

基于近红外光谱的英红九号红茶快速定级方法

梁建华^{1,2}, 郭嘉明², 夏红玲¹, 马成英¹, 胡海涛³, 乔小燕¹

(1.广东省农业科学院 茶叶研究所/广东省茶树资源创新利用重点实验室, 广州 510640;
2.华南农业大学 工程学院, 广州 510642; 3.广东鸿雁茶业有限公司, 广东 英德 513042)

摘要: **目的** 以英红九号红茶为研究对象, 提出一种基于近红外光谱的红茶质量快速定级方法。**方法** 首先使用湿化学法对英红九号红茶进行内质成分含量检测, 并通过感官审评对参试红茶进行定级, 基于内质成分含量建立英红九号红茶定级模型, 然后利用近红外光谱构建红茶内质成分的定量模型, 以快速预测英红九号红茶的内质成分含量。将内质成分含量预测值输入定级模型, 以预测英红九号红茶的质量等级。**结果** 建立了红茶茶多酚、可溶性糖、游离氨基酸和咖啡碱4个内质成分的偏最小二乘法定量模型, 其测试集的决定系数分别为0.974 5、0.887 6、0.963 6、0.860 6, 基于感官审评和内质成分的随机森林定级模型测试集的准确率为90.48%。**结论** 为红茶质量快速定级提供了一种可行方案, 增强了基于近红外光谱的红茶定级方法的解释力。

关键词: 近红外光谱; 红茶; 定级; 内质成分

中图分类号: TS207.3; O657.33 文献标识码: A 文章编号: 1001-3563(2023)13-0157-09

DOI: 10.19554/j.cnki.1001-3563.2023.13.019

Rapid Grading Method of Black Tea 'Yinghong 9' Based on Near-infrared Spectroscopy

LIANG Jian-hua^{1,2}, GUO Jia-ming², XIA Hong-ling¹, MA Cheng-ying¹, HU Hai-tao³, QIAO Xiao-yan¹

(1. Tea Research Institute, Guangdong Academy of Agricultural Science/Guangdong Provincial Key Laboratory of Tea Plant Resources Innovation & Utilization, Guangzhou 510640, China; 2. College of Engineering, South China Agricultural University, Guangzhou 510642, China; 3. Guangdong Hongyan Tea Industry Co., Ltd., Guangdong Yingde 513042, China)

ABSTRACT: The tea production is developing towards large-scale currently, and the quality evaluation of tea after processing still relies on highly subjective artificial sensory evaluation, which is not suitable for the large-scale development of tea. Near-infrared spectroscopy (NIRS) has rich structural and compositional information, which is suitable for the detection of physicochemical parameters of hydrogen-containing organic substances. So it is widely used in the detection of biochemistry components of tea and the classification, such as authenticity discrimination and origin traceability. The work aims to take 'Yinghong 9' black tea as the research object, and propose a rapid grading method for tea quality based on NIRS. Firstly, a total of 42 samples of black tea processed from the fresh tea leaves of various grades were collected, a sub-sample was taken from each sample and ground into powder. A NIR spectrometer was used to scan tea

收稿日期: 2023-02-02

基金项目: 广东省乡村振兴战略专项资金(农业科技能力提升)(403-2018-XMZC-0002-90); 广东省农业科学院“中青年学科带头人”培养项目(R2020PY-JX016)

作者简介: 梁建华(1997—), 硕士生, 主攻农业机械。

通信作者: 乔小燕(1982—), 女, 博士, 副研究员, 主要研究方向为茶叶加工。

powder to collect the spectrum of each sample. Secondly, quantitative models for the biochemistry components were constructed based on NIRS to gain the biochemistry component information of black tea. Thirdly, 5 professional tea tasters were invited to conduct sensory evaluation on all samples. Based on the opinions of the tea tasters, the quality grade of tea samples were determined. Finally, the relationship between sensory evaluation results and biochemistry components were established to achieve the quality grading of 'Yinghong 9' black tea. In particular, when establishing the grading model, only the black tea processed from the second grade fresh leaves was selected and divided into three grades according to the sensory evaluation results. The quantitative models of four biochemistry components including tea polyphenol, soluble sugar, free amino acid and caffeine in black tea were established while these four quantitative models were preprocessed by combination data correction and normalization to reduce noise, drift as well as other interference and improve the difference between samples. These quantitative models were uniformly built using Partial Least Squares algorithm after using Genetic Algorithm, Successive Projections Algorithm, Variable Combination Population Analysis combined with Genetic Algorithm and other algorithms respectively to extract features. In order to ensure the reliability and stability of the model, Kennard-Stone algorithm was used to divide the samples into calibration set and test set before modeling, and K-fold verification was used in the modeling process. The principal components of the four quantitative models were all less than 10. The coefficients of determination on calibration set were tea polyphenol 0.974 5, soluble sugar 0.887 6, free amino acid 0.963 6 and caffeine 0.860 6 and the Root Mean Squared Error were 0.630 0, 0.298 3, 0.045 6, 0.162 6, respectively. The grading model based on sensory evaluation and biochemistry components had an accuracy of over 85%, which was built using Random Forest algorithm with 35 trees. The research results provide a feasible scheme for rapid grading of processed black tea based on specially graded fresh tea leaves, and effectively improve the interpretability of black tea grading.

