

# 基于卷积神经网络的茶鲜叶主要内含物的光谱快速检测方法

李晓丽 张东毅 董雨伦 金娟娟 何勇\*

(浙江大学 生物系统工程与食品科学学院/农村农业部光谱传感重点实验室,杭州 310058)

**摘要** 儿茶素和咖啡碱是茶叶品质的重要评价指标,为了探索深度卷积神经网络(CNN)结合可见近红外光谱(Vis/NIR)用于茶叶儿茶素和咖啡碱无损快速检测的可行性,本研究通过高效液相色谱来测定茶叶中的儿茶素和咖啡碱含量,并与样本的光谱信息建立对应关系;采用回归分析和CNN建模构建了光谱与茶叶内含物的定量关系;采用竞争自适应重加权采样(CARS)和连续投影算法(SPA)选择特征波长,用于开发基于这些特征波长的简单模型。结果表明:4种儿茶素和咖啡碱含量从第1叶位到第6叶位呈现出逐渐降低的趋势;提取特征波长不仅减少了光谱变量数,还获得了比全谱更优或接近的模型性能;CNN在回归分析和特征提取中均表现出良好的性能,预测儿茶素和咖啡碱最优模型的决定系数( $R^2$ )和残余预测偏差(RPD)分别达到了0.93和3.28以上。因此,卷积神经网络结合可见近红外光谱可以对儿茶素和咖啡碱的含量进行快速无损检测。

**关键词** 卷积神经网络;可见近红外光谱;新鲜茶叶;儿茶素;咖啡碱

中图分类号 S571.1;O657.33

文章编号 1007-4333(2021)11-0113-10

文献标志码 A

## Spectral rapid detection of phytochemicals in tea (*Camellia sinensis*) based on convolutional neural network

LI Xiaoli, ZHANG Dongyi, DONG Yulun, JIN Juanjuan, HE Yong\*

(College of Biosystem Engineering and Food Science/Key Laboratory of Spectroscopy Sensing of Ministry of Agriculture and Rural Affairs, Zhejiang University, Hangzhou 310058, China)

**Abstract** Catechins and caffeine are the key indicators of tea quality. In order to explore the feasibility of visible/near infrared spectroscopy (Vis/NIR) combined with convolutional neural network (CNN) for nondestructive and rapid detection of catechins and caffeine of fresh tea, this study used high performance liquid chromatography to determine the content of catechins and caffeine in tea, and established a corresponding relationship with the spectral information of the sample. Regression analysis and CNN modeling were used to construct the quantitative relationship between the spectrum and contents of tea. In addition, the competitive adaptive reweighted sampling (CARS) and successive projections algorithm (SPA) were used to select the characteristic wavelengths for the development of simple models. The result showed that: The contents of four catechins and caffeine displayed a gradual decrease trend from the first leaf position to the sixth leaf position. The extracting characteristic wavelengths not only reduced the number of spectral variables, but also obtained models performance that were better or close to that of the full spectrum. CNN performed well in the regression analysis and feature extraction, the coefficient of determination ( $R^2$ ) and residual prediction deviation (RPD) of optimal models for predicting catechins and caffeine reached 0.93 and 3.28, respectively. Therefore, the convolutional neural network combined with visible/near infrared spectroscopy could quickly and non-destructively detect contents of catechins and caffeine.

**Keywords** convolutional neural network; visible/near infrared spectroscopy; fresh tea; catechin; caffeine

收稿日期:2021-01-29

基金项目:国家自然科学基金项目(31771676);国家重点研发计划项目(2018YFD0700501);浙江省科技计划项目(2017C02027-01);高校基本科研业务费专项资金项目(2015QNA6005)

第一作者:李晓丽,教授,主要从事农业信息化、农产品品质无损测量研究,E-mail:xiaolili@zju.edu.cn

通讯作者:何勇,教授,主要从事数字农业与精细农业研究,E-mail:yhe@zju.edu.cn

茶是世界三大非酒精饮料之一,具有特殊的风味和保健功能<sup>[1]</sup>。茶由茶树的嫩叶制成,茶叶中的茶多酚和咖啡碱是决定茶叶色、香和味的重要成分<sup>[2]</sup>。茶多酚是一类多羟基化合物,主要由儿茶素组成<sup>[3]</sup>,被证明具有各种有益的药用特性,如抗氧化能力、抗肿瘤活性、预防心血管疾病和增强免疫力<sup>[4]</sup>。表没食子儿茶素没食子酸酯(EGCG)、表儿茶素没食子酸酯(ECG)、表没食子儿茶素(EGC)和表儿茶素(EC)是新鲜茶叶中具有不同生物活性的4种主要儿茶素<sup>[5]</sup>,在加工过程中,这些儿茶素可分别异构化为没食子儿茶素没食子酸酯(GCG)、儿茶素没食子酸酯(CG)、没食子儿茶素(GC)和儿茶素(C)<sup>[6]</sup>。咖啡碱(CAF)也是茶叶的重要成分<sup>[7]</sup>,它可以对人体产生刺激作用<sup>[8]</sup>。

研究表明,不同品种茶叶的茶多酚和CAF有较大的差异<sup>[9]</sup>,此外,茶树在生长过程中不同叶位的茶多酚和CAF也会有所不同<sup>[10]</sup>。通过研究鲜叶中茶多酚和CAF的含量一方面可以了解茶园中茶树的生长状况,进而对茶树进行更加有效的管理,同时也可以为选择具有更多生物活性成分的叶片提供指导,保证采摘茶叶的品质,以此促进我国茶产业的高质量发展。因此,迫切需要1种能够快速检测新鲜茶叶内含物含量的方法,为以后茶园智能采摘茶叶和精细化管理提供帮助。

