基于近红外光谱技术结合改进的 CS-BPNN 樱桃番茄 SSC 和 Vc 含量检测

康明月^{1,2},罗斌²,周亚男²,王成²,孙鸿雁^{1*}

(1.中国地质大学 (北京) 数理学院, 北京 100000)

(2.北京市农林科学院智能装备技术研究中心,信息技术研究中心,北京 100097)

摘要:为实现精确预测樱桃番茄中 SSC 和 Vc 含量,该研究提出一种改进杜鹃鸟搜索算法优化的 BP 神经网络(Back Propagation Neural Network Optimized by Improved Cuckoo Search Algorithm, ICS-BPNN)模型。采集样品在1350~1800 nm 的近红外光谱数据, 首先采用不同方法进行预处理; 然后利用稳定性竞争性自适应重加权算法(Stability Competitive Adaptive Reweighting Algorithm, SCARS)、遗传算法(Genetic Algorithm, GA)和自动有序预测因子选择算法(Automatic Ordinal Predictor Selection Algorithm, Auto OPS) 3 种方法进行特征波长提取; 最后结合机器学习方法建立了 BP 神经网络(Back Propagation Neural Network, BPNN)和基于杜鹃鸟搜 索的 BP 神经网络模型(Back Propagation Neural Network Optimized by Cuckoo Search Algorithm, CS-BPNN)。为进一步提高模型精度 与收敛性,引入自适应算法对杜鹃鸟蛋被淘汰的概率进行改进并对越界鸟窝进行新处理-基于改进杜鹃鸟搜索算法优化的 BP 神经网 络。结果表明:优化后模型效果最好,SSC 含量使用该模型决定系数 $R^2_c \rightarrow R^2_p$ 是 0.83 和 0.85, RMSEC 和 RMSEP 为 0.85 和 0.79; Vc 含量使用此模型 $R^2_c \rightarrow R^2_p$ 为 0.91 和 0.91, RMSEC 和 RMSEP 分别是 0.48 和 0.45。因此,采用近红外光谱技术与改进的机器学习 方法结合可实现对樱桃番茄内部品质的快速无损预测分析。

关键词:樱桃番茄;杜鹃鸟搜索算法;BP神经网络 文章编号:1673-9078(2023)08-287-295

DOI: 10.13982/j.mfst.1673-9078.2023.8.0974

Soluble Solids Content and Vitamin C Detection in Cherry Tomatoes Based

on Near Infrared Spectroscopy Combined with Improved CS-BPNN

KANG Mingyue^{1,2}, LUO Bin², ZHOU Ya'nan², WANG Cheng², SUN Hongyan^{1*}

(1.China University of Geosciences School (Beijing) of Mathematics and Physics, Beijing 100000, China) (2.Research Center of Intelligent Equipment, Research Center of Information Technology, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China)

Abstract: To accurately predict soluble solids and vitamin C content in cherry tomatoes, a backpropagation neural network model optimized using the improved cuckoo search algorithm (ICS-BPNN) is proposed. The near-infrared spectra of the samples at 1 350~1 800 nm were collected and pre-processed using different methods. The stability competitive adaptive reweighted sampling (SCARS), genetic (GA), and automatic ordinal predictor selection (Auto OPS) algorithms were then employed to extract the characteristic wavelength. BPNN and CS-BPNN models were established using machine learning methods. To further enhance accuracy and convergence of the models, an adaptive algorithm was introduced to improve the probability of cuckoo egg elimination, and the cross-border nests were newly processed via ICS-BPNN. The optimized models demonstrated ideal results. The results showed that the coefficients of determination, R_c^2 and R_p^2 of the soluble solid content were 0.83 and 0.85, respectively; the root mean

引文格式:

