

分类号:

学 号: 20192109015

密 级: 公开

单位代码: 10759

石河子大学

硕士学位论文



基于光谱的绿茶加工过程中水分信息检测方法研究

学 位 申 请 人

刘中原

指 导 教 师

罗 昕 教 授

董春旺 副研究员

申请学位类别

专 业 硕 士

专 业 名 称

机 械 工 程

研 究 领 域

机 械 工 程

所 在 学 院

机械电气工程学院

中国·新疆·石河子

2022 年 6 月

分类号:

学 号: 20192109015

密 级: 公开

单位代码: 10759

石河子大学

硕士学位论文



基于光谱的绿茶加工过程中水分信息检测方法研究

学 位 申 请 人

刘中原

指 导 教 师

罗 昕 教 授

董春旺 副研究员

申请学位类别

工 程 硕 士

专 业 名 称

机 械 工 程

研 究 领 域

机 械 工 程

所 在 学 院

机械电气工程学院

中国·新疆·石河子

2022 年 6 月

Research on Detection Method of Moisture Information in the Green Tea Processing Based on Spectrum

A Dissertation Submitted to

Shihezi University

In Partial Fulfillment of the Requirements

for the Degree of

Master of Engineering

By

Liu zhong-yuan

(Mechanical Engineering)

Dissertation Supervisor: **Prof. Luo Xin**

& A/Prof. Dong Chun-wang

June,2022

石河子大学学位论文独创性声明及使用授权声明

学位论文独创性声明

本人所呈交的学位论文是在我导师的指导下进行的研究工作及取得的研究成果。据我所知，除文中已经注明引用的内容外，本论文不包含其他个人已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中作了明确的说明并表示谢意。

研究生签名：刘中原

时间：2022年6月14日

使用授权声明

本人完全了解石河子大学有关保留、使用学位论文的规定，学校有权保留学位论文并向国家主管部门或指定机构送交论文的电子版和纸质版。有权将学位论文在学校图书馆保存并允许被查阅。有权自行或许可他人将学位论文编入有关数据库提供检索服务。有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

研究生签名：刘中原

时间：2022年6月14日

导师签名：罗昕

时间：2022年6月14日

摘 要

绿茶加工过程主要包括摊青、杀青、揉捻、干燥等工序，水分含量是设定工艺流程和工艺参数、控制原料的物理状态变化和化学反应进程的重要依据，但目前大多茶企对绿茶加工过程中水分含量的把控主要依靠经验法和水分测定仪判断水分含量，无法实现水分含量的快速、无损、准确的检测，从而导致茶叶品质不统一。若要实现绿茶加工过程中水分含量的精准测量，则需依靠烘干称重法，但是此方法耗时较长，显然无法实现快速感知的需求，难于应用于实际生产中。因此，急需一种快速、精准的无损检测方法以实现绿茶加工过程中水分含量的检测。本文选取不同品种的茶叶样品作为实验对象，利用近红外光谱技术，并结合机器视觉技术和化学计量学方法，建立绿茶不同加工工艺下的水分含量预测和判别模型，以期为提高成品绿茶的品质，推动绿茶加工标准化和智能化发展提供理论基础。本文主要研究内容如下：

（1）基于近红外光谱的绿茶摊青过程中的水分含量研究

水分含量是绿茶摊青过程中的重要指标。为了实现摊青过程中含水率的快速有效检测，本文采集了绿茶摊青过程中 200 个近红外光谱数据。为了消除光谱内的噪声信息，采用 5 种不同的预处理方法，并建立偏最小二乘回归（PLSR）预测模型，从而优选出最佳预处理方法。为了减少光谱中的冗余信息，提高模型的预测精度和速度，基于最佳预处理方法提取光谱中特征波长，并建立了偏最小二乘回归和支持向量回归（SVR）预测模型，并分析基于不同特征波长筛选方法所建立的模型效果，确定了 VCPA-GA+SVR 作为最终的绿茶摊青水分含量预测模型。模型的校正集相关系数 R_c 、预测集相关系数 R_p 均高于 95%，相对标准偏差（RPD）大于 2，表明模型预测性能较好，可以实现绿茶摊青过程中的水分含量的精准预测。

（2）基于多源信息融合的绿茶杀青叶水分含量预测模型建立

为了实现绿茶杀青过程中水分含量的快速有效检测，利用机器视觉结合近红外光谱技术，构建了绿茶杀青过程中水分含量变化的定量预测模型。首先采集杀青过程中在制品的光谱和图像信息，然后不同变量筛选方法提取光谱中的特征波长，并融合图像中的 15 个色泽和纹理特征建立线性偏最小二乘回归（PLSR）和非线性支持向量回归（SVR）预测模型。结果表明，与单一数据相比，基于融合数据所建立的模型能有效的提高预测精度，其中基于竞争性自适应重加权法（CARS）算法提取光谱特征波长融合图像的 15 个颜色特征，并结合归一化（normalize）预处理和主成分分析法（PCA）所建立的 SVR 模型效果最佳，其中校正集相关系数（ R_c ）为 0.9742，预测集相关系数（ R_p ）值为 0.9719，相对标准偏差（RPD）值为 4.1546，表明模型具有极好的预测性能。综上，本研究证明了融合光谱和图像技术对绿茶杀青过程中水分含量预测的可行性，克服了单一传感器预测精度低的问题，为实现绿茶杀青叶水分含量的快速无损检测和精准把控杀青质量奠定了理论基础。

（3）基于近红外光谱的绿茶干燥程度判别模型建立。

以不同干燥阶段的绿茶在制品为研究对象，建立绿茶干燥程度判别模型。根据干燥样本的水分含量将其分为3类（干燥不足、干燥适度、干燥过度），并通过提取不同干燥程度的茶叶样品光谱吸光度值，结合4种光谱预处理方法，建立偏最小二乘线性判别（PLS-DA）模型，得出均值中心化（MC）判别效果最佳。进而采用竞争性自适应重加权法（CARS）、变量组合集群分析法（VCPA）、变量组合集群分析法结合迭代保留信息变量法（VCPA-IRIV）和变量组合集群分析法结合遗传算法（VCPA-GA），筛选出与水分含量相关的最优特征波长，然后结合主成分分析（PCA），建立K-邻近算法（KNN）和极限学习机（ELM）判别模型。分析不同模型判别效果，确定了VCPA-IRIV+PCA+ELM判别效果最佳，校正集和预测集判别效果均达到100%，可以实现绿茶干燥程度的准确判别，为绿茶干燥程度在线判别设备的研制奠定了理论基础

（4）基于机器视觉和近红外光谱的成品绿茶水分含量预测模型建立

在绿茶加工过程中，为了使预测模型更加准确、全面的预测绿茶不同加工阶段下的水分信息，以诸叶齐茶为研究对象，采集不同加工阶段（摊青、杀青、初烘、理条、足烘）的近红外光谱数据和图像信息通过融合两个传感器的数据信息，采用竞争性自适应重加权法（CARS）提取光谱中的特征波长，结合图像的9个颜色特征和6个纹理特征，建立非线性的SVR预测模型。结果表明，基于数据融合所建立的SVR模型获得了极好的预测效果，校正集相关系数 R_c 值为0.9804，预测集相关系数 R_p 值0.977，相对标准偏差（RPD）值为4.5002。表明，基于机器视觉和近红外光谱技术的数据融合可以有效的预测绿茶从鲜叶摊青至干燥成品整个加工过程的水分含量。

关键词：绿茶水分含量；近红外光谱；机器视觉；判别；定量预测

Abstract

The processing of green tea mainly includes spreading, fixation, rolling, drying and other processing technology. The moisture content is an important basis for setting the process flow and process parameters and controlling the physical state change and chemical reaction process of raw materials. However, at present, most tea companies mainly rely on empirical methods and moisture analyzers to determine the moisture content in the green tea processing process, and cannot achieve rapid, non-destructive and accurate detection of moisture content, resulting in inconsistent tea quality. In order to realize the accurate measurement of moisture content in the green tea processing, it is necessary to rely on the oven drying method, but this method takes a long time, obviously cannot meet the needs of rapid perception, and is difficult to apply in actual production. Therefore, a rapid and accurate nondestructive testing method is urgently needed to detect the moisture content in the process of green tea processing. In this thesis, tea samples of different varieties were selected as the experimental objects, and the near-infrared spectroscopy technology, combined with machine vision technology and chemometric methods, was used to establish the moisture content prediction and discrimination model of green tea under different processing techniques, in order to improve the quality of finished green tea, Promote the standardization and intelligent development of green tea processing to provide a theoretical basis. The main research contents of this thesis are as follows:

(1) Research on the moisture content of green tea in the green tea spreading process based on near-infrared spectroscopy

Moisture content is an important indicator in the process of green tea spreading. In order to realize the rapid and effective detection of the moisture content during the spreading process, this thesis collected 200 near-infrared spectral data during the spreading process. In order to eliminate the noise information in the spectrum, five different preprocessing methods were adopted, and the PLSR prediction model was established to optimize the best preprocessing method. In order to reduce the redundant information in the spectrum and improve the prediction accuracy and speed of the model, the characteristic wavelengths in the spectrum were extracted based on the best preprocessing method, and the partial least squares regression (PLSR) and support vector regression (SVR) prediction models were established, and the effects of the models established based on different characteristic wavelength screening methods were analyzed, and determined VCPA-GA+SVR as the final prediction model of moisture content in green tea spreading process. The R_c and R_p of the model are both higher than 95%, and the RPD is greater than 2, indicating that the model has good prediction performance and can accurately predict the moisture content in the process of green tea spreading.

(2) Establishment of a prediction model for the moisture content of green tea fixation leaves based on multi-source information fusion

In order to realize the rapid detection of moisture content in the green tea fixation process, a quantitative prediction model of moisture content change in the process of green tea fixation was constructed by using machine vision and near infrared spectroscopy. By collecting the spectrum and image information in the fixation process, and then fused the data information from the two sensors, using competitive adaptive reweighted sampling (CARS), variable combination population analysis (VCPA), variable combination population analysis and the iterative retained information variable algorithm (VCPA-IRIV) and random frog (RF) algorithm to extract the characteristic wavelengths in the spectrum and combined the 15 colors and texture features in the image, the prediction models of partial least squares regression (PLSR) and support vector regression (SVR) were established. The results show that the model based on fusion data can effectively improve the prediction accuracy. Among them, the spectral feature wavelength extracted based on the CARS algorithm was combined with 15 color features of the image, the established SVR model had the best effect, in which the correlation coefficient of the calibration set (R_c) was 0.9742, prediction set correlation coefficient (R_p) value was 0.9719, and relative percent deviation (RPD) value was 4.1546, indicating the model had excellent predictive performance. In conclusion, this study proves the feasibility of integrating spectrum and image technology to predict the moisture content in the process of green tea fixation, overcame the problem of low prediction accuracy of a single sensor, and laid a theoretical foundation for realizing the rapid nondestructive detection of the moisture content of green tea fixation leaves and accurately controlling the fixation quality.

(3) The establishment of a green tea drying degree discrimination model based on near-infrared spectroscopy.

Taking the green tea products in different drying stages as the research object, a discriminant model of green tea drying degree was established. According to the moisture content of dried samples, they were divided into three categories (under-drying, moderate drying, over-drying). By extracting the spectral absorbance values of tea samples with different drying degrees, combined with four spectral pretreatment methods, a partial least squares discrimination analysis (PLS-DA) model was established, and it was concluded that the discrimination model based on mean center (MC) pretreatment method has the best effect. Then, the competitive adaptive reweighted sampling (CARS), variables combination population analysis (VCPA), variables combination population analysis combined with iterative retained information variable algorithm (VCPA-IRIV) and variables combination population analysis combined with genetic algorithm (VCPA-GA) were used to screen the optimal characteristic wavelength related to moisture content. Then, combined with principal component analysis (PCA), k-nearest neighbor (KNN) and extreme learning machines (ELM) discrimination models were established. By analyzing the discrimination effects

of different models, it is determined that MC+VCPA-IRIV+PCA+ELM has the best discrimination effect, and the discrimination effects of correction set and prediction set both reached 100%, which can realize the accurate discrimination of green tea drying degree. It lays a theoretical foundation for the development of online discrimination equipment for the drying degree of green tea.

(4) Research on online detection method of moisture content of finished green tea based on multi-source information fusion

In the process of green tea processing, in order to make the prediction model more accurate and comprehensive to predict the moisture information of green tea in different processing stages, taking zhuyeqi tea as the research object, collect the near-infrared spectral data and image information in different processing stages (spreading, fixation, first-drying, carding, and secondary-drying). By fusing the data information of the two sensors, the CARS was used to extract the characteristic wavelength in the spectrum, and combined with 9 color features and 6 texture features of the image, a nonlinear SVR prediction model was established. The results show that the SVR model based on data fusion has the good effect, the R_c value was 0.9804, the R_p value was 0.977, and the relative percent deviation (RPD) was 4.5002. It is shown that the data fusion based on machine vision and near-infrared spectroscopy technology can effectively predict the moisture content of green tea during the whole processing process from fresh leaves to dry finished product, and compared with the results of the moisture content prediction model for each processing stage of Gaoqiao Yinfeng tea.