KEY WORDS: near-infrared spectroscopy; black tea; grade; biochemistry components

英德是广东的传统优势红茶产区,截至2021年底,英德市的茶园面积为113.5 km²,全市全年干茶产量超过1.35万t。英德红茶产业一直向着标准化、规模化发展,以英红九号红茶为主要产品。企业通常会根据芽叶比例确定茶青等级,但不同批次的同等级茶青间有所差异,且在实际生产过程中温湿度等环境因素对生产有着重要影响。在内外诸多因素的干扰下,同一等级茶青在标准化生产线上生产的茶叶质量参差不齐,生产加工后茶叶的质量评定成为一项重要生产要求。通过有效地对生产质量进行把控,为后续的茶叶精制、定价包装等流程提供依据。目前,生产线所用评定茶叶质量的方法为人工感官审评,感官审评极度依赖人工,且具有较强的主观性和模糊性^[1],并不适应当前茶叶生产规模化发展的要求。茶叶现有的快速质量评价方法主要围绕光谱、机器视觉、电化学等技术手段展开^[2-3]。其中,近红外光谱技术具有无损和快捷的优点,最早应用于内质成分的检测,可有效克服传统成分理化分析方法费时、耗力的缺点,广泛应用于茶叶的品质研究,如茶叶的真实性判别及产地溯源等^[4-6]。目前,已开展基于近红外光谱对茶叶质量定级的研究,部分学者通过选择近红外光谱中的特征变量建立定性模型,实现茶叶的质量定级,如Ren等^[7]建立了祁门红茶不同等级的模型。也有学者通过近红外光谱与感官审评得分建立定量模型,以确定茶叶的质量得分,如Zuo等^[8]建立了贵州不同海拔

绿茶感官审评得分的定量模型;周小芬等^[9]建立了大佛龙井茶的品质评价定量模型。这些模型基于近红外光谱对茶叶进行快速质量评价,但越过了决定茶叶质量的内质成分,未能充分解释茶叶内质成分、茶叶质量和近红外光谱三者间的关系。在实际生产中,茶企业会控制茶鲜叶的品种、芽叶比例、嫩度等因素,以确保茶鲜叶的质量相对稳定,进而在很大程度上减小了所产茶叶质量的区分度。已有研究表明,针对不同品种、不同海拔、不同等级原料的茶叶,样本的组间区分度远大于生产线上所加工的不同质量等级茶叶区分度。

有研究^[10]表明,游离氨基酸、茶多酚、可溶性糖和咖啡碱作为红茶的主要成分,其含量与红茶的感官品质显著相关,并且这些内质成分可使用近红外光谱技术进行准确检测^[11],因此选用这些对红茶品质具有重要影响的内质成分作为评价同一等级茶鲜叶所加工红茶品质的变量。由此,文中提出以英红九号红茶所制茶叶为研究对象,利用近红外光谱仪收集茶叶样品的光谱,建立近红外光谱与红茶中内质成分间的定量模型,通过感官审评确认红茶的质量等级,并建立感官审评定级结果与内质成分间的定级模型,利用近红外光谱预测红茶内的内质成分信息,并基于此信息通过定级模型获知英红九号红茶的质量等级。此方法解释了茶叶内质成分、茶叶品质和近红外光谱三者间的直接关系,拟为生产线上的茶叶质

量定级提供科学依据。

1 实验

1.1 材料

实验所用茶叶由广东省英德市广东鸿雁茶业有限公司、广东英九庄园绿色产业发展有限公司、广东石门山生态科技茶业有限公司 3 家企业提供, 每份约 250 g。茶叶的生产时间为 4—5 月, 在生产前已确认茶青等级。其中, 一级、二级、三级茶青均为以一芽二叶为主的茶鲜叶, 详细情况如表 1 所示, 各样品为企业不同批次所制茶叶。

表 1 样本详情
Tab.1 Sample details

茶青等级	数量/份	芽叶占比
特级	3	100% (一芽一叶)
一级	13	>90% (一芽二叶)
二级	21	80%~90% (一芽二叶)
三级	5	<80% (一芽二叶)

1.2 方法

实验旨在通过近红外光谱技术获知红茶的内质成分信息, 并根据获知信息实现红茶的质量定级, 具体流程如图 1 所示。

1.3 内质成分检测与茶叶品质评定

参照茚三酮比色法检测试验材料的游离氨基酸含量^[12], 采用福林酚法检测试验材料中的茶多酚含

量^[13], 采用萘酚比色法检测试验材料中的可溶性糖^[14], 采用 HPLC 法检测试验材料的咖啡碱含量^[15]。每个样品重复测定 3 次。

按照《茶叶感官审评方法》^[16], 邀请 5 位熟悉茶叶市场的专业品茶师, 根据内质评审法对红茶品质进行质量综合等级评定。

1.4 光谱数据的获取

采用二级茶青制作红茶样本, 每份制作 3 个子样本, 制作红茶子样本共计 63 份。将各红茶样本研磨成粉末, 并通过网格筛, 每次称量 13 g 茶粉, 均匀铺于检测容器中。利用傅里叶光谱仪 (Thermo Fisher Scientific inc 5225 Verona Rd, Madison, WI) 扫描粉末样本, 以获取样本的光谱数据。设置仪器光谱的扫描范围为 4 000~10 000 cm^{-1} , 分辨率为 8 cm^{-1} , 扫描 64 次, 采点数量为 1 557。近红外光谱仪采用积分球采样模式, 每份样本扫描 3 遍, 使用平均光谱作为样本光谱, 一共采集 84 份光谱。

1.5 分析与建模软件

使用 Origin 2021 对内质成分含量进行分析。使用 Matlab R2021a 软件建立近红外定量模型和定级模型。

2 内质成分预测模型与定级模型构建

2.1 预处理方法

为了减小光谱数据受外界因素干扰而产生的噪声, 对光谱数据统一采用卷积平滑 (Savitzky-Golay, SG) 算法进行降噪, 平滑窗口的宽度为 17, 多项式的阶数为 3^[17]。同时, 为了去除光谱收集时因散射