康明月,罗斌,周亚男,等.基于近红外光谱技术结合改进的CS-BPNN樱桃番茄SSC和Vc含量检测[J].现代食品科技,2023,39(8):287-295

KANG Mingyue, LUO Bin, ZHOU Ya'nan, et al. Soluble solids content and Vitamin C detection in cherry tomatoes based on near infrared spectroscopy combined with improved CS-BPNN [J]. Modern Food Science and Technology, 2023, 39(8): 287-295

作者简介: 康明月(1997-), 女, 硕士生, 研究方向: 机器学习算法, E-mail: 774404711@qq.com

收稿日期: 2022-08-03

基金项目:国家自然科学基金项目(11601494);广东省重点领域研发计划(2019B020214005);江苏省科技计划重点及面上项目(BE2021379)

通讯作者: 孙鸿雁 (1985-), 女, 博士, 副教授, 研究方向: 多元统计, E-mail: sun_hy@cugb.edu.cn

现代食品科技

square error of calibration (RMSEC) and prediction (RMSEP) sets were 0.85 and 0.79, respectively. The vitamin C content obtained using the optimized model had $R_{\rm c}^2$ and $R_{\rm p}^2$ of 0.91 and 0.91, respectively. The RMSEC and RMSEP values were 0.48 and 0.45, respectively. Thus, a combination of near-infrared spectroscopy and improved machine learning methods can achieve the rapid and non-destructive predictive analysis of the internal quality of cherry tomatoes.

Key words: cherry tomatoes; cuckoo search algorithm; backpropagation neural network

樱桃番茄是一年生长草本植物,色泽鲜红,口感 清甜,具有丰富的营养素,是一种老少皆宜的水果蔬 菜。果蔬的内部品质包括硬度、可溶性固形物、糖度、 酸度、维生素等成分,其中 SSC 含量是一项综合指标, 是所有溶解于水的化合物的总称,反映果蔬风味品质。 樱桃番茄中的 Vc 含量较高,Vc 是一种水溶性维生素, 有降低毛细血管的破裂、预防高血压等等功效,体现 了番茄的营养价值。因此本文将选择 SSC 含量和 Vc 含量对樱桃番茄的品质进行综合评价。

传统的 SSC 检测方法是用折射仪法, Vc 检测方 法主要是荧光法、二硝基苯肼法等物化方法。这些方 法检测过程复杂,耗时长,难以满足大量果蔬分级分 选的要求,因此寻找快速,简便且精准的检测方法具 有重要意义。

近年来,近红外光谱技术在食品、医药等领域发 展迅猛,该技术摒弃了传统繁琐的物化实验,具有绿 色,无损等优点。在果蔬品质检测方面,国内外已有 大量研究,郭阳等[1]利用近红外光谱技术结合反向区 间偏最小二乘算法预测哈密瓜 SSC 含量。ÉgeiMárton 等^[2]通过可见-近红外吸光度和反射率数据估算 SSC 和番茄红素含量。陈帅帅等[3]为提高可见/近红外光谱 对苹果 SSC 含量的检测精度,利用主成分回归、偏最 小二乘法和改进偏最小二乘法三种方法进行比较。上 述研究均是利用偏最小二乘回归和主成分回归等线性 模型,而有大量研究证明近红外光谱数据与果蔬品质 之间存在显著的非线性关系,因此机器学习模型更具 优势。机器学习模型中 BPNN 和极限学习机 (Extreme Learning Machine, ELM) 最为经典, 但其超参数选择 和收敛速度等方面有较大的局限性。针对这些问题, 纪然仕等[4]采用迭代保留信息变量与基于果蝇算法优 化的 ELM 算法对花椒挥发油含量进行预测,该模型 具有较高的收敛性和预测精度,为光谱检测提供新思 路。葛春靖等^[5]提出 GA-BPNN 的玉露香梨 SSC 预测 方法,该优化模型预测决定系数达到0.86,可准确检 测可溶性固形物。因此采用寻优算法对机器学习方法 进行优化有显著优势,但这些算法有参数多,操作复 杂等缺点。而杜鹃鸟搜索算法参数少、操作简单、易 实现、随机搜索路径优和寻优能力强,且该算法鲜少

