

基于 Stacking 框架的蓝莓硬度预测可溶性固形物含量及维生素 C 模型构建

付晟宏, 朱玉杰*, 冯国红*

(东北林业大学工程技术学院, 哈尔滨 150040)

摘 要: 目的 探索一种基于蓝莓的质量和硬度快速、无损预测其可溶性固形物含量(soluble solid content, SSC)与维生素 C (vitamin C, VC)含量的方法, 为蓝莓化学成分的预测提供一种新思路。**方法** 通过对蓝莓质量、硬度与 SSC、VC 相关性分析后, 建立基于一维特征的质量和硬度预测 SSC 与 VC 模型。其次, 对硬度添加多项式特征做升维处理, 同一维进行相同研究。最后, 对比 Stacking 框架与单一模型、及添加多项式特征的预测效果。**结果** 一维特征条件下, 基于 Stacking 框架的硬度预测 SSC 与 VC 的决定系数(coefficient of determination, R^2)分别为 0.873、0.875, 预测效果优于质量与单模型预测; 多维特征条件下, 硬度添加到 3 次方时预测 SSC 效果最佳, R^2 为 0.889; 添加到 12 次方时预测 VC 含量效果最佳, R^2 为 0.890, 预测效果均好于一维特征。**结论** Stacking 框架结合添加多项式特征在蓝莓硬度快速、无损预测其 SSC 及 VC 方面具有良好的潜力, 为蓝莓品质检测提供新途径。

关键词: Stacking 框架; 添加多项式特征; 蓝莓; 理化指标

Construction of blueberry hardness prediction soluble solid content and vitamin C model based on Stacking framework

FU Sheng-Hong, ZHU Yu-Jie*, FENG Guo-Hong*

(College of Engineering and Technology, Northeast Forestry University, Harbin 150040, China)

ABSTRACT: Objective To explore a fast and nondestructive method to predict the soluble solid content (SSC) and vitamin C (VC) content of blueberry based on its mass and hardness, and provide a new idea for the prediction of chemical composition of blueberry. **Methods** Through the correlation analysis of blueberry mass, hardness with the SSC and VC, a prediction model of the SSC and VC based on one-dimensional features of mass and hardness was established. Secondly, polynomial features were added to the hardness to do the dimensional lift and the same study was carried out in the same dimension. Finally, the prediction effects of Stacking framework with a single model and

基金项目: 黑龙江省自然科学基金项目(LH2020C050)、中央高校基金科研业务费专项基金项目(2572020BL01)

Fund: Supported by the Science Foundation of Heilongjiang Province (LH2020C050), and the Special Fund for Scientific Research Business Expenses of the Central University Fund (2572020BL01)

***通信作者:** 朱玉杰, 博士, 教授, 主要研究方向为工业工程。E-mail: zhuyujie004@126.com

冯国红, 博士, 副教授, 主要研究方向为应用近红外光谱从事蓝莓主要营养成分的模型构建和木材种类的识别研究。E-mail: fgh_1980@126.com

***Corresponding author:** ZHU Yu-Jie, Ph.D, Professor, Northeast Forestry University, No.26 Hexing Road, Xiangfang District, Harbin 150040, China. E-mail: zhuyujie004@126.com

FENG Guo-Hong, Ph.D, Associate Professor, Northeast Forestry University, No.26 Hexing Road, Xiangfang District, Harbin 150040, China. E-mail: fgh_1980@126.com

the addition of polynomial features were compared. **Results** Under one-dimensional feature condition, coefficient of determination (R^2) of the hardness prediction of SSC and VC based on Stacking framework was 0.873 and 0.875, respectively, which was better than the prediction of mass and single model; under the condition of multi-dimensional features, when hardness was added to the 3rd power, it was the best to predict the SSC, R^2 was 0.889; when hardness was added to the 12th power, it was the best to predict the content of VC, R^2 was 0.890, which was better than one-dimensional features. **Conclusion** The Stacking framework combined with the addition of polynomial features has good potential for rapid and nondestructive prediction SSC and VC based on blueberry hardness, which provides a new way for blueberry quality detection.