目前,高效液相色谱(HPLC)<sup>[11]</sup>和毛细管电泳(CE)<sup>[12-13]</sup>等多种技术常用于测定茶叶中的儿茶素和CAF含量,但这些方法耗时且耗材,不能实现茶叶品质的在线实时检测。可见/近红外光谱(Vis/NIR)是1种快速、简单和非破坏性的技术,它已被广泛应用于许多领域<sup>[14]</sup>。随着光学仪器的改进和计算机科学与化学计量学的发展,Vis/NIR已广泛应用于茶叶质量分析领域<sup>[15-18]</sup>。Vis/NIR可以反映C-H,N-H,O-H主要化学键的吸收特性,因此被广泛用于有机物质组成的定量分析<sup>[19]</sup>。近年来,由于其特征学习能力,深度学习在分类和回归分析中具有更好的模型性能。深度学习最初是针对分类问题而开发的,最近有研究报道将深度学习用于光谱数据分析中的回归问题<sup>[20]</sup>。此外,深度学习方法不仅可以用作建模回归还可以进行特征提取<sup>[21]</sup>。研究表明,与其他机器学习方法相比,用于回归的卷积神经网络(CNN)可以获得相同甚至更好的模型性能。因此,本研究旨在探讨不同品种和不同叶片位置

的茶鲜叶光谱和内含物的差异性,并基于Vis/NIR技术结合CNN算法来构建鲜叶中4种主要儿茶素(EGCG, ECG, EGC, EC)和CAF的定量模型,以期实现不同品种和叶位儿茶素单体和CAF的快速无损检测。

## 1 材料与方法

### 1.1 茶鲜叶样品采集

2019年4月25日在浙江大学华家池校区(30.3°N, 120.2°E)茶园进行样品采集。茶树树龄为5年,采摘3个不同品种茶树的鲜叶,分别为‘菊花春’(JHC), ‘浙农25’(ZN25)和‘迎霜’(YS),每个品种分别采摘6个叶位(叶梢第1叶至第6叶),每个叶位随机收集9个样品。最终,共获得了162个新鲜茶叶样品。

### 1.2 可见/近红外光谱采集

使用XDS近红外光谱仪(FOSS, Denmark)在400~2498 nm范围内以2 nm的分辨率进行叶面的Vis/NIR反射光谱采集。样本的每条光谱值均为光谱仪扫描32次的平均光谱值,每条光谱产生总共1050个反射率值。吸光度的计算公式如下:

$$A = \log(1/R) \quad (1)$$

式中:A为光谱的吸光度;R为光谱的反射率,%。

为了增加信噪比,每个茶叶样品测量3次,以3个光谱的平均值作为样品的光谱值,通过“ISIScan v1.50”软件获取光谱数据。

### 1.3 儿茶素和咖啡碱含量的化学检测

采集完新鲜茶叶样本的光谱后,将所有新鲜茶叶在冷冻干燥器(FreeZone6, Labconco Corp, Kansas City, MO, USA)中干燥超过24 h。冷冻干燥后,用研磨机(FW100, Taisite Instrument Co., Ltd., Tianjin, China)将样品研磨成粉末,并通过60目筛子筛分。随后,用25 mL水萃取0.1 g筛分的茶粉在85 °C的温度下保持20 min。然后将上清液通过0.22 μm膜过滤器过滤用于HPLC分析。整个操作在室温(25 °C)和黑暗环境下进行,以防止儿茶素的分解。

通过配有紫外可见检测器(波长范围:190~600 nm)的日本岛津LC-20AD高效液相色谱仪来测定4种主要儿茶素和咖啡碱的含量<sup>[22]</sup>。流动相为:(A)V(乙酸):V(乙腈):V(蒸馏水)=0.5:3:96.5和(B)V(乙酸):V(乙腈):V(蒸馏水)=

0.5 : 30 : 69.5。梯度洗脱程序如下:0~35 min,线性梯度从 20%至 65%B 流动相;35~45 min,等度 65%B 流动相。流速为 1.0 mL/min,注射体积为 10  $\mu$ L。检测波长为 280 nm,柱温保持在 35  $^{\circ}$ C。

## 1.4 常规数据分析方法

### 1.4.1 回归分析的样本划分

将所有茶叶样品根据 4 种儿茶素和 CAF 的化学含量值以升序排列,每 3 个样本分为 1 组,位于每组中间的样本被选为预测组,而剩余的 2 个样品被选为校准组。因此,校准组的化学值范围可以覆盖预测组的范围,使得模型更稳定。本次试验共有 162 个样本,因有 1 个异常样本将其剔除,最后 161 样本中 107 个样本用于校准集以建立检测模型,而剩余的 54 个样本用于预测集以测试模型的稳健性。

### 1.4.2 建立回归模型

偏最小二乘回归(PLSR)是一种线性回归方法,它同时将自变量(X)和因变量(Y)投影到一个新的空间,其约束条件是可以尽可能多地解释 X 和 Y 之间的方差。所以它不仅是尽可能多地提取有用信息,还能加强 X 和 Y 之间的相关性<sup>[23]</sup>。通过对光谱进行预处理,比较光谱 PLSR 模型的预测能力,进而评估预处理方法的性能。

### 1.4.3 特征波长选择

Vis/NIR 光谱通常由数千个光谱变量组成,其中包含许多非信息量或共线变量。全谱(Original)建模不仅会增加模型的复杂度和计算时间,还会影响模型的预测能力,从而阻碍了模型在实际生产中的应用。为了解决该问题,需要从 Original 中选择特征波长以减小 Vis/NIR 光谱数据的维数。CARS 是一种基于偏最小二乘法(PLS)模型回归系数的波长选择方法,使用蒙特卡洛采样建立 PLS 模型,以获得 PLS 模型的回归系数的绝对值。计算每个波长的绝对回归系数的权重,并根据指数递减函数删除权重小的波长。选择与具有最小交叉验证均方根误差(RMSECV)的 PLS 模型相对应的候选子集<sup>[24]</sup>。竞争自适应加权抽样(Competitive adaptive reweighted sampling, CARS)是一种很好的变量选择方法,这些变量在 PLS 模型中具有绝对值更大的回归系数,以此来提高模型的预测能力。连续投影算法(Successive projections algorithm, SPA)是通过比较变量投影来选择特征波长的方法,它能够从光谱信息中充分寻找含有最低冗余信息的变量组,

