

DOI: 10.19812/j.cnki.jfsq11-5956/ts.20231221009

近红外光谱技术在茶叶生产质控与品质评价中的研究进展

梁建华^{1,2}, 郭嘉明², 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)

摘要: 茶是世界上最受欢迎的饮料之一, 在国内外均有庞大的市场。市场的发展对茶叶的品质要求不断提高, 而快速、可靠的茶叶质量控制与品质评价技术是辅助茶叶品质提升的重要手段, 一直是研究者关注的热点。传统的茶叶质量与品质检测方法费力、耗时或极具主观性。近红外光谱技术以其简单、无损、快速的特点, 在茶叶生产质控与品质评价中越来越受到人们的欢迎。基于近红外光谱的检测技术是一种间接的检测方法, 其检测结果容易在数据传递过程中受到各种影响因素的干扰, 本文总结了近红外光谱技术在茶叶生产质控与品质评价中的重要影响因素, 分析了近红外光谱技术在茶叶生产质控与品质评价中的应用状况, 并对研究前景进行展望, 为应用近红外光谱技术于茶叶生产质控与品质评价中的研究方向提供参考。

关键词: 近红外光谱技术; 茶叶; 质量控制; 信息融合

Research progress of near-infrared spectroscopy in quality control and evaluation of tea production

LIANG Jian-Hua^{1,2}, GUO Jia-Ming², QIAO Xiao-Yan^{1*}

(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)

ABSTRACT: Tea is one of the most popular beverages in the world and has a huge market both at home and abroad. With the development of the market, the quality requirements of tea are constantly improved, rapid and reliable tea quality control and quality evaluation technology is an important means to assist the quality improvement of tea, which has been the focus of researchers. Traditional methods used for assessing tea quality are laborious, time-consuming, or subjective. The near-infrared spectroscopy technology is more and more popular in tea production quality control and quality evaluation because of its simple, non-destructive and rapid characteristics.

基金项目: 广东省乡村振兴战略专项资金项目(农业科技能力提升)(403-2018-XMZC-0002-90)、2022年度南雄市科技支撑乡村振兴项目、黔东南州科技计划项目[黔东南科合成果(2023)3号]

Fund: Supported by the Project of Promoting the Scientific and Technological Capacity of Tea in Cities and Counties to Promote Industrial Development (403-2018-XMZC-0002-90), the Project of Promoting The Scientific and Technological Capacity of Tea in Nanxiong City to Promote Industrial Development (2022), and the Project of Science and Technology Planning in Qiandongnan Prefecture [Qiandongnan Science and Technology Achievements (2023) No.3]

*通信作者: 乔小燕, 博士, 研究员, 主要研究方向为茶叶加工。Email: qiaoxiaoyan@tea.gdaas.cn

*Corresponding author: QIAO Xiao-Yan, Ph.D, Professor, Tea Research Institute, Guangdong Academy of Agricultural Science, No.6, Dafeng Road, Tianhe District, Guangzhou 510640, China. Email: qiaoxiaoyan@tea.gdaas.cn

Nevertheless, since near-infrared spectroscopy is an indirect detection method that can be easily influenced by various factors during data transmission processes. This paper summarized the important influencing factors of near-infrared spectroscopy technology in tea production quality control and quality evaluation, analyzed the application status of near-infrared spectroscopy technology in tea production quality control and quality evaluation, and prospected the research prospect, providing a reference for the research direction of the application of near-infrared spectroscopy technology in tea production quality control and quality evaluation.

KEY WORDS: near-infrared spectroscopy; tea; quality control; data fusion

0 引言

茶叶是我国重要的经济作物,市场的发展对茶叶的质量和um安全要求不断提高。有效可靠的茶叶质量控制方法与品质评价技术是辅助茶叶品质提升的重要手段,现有的茶叶质量控制与品质评价方法中,化学检测法费时费力且依赖昂贵的化学仪器,产业中常用的感官评审法极具主观性。这些方法无法满足当前茶叶自动化,标准化发展的要求。

近红外光谱(near-infrared spectroscopy, NIRS)是一种介于可见光和中红外光之间的波长范围为780~2526 nm的电磁波谱,可与有机分子中含氢基团(如C-H、O-H、N-H)产生合频、倍频效应,使样品光谱携带含氢基团的特征信息,而这些基团在一定程度上与样品中的特定成分有密切的联系,因此NIRS可被作为获取样品信息的载体^[1],通过分析样本光谱即可获得样本信息。该技术在应用过程中具有高效、准确、无损、成本低等特点,因而被广泛使用于各领域的检测工作中^[2]。

近年来NIRS技术的应用发展快速,基于NIRS技术的茶叶质量控制与品质评价研究成果不断发表,其发展状况及方向日渐明朗^[3]。因此,本文分析了NIRS技术在当前茶叶质量控制与品质评价应用过程中的重要影响因素,总结了国内外NIRS在茶叶质量控制与品质评价中的应用进展,并对NIRS技术在茶学质量控制与品质评价中的研究前景进行展望,为相关研究的进一步发展提供理论依据。

1 NIRS在茶叶生产质控与品质评价应用中的影响因素

基于NIRS技术的检测方法是通过对样本的光谱进行分析进而获取其中的样本信息,是一种间接检测的方法。其检测结果容易在数据传递过程中受到各种影响因素的干扰,需要弱化或消除这些干扰以提升检测结果的可靠性。