KEY WORDS: Stacking framework; adding polynomial features; blueberry; physical and chemical indicators

0 引言

蓝莓果实营养丰富, 富含花青素、维生素 C (vitamin C, VC)、微量元素及可溶性固形物(soluble solid content, SSC)等多种化学成分^[1-2], 具有抗癌、保护视力、美白皮肤、延缓神经衰老等保健功能^[3-4], 但其品质也易受到温度、自身代谢、机械损伤、微生物侵染^[5-6]等影响。随着贮藏时间的增加, 蓝莓硬度逐渐下降, 质量损失, 导致 SSC、VC 等化学成分含量下降, 直接影响蓝莓的功效与货架期^[7-9]。因此, 及时监测蓝莓的理化指标具有重要的应用价值。

目前, 蓝莓 SSC 的检测主要采用折光仪实现; VC 的检测主要采用化学方法; 而折光仪对环境适应能力较弱, 化学方法存在检测时间长、流程烦琐等问题。针对这些问题, 学者们基于光谱分析具有分析高效、无损检测等优点, 将光谱分析广泛应用到蓝莓理化指标的检测上, 并取得了一定成果^[10-12]。LEIVA-VALENZUELA 等^[13]利用高光谱结合偏最小二乘法实现了蓝莓硬度与 SSC 的检测。朱金艳等^[14]通过光谱技术与极限学习机检测了蓝莓的 SSC、VC 等。BAI 等^[15]通过最小二乘回归采用近红外光谱与液相色谱测定了蓝莓总糖、总酸及花青素含量。虽然近红外光谱分析法较为热门, 但其分析结果对模型的建立有较高的要求^[16], 而高光谱对于光源平衡的问题还有待解决^[17], 且二者均较为昂贵。

Stacking 框架是一种将多种学习器组合起来的模型, 与传统机器学习模型相比, 具有更好的稳定性^[18], 已广泛应用在各个领域^[19-24]。CUI 等^[25]利用 Stacking 框架实现了对汽油辛烷值的预测。MASHHADI 等^[26]结合深度学习将 Stacking 框架应用到卡车的预测维护, 取得了优越的性能。WANG 等^[27]对比了 Stacking 框架与 Bagging、Boosting 集成在信用评估上的预测精度。这些研究虽然取得了良好的效果, 但对于如何选择学习器, 不同原理的学习器组合起来是否会降低精度等问题没有做过多的探索。

蓝莓物理指标的检测较易实现且检测准确性高, 如电子天平对质量的测定, 质构仪对硬度的测定等; 本研究充分考虑了 Stacking 框架中模型差异性与多样性, 基于

Stacking 框架对蓝莓质量、硬度与 SSC、VC 的关系进行了探索, 构建了最优模型, 为蓝莓贮藏品质快速定量检测提供参考。

1 材料与方法

1.1 实验材料、试剂与仪器

绿宝石, 2021 年 6 月购置于辽宁丹东。将表面无损伤、果实饱满、大小成熟度相同的蓝莓分组, 每 6 个一组, 共 300 组, 装于专用保鲜盒, 放置于 10℃的恒温箱内进行贮藏。

抗坏血酸、盐酸(分析纯, 天津市富宇精细化工有限公司)。

LYT-330 型手持式折光仪(上海淋誉公司); Universal TA 型质构仪(上海腾拔公司); UV-1801 型紫外分光光度计(北京北分瑞利公司); PX-70BIII 型生化培养恒温箱(天津泰斯特公司); LT202C 型电子天平(精度 0.01 g, 常熟市天量仪器有限公司)。

1.2 实验方法

从贮藏当天起, 每天对每组蓝莓质量值、硬度值、SSC 及 VC 含量各测定一次, 共测 10 d。每次实验前取 30 组于室温下测定。

1.2.1 质量值的测定

质量值采用电子天平测定每组蓝莓单果质量并取平均作为该组质量值。

1.2.2 硬度值的测定

硬度值采用质构仪柱形探头 P5 以测前、测试和测后速度分别为 2.00、1.00、2.00 mm/s, 接触力为 0.5 N, 下压距离为 7.0 mm, 进行测量, 测试时以第 1 次下压曲线上最大力量值作为单次硬度值, 以每组平均值作为该组硬度值。

1.2.3 SSC 的测定

测定 SSC 时, 取一组蓝莓研捣均匀, 用纱布包裹挤汁于折光仪进行测定, 重复 3 次取平均值作为该组 SSC 的含量。

1.2.4 VC 含量的测定

测定 VC 含量时, 取一定量上述研捣后的蓝莓碎渣于容量瓶, 加入 2 mL 1%盐酸混合后, 加蒸馏水定容至 25 mL,

用滤纸滤出滤液 2 mL 后, 加 0.2 mL 10% 盐酸混合, 再加水稀释至 10 mL。以蒸馏水作为空白对照, 借助紫外分光光度计在 243 nm 处测得吸光度值(C), 并查标准曲线($C=15.12A+0.1222$, $R^2=0.9833$, 由抗坏血酸标准液测得), 计算出该组 VC 的含量。

1.3 数据预处理

在 SPSS 22 软件中进行相关性分析, 通过标准分数法对数据进行异常值分析, 采用直方图对 SSC 和 VC 进行正态分布分析。其中, 直方图分布采用 Origin 2021 进行绘图。

