏³,周兴虎²,钱津²,王鹏^{3,*}

(1.中国海洋大学海德学院,山东青岛 266104;

2.江苏省沿海开发集团有限公司,江苏省沿海开发投资有限公司,
江苏海洋生物资源创新中心,江苏南京 210019;

3.中国海洋大学食品科学与工程学院,山东青岛 266104)

摘要:为探索近红外光谱结合深度学习网络对紫菜水分定量检测的可行性,本研究检测并收集了479组干条斑紫菜的光谱数据和水分含量数据,分别使用四种方法对其中的光谱数据进行了预处理,并在全波段下建立了四种传统定量水分预测模型和一种卷积神经网络(Convoluted Neural Networks, CNN)深度学习水分预测模型。对比五种模型预测结果后发现,在S-G平滑结合二阶导数的预处理方法下所建立的CNN模型预测效果最佳,其预测均方根误差(Root-Mean-Square Error of Prediction, RMSEP)值为0.456,预测集决定系数(Coefficient of Determination of Prediction, R_p^2)值为0.990,优化后,该模型的RMSEP值降至0.342, R_p^2 值可以达到0.994(>0.8),同时,外部验证相对误差(Ratio of Performance to Deviation for Validation, RPD)值达6.155(>3),证明了模型实际应用于农业和食品工业的可能性。该CNN模型能够快速、准确、无损地预测条斑紫菜的水分含量,提高了紫菜水分检测的效率和准确性,为相关干制水产品的质量控制提供了重要的参考依据。

关键词:条斑紫菜,水分含量,近红外光谱,深度学习,卷积神经网络

中图分类号:TS201.1 **文献标识码:**A **文章编号:**1002-0306(2024)21-0190-08

DOI: 10.13386/j.issn1002-0306.2023100153

本文网刊:



Non-destructive Detection of Water Content in *Porphyra* Based on Near-infrared Spectroscopy and Deep Learning

ZHANG Wenwen¹, HU Yadong², SUN Wenke³, PAN Mingxuan², SHEN Zhaopeng³, ZHOU Xinghu², QIAN Jin², WANG Peng^{3,*}

(1.Haide College, Ocean University of China, Qingdao 266104, China;

2.Jiangsu Coast Development Group Co., Ltd., Jiangsu Jinniu Industry & Trade Co., Ltd., Jiangsu Innovation Center of Marine Resources, Nanjing 210019, China;

3.College of Food Science and Engineering, Ocean University of China, Qingdao 266104, China)

Abstract: In order to explore the feasibility of combining near-infrared (NIR) spectroscopy and deep learning network for quantitative moisture detection, the dried *Porphyra* was divided into 479 groups, which detected the NIR spectra and moisture content. Four traditional quantitative moisture prediction models and a convolution neural networks (CNN) deep-learning moisture prediction model were finally established at full spectrum by preprocessing and analyzing the experimental data. After comparing the prediction results of the five models, it was found that the CNN model established

收稿日期: 2023-10-23

基金项目: 江苏省沿海开发集团有限公司江苏海洋生物资源创新中心自主立项项目(2023YHTZZ01)。

作者简介: 张雯雯(1992-),女,大学本科,研究方向:食品科学与工程,E-mail:wwz_mail_02@163.com。

*通信作者: 王鹏(1980-)(ORCID: 0000-0002-2337-5812),男,博士,教授,研究方向:海洋应用微生物,E-mail:pengwang@ouc.edu.cn。

by the preprocessing method of S-G smoothing combined with the second derivative had the best prediction effect. Its root-mean-square error of prediction (RMSEP) value was 0.456 and the coefficient of determination of prediction (R_p^2) value was 0.990. After optimization, the RMSEP value of the model was reduced to 0.342 and the R_p^2 value could reach 0.994 (>0.8). At the same time, the ratio of performance to deviation for validation (RPD) was 6.155 (>3), which proved the possibility of practical application of the model in agriculture and food industry. The CNN model could predict the moisture content quickly, accurately, and non-destructive, improve the efficiency and accuracy of moisture detection, and provide an important reference for the quality control of related dry aquatic products.

Key words: *Porphyra*; moisture content; near infrared spectroscopy; deep learning; convolution neural networks

条斑紫菜(*Porphyra yezoensis*)隶属于红藻门(Rhodophyta)、原红藻纲(Protoflorideophyceae)、红毛菜目(Bangiales)、红毛菜科(Bangiaceae)、紫菜属(*Pyropia*)^[1],是一种重要的海洋经济作物。条斑紫菜主要分布于黄海、渤海、东海北部沿岸,其主产区在长江以北沿海的山东、江苏两省^[2]。作为海洋农业和出口创汇的主导品种之一,我国每年生产的条斑紫菜有70%销往国外,出口量占国际紫菜市场贸易总量的50%以上^[3]。条斑紫菜营养丰富,若干燥不到位,极易在运输和贮存环节因水分含量过高而导致腐败变质^[4-5]。因此有必要在紫菜烘干工艺之后、封装工艺之前^[6-7],对生产线上的所有紫菜产品的水分进行在线监测,从而及时去除不合格的产品。但目前常用的水分检测方法主要为直接烘干法^[8-9],该法耗时且损耗样品,不适用于大批量检测,也不能满足生产现场对水分检测的快速、无损等测定要求^[10]。在国际紫菜标准的制定中,水分是评价紫菜质量的重要参数之一,若缺少高效精准的水分检测技术,将不利于我国紫菜的质量管理,从而进一步影响我国紫菜进出口贸易发展^[11]。在此背景下,近年来,大批新型水分检测技术应运而生。以测试简易、分析快速为特点的近红外光谱技术便是其中之一^[12]。