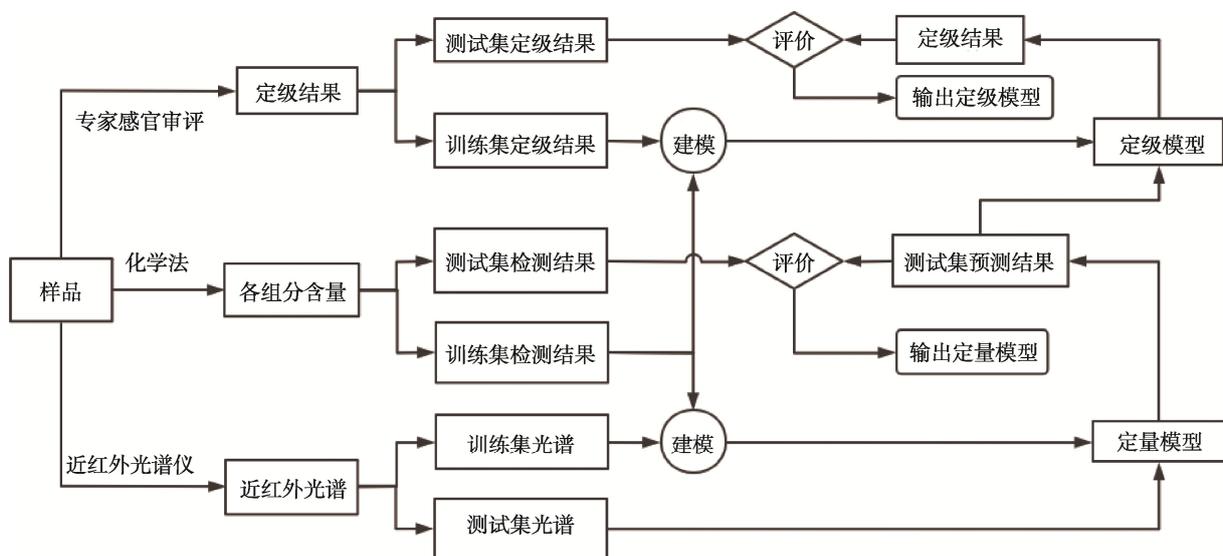


图 1 实验流程
Fig.1 Experiment process

光、颗粒不均等因素导致的光谱漂移,以及增加样本间的差异性^[18],在实验中对不同茶叶的内质成分进行建模前,采用不同的数据校正和标准化组合进行预处理,尽量减小预处理弱化关键变量的可能性。其中,数据校正方法包括多元散射校正(Multiplicative Scatter Correction, MSC)、标准正态变量变换(Standard Normal Variate Transformation, SNV)和迭代自适应加权惩罚最小二乘法(Adaptive Iterative Re-weighted Penalized Least Squares, airPLS)^[19]。标准化方法包括自标度化(autoscaling)、中心化(center)、最小-最大归一化(minmax),各标准化的定义见式(1)~(3)。

$$\text{Autoscaling } X=(x-\text{mean}(x)^a)/\text{std}(x)^b \quad (1)$$

$$\text{Center } X=x-\text{mean}(x)^a \quad (2)$$

$$\text{Minmax } X=(x-\text{min}(x)^b)/(\text{max}(x)^a-\text{min}(x)^b) \quad (3)$$

式中: x 为样本某波长对应的吸收度; $\text{mean}(x)$ 为训练集某波长对应吸收度的均值; $\text{std}(x)$ 为训练集某波长对应吸收度的标准差; $\text{max}(x)$ 、 $\text{min}(x)$ 分别为训练集某波长对应吸收度的最大值、最小值。测试集的标准化处理调用训练集中的 a 、 b 参数,而非以测试集自身产生的 a 、 b 参数。

2.2 数据集划分

采用KS(Kennard-Stone)算法^[20],将样本光谱集以7:3的比例划分为训练集与测试集。采用KS算法,先计算在 m 维空间中任意2个样本之间的真实距离,即欧氏距离,选择欧氏距离最远的2个样本进入训练集。然后,通过计算剩下的每个样品到训练集内每个已知样品的欧式距离,找到拥有最大距离的待选样本,并放入训练集,反复此操作,直到训练集数目满足要求。该方法使得训练集中的样本按照空间距离分布均匀,能有效提高训练样本的代表性。

2.3 特征选择

由于光谱的数据维度较高,其中包含了大量无关信息,不仅极大地延长了数据分析时间,还可能出现问题数据过拟合、模型泛化能力极差、维数灾难等问题^[21]。文中采用特征选择的方式,获取优选波长变量。特征选择变量基于偏最小二乘法(Partial Least Squares, PLS)拟合评价效果获取,在此次实验中使用的特征选择算法包括区间偏最小二乘法(interval PLS, iPLS)、连续投影法(Successive Projections Algorithm, SPA)^[22]、遗传算法(Genetic Algorithm, GA)、迭代变量子集优化(Iteratively Variable Subset Optimization, IVSO)、随机雾化法(Randomfog),以及基于GA、迭代的保持信息变量(Iteratively Retaining Informative Variables, IRIV)的变量组合总体分析(Variable Combination Population Analysis, VCPA—

GA、VCPA-IRIV)^[23-24]。各方法选择变量迭代的方式不同,如iPLS是将寻找的最佳变量区间作为特征变量;GA是从全集中挑选若干个最适波长段作为特征变量。针对不同的内质成分,需要选择合适的特征选择方式,以获取关键的优选特征。

2.4 模型训练和评价方法

基于特征选择采用PLS作为评价函数,利用PLS将训练集选择的特征变量和其对应的内质成分建立多元线性回归模型。使用K-fold验证方式进行交叉验证, $K=5$,防止PLS模型过拟合^[25-26]。针对测试集,采用与训练集相同的预处理方式获取优选特征变量,然后代入回归模型,得到测试集的对应内质成分预测含量。以湿化学法检测的4个内质成分含量为特征,以感官审评定级结果为输出,建立红茶质量定级模型。将测试样本光谱通过4个定量模型,得到测试样本的4个内质成分含量,进而通过定级模型获取预测等级。