以减少变量之间的共线性。该算法可以通过设定目标变量数范围,利用某个波长在其它所有波长上的投影向量的大小来挑选引入的待选波长,最后基于校正模型,从所有待选波长集合中选出符合设定数量范围的特征波长<sup>[25]</sup>。最后分别对 Original 以及 SPA 和 CARS 选择的特征波长进行 PLS 回归,如图 1 中的①所示。

## 1.5 深度学习方法

深度学习在分类任务中取得了很好的效果之后,也开始应用于回归建模。在这项研究中,采用了基于自定义分类的 CNN 架构进行回归<sup>[26]</sup>。CNN 架构是使用几个基本单元设计的,包括一维(1D)卷积层、最大池化层、relu 层和完全连接层。一维卷积和最大池的组合用于从原始输入中自动进行特征表示,使用完全连接的块来建立提取的特征和相应标签之间的关系。训练任务的目的是使 CNN 模型的输出值和所测量的化学值之间的误差最小化。

在本研究中,通过 CNN 方法对 4 种儿茶素和 CAF 进行检测的整体框架分别如图 1 中的②和③所示。第 1 种是将 CNN 用作建模方法,通过训练 CNN 模型来对 4 种主要儿茶素和 CAF 进行预测。第 2 种 CNN 被用作特征提取,Original 以及 SPA 和 CARS 选择的波长分别进行 CNN 特征提取,总共提取了 64 个特征,然后提取的特征被用作 PLS 回归模型的输入,以探索特征与化学成分之间的关系。CNN 架构通过不同堆叠的矩形来表示,紫色的小矩形表示卷积核,圆圈表示完全连接层中的神经元,堆叠矩形旁边标记的数字描述了卷积核的数量,圆圈旁边标记的数字描述了完全连接层中的神经元数量,L 为卷积层展开的全部神经元数量。将卷积内核大小设置为 3,步长为 1,最大池化层大小设置为 2,步长为 2。

## 1.6 模型评估

光谱数据分析和特征波长选择基于 Matlab R2019b(The MathWorks, Natick, MA, USA)进行。CNN 在 Python3 和 Pytorch 框架(Facebook, Menlo Park, California, USA)上执行。PLSR 是基于 Unscrambler 软件实现的(版本 10.1, CAMO, Inc, Oslo, Norway)。开发模型的性能通过决定系数( $R^2$ ),均方根误差(RMSE)和残余预测偏差(RPD)进行评估,良好性能的模型应该包含较高的  $R^2$  和 RPD 值,并有较低的 RMSE 值<sup>[27]</sup>。

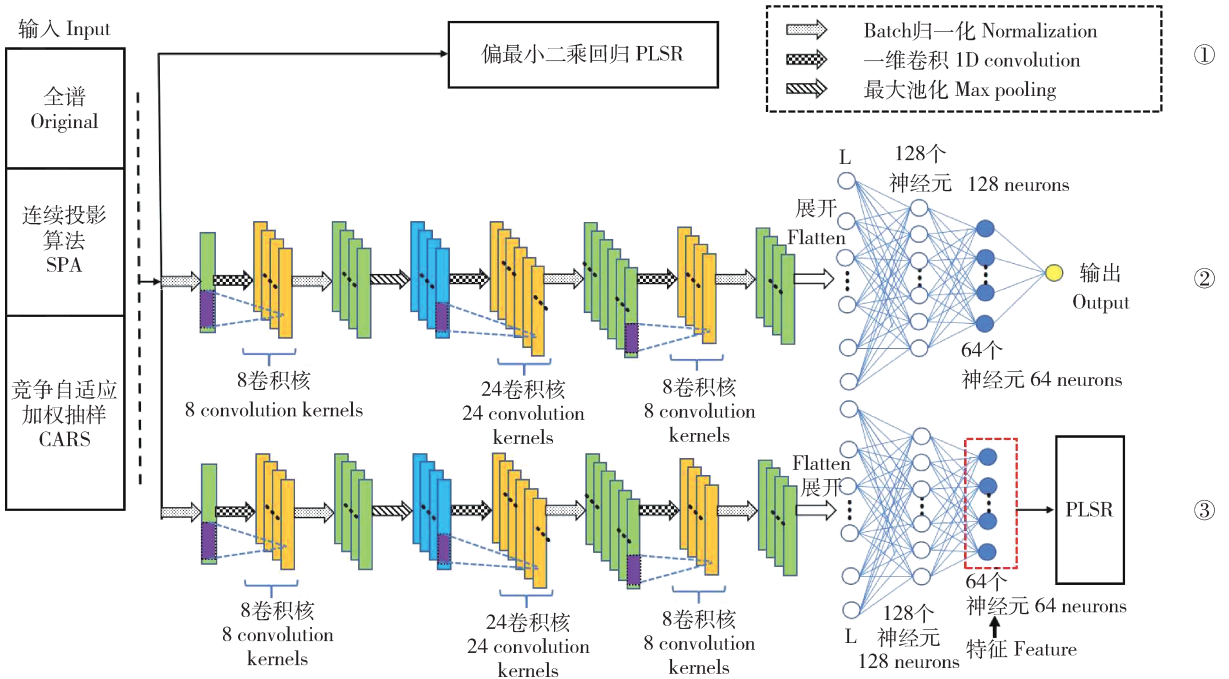


图1 光谱检测4种主要儿茶素和咖啡碱的整体框架

Fig. 1 The overall framework of spectral detection of the four main catechins and caffeine

## 2 结果与分析

### 2.1 茶鲜叶的 Vis-NIR 光谱响应特性

茶叶样品的原始光谱采集如图2所示,波长范围为400~2 498 nm,所有茶叶样品在整个光谱区域具有相似的趋势,但吸光度的大小不同。在位于400~800 nm的可见光区域中存在480 nm和670 nm 2个明显的吸收峰。表明茶叶样品主要在可见光谱区范围内吸收蓝紫色(455~492 nm)和红

色(622~770 nm)光,而很少吸收绿色(492~577 nm)光,这是茶叶显示绿色的原因。在近红外区域(800~2 500 nm)有7个吸收带,1 200 nm处的平坦吸收峰归因于-CH<sub>2</sub>的C-H伸缩振动的第二倍频;1 464 nm处的峰值带归因于O-H的振动;1 736 nm范围内的信号与-CH<sub>3</sub>的C-H振动相关;1 930 nm处的吸收峰来自O-H拉伸和HOH转变的组合振动;2 130 nm处的吸收带是由N-H弯曲和C=O拉伸的组合振动引起;在2 310 nm和2 350 nm