近红外光谱是介于可见光和中红外光之间的电磁波谱,其波长范围为780~2526 nm^[13],它能反映O-H、C-H、N-H、S-H等不同化学键的信息^[14],这使得近红外光谱应用于紫菜水分检测成为可能。目前,近红外分析技术在紫菜质量检测方面已有较多应用。Guan等^[15]利用近红外光谱技术建立了紫菜最佳收获时间的识别模型;Wu等^[16]采用偏最小二乘法结合无信息变量消除算法构建了紫菜蛋白含量近红外检测模型。除利用传统定量分析模型之外,部分研究将近几年以突出的特征提取能力而得到广泛应用的深度学习模型引入近红外光谱检测领域,如方明明等^[17]基于卷积神经网络建立了苹果脆片水分近红外预测模型;Benmouna等^[18]基于近红外光谱和一维卷积神经网络建立了富士苹果的质量检测模型。而利用深度学习算法对紫菜质量进行检测的相关研究仍较为欠缺。本文为探索深度学习模型对于紫菜近红外光谱的特征学习能力和水分的回归分析能力,对相关光谱预处理方法进行了优化,同时在全波段下比较了传统定量模型与深度学习模型的预测效果,并通过

调整模型超参数进一步挖掘深度学习模型的预测潜力。

1 材料与方法

1.1 材料与仪器

干紫菜薄片 实验共选取479组干条斑紫菜样本,样品采购自赣榆、秦皇岛、青岛、威海等全国各主要产地,分别来自于2022年年末一茬和2023年四茬紫菜样品(表1)。测定时,每组取10张条斑紫菜作为平行样,每张紫菜的尺寸为19 cm×21 cm。

表1 紫菜样品批次统计
Table 1 Batch statistics of *Porphyra* samples

紫菜样品批次	数目
2022年年末	145
2023年头茬	150
2023年二茬	54
2023年三茬	49
2023年五茬	81

MicroNIR 1700ES VIAVI 近红外光谱仪 美国 VIAVI,采用集成真空钨灯作为照明光源,波长范围为908~1676 nm,样品工作距离(即样品与探头距离)范围为0~15 mm;LRH-70 恒温干燥箱 上海一恒科学仪器有限公司;ME203E 电子天平 上海梅特勒-利特多有限公司。

1.2 实验方法

1.2.1 紫菜水分检测 按照GB 5009.3-2016《食品安全国家标准 食品中水分的测定》中直接测定法对干条斑紫菜样本进行水分测定。

1.2.2 光谱采集 如图1所示,将每张样品均分为4×5的20份均等区域,竖立置于样品台上。分别采集每组样品每张条斑紫菜片四个角落和中间2点,共6个点的光谱数据,以消除紫菜不均匀性所造成的误差。

采样方式为漫反射积分球模式,波长范围为900~1650 nm、光谱分辨率为6.2 nm、样品工作距离为3 mm、信噪比为23000(扫描次数为100次,采集后取平均值)、工作温度为25 °C。预热设备15 min,校正暗电流并以聚四氟乙烯白板为参比。

1.3 数据处理

1.3.1 样本集划分 目前最常用的样本集划分方法是基于光谱变量的 Kennard-Stone(K-S)法以及在



图 1 近红外光谱采集装置图

Fig.1 Schematic diagram of near-infrared acquisition device

注: a.薄片型干条紫菜样品; b.MicroNIR 1700ES 探测头; c.标准白板; d.外接平板与内置软件。

K-S 法的基础上提出的 Sample set Partitioning based on joint X-Y distances(SPXY)法。本文采用 SPXY 法,相对于 K-S 法该法能够同时考量理化值和光谱两个变量,有效覆盖了多维向量空间,并增加了样本之间的差异性和代表性^[19-20]。通过 SPXY 法,本文将 2022 年年末和 2023 年头茬、二茬和三茬采集的紫菜数据集划分训练集和测试集。同时,因 2023 年采集的五茬紫菜水分含量分布最广泛,故将其作为外部验证集评价最终模型的预测能力。

1.3.2 光谱数据预处理 在进行近红外光谱检测时,环境背景噪声和仪器误差会降低光谱数据的解释性,进而影响建模精度。因此,必须对已获取的光谱信息进行预处理操作。本研究采用多元散射校正(Multiplicative Scatter Correction, MSC)、标准正态变量校正(Standard Normal Variate Correction, SNV)、一阶导数校正(First Derivative Correction, FD)、二阶导数校正(Second Derivative Correction, SD)和 S-G 平滑(Savitzky-Golay, S-G)共五种单类方法和四种组合方法对光谱数据进行预处理。对于不同预处理方法,统一建立卷积神经网络模型,并通过决定系数 R_p^2 的大小评价各预处理方法的效果。

1.3.3 深度学习预测模型的建立 卷积神经网络(Convolution Neural Networks, CNN)是一个多层非全连接的深度神经网络^[21],其特点是通过卷积层和池化层的多次反复交替学习来提取数据特征。相比于普通全连接神经网络,卷积神经网络凭借其降低模型

复杂度、高效处理数据等优势,近年来被应用于目标检测、图像分类等计算机视觉领域。

CNN 基本架构由输入层、卷积层、池化层、全连接层和输出层堆叠而成,本研究使用的 CNN 结构如图 2 所示,由 3 层一维卷积层(conv1d)、2 层最大池化层(maxpooling1d)和 2 个全连接层(dense)组成。