1.1 光谱收集环境和方法

NIRS属微弱信号,通过NIRS光谱仪测出的光谱为表观光谱,其包含真实光谱和背景光谱两部分^[4]。真实光谱是确定的,而背景光谱是不确定的,一般携带较多的不确

定环境信息,如由某些设备或外部环境引起的噪声污染导致光谱产生噪声^[5]。相比于小型的或者手持型的NIRS仪,昂贵的台式NIRS仪有效降低了光谱受到的环境干扰,通常具有更高的光谱测量精度。除收集系统因素外,样本本身属性也会对光谱测量值造成干扰,如鲜叶自身的水分、厚度及内部结构分布不均以及干茶的不规则索条形和颗粒不均等因素都会严重影响光谱的测量值。因此可将干茶研磨成粉末以提升样本的均匀性和检测光谱的代表性^[6]。

1.2 检测指标

检测指标与NIRS间的相关性决定了使用NIRS技术对茶叶靶向指标检测的可行性。通常,茶叶中活性成分含量越高,使用NIRS检测的可行性越高,如茶叶中含量丰富的茶多酚、咖啡碱、茶氨酸等成分,在大量的研究中都建立了良好的检测模型^[7];近年来,有研究利用NIRS技术对茶叶中含量浓度低的活性成分进行检测,也获得了良好的检测结果,如儿茶素、茶色素等^[8]。在建立数学模型的过程中,通常直接将检测指标的结果作为检测模型的参考,导致检测指标中的所有误差都会直接传递给模型而影响后续检测。因此,可靠的指标参考数据尤为重要。当前,较多研究以茶叶的属性和品质等为检测指标,这些指标通常依据人的经验与感官评价获取,存在可重复性差和模糊等问题,损害了相应检测模型的可靠性。因此,在获取参考数据时,需确保指标参考检测数据的可靠性。

1.3 数据预处理和建模方法

在对NIRS分析前剔除异常数据对提升模型的可靠性是有益的,异常的样本数据不仅会降低样本的代表性,还会降低模型的可靠性^[9]。李浩光等^[10]提出的支持向量数据描述(support vector data description, SVDD)判别方法可有效帮助剔除异常光谱,石鲁珍等^[11]采用马氏距离法对异常光谱样品进行剔除提高了模型的预测能力。样本检测指标的异常值剔除一般通过经验法和寻找重复测量值中的突出值完成。在建模时需要将样本数据划分为训练集和测试集,测试集充当未知样本集用于检验模型的性能。为降低抽样过程引入的不确定度分量^[12],样本划分需要满足样本分布条件,有效的方法是使用随机算法进行随机抽样,其中蒙特卡罗方法、拟蒙特卡罗方法较为常用。

测量光谱通常携带噪声和偏差,有效的光谱预处理可以消除固体样品中光散射和不同有效光程长度引入的物理干扰,常用的预处理方法有:平滑、校正、标准化。平滑处理可以有效去除光谱中的噪声,其中的卷积平滑(Savitzky-Golay, S-G)在 NIRS 的预处理中较为常用,此外还有滑动平均平滑、高斯平滑等方法。校正处理则可以有效去除光谱中的基线漂移,常用的方法包括位移校正、趋势校正、多元散射校正(multiplicative scatter correction, MSC)、标准正态变量校正(standard normalized variate, SNV)等。在 CHEN 等^[13]建立的预测绿茶多酚和咖啡因模型中,相比于采用 SNV 预处理模型,不采用预处理的咖啡因预测模型的预测残差提高了约 100%,多酚预测模型的预测残差提高了约 50%。因此对光谱进行有效的预处理是必要的。周昌海等^[14]在检测模型分析中使用高斯平滑预处理,获得了相对使用求导、SNV 等预处理方法更好的预测铁观音中茶多酚含量模型;沈诗钰等^[15]在白茶可溶性糖分析实验中选用了导数处理的方法找到了肩峰和吸收峰的位置并消除了基线漂移的影响,赵静远等^[16]在使用连续投影法建立茶叶中咖啡碱的 NIRS 分析模型时,结合小波变换使模型得到了优化。有效的光谱预处理还能在一定程度上消除茶鲜叶的水分干扰,GUO 等^[17]提出了使用外部参数正交化(external parameter orthogonalization, EPO)以消除水分对光谱的影响,相比于从未经预处理的全光谱偏最小二乘法回归(partial least squares regression, PLSR)模型,他们建立的结合 EPO 与基于迭代的保持有信息变量的变量组合总体分析(variable combination population analysis-iteratively retaining informative variables, VCPA-IRIV)的氮含量检测模型,预测模型的相关系数由 0.5846 上升为 0.9371。另外,数据变换如标准化、中心化、归一化等处理通过对数据进行缩放,可以统一特征量纲,并在一定程度上放大特征差异,已被广泛用于光谱预处理。

然而经预处理的光谱仍然存在高共线性度和特征冗余的问题,研究者们对该问题的常用解决方法是数据降维^[18]。数据降维的方法包括特征提取和特征选择^[19]。在基于 NIRS 的茶叶定量检测研究中,特征选择的应用最广,其核心是通过去除光谱中的低信噪比波段和优选其中的最有效特征以降低特征变量的数量,主要方法有遗传基因算法(genetic algorithms, GA)、粒子群算法(particle swarm optimization, PSO)、连续投影法(successive projections algorithm, SPA)等。通常结合特征选择建立的近红外光谱模型具有更优的性能,苏丹等人在基于知识图谱分析近红外光谱技术在茶叶分析中的研究进展中提出特征提取保障茶叶判别性能的性能^[20]。