建立模型前, 对数据进行了标准化和卷积平滑(savitzky-golay, SG)处理, 其中 SG 处理采用了“mirror”进行末端补齐。本研究还采用了添加多项式特征的方法将一维特征处理成多维特征, 实现特征升维, 即将特征 X (本研究中对质量、硬度)进行乘方, 如 X^2 、 X^3 、 X^4 等, 从而使模型的输入由特征 X 转变为如 $X+X^2+X^3+X^4$ 的多维特征的形式。

1.4 Stacking 框架建立与评估

本研究中 Stacking 框架由第一层的多种基学习器与第二层的元学习器构成。本研究在搭建 Stacking 框架时, 从模型的多样性与差异性的角度出发^[28], 考虑了如下模型: 随机森林算法(random forest, RF)、自适应增强算法(adaptive boosting, ADA)、梯度提升回归树算法(gradient boosting regression tree, GBRT)、极端梯度提升树算法(extreme gradient boosting, XGB)、轻量梯度提升机算法(light gradient boosting machine, LGBM)、支持向量回归(support vector regression, SVR)、多层感知机(multi-layer perceptron, MLP)、K-近邻(k-nearest neighbor, KNN)等, 以及已被证实是一种提升模型有效性的线性回归模型(linear regression, LR)^[29-30]。其中, 模型构建基于 Python 3.8。

本研究采用决定系数(coefficient of determination, R^2)、均方根误差(root mean squared error, RMSE)、平均绝对误差(mean absolute error, MAE) 3 种指标来评价模型的精度。

2 结果与分析

2.1 异常值分析结果

由图 1 与图 2 可知, VC 和 SSC 的 300 个样本基本呈正态分布; VC 含量平均值为 35.0209 mg/100 g, 标准差为 5.7870 mg/100 g; SSC 平均值为 13.3479%, 标准差为 1.8569%。由切比雪夫定理可知, 至少 89% 的数据值与平均值的距离在 3 个标准差之内, 故将标准分数大于阈值 ± 3 (即计算后超出 3 个标准差)的测量值作为异常值。本研究在计算理化指标的标准分数后, 对标准分数结果取绝对值降序排序后以前 5 行为例进行说明, 结果见表 1。由表 1 可知, 质量值有一个异常值, 硬度值处于正常状态, SSC 有 3 个异

常值, VC 含量有 2 个异常值。同时, 由于均值和标准差对异常值都很敏感, 可能会导致少量数据发生异常, 同时样本数据符合正态分布, 且异常值数量相对样本数量可以忽略不计, 故不对异常值进行剔除。

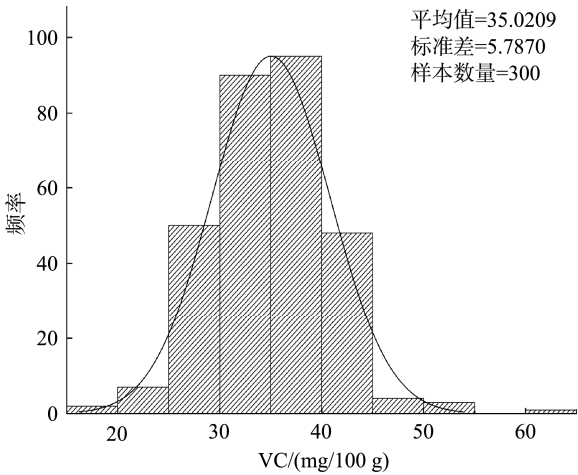


图 1 VC 直方图分布
Fig.1 VC histogram distribution

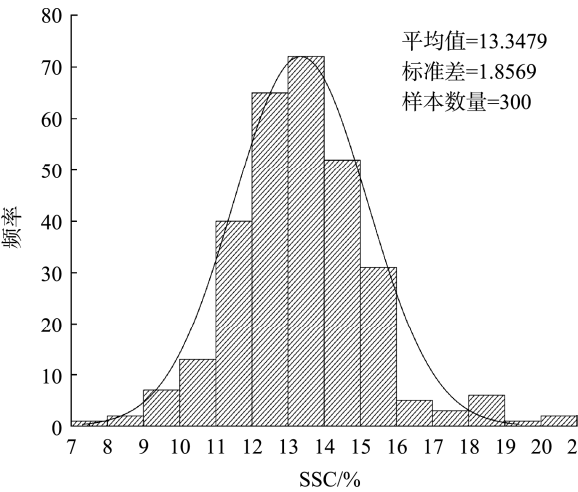


图 2 SSC 直方图分布
Fig.2 SSC histogram distribution

表 1 理化指标标准分数计算结果

Table 1 Physical and chemical indicator results of standard score calculations			
质量值	硬度值	SSC	VC
3.2923 (70)	2.7107	4.0132 (281)	4.5866 (10)
2.1905	2.7107	3.8516 (270)	3.0540 (26)
2.1497	2.2051	3.1336 (238)	2.9633
1.9457	2.1964	2.8464	2.7521
1.7824	2.0063	2.7746	2.3906