Key words: green tea moisture content; near-infrared spectroscopy; machine vision; judgement; quantitative prediction

目 录

摘 要	I
Abstract	III
第 1 章 绪论	1
1.1 研究背景	1
1.2 研究目的及意义	2
1.3 国内外研究现状	2
1.3.1 绿茶加工设备的研究现状	2
1.3.2 茶叶中水分含量检测研究现状	3
1.3.3 近红外光谱技术在含水率检测方面研究现状	4
1.3.4 近红外光谱融合机器视觉技术茶叶成分检测方面研究现状	5
1.4 研究内容	6
1.5 本章小结	7
第 2 章 基于近红外光谱的绿茶摊青过程中水分含量检测模型建立	8
2.1 引 言	8
2.2 材料与方法	8
2.2.1 试验材料	8
2.2.2 试验设备	8
2.2.3 试验方案	10
2.2.4 光谱预处理方法	11
2.2.5 特征波长筛选方法	12
2.2.6 定量预测模型建立	13
2.2.7 模型性能评价指标	13
2.3 结果与讨论	14
2.3.1 样品集的划分及水分含量分析	14
2.3.2 近红外光谱特征和预处理优化	15
2.3.3 特征波长优选	17
2.3.4 水分含量相关的变量响应	20
2.3.5 模型的建立与优选	20
2.4 本章小结	22

第3章 基于多源信息融合的绿茶杀青叶水分含量预测模型建立	24
3.1 引言	24
3.2 材料和方法	24
3.2.1 样品制备	24
3.2.2 仪器仪表和数据采集	24
3.3 特征提取和数据融合	26
3.3.1 近红外光谱处理	26
3.3.2 图像特征	27
3.3.3 数据融合	28
3.4 多变量数据分析方法	28
3.4.1 主成分分析	28
3.4.2 定量预测建模方法	28
3.4.3 模型评价指标	28
3.5 结果与分析	29
3.5.1 光谱特征和预处理优选	29
3.5.2 颜色特征变化	31
3.5.3 光谱特征波长筛选	33
3.5.4 基于单一传感器的预测模型	34
3.5.5 基于数据融合的预测模型	36
3.6 本章小结	37
第4章 基于近红外光谱的绿茶干燥程度判别模型建立	39
4.1 引言	39
4.2 材料和方法	39
4.2.1 试验材料	39
4.2.2 试验设备及工艺流程	39
4.2.3 光谱采集	40
4.2.4 数据分析方法	40
4.3 数据分析与处理	41
4.3.1 光谱特征与预处理方式	41
4.3.2 主成分分析	43
4.3.3 基于不同预处理方法的线性判别	44
4.3.4 特征波长筛选	45
4.3.5 基于不同变量筛选的 KNN 判别模型	47
4.3.6 基于不同变量筛选的 ELM 判别模型	49

4.3.7 PLS-DA、KNN 和 ELM 模型的比较	50
4.4 本章小结	51
第 5 章 基于机器视觉和近红外光谱的成品绿茶水分含量预测模型建立	52
5.1 引言	52
5.2 材料和方法	52
5.2.1 样品制备	52
5.2.2 光谱数据获取与水分含量测定	52
5.2.3 机器视觉系统的搭建和光谱、图像采集	53
5.2.4 NIR 数据预处理与图像特征提取	54
5.2.5 NIR 特征波长筛选与特征提取	54
5.2.6 光谱与图像特征信息数据融合	55
5.2.7 模型的建立和性能评估	55
5.3 结果和讨论	55
5.3.1 光谱特征和预处理结果	55
5.3.2 加工过程中颜色变量响应	57
5.3.3 NIR 特征波长提取	60
5.3.4 基于单个传感器数据的预测模型	61
5.3.5 基于数据融合的预测模型	62
5.4 本章小结	64
第 6 章 结论与展望	66
6.1 结论	66
6.2 展望	67
参考文献	68
致 谢	73
作 者 简 介	74
石河子大学硕士研究生学位论文	76
导师评阅表	76

第1章 绪论

1.1 研究背景

茶叶是我国重要的经济作物，茶产业是传统的优势特色农业产业^[1]，茶叶加工业是我国传统特色食品加工业的重要组成部分，茶叶已深深的融入了人民的生活，成为传承中华文化的重要载体^[2]。随着人们生活水平的不断提高，对健康的重视程度日益提升，饮茶作为一种健康的生活习惯，符合现阶段消费者对健康和高品质生活的追求。2021年我国茶叶产量为318万吨，较2020年的293.18万吨，同比增长8.47%，其中绿茶产量为184.94万吨，出口量为31.23万吨，占总口量的84.5%。无论是产量、消费量还是国民接受度，绿茶都是我国茶叶第一品类^[3, 4]。我国绿茶产区分布广泛，主要分为四大产区：华南茶区、江南茶区、西南茶区和江北茶区，绿茶细分品种众多，涌现出西湖龙井、碧螺春、信阳毛尖、黄山毛峰、太平猴魁、六安瓜片、庐山云雾等优质名茶。

绿茶中富含茶多酚、咖啡碱、儿茶素等营养成分，具有防癌、减肥、灭菌、抗衰老等功效，为其他茶类所不及^[5, 6]。绿茶属于不发酵茶，其加工的实质是逐渐失水并成型的过程^[7]，采摘鲜叶需经过摊青、杀青、揉捻、干燥等加工工艺最终形成成品茶，但无论是哪种加工工序，水分含量是设定工艺流程的重要参数。以摊青工艺为例，该过程不仅为杀青工艺中钝化茶叶中酶活性做准备，还能使鲜叶中的理化成分发生一定的变化^[8]，而水分含量是控制绿茶摊青程度的主要指标，然而目前绿茶摊青适度的主要判断标准包括摊青叶的水分含量和感官品质，感官品质的判定主要依赖于人工判别，而人工判别存在一定的误差，水分含量主要以抽样检测为主，存在效率低下无法实现快速、无损检测的需求。以绿茶杀青工艺为例，杀青是绿茶初制的关键工序，对成品绿茶的外形、色泽、香气和滋味等风味品质的形成影响显著。杀青工艺是采取高温的方法，减少茶叶中的水分含量，钝化杀青叶中的酶活性，并使加工叶中的内质成分发生一定的化学反应，从而形成绿茶所需的品质特征。除高级名优绿茶采用手工杀青之外，茶叶杀青一般采用滚筒杀青等方式进行机械杀青。绿茶杀青时多根据杀青叶色泽变化多根据杀青叶色泽变化和手握的触感，判断杀青叶水分含量情况和杀青程度，如适度的杀青叶色泽暗绿，手捏叶质柔软，略有黏性。但是人工经验法属于专家行为，易导致在制品发生红边、焦叶或杀不透现象，造成品质不稳定和鲜叶原料的浪费。以干燥工序为例，干燥的目的是使茶叶水分降至7%以下，如果茶叶水分含量较高，即使茶叶

在密封状态下，也会造成茶叶品质的变化，不利于茶叶的存储，所以成品茶中水分含量对保持茶叶的品质成分具有十分重要的作用。因此想要提高绿茶的加工质量并精准的掌握整个绿茶加工过程，必须实现绿茶加工过程中水分含量的精准、快速检测。

近红外光谱技术具有高效、无损和精准等优点^[9, 10]，已普遍应用于茶叶内质成分的无损检测中，对摆脱人工操作和经验判别的看茶做茶模式，以及缺乏对整个加工过程中水分含量进行量化的科学客观评价标准，提供了重要理论以及，对实现绿茶加工自动和智能化具有重要的促进作用。

1.2 研究目的及意义

本研究将从绿茶加工过程和工艺技术理论出发，搭建绿茶加工过程中水分检测智能感知系统，以不同加工工艺流程中的摊青、杀青、干燥叶等为研究对象，进行基于近红外光谱技术的水分含量定量预测模型构建，实现对绿茶加工过程中水分含量的智能感知。为提高绿茶加工品质，满足现代名优绿茶的智能化、自动化加工需求提供了先决条件。

本研究成果将为今后绿茶加工过程水分检测智能化监控设备的研制，提供理论基础与科学依据。为提升我国绿茶加工技术的智能化和信息化发展建立基础，对提高我国绿茶的品质在国际市场中的经济地位有深远意义。

1.3 国内外研究现状

1.3.1 绿茶加工设备的研究现状

在绿茶加工技术和装备领域，日本茶叶生产机械化和自动化程度最高。在上世纪80年代，已建立120 K(产量120kg/h)、180 K和240 K系统配套的系列蒸青绿茶（煎茶）自动生产线，具有茶鲜叶质量评价装置和全过程水分、温度、风量等关键参数的在线检测，实现了煎茶生产全过程的自动检测与控制；近年来，日本茶机生产企业还针对中国市场开发了炒青绿茶全自动生产线、针芽形名优绿茶全自动生产线、颗粒形名优绿茶全自动生产线以及工夫红茶自动生产线，不断抢占我国茶机装备的市场。近年来我国以促进茶叶加工省力化、产品优质化为目标，在茶叶加工技术理论、茶叶加工工艺升级、设备研制等方面展开了系统研究，茶叶加工关键工序的工艺特性及在制品理化变化规律日益明确；加工技术得到了多层次、多方位的快速发展，如设施摊青萎凋

技术、电磁内热杀青技术、全自动揉捻技术、远红外提香技术、精制色选技术等取得了突破；加工装备不断升级，已实现手工制茶向机械化、连续化制茶方式的转变，研制了适应不同茶类的关键装备和生产线，实现了大宗茶连续化自动化加工，正逐步向智能化加工技术与装备的推进和转型升级中。

随着自然资源的日益紧缺和人力成本的不断提高，采用高效、节能、智能型设备代替人工作业，已成为茶叶加工业发展的必然趋势。传统茶机生产企业因受技术实力、创新理念等的局限，难以单独实现加工设备的提档升级，需要借助食品机械、机电一体化、电子信息、物理化学等多学科的融合，通过创新研究在热能转化、数据信息采集、智能调控等领域取得突破进展，才能够让高端智能加工设备真正实现产业化应用。

1.3.2 茶叶中水分含量检测研究现状

在绿茶生产加工过程中，水分含量是设定加工工艺流程和工艺参数、控制茶叶原料形态变化和内质成分反应的重要依据^[1]。因此，实现绿茶水分含量的实时、快速、准确和无损检测，对促进标准化绿茶生产加工和提高成品茶品质具有重要作用。目前，对茶叶含水率检测主要分为直接检测法和间接检测法。

直接检测法可分为经验法、国标法和快速水分测定法。经验法是制茶师傅通过用手触摸茶叶，感受茶叶的质感和干湿程度，从而判断茶叶中的水分含量。经验法在传统手工制茶方面较为普遍，但是该方法受主观因素影响和外界环境影响较大，水分含量的判定存在较大的误差，无法实现茶叶的标准化生产^[12, 13]。国标法是将茶叶样品置于 $103\text{ }^{\circ}\text{C} \pm 2\text{ }^{\circ}\text{C}$ 的恒温干燥箱之中加热至恒重，称量，通过计算茶叶样品的质量损失从而得出样品含水率^[14]。国标法可以测量水分含量 3 %-75 %范围内的茶叶样品，测量的准确性较高，重复性较好，误差较小，但是国标法测量时间较长，每测量一次需要耗时 4 h 左右，测量时间过长，无法实现实际生产中快速检测的需求。快速水分测定法主要通过快速水分测定仪来实现，该方法相比于国标法可以显著减少测量时间，并且操作简单，测量精度和准确性较高，但是实时性较差，测量时间在 5-15 min 之间^[14]，且茶叶样品水分含量越高，测量时间越长，且存在有损的缺点，无法满足时间生产中快速、无损检测的需求。

间接检测法包括电测法、微波法、低场核磁共振技术测定法、近红外光谱法和机器视觉法等方法。电测法是通过传感器将水分转化为电量的方式进行测量，主要分高频阻抗法、交流电容法和直流电阻法。高频阻抗法是根据在敏感频带（100 k-250 kHz）施以外加电场的情况下茶叶样品水分与其交流阻抗呈现对数关系这一理论来测量其水分的，该方法可以检测水分含量在 3 %- 75 %范围内的茶叶样品，但是该方法无法实现茶叶样品的在线测量，且易受茶叶等级和类型的影响，检测效果较低。交流电容法是

根据茶叶中水分含量的介电常数比茶叶中其它内质成分的介电常数较大,因此水分含量在产生变化的同时必将引起电容量的变化,所以,通过检测茶叶样品中水分含量的变化引起的电容变化即可得到茶叶样品中的水分含量。此方法可进行在线测量,但是该方法适用于检测水分含量降低的茶叶样品,对于水分含量较高的加工过程样难于进行检测。直流电阻法根据茶叶中水分含量,从而测量出样品的电阻,即可以间接的测量水分含量,该方法水分含量越高则电阻越小,存在的检测信号就越弱,抗干扰能力就越弱,因此不适用检测含量较高的样品。微波法主要通过茶叶样品中水分含量对能量的衰减,从而间接检测出茶叶中的水分含量^[14]。微波法测量含水分含量具有快速、无损的优点,但是该方法检测茶叶水分含量范围在 0.3 %-30 %,对于绿茶摊青、杀青和揉捻过程中的水分检测并不适用。低场核磁共振技术测定法是将磁场的强度保持在 0.5 T 以下,该方法首先将标定不同重量纯水标样 ($n \geq 5$) 的自由感应衰减 (Free induction decay FID) 信号幅值,拟合出水分含量与 FID 信号幅值之间的直线方程,然后取一定质量 (1-2 g) 茶叶标样 ($n \geq 5$),分别测量茶叶标样的 FID 信号幅值,再根据拟合的直线方程换算成茶叶含水率^[15, 16],低场核磁共振可以实现无损检测的需求,精度和稳定性相对较高,但是该方法实时性较差,且每次检测的样本量有限,并不适用于实际生产中^[17]。

1.3.3 近红外光谱技术在含水率检测方面研究现状

近红外光谱技术是一种常用的无损检测方法,具有快速、高效、无损的优点,已在农产品和茶叶领域得到了广泛应用^[18, 19]。王文霞等^[20]为实现对哈密大枣中水分含量的准确预测,运用近红外 (869.19~2501.14 nm) 检测技术检测干制哈密大枣中的水分含量,为减少谱线重叠、噪音信号和基线漂移等问题,该研究采用标准正态变换 (standard normal variate, SNV)、多元散射校正 (multiplicative scatter correction, MSC) 和矢量归一化 (normalization, Norm) 三种光谱预处理方法对原始光谱进行校正,为了进一步提高模型的预测精度、减少光谱中的冗余信息、提高检测速度,通过采用联合区间偏最小二乘 (synergy interval partial least squares, Si-PLS)、遗传算法 (genetic algorithm, GA) 和连续投影算法 (successive projections algorithm, SPA) 及其组合算法从全波段中提取与哈密大枣水分含量相关的特征波长,随后建立 GA-ELM 预测模型。经比较发现,基于 SNV 预处理和 SPA 特征波长筛选方法所建立的额 GA-ELM 模型效果最佳,其校正集相关系数 (R_c) 和预测集相关系数 (R_p) 分别为 0.9842 和 0.9675,校正均方根误差 (RMSEC) 和预测均方根误差 (RMSEP) 分别为 0.0061 和 0.0079,相对标准偏差 (RPD) 为 3.6788,表明模型的预测效果较好。Duthen Simon 等^[21]为了快速、无损检测明胶中的的水分含量,利用近红外光谱仪 (800-2781 nm) 采集 135 种工