建立预测模型的过程可以理解为将光谱特征与样品待测属性之间建立关联的过程,这个预测模型是两者联系的桥梁。主成分回归(principal component regression, PCR)、

偏最小二乘法(partial least squares, PLS)、支持向量机(support vector machines, SVM)等经典算法以及如今大热的神经网络算法都是建模时常用的算法。PLS 及其改进的算法通常具有良好的性能因而被广泛使用,NOBARI 等^[21]对关于食品欺诈的多变量分析综述中也强调了 PLS 的使用。其他建模算法在具体研究中也有突出表现,如 ONG 等^[22]在使用 NIRS 技术检测乌龙茶游离氨酸的实验中发现,高斯过程回归(gaussian process regression, GPR)取得了优于 PLSR 的检测精度。

2 NIRS 技术在茶叶生产质控与品质评价中的应用

2.1 内含物、感官得分等定量分析

NIRS 技术在茶叶生产质控与品质评价中最常用于定量分析,表 1 显示了有关基于 NIRS 技术在茶叶生产质控与品质评价中的应用。定量分析的核心是建立回归模型,将 NIRS 中的光谱特征与通过常规方法检测的目标指标数值联系起来,后续对未知样本光谱的检测时通过利用该模型表达的联系可以获取目标指标的量化值。

基于 NIRS 技术对茶叶进行定量分析的检测指标主要有两种,一是茶叶内部化学成分含量,二是感官审评得分等属性值。其中许多研究对茶叶样本进行了研磨处理, PARK 等^[33]的研究发现对固体样本进行精细研磨有利于建立良好的 NIRS 模型。活性化学成分的检测以茶多酚、生物碱、可溶性糖、色素等为主要检测指标,回归模型的决定系数一般在 0.85~0.99 之间,这证明了 NIRS 技术可以很好完成茶叶中活性化学成分检测的任务。经典的 PLSR 通过非迭代偏最小二乘算法建立,并结合了更新分数和权重的算法,在大量的回归分析中均有优异的表现,在活性化学成分检测中应用频率最高。此后发展了不少基于 PLSR 改进的算法,它们在一定程度上可以进一步提升检测的精度,如 LIU 等^[34]在对普洱茶的检测研究中提出的使用加权 PLSR 建立的多种组分检测模型其性能均比使用原始 PLSR 明显提升。在这些研究中,绝大部分的定量模型都使用了预处理和特征选择而建立良好的检测模型。但使用新的预处理和特征选择这并不总是能提高模型的性能,如在 SANAEIFAR 等^[23]建立的基于 NIRS 的维生素 C 检测模型中,使用全光谱的 PLSR 模型相比于其他算法组合构建的模型有更优异的表现,这可能是由于预处理、特征选择在削弱噪声的同时弱化了关键特征。另一方面,近年来基于 NIRS 技术的茶叶感官等属性检测研究增加,检测模型的相关系数在 0.80~0.97 间。在这些研究中 PLSR 不再是首选,PCR 和 MLR 等算法亦有优异的表现,这些研究证明了 NIRS 技术对各类茶叶的品质评价、感官审评以及抗氧化活性上的可行性。实际上以抗氧化活性为例,多酚类与黄

表 1 关于使用 NIRS 技术的茶叶定量分析研究
Table 1 Studies on quantitative analysis of tea using near infrared spectroscopy

样本	样本处理	检测指标	建模方法	R_p^2	参考文献	
茶鲜叶	无	叶绿素 a	CARS+PLSR	0.583	[23]	
		叶绿素 b	SNV+CARS+PCR	0.819		
		类胡萝卜素	NOR+CARS+PCR	0.798		
		维生素 C	PLSR	0.803		
		谷胱甘肽	NOR+CARS+SPA+PLSR	0.917		
		含水量	SNV+CARS+PCR	0.931		
绿茶	研磨	(-)-表没食子儿茶素(-)-表儿茶素表 没食子儿茶素没食子酸酯(-)-表儿茶 素没食子酸酯	MC+PLSR MC+PLSR SNV+PLSR SNV+PLSR	0.985 0.960 0.976 0.976	[24]	
		感官评分	LSTA+RVM	0.963		[25]
		抗氧化能力	SA+PLSR	0.800		[26]
		含水量	SNV+PCA+ENN	0.993		[27]
红茶萎凋叶	无	含水量	SNV+PCA+ENN	0.993	[27]	
红茶发酵叶	研磨	茶褐素	FD+CARS+PLSR	0.970	[28]	
	研磨	葡萄糖 蔗糖	SNV+AS+PLSR SNV+PLSR	0.890 0.960	[29]	
红茶	无	外形得分	S-G+MSC+PCR	0.930	[30]	
		质感得分	S-G+MSC+PLSR	0.970		
		茶汤颜色得分	NW+PCR	0.840		
		品质得分	NW+PCR	0.880		
黑茶, 乌龙茶, 红茶, 绿茶	研磨	苦味	CARS+MLR	0.944	[31]	
	收敛性	CARS+PLSR	0.912			
黑茶, 乌龙茶, 红茶, 绿茶	研磨	茶多酚	MSC+FD+S-G+	0.994	[32]	
		咖啡碱	RF/CARS+PLSR	0.986		
		游离氨基酸	RF/CARS+PLSR	0.993		