用于检测樱桃番茄的品质信息。

本文运用近红外光谱分析技术结合 BPNN 方法和 基于杜鹃鸟搜索算法优化 BPNN 方法开展樱桃番茄内 部品质含量的检测研究。杜鹃鸟搜索算法虽有诸多优 势,但有全局优化能力弱,收敛速度慢等缺点,因此 针对这些缺陷进行改进,引入了改进的杜鹃鸟搜索算 法来优化 BPNN 的方法。在这项研究中,比较多种预 处理和特征波长建模方法,以建立基于近红外光谱检 测樱桃番茄 SSC 和 Vc 含量的预测模型。

1 材料与方法

1.1 材料与仪器

以千禧果、粉圆圣女果、荷兰小番茄和极星农业红 色串装小番茄四个樱桃番茄品种为实验对象,均为在超 市选购所得。每个品种选取了30个样本,共计120个。

采用型号为 BIO-NIRONE-HEM 的手持式近红外 光谱仪来采集樱桃番茄光谱数据。设置波长测定范围 为1350~1800 nm,间隔为1.5 nm,扫描时间预计9 s。 测量参数为吸光度,每个样本测量三次光谱曲线并取 平均值,作为该样本的分析光谱。使用 MEMS-FPI-Shortcut 软件,可以将采集的光谱数据文件导入到外 部 U 盘中储存备用。

1.2 内部品质化学测定

1.2.1 SSC 测定

将采集完光谱的样品进行榨汁,用绒布擦净 2WA-J 阿贝折射仪棱镜表面后,取 2~3 滴汁液在中央, 闭合棱镜,调节视度圈,使得视野分为明暗两部分, 且明暗界线在十字交叉点上,测定并读数,从而通过 折射仪法^[6]获得样品的 SSC 含量。

1.2.2 Vc 测定

利用分析天平准确称量樱桃番茄样品 5 g, 配制质 量分数为 1%草酸溶液、2%草酸溶液、0.01% 2,6-二氯 酚靛酚溶液和标准抗坏血酸溶液,利用 2,6-二氯酚靛 酚滴定法^[7]进行测定,当染料溶液滴定至淡红色为终 点,记录数值,利用 1%草酸溶液作为空白对照,利 用公式^[8]计算数值作为样本的 Vc 含量。

1.3 数据处理

1.3.1 预处理方法

首先将样本按照三倍标准差剔除异常值后,采用 Kennard-Stone(K-S)分类算法^[9]将样本按 3:1 的比例 划分为校正集和预测集。由于近红外光谱存在人为或 自身的噪声影响,本文采用多元散射校正(Multivariate Scattering Correction, MSC)^[10]、Savitzky-Golay 卷积 平滑(Savitzky-Golay Convolution Smoothing, SG)^[11]、 去趋势化(De-trending)^[12]、变量标准化(Standard Normal Variate, SNV)^[13]四种经典的方法进行预处理, 提升数据精度。

1.3.2 特征波长选取

光谱间存在大量重复冗余信息,会明显降低模型的计算速度。因此,为提升模型的稳定性,选择两种常用的特征波长选择的方法和一种新兴的特征变量选择方法。特征波长提取方法有稳定性竞争性自适应重加权算法、遗传算法和自动有序因子选择算法^[14]。

1.3.3 回归模型

采用 BP 神经网络、基于杜鹃鸟搜索算法优化的 BP 神经网络以及改进的杜鹃鸟搜索算法优化的 BP 神 经网络建立樱桃番茄内部品质含量回归模型。BP 神经 网络在预测时,依赖于初始权值和阈值,将杜鹃鸟搜 索算法引入 BP 神经网络中,避免初始权值和阈值陷 入局部最小,同时也可以改进收敛速度。其中杜鹃鸟 搜索算法是基于杜鹃鸟的寄生行为^[15]和鸟类的 Lévy 飞行行为^[16]。但传统的杜鹃鸟算法全局搜索能力和收 敛速度都有待提高,因此本文通过引入自适应算法和 变更边界值进行改进。