注: 表中()内数值为测量值标号, 70 即代表第 70 个测量值。

2.2 一维特征预测结果

2.2.1 一维特征相关性分析结果

对测定的蓝莓质量值、硬度值分别与 SSC、VC 进行皮尔森相关性分析。由分析结果可知,质量值与 SSC 之间呈极显著正相关,相关系数为 0.334 ($P<0.01$);与 VC 之间呈极显著负相关,相关系数为-0.308 ($P<0.01$)。硬度值与 SSC 之间呈极显著正相关,相关系数为 0.704 ($P<0.01$);与 VC 之间呈极显著负相关,相关系数为-0.639 ($P<0.01$)。本研究分别以质量和硬度建立 SSC、VC 的预测模型。

2.2.2 模型参数设置

利用网格搜索对参数较多的模型进行参数优化,参数较少的模型采用手动调参,以提高训练速度与精度。经调试后,一维特征条件下各模型参数选择见表 2。

表 2 基于质量(硬度)预测模型的各学习器参数
Table 2 Parameters of each learner based on mass (hardness) prediction model

学习器	主要参数设置
ADA	learning_rate=0.3, n_estimators=25(30)
LGBM	learning_rate=0.5, num_iterations=300 max_depth=3(5), min_child_samples=10
XGB	learning_rate=0.5, n_estimators=50, max_depth=10
GBRT	learning_rate=0.3, n_estimators=200 max_depth=5
SVR	c=100, epsilon=0.1, gamma=30
MLP	hidden_layer_sizes=13(3), activation='relu' learning_rate_init=0.001, solver='adam'
RF	n_estimators=150, max_depth=15
KNN	n_neighbors=4(5)

2.2.3 学习器的选择

选取预测结果之间相关性较小且欧式距离较大的模型作为基学习器。基学习器的选取过程以一维硬度预测 VC 的模型为例说明。其中,相关性分析以热力图呈现,相

关性与欧式距离分析分别见图 3 与表 3。

基学习器选取完成后,将集成好的第一层模型分别与各模型进行组合训练,综合平均决定系数与平均均方根误差后,完成对元学习器的选取与第二层模型的构建及整个 Stacking 框架的搭建。不同条件下的学习器选取结果见表 4。

2.2.4 Stacking 框架预测结果

在 2.2.2 和 2.2.3 的 Stacking 框架的设计下,本研究以实验获取的数据集为输入,分别基于质量和硬度对 SSC、VC 进行预测。其中,数据集采用训练集:测试集为 4:1 的比例进行划分。

预测结果如下。以硬度预测 SSC 时,Stacking 框架得到的 R^2 为 0.873,比质量预测(R^2 为 0.207)时提高了 0.666;RMSE 为 0.250,比质量预测(RMSE 为 0.623)时降低了 0.373;MAE 为 0.194,比质量预测(MAE 为 0.505)时降低了 0.311。

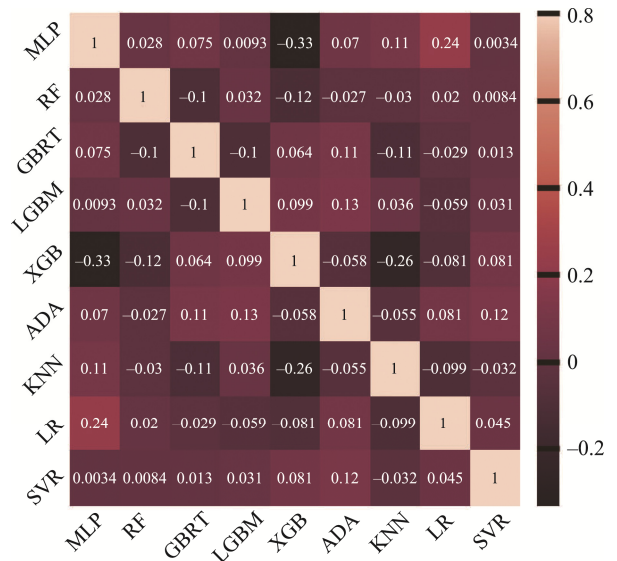


图 3 硬度预测 VC 模型相关性热力图

Fig.3 Heat map of model correlations for hardness prediction of VC

表 3 硬度预测 VC 模型欧式距离分析
Table 3 Hardness prediction VC model euclidean distance analysis