1.3.4 传统定量模型的建立 本文同时选用多元线性回归(Multiple Linear Regression, MLR)法、偏最小二乘(Partial Least Squares, PLS)法、支持向量回归(Support Vector Regression, SVR)和人工神经网络(Artificial Neural Network, ANN)共四种传统定量校正方法进行预测模型的建立,并将其与卷积神经网络所建立的模型相比较,探究使用卷积神经网络建立紫菜水分预测模型的优势。

用于建立模型的数据分析软件为开源 Python 3.10.9 与相关数据科学包,数据科学包包括: scikit-learn 0.22.1、scipy 1.4.1、NumPy 1.19.4、pandas 1.0.1、Matplotlib 3.1.3、TensorFlow 2.11.0 等。

1.3.5 模型评价标准 引入预测决定系数(Coefficient of Determination of Prediction, R_p^2)、外部验证相对误差值(Ratio of Performance to Deviation for Validation, RPD)和预测均方根误差(Root-Mean-Square Error of Prediction, RMSEP)3 个指标对预测模型进行评价。其中,决定系数值 R_p^2 越接近于 1,说明模型的拟合效果越好; RPD 值用于外部验证,其值越大则预测模型的效果越好; RMSEP 的值越低,

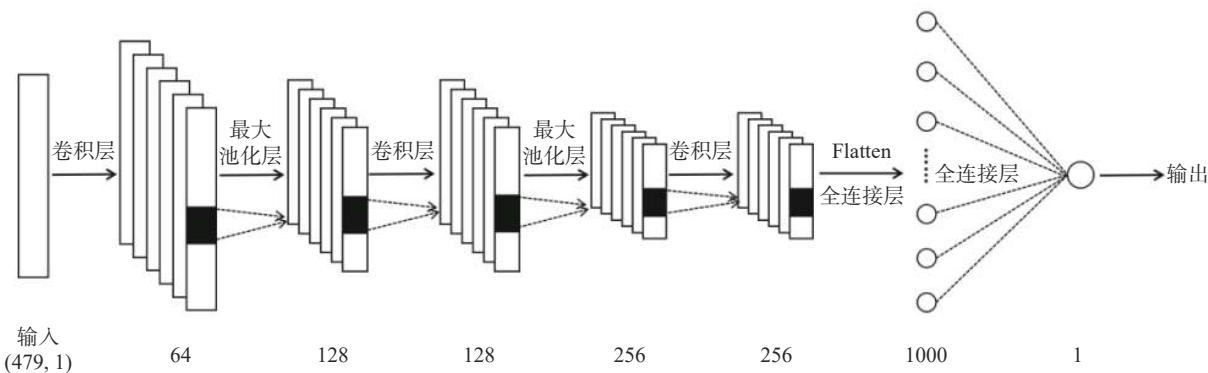


图 2 CNN 基本架构模式图

Fig.2 CNN basic architecture model diagram

说明模型预测的准确率越高。 R_p^2 、RPD 及 RMSEP 计算公式如下。

$$R_p^2 = 1 - \frac{\sum_{i=1}^n (y_{i,actual} - y_{i,predicted})^2}{\sum_{i=1}^n (y_{i,actual} - \bar{y}_{i,predicted})^2}$$

$$RPD = \frac{1}{\sqrt{1 - R_p^2}}$$

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (y_{i,actual} - y_{i,predicted})^2}{n - 1}}$$

式中, $y_{i,actual}$ 为验证集样品水分实际测定值; $y_{i,predicted}$ 为验证集样品水分预测值; n 为验证集的样品数。

2 结果与分析

2.1 样品水分检测结果

采用规定方法对紫菜水分进行测定, 得到样品水分含量频率分布直方图和正态分布曲线见图 3。由图可见 2022 年年末一茬紫菜以及 2023 年四茬紫菜的水分含量数据均呈正态分布, 且数据集具有代表性, 适合进行模型训练及验证。

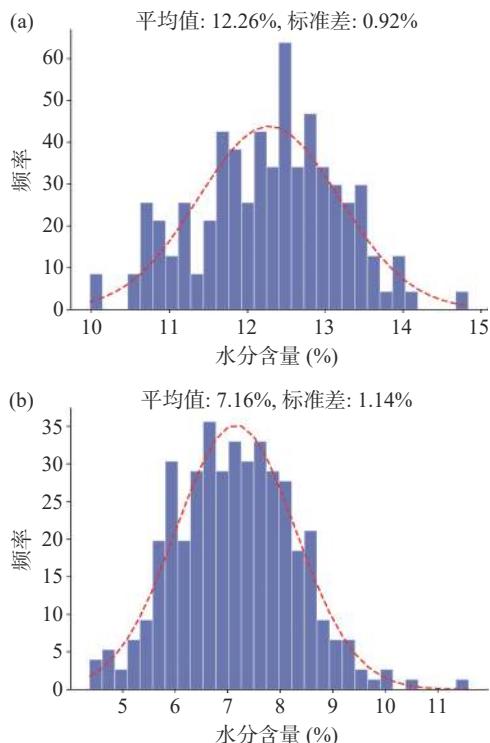


图 3 2022 年紫菜水分含量分布图(a)和 2023 年紫菜水分含量分布图(b)

Fig.3 Distribution of water content of samples in 2022 (a) and 2023 (b)

2.2 光谱采集结果

使用近红外光谱仪对紫菜样品进行光谱采集, 得到光谱数据如图 4 所示。由于光谱数据受随机噪声、样品表面散射及光程变化等因素的影响, 吸收光谱线带较宽, 特征吸收峰不明显, 无法直接建模分析。