传统杜鹃鸟算法中的鸟蛋被淘汰的概率 *P_a* 固定 值 0.25,但这不利于局部与全局搜索的战略,经已有 研究发现^[17],*P_a*在 0.1~0.75 之间随着迭代次数的增加 而增大,因此引入自适应策略公式(1)为:

$$P_a = 1.1 - e^{\left(-(N_i/N_{im})^{0.6}\right)}$$
(1)

式中:

 P_a ——乌蛋被淘汰的概率;

Ni——算法当前迭代次数;

Nim——设定的最大迭代次数。

在传统杜鹃鸟搜索算法中,越界值普遍采用边界 值,这丧失了鸟窝不断变化的灵活性,因此提出一种 按照随机数的不同处理鸟窝的新方法,这将进一步提 高算法收敛性,新方法可由式实现:

$$\begin{cases} X(i,j) = rand \times (U_b - L_b) + L_b, X(i,j) \ge U_b \\ X(i,j) = rand \times (U_b - L_b) + L_b, X(i,j) \le L_b \end{cases}$$
(2)

式中: X(i, j) ——最新迭代鸟窝的位置; Ub——边界最大值; Lb——边界最小值。 流程图如图1所示。 初始参数 随机产生—组鸟巢对应的 初始权值和阈值



图 1 基于杜鹃鸟搜索算法的 BP 神经网络流程图

Fig.1 BP neural network flow chart based on cuckoo optimization algorithm

该算法步骤如下:

(1)初始化鸟巢数 n,以及最大迭代次数 N_{iter max};

(2)随机产生 n 个鸟巢的初始位置,该位置与神 经网络的权值和阈值相对应;

(3) 计算初始位置的适应度值, 其中适应度函数 为 BPNN 中训练集的均方误差;

(4)利用公式(1)计算鸟蛋淘汰率,并不断更 新鸟巢,根据公式(2)判断是否越界,从而更新位置;

(5) 根据适应度值找到最佳位置;

(6) 若迭代次数达到最大,则将最优输出为模型 初始权值和阈值,否则返回步骤4;

(7) 计算误差更新权值和阈值, 对样本含量进行 预测。

1.3.4 模型评价指标

校正集效果根据校正集决定系数 (R^2_c) 和校正集 均方根误差 (RMSEC) 来评价,校正集决定系数越大, 校正集均方根误差越小,校正模型效果越好;预测效果 由预测集决定系数(R^2_p)和预测集均方根误差(RMSEP) 进行综合评价,其中预测集决定系数越接近于 1,其 均方根误差越接近于 0,则模型预测性能越好。

1.3.5 数据分析

设置近红外光谱上述参数,选择樱桃番茄的果底和赤道相对的两点作为感兴趣区域,获取光谱数据后, 其原始光谱图如图2所示,在 Matlab 2019a 软件上编 写程序对该区域光谱数据进行预处理,特征波长提取 和建立预测模型,从而获得光谱预处理后光谱曲线和 特征波长提取过程,最终采用 Origin 7.5 画出建模结 果的对比图。



2 结果与讨论

2.1 预处理方法

表1 预处理方法比较

Table 1 Comparison of pretreatment methods

山如日氏	石山田二山	枢	交正集	予	预测集		
闪印印顶	顶处理方法	$R^2_{\rm c}$	RMSEC	R^2_{p}	RMSEP		
	Original	0.93	0.49	0.78	0.95		
	MSC	0.92	0.51	0.78	0.95		
SSC	SG	0.77	0.87	0.46	1.66		
	De-trending	0.95	0.41	0.84	0.79		
	SNV	0.93	0.49	0.78	0.98		
	Original	0.91	0.49	0.89	0.49		
	MSC	0.92	0.46	0.69	0.82		
Vc	SG	0.91	0.49	0.89	0.48		
	De-trending	0.88	0.55	0.77	0.71		
	SNV	0.92	0.47	0.70	0.81		