业明胶的反射光谱。为了减少光散射影响,突出一些波峰并且消除基线漂移噪声,分别采用无预处理、一阶导数变换、二阶导数变换、标准正态变换和标准正态变换加一阶导数变换五种预处理方法对原始光谱进行预处理操作,为了降低数据维度,对预处理后的光谱信息采用主成分分析(PCA),并结合偏最小二乘回归算法,建立明胶水分含量预测模型,其中 Raw+PCA+PLSR 获得了最佳的模型预测效果, R^2 值为 0.93, RMSEP 值为 0.18%, RPD 值为 3.9, RPD 较高,因此水分模型可用 ZA 于定量应用。Bing Lu 等^[22]以可可豆基质为研究对象,运用近红外光谱仪快速采集可可豆基质的反射光谱,利用烘干法制备了 120 个不同水分梯度的样品。为了减少背景噪声、基线漂移和光谱散射对光谱检测的影响,采用 Savitzky-Golay (S-G) 平滑、标准正态变量变换 (SNV)、多元散射校正 (MSC) 和导数结合 S-G 平滑对光谱数据进行预处理。为了减少无关的冗余信息,提高检测精度,采用 SPA, UVE 和 Si-PLS 三种方法筛选与水分相关的特征波长,通过比较发现,基于 S-G 平滑结合 SPA 特征波长筛选方法所建立的 MLR 获得了最好的模型效果,其中 R_c 值为 0.9976, R_p 值为 0.9963, RPD 值为 11.28。结果表明,采用近红外光谱技术可以实现可可豆基质中水分含量的精准检测。

在茶叶水分含量检测方面,陈琳等^[23]为了快速检测工夫红茶干燥过程种中的水分含量,通过利用近红外光谱仪器,采集工夫红茶 6 次干燥过程中的 226 个样品的光谱信息为研究对象,结合标准正态变量变换 (SNVT) 预处理方法、偏最小二乘法 (PLS) 及联合区间偏最小二乘法 (Si-PLS) 建模方法建立工夫干燥过程中的水分含量预测模型。结果表明,基于 SNVT 变量筛选方法所建立的 Si-PLS 模型获得了最佳的预测效果,其预测集相关系数 (R_p) 为 0.9593,预测集均方根误差 (RMSEP) 为 0.0395,表明模型具有较好的预测效果,可以实现工夫红茶干燥过程中水分含量的精准、在线检测。白晓丽等^[24]采用近红外光谱仪器,采集随机选取的 42 个市售速溶茶普洱茶的近红外光谱信息,根据国标法测定茶叶样品的含水率,采用平滑 (smoothing)、导数 (derivative)、标准正态变量变换 (SNV) 和多元散射校正 (MSC) 四种预处理方法减低原始光谱中的噪声信息,然后建立偏最小二乘 (PLS) 水分含量预测模型。结果表明,校正集相关系数 (R_c) 和校正集均方根误差 (RMSEC) 值分别为 0.9266 和 0.6439,验证集相关系数 (R_p) 值和为验证集均方根误差 (RMSEP) 值预测相关系数分别为 0.9050 和 0.6786,表明模型可以满足速溶普洱茶水分含量定量预测要求,具有较好的预测效果。以上研究表明,近红外光谱结合化学计量学方法和机器学习算法,可以实现茶叶中水分含量的在线检测。

1.3.4 近红外光谱融合机器视觉技术茶叶成分检测方面研究现状

近红外光谱和机器视觉技术已成为两种应用较多的农产品品质成分无损检测方法,

为了提高模型预测精度,克服单一传感器预测精度不够的问题,近红外光谱结合机器视觉技术在食品的品质及成分无损检测方面得到了更为广泛的应用。王巧华等^[25]提出了利用机器视觉和近红外光谱技术检测皮蛋品质的可行性,以采集的皮蛋透射光图像,提取图像的18个颜色特征值作为模型的输入,建立GA-SVM分级模型,把皮蛋样本分为可食用蛋与不可食用蛋,劣质蛋测试集识别率为100%,然后利用采集的可食用蛋近红外光谱建立MSC+CARS+SVM模型区分出优质蛋与次品蛋,优质蛋测试集识别率为96.49%,次品蛋识别率为94.12%。Yan Song等^[26]提出一种联合使用近红外光谱技术和计算机视觉系统评估茶叶风味和外观质量的快速分级方法,使用Bruker-MPA傅里叶变换近红外光谱仪记录样品的光谱,使用计算机视觉系统对7个不同等级的茶叶样品采集图像,结果表明,近红外光谱和CVS特征的融合,校准、验证和测试集的准确性从99.29%、96.67%和98.57%提高到100.00%、99.29%和100.00%。Yujie Wang等^[27]为了提高多酚和儿茶素在加工阶段的快速检测,基于微型近红外光谱和自建计算机视觉系统,采集茶叶的萎凋、揉捻、发酵和干燥步骤的样品数据,将来自两个传感器的数据进行融合,分别采用CARS和PCA从光谱和颜色变量中选择有效变量,采用PLS进行建模,结果表明,预测的平均均方根误差分别为 0.66 ± 0.12 和 1.06 ± 0.11 g/100 g,剩余预测偏差分别为 5.41 ± 0.99 和 4.03 ± 0.38 。然而,融合近红外光谱和机器视觉技术在绿茶加工过程中水分含量的定量预测方面相关研究较少,研究成果大多集中于红茶品质的定性判别上,尚未评估对机器视觉与近红外光谱进行数据融合预测绿茶加工过程中含水率定量预测的可行性。

1.4 研究内容

本课题的主要研究内容是如何利用近红外光谱技术实现绿茶加工过程中每个加工阶段水分含量的实时在线检测,以近红外光谱仪器检测绿茶加工过程中茶叶水分含量代替人工经验判别所获得的结果,作为绿茶加工过程中的重要工艺参数,为解决绿茶加工工艺参数不统一以及克服茶产业无法满足标准化生产的局限,对改善绿茶加工质量和提高产量具有重要的意义。本文的主要研究内容如下:

(1) 建立绿茶摊青过程中水分含量在线检测模型

利用近红外光谱仪器(900-1700 nm)采集不同摊青时序下样品的近红外光谱信息,为了提高模型发的预测精度,消除光谱中噪声的影响,对光谱信息采用多种预处理方法,并通过建立线性模型,优选出最佳预处理方法;为了简化模型,提高茶叶的检测效率,采用不同的变量筛选方法筛选出与水分含量相关的光谱特征波长,并建立PLSR

和 SVR 定量预测模型，对比模型性能，优选出摊青过程中水分含量最佳定量预测模型，以后续杀青工序的进行奠定研究基础。

(2) 建立基于光谱特征和图像色泽、纹理特征杀青水分关联模型

利用近红外光谱仪器和搭建的机器视觉采集系统，对杀青过程中的茶叶样品进行光谱和图像信息采集，并分析杀青过程中的光谱信息特征和图像的色泽、纹理变化与水分含量之间的相关关系。对采集的光谱信息和图像信息进行数据融合，探索近红外光谱融合和机器视觉技术对实现绿茶杀青过程中水分含量精准预测的可行性，通过不同的数据融合方式和人工智能算法，探索最佳的水分含量预测模型。

(3) 建立不同干燥程度的水分含量判别模型

利用近红外光谱仪器采集不同干燥程度下的光谱信息，并根据茶叶的水分含量将绿茶干燥程度分为3类（干燥不足、干燥适度、干燥过度）。分析了不同预处理方法、特波长筛选方法和判别模型对绿茶干燥程度判别效果影响。

(4) 建立基于近红外光谱和机器视觉技术的成品茶水预测模型

以另一茶叶品种的不同工序下的茶叶样品为研究对象，通过搭建的近红外光谱和机器视觉系统，实现从鲜叶到成品茶的水分含量预测模型，验证基于数据融合方式对绿茶加工过程水分含量预测的可行性。

1.5 本章小结

本章介绍了研究背景和研究意义，介绍了现阶段绿茶加工装备以及加工过程中水分含量检测方式存在的问题，通过对基于近红外光谱技术及机器视觉技术检测含水率的相关文献的研究和概括，论证了近红外光谱技术及近红外光谱融合机器视觉技术对茶叶水分含量实现快速、无损检测的可行性。并基于绿茶加工工序中水分检测存在的问题，提出了本文的研究内容，以不同加工工序下的光谱和图像信息作为研究对象，实现各个工艺条件下的水分检测，为绿茶加工品质的提高和标准化茶叶生产奠定了理论基础。

第2章 基于近红外光谱的绿茶摊青过程中水分含量检测模型建立

2.1 引言

摊青作为名优绿茶加工过程中的首道工序，良好的摊青方式不仅可以提高成品绿茶的品质，还可以延长鲜叶的贮藏保鲜时间，对后续加工工序有着重要影响。绿茶摊青以叶面开始萎蔫，叶质由硬变软，叶色由鲜绿转暗绿，青气消失，清香显露，摊放叶含水率降至 $(70 \pm 2)\%$ 为适度。水分含量作为衡量绿茶摊青的重要品质参数，其含量将直接影响茶叶中营养成分的变化，对下一步杀青工序也会产生的，最终影响成品茶的质量。因此实现绿茶摊青过程中，水分含量的实时、准确、无损检测，对于提高在制茶含水率的控制水平、标准化茶叶生产具有重要的意义。

2.2 材料与方法

2.2.1 试验材料

本研究选用的茶叶样品采摘于湖南省长沙市高桥镇，茶鲜叶品种为高桥银峰茶，鲜叶嫩度为单芽、一芽一叶和一芽二叶，其中以一芽一叶和一芽二叶为主，采摘时间为2020年5月25日

高桥银峰茶，产于湖南省长沙市东郊玉皇峰下的高桥，是湖南省茶叶研究所于1957年创制，其外形条索匀净紧细而略微卷曲、外披茸毛、色泽翠绿欲滴、香气持久鲜嫩、茶汤清亮、滋味醇鲜、叶底明亮嫩绿等品质特色，适用于制作绿茶，属于绿茶类。

2.2.2 试验设备

(1) 摊青设备

本研究所用绿茶摊青设备为长沙湘丰智能装备有限公司研制，已广泛应用于绿茶连续化生产作业中。该摊青机为网带式框架结构，在网带下方具有独立的供风腔体，

在茶叶生产过程中，可以通过提供不同的供风方式（自然风、暖风）改变茶叶的摊青状态，同时，通过操作面板可以监控绿茶摊青过程中的茶叶温度，根据摊青温度，保持合理的摊青方式。本研究在绿茶摊青过程中适当的提供自然风，以达到鲜叶摊放，实现摊青的功能，摊青装置如图 2-1 所示。



图 2-1 茶叶摊青设备

Fig.2-1 The tea spreading equipment

(2) 近红外光谱采集设备

本研究采用 IAS-3100 近红外光谱仪（无锡迅杰光远科技有限公司）对摊青时序下的样品进行光谱信息采集。仪器波长范围为 900~1700 nm，光谱精度为 ± 1 nm，采样方式为下照漫反射，样品池尺寸为直径 12.5cm。每个时序下取 20 个样品，每个样品重约 80 g 置于样品池中，经“翻拌-扫描”进行 3 次光谱采集，取 3 次光谱均值作为该样品的光谱信息，共摊青 10 个小时，获得 200 条光谱信息，近红外光谱仪器如图 2-2 所示：



图 2-2 近红外光谱仪器

Fig.2-2 Near-infrared spectrometer

(3) 水分检测装备

本研究采用赛多利斯水分测定仪（MA35M-000230V1，Sartorius）测定绿茶摊青过程中的水分含量，其工作原理为干燥失重法，通过恒温加热处理使进行水分失重，并通过水分测定仪内部的电子称重系统，计算出水分值含量。每次采集完光谱数据后取3克茶叶样品放入到水分测定仪中测量其水分值，记录每次所得的测量结果，该过程重复三次，取三次所测定结果的平均值作为该时刻下的样品的水分值。赛多利斯水分测定仪如图2-3所示。



图 2-3 水分测定仪

Fig.2-3 The moisture analyzer

2.2.3 试验方案

本研究以绿茶鲜叶为研究对象，茶叶品种为高桥银峰茶，嫩度主要为一芽一叶和一芽二叶。摊青过程在摊青机上进行，保持摊青房室内温度为20℃左右，相对湿度为60%左右。摊青过程中，每隔一小时取一次样品，每次取样分为20组，每组样品采集三次光谱数据，以三次采集的光谱数据的平均值作为该组的光谱信息。将每个时刻采完光谱数据后的茶叶样品使用水分测定仪测量该时刻下的水分含量，该过程重复三次，取每时刻三次测定的水分含量的平均值作为该时刻的水分值。摊青10小时后，摊青水分76.52%降至65.05%，共获得10个时间节点下的200条光谱数据。本章整体研究思路如下图2-4所示。

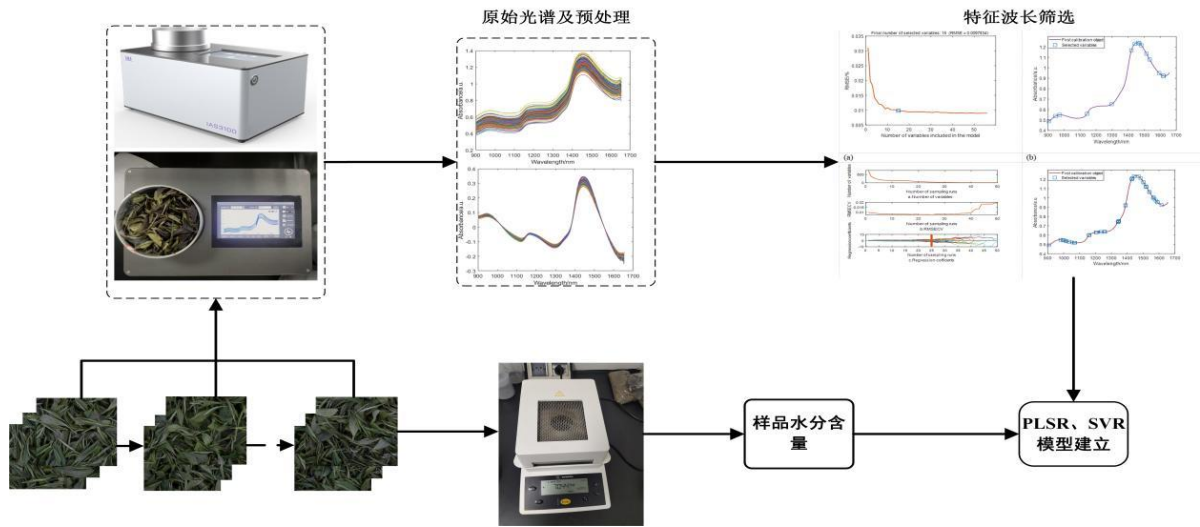


图 2-4 技术路线

Fig.2-4 Echnology roadmap

2.2.4 光谱预处理方法

在采集近红外光谱时，会受到仪器的基线漂移、茶叶样本表面不均匀、谱线重叠、随机噪声、光散射等问题使得采集的原始光谱中包含了一些无用的信息，影响建立的模型效果。为了消除干扰，提高模型的预测精度和建模效果，本研究应用采用标准正态变量（standard normal variables, SNV），均值中心化（mean centering, MC），最大最小归一化（max-min normalization, Max-Min），去趋势（De-trending, DT），最小二乘平滑滤波（Savitzky-Golay filtering, SG FIR）5 种不同预处理方法。