注: R_p^2 : determination coefficient of predicted set, 预测相关系数; CARS: competitive adaptive reweighted sampling, 竞争性自适应重加权抽样; NOR: normalization, 标准化; MC: mean center, 平均中心化; LSTA: local tangent space alignment, 局部切空间对齐算法; RVM: relevance vector machine, 关联向量机; ENN: Elman neural network, 基于 Elman 的神经网络; FD: first derivative, 一阶导数; AS: auto scaling, 自标度化; NW: Norris-Williams filter, 诺里斯-威廉姆斯滤波器; MLR: multiple linear regression, 多元线性回归; RF: random frog, 随机蛙跳。

酮类的含量与抗氧化性是正相关的^[35], 而多酚类和黄酮类化合物能被 NIRS 有效检测, 因此抗氧化性与 NIRS 的间接关系可有效建立, 这在一定程度上说明了 NIRS 技术可以对茶叶属性进行检测的原因。

随着茶叶市场的发展, 茶叶精制越发成为一种必要的生产方式, 其中拼配茶可以均衡茶的香气和滋味, 提升茶叶的品质和经济效益, 而如何掌控拼配比例是该工作的重点。宋彦等^[36]提出的基于 NIRS 技术的眉茶拼配比例预测方法为此问题提供了解决思路。在该研究中卷积神经网络比传统的回归表现出更好的性能和非线性回归能力, 其构建的 CNN 模型在识别 25 组拼配茶中测试集的决定系数达 0.9696, 这反映出如今卷积神经网络成功应用于 NIRS 的茶叶定量分析中, 但神经网络在基于 NIRS 的茶叶检测应用中的效用仍需考量。MAYR 等^[37]认为神经网络可以有效增强模型的鲁棒性, 但为确定给定光谱组的最佳隐藏神

经元层数, 神经网络校准需要更高层次的监督, 这就要求训练的样本量相对较大, 而当前基于 NIRS 的茶叶生产质控与品质评价研究大多数是基于小型数据集的。

2.2 品质等级、产地溯源等定性分析

除定量分析外, NIRS 技术在茶叶生产质控与品质评价中还应用于定性分析, 包括茶叶品质等级、地理起源、品种、真实性、掺假等鉴别, 主要采取线性判别分析(linear discriminant analysis, LDA)、SVM、人工神经网络等分类算法实现。传统采用的茶叶品质等级鉴别方法为感官审评法, 由于极度依赖品茶师的经验及实时感官感受, 因此审评结果通常具有主观性和模糊性。基于 NIRS 技术快速高效的特性, 其已被用作茶叶品质研究的重要技术手段。REN 等^[38]在研究 NIRS 对祁门红茶分级的实验中, 建立的结合 CARS 与最小二乘支持向量机(least squares support

vector machine, CARS-LSSVM)模型在测试集上表现优异,准确率达 99.01%,这证明了 NIRS 技术可实现茶叶品质等级的准确检测。研究地理起源,品种,真实性,掺假对维持茶叶质量和行业发展至关重要,因为地理起源和品种会显著影响消费者的购买力,真实性与掺假则直接影响消费者的权益^[39]。NIRS 技术在这类问题的定性研究中均有突出的表现;LI 等^[40]在识别 9 个产地的红茶实验中,基于 NIRS 技术的无监督学习 PCA 模型显示出了组间分离,监督学习 LDA 取得 100%的正确率;SHI 等^[41]对红茶、绿茶、白茶种类识别准确率可达 99.87%;有研究在对大吉岭红茶的真实性判别实验中,对大吉岭红茶与其他红茶进行区分,建立的基于近红外偏最小二乘判别分析(partial least squares discriminant analysis, PLSDA)模型准确率达 98.57%^[42]。显然,利用 NIRS 技术对茶叶进行定性分析是可行的。但注意到这些实验中使用的样本在组间存在较大的差异,表现在使用无监督学习方法 PCA 进行分析即已显示出组间差异,而在实际应用过程中检测分类的应用场景经常会遇到组间差异较小的样本,多源信息融合方法应运而生。

2.3 多源信息融合方法

随着茶叶研究的深入,人们发现仅依赖单一传感器进行茶叶检测不再满足进一步提升检测精度和细化应用场景的需要;在对具有高度相似性的茶叶进行检测时,检测难度较大,因为单一传感器的信息无法完全表征茶叶的需检测要素。由此,研究者们提出了利用两个及以上的传感器来完成检测。融合两个及以上的传感器信息进行分析的过程被称为多源信息融合,亦称多模态融合、多传感器融

合,其核心是整合两类或以上的模态信息以提升模型的预测性能^[43]。其通过融合多种模态中的信息来实现准确预测所需的有效信息的补充,提升预测精度和模型的鲁棒性^[44]。在食品研究中,这种方法已逐渐成为各类研究的重点,如 WU 等^[45]认为多模态深度学习将作为营养研究的下一个挑战。多源信息融合方法在茶叶生产质控与品质评价中表现优异,AN 等^[46]在红茶发酵香气的质量评价研究中提出的融合高光谱和比色传感阵列(colorimetric sensor array, CSA)信息的方法相比于传统的 CSA 方法显著提升了准确率;而近年来利用 NIRS 技术结合多源信息融合方法在茶叶检测的研究正逐步开展,如在 LI 等^[47]在祁门红茶等级评价研究中相比于仅使用 NIRS 光谱信息的模型,将光谱信息与其他模态信息融合后,内含物检测及品质定级模型性能有所提高。