将原始光谱作上述方法的预处理后,将预处理变 量和原始变量分别作为输入变量,内部品质含量作为 输出变量,建立偏最小二乘回归^[18]。继而选出适合不同内部品质的最优预处理方法。结果如表 1。

由表 1 可知,在预测 SSC 含量时,去趋势法的 *R*²_p最高,均方根误差均低于其他预处理方法,分别为 0.95 和 0.84。去趋势法预处理后的光谱图如图 3 所示。 0.04 Γ



图 3 SSC 最优预处理光谱图(De-trending)

Fig.3 SSC optimal preprocessing spectrum (De-trending)

基于原始光谱建立的 Vc 含量预测模型, Savitzky-Golay 卷积平滑在预测集中有显著优势,虽然 原始光谱的 R^2_p 同为 0.89,但 RMSEP 略低于原始光谱 为 0.49,因此, Savitzky-Golay 卷积平滑为预测的最优 方法。经 SG 算法处理后的光谱图由图 4 所示。



2.2 BP 神经网络建模

Table 2 Comparison of BPNN modeling feature variable selection methods											
内部品质	特征提取方法	<u> </u>	枢	交正集	预测集						
		波长效	R ² _c	RMSEC	R ² _p	RMSEP					
	SCARS	38	0.93	0.50	0.76	1.06					
SSC	GA	36	0.90	0.57	0.74	1.13					
	Auto OPS	60	0.85	0.70	0.71	1.06					
Vc	SCARS	10	0.89	0.87	0.86	0.85					
	GA	31	0.82	0.88	0.71	0.95					
	Auto OPS	200	0.90	0.54	0.74	0.91					

表 2 BPNN 建模特征变量选取方法比较



Modern Food Science and Technology



Fig.5 Vc optimal characteristic variables selection process 采用偏最小二乘方法建模时,校正集结果较好, 但预测效果相对一般,因此为减少二者差距,进一步 提高预测效果,将最佳预处理变量分别应用 SCARS、 GA 和 Auto OPS 算法处理,建立的 BPNN 预测模型如 表 2 所示。设置 BP 网络训练目标最小误差为 0.1,训 练次数为 1000,训练所要达到的精度为 1e⁻⁶,学习速 率为 0.01,训练算法为 Levenberg-Marquardt 算法^[19]。

由表2可知,根据综合评价指标,基于该方法分 别建模时,SCARS 算法具有显著优势,其波长数基本 少于其他方法。当用该算法检测 SSC 含量时,校正集 的决定系数最高为0.93,均方根误差最小为0.50,其 预测集的决定系数为0.76,均方根误差为1.06。

应用于樱桃番茄 Vc 含量测定时,同样是 SCARS 算法的预测效果最好。其预测精度最高为 0.86,远大 于剩余 GA 和 Auto OPS 两算法决定系数即 0.71 和 0.74。该方法的特征波长提取过程和提取结果如图 5 和图 6 所示。从选择结果上看,在迭代到第 202 次时, 交叉验证的均方根误差达到最小为 0.76,选择变量数 为 10 个,仅占全波长的 3.3%,明显降低模型复杂度。



将最佳预处理变量分别应用 SCARS、GA 和 AutoOPS 算法处理后,建立的樱桃番茄内部品质含量 ELM 预测模型如表 2 所示。ELM 建模时,樱桃番茄 SSC 含量定义的隐含层神经元个数为 35,而 Vc 含量定 义个数为 30,采用的隐含层激活函数为 Sigmoid 函数。