定量模型评价指标主要有决定系数 R^2 、测试集决定系数 R_p^2 、建模集的均方差(Root Mean Square Error of Cross Validation, RMSECV)、测试集的均方差(Root Mean Square Error of Prediction, RMSEP),定级模型的评价指标为测试集的预测正确率。最终模型的性能依据测试集的预测结果决定^[27]。

3 结果与分析

3.1 内质成分分析及感官审评结果

采用各等级茶青所制红茶的内质成分含量检测结果如图2所示,每个分布框内的横线表示该等级茶青的内质成分含量(质量分数)均值。特级茶青,即一芽一叶所制红茶的游离氨基酸和咖啡碱含量分布于3.0%~4.0%、2.0%~2.5%,含量均值分别为3.37%、2.16%,高于以一芽二叶为主的加工红茶(一级、二级、三级);可溶性糖的含量(质量分数)较低,均值为4.26%,比一芽二叶所制红茶低约25%;茶多酚含量相当。在以一芽二叶为主的加工红茶中,茶多酚、咖啡碱、游离氨基酸含量分布总体随茶青等级的升高而逐级增加,各级茶青所制红茶的可溶性糖含量较接近,约为5.7%。所述4种内质成分在4个等级茶青所制红茶中的含量有较大差异,而实际生产中茶青等级在厂商收购茶青时已确定,因此茶叶的质量定级通常建立在特定等级茶青的基础上。

感官审评的结果如表2所示,绝大部分样品的质量等级基于茶青等级(特级、一级、二级、三级)决定,仅一级茶青出现了越级(降级)现象。英德生产的茶叶产品以二级茶青所制茶叶为主,占全年英德红

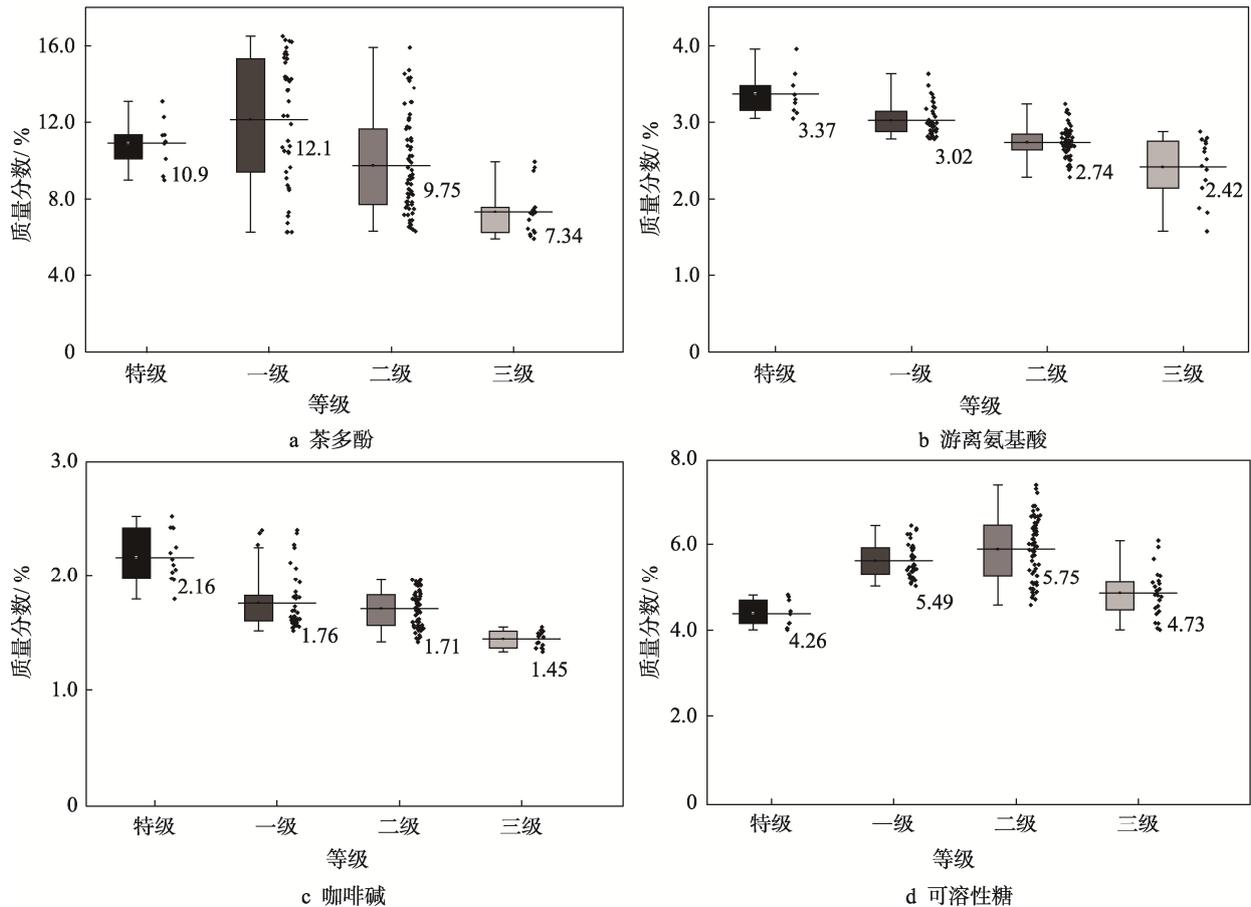


图 2 不同等级茶青所制红茶内质成分箱线图

Fig.2 Box diagram of contents of black tea made from fresh leaves of tea of different grades