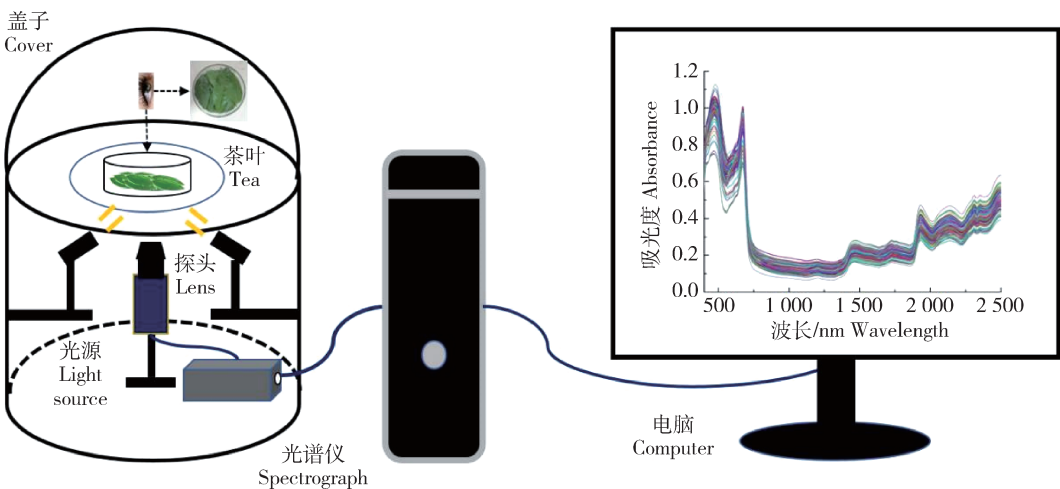


图2 茶样品的可见近红外光谱采集系统

Fig. 2 The visible and near infrared spectrum acquisition system of tea samples

处的 2 个峰分别较弱,分别对应于 C-H 弯曲和 C-H 对称拉伸<sup>[28]</sup>。这些条带与多酚和咖啡碱的各种官能团(CH、CH<sub>2</sub>、CH<sub>3</sub>、C=O、OH 和 NH)密切相关<sup>[29]</sup>。

### 2.2 茶鲜叶内含物统计分析

3 个品种茶鲜叶不同叶位对 4 种主要儿茶素和咖啡碱的含量影响见图 3。可以发现,‘JHC’和‘ZN25’的 EGCG,EGC,ECG 和 EC 在第 1 叶位的含量高于其他 5 个叶位,且‘YS’的 EGCG,EGC 和

ECG 在第 1 叶位的含量最高,每个品种茶叶的 CAF 从第 1 叶位到第 3 叶位含量基本一致,第 4 叶位到第 6 叶位含量则较低。此外,新鲜茶叶中 4 种主要儿茶素和咖啡碱的含量随茶树品种的变化而不同。在每个品种鲜叶的第 1 叶位中,儿茶素单体 EGCG 高于其他 3 种儿茶素单体含量,具体遵循以下的顺序:EGCG>EGC>ECG>EC。总之,4 种儿茶素单体和 CAF 含量随茶树品种和叶位的变化而变化,而且品种和叶位对其含量的影响显著。