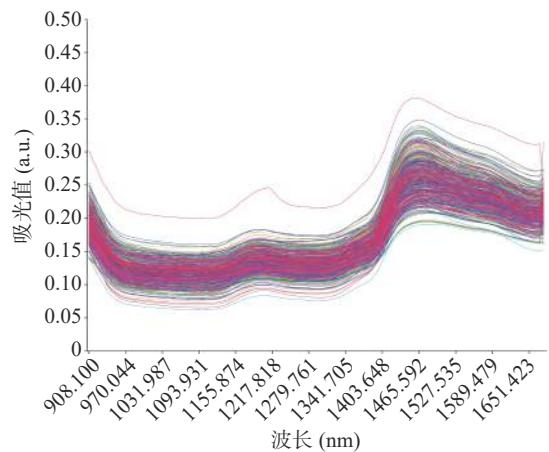


图 4 紫菜样品近红外光谱采集

Fig.4 Near-infrared spectrum acquisition of samples

2.3 光谱数据预处理结果

将采集到的光谱数据输入 CNN 模型进行训练之前, 需要先对数据进行预处理。本文共选择 MSC、SNV、FD、SD 和 S-G 平滑共 5 种单类预处理方法, 并在此基础上进行优化。本文所选五种方法对于光谱数据的预处理效果各有不同, 其中, MSC 和 SNV 法可以消除样品因颗粒分布不均或颗粒大小不同而导致的散射^[22]; S-G 平滑法可用于消除光谱中的随机噪声^[23]; 一、二阶导数变换能够消除光谱中基线的平移和漂移, 提高分辨率和灵敏度^[24]。

由图 5(f)可以看出, 除 MSC 之外, 其他四种单类预处理方法处理后的数据相对于未经过预处理的数据来说所建模型的预测效果表现均有提升。经过预处理后, 模型的 RMSEP 值均有所降低, 且 R_p^2 值均有所升高。从图 5(d)可以看出, 经过 MSC 处理后的光谱特征峰并不明显, 同时线带宽度大, 散射影响未被有效清除, 故造成其模型训练效果并不理想。

为探索两种预处理方法组合使用的方式对于模型训练的叠加效果, 本研究采用了 S-G 平滑+FD、S-G 平滑+SD、SNV+FD、SNV+SD 共四种组合预处理方法, 以达到“清除随机噪声+消除基线平移”或“清除散射影响+消除基线平移”的双重效果。由图 5(f)可以看出, 经过组合方法处理过的数据所建模型的预测效果显著优于单类方法数据所建模型。其中, 通过 S-G 平滑+SD 处理所得到的模型决定系数 R_p^2 值最高, 为 0.990; 同时其预测均方根误差 RMSEP 最小, 为 0.456, 表明利用该方法预处理后的光谱数据能够较好地适用于 CNN 模型的特征识别。由图 5(h)也可以看出, 在 S-G 平滑方法上进行二阶导数的操作可以明显突出特征吸收峰, 同时在一定程度上清除毛刺噪声。

2.4 传统定量模型结果比对

根据 1.3.1 中的样本集划分方法, 将 2022 年和 2023 年采集的三茬紫菜(2023 年五茬紫菜除外, 2023 年五茬紫菜单独作为验证集)按照 5:1 的比例划分为训练集和测试集。样本集划分结果见表 2。