	值向量之间的相关性								
	MLP	RF	GBRT	LGBM	XGB	ADA	KNN	LR	SVR
MLP	0.000								
RF	2.236	0.000							
GBRT	2.214	2.599	0.000						
LGBM	2.351	2.508	2.689	0.000					
XGB	2.560	2.556	2.350	2.382	0.000				
ADA	2.053	2.364	2.220	2.262	2.355	0.000			
KNN	2.262	2.587	2.638	2.591	2.800	2.491	0.000		
LR	2.169	2.605	2.652	2.793	2.686	2.414	2.798	0.000	
SVR	2.344	2.520	2.512	2.576	2.385	2.260	2.632	2.622	0.000

表 4 不同条件下学习器选取结果
Table 4 Results of learner selection under different conditions

条件	基学习器选取结果	元学习器选取结果
质量预测 SSC 模型	GBRT+KNN+LGBM+RF+XGB	SVR
质量预测 VC 模型	GBRT+KNN+RF+SVR+XGB	ADA
硬度预测 SSC 模型	ADA+GBRT+MLP+SVR+XGB	SVR
硬度预测 VC 模型	GBRT+KNN+LR+RF+XGB	SVR

以硬度预测 VC 含量时, Stacking 框架得到的 R^2 为 0.875, 比质量预测(R^2 为 0.234)时提高了 0.641; RMSE 为 0.263, 比质量预测(RMSE 为 0.661)时降低了 0.398; MAE 为 0.199, 比质量预测(MAE 为 0.557)时降低了 0.358。

2.3 多维特征预测结果

由 2.2.4 的结果可知, 硬度对于 SSC、VC 的预测明显优于质量, 故仅针对硬度添加多项式特征进行预测研究。

2.3.1 多维特征相关性分析结果

对硬度做 X^2 、 X^3 、....等多项式特征处理后, 再分别与 SSC 和 VC 进行相关性分析, 结果如表 5 所示。

通过表 5 可知, 随着维度的升高, 硬度与 SSC 相关性

呈短暂下降趋势, 在 X^4 至 X^{10} 期间, 相关性逐渐上升, 此时仍处于强相关状态, 从 X^{11} 开始, 相关性虽呈继续上升趋势, 但已由强相关变为弱相关, 最后直到不相关; 硬度与 VC 含量相关性先呈现下降趋势, 到 X^{12} 时相关性小幅度提升, 此时仍处于强相关状态, 从 X^{13} 开始相关性再次下降, 由弱相关直到不相关。

2.3.2 Stacking 框架预测结果

模型参数设置方式同 2.2.2, 模型组合过程同 2.2.3, 经调试后, 添加多项式特征时硬度对 SSC 与 VC 的预测结果见表 6。

通过表 6 可知, 当硬度添加到 3 次方时, 对 SSC 的预测效果最佳, R^2 为 0.889, 比添加到其他维度时平均提高了 0.015; RMSE 为 0.232, 比添加到其他维度时平均降低了 0.016; MAE 为 0.181, 比添加到其他维度时平均降低了 0.013。

当硬度添加到 12 次方时, 对 VC 含量预测的效果最佳, R^2 为 0.890, 比添加到其他维度时平均提高了 0.012; RMSE 为 0.250, 比添加到其他维度时平均降低了 0.013; MAE 为 0.191, 比添加到其他维度时平均降低了 0.017。

2.4 对比分析

2.4.1 Stacking 框架与单一模型对比分析

在一维特征条件下, 基于硬度的预测结果作单一模型与 Stacking 框架的对比分析, 对比结果见表 7。

表 5 多项式特征处理后硬度与 SSC、VC 含量相关性变化表
Table 5 Correlation change between hardness and SSC, VC content after polynomial feature processing

维度		X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}
指标	SSC	-0.639	-0.640	-0.641	-0.640	-0.639	-0.637	-0.632	-0.631	-0.621	-0.608	-0.581	-0.570	-0.353	-0.075	0.000
	VC	0.704	0.703	0.702	0.699	0.696	0.691	0.686	0.683	0.672	0.661	0.609	0.624	0.344	0.093	0.000