（1）标准正态变化（standard normal variables, SNV），适用于对茶叶样品因散射而引起误差进行校正，该方法对每条光谱都进行标准化处理^[28]。标准正态变化与标准化预处理有着相似的计算方式，主要区别在于标准化是对所获得的的光谱数据矩阵进行列处理，而SNV则是对光谱数据矩阵进行行处理^[28]。SNV的计算公式如下：

$$x_{mk, SNV} = \frac{x_{mk} - \overline{x_m}}{\sqrt{\frac{\sum_{k=1}^n x_{mk} - \overline{x_m}}{n-1}}} \quad (2-1)$$

式中 $x_{mk, SNV}$ 表示经过SNV预处理后第 m ($m=1, 2, \dots, n$) 个光谱样本在第 k

($k=1,2,\dots,n$) 个波数点处的光谱值; x_{mk} 则表示第 m 个光谱样本在第 k 个波数点处的光谱数据; $\bar{x}_m$ 则表示第 m 个光谱样本的光谱在全部 n 个波数点处光谱数据的平均值。

(2) 均值中心化 (mean centering, MC) 是指每个茶叶样品的光谱数据减去该样品的平均光谱, 以消除光谱的绝对吸光度值。采用均值中心化预处理方式不仅可以增强光谱数据之间的差异, 而且还可以将光谱数据特征变化与待测茶叶水分含量之间的变化相关联, 突出光谱的变化特征, 因而提高模型的预测精度^[29]。在对茶叶样品的理化成分进行定量和定性分析时, 光谱均值中心化最为常用。MC的计算公式如下:

$$x'_i = x_i - \bar{x} \quad (2-2)$$

其中, $\bar{x}$ 表示 $n \times m$ 维的茶叶样品原始光谱矩阵的均值, x_i 表示第 i 个采集的茶叶样品的光谱数据。

(3) 最大最小归一化 (max-min normalization, Max-Min) 也称为离散标准化, 是对原始光谱数据的线性变换, 使得光谱数据结果映射到 $[0,1]$ 之间, 转换函数如下:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (2-3)$$

其中, $x_{\max}$ 为茶叶样本原始光谱数据的最大值, $x_{\min}$ 为茶叶样本原始光谱数据的最小值。

(4) 去趋势 (De-trending, DT) 方法主要用于消除光谱中的基线漂移, 采用多项式最小二乘方法将茶叶光谱的吸光度值与波长进行拟合从而消除光谱数据中的非线性变化, 通常在进行去趋势预处理之前首先进行SNV预处理, 但可以单独使用。

(5) 最小二乘平滑滤波 (Savitzky-Golay filtering, SG FIR) 是光谱分析中常用的数据降噪方法, 该方法在保证原始光谱数据尽量不失真的情况消除光谱中的噪声信息。该算法对茶叶样品原始光谱数据矩阵每一列进行操作, 平滑效果随着选取的窗宽不同而发生改变, 但是多项式阶次 (order) 必须小于窗宽 (framelen) 的长度, 因此framelen必须为奇数, 如果order=framelen-1, 则滤波器不进行平滑。

2.2.5 特征波长筛选方法

本实验中光谱包含了 755 个特征变量, 数据维度较高, 在建立预测模型时, 若模型中存在较多与绿茶摊青过程中水分含量无关的信息, 将大大降低预测模型的精度, 减低模型预测效果。因此需要从大量的光谱变量中提取与水分信息存在相关性的特征波长, 本研究采用连续投影算法 (SPA)、竞争性自适应加权算法 (CARS)、迭代保留信息变量 (IRIV)、变量组合聚类分析法 (VCPA) 以及结合遗传算法的

变量组合集群分析法（VCPA-GA）等方法来提取特征波长，达到简化模型提高模型精度的目的。

2.2.6 定量预测模型建立

本文采用偏最小二乘回归（PLSR）和支持向量回归（SVR）算法建立定量预测模型。PLSR 模型是用于解决回归问题最广泛使用的线性算法之一，具有主成分分析以及典型相关分析的特点，通过从自变量（待测样品光谱数据）和因变量数据（待预测理化成分）中获取包含原数据变异信息的主成分来建立回归模型，PLSR 已在食品及农产品内质理化成分的定量预测方面得到了广泛应用；SVR 模型是一种基于 SVM 理论的非线性数据处理方法，其基本思想是通过在高维空间中进行线性回归，从而实现函数逼近。该方法通过引入非线性映射函数（核函数）的方式，将原始空间映射到高维特征空间。因此，低维空间中的非线性问题可以转化为高维空间中的线性问题，在这个空间中，构造了特殊类型的超平面，然后在超平面中建立回归模型。

2.2.7 模型性能评价指标

本研究采用校正集相关系数（ R_c ）、预测集相关系数（ R_p ）、交互验证均方根误差（RMSECV）、预测均方根误差（RMSEP）以及相对标准偏差（RPD）作为绿茶摊青过程中水分含量预测模型的评价指标。其中 R_c 与 R_p 值越大 RMSEC 与 RMSEP 越小则模型的预测和泛化性能越好^[28]。当 RPD 值大于 1.4 时，说明模型预测效果较差，无法用于定量预测；当 RPD 值在 1.4~1.8 之间，表明模型对检测指标有一定的预测能力；当 RPD 在 1.8~2 之间，表明模型具有较好的预测效果，可以用于品质成分的定量预测；当 $RPD > 2$ 时，表明所建立的模型具有极好的预测能力^[30]，计算公式如（1）~（5）所示^[28, 31, 32]。本实验所有数据处理与结果均在 Matlab R2019a 软件中进行。

$$R_c = \frac{\sqrt{\sum_{i=1}^{n_c} (y_{pi} - y_{mi})^2}}{\sqrt{\sum_{i=1}^{n_c} (y_{pi} - y_{mean})^2}} \quad (2-4)$$

$$R_p = \frac{\sqrt{\sum_{i=1}^{n_p} (y_{pi} - y_{mi})^2}}{\sqrt{\sum_{i=1}^{n_c} (y_{pi} - y_{mean})^2}} \quad (2-5)$$

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{n_c} (y_{pi} - y_{mi})^2}{n_c}} \quad (2-6)$$

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n_p} (y_{pi} - y_{mi})^2}{n_p}} \quad (2-7)$$

$$RPD = \frac{std_{prediction}}{RMSEP} \quad (2-8)$$

式中： y_i ：表示在划分的预测集中对第 i 个样品的预测值；

y_{mi} ：表示在划分的校正集中对第 i 个样品的实际测量值；

y_{mean} ：表示 n_c 个校正集样品或 n_p 个预测集样品实际测量值的平均值；

$std_{prediction}$ ：表示预测集的标准差。

2.3 结果与讨论

2.3.1 样品集的划分及水分含量分析

200 个样品集采用 K-S (Kennard-Stone) 算法进行校正集和预测集的划分, 按照校正集和预测集为 3:1 的比例挑选出 150 个样品作为校正集, 50 个样品作为预测集。通过水分测定仪测得的绿茶摊青过程中的全部样品, 以及校正集和预测集的水分信息如表 2-1 所示。水分含量与摊青时间呈负相关, 在初始摊青时含水率最高, 为 76.52%, 随着摊青时间的延长, 含水率逐渐降低, 到第 10 个小时含水率降至最低, 为 65.05%。

表 2-1 试验样本水分含量标准理化值统计结果

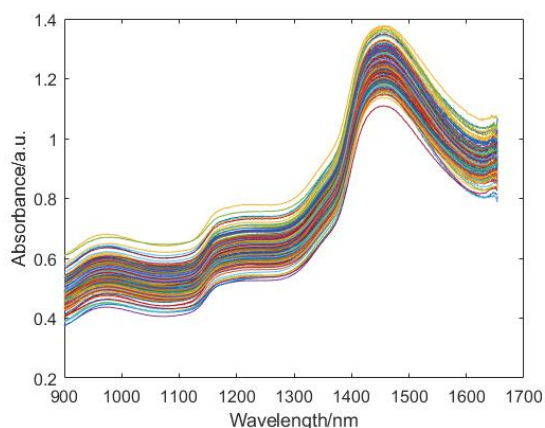
Table 2-1 Statistical results of standard physical and chemical contents of water content
in experimental sample

Sample set division	Moisture range/%	Moisture mean/%	Moisture variance/%	Number of samples
All sample	76.52–65.05	71.09	3.23	200
Calibration set	76.52–65.05	71.41	3.16	150
Prediction set	76.52–65.05	70.12	3.24	50

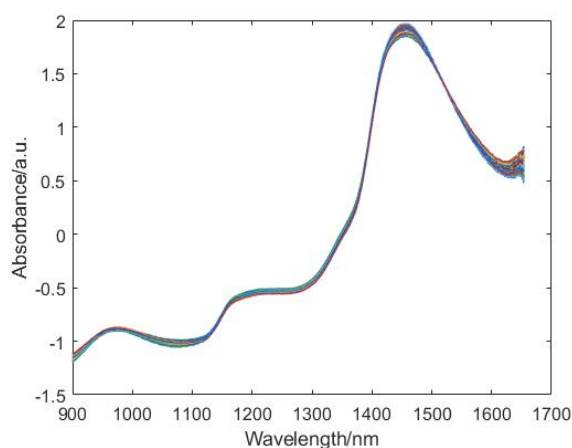
2.3.2 近红外光谱特征和预处理优化

由于1656 nm - 1700 nm范围内噪声影响明显，信噪比较低，因此，本研究仅选用900-1655 nm范围内的光谱作下一步分析。图2-5（a）为绿茶摊青过程中各个时序下的光谱曲线，由图2-5（a）可知，在960 nm、1180 nm和1440 nm处存在明显的波峰，这可能有由于鲜叶中水分子O-H基团的伸缩振动引起的。光谱曲线在1369 - 1440 nm范围内急剧上升，这可能是由于茶叶样品在此波段的吸收较少。

为了消除光谱中噪声信息的影响，优选出最佳的预处理方法，本研究基于原始光谱数据和预处理后的光谱数据建立PLSR预测模型，其模型效果如表2-2所示，预处理后的光谱如图2-5（b）、图2-5（c）、图2-5（d）、图2-5（e）和图2-5（f）所示。其中基于原始光谱所建立的PLSR模型中 R_c 和 R_p 值分别为0.9366和0.9382，RMSECV和RMSEP分别为0.0111和0.0113，RPD为2.6511。对比于基于原始光谱所建立的PLSR模型性能，经不同预处理方法所获得的模型效果中， R_c 、 R_p 和RPD值均有所提高，说明光谱经过预处理后，模型性能有了进一步的提升。其中，基于SG FIR预处理后所建立的PLSR模型效果最佳，其 R_c 和 R_p 值分别为0.9558和0.9507，RPD为3.2810。因此本章选用SG FIR作为最终的预处理方法，并进行下一步的建模分析。其中SG FIR中多项式阶数取3，框长度取11。



(a)



(b)

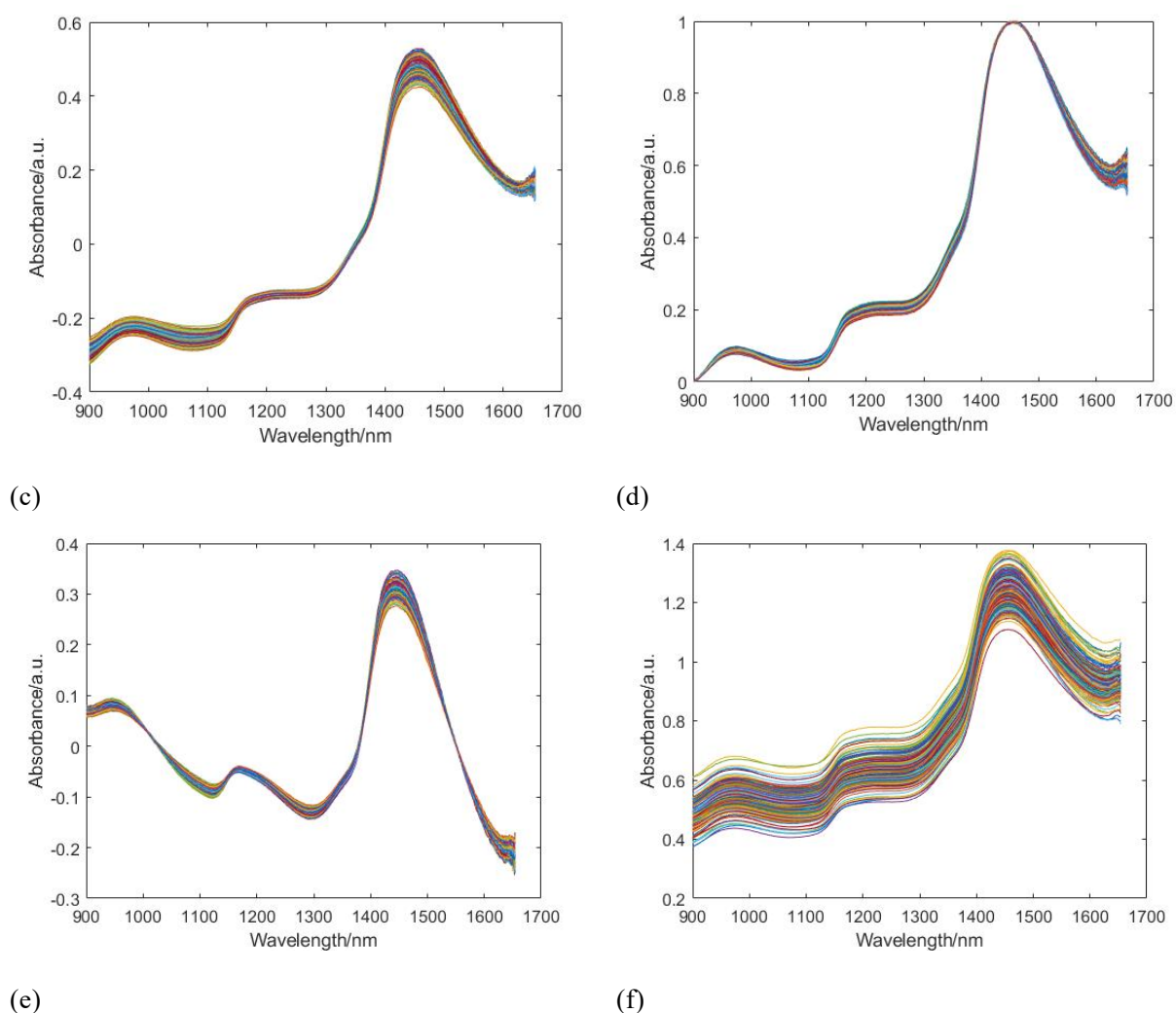


图 2-5 近红外原始光谱及预处理后的光谱图

Fig.2-5 NIR raw spectra and various preprocessed spectra

表 2-2 不同预处理方法下的 PLSR 水分含量预测结果

Table 2-2 Prediction results of PLS moisture content of different pretreatment methods