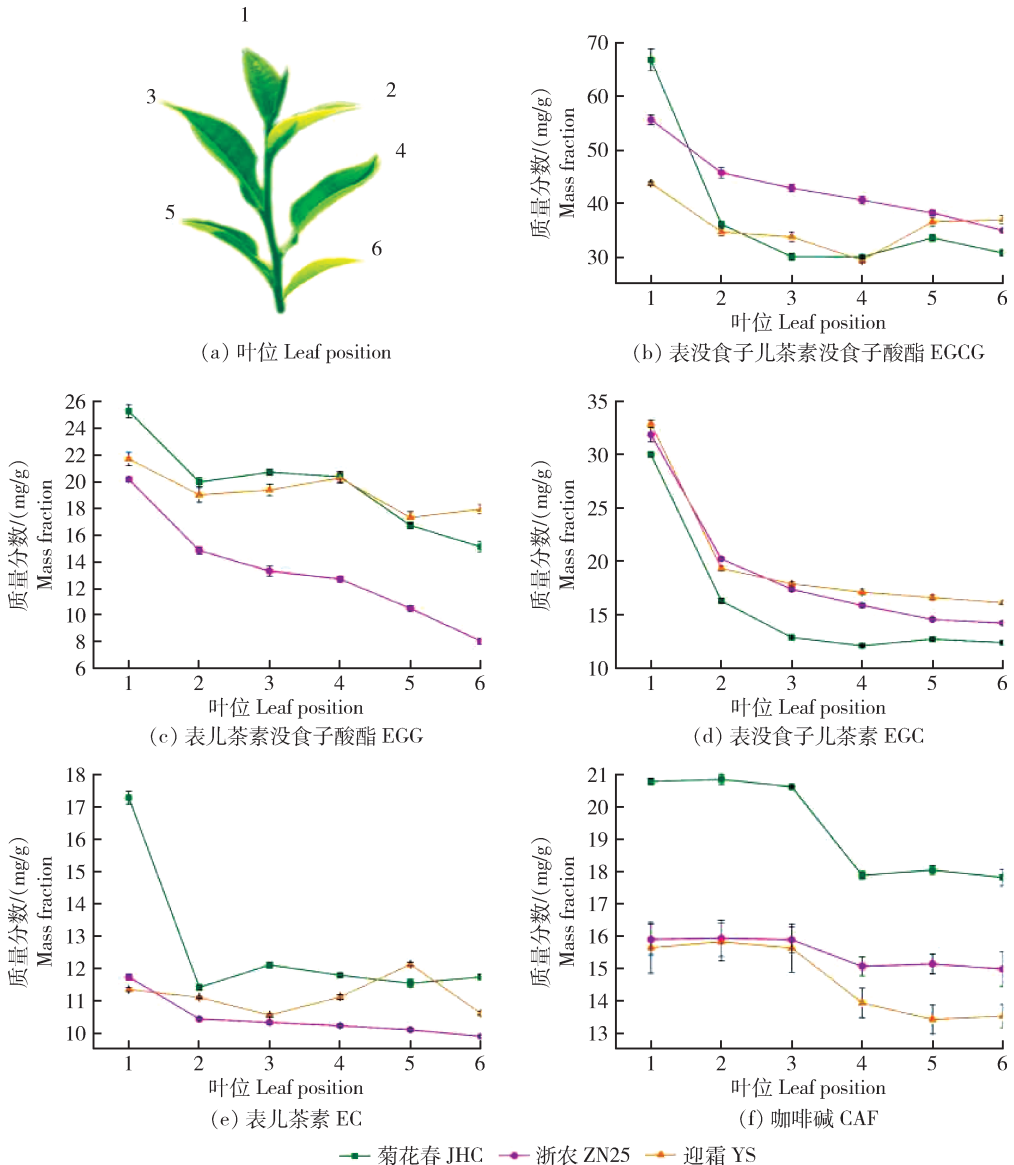


图 3 鲜茶叶的叶位分布及不同叶位对 4 种主要儿茶素和咖啡碱含量的影响

Fig. 3 The leaf position distribution of fresh tea leaves and the effects of different leaf positions on contents of four main catechins and caffeine

### 2.3 基于 PLS 的茶鲜叶中生化成分预测模型

PLS 方法经常被用于各种化学成分的分析中,

作为常规的回归方法,它的应用十分广泛。在本研究中,通过 Original、SPA 和 CARS 选择的特征波长

作为输入变量,然后使用 PLS 对 4 种主要儿茶素和 CAF 进行回归分析建模。通过 CARS 和 SPA 选择光谱变量后,变量数从 1 050 个减少到 20~75 个,

大大提高了检测效率。Original、SPA 和 CARS 挑选的特征波长分别作为输入变量构建 PLS 回归模型的性能比较如表 1 所示。

表 1 PLS 回归模型的性能比较

Table 1 Performance comparison of PLS regression model

成分 Component	输入变量 Input variable	变量维度 Variable dimension	建模决定 系数 $R_c^2$	建模均方 根误差 RMSEC	预测决定 系数 $R_p^2$	预测均方 根误差 RMSEP	残余预测 偏差 RPD
EGCG	Original	1 050	0.938 5	2.308 7	0.921 6	2.668 2	3.532 7
	SPA	26	0.938 3	2.311 5	0.910 1	2.857 1	3.299 1
	CARS	50	0.947 9	2.123 8	0.929 8	2.524 7	3.733 5
ECG	Original	1 050	0.902 7	1.310 7	0.892 2	1.392 1	3.047 9
	SPA	26	0.899 9	1.329 7	0.897 0	1.360 9	3.117 7
	CARS	70	0.903 7	1.304 2	0.899 9	1.341 6	3.162 6
EGC	Original	1 050	0.959 5	1.274 5	0.950 1	1.422 7	4.489 3
	SPA	29	0.964 6	1.191 2	0.951 4	1.402 6	4.553 6
	CARS	33	0.970 3	1.091 1	0.966 7	1.161 8	5.497 5
EC	Original	1 050	0.949 1	0.358 6	0.916 0	0.464 0	3.469 8
	SPA	29	0.935 1	0.405 0	0.910 3	0.479 6	3.356 9
	CARS	75	0.934 8	0.405 8	0.916 2	0.463 7	3.472 1
CAF	Original	1 050	0.882 2	0.986 2	0.891 6	0.951 8	3.034 2
	SPA	20	0.879 9	0.995 9	0.879 0	1.005 7	2.871 6
	CARS	59	0.953 3	0.620 4	0.953 1	0.625 9	4.614 1

注:EGCG 为表没食子儿茶素没食子酸酯,ECG 为表儿茶素没食子酸酯,EGC 为表没食子儿茶素,EC 为表儿茶素,CAF 为咖啡碱。下同。

Note: EGCG is epigallocatechin gallate, ECG is epicatechin gallate, EGC is epigallocatechin, EC is epicatechin, and CAF is caffeine. The same below.

对于 EGCG, EGC, EC 和 CAF, 基于 CARS 特征波长的 PLS 模型的性能优于基于 Original 的性能, 具有更高  $R_p^2$  值, 更低 RMSEP 值和更高 RPD 值。对于 ECG, CARS 和 SPA 两种选择特征波长的 PLS 模型与 Original 的模型保持相当的预测效果, 但是变量的数量也减少了许多。可以看出, CARS 和 SPA 不仅可以选择与物质成分相对应的特征波长, 还可以在在一定程度上提高模型的性能。总的来说, CARS 和 SPA 两种选择特征波长的方法和 PLS 的结合不仅可以大大简化模型, 还可以保持模型的稳健性和准确性。

## 2.4 基于 CNN 的茶鲜叶中生化成分预测模型

将深度学习用作建模方法, 使用 CNN 对 Original、SPA 和 CARS 提取的特征波长进行回归

分析建模, 来预测各个生化成分。表 2 显示了使用 Original 和提取的特征波长通过 CNN 回归模型对 4 种主要儿茶素和 CAF 的预测结果。可以看出, 无论是 Original 还是通过 SPA 或 CARS 选择的特征波长, EGCG, EGC, EC 和 CAF 的  $R_p^2$  都超过了 0.934 7, ECG 结果相对较差, 但通过 SPA 选择的特征波长得到的  $R_p^2$  和 RPD 分别达到了 0.932 2 和 3.621 2。综上, CNN 回归分析取得了很好的效果, 展现出对这 4 种主要儿茶素和 CAF 定量检测的可行性。

## 2.5 使用 CNN 提取特征

通过 CNN 提取特征(图 1), 然后使用 PLS 进行回归分析的建模效果见表 3。可以看出, 所有模型均获得了良好的效果, 所有成分最低的  $R_c^2$  都超过