表 2 展示了将 NIRS 技术与其他技术融合在茶叶生产质控与品质评价中的研究,其研究对象主要以红茶和绿茶为主,而与 NIRS 融合的技术以计算机视觉为主,定性分析准确率在 89%~100%,定量分析的相关系数在 0.96~0.98 间。对比仅使用 NIRS 技术进行检测,使用多模态融合并不一定能显著提升检测的准确度;当仅使用 NIRS 进行检测时,若其检测精度已相对较高,联合其他模态数据的检测难以产生较大的提升,较大的提升只在仅使用 NIRS 检测获得较低的准确率时出现,如在 LI 等^[56]对红茶的品质定级时,仅使用 NIRS 进行定级时的模型最优准确率仅有 68.57%,而联合计算机视觉技术进行融合后,准确度达 96.29%。因此在仅使用 NIRS 或单一传感器进行检测时效果不佳时,融合其他模态数据会是一种可行的方法。

表 2 NIRS 结合多源信息融合在茶叶生产质控与品质评价中的应用

Table 2 Application of multi-source information fusion involving NIRS in quality control and evaluation of tea production

样品	应用场景	模态信息技术	融合方法	相关系数*/准确率		参考文献
				NIRS	融合后	
6 大茶类	茶类辨别	NIRS+UV-Vis+SF	特征级融合	97.30%	100.00%	[48]
藤茶	品质定级	NIRS+GC-MS	特征级融合	86.19%	92.38%	[49]
绿茶	加工环节水分检测	NIRS+CV	特征级融合	0.9689*	0.9777*	[50]
	杀青过程水分检测	NIRS+CV	特征级融合	0.9212*	0.9719*	[51]
	萎凋程度	NIRS+CV+CSA	特征级融合	82.50%	97.50%	[52]
红茶	发酵程度	NIRS+CV	特征级融合	62.16%	89.19%	[53]
	干燥过程水分检测	NIRS+CV	特征级融合	0.9457*	0.9696*	[54]
	加工环节茶多酚/儿茶素检测	NIRS+CV	特征级融合	0.9800/0.9700*	0.9800/0.9700*	[55]
	品质定级	NIRS+CV	特征级融合	68.57%	94.29%	[56]
	品质定级	NIRS+CV+CSA	特征级融合	91.43%	98.57%	[47]

注: UV-Vis: ultraviolet-to-visible, 紫外可见光谱; SF: synchronous fluorescence, 同步荧光光谱; GC-MS: gas chromatography-mass spectrometry, 气相色谱-质谱; CV: computer vision, 计算机视觉。

多源信息融合的方法可以分为 4 类: 特征级融合、决策级融合、混合级融合和模型级融合。特征级融合在数据分析前期的特征提取阶段进行, 在对各源信息完成融合后再利用机器学习方法进行分析, 属于早期融合; 而决策级融合是对决策结果进行, 属于后期融合, 这两者类似于集成学习。混合级融合结合了前期融合与后期融合, 模型级融合则是进一步的联合前三种融合, 混合级融合以及模型级融合通常用于解决自然语言处理中话语层面的多模态融合问题。从融合方法上看, 统计的基于 NIRS 与其他模态信息融合的茶学研究只使用了特征级融合, 这反映了特征融合方法在茶叶检测领域的研究不够深入。尽管这些研究均进行了低级融合(直接将模态信息拼合)与中级融合(提取特征后融合特征)的对比^[57], 但这都属于前期的特征级融合方法, 进一步的研究可以探讨其他的融合方法的实用性。从统计的研究应用上看, 使用数据融合的研究在应用上更趋于生产过程中茶产品的检测, 这区别于依赖单一 NIRS 技术时针对的单类型样本的定量分析或具有较大组间差异的茶叶鉴别, 反映出茶叶检测应用场景的细化。

3 结束语

NIRS 技术现已成功应用于茶叶的定量和定性分析, 且趋于生产场景的应用。其核心是建立一个可靠的广泛应用的预测模型, 其中样本制备的方法条件, 数据分布, 数据预处理和回归/分类算法是影响模型性能的重要因素。多模态融合方法的引入, 拓宽了 NIRS 的检测范围, 提升其检测能力。然而, 同时需要认识到 NIRS 技术和当前近红外在茶叶检测研究中的局限性, 如样本制备时的研磨操作, 尽管可以提升模型的性能, 但研磨操作无疑导致了额外的样本制备时间和劳动力; NIRS 模型大多数是基于小样本集的, 模型的代表性无法保证; 由于 NIRS 属于精密仪器, 不同 NIRS 仪间的测量误差有别, 导致已搭建的模型在迁移到另一 NIRS 仪应用时无法满足应有的性能; 并且台式光谱仪的昂贵价格仍是限制其实际应有的重要因素。

因此, 基于当前 NIRS 技术在茶叶检测中应用的局限性, 今后的研究可围绕以下方面展开: 1) 提高小型 NIRS 仪器的实用性。当前各种价格相对较低廉的手持小型 NIRS 仪不断被开发, 但仍存在开机时间长, 精度较低等问题, 仪器生产技术的提升和算法的优化成为提升小型 NIRS 仪的实用性的重要研究内容。2) 多源信息融合技术在茶叶研究中的应用场景越来越广。从生产监管到产品质量检测, 融合 NIRS 技术与其他传感器的方法在实际应用场景中表现出了很大潜力, 未来的研究可以围绕更细化的应用场景以及更有效的融合策略展开。3) 构建大型的公共数据库来增强模型的稳健性和可靠性。当前大多数的近红外研究是基于有限的样本集, 机器学习模型的稳健性需要依赖大型的数据集支撑。通过建立标准化的数据上传, 数据管理来实