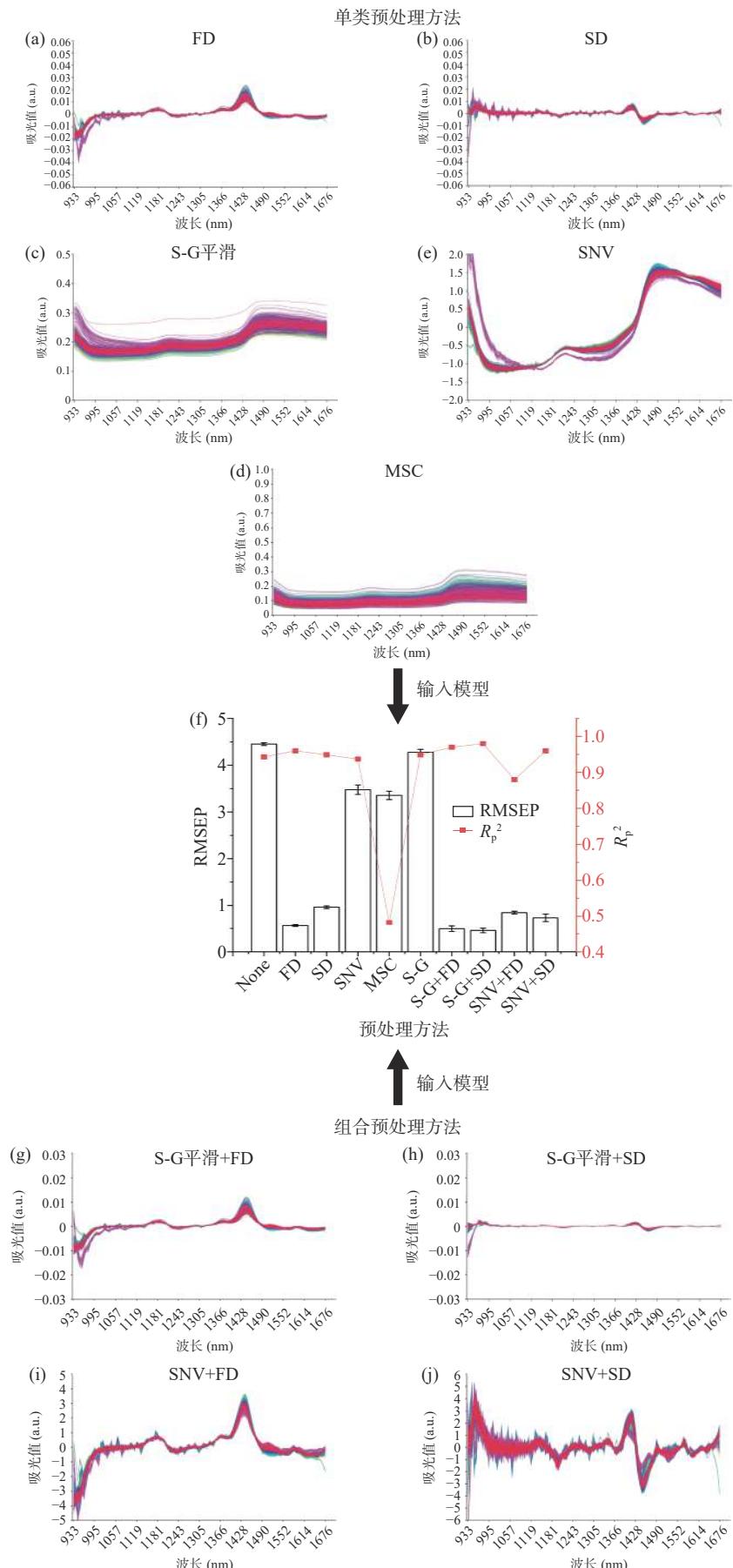


图 5 通过不同方法预处理后的近红外光谱及其建立的 CNN 模型的预测结果

Fig.5 NIR spectra pretreated by different methods and the prediction results of the established CNN model

注: (a)FD 处理; (b)SD 处理; (c)S-G 平滑处理; (d)MSC 处理; (e)SNV 处理; (f)不同预处理方法下建立水分含量预测模型的结果; (g)S-G 平滑+FD 处理; (h)S-G 平滑+SD 处理; (i)SNV+FD 处理; (j)SNV+SD 处理。

表 2 样本集划分结果

Table 2 Results of sample set partitioning

样本集	样本数目
训练集	332
测试集	66
验证集	81

不同的模型对应有不同的最佳预处理方法, 本研究在利用不同方法预处理光谱数据后, 分别输入四种传统定量分析模型(SVR、PLS、MLR、ANN)进行水分含量预测, 表 3 是四种传统模型及 CNN 模型各自所对应的三种最佳预处理方法处理后所得到的模型预测效果。

表 3 最佳预处理方法下不同建模方法建立水分含量预测模型的结果

Table 3 Results of water content prediction models established by different modeling methods under the optimal pretreatment method

模型	预处理方法	RMSEP	R_p^2
MLR	FD	0.629	0.974
	SD	0.632	0.974
	S-G+FD	0.629	0.974
SVR	FD	0.556	0.980
	S-G+FD	0.553	0.980
	S-G+SD	0.556	0.980
ANN	FD	1.341	0.879
	SD	1.339	0.883
	S-G+FD	1.440	0.859
PLS	FD	0.753	0.963
	SD	0.715	0.967
	S-G+FD	0.764	0.962
CNN	FD	0.562	0.980
	S-G+FD	0.494	0.985
	S-G+SD	0.456	0.990

在四种传统定量分析模型中, SVR 模型的预测优势十分显著。在利用 S-G 平滑+FD 预处理后, 其预测误差 RMSEP 降至 0.553, R_p^2 值达到 0.980, 与 Li 等^[25]基于土壤光谱数据所建立的 SVR 模型 (R_p^2 值为 0.674) 相比, 预测准确率更高且模型稳定性强。同时根据 Williams 等^[26]的研究结果, 模型决定系数 R_p^2 大于 0.8 时, 可认为模型在农业和食品工业实际可用, 本文 SVR 模型的 R_p^2 为 0.980, 说明基于传统化学计量学方法的条斑紫菜水分定量模型建立成功。

将 SVR 模型与深度学习类模型 CNN 对比, 可以看出在 S-G 平滑+SD 预处理后的 CNN 模型, 虽然没有经过优化调整, 但其 RMSEP 值比 SVR 模型低 0.097, 决定系数 R_p^2 比 SVR 模型高, 模型预测效果更好。此 CNN 模型相较于 Liu 等^[27]基于组合传统定量模型——偏最小二乘和支持向量机(PLS-SVR)所建立的龙香梨水分定量模型(R_p^2 值为 0.760)来说, 表现出更好的预测性能。因此, 可以看出深度学习在建立水分预测模型的方面相较于传统化学计量法具有明显优势。

2.5 CNN 模型分析

根据 2.3 中对于预处理方法的评价, 本研究选择 SG+SD 的方法作为 CNN 模型最终的预处理方法。并在此基础上, 通过调整 CNN 模型的学习率和迭代次数来进一步优化 CNN 的最终预测效果。

2.5.1 学习率对模型预测效果的影响 学习率在模型训练中起着关键作用。过大的学习率会导致训练过程不稳定、振荡, 并使模型难以收敛, 预测误差大^[28]; 过小的学习率则会使模型出现训练速度缓慢, 陷入局部最优解等问题^[29]。