表 6 添加多项式特征时预测结果
Table 6 Prediction results when adding polynomial features

维度	预测 SSC 结果			预测 VC 结果		
	R^2	RMSE	MAE	R^2	RMSE	MAE
X^2	0.875	0.244	0.191	0.881	0.261	0.204
X^3	0.889	0.232	0.181	0.878	0.267	0.206
X^4	0.872	0.249	0.194	0.879	0.264	0.200
X^5	0.883	0.241	0.192	0.872	0.272	0.209
X^6	0.871	0.245	0.194	0.871	0.262	0.210
X^7	0.877	0.244	0.187	0.876	0.266	0.209
X^8	0.874	0.248	0.196	0.880	0.261	0.193
X^9	0.871	0.243	0.189	0.881	0.262	0.201
X^{10}	0.883	0.258	0.198	0.877	0.265	0.202
X^{11}	0.866	0.255	0.197	0.885	0.262	0.203
X^{12}	0.877	0.249	0.197	0.890	0.250	0.191
X^{13}	0.878	0.245	0.196	0.881	0.260	0.201
X^{14}	0.860	0.254	0.198	0.881	0.257	0.201

表 7 单一模型与 Stacking 框架对比结果
Table 7 Comparison results of single model and Stacking framework

硬度预测 SSC 结果				硬度预测 VC 结果			
模型	R ²	RMSE	MAE	模型	R ²	RMSE	MAE
ADA	0.871	0.253	0.193	LR	0.856	0.286	0.228
SVR	0.816	0.290	0.228	KNN	0.850	0.283	0.221
GBRT	0.851	0.268	0.207	RF	0.865	0.278	0.202
MLP	0.795	0.319	0.254	GBRT	0.864	0.276	0.201
XGB	0.849	0.266	0.208	XGB	0.861	0.273	0.200
Stacking 框架	0.873	0.250	0.194	Stacking 框架	0.875	0.263	0.199

通过表 7 可知,以硬度预测 SSC 时,单一模型 ADA 表现较优,MLP 表现较差;以硬度预测 VC 时候,单一模型 RF、GBRT 表现较优,KNN 表现一般。即使在预测效果优劣均有的条件下,Stacking 框架依旧取得了最佳预测效果;一方面说明,Stacking 框架可以发挥预测效果较好模型的优势;另一方面说明,Stacking 框架具有良好的协同性,使整个融合具有较强的稳定性。以硬度预测 SSC 时,Stacking 框架得到的 R^2 为 0.873,比单一模型预测平均提高了 0.034;RMSE 为 0.250,比单一模型预测平均降低了 0.029;MAE 为 0.194,比单一模型预测平均降低了 0.024。以硬度预测 VC 时,Stacking 框架得到的 R^2 为 0.875,比单一模型预测平均提高了 0.016;RMSE 为 0.263,比单一模型预测平均降低了 0.016;MAE 为 0.199,比单一模型预测平均降低了 0.012。

2.4.2 一维特征与多维特征对比分析

通过表 6 可知,硬度对于 SSC 的预测,添加到 3 次方时,模型已达到最优,过多的特征添加反而使模型精度下降;对于 VC 的预测,添加到 12 次方时,模型效果最佳。结合多维特征相关性分析可知,模型达到最优效果时均在相关性发生波动的情况下。

在考虑参数对模型精度影响的前提下,分别对一维与添加多项式特征的模型参数进行寻优,再进行预测后结果的对比。综合一维特征与多维特征预测结果可知,当硬度预测 SSC 时,最佳添加多项式特征的 Stacking 框架得到的 R^2 为 0.889,比一维特征提高了 0.016, RMSE 为 0.232,比一维特征降低了 0.018, MAE 为 0.181,比一维特征降低了 0.013;当硬度预测 VC 时,最佳添加多项式特征的 Stacking 框架得到的 R^2 为 0.890,比一维特征提高了 0.015, RMSE 为 0.250,比一维特征降低了 0.013, MAE 为 0.191,比一维特征降低了 0.008。

3 结 论

本研究针对蓝莓品质检测的现状,提出了一种基于 Stacking 框架的蓝莓理化指标间的预测模型。通过对质量、硬度与 SSC、VC 进行相关性分析,建立了基于一维特征的质量、硬度对 SSC 与 VC 预测的 Stacking 框架模型。结果

表明,硬度对 SSC 与 VC 的预测效果远远好于质量,Stacking 框架预测精度也高于单一模型。在添加多项式特征且模型参数重新寻优后,预测效果得到了进一步提升,当以硬度预测 SSC 时,添加到 3 次方时,Stacking 框架获得的最佳 R^2 为 0.889, RMSE 为 0.232, MAE 为 0.181;当以硬度预测 VC 时,添加到 12 次方时,Stacking 框架获得的最佳 R^2 为 0.890, RMSE 为 0.250, MAE 为 0.191。综上,Stacking 框架结合添加多项式特征所建立的模型具有良好的预测潜力;在实际应用方面,本研究所建立的理化指标间预测模型高效便捷,为预测蓝莓 SSC 与 VC 提供了新途径。

参考文献

[1] 王姗姗,孙爱东,李淑燕. 蓝莓的保健功能及其开发应用[J]. 中国食物与营养, 2010, (6): 17-20.
WANG SS, SUN AID, LI SY. Health function of blueberry and its development and application [J]. Food Nutr China, 2010, (6): 17-20.