表 2 CNN 回归模型的性能比较

Table 2 Performance comparison of CNN regression models

成分 Component	输入变量	变量维度	建模决定	建模均方	预测决定	预测均方	残余预测
	Input variable	Variable dimension	系数 $R_c^2$	根误差 RMSEC	系数 $R_p^2$	根误差 RMSEP	偏差 RPD
EGCG	Original	1 050	0.993 7	1.506 0	0.971 8	2.117 8	4.450 8
	SPA	26	0.972 9	1.242 2	0.934 7	2.517 5	3.744 1
	CARS	50	0.990 0	1.148 5	0.963 1	1.831 1	5.147 7
ECG	Original	1 050	0.990 4	1.226 4	0.860 0	1.581 0	2.683 7
	SPA	26	0.977 7	1.373 0	0.932 2	1.171 7	3.621 2
	CARS	70	0.982 3	0.698 7	0.894 6	1.314 9	3.226 8
EGC	Original	1 050	0.992 2	0.911 7	0.981 5	1.189 0	5.371 7
	SPA	29	0.993 3	0.514 5	0.983 5	0.822 0	7.770 0
	CARS	33	0.992 2	0.611 8	0.986 6	0.855 0	7.470 1
EC	Original	1 050	0.982 0	0.381 2	0.969 6	0.313 4	5.137 2
	SPA	29	0.986 9	0.237 2	0.945 8	0.379 9	4.237 9
	CARS	75	0.961 8	0.261 0	0.940 1	0.379 3	4.244 6
CAF	Original	1 050	0.976 0	0.557 1	0.953 1	0.700 9	4.120 4
	SPA	20	0.986 9	0.677 9	0.944 6	0.759 9	3.800 5
	CARS	59	0.953 6	0.548 6	0.940 7	0.698 6	4.133 9

表 3 通过 CNN 提取特征的 PLS 建模效果

Table 3 PLS modeling effect of extracting features through CNN

成分 Component	输入变量	变量维度	建模决定	建模均方	预测决定	预测均方	残余预测
	Input variable	Variable dimension	系数 $R_c^2$	根误差 RMSEC	系数 $R_p^2$	根误差 RMSEP	偏差 RPD
EGCG	Original	1 050	0.998 2	0.379 8	0.984 0	1.789 1	5.268 5
	SPA	26	0.998 1	0.576 8	0.986 5	1.619 1	5.821 7
	CARS	50	0.998 5	0.512 6	0.984 9	1.736 2	5.429 1
ECG	Original	1 050	0.999 6	0.125 0	0.923 8	1.630 9	2.601 6
	SPA	26	0.999 2	0.164 9	0.952 8	1.293 4	3.280 5
	CARS	70	0.998 9	0.192 7	0.899 2	1.874 8	2.263 1
EGC	Original	1 050	0.999 2	0.256 9	0.987 9	1.047 3	6.098 5
	SPA	29	0.999 1	0.269 3	0.991 1	0.881 3	7.247 2
	CARS	33	0.999 1	0.261 4	0.994 4	0.677 7	9.424 5
EC	Original	1 050	0.997 8	0.104 3	0.966 0	0.436 1	3.691 8
	SPA	29	0.998 4	0.088 7	0.966 5	0.440 1	3.658 2
	CARS	75	0.998 3	0.091 7	0.981 6	0.344 1	4.678 8
CAF	Original	1 050	0.989 0	0.300 7	0.985 4	0.841 9	3.430 3
	SPA	20	0.995 5	0.272 8	0.972 2	0.711 6	4.058 4
	CARS	59	0.995 0	0.286 1	0.950 8	0.911 9	3.167 0

了 0.989 0, 并且 EGCG, EGC, EC 和 CAF 最低的  $R_p^2$  和 RPD 均分别超过了 0.950 8 和 3.167 0; ECG 建模效果较差, 最低的  $R_p^2$  为 0.899 2, RPD 值超过了 2.263 1, 总体上建模效果较好。表明通过 CNN 提取的特征可以很好的用于回归分析中, 并能取得理想的建模效果。

## 2.6 传统方法与深度学习方法之间的比较

如表 1 和表 2 所示, 4 种主要儿茶素和 CAF 含量预测的常规方法和深度学习方法均具有良好的性能。将 CNN 用于回归建模时, 使用 Original 和选定的波长进行 CNN 建模可获得比 PLS 模型更好或接近的性能。这说明, 深度学习具有巨大的潜力, 可以对 4 种主要儿茶素和 CAF 的可见近红外光谱进行回归分析。此外, 通过 CNN 提取特征也获得了很好的建模效果(表 3), 深度学习提取特征具有较大优势, 并且基于 SPA 和 CARS 方法选择特征波长后的模型性能优于 Original, 这 2 种特征选择波长方法结合深度学习可以提高模型的准确性。深度学习作为建模和特征提取方法可用于回归问题的分析。

## 3 讨论

在鲜叶内含物统计分析中可以发现, ‘JHC’ 和 ‘ZN25’ 的 EGCG、EGC、ECG 和 EC 在第 1 叶位的含量高于其他 5 个叶位, ‘YS’ 的 EGCG、EGC 和 ECG 在第 1 叶位的含量最高, 每个品种茶叶的 CAF 从第 1 叶位到第 3 叶位含量基本一致, 第 4 叶位到第 6 叶位含量则较低, 这与 Lin 等<sup>[30]</sup> 和 Turkmen 等<sup>[31]</sup> 等的研究结果一致, 茶树中的生化物质具有向上积累的趋势并且大部分储存在嫩叶中, 在较老的叶子中很少。王峰等<sup>[32]</sup> 在研究中也发现, 不同品种的茶鲜叶 6 个叶位的儿茶素各个单体含量变化情况大致相同, 从第 1 叶位至第 6 叶位依次递减。以上结果表明, 茶树第 1 叶位叶片可为优质茶叶提供材料。

本研究通过深度学习方法对 4 种儿茶素和 CAF 含量进行建模预测, 获得了比 PLS 模型更好或接近的性能。深度学习非常适合大量的数据集, 本文只有 161 个样本, 这限制了深度学习的使用。但在本研究中, 将 CNN 用于建模, 各个内含物最优模型的  $R_p^2$  均高于 0.932 2; 将 CNN 用于提取特征, EGCG、ECG、EC 和 CAF 最优模型的  $R_p^2$  均高于