由图 6 可知, 当学习率逐渐减小时, 模型 RMSEP 值逐渐减小, R_p^2 值逐渐增大, 这是因为当学习率逐渐减小时, 模型的步长减小, 可以更精确地搜索参数空间, 从而更有可能找到最优解^[30]。而当学习率降至 0.00001 时, 过小的步长导致模型在参数空间中移动的速度变慢, 使模型陷入局部最优解, 无法达到全局最优解, 故 RMSEP 值和 R_p^2 值变化趋势发生相反的变化, 模型的预测性能会下降。根据训练结果, 对于此 CNN 模型来说, 其学习率在 0.0001 左右时预测效果最好。

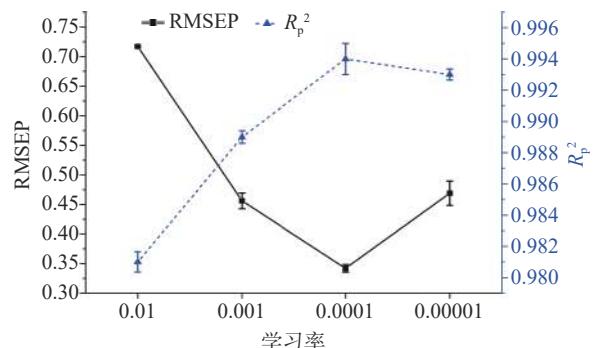


图 6 学习率对水分含量预测模型的影响

Fig.6 Influence of learning rate on moisture content prediction model

2.5.2 迭代次数对模型预测效果的影响 迭代次数是模型训练过程中的重要参数, 对训练效果产生显著影响。适当的迭代次数可以帮助模型更好地理解数据模式, 提高训练精度^[29]; 然而, 过多的迭代次数可能引发过拟合问题, 导致模型在训练集上表现良好但无法泛化到新数据^[31]。

由图 7 可知, 随着迭代次数的增加, 模型可以逐渐减少在训练数据上的预测误差, 从而使 RMSEP 值逐渐减小; 同时, 模型也可以更好地拟合训练数据的目标值, 使得 R_p^2 值逐渐增大。当迭代次数增加到 1200 时, 模型可能过度学习了训练数据中的噪声和随机变动, 而忽略了真实的数据模式和规律, 造成过拟合, 预测性能下降^[31]。

为了平衡模型的拟合能力和泛化能力, 本模型最终确定合适的迭代次数大致为 800 左右。

最终通过反复训练可得, 当学习率为 0.0001, 迭代 800 次后模型训练效果最好, 其预测均方根误差

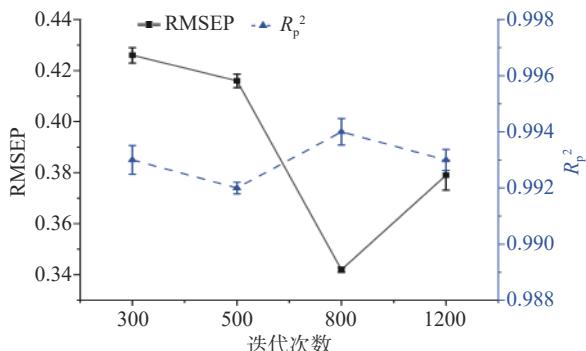


图7 迭代次数对水分含量预测模型的影响

Fig.7 Influence of epochs on moisture content prediction model

RMSEP降低到0.342,决定系数 R_p^2 达到了0.994,说明在全波段下构建CNN水分预测模型可行。

相较于谈爱玲等^[32]基于1D-CNN和串行融合光谱预处理方法所建立的水分预测模型(R_p^2 为0.906)和沈晔^[33]通过复杂的特征波段提取后建立的干贝水分定量分析模型(R_p^2 为0.967)来说,本研究所建立的深度学习模型虽未经过特定的特征波段提取环节,但训练效果并未逊于前者。只经过简单光谱预处理并基于全波段所建立的CNN模型整体拟合度较优,特征选择较为准确。

2.6 外部验证

在对比了不同模型的预测效果后,最后将2023年采集的第五茬紫菜光谱数据输入CNN模型,进行外部验证,验证结果见图8。Jiang等^[34]发现外部验证相对误差值RPD越大则预测模型的效果越好,当 $RPD > 3$ 时则表明模型可以用在生产实践中的检测。经测验,该CNN模型的RPD值为6.155,说明该CNN模型可以在实际使用场景中拥有较好的预测效果,可以实现快速准确地预测条斑紫菜水分含量。

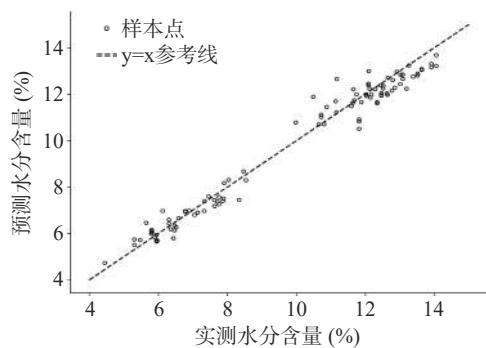


图8 外部样本模型验证结果

Fig.8 Verification results of external sample model

3 结论

本文分别采用多元线性回归、偏最小二乘、支持向量回归、人工神经网络和卷积神经网络基于近红外光谱全波段建立了紫菜水分预测模型。结果表明,在四种传统定量水分预测模型中,SVR模型的预测效果最佳,经过S-G平滑+FD预处理后,其预测误差RMSEP降至0.553, R_p^2 值达到0.980。对比传统