[2] 卜庆雁,周晏起. 浅析蓝莓的营养保健功能及开发利用前景[J]. 北方园艺, 2010, (8): 215-217.
BU QY, ZHOU YQ. Analysis of the nutritional health function and development and utilization prospect of blueberry [J]. North Hortic, 2010, (8): 215-217.

[3] 金永学,冯建国,郑淘,等. 蓝莓的功能成分、保健作用及其开发利用[J]. 食品与机械, 2020, 36(5): 231-236.
JIN YX, FENG JG, ZHENG T, et al. Functional components, health function and utilization of blueberry [J]. Food Mach, 2020, 36(5): 231-236.

[4] 贺强,吴立仁. 蓝莓果实中营养成分的生物学功能[J]. 北方园艺, 2010, (24): 222-224.
HE Q, WU LR. Discussions on the biological function of nutrients in blueberry fruit [J]. North Hortic, 2010, (24): 222-224.

[5] 杨丽. 蓝莓采后品质与综合保鲜技术[J]. 现代园艺, 2018, (18): 32-33.
YANG L. Postharvest quality and comprehensive preservation technology of blueberry [J]. Contemp Hortic, 2018, (18): 32-33.

[6] 郑秀艳,孟繁博,黄道梅,等. 蓝莓采后贮藏保鲜技术研究进展[J]. 食品安全质量检测学报, 2016, 7(9): 3560-3565.
ZHENG XY, MENG FB, HUANG DM, et al. Research advances on technologies for postharvest storage and preservation of blueberry [J]. J Food Saf Qual, 2016, 7(9): 3560-3565.

[7] CAO XH, ZHANG FF, ZHAO DY, et al. Effects of freezing conditions on quality changes in blueberries [J]. J Sci Food Agric, 2018, 98(12): 4673-4679.