茶产量的 90%以上,对二级茶青所制茶叶的质量定级具有产业意义。由此,从二级茶青所制茶叶出发,研究基于特定茶青等级所制茶叶的快速质量定级方法。

表 2 样品感官审评定级结果
Tab.2 Sensory evaluation grading results of samples

茶青等级	感官审评等级	数量	组内比例/%
特级	特级一等	2	66.7
	特级二等	1	33.3
一级	一级一等	6	46.1
	一级二等	5	38.5
	一级三等	1	7.7
	二级一等	1	7.7
二级	二级一等	5	23.8
	二级二等	9	42.9
	二级三等	7	33.3
三级	三级一等	2	40
	三级二等	2	40
	三级三等	1	20

3.2 原始光谱与预处理光谱

观测图 3 中的原始光谱可以发现,各样本光谱间具有高度的相似性,且光谱在吸收度上存在漂移。通过放大观测发现,光谱在不同区间均存在较严重的噪声,因此需对原始光谱进行预处理,图 3 展示了部分预处理后的光谱。如图 3b 所示,在采用 MSC 校正预处理后,光谱漂移现象明显减少,但样本光谱间仍存在高度相似性。为了突出样本间的差异,对光谱进行了标准化处理。图 3c 显示了 MSC+autoscaling 处理后的光谱,经处理后光谱间的差异将显著放大,但光谱中携带的噪声同样会被放大,因此在进行数据校正和标准化前,需进行降噪处理,这里采用 SG 卷积平滑方法。经过 SG+SMC+autoscaling 处理后,光谱噪声明显降低,对比图 3c、d 中的局部放大图可知,经平滑处理后光谱噪声被抑制,且光谱主成分信息被有效保留。

3.3 定量预测结果

考虑到各算法之间相互影响的不确定性,选择全局试验。通过将预处理方式(卷积平滑+校正+标准化)

与特征选择算法自由组合,并基于 PLS 建立回归模型,针对每类内质成分分别建立 63 (3×3×7) 个预测模型。为了观测预处理对模型的影响,开展不进行预处理的建模实验。同时,为了考量特征选择的作用,

进行了基于全光谱的建模实验。各内质成分最佳预测模型的方法及实验结果如表 3 所示,对比了不进行预处理但进行特征选择,以及进行预处理但不进行特征选择 (ALL) 建模的结果。