中部日氏	サイントレークテント	1 V 4	杉	交正集	预测集		
内部四顶	将征波大灰权力法	波大致	R^2_{c}	RMSEC	R^2_{p}	RMSEP	
	SCARS	38	0.89	0.60	0.64	1.46	
SSC	GA	36	0.90	0.57	0.69	1.16	
	Auto OPS	60	0.84	0.72	0.63	1.76	
	SCARS	10	0.85	0.62	0.80	0.86	
Vc	GA	31	0.64	0.98	0.61	0.89	
	Auto OPS	200	0.87	0.58	0.69	0.95	

表 3 ELM 建模特征变量选取方法比较

由表 2 可知, ELM 作为机器学习算法的一种, 应 用于樱桃番茄内部品质含量测定时, SSC 含量的 R_p^2 均在 0.70 以下, Vc 最高 R_p^2 为 0.8, 预测精度相对一般。基于该方法建模时, SCARS 算法具有显著优势,







由表 2、3 对比得知, BPNN 效果优于 ELM 算法, 因此进一步对 BPNN 进行优化,以期获得更高的预测 精度。杜鹃鸟搜索算法的优点是提高算法的局部和全 局寻优,本文初始化鸟巢数 n 为 25,迭代次数为 100 次,以均方差作为适应度函数。运行结果如表 4 所示。 预测 SSC 含量时, De-trending-Auto OPS-CS-BPNN 模型最佳, *R*²_c 和 *R*²_p分别为 0.83 和 0.80, RMSEC 和 RMSEP 分别为 0.80 和 1.02, 迭代曲线如图 7 所示。 当迭代次数为 93 次时,迭代基本稳定,此时适应度值 为 0.14,表明预测值和测量值误差逐渐稳定。

预测 Vc 含量时, SG-SCARS-CS-BPNN 模型最优, R_c^2 , R_p^2 分别为 0.90 和 0.86, RMSEC 和 RMSEP 分别 为 0.54 和 0.54。迭代曲线如图 8 所示。随着迭代次数 增加,适应度曲线逐渐平缓,当迭代次数为 69 次时, 迭代基本稳定,适应度值为 0.06。



Fig.8 Iterative curve of CS algorithm (Vc)

内部品质	妹红沙と坦取ナン	ンセレキレ	校	正集	予	预测集		
	将征波下延取方法	波大致	$R^2_{\rm c}$	RMSEC	R^2_{p}	RMSEP		
SSC	SCARS	38	0.73	1.08	0.77	1.25		
	GA	36	0.74	0.89	0.78	1.34		
	Auto OPS	60	0.83	0.80	0.80	1.02		
	SCARS	10	0.90	0.54	0.86	0.54		
Vc	GA	31	0.87	0.88	0.85	0.84		
	Auto OPS	200	0.88	0.57	0.75	0.67		

Table 4 Comparison of characteristic variable selection methods for CS-BP neural network modeling

表5	ICS-BP 神经网络建模特征变量选取方法比较
----	-------------------------

ľ٤	ıbl	e 5	5 C	omr	arisoi	1 of	characteristic	variable	e selection	methods	for	· ICS-B	P neural	network	modeling
			-	r											

$\overline{\mathbf{X}}$	山朝口氏	時行地と相面子は	计上步	杉	交正集	予	页测集		
	內미이贝	讨证成认证收方法	灰 下	$R^2_{\rm c}$	RMSEC	R^2_{p}	RMSEP		
		SCARS	38	0.82	0.85	0.82	0.89		
\frown	SSC	GA	36	0.80	0.97	0.82	1.08		
	<u> </u>	Auto OPS	60	0.83	0.85	0.85	0.79		
		SCARS	10	0.91	0.48	0.91	0.45		
	Vc	GA	31	0.89	0.55	0.88	0.45		
		Auto OPS	200	0.87	0.60	0.79	0.64		

2.5 ICS-BP 神经网络建模

采用改进的杜鹃鸟搜素算法优化 BP 神经网络方 法来预测樱桃番茄的品质含量结果如表 5 所示。该方 法的预测效果较优。 综合对比三种不同特征波长提取的建模方法可知, De-trending-Auto OPS-ICS-BPNN 模型建立的樱桃番茄 SSC 含量预测性能最优,其预测精度为 0.85, De-trending-SCARS-ICS-BPNN 模型次之, De-trending-GA-ICS-BPNN 模型最差,其预测精度均为 0.82,均