0.950 8。Li 等<sup>[33]</sup> 基于红外光谱技术建立的茶多酚最优回归模型验证相关性为 0.950 9, 均方根误差为 1.027 7, 能方便有效的测定 14 个茶树品种中茶多酚含量。周小芬<sup>[34]</sup> 采用 NIRS-PLS 法, 对不同光谱变量数和光谱预处理方法进行优化选择, 建立了大佛龙井茶 CAF 定量分析模型, 主成分因子数为 7 时, 校正相关系数和预测相关系数分别为 0.964 8 和 0.953 6, 模型预测效果较好。王玉霞等<sup>[35]</sup> 建立了绿茶茶汤中茶多酚和 CAF 的定量分析模型, 茶多酚交叉检验和预测模型的决定系数分别为 0.992 1 和 0.988 9, CAF 交叉检验和预测模型的决定系数分别为 0.952 7 和 0.962 3。赵峰等<sup>[36]</sup> 采用 TQ Analyst 8.5 光谱分析计量学软件, 比较不同数据预处理方法的建模结果, 优选并成功建立了武夷岩茶茶多酚和 CAF 的定量分析模型, 预测模型相关系数分别为 0.742 7 和 0.860 1。本研究基于深度学习方法获得了比以上研究更好或接近的模型性能, 样本集的扩增更将扩大深度学习方法的优势。

## 4 结论

本研究使用 PLS 方法对化学含量进行回归分析, 并通过 SPA 和 CARS 方法简化建立的模型, 然后将 CNN 用于建模和特征提取, 来预测 4 种主要儿茶素和 CAF。结果表明, 深度学习方法可以获得更好的模型效果, 展现出巨大的优势。此外, 试验是在没有任何处理的情况下对新鲜茶叶进行无损快速光谱扫描, 实现了儿茶素和 CAF 的快速和无损测定, 为 Vis/NIR 光谱在茶园精细化管理和嫩叶智能采摘中的应用奠定了基础。

## 参考文献 References

- [1] Liu L, Fan Y, Fu H Y, Chen F, Ni C, Wang J X, Yin Q B, Mu Q L, Yang T M, She Y B. “Turn-off” fluorescent sensor for highly sensitive and specific simultaneous recognition of 29 famous green teas based on quantum dots combined with chemometrics[J]. *Analytica Chimica Acta*, 2017, 963: 119-128
- [2] Le Gall G, Colquhoun I J, Defernez M. Metabolite profiling Using 1H NMR spectroscopy for quality assessment of green tea, *Camellia sinensis* (L) [J]. *Journal of Agricultural and Food Chemistry*, 2004, 52(4): 692-700
- [3] Zaveri N T. Green tea and its polyphenolic catechins: medicinal uses in cancer and noncancer applications[J]. *Life*



- Sciences, 2006, 78(18): 2073-2080
- [4] Khan N, Mukhtar H. Tea polyphenols for health promotion [J]. *Life Sciences*, 2007, 81(7): 519-533
- [5] Singh B N, Shankar S, Srivastava R K. Green tea catechin, epigallocatechin-3-gallate (EGCG): mechanisms, perspectives and clinical applications [J]. *Biochemical Pharmacology*, 2011, 82(12): 1807-1821
- [6] Bradfield A E, Penny M, Wright W B. The catechins of green tea[J]. *Journal of the Acoustical Society of America*, 1947, 103(5): 3-5
- [7] Dutta R, Stein A, Bhagat R M. Integrating satellite images and spectroscopy to measuring green and black tea quality[J]. *Food Chemistry*, 2011, 127(2): 866-874
- [8] Diniz P H G D, Barbosa M F, de Melo Milanez K D T, Pistonesi M F, de Araújo M C U. Using UV-Vis spectroscopy for simultaneous geographical and varietal classification of tea infusions simulating a home-made tea cup[J]. *Food Chemistry*, 2016, 192: 374-379
- [9] Zhou R Q, Li X L, He Y, Jin J J. Determination of catechins and caffeine content in tea (*Camellia Sinensis* L) leaves at different positions by fourier-transform infrared spectroscopy [J]. *Transactions of the ASABE*, 2018, 61(4): 1221-1230
- [10] 任明强, 赵宾, 赵国宣, 陈旭晖. 不同叶位新梢绿茶的品质及其影响因素探讨[J]. *贵州农业科学*, 2010, 38(12): 77-79
- Ren M Q, Zhao B, Zhao G X, Chen X H. Quality of green tea from different leaf position on new shoots and its influence factors[J]. *Guizhou Agricultural Sciences*, 2010, 38(12): 77-79 (in Chinese)
- [11] El-Shahawi M S, Hamza A, Bahaffi S O, Al-Sibaai A A, Abduljabbar T N. Analysis of some selected catechins and caffeine in green tea by high performance liquid chromatography [J]. *Food Chemistry*, 2012, 134(4): 2268-2275
- [12] Chen C N, Liang C M, Lai J R, Tsai Y J, Tsay J S, Lin J K. Capillary electrophoretic determination of theanine, caffeine, and catechins in fresh tea leaves and oolong tea and their effects on rat neurosphere adhesion and migration[J]. *Journal of Agricultural and Food Chemistry*, 2003, 51(25): 7495-7503
- [13] Lee B L, Ong C N. Comparative analysis of tea catechins and theaflavins by high-performance liquid chromatography and capillary electrophoresis[J]. *Journal of Chromatography A*, 2000, 881(1/2): 439-447
- [14] Luybaert J, Zhang M H, Massart D L. Feasibility study for the use of near infrared spectroscopy in the qualitative and quantitative analysis of green tea, *Camellia sinensis* (L)[J]. *Analytica Chimica Acta*, 2003, 478(2): 303-312
- [15] Li X L, He Y, Wu C Q, Sun D W. Nondestructive measurement and fingerprint analysis of soluble solid content of tea soft drink based on Vis/NIR spectroscopy[J]. *Journal of Food Engineering*, 2007, 82(3): 316-323
- [16] Diniz P H G D, Gomes A A, Pistonesi M F, Band B S F, Araújo M C U. Simultaneous classification of teas according to their varieties and geographical origins by using NIR spectroscopy and SPA-LDA[J]. *Food Analytical Methods*, 2014, 7(8): 1712-1718
- [17] 周健, 成浩, 王丽鸳. 近红外技术在茶叶上的研究进展[J]. *茶叶科学*, 2008, 28(4): 294-300
- Zhou J, Cheng H, Wang L Y. Recent advance on the application of near-infrared spectroscopy in tea[J]. *Journal of Tea Science*, 2008, 28(4): 294-300 (in Chinese)
- [18] 胡茶根, 赵红霞, 边文亮. 近红外技术在茶叶快速无损检测方面的研究与应用[J]. *食品科学*, 2007, 28(10): 638-643
- Hu C G, Zhao H X, Bian W L. Research and applications of near infrared spectroscopy in tea[J]. *Food Science*, 2007, 28(10): 638-641 (in Chinese)
- [19] Cen H, He Y. Theory and application of near infrared reflectance spectroscopy in determination of food quality[J]. *Trends in Food Science & Technology*, 2007, 18: 72-83
- [20] Ahn D, Choi J Y, Kim H C, Cho J S, Moon K D, Park T. Estimating the composition of food nutrients from hyperspectral signals based on deep neural networks [J]. *Sensors*, 2019, 19(7): 1560
- [21] Yu X J, Wang J P, Wen S T, Yang J Q, Zhang F F. A deep learning based feature extraction method on hyperspectral images for nondestructive prediction of TVB-N content in Pacific white shrimp (*Litopenaeus vannamei*) [J]. *Biosystems Engineering*, 2019, 178: 244-255
- [22] Li X L, Jin J J, Sun C J, Ye D P, Liu Y F. Simultaneous determination of six main types of lipid-soluble pigments in green tea by visible and near-infrared spectroscopy[J]. *Food Chemistry*, 2019, 270: 236-242
- [23] Kamruzzaman M, Makino Y, Oshita S. Parsimonious model development for real-time monitoring of moisture in red meat using hyperspectral imaging[J]. *Food Chemistry*, 2016, 196: 1084-1091
- [24] Fan W, Shan Y, Li G Y, Lv H, Li H D, Liang Y Z. Application of competitive adaptive reweighted sampling method to determine effective wavelengths for prediction of total acid of vinegar[J]. *Food Analytical Methods*, 2012, 5(3): 585-590
- [25] Soares S, Gomes A A. The successive projections algorithm [J]. *Trends in Analytical Chemistry*, 2013, 42: 84-98
- [26] Qiu Z J, Chen J, Zhao Y Y, Zhu S S. Variety identification of single rice seed using hyperspectral imaging combined with