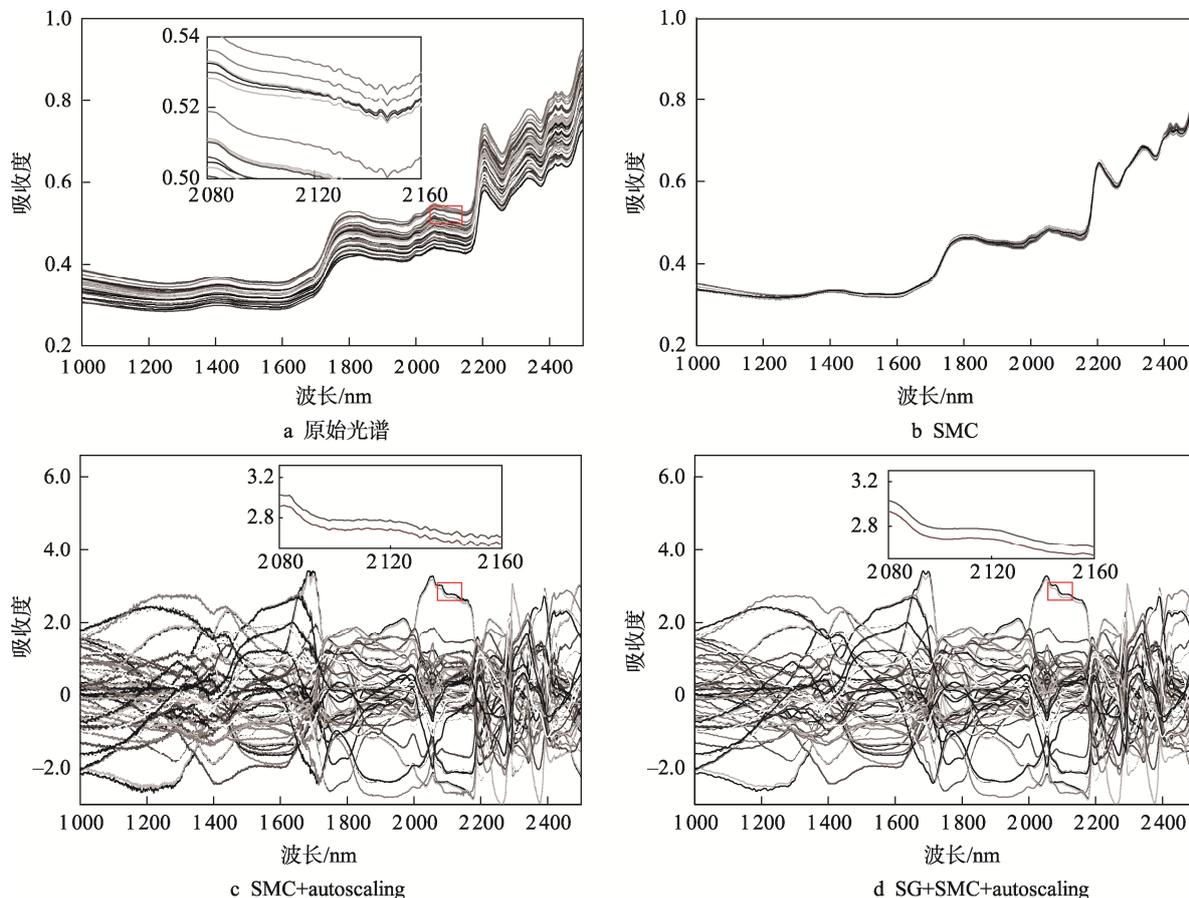


图 3 原始光谱图与预处理光谱
Fig.3 Diagram of raw and preprocessed spectra

表 3 部分定量模型及其结果
Tab.3 Partial quantitative models and their results

内质成分	预处理方式	特征选择	R^2	R_p^2	RMSECV	RMSEP	优选波长数量
茶多酚	SG+MSC+minmax	iPLS	0.960 3	0.947 5	0.695 9	0.880 5	155
	none	iPLS	0.942 3	0.926 5	0.966 4	1.086 9	155
	SG+MSC+Minmax+SG	ALL	0.988 2	0.974 5	0.382 2	0.630 0	1 557
可溶性糖	SG+airPLS+autoscaling	GA	0.918 1	0.858 9	0.381 1	0.402 6	15
	none	GA	0.916 8	0.887 6	0.294 6	0.298 3	15
	SG+airPLS+autoscaling+SG	ALL	0.937 2	0.509 1	0.181 3	0.849 7	1 557
游离氨基酸	SG+SNV+autoscaling	VPCA_GA	0.954 3	0.963 6	0.045 3	0.045 6	14
	none	VPCA_GA	0.959 6	0.878 8	0.043 5	0.074 9	81
	SG+SNV+autoscaling+SG	ALL	0.997 6	0.942 2	0.011 5	0.106 7	1 557
咖啡碱	SG+SNV+autoscaling	SPA	0.872 3	0.860 6	0.169 4	0.162 6	15
	none	SPA	0.439 0	0.47	0.323 1	0.295 0	15
	SG+SNV+autoscaling+SG	ALL	0.922 7	0.829 7	0.133 5	0.225 6	1 557

由实验结果可知, 茶多酚采用 SG+MSC+minmax+iPLS+PLS 算法建立预测模型的效果较佳, RMSEP 值为 0.880 5, R_p^2 为 0.947 5, 相较于不进行预处理, R^2 和 R_p^2 均有所提高, 均方残差也有所下降, 这表明经过预处理, 此模型的综合性能得到提高。采用各预处理方式后, 可溶性糖的结果 (最佳 R_p^2 为 0.858 9) 比不进行预处理的结果 (R_p^2 为 0.887 6) 稍差, 可能因实验中采用的预处理方式并不能很好地去除变量中的噪声等无关信息, 或是在预处理过程中有效信息被弱化。游离氨基酸采用 SG+SNV+autoscaling+VPCA_GA+PLS 建立的预测模型性能比采用其他算法组合建立的模型具有更高的稳定级和预测性能, 对比不进行预处理进行特征选择后建模, R_p^2 和 R^2 更稳定, R_p^2 更高, 这意味着此模型的泛化能力比不进行预处理的高, 且预测性能更好。在实验中, 建立咖啡碱的预测模型的最佳算法组合为 SG+SNV+autoscaling+SPA, R_p^2 为 0.860 6, 对比不进行预处理模型, 其性能得到巨大提升。在建模时, 采用正确的预处理方式很大程度地提升了预测模型的综合性能, 不匹配的预处理方式会导致有效信息被弱化, 以致特征选择效率降低。

在不进行特征选择而将完整的光谱作为有效特征建立的全光谱模型中, 可溶性糖、游离氨基酸、咖啡碱 3 种内含物模型在训练时验证得到的 R^2 较高, 且 RMSECV 值较小, 但使用测试集测试得到的 R_p^2 远低于训练集, RMSEP 值也明显增大, 出现了过拟合现象。说明利用全光谱模型的泛化能力较弱, 这并非意味着全光谱建模具有不可行性。在茶多酚建模实验中, 利用 PLS 基于全光谱建立的模型, 其预测表现相较于 SG+MSC+minmax+iPLS+PLS 模型, 其测

试集的 RMSEP 值降低了约 28%, 相关系数提升了约 3%。综上分析, 茶叶内质成分的预测模型采用的方案如表 4 所示。

表 4 各内质成分选用建模方法
Tab.4 Modelling method selected for each biochemistry component

内质成分	模型算法
茶多酚	SG+MSC+minmax+SG+ALL_PLS
可溶性糖	SG+GA+PLS
游离氨基酸	SNV+autoscaling+VPCA_GA+PLS
咖啡碱	SNV+autoscaling+SPA+PLS

3.4 定级模型

基于内质成分 (化学法测定) 与感官审评定级结果, 在 Matlab 软件的分学习器中寻找合适的分级模型, 验证方法为 K-折交叉验证, 设定 K=5。在该分类器中的分类模型包括决策树、判别分析、支持向量机、逻辑回归、最近邻、朴素贝叶斯、核逼近、集成和神经网络分类等分类器, 通过执行自动训练搜索最佳分类模型类型。最终选择随机森林法^[28]建立树的数量为 35 的红茶质量定级模型, 定级模型验证集的正确预测率为 85.71%。

利用随机抽样方法, 从二级茶青所制红茶子样本光谱中挑选 21 个光谱作为测试集, 将测试集数据通过定量模型, 获取样品内质成分预测含量。将内质成分预测含量输入定级模型, 得到测试集样本的质量等级, 其准确率为 90.48%, 测试集的定级结果如表 5 所示, 错误定级样品的内质成分信息如表 6 所示。

表 5 感官审评与模型定级结果
Tab.5 Grading Result of sensory evaluation and our model

定级方法	样本序号																				
	1	2*	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20*	21
感官审评	2	1	2	2	1	2	3	2	3	3	2	1	2	1	3	3	3	3	3	1	3
定级模型	2	2	2	2	1	2	3	2	3	3	2	1	2	1	3	3	3	3	3	2	3

注: 1、2、3 分别表示二级一等、二级二等、二级三等; *表示错误预测样本。

表 6 错误定级样品的内质成分信息
Tab.6 Biochemistry component information for mis-graded samples

错误定级样本 序号	方法	质量分数/%			
		茶多酚	可溶性糖	游离氨基酸	咖啡碱
2	化学法	8.80	2.92	1.92	2.45
	光谱预测	7.79	2.89	1.88	2.48
20	化学法	11.39	3.05	1.74	2.74
	光谱预测	9.61	3.13	1.72	2.79