图 9 基于 De-trending-Auto OPS-ICS-BPNN 模型 SSC 含量回归 效果

Fig.9 Regression effect of SSC content based on

De-trending-Auto OPS-ICS-BPNN model

为防止 BP 神经网络有过拟合现象^[20],采用的方 法是把数据 6:2:2 的比例划分成三份,分别为 Training (训练)、Validation(验证)和 Test(测试)。其中 Training 数据参加训练,其他两部分数据用于检验。在训练进行过程中,目标(Target)即樱桃番茄内部品质含量的化学值和测试(Test)数据之间的误差会越来越小,它们之间的相关系数越接近于1时,效果越好。使用Matlab对模型进行的训练回归效果如图 9 所示,Training、Validation、Test和All(整体)与目标值的相关系数均在 0.87 以上。因此可知,De-trending-AutoOPS-ICS-BPNN模型可以作为樱桃番茄 SSC含量的预测模型。

用该建模方法预测樱桃番茄 Vc 含量时,SG-SCARS-ICS-BPNN 模型效果最佳,预测精度高达0.91,均方根误差相比最小为0.45。对 SG-SCARS-ICS-BPNN 模型进行的训练回归效果如图 10 所示,Training、 Validation、Test 和 All (整体)与目标值的相关系数分 别是 0.97、0.98、0.85 和 0.96。因此可知,预测值和 参考值之间具有较好的线性关系,表明近红外光谱技 术对樱桃番茄 Vc 含量的检测是可行的。



Modern Food Science and Technology





2.6 不同建模方法比较



Fig.12 Comparison of Vc optimal models

图 11 和图 12 分别为 SSC 和 Vc 含量基于不同建 模方法,分别选择最优结果的对比结果。X 轴以上为 预测集 R²,以下为 RMSEP。柱状的长度对应数值的 呈正比关系。从图 11 可看出,上柱状 R²随着模型不 同显著提升,下柱状 RMSEP 显著减小,ICS-BPNN 预测集 R²达到 0.85,相较 ELM、BPNN 和 CS-BPNN 分别提升了 16%、9%和 5%, RMSEP 达到 0.79,相 比 ELM、BPNN 和 CS-BPNN 模型降低了 0.37、0.27 和 0.23,所以应用 ICS-BPNN 时,模型效果最好。从 图 12 可以看出,上柱状 R²随着模型不同柱状长度增加不明显,但下柱状 RMSEP 的柱状长度有明显减小, Vc 含量预测集均方根误差从最大值 0.85 降到最小值 0.45。因此,经改进的杜鹃鸟搜索算法优化 BPNN 模 型具有最好的预测效果。

3 结论

本文利用近红外光谱技术对四个不同品种的樱桃 番茄进行测量,通过多种不同算法建立的樱桃番茄 SSC 和 Vc 含量预测模型。同时经过比较分析得出, 经改进后的 CS-BPNN 模型预测精度最高。在 SSC 含 量应用中该模型预测集决定系数达到 0.85,预测集均 方根误差为 0.79; Vc 含量应用该模型预测集决定系数 高达 0.91,预测集均方根误差为 0.45。因此,改进的 机器学习算法可有效提高检测精度,为樱桃番茄品质 检测提供理论依据。