Pretreatment Methods	PCs	Calibration set		Prediction set		RPD
		Rc	RMSECV	Rp	RMSEP	
Raw	10	0.9366	0.0111	0.9382	0.0113	2.6511
SNV	9	0.9575	0.0094	0.9491	0.0102	3.0590
MC	7	0.9503	0.0099	0.9421	0.0112	2.9875

Max-Min	7	0.9575	0.0094	0.9491	0.0102	3.0590
Detrend	6	0.9543	0.0097	0.9451	0.0102	3.1143
SG FIR	6	0.9558	0.0093	0.9507	0.0093	3.2810

2.3.3 特征波长优选

1、基于 SPA 方法提取特征波长

SPA 是一种使矢量空间共线性最小化的前向变量选择算法^[33], 是近年来新兴的一种特征波长筛选方法, 它不仅可以有效地减少谱带数, 而且可以消除自变量的共线性, 避免模型过拟合问题^[34]。图 2-5 (a) 为交互均方根误差 (RMSE) 值随 SPA 选择变量增加变化图, 当特征波长增加至 15 后, RMSE 值基本保持不变, 按照特征波长被选中的概率降序排列: 946、1296、904、1592、972、1617、1470、1513、1419、1142、1533、1463、1620、1441、1481nm, 占全波段的 1.99%。特征波长筛选结果如图 2-6 (b) 所示。

2、基于 CARS 方法提取特征波长

CARS 是通过迭代蒙特卡罗重采样实现的, 该方法根据 PLS 回归系数 (RCs) 的绝对值来评价各个变量的重要性, 在 n 次采样过程中共选出 n 个变量子集, 并且通过交互验证的方式选出 RMSECV 值最低的子集, 从而有效寻出最优变量组合^[35]。在每次迭代过程中包括以下四个连续步骤: (1) 通过蒙特卡罗抽样建立 PLSR 预测模型; (2) 通过使用指数递减函数 (exponentially decreasing function, EPD) 消除 RCs 值相对较小的变量, (3) 在竞争模式下通过自适应加权抽样筛选变量; (4) 根据最小 RMSECV 计算每个迭代的子集^[36]。本研究设置采样次数为 50 次, 波长变量采用 10 折 PLS 交叉验证建模, 选取 RMSECV 最小的一次为最佳采样点, 当采样次数为 29 次时 RMSECV 最小, 筛选出的特征波长为 41 个 (6、7、8、47、48、61、62、80、95、97、216、225、260、334、335、336、357、359、399、447、491、519、520、521、522、532、544、564、570、571、609、617、618、619、649、650、666、682、693、717、744) 占全波段的 5.43%。特征波长筛选结果如图 2-6 (c) 所示。

3、基于 IRIV 方法提取特征波长

IRIV 算法是由 Yun 等^[37]提出的一种基于二进制矩阵变换滤波器 (binary matrix shuffling filter, BMSF) 的变量选择方法。该算法采用模型总体分析 (model population

analysis, MPA) 将所有变量划分为强信息变量、弱信息变量、非信息变量和干扰变量, 并以迭代方式去除非信息变量和干扰变量, 最后, 选择反向消除后的变量作为特征变量^[38]。本研究确定用 5 折交叉验证最大主成分因子数为 10 建立 PLS 模型, 以 RMSECV 作为评价指标, 从而选出最优变量。对于筛选出的新变量, 采用同样的方法保留变量, 最后通过反向消除法得到所需的变量。通过 10 此迭代后共选出特征波长 58 个 (7、81、82、95、96、103、114、115、132、133、155、166、167、169、170、259、260、261、304、305、306、307、334、335、336、337、356、358、359、442、443、444、445、446、447、489、490、529、530、531、532、544、571、585、602、603、618、619、620、639、640、664、665、670、681、690、691、717) 占全波段的 7.68%。特征波长筛选结果如图 2-6 (d) 所示。

4、基于 VCPA 方法提取特征波长

VCPA 是一种基于模型集群分析 (model population analysis, MPA) 思想的特征波长筛选方法, 该方法利用二进制矩阵采样法 (binary matrix sampling, BMS) 使每个变量有同等被选中的机会, 同时利用指数衰减函数 (exponentially decreasing function, EDF) 迭代进行特征波长筛选, 缩小变量空间。本文设置 BMS 采样运行次数为 1000, 设置 EDF 运行数为 50, 通过 VCPA 算法共筛选出特征变量 12 个 (7、49、51、61、335、359、446、531、532、618、691、717) 占全波段的 1.59%。特征波长筛选结果如图 2-6 (e) 所示。

5、基于 VCPA-GA 方法提取特征波长

VCPA-GA 算法是由 Yun 等^[37]提出的一种新的变量选择策略, 也是基于进化论中的自然选择和遗传机制。VCPA 可以通过不断将变量空间从大到小来优化变量, 然后使用 GA 进行进一步优化^[39]。本文设置指数衰减函数 (EDF) 为迭代次数为 50、最终剩余变量数为 100, 最优子模型的比例为 10%, BMS 运行次数为 1000, 并采用 5 折交叉验证的方式进行变量重要性评价。通过 VCPA-GA 算法共筛选出特征变量 54 个 (619、691、334、6、82、532、717、81、608、690、7、633、45、400、531、609、533、617、8、83、618、46、331、519、639、95、361、445、681、337、359、63、442、542、682、44、64、80、544、716、160、166、303、336、446、462、491、520、529、571、603、629、692、718) 占全波段的 5.15%。特征波长筛选结果如图 2-6 (f) 所示。

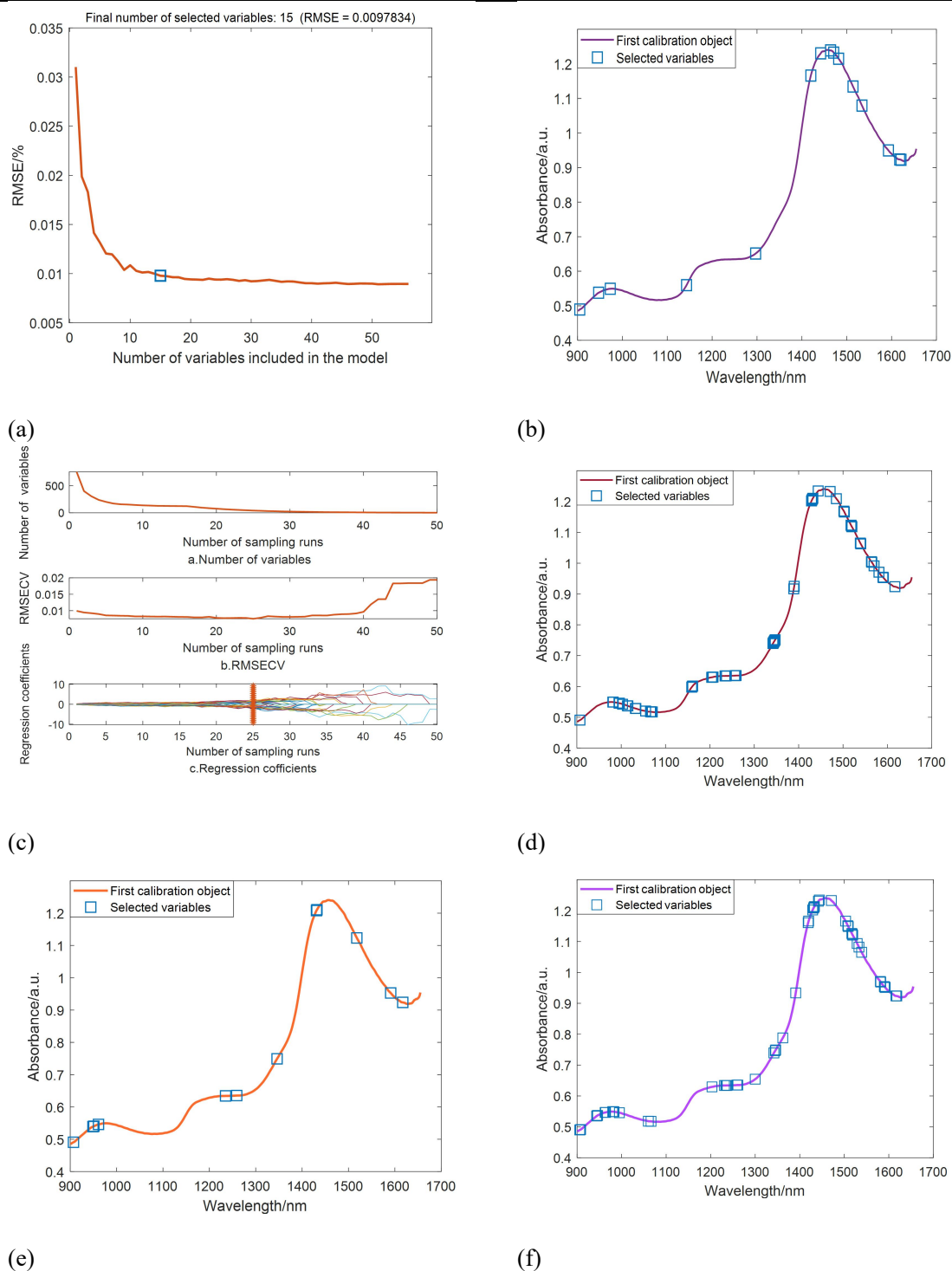


图 2-6 利用不同的特征波长筛选算法提取的特征波段

Fig.2-6 Feature bands extracted utilizing different characteristic wavelength screening algorithms: (a,b) Successive projections algorithm; (c) competitive adaptive reweighted sampling; (d) iterative retained information variables; (e) Variable combination population analysis; (f) variable combination population analysis combined with genetic algorithm

2.3.4 水分含量相关的变量响应

图 2-7 为不同特征波长筛选方法选择的特征变量在全波段光谱中的分布情况。从图 3 中可以看出, 950 (10530cm^{-1})、1430nm (6993cm^{-1}) 和 1600nm (6250cm^{-1}) 附近处的特征波长较为集中, 这是因为水分子中 OH 基团伸缩振动第一倍频在 1430nm (6993cm^{-1}) 附近, 第二倍频在 950nm (10530cm^{-1}) 附近。并且绿茶的摊青过程相对复杂, 当水分含量发生变化时, 也会产生氨基酸和其他有机化合物^[40]。NH 键是有机分子中的重要结构, 存在于氨基酸和蛋白质中。在摊青过程中, 由于蛋白水解酶活性的增加, 蛋白质被水解, 摊青叶中可溶性蛋白质和游离氨基酸的含量增加^[41]。NH 基团的第一个倍频吸收带位于 1500 (6666cm^{-1}) 附近, 因此在 1500 nm 筛选出一些特征波长。在绿茶摊青过程中, CH 可能产生一定的影响, 且 CH 键的一阶倍频为 1600 nm-1800 nm (6250cm^{-1} - 5556cm^{-1}), 二阶倍频为 1150 nm-1250 nm (8696cm^{-1} - 8000cm^{-1}), 三阶倍频为 890 nm-940 nm (11236cm^{-1} - 10638cm^{-1})。因此, 在 1150nm 和 1600nm 附近筛选出一些特征波长。

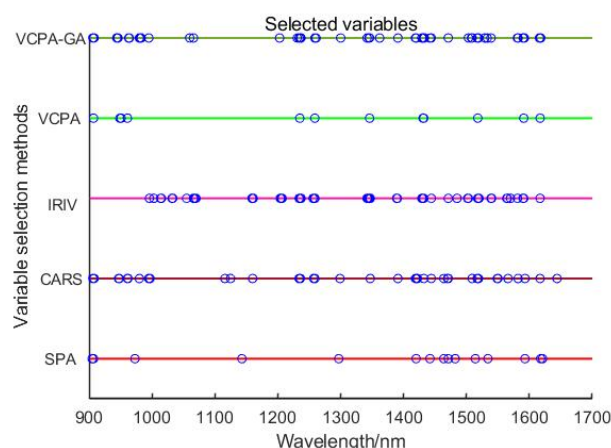


图 2-7 采用不同特征波长筛选方法选择的特征波长变量

Fig.2-7 Characteristic wavelength variables selected by different variable screening methods

2.3.5 模型的建立与优选

由于近红外光谱采集过程中受到仪器本身和外部环境的干扰, 存在一些非线性的因素, 因此在建立线性 PLSR 模型的同时需要建立非线性的模型, 以寻求较优的建模效果。SVR 可以有效地解决一些光谱数据中存在的非线性问题。基于上文介绍的特征波长优选方法, 在完成特征波长筛选之后, 将特征波长作为 SVR 预测模型的输入, 建立绿茶摊青含水率的预测模型, 并与所建立的 PLSR 模型相比较, 得出最优模型, 结果如表 2-3 所示。