现将多个独立样本整合开发一个公共数据库, 以供近红外技术在茶叶检测领域的研究人员使用。构建大型数据集的可行性与有效性已在图像识别领域被证明, 因此构建茶叶的 NIRS 大型数据集是必要的。

参考文献

- [1] 王胜鹏, 高士伟, 滕靖, 等. 近红外光谱技术在茶叶中的研究进展[J]. 华中农业大学学报, 2021, 40(5): 226–232.
WANG SP, GAO SW, TENG J, *et al.* Progress of using near infrared spectroscopy in tea [J]. *J Huazhong Agric Univ*, 2021, 40(5): 226–232.
- [2] LIN XH, SUN DW. Recent developments in vibrational spectroscopic techniques for tea quality and safety analyses [J]. *Trends Food Sci Technol*, 2020, 104: 163–176.
- [3] SHUAI MY, PENG CY, NIU HL, *et al.* Recent techniques for the authentication of the geographical origin of tea leaves from *camellia sinensis*: A review [J]. *Food Chem*, 2022, 374: 131713.
- [4] WEI JJ, ZHU C, ZHANG ZM, *et al.* Two-stage iteratively reweighted smoothing splines for baseline correction [J]. *Chemometr Intell Lab*, 2022, 227: 104606.
- [5] LUO W, TIAN P, FANG Z, *et al.* Non-destructive determination of four tea polyphenols in fresh tea using visible and near-infrared spectroscopy [J]. *Infrared Phys Technol*, 2022, 123: 104037.
- [6] TIAN WF, LI YH, GUZMAN C, *et al.* Quantification of food bioactives by NIR spectroscopy: Current insights, long-lasting challenges, and future trends [J]. *J Food Compos Anal*, 2023, 124: 105708.
- [7] NAGY MM, WANG S, FARAG MA. Quality analysis and authentication of nutraceuticals using near IR (NIR) spectroscopy: A comprehensive review of novel trends and applications [J]. *Trends Food Sci Technol*, 2022, 123: 290–309.
- [8] LI XL, JIN JJ, SUN CJ, *et al.* Simultaneous determination of six main types of lipid-soluble pigments in green tea by visible and near-infrared spectroscopy [J]. *Food Chem*, 2019, 270: 236–242.
- [9] 林俊. 计量测试中异常数据的处理方式[J]. 电子技术与软件工程, 2019, (4): 153.
LIN J. Processing mode to handle abnormal data in measurement testing [J]. *Electron Technol Softw Eng*, 2019, (4): 153.
- [10] 李浩光, 于云华, 沈学锋, 等. SVDD 的近红外光谱定性分析光谱质量判定方法[J]. 光谱学与光谱分析, 2019, 39(12): 3783–3787.
LI HG, YU YH, SHEN XF, *et al.* Research on NIR spectra quality detection method based on support vector data description [J]. *Spectrosc Spect Anal*, 2019, 39(12): 3783–3787.
- [11] 石鲁珍, 张景川, 王彦群, 等. 马氏距离与浓度残差剔除近红外异常样品研究[J]. 中国农机化学报, 2016, 37(6): 99–103.
SHI LZ, ZHANG JC, WANG YQ, *et al.* Research for eliminating outlier samples in near-infrared spectroscopy by method of Mahalanob distance-concentration residual [J]. *J Chin Agric Mechan*, 2016, 37(6): 99–103.
- [12] 朱顺, 姚佳爱. 抽样引入的测量不确定度分量的评估[J]. 上海计量测试, 2021, 48(5): 21–22.
ZHU S, YAO JAI. Evaluation of measurement uncertainty components caused by sampling [J]. *Shanghai Meas Test*, 2021, 48(5): 21–22.
- [13] CHEN QS, ZHAO JW, HUANG XY, *et al.* Simultaneous determination of