由表6可知,错误的2个样本在使用光谱进行定量预测时,其茶多酚预测值与化学法测量值的偏差约为11%,其他内质成分的测量值与真实值较接近,偏差小于3%。预测结果表明,茶多酚对茶叶质量等级具有较大影响,也证明基于感官审评和内质成分的红茶定级方法具有可行性。

3.5 讨论

红茶中的有效成分多种多样,影响红茶品质的内质成分远不止4种。有研究者认为茶多酚中的酯型儿茶素与非酯型儿茶素间的含量比值也可作为评价红茶茶汤滋味的重要指标^[29],今后可进一步完善该方法,通过近红外光谱检测红茶或其他茶类中的有效成分,提高此方法的准确性和适用性。此外,对茶进行人工感官审评定级时,除内质成分外,还需考虑茶叶外观等因素^[30],这些因素无法采用近红外光谱表征。后续可进一步利用近红外光谱技术与其他技术(如机器视觉)相结合,以识别茶叶的外观质量,通过数据融合实现红茶快速精准的质量定级^[31]。

4 结语

以英红九号红茶为研究对象,根据红茶规模化发展的要求,从研究近红外光谱、内质成分与茶叶等级的关系出发,提出一种基于近红外光谱的红茶质量快速定级方法。利用近红外光谱仪收集红茶样品的光谱,通过卷积平滑与校正、标准化对光谱进行预处理后,对光谱进行特征选择,建立优选特征波长变量与红茶中内质成分间的定量模型。通过实验得到内质成分定量模型验证集的决定系数:茶多酚为0.974 5,可溶性糖为0.887 6,游离氨基酸为0.963 6,咖啡碱为0.860 6。通过感官审评确定红茶的质量等级,建立基于内质成分含量的红茶质量定级模型,其中使用随机森林算法搭建模型的性能较好。将光谱预测的红茶内质成分含量输入定级模型,其输出准确率超过85%,测试集中误判的2个样本说明,定级模型对定量模型的精度要求较高。此方法有效统一了近红外光谱、内质成分、感官审评定级之间的关系,为红茶加工的质量定级提供了一种可行方案。

参考文献:

- [1] CHEN Quan-sheng, ZHANG Dong-liang, PAN Wen-xiu, et al. Recent Developments of Green Analytical Techniques in Analysis of Tea's Quality and Nutrition[J]. Trends in Food Science & Technology, 2015, 43(1): 63-82.
- [2] 刘奇, 欧阳建, 刘昌伟, 等. 茶叶品质评价技术研究进展[J]. 茶叶科学, 2022, 42(3): 316-330.
- [3] LIU Qi, OUYANG Jian, LIU Chang-wei, et al. Research Progress of Tea Quality Evaluation Technology[J]. Journal of Tea Science, 2022, 42(3): 316-330.
- [4] HIDAYAT S N, TRIYANA K, FAUZAN I, et al. The Electronic Nose Coupled with Chemometric Tools for Discriminating the Quality of Black Tea Samples in Situ[J]. Chemosensors, 2019, 7(3): 29.
- [5] 宋彦, 汪小中, 赵磊, 等. 基于近红外光谱技术的眉茶拼配比例预测方法[J]. 农业工程学报, 2022, 38(2): 307-315.
- [6] SONG Yan, WANG Xiao-zhong, ZHAO Lei, et al. Predicting the Blending Ratio of Mee Tea Based on near Infrared Spectroscopy[J]. Transactions of the Chinese Society of Agricultural Engineering, 2022, 38(2): 307-315.
- [7] 王淑贤, 肖航, 杨振发, 等. 香精掺假普洱茶的近红外光谱检测[J]. 激光与光电子学进展, 2020, 57(23): 392-397.
- [8] WANG Shu-xian, XIAO Hang, YANG Zhen-fa, et al. Detection of Flavor Adulterated Pu'er Tea by Near-Infrared Spectroscopy[J]. Laser & Optoelectronics Progress, 2020, 57(23): 392-397.
- [9] 陈琦, 潘天红, 李鱼强, 等. 基于卷积神经网络和近红外光谱的太平猴魁茶产地鉴别分析[J]. 光谱学与光谱分析, 2021, 41(9): 2776-2781.
- [10] CHEN Qi, PAN Tian-hong, LI Yu-qiang, et al. Geographical Origin Discrimination of Taiping Houkui Tea Using Convolutional Neural Network and Near-Infrared Spectroscopy[J]. Spectroscopy and Spectral Analysis, 2021, 41(9): 2776-2781.
- [11] REN Guang-xin, WANG Yu-jie, NING Jing-ming, et al. Highly Identification of Keemun Black Tea Rank Based on Cognitive Spectroscopy: Near Infrared Spectroscopy Combined with Feature Variable Selection[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2020, 230: 118079.
- [12] ZUO Ya-min, TAN Gao-hao, XIANG Di, et al. Development of a Novel Green Tea Quality Roadmap and the Complex Sensory-Associated Characteristics Exploration Using Rapid Near-Infrared Spectroscopy Technology[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2021, 258: 119847.
- [13] 周小芬, 叶阳, 周竹定, 等. 基于近红外光谱法的大佛龙井茶品质评价研究[J]. 光谱学与光谱分析, 2012, 32(11): 2971-2975.
- [14] ZHOU Xiao-fen, YE Yang, ZHOU Zhu-ding, et al. Study on Quality Evaluation of Dafo Longjing Tea Based on near Infrared Spectroscopy[J]. Spectroscopy and Spectral Analysis, 2012, 32(11): 2971-2975.
- [15] 吴英. 万州燕山红茶鲜叶产量及品质动态变化研究