[8] PARK B, SHIN TS, CHO JS, et al. Characterizing hyperspectral

- microscope imagery for classification of blueberry firmness with deep learning methods [J]. *Agronomy-Basel*, 2022, 12(1): 85.
- [9] K TENIOUDAKI A, O'DONNELL CP, EMOND JP, *et al.* Blueberry supply chain: Critical steps impacting fruit quality and application of a boosted regression tree model to predict weight loss [J]. *Postharvest Biol Technol*, 2021, 179(4): 111590.
- [10] 王多加, 周向阳, 金同铭, 等. 近红外光谱检测技术在农业和食品分析上的应用[J]. *光谱学与光谱分析*, 2004, 24(4): 447-450.
- WANG DJ, ZHOU XY, JIN TM, *et al.* Application of near-infrared spectroscopy to agriculture and food analysis [J]. *Spectrosc Spect Anal*, 2004, 24(4): 447-450.
- [11] 张保华, 李江波, 樊书祥, 等. 高光谱成像技术在果蔬品质与安全无损检测中的原理及应用[J]. *光谱学与光谱分析*, 2014, 34(10): 2743-2751.
- ZHANG BH, LI JB, FAN SX, *et al.* Principles and applications of hyperspectral imaging technique in quality and safety inspection of fruits and vegetables [J]. *Spectrosc Spect Anal*, 2014, 34(10): 2743-2751.
- [12] 彭彦颖, 孙旭东, 刘燕德. 果蔬品质高光谱成像无损检测研究进展[J]. *激光与红外*, 2010, 40(6): 586-592.
- PENG YY, SUN XD, LIU YD. Research progress of hyperspectral imaging in nondestructive detection of fruits and vegetables quality [J]. *Laser Infrared*, 2010, 40(6): 586-592.
- [13] LEIVA-VALENZUELA GA, LU RF, AGUILERA JM. Prediction of firmness and soluble solids content of blueberries using hyperspectral reflectance imaging [J]. *J Food Eng*, 2013, 115(1): 91-98.
- [14] 朱金艳, 朱玉杰, 冯国红, 等. 基于近红外光谱技术联合极限学习机的蓝莓贮藏品质定量模型建立[J]. *食品与发酵工业*, 2022, 48(16): 270-276.
- ZHU JY, ZHU YJ, FENG GH, *et al.* Establishment of quantitative models for blueberry storage quality based on near infrared spectroscopy combined with extreme learning machine [J]. *Food Ferment Ind*, 2022, 48(16): 270-276.
- [15] BAI WM, YOSHIMURA N, TAKAYANAGI M, *et al.* Construction of models for nondestructive prediction of ingredient contents in blueberries by near-infrared spectroscopy based on HPLC measurements [J]. *Jove-J Vis Exp*, 2016, 1(112): 53981.
- [16] 徐广通, 袁洪福, 陆婉珍. 现代近红外光谱技术及应用进展[J]. *光谱学与光谱分析*, 2000, (2): 134-142.
- XU GT, YUAN HF, LU WZ. Modern near-infrared spectroscopy technology and its application progress [J]. *Spectrosc Spect Anal*, 2000, (2): 134-142.
- [17] 申亚其, 李松林, 何杰, 等. 高光谱图像技术在水果品质无损检测中的研究进展[J]. *林业机械与木工设备*, 2021, 49(3): 4-9.
- SHEN YQ, LI SL, HE J, *et al.* Research progress of hyperspectral image technology in nondestructive testing of fruit quality [J]. *For Mach Woodwork Equip*, 2021, 49(3): 4-9.
- [18] 韩腾飞, 李亚平. 基于 Stacking 集成学习的剩余使用寿命预测[J/OL]. *计算机集成制造系统*: 1-18. [2022-10-19]. <http://kns.cnki.net/kcms/detail/11.5946.TP.20220314.1224.016.html>
- HAN TF, LI YP. Remaining useful life prediction based on Stacking ensemble learning [J/OL]. *Comput Integr Manuf Syst*: 1-18. [2022-10-19]. <http://kns.cnki.net/kcms/detail/11.5946.TP.20220314.1224.016.html>
- [19] ANIFOWOSE F, LABADIN J, ABDULRAHEEM A. Improving the prediction of petroleum reservoir characterization with a stacked generalization ensemble model of support vector machines [J]. *Appl Soft Comput*, 2015, 26: 483-496.
- [20] WU C, HWANG M, HUANG TH, *et al.* Application of artificial intelligence ensemble learning model in early prediction of atrial fibrillation [J]. *Bmc Bioinform*, 2021, 22(5): 1-12.
- [21] NTI IK, ADEKOYA AF, WEYORI BA. A comprehensive evaluation of ensemble learning for stock-market prediction [J]. *Big Data*, 2022, 7(1): 1-40.
- [22] BUYRUKOGLU S, SAVAS S. Stacked-based ensemble machine learning model for positioning footballer [J]. *Arab J Sci Eng*, 2020, 71(4): 360-365.
- [23] DUMANCAS G, ADRIANTO I. A stacked regression ensemble approach for the quantitative determination of biomass feedstock compositions using near-infrared spectroscopy [J]. *Spectrochim Acta A*, 2022, 276: 121231.
- [24] JIANG ZW, YANG SH, LIU ZY, *et al.* Can ensemble machine learning be used to predict the groundwater level dynamics of farmland under future climate: A 10-year study on Huaibei Plain [J]. *Environ Sci Pollut Res*, 2022, 29(29): 44653-44667.
- [25] CUI SZ, QIU HX, WANG ST, *et al.* Two-stage stacking heterogeneous ensemble learning method for gasoline octane number loss prediction [J]. *Appl Soft Comput*, 2021, 113(Part B): 107989.
- [26] MASHHADI PS, NOWACZYK S, PASHAMI S. Stacked ensemble of recurrent neural networks for predicting remaining useful life of a turbocharger [J]. *Appl Sci-Basel*, 2020, 10(1): 69.
- [27] WANG G, HAO JX, MA JA, *et al.* A comparative assessment of ensemble learning for credit scoring [J]. *Expert Syst Appl*, 2011, 38(1): 223-230.
- [28] AGARWAL S, CHOWDARY CR. A-stacking and a-bagging: Adaptive versions of ensemble learning algorithms for spoof fingerprint detection [J]. *Expert Syst Appl*, 2019, 146(5): 113160.
- [29] KSHATRI SS, SINGH D, NARAIN B, *et al.* An empirical analysis of machine learning algorithms for crime prediction using stacked generalization: An ensemble approach [J]. *Ieee Access*, 2021, 9: 67488-67500.
- [30] ROONEY N, PATTERSON D, NUGENT C. Non-strict heterogeneous Stacking [J]. *Pattern Recogn Lett*, 2007, 28(9): 1050-1061.

(责任编辑: 张晓寒 于梦娇)

作者简介



付晟宏, 硕士研究生, 主要研究方向为蓝莓化学成分预测模型的构建。
E-mail: fushenghong1998@163.com



朱玉杰, 博士, 教授, 主要研究方向为工业工程。
E-mail: zhuyujie004@126.com

冯国红, 博士, 副教授, 主要研究方向为应用近红外光谱从事蓝莓主要营养成分的模型构建和木材种类的识别研究。
E-mail: fgh_1980@126.com