- total polyphenols and caffeine contents of green tea by near-infrared reflectance spectroscopy [J]. *Microchem J*, 2006, 83(1): 42–47.
- [14] 周昌海, 周海军. 铁观音中茶多酚含量的近红外光谱检测模型分析[J]. *宜宾学院学报*, 2019, 19(12): 121–124.
ZHOU CH, ZHOU HJ. Study on nearinfrared spectroscopic detection model of tea polyphenols in Tieguanyin [J]. *J Yibin Univ*, 2019, 19(12): 121–124.
- [15] 沈诗钰, 孙威江, 唐琴, 等. 基于近红外光谱技术的白茶可溶性糖总量快速测定研究[J]. *天然产物研究与开发*, 2019, 31(1): 16–23.
SHEN SY, SUN WJ, TANG Q, *et al*. Rapid determination of total water-soluble sugars in white tea by near-infrared spectroscopy [J]. *Nat Prod Res Dev*, 2019, 31(1): 16–23.
- [16] 赵静远, 熊智新, 宁井铭, 等. 小波变换结合连续投影算法优化茶叶中咖啡碱的近红外分析模型[J]. *分析科学学报*, 2021, 37(5): 611–617.
ZHAO JY, XIONG ZX, NING JM, *et al*. Wavelet transform combined with SPA to optimize the near-infrared analysis model of caffeine in tea [J]. *J Anal Sci*, 2021, 37(5): 611–617.
- [17] GUO JM, HUANG H, HE XL, *et al*. Improving the detection accuracy of the nitrogen content of fresh tea leaves by combining FT-NIR with moisture removal method [J]. *Food Chem*, 2023, 405: 134905.
- [18] VACHHARAJANI B, PANDYA D. Dimension reduction techniques: Current status and perspectives [J]. *Mater Today: Proc*, 2022, 62(13): 7024–7027.
- [19] ABE S. Feature selection and extraction. In: *Support vector machines for pattern classification* [M]. London: Springer, 2010.
- [20] 苏丹, 王志霞, 周佳, 等. 基于知识图谱分析近红外光谱技术在茶叶分析中的研究进展[J]. *食品安全质量检测学报*, 2022, 13(4): 1193–1200.
SU D, WANG ZX, ZHOU J, *et al*. Research progress of near infrared spectroscopy in tea analysis based on knowledge mapping [J]. *J Food Saf Qual*, 2022, 13(4): 1193–1200.
- [21] NOBARI MH, TAMJJI Z, AKBARI LM, *et al*. Multivariate analysis of food fraud: A review of NIR based instruments in tandem with chemometrics [J]. *J Food Compos Anal*, 2022, 107: 104343.
- [22] ONG P, CHEN S, TSAI CY, *et al*. Prediction of tea theanine content using near-infrared spectroscopy and flower pollination algorithm [J]. *Spectrochim Acta A*, 2021, 255: 119657.
- [23] SANAEIFAR A, ZHU FL, SHA JJ, *et al*. Rapid quantitative characterization of tea seedlings under lead-containing aerosol particles stress using Vis-NIR spectra [J]. *Sci Total Environ*, 2022, 802: 149824.
- [24] CHEN QS, ZHAO JW, CHAITSP SP, *et al*. Simultaneous analysis of main catechins contents in green tea [*Camellia sinensis* (L.)] by fourier transform near infrared reflectance (FT-NIR) spectroscopy [J]. *Food Chem*, 2009, 113(4): 1272–1277.
- [25] LIU P, ZHU XY, HU X, *et al*. Local tangent space alignment and relevance vector machine as nonlinear methods for estimating sensory quality of tea using NIR spectroscopy [J]. *Vib Spectrosc*, 2019, 103: 102923.
- [26] GUO ZM, BARIMAH AO, SHUJAT A, *et al*. Simultaneous quantification of active constituents and antioxidant capability of green tea using NIR spectroscopy coupled with swarm intelligence algorithm [J]. *LWT*, 2020, 129: 109510.
- [27] SHEN S, HUA JJ, ZHU HK, *et al*. Rapid and real-time detection of moisture in black tea during withering using micro-near-infrared spectroscopy [J]. *LWT*, 2022, 155: 112970.
- [28] 夏晶晶, 张敏, 王飞仁, 等. 英红九号发酵叶中茶褐素近红外定量模型的优化与验证 [J]. *现代食品科技*, 2023, 39(6): 313–320.
XIA JJ, ZHANG M, WANG FR, *et al*. Optimization and verification of a near infrared quantitative model for the theabrownin in Yinghong No.9 fermented leaves [J]. *Mod Food Sci Technol*, 2023, 39(6): 313–320.
- [29] NELUM KG, PIYASENA P, RANATUNGA MAB, *et al*. Prediction of glucose and sucrose values of black tea samples using NIR spectroscopy and chemometrics [J]. *Food Humanity*, 2023, 1: 1482–1493.
- [30] TURGUT SS, ENTRENAS JA, TAŞKIN E, *et al*. Estimation of the sensory properties of black tea samples using non-destructive near-infrared spectroscopy sensors [J]. *Food Control*, 2022, 142: 109260.
- [31] WANG YJ, LI TH, LI LQ, *et al*. Evaluating taste-related attributes of black tea by micro-NIRS [J]. *J Food Eng*, 2021, 290: 110181.
- [32] WANG JH, WANG YF, CHENG JJ, *et al*. Enhanced cross-category models for predicting the total polyphenols, caffeine and free amino acids contents in Chinese tea using NIR spectroscopy [J]. *LWT*, 2018, 96: 90–97.
- [33] PARK JR, KANG HH, CHO JK, *et al*. Feasibility of rapid piperine quantification in whole and black pepper using near infrared spectroscopy and chemometrics [J]. *J Food Sci*, 2020, 85(10): 3094–3101.
- [34] LIU Z, XIE HL, CHEN L, *et al*. An improved weighted partial least squares method coupled with near infrared spectroscopy for rapid determination of multiple components and anti-oxidant activity of Pu-Erh tea [J]. *Molecules*, 2018, 23(5): 1058.
- [35] GULSUNOGLU ZH, KARBANCIOGLU-GULER FD, RAES KL, *et al*. Soluble and insoluble-bound phenolics and antioxidant activity of various industrial plant wastes [J]. *Int J Food Prop*, 2019, 22(1): 1501–1510.
- [36] 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.
SONG Y, WANG XZ, ZHAO L, *et al*. Predicting the blending ratio of Mee tea based on near infrared spectroscopy [J]. *Trans Chin Soc Agric Eng*, 2022, 38(2): 307–315.
- [37] MAYR S, BEĆ KB, GRABSKA J, *et al*. Challenging handheld NIR spectrometers with moisture analysis in plant matrices: Performance of PLSR vs. GPR vs. ANN modeling [J]. *Spectrochim Acta A*, 2021, 249: 119342.
- [38] REN GX, WANG YJ, NING JM, *et al*. Highly identification of Keemun black tea rank based on cognitive spectroscopy: Near infrared spectroscopy combined with feature variable selection [J]. *Spectrochim Acta A*, 2020, 230: 118079.
- [39] LIU Y, PENG QW, YU JC, *et al*. Identification of tea based on CARS-SWR variable optimization of visible/near-infrared spectrum [J]. *J Sci Food Agric*, 2020, 100(1): 371–375.
- [40] LI YC, LOGAN N, QUINN B, *et al*. Fingerprinting black tea: When spectroscopy meets machine learning a novel workflow for geographical origin identification [J]. *Food Chem*, 2024, 438: 138029.
- [41] SHI WM, LIU Y, KONG W, *et al*. Tea Classification by near infrared spectroscopy with projection discriminant analysis and gene expression programming [J]. *Anal Lett*, 2015, 48(18): 2833–2842.
- [42] FIRMANI P, DE LS, BUCCI R, *et al*. Near infrared (NIR) spectroscopy-based classification for the authentication of Darjeeling black tea [J]. *Food Control*, 2019, 100: 292–299.