- [D]. 重庆: 重庆三峡学院, 2021: 46-50.
- WU Ying. Study on Dynamic Change of Fresh Leaf Yield and Quality of Yanshan Black Tea in Wanzhou[D]. Chongqing: Chongqing Three Gorges University, 2021: 46-50.
- [11] LIN Xiao-hui, SUN Da-wen. Recent Developments in Vibrational Spectroscopic Techniques for Tea Quality and Safety Analyses[J]. Trends in Food Science & Technology, 2020, 104: 163-176.
- [12] GB/T 8314—2013, 茶 游离氨基酸总量的测定[S]. GB/T 8314-2013, Tea—Determination of Free Amino Acids Content[S].
- [13] GB/T 8313—2018, 茶叶中茶多酚和儿茶素类含量的检测方法[S]. GB/T 8313-2018, Determination of Total Polyphenols and Catechins Content in Tea[S].
- [14] 傅博强, 谢明勇, 聂少平, 等. 茶叶中多糖含量的测定[J]. 食品科学, 2001, 22(11): 69-73. FU Bo-qiang, XIE Ming-yong, NIE Shao-ping, et al. Method Simplified in Assaying Tea Polysaccharide[J]. Food Science, 2001, 22(11): 69-73.
- [15] GB/T 8312—2013, 茶咖啡碱测定[S]. GB/T 8312-2013, Tea—Determination of Caffeine Content[S].
- [16] GB/T 23776—2018, 茶叶感官审评方法[S]. GB/T 23776-2018, Methodology for Sensory Evaluation of Tea[S].
- [17] DING Yu-han, YAN Yu-li, LI Jun, et al. Classification of Tea Quality Levels Using Near-Infrared Spectroscopy Based on CLPSO-SVM[J]. Foods, 2022, 11(11): 1658.
- [18] HUANG Yi-feng, DONG Wen-tao, SANAEIFAR A, et al. Development of Simple Identification Models for Four Main Catechins and Caffeine in Fresh Green Tea Leaf Based on Visible and Near-Infrared Spectroscopy[J]. Computers and Electronics in Agriculture, 2020, 173: 105388.
- [19] WEI Jia-jin, ZHU Chen, ZHANG Zhi-min, et al. Two-Stage Iteratively Reweighted Smoothing Splines for Baseline Correction[J]. Chemometrics and Intelligent Laboratory Systems, 2022, 227: 104606.
- [20] KENNARD R W, STONE L A. Computer Aided Design of Experiments[J]. Technometrics, 1969, 11(1): 137-148.
- [21] YUN Yong-huan, LI Hong-dong, DENG Bai-chuan, et al. An Overview of Variable Selection Methods in Multivariate Analysis of Near-Infrared Spectra[J]. TrAC Trends in Analytical Chemistry, 2019, 113: 102-115.
- [22] ARAÚJO M C U, SALDANHA T C B, GALVÃO R K H, et al. The Successive Projections Algorithm for Variable Selection in Spectroscopic Multicomponent Analysis[J]. Chemometrics and Intelligent Laboratory Systems, 2001, 57(2): 65-73.
- [23] LI Peng, MA Jun-chao, ZHONG Nan. Fourier Transform Near-Infrared Spectroscopy Coupled with Variable Selection Methods for Fast Determination of Salmon Fillets Storage Time[J]. Journal of Molecular Structure, 2022, 1264: 133223.
- [24] YUN Yong-huan, BIN Jun, LIU Dong-li, et al. A Hybrid Variable Selection Strategy Based on Continuous Shrinkage of Variable Space in Multivariate Calibration[J]. Analytica Chimica Acta, 2019, 1058: 58-69.
- [25] GEISSER S. The Predictive Sample Reuse Method with Applications[J]. Journal of the American Statistical Association, 1975, 70(350): 320-328.
- [26] DENG Bai-chuan, YUN Yong-huan, LIANG Yi-zeng, et al. A New Strategy to Prevent Over-Fitting in Partial Least Squares Models Based on Model Population Analysis[J]. Analytica Chimica Acta, 2015, 880: 32-41.
- [27] ERICKSON B. Basic Artificial Intelligence Techniques: Machine Learning and Deep Learning[J]. Radiologic Clinics of North America, 2021, 59(6): 933-940.
- [28] DENG Xun-fei, LIU Zhi, ZHAN Yu, et al. Predictive Geographical Authentication of Green Tea with Protected Designation of Origin Using a Random Forest Model[J]. Food Control, 2020, 107: 106807.
- [29] 卢丹敏, 欧阳建, 熊一帆, 等. 湖南红茶品质特征分析及其等级判别模型的构建[J]. 食品工业科技, 2021, 42(10): 255-260. LU Dan-min, OUYANG Jian, XIONG Yi-fan, et al. Analysis of the Quality Characteristics of Hunan Black Tea and the Establishment of Its Grade Discrimination Model[J]. Science and Technology of Food Industry, 2021, 42(10): 255-260.
- [30] 徐立恒, 吕进, 林敏, 等. 茶叶中 3 类主要组分的近红外光谱分析作为茶叶质量的快速评定方法[J]. 理化检验-化学分册, 2006, 42(5): 334-336. XU Li-heng, LYU Jin, LIN Min, et al. NIR Spectrometric Analysis of 3 Kinds of Main Components in Tea as a Rapid Method of Qualification of Tea[J]. Physical Testing and Chemical Analysis (Part B (Chemical Analysis)), 2006, 42(5): 334-336.
- [31] 金山峰, 王冬欣, 黄俊仕, 等. 基于计算机视觉的茶叶品质在线评价系统[J]. 食品工业科技, 2021, 42(14): 219-225. JIN Shan-feng, WANG Dong-xin, HUANG Jun-shi, et al. Online Evaluation System of Tea Quality Based on Computer Vision[J]. Science and Technology of Food Industry, 2021, 42(14): 219-225.