表 2-3 基于不同变量筛选方法含水率的PLSR和SVR模型预测结果

Fig.3 Prediction results of moisture content PLSR and SVR models based on different variable screening methods

Model	Variable selection		Calibration set		Prediction set		RPD
	Methods	Numbers	Rc	RMSECV	Rp	RMSEP	
PLSR	全光谱	755	0.9558	0.0093	0.9507	0.0093	3.2810
	SPA	15	0.9622	0.0086	0.9544	0.0098	3.1227
	CARS	41	0.9760	0.0069	0.9615	0.0090	3.5030
	IRIV	58	0.9714	0.0075	0.9622	0.0090	3.4806
	VCPA	12	0.9658	0.0082	0.9532	0.0099	3.0578
	VCPA-GA	54	0.9716	0.0075	0.9601	0.0093	3.3546
SVR	全光谱	755	0.9711	0.0078	0.9598	0.0095	3.0293
	SPA	15	0.9652	0.0085	0.9575	0.0099	2.8415
	CARS	41	0.9735	0.0076	0.9627	0.0093	3.0690
	IRIV	58	0.9792	0.0067	0.9623	0.0091	3.2017
	VCPA	12	0.9720	0.0076	0.9619	0.0093	3.0696
	VCPA-GA	54	0.9785	0.0067	0.9658	0.0089	3.6787

结果显示, 对比 PLSR 模型效果可知, 基于全光谱所建立的 PLSR 模型结果, R_c 值和 R_p 值分别为 0.9558、0.9507, RMSECV 和 RMSEP 均为 0.0093, RPD 为 3.2810 (>2) 说明模型的预测性能较好。通过不同的特征波长选择方法, 模型性能会发生一定的变化, 其中基于 SPA 和 VCPA 特征波长提取方法所建立的 PLSR 模型性能相比于 CARS, IRIV 和 VCPA-GA 性能较低, 但 SPA、VCPA 的 R_c 值分别为 0.9622 和 0.9658, R_p 值分别为 0.9544、0.9563、0.9532, 均高于基于全波段所建立的模型性能 R_c 值和 R_p 值, 且 RPD 均大于 3, 说明模型性能较好。其中 SPA 和 VCPA 变量压缩率分别为 98.01%和 98.41%, 可能在筛选变量时剔除了一些有用波长, 导致 SPA 和 VCPA 模型性能低于 CARS, IRIV 和 VCPA-GA。其中基于 IRIV 所建立的 PLSR 模型性能最优, R_p 从 0.9507 提高到 0.9622, 均高于其他特征波长筛选方法所获得的 R_p 值, 且 RPD 为 3.4806 (>2), 表明模型预测效果极好。参与建模的变量数从 755 个减少到 58 个, 仅占全波段的 7.68%, 既可以在保证简化模型的前提下, 又可以更好的提高模型的预测

能力。其预测模型效果如图 2-8 (a) 所示。

对比 SVR 预测模型效果可知, 基于全光谱所建立的 SVR 模型, R_c 值和 R_p 值分别为 0.9711、0.9598, RMSEC 和 RMSEP 分别为 0.0078 和 0.0095, RPD 为 3.0293, 表明模型的预测效果较好。分析 SPA-SVR, CARS-SVR, IRIV-SVR, VCPA-SVR 和 VCPA-GA-SVR 五种预测模型可知, 其中 VCPA-GA-SVR 模型的预测效果最好, 其中 R_c 值和 R_p 值分别为 0.9785 和 0.9658, RMSEC 和 RMSEP 分别为 0.0067 和 0.0089, RPD 为 3.2293, 说明模型预测效果极好。其中 SVR 模型是在 Matlab 平台下使用 LIBSVM 工具箱创建的, 选取 RBF 作为核函数, 以不同特征波长选择方法所取得的光谱变量作为模型的输入, 以水分值作为模型的输出, 采用交叉验证函数 `tunelsssvm` 对惩罚因子 c 和核函数参数 g 进行参数优化, VCPA-GA-SVR 模型参数寻优结果如图 2-8 (b) 和预测效果图 2-8 (c) 所示。

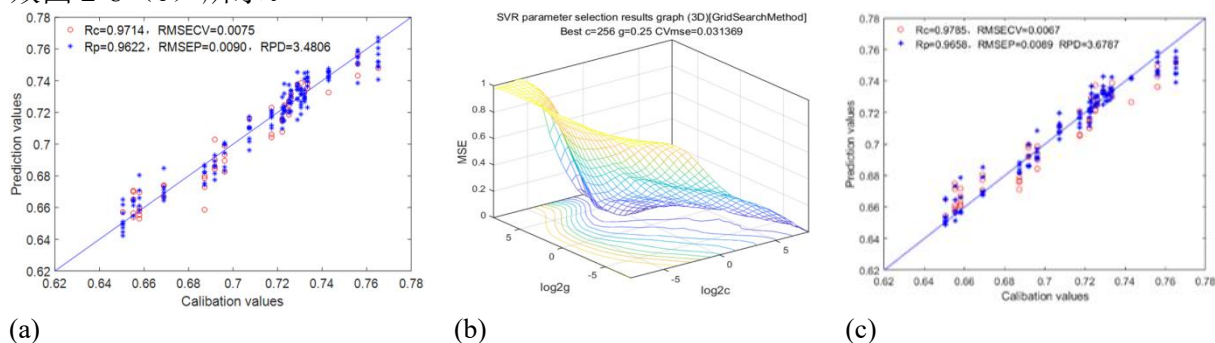


图 2-8 基于 PLSR 和 SVR 的模型优化: (a) PLSR 模型的校正值与预测值;

(b) SVR 模型参数优化 (c) SVR 模型的校正值与预测值

Fig.2-8 Model optimization based on PLSR and SVR: (a) Corrected values versus predicted values of PLSR models; (b) SVR model parameters optimization; (c) corrected values versus predicted values of SVR models

2.4 本章小结

本研究采用 NIR 技术对绿茶摊青过程中的含水率进行了定量检测, 基于 5 种预处理方法和 7 种特征波长筛选方法, 分别建立线性 PLSR 和非线性 SVR 定量预测模型, 为利用 NIR 技术实现绿茶摊青过程中含水率的定量预测提供了参考。结论如下:

(1) 为了减少光谱噪声, 分别采用了 SNV, Center, Max-Min, Detrend, SG FIR 等光谱预处理方法, 然后建立 PLSR 模型, 通过对比所获得的 PLSR 模型预测效果, 得出 SG FIR 为最优预处理方法, 其 R_c 值从 0.9366 提高到 0.9558, R_p 值从 0.9382 提高到 0.9507, RPD 值从 2.6511 提高到 3.2810, 表明光谱可以有效的提高模型的预测精度,

减少噪声信息。

(2) 为了进一步提高模型精度,减少检测周期,分别采用了 SPA、CARS、IRIV、VCPA、VCPA-GA 五种不同的特征波长筛选方法。结果表明:变量筛选方法不仅有效减少光谱中的冗余变量,还能进一步的模型的预测精度,其中建立的 PLSR 模型中 IRIV 为最优的特征波长筛选方法,建立的 SVR 模型中 VCPA-GA 为最优的特征波长筛选方法。IRIV-PLSR 和 VCPA-GA-PLSR 的 R_c 分别为 0.9714 和 0.9785, R_p 值分别为 0.9622 和 0.9658。

(3) 对比线性的 PLSR 模型和非线性的 SVR 模型的预测精度发现,非线性的 SVR 模型性能整体要优于线性的 PLSR 模型,其中所建立的 SG FIR+VCPA-GA-SVR 模型性能最优, R_p 为 0.9658, RPD 为 3.6787,表明模型具有很好的预测性能,近红外光谱技术可以实现绿茶摊青过程中含水率的快速检测。

第3章 基于多源信息融合的茶杀青叶水分含量预测模型建立

3.1 引言

杀青是绿茶加工过程中的关键工序，水分含量影响杀青叶物理状态和化学反应的进程，对成品茶叶的品质形成具有重要的作用。传统的茶叶水分检测方法为烘干称重法，该方法测量精度较高，但存在耗时和有损的缺点，无法满足茶叶加工过程中在线、快速和无损伤检测的需求。在绿茶生产实际中，多根据杀青叶色泽变化和手握的触感，判断杀青叶水分含量情况和杀青程度，如适度的杀青叶色泽暗绿，手捏叶质柔软，略有黏性。人工经验法属于专家行为，易导致在制品发生红边、焦叶或杀不透现象，造成品质不稳定和鲜叶原料的浪费。因此本研究以针芽形绿茶自动化生产线为载体，同步采集杀青过程中在制品的图像与光谱信息，提取图像纹理、颜色特征及特征光谱变量。探讨有效的数据融合方式以提高模型的预测精度，克服单一传感器预测精度不够的问题，为今后在标准化、数字化生产线中杀青模块的专家系统研发，提供理论依据和数据支撑。

3.2 材料和方法

3.2.1 样品制备

本实验于2020年4月7日在湖南省长沙市湘丰智能装备有限公司实施，茶叶品种为高桥银峰茶，嫩度为一芽一叶和一芽二叶。将采集好的鲜叶原料放置于鲜叶摊青机中，摊叶厚度为5-6 cm，保持室内空气温度在20℃左右，相对湿度在60%左右，在绿茶摊青机上采用自然摊青的方式摊放10 h左右，在水分含量接近65.05%处，摊青结束，随即进行杀青。杀青在6CS-60型滚筒式杀青机内进行，滚筒杀青温度为270+250+230℃（前段筒温+中段筒温+后段筒温），滚筒转速为22 r/min，杀青时间为6分钟，水分含量为54.24%。

3.2.2 仪器仪表和数据采集

如图3-1所示，本研究采用的检测系统主要由三部分组成：近红外光谱部分、机器

视觉部分、水分采集部分。在近红外光谱部分使用 IAS-3100 采集光谱数据，测样方式为下照漫反射；波长范围为 900-1700 nm；光谱精度为±1 nm，分辨率为 4 cm⁻¹；机器视觉部分包括图像传感器、专业工业相机、弧形光源、暗箱、计算机和 GUI 软件处理系统组成，专业工业相机型号为 FI-S200C-G，其中镜头为 4 nm 低畸变镜头，传感器为 1/2.8CMOS 图像传感器，分辨率为（1080*1080），曝光时间设为 0.09ms；光源采用 DOME 单色纯白弧形光源，强度为 100 勒克斯（lx），图片类型为 bmp 格式；相机镜头被固定在暗箱顶部，以确保拍摄角度和距离恒定。水分采集部分使用三台水分测定仪（MA35M-000230V1，Sartorius）对茶叶样品含水率进行采集，其量程为 35g，称重系统精度为 1mg。

为了实时检测杀青过程中水分的变化，采用每隔一分钟取一次样品，每次取 15 个样品，每个样品重约 80g，均匀平铺在规格为Φ70 mm×10 mm 的石英培养皿中，每个茶叶样品采集 3 次光谱数据，将 3 次光谱数据取平均值，作为该样品的光谱信息，共获得 105 条绿茶杀青过程中的光谱信息。将每个采集完光谱数据的样品放置于相机的视野中，采集样品图像信息，该研究共获得 105 张图像用于数据的分析和模型的建立。将采集完光谱和图像信息的样品测量水分含量，水分值如表 3-1 所示。

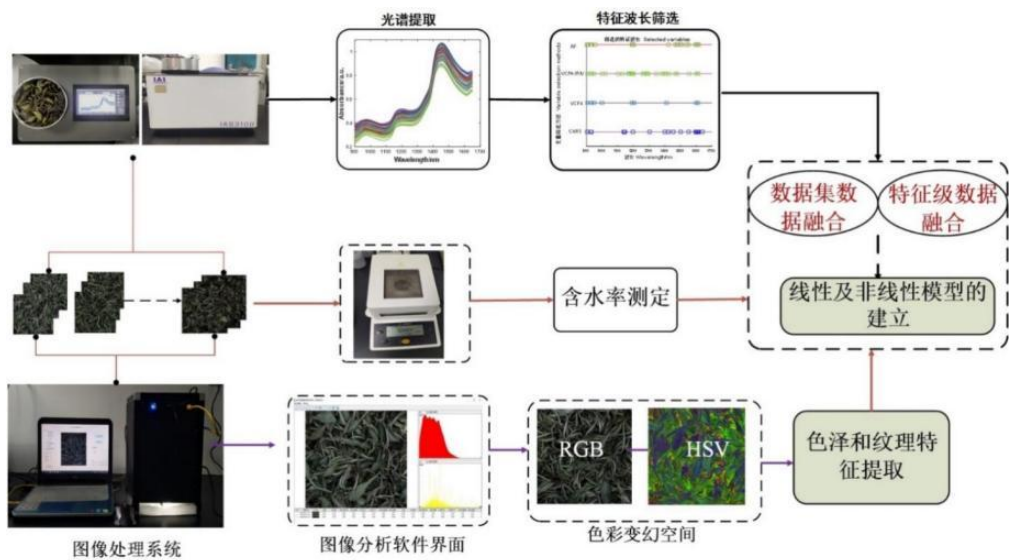


图 3-1 数据融合流程图

Fig.3-1 Flowchart of the data fusion used in this study

表 3-1 绿茶杀青样本水分分布情况

Table 3-1 Information of samples in the green tea fixation process

工序	含水率范围/%	含水率平均值/%	含水率方差	样品数目
杀青 1	65.01~64.07	64.84	0.07978	15
杀青 2	63.21~62.06	62.41	0.14011	15

杀青 3	61.85~60.98	61.35	0.14881	15
杀青 4	60.73~59.13	60.23	0.30934	15
杀青 5	58.96~58.41	58.75	0.04916	15
杀青 6	56.94~55.73	56.24	0.16356	15
杀青 7	54.82~53.46	54.24	0.22599	15