- [43] MENG T, JING X, YAN Z, *et al.* A survey on machine learning for data fusion [J]. *Inform Fusion*, 2020, 57: 115–129.
- [44] GUO WZ, WANG JW, WANG SP. Deep Multimodal representation learning: A survey [J]. *Ieee Access*, 2019, 7: 63373–63394.
- [45] WU XX, JIA W. Multimodal deep learning as a next challenge in nutrition research: Tailoring fermented dairy products based on cytidine diphosphate-diacylglycerol synthase-mediated lipid metabolism [J]. *Crit Rev Food Sci*, 2023, 2023: 1–12.
- [46] AN T, HUANG WQ, TIAN X, *et al.* Hyperspectral imaging technology coupled with human sensory information to evaluate the fermentation degree of black tea [J]. *Sens Actuators B Chem*, 2022, 366: 131994.
- [47] LI LQ, CHEN YR, DONG S, *et al.* Rapid and comprehensive grade evaluation of Keemun black tea using efficient multidimensional data fusion [J]. *Food Chem: X*, 2023, 20: 100924.
- [48] DANKOWSKA A, KOWALEWSKI W. Tea types classification with data fusion of UV-Vis, synchronous fluorescence and NIR spectroscopies and chemometric analysis [J]. *Spectrochim Acta A Mol Biomol Spectrosc*, 2019, 211: 195–202.
- [49] LI QQ, ZHANG Y, WANG HW, *et al.* Machine learning technique combined with data fusion strategies: A tea grade discrimination platform [J]. *Ind Crop Prod*, 2023, 203: 117127.
- [50] LIU ZY, ZHANG RT, YANG CS, *et al.* Research on moisture content detection method during green tea processing based on machine vision and near-infrared spectroscopy technology [J]. *Spectrochim Acta A Mol Biomol Spectrosc*, 2022, 271: 120921.
- [51] 董春旺, 刘中原, 杨明, 等. 基于多源信息融合的绿茶杀青叶水分含量智能感知方法[J]. *食品科学*, 2022, 43(20): 242–251.
DONG CW, LIU ZY, YANG M, *et al.* Intelligent sensing method for detecting moisture content in fixed tea leaves for green tea based on multi-source information fusion [J]. *Food Sci*, 2022, 43(20): 242–251.
- [52] WANG YJ, LIU Y, CUI QQ, *et al.* Monitoring the withering condition of leaves during black tea processing via the fusion of electronic eye (E-eye), colorimetric sensing array (CSA), and micro-near-infrared spectroscopy (NIRS) [J]. *J Food Eng*, 2021, 300: 110534.
- [53] JIN G, WANG YJ, LI MH, *et al.* Rapid and real-time detection of black tea fermentation quality by using an inexpensive data fusion system [J]. *Food Chem*, 2021, 358: 129815.
- [54] SHENG XF, ZAN JZ, JIANG YW, *et al.* Data fusion strategy for rapid prediction of moisture content during drying of black tea based on micro-NIR spectroscopy and machine vision [J]. *Optik*, 2023, 276: 170645.
- [55] WANG YJ, LI LQ, LIU Y, *et al.* Enhanced quality monitoring during black tea processing by the fusion of NIRS and computer vision [J]. *J Food Eng*, 2021, 304: 110599.
- [56] LI LQ, WANG YJ, JIN SS, *et al.* Evaluation of black tea by using smartphone imaging coupled with micro-near-infrared spectrometer [J]. *Spectrochim Acta A Mol Biomol Spectrosc*, 2021, 246: 118991.
- [57] CHEN QS, ZHANG DL, PAN WX, *et al.* Recent developments of green analytical techniques in analysis of tea's quality and nutrition [J]. *Trends Food Sci Technol*, 2015, 43(1): 63–82.

(责任编辑: 韩晓红 张晓寒)

作者简介



梁建华, 硕士, 主要研究方向为农业机械/茶叶检测。

E-mail: Mrliang8015@163.com



乔小燕, 博士, 研究员, 主要研究方向为茶叶加工。

E-mail: qiaoxiaoyan@tea.gdaas.cn