参考文献

- [1] 郭阳,史勇,郭俊先,等.近红外光谱技术结合反向区间偏最 小二乘算法-连续投影算法预测哈密瓜可溶性固形物含量
 [J].食品与发酵工业,2022,48(2):248-253.
- [2] Égei Márton, Takács Sándor, Palotás Gábor, et al. Prediction of soluble solids and lycopene content of processing tomato cultivars by Vis-NIR spectroscopy [J]. Frontiers in Nutrition, 2022, 9(9): 1-10
- [3] 陈帅帅,张鹏,李江阔,等.寒富苹果可溶性固形物可见/近红外 光谱无损检测模型的优化[J].保鲜与加工,2018,18(2):86-93.
- [4] 纪然仕,陈晓燕,刘素珍,等.基于高光谱技术和 IRIV-FOA-ELM 算法的花椒挥发油无损检测[J].激光与光电子学进 展,2020,57(20):394-400.
- [5] 葛春靖,张淑娟,孙海霞.基于 GA-BP 神经网络玉露香梨可 溶性固形物高光谱技术检测[J].现代食品科技,2021,37(5): 296-302,278.
- [6] 黄光莉.折射仪法快速检测卷烟的白利糖度[J].现代科学仪器,2012,1:95-97.
- [7] 徐朝阳.2,6-二氯酚靛酚滴定法与碘量法测定蔬菜水果中 Vc方法的准确度比较[J].食品安全导刊,2021,25:100-101.
- [8] 吴晓伟,杨剑婷,王俊.不同蔬菜热烫对 Vc 的影响[J].食品工 业科技,2012,33(11):238-240.
- [9] Li T Y, Wu Y Y, Wu F, et al. Sleep pattern inference using IoT sonar monitoring and machine learning with Kennard stone balance algorithm [J]. Computers and Electrical Engineering, 2021, 93(93): 1-19.
- [10] Li Y, Wang G Z, Guo G S, et al. Spectral pre-processing and multivariate calibration methods for the prediction of wood

Modern Food Science and Technology

density in Chinese white poplar by visible and near infrared spectroscopy [J]. Forests, 2022, 13(1): 61-72.

- [11] Krishna C M, Sharma L D. Electrocardiogram signal filtering using circulant singular spectrum analysis and cascaded Savitzky-Golay filter [J]. Biomedical Signal Processing and Control, 2022, 75: 1-15.
- [12] Emirmahmutoglu F, Omay T, Shahzad S J H, et al. Smooth break detection and de-trending in unit root testing [J]. Mathematics, 2021, 9(4): 371-371.
- [13] Xiao Q L, Tang W T, Zhang C, et al. Spectral preprocessing combined with deep transfer learning to evaluate chlorophyll content in cotton leaves [J]. Plant Phenomics, 2022, 2022: 1-42.
- [14] Jussara V, Roque, Wilson C. et al. Comprehensive new approaches for variable selection using ordered predictors selection [J]. Analytica Chimica Acta, 2019, 1075: 57-70.
- [15] Wei L, Canchao Y, Fugo T. Modeling the cuckoo's brood parasitic behavior in the presence of egg polymorphism [J]. Journal of Ethology, 2016, 34(2): 127-132.

- [16] 张慧颖,王凯,于海越,等.基于自适应 Levy 飞行的黄金正弦 可见光定位研究[J].激光技术,2022,46(4):519-524.
- [17] 贺淼,阮奇,郑晓桂,等.自适应布谷鸟搜索算法[J].计算机与应用化学,2014,31(8):961-968.
- [18] Miloš B, Bensa A, Japundžić-Palenkić B. Evaluation of Vis-NIR preprocessing combined with PLS regression for estimation soil organic carbon, cation exchange capacity and clay from eastern Croatia [J]. Geoderma Regional, 2022, 30: 1-12.
- [19] Wang M Y, Xu X Y, Yan Z, et al. An online optimization method for extracting parameters of multi-parameter PV module model based on adaptive Levenberg-Marquardt algorithm [J]. Energy Conversion and Management, 2021, 245(1): 114611.
- [20] Nugroho K, Noersasongko E, Purwanto, et al. Enhanced indonesian ethnic speaker recognition using data augmentation deep neural network [J]. Journal of King Saud University - Computer and Information Sciences, 2022, 34(7): 4375-4384.