3.3 特征提取和数据融合

3.3.1 近红外光谱处理

在杀青叶光谱采集的过程中，由于培养皿中样品不平整，光谱信息容易受到散光、噪声等因素干扰，故使用多元散射校正（Multiplicative scatter correction, MSC）、均值中心化（Mean Centering, MC）、标准差归一化（Z-Score）、极差归一化（R-score）和一阶微分（first derivative, FD）五种方式对原始光谱进行预处理，以减少光谱中的噪声信息^[30, 42]。为了消除与光谱信息无关的冗余信息，使用竞争性自适应加权算法（competitive adaptive reweighted sampling, CARS）、变量组合集群分析法（variables combination population analysis, VCPA）、变量组合集群分析法结合迭代保留信息变量法（variables combination population analysis and iterative retained information variable algorithm, VCPA-IRIV）和随机蛙跳法（random frog, RF）四种变量筛选算法从全光谱中提取特征波段。由于第二章中已对 MC 预处理做过介绍，本章主要阐述 MSC、Z-Score、R-score 和 Derive 预处理方法。

MSC 是目前最为常用的一种光谱预处理方法，该方法用于多波长校准和建模的化学计量学方法，可以有效的减少基线漂移和光散射所产生的影响，并增强了组分含量相关的光谱吸收信息。以下为具体的算法过程：

$$\text{计算平均光谱: } \bar{x} = \frac{\sum_{i=1}^n x_{ij}}{n} \quad (3-1)$$

$$\text{一元线性回归: } x_i = k_i \bar{x} + b_i \quad (3-2)$$

$$\text{多元散射校正: } x_{i(MSC)} = \frac{(x_i - b_i)}{k_i} \quad (3-3)$$

Z-Score 也称为零-均值规范化，经过处理的光谱数据均值为零，标准差为 1。具体转换公式如下所示：

$$x^* = \frac{x - \bar{x}}{\sigma} \quad (3-4)$$

其中, $\bar{x}$ 为原始光谱数据的均值, σ 为原始光谱数据的标准差。

R-Score 处理后的光谱数据在 (0,1) 之间, 数据分布更加的均衡, 其目的一方面是为了降低同一茶叶样品多次采集的光谱数据之间的差异性, 减小因不同茶叶样品之间差异引起的光散射与微小光程差变化产生的影响, 为后续建模分析提供可靠的光谱数据; 另一方面是为了消除光谱内的冗余信息, 加快模型的收敛速度, 从而提高模型的稳健性和预测能力^[43]。极差归一化计算公式如下:

$$\dot{x}_{ij} = \frac{x_{ij} - \min x_{ij}}{\max x_{ij} - \min x_{ij}} \quad (3-5)$$

其中, $i = 1, 2 \dots n$, n 为校正集样品数; $j = 1, 2 \dots m$, m 为波长点数。

一阶微分 (FD) 可以提高近红外光谱与茶叶水分含量值之间的相关性, 从而快速确定光谱的拐点和最大值、最小值反射率所在的位置。其公式如下:

$$\rho'(\lambda_i) = \frac{\rho(\lambda_{i+1}) - \rho(\lambda_{i-1})}{2\Delta\lambda} \quad (3-6)$$

式中, $\rho'(\lambda_i)$ 是在光谱波段 λ_i 处光谱的一阶导数值, λ_i 是光谱波段 i 的波长数值, $\Delta\lambda$ 是波长 λ_i 到 λ_{i+1} 的差值。

3.3.2 图像特征

软件系统部分以茶叶图像中像素点 (1728,1152) 为中心, 采用自动的方式分割出 1000×1000 像素区域, 并提取该区域内的色泽和纹理特征^[44]。通过 RGB 和 HSV 之间的颜色模型转换, 最终获得了样品图样的 9 个颜色特征以及 6 个纹理特征。9 个颜色特征分别为红色通道均值 (R)、绿茶通道均值 (G)、蓝色通道均值 (B)、色点均值 (H)、饱和度均值 (S)、亮度均值 (V)、超绿变换 ($2G-R-B$)、红色通道均值与绿色通道均值的比值 (R/G) 和颜色角 (hab^*); 6 个纹理特征分别为: 平均灰度值 (m)、标准差 (δ)、平滑度 (r)、三阶距 (μ)、一致性 (U)、熵 (e), 共计得到 15 个图像特征^[45]。

3.3.3 数据融合

数据融合可以分为三个不同的层次：数据级、特征级和决策级^[46]。在数据级融合过程中，直接将近红外光谱和机器视觉系统所采集的数据融合起来为数据级融合；特征级数据融合是将来自不同传感器之间的特征变量进行数据融合；在决策级融合过程中，从每个数据源中计算出单独的回归模型，并将其结果合并，得到最终的回归模型。然而，决策级数据融合需要准确的数据处理方式，如果没有考虑不同传感器之间的数据相关性，此过程可能会导致大量信息丢失。因此，本研究选择数据级和特征级融合进行数据处理。为了消除数据的大小对建模结果的影响，采用数据归一化对融合的数据进行预处理。

3.4 多变量数据分析方法

3.4.1 主成分分析

主成分分析（Principal component analysis, PCA）作为特征提取的一种方式，在保证光谱和图像主要特征不丢失的情况下，降低数据集的维数，将多特征转化为少量主要特征的方法。主成分分析通过将原始变量转换为一组成分（PCs）的线性不相关的新变量来汇总数据,累积贡献率表示相应的主成解释原始变量的能力^[47]，本研究中第 10 个 PCs 的 RMSECV 值最低，且前 10 个 PCs 的累计贡献率达到了 99%以上，因此本研究采用前十个 PCs 建立定量预测模型。

3.4.2 定量预测建模方法

在建立模型之前，使用 Kennard-Stone（K-S）算法对光谱数据集进行校正集和预测集的划分，其中校正集样本数为 79 个，预测集样本数为 26 个，校正集与预测集比例为 3:1^[48]。然后采用偏最小二乘回归（PLSR）和支持向量回归（SVR）建立绿茶杀青过程中水分含量定量预测模型。

3.4.3 模型评价指标

本研究所采用的模型评价指标与评价标准与第二章保持一致，及采用校正集相关系数（ R_c ）、交互验证均方根误差（RMSECV）和预测集的相关系数（ R_p ）、预测

均方根误差 (RMSEP) 分别作为预测模型的评价指标。本研究所有的数据处理分析均在 Origin2018 和 Matlab 2019a 中完成。

3.5 结果与分析

3.5.1 光谱特征和预处理优选

图 3-2 (a) 为绿茶杀青试验样本的近红外光谱图, 记录了 900-1650nm 波长样本的光谱曲线, 剔除噪声影响较大的 1651-1700nm 范围内的光谱数据, 不同杀青时间样本的光谱曲线存在细微差距。发现主要吸收峰位于 960nm、1225nm 和 1440nm 处。其中, 960nm 处的吸收峰主要与水分子中 O-H 基团的二级倍频有关; 1220nm 处的吸收峰受水分子 O-H 基团的合频吸收带以及 C-H 键二级倍频的影响, 其中 C-H 的影响来自杀青过程中, 在制品发生酶钝反应, 引起内质成分变化的影响; 1440nm 处的吸收峰受水分子中 OH 基团在近红外区的伸缩振动一级倍频的影响^[49]。

由于采集的近红外原始光谱中包含了除光谱信息之外的其他噪声信息, 因此使用 5 种不同的光谱预处理方法校正光谱信息, 校正后的光谱如图 2(b)-2(f) 所示。基于原始光谱和预处理后的光谱数据建立 PLSR 预测模型, 并进行模型比较, 以确定最优的预处理方法, 结果如表 3-2 所示。由表 3-2 可知, 基于原始光谱所建立的预测模型效果较差, $RPD < 1.8$, 只可以对绿茶杀青过程中水分含量实现粗略的预测; 使用预处理后所建立的预测模型性能相比原始光谱所建立的模型效果明显较优, 其中 MC 为最佳的预处理方法, 其中 R_c 值为 0.9385, R_p 值为 0.9071, RPD 为 2.1804, 可以对绿茶杀青过程中水分含量实现较为准确的预测。

表 3-2 基于不同预处理方法水分含量预测模型

Table 3-2 Result of moisture content prediction model for different pretreatment methods

预处理方法	主成分数	校正集		预测集		RPD
		R_c	RMSECV	R_p	RMSEP	
原始光谱	10	0.9286	0.0106	0.8699	0.0133	1.5591
MSC	9	0.94249	0.0116	0.8864	0.0143	2.0678
MC	8	0.9385	0.0119	0.9071	0.0143	2.1804
Z-score	8	0.9447	0.0111	0.8920	0.0141	2.1128

R-score	9	0.9428	0.0112	0.8740	0.0141	1.8187
FD	8	0.88947	0.0149	0.8748	0.0123	1.8252

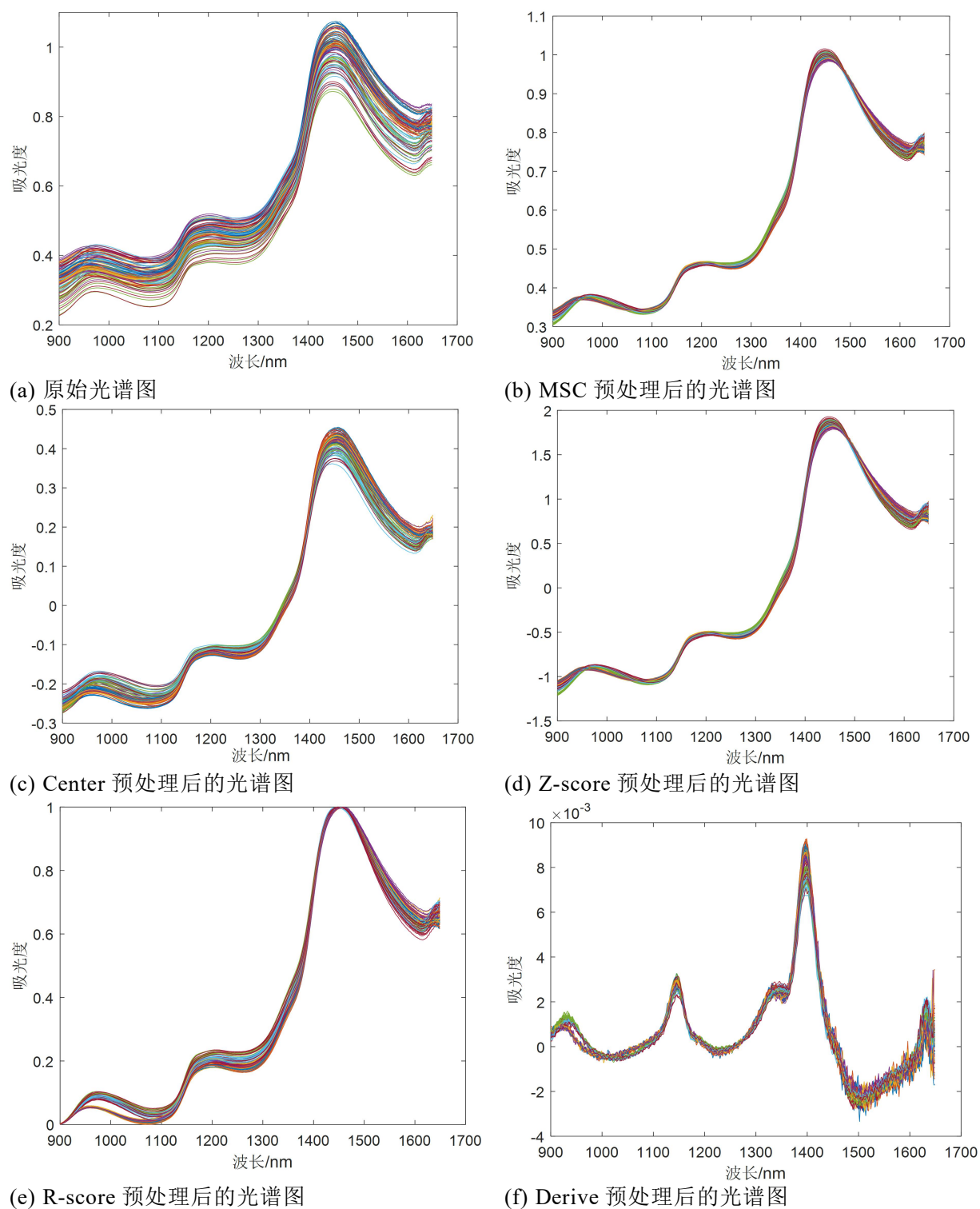


图 3-2 不同预处理方法对光谱的影响

Fig.3-2 NIR and various preprocessed spectrum

3.5.2 颜色特征变化

为了更好的研究杀青过程中叶片颜色变化规律，本研究首先对整体视觉变化进行分析。随机拍摄不同杀青时间的图像，根据杀青时间顺序排列。然后提取图像的平均颜色，结果如图 3-3(A)和 3-3(B)所示，杀青时间不同，颜色也存在一定的差异。为了更好的而区分不同加工时间图像之间的差异性，本研究将 RGB 图像转化为 HSV 颜色模型然后取平均色，结果如图 3-3(C)和 3-3(D)所示。从图中可以看出，在绿茶杀青过程中，叶子变得柔软卷曲，颜色逐渐由青绿色变成深绿色。随后，提取图像的不同颜色特征参数，从微观角度呈现叶子颜色的变化规律，结果如图 3-4(a)~3-4(o)所示。其中 RGB 值随着绿茶杀青过程而显著降低，导致颜色从青绿色变成暗绿色，这是因为在高温钝化了酶的活性，终止了各种酶促变化途径，同时叶绿素总量、叶绿素 a 与叶绿素 b 的比例发生了变化，鲜叶中墨绿色的叶绿素 a 与黄绿色的叶绿素 b 的比例大约是 2:1，通过杀青工艺，叶绿素 a 的破坏较多，剩下 25%左右，叶绿素 b 的破坏较少，在杀青叶中约 50%~60%，由此可见，杀青之前墨绿色的叶绿素 a 比黄绿色的叶绿素 b 大约多一倍，杀青之后，则叶绿素 b 比叶绿素 a 多一倍，使得叶色由鲜绿色变为暗绿^[50-52]。对于 HSV 颜色空间，在杀青过程中 H 和 V 值在杀青过程中逐渐降低，S 值在杀青过程中逐渐增大，这可以通过图 3-3 中描绘的颜色得以证明。其中，纹理变化过程中除了三阶矩值和一致性值逐渐增大之外，其他纹理特征均呈现逐渐降低的趋势。