定量水分预测模型和深度学习模型来看,在S-G平滑+SD的预处理方法下所建立的CNN模型预测效果最佳,其RMSEP值为0.456, R_p^2 值为0.990,优化后,该模型的RMSEP值降至0.342, R_p^2 值可以达到0.994,大于0.8,同时外部验证RPD值为6.115,大于3,充分证明了全波段卷积神经网络(CNN)模型实际用于农业和食品工业中的可能性。同时,也可以看出深度学习模型在通过近红外光谱识别干紫菜中的非主要成分,即水分时所展现出的巨大潜力。在实际生产现场中,厂家可在条斑紫菜烘干工艺结束后的生产线上安装内置有该模型的近红外光谱仪,对烘干结束后的紫菜样品实时进行水分监测,从而实现对紫菜品质的高效控制和管理。

此次实验所测量的波长范围较小,故如游离OH组合频波段(1920~1980 nm)^[24]等可能与本次实验相关的波段没有被包含到数据集中。同时,近年来集成学习和融合策略的概念逐渐被应用于近红外光谱分析领域中。因此,之后的研究应当基于更完整的光谱进行建模,并结合集成模型的应用,探讨更精准的预测效果的实现。

© The Author(s) 2024. 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] 曹荣,胡梦月,谭志军,等.基于电子舌和气相色谱-离子迁移谱分析坛紫菜与条斑紫菜的风味特征[J].*食品科学*,2021,42(8):186~191. [CAO R, HU M Y, TAN Z J, et al. Flavor characteristics of *Porphyra haitanensis* and *Porphyra yezoensis*: A comparison study using electronic tongue and gas chromatography-ion mobility spectrometry[J]. *Food Science*, 2021, 42(8): 186~191.]
- [2] 侍江春,肖宇雪,俸贵强,等.不同干条斑紫菜和烤条斑紫菜光合色素含量分析[J].*食品与发酵工业*,2020,46(23):169~177. [SHI J C, XIAO Y X, FENG G Q, et al. Analysis of photosynthetic pigment content of different dried and roasted laver[J]. *Food and Fermentation Industries*, 2020, 46(23): 169~177.]
- [3] 于宁,戴卫平,高永刚,等.中国紫菜国际贸易现状及主要问题研究[J].*中国渔业经济*,2018,36(5):39~45. [YU N, DAI W P, GAO Y G, et al. Analysis and countermeasures of Chinese laver international trade[J]. *Chinese Fisheries Economics*, 2018, 36(5): 39~45.]
- [4] 赵炜.探讨干紫菜质量安全卫生控制体系的建立[J].*食品研究与开发*,2014,35(23):134~137. [ZHAO W. Discussion on the establishment of quality safety and sanitary control system of dried seaweed[J]. *Food Research and Development*, 2014, 35(23): 134~137.]
- [5] 林鹏程,张钟元,江宁,等.紫菜热风/微波联合干燥工艺优化及品质分析[J].*食品工业科技*,2022,43(2):215~225. [LIN P C, ZHANG Z Y, JIANG N, et al. Optimization of two step drying process of *Porphyra* by hot air and microwave and quality evaluation[J]. *Science and Technology of Food Industry*, 2022, 43(2): 215~225.]
- [6] 陈洋,张建,陈炳屹,等.条斑紫菜加工技术研究进展[J].*特种经济动植物*,2022,25(3):113. [CHEN Y, ZHANG J, CHEN B Y, et al. Research progress on processing technology of *Porphyra striata*[J]. *Special Economic Animals and Plants*, 2022, 25(3): 113.]