- convolutional neural network[J]. *Applied Sciences*, 2018, 8(2): 212
- [27] Li X L, Sun C J, Zhou B X, He Y. Determination of hemicellulose, cellulose and lignin in moso bamboo by near infrared spectroscopy[J]. *Scientific Reports*, 2015, 5: 17210
- [28] Workman J, Weyer L. *Practical Guide to Interpretive Near-infrared Spectroscopy*[M]. Landenberg: CRC Press, 2007
- [29] Xiong C, Liu C, Pan W, Ma F, Xiong C, Qi L, Chen F, Lu X, Yang J, Zheng L. Non-destructive determination of total polyphenols content and classification of storage periods of Iron Buddha tea using multispectral imaging system [J]. *Food Chemistry*, 2015, 176: 130-136
- [30] Lin Y S, Tsai Y J, Tsay J S, Lin J K. Factors affecting the levels of tea polyphenols and caffeine in tea leaves[J]. *Journal of Agricultural and Food Chemistry*, 2003, 51(7): 1864-1873
- [31] Turkmen N, Sari F, VeliOglu Y S. Factors affecting polyphenol content and composition of fresh and processed tea leaves[J]. *Akademik Gida*, 2009, 7: 29-40
- [32] 王峰, 陈玉真, 王秀萍, 尤志明, 陈常颂. 茶树不同叶位叶片功能性状与光合特性研究[J]. *茶叶科学*, 2016, 36(1): 77-84  
Wang F, Chen Y Z, Wang X P, You Z M, Chen C S. Leaf functional and photosynthetic characteristics in different leaves positions of tea plant[J]. *Journal of Tea Science*, 2016, 36(1): 77-84 (in Chinese)
- [33] Li X L, Sun C J, Luo L B, He Y. Determination of tea polyphenols content by infrared spectroscopy coupled with iPLS and random frog techniques [J]. *Computers and Electronics in Agriculture*, 2015, 112: 28-35
- [34] 周小芬. 基于近红外的大佛龙井茶品质评价研究[D]. 北京: 中国农业科学院, 2012  
Zhou X F. Study on quality evaluation of Dafo Longjing tea based on near infrared spectroscopy [D]. Beijing: Chinese Academy of Agricultural Sciences, 2012 (in Chinese)
- [35] 王玉霞, 徐荣荣, 任广鑫, 王胜鹏, 宛晓春, 张正竹. 绿茶茶汤中主要品质成分近红外定量分析模型的建立[J]. *茶叶科学*, 2011, 31(4): 355-361  
Wang Y X, Xu R R, Ren G X, Wang S P, Wan X C, Zhang Z Z. Establishment of the predictive models for quantitative analysis of major components in green tea infusion by near infrared spectroscopy (NIRS) [J]. *Journal of Tea Science*, 2011, 31(4): 355-361 (in Chinese)
- [36] 赵峰, 林河通, 杨江帆, 叶乃兴, 俞金朋. 基于近红外光谱的武夷岩茶品质成分在线检测[J]. *农业工程学报*, 2014, 30(2): 269-277  
Zhao F, Lin H T, Yang J F, Ye N X, Yu J P. Online quantitative determination of Wuyi rock tea quality compounds by near infrared spectroscopy[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2014, 30(2): 269-277 (in Chinese)

责任编辑: 董金波