- [7] 冯智银. 条斑紫菜干紫菜加工技术[J]. *渔业现代化*, 2002(5): 29–30. [FENG Z Y. Processing technology of dried laver [J]. *Fishery Modernization*, 2002(5): 29–30.]
- [8] 国家市场监督管理总局, 国家标准化管理委员会. GB/T 23597-2022 干紫菜质量通则[S]. 2022; 12. [State Administration for Market Regulation, National Standardization Administration. GB/T 23597-2022 General principles of quality of dried laver [S]. 2022; 12.]
- [9] 中华人民共和国国家卫生健康委员会. GB 5009.3-2016 食品安全国家标准食品中水分的测定[S]. 2016; 12. [National Health Commission of the People's Republic of China. GB 5009.3-2016 Determination of moisture in food of National Standard for Food Safety [S]. 2016; 12.]
- [10] 周熙, 秦文. 浅谈食品水分含量快速检测法[J]. 现代测量与实验室管理, 2007(3): 15–17, 21. [ZHOU X, QIN W. Rapid detection of food moisture content [J]. *China Inspection Body & Laboratory*, 2007(3): 15–17, 21.]
- [11] 王联珠, 殷邦忠, 戴卫平, 等. 紫菜国际标准制定对我国紫菜产业的影响[J]. *渔业科学进展*, 2013, 34(6): 143–148. [WANG L Z, YIN B Z, DAI W P, et al. The impact of formulation of international standards for laver products on Chinese laver industry [J]. *Progress in Fishery Sciences*, 2013, 34(6): 143–148.]
- [12] 王会, 卢雅婷. 近红外水分测定仪的发展及在食品水分检测中的应用[J]. *农业工程*, 2016, 6(1): 58–61. [WANG H, LU Y T. Development of near infrared moisture measuring instrument and its application in food moisture detection [J]. *Agricultural Engineering*, 2016, 6(1): 58–61.]
- [13] 刘爽, 柴春祥. 近红外光谱技术在水产品检测中的应用进展[J]. 食品安全质量检测学报, 2021, 12(21): 8590–8596. [LIU S, CHAI C X. Application progress of near infrared spectroscopy in aquatic products detection [J]. *Journal of Food Safety & Quality*, 2021, 12(21): 8590–8596.]
- [14] AFARA I O, PRASADAM I, ARABSHAHI Z, et al. Monitoring osteoarthritis progression using near infrared (NIR) spectroscopy [J]. *Scientific Reports*, 2017, 7(1): 11463.
- [15] GUAN X C, CHEN X J, JIANG J. Integration of modified uninformative variable elimination and successive projections algorithm for determination harvest time of laver by using visible and near infrared spectra [J]. *African Journal of Agricultural Research*, 2011, 6(27): 5987–5991.
- [16] WU D, CHEN X, ZHU X, et al. Uninformative variable elimination for improvement of successive projections algorithm on spectral multivariable selection with different calibration algorithms for the rapid and non-destructive determination of protein content in dried laver [J]. *Analytical Methods*, 2011, 3(8): 1790–1796.
- [17] 方明, 刘静. 基于回归卷积神经网络的近红外光谱苹果脆片品质评价方法研究[J]. *食品科技*, 2020, 45(7): 303–308, 16.
- [18] BENMOUNA B, GARCÍA-MATEOS G, SABZI S, et al. Convolutional neural networks for estimating the ripening state of fuji apples using visible and near-infrared spectroscopy [J]. *Food and Bioprocess Technology*, 2022, 15(10): 2226–2236.
- [19] 门昌骞, 孟晓超, 姜高霞, 等. 一种利用 SPXY 采样的标签噪声主动清洗方法[J]. *小型微型计算机系统*, 2021, 42(9): 1865–1870. [MEN C Q, MENG X C, JIANG G X, et al. Active label noise cleaning method based on SPXY sampling [J]. *Journal of Chinese Computer Systems*, 2021, 42(9): 1865–1870.]
- [20] WANG S F, HAN P, CUI G L, et al. The NIR detection research of soluble solid content in watermelon based on SPXY algorithm [J]. *Spectroscopy and Spectral Analysis*, 2019, 39(3): 738.
- [21] LU M Y, YANG K, SONG P F, et al. The study of classification modeling method for near infrared spectroscopy of tobacco leaves based on convolution neural network [J]. *Spectroscopy and Spectral Analysis*, 2018, 38(12): 3724–3728.
- [22] 陈方方, 丁跃武, 杨友, 等. 近红外光谱样本集划分以及预处理的方法研究[J]. *甘肃科技*, 2023, 39(1): 46–50. [CHEN F F, DING Y W, YANG Y, et al. Research on near infrared spectroscopy sample set partitioning and preprocessing [J]. *Gansu Science and Technology*, 2023, 39(1): 46–50.]
- [23] SAVITZKY A, GOLAY M J E. Smoothing and differentiation of data by simplified least squares procedures [J]. *Analytical Chemistry*, 1964, 36(8): 1627–1639.
- [24] 褚小立, 刘慧颖, 燕泽程. 近红外光谱分析技术实用手册 [M]. 北京: 机械工业出版社, 2016: 956. [CHU X L, LIU H Y, YAN Z C. 近红外光谱分析技术实用手册 [M]. Beijing: China Machine Press, 2016: 956.]
- [25] LI X X, ZHU C G, FU Z T, et al. Rapid detection of soil moisture content based on UAV multispectral image [J]. *Spectroscopy and Spectral Analysis*, 2020, 40(4): 1238.
- [26] WILLIAMS P C, NORRIS K H. Near-infrared technology in the agricultural and food industries [J]. *Starch-Stärke*, 1988, 40(10): 408.
- [27] LIU D, WANG E, WANG G, et al. Nondestructive determination of soluble solids content, firmness, and moisture content of "Longxiang" pears during maturation using near-infrared spectroscopy [J]. *Journal of Food Processing and Preservation*, 2022, 46(3): e16332.
- [28] SHIRISH KESKAR N, MUDIGERE D, NOCEDAL J, et al. On large-batch training for deep learning: Generalization gap and sharp minima [J/OL]. 2016, arXiv:1609.04836 https://ui.adsabs.harvard.edu/abs/2016arXiv160904836S. 10.48550/arXiv.1609.04836.
- [29] HEATON J. Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning; The MIT Press, 2016, 800 pp, ISBN: 0262035618 [J]. *Genetic Programming and Evolvable Machines*, 2017, 19: 305–307.
- [30] BOUTAMI S, ZHAO N, FAN S. Determining the optimal learning rate in gradient-based electromagnetic optimization using the Shanks transformation in the Lippmann-Schwinger formalism [J]. *Optics Letters*, 2020, 45(3): 595–598.
- [31] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: A simple way to prevent neural networks from overfitting [J]. *Journal of Machine Learning Research*, 2014, 15(1): 1929–1958.
- [32] 谈爱玲, 王晓斯, 楚振原, 等. 基于近红外光谱融合与深度学习的玉米成分定量建模方法[J]. *食品与发酵工业*, 2020, 46(23): 213–219. [TAN A L, WANG X S, CHU Z Y, et al. Research on quantitative modeling method of maize composition based on near infrared spectrum fusion and deep learning [J]. *Food and Fermentation Industries*, 2020, 46(23): 213–219.]
- [33] 沈晔. 基于高光谱成像技术的干贝水分含量快速检测研究 [D]. 杭州: 浙江大学, 2017. [SHEN Y. Rapid determination of the moisture content in dried scallop based on hyperspectral imaging [D]. Hangzhou: Zhejiang University, 2017.]
- [34] JIANG H, ZHOU Y, ZHANG C, et al. Evaluation of dual-band near-infrared spectroscopy and chemometric analysis for rapid quantification of multi-quality parameters of soy sauce stewed meat [J]. *Foods*, 2023, 12(15): 2882.