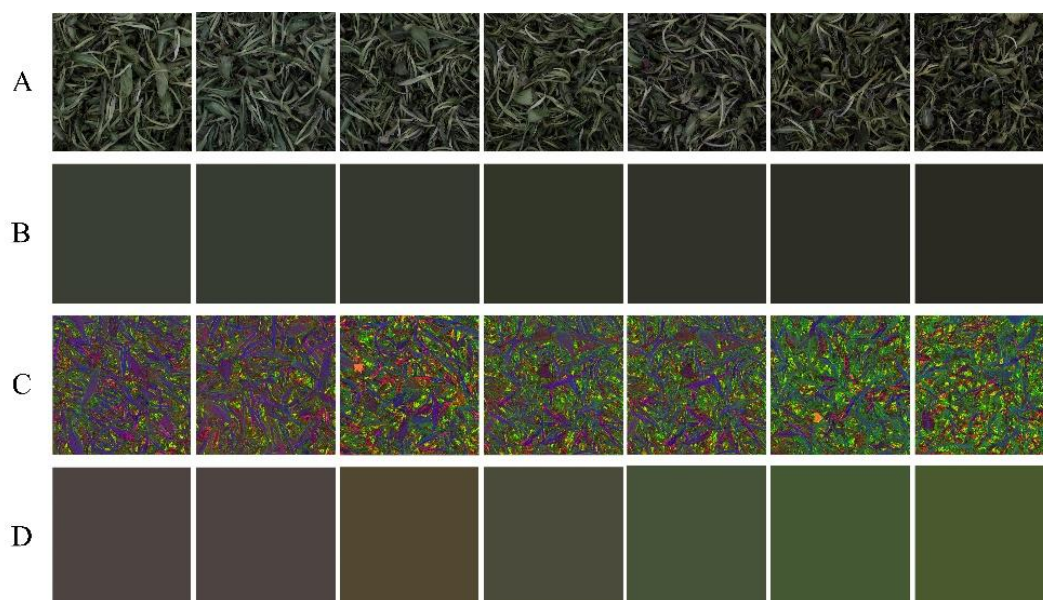


图 3-3 杀青过程中颜色变化图

Fig.3-3 Color change diagram during the fixation process

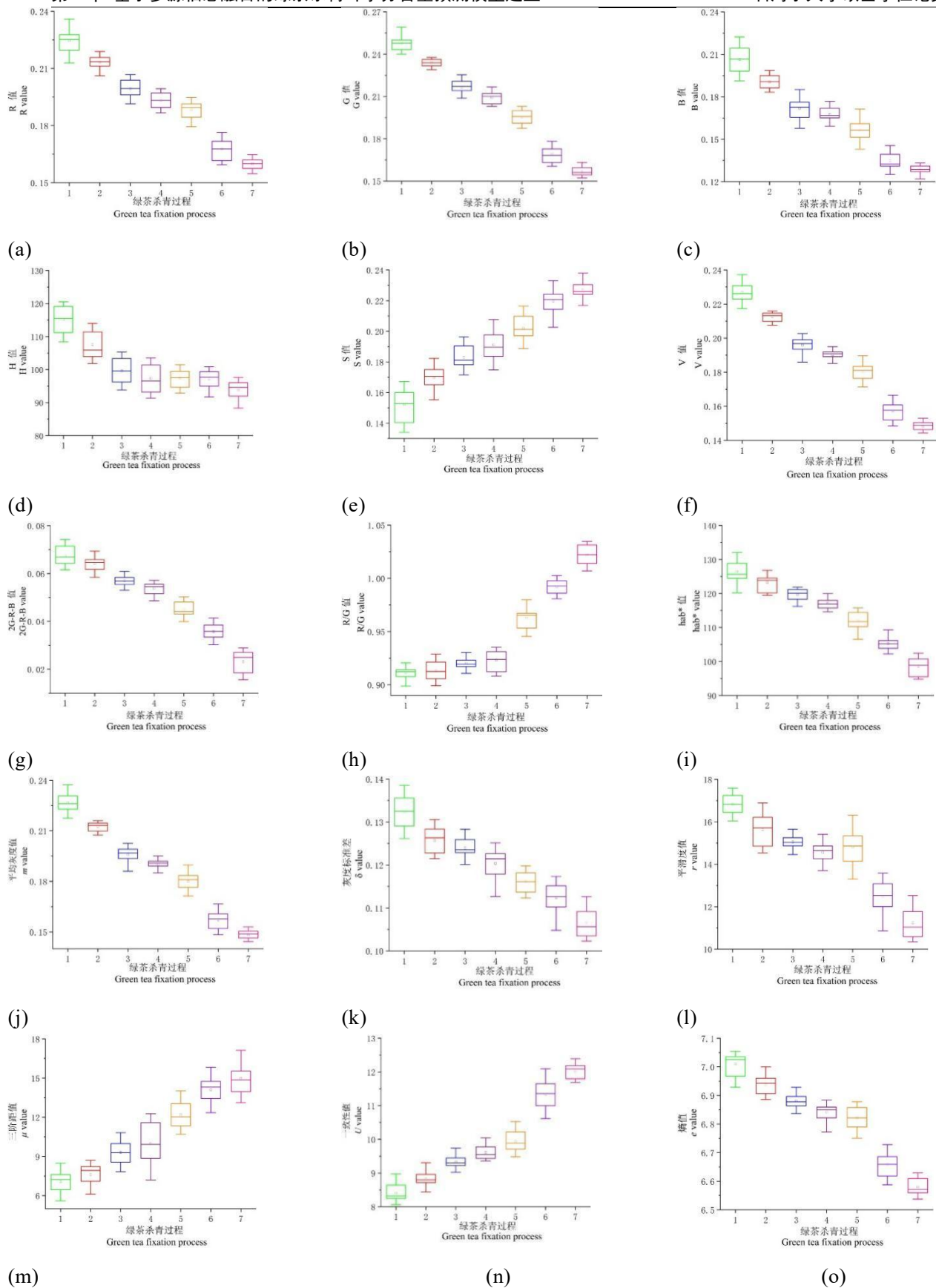
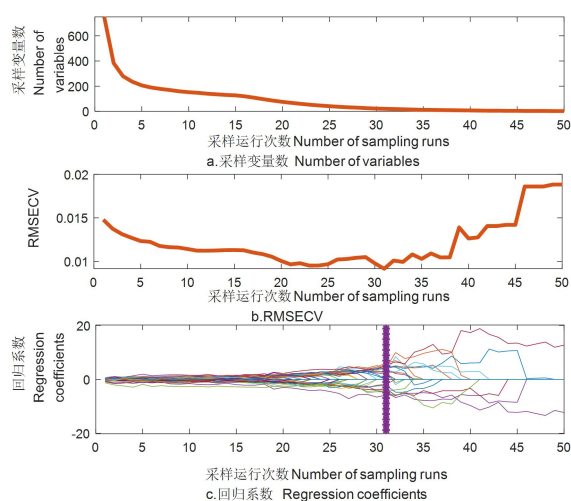


图 3-4 杀青过程中色泽和纹理特征值的变化结果

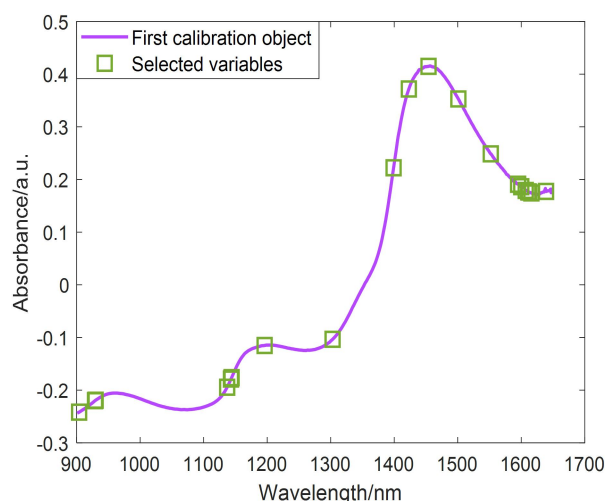
Fig.3-4 The result of the change of color and texture characteristic value during the curing process

3.5.3 光谱特征波长筛选

为了进一步提高模型的预测精度,将采集的近红外光谱数据使用最佳预处理后,分别使用 CARS、VCPA、VCPA-IRIV 和 RF 四种特征波长筛选方法提取特征波长。其中,采用 CARS 算法时迭代次数与蒙特卡罗采样运行次数分别设为 1000 和 50^[52],因此 CARS 算法共筛选出 20 个特征波长,筛选结果如图 3-5 (a)、3-5 (b) 所示;运行 VCPA 算法时,最佳子集占比为 0.1,二进制矩阵采样 (binary matrix sampling, BMS) 运行次数设为 1000,指数衰减函数 (EDF) 运行的次数设为 50,剩余变量数目设为 14^[53],因此 VCPA 算法筛选出 9 个特征波长,筛选结果如图 3-5 (c) 所示;VCPA-IRIV 算法首先利用 VCPA 算法对变量空间快速缩小,并将最终剩余变量数设为 N;然后再通过 IRIV 算法对剩余变量空间中每个变量进行重要性评估,从而选取出最佳变量子集^[38],本研究最终优选出 33 特征波长,筛选结果如图 3-5 (d) 所示;RF 算法是基于可逆跳跃马尔可夫链蒙特卡罗 (reversible jump Markov Chain Monte Carlo, RJMCMC) 的框架进行工作^[54],采用 RF 算法进行特征波长筛选时,为了减少算法产生的不稳定性,应将增大蒙特卡罗采样次数,本研究设置蒙特卡罗采用次数为 10000 次,共筛选出 25 个特征波长,筛选结果如图 3-5 (e) 所示。图 3-5 (f) 为四种变量筛选方法提取的特征波长分布图,可以看出特征波长在 960nm、1220nm、1440nm 和 1620nm 附近,这可能与 OH 基团、NH 基团和 CH 基团的伸缩振动有关^[49]。



(a)



(b)

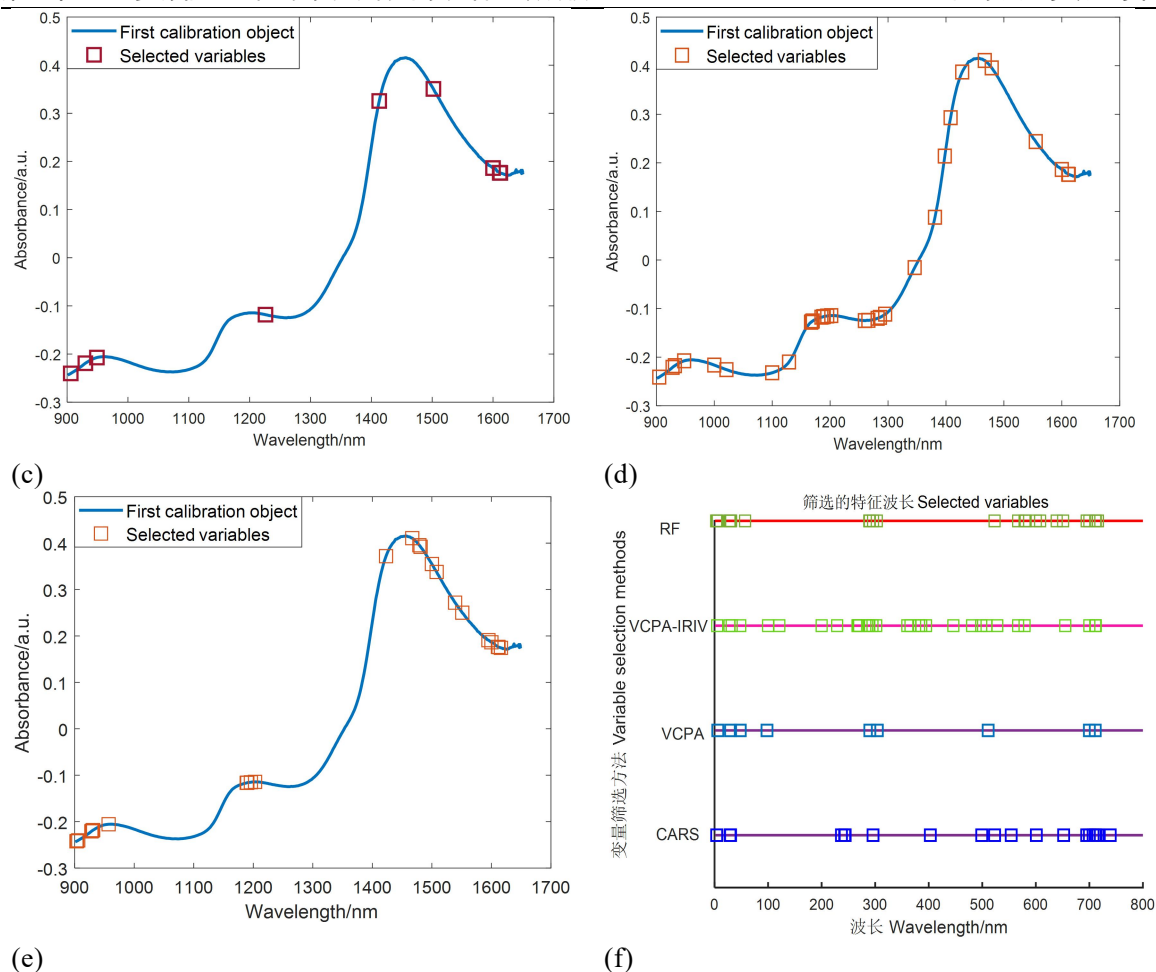


图 3-5 通过不同的变量筛选方法选择的特征波长

Fig.3-5 Characteristic wavelength variables selected by different variable screening methods

3.5.4 基于单一传感器的预测模型

为获取较优模型以实现绿茶杀青过程中水分含量的定量预测，基于最优预处理方法进行特征波长的筛选，然后结合 PCA 建立 SVR 非线性预测模型，并与线性的 PLSR 模型相比较。表 3 列出了基于近红外光谱数据和颜色特征数据预测绿茶杀青过程中水分含量的模型性能指标。

对比线性的 PLSR 模型结果可知，全光谱-PLSR 模型的 R_c 值为 0.9385， R_p 值为 0.9071，RPD 值为 2.1804，表明模型的预测效果较好。基于筛选出的特征波长所建立的模型效果明显基于全波段所建立的模型性效果，因为在筛选特征波长过程中剔除了很多冗余和无关的信息变量，其中基于 VCPA-IRIV 算法所建立的模型效果最优， R_c 值为 0.9550，RMSECV 值为 0.0102， R_p 值为 0.9422，RMSEP 值为 0.0103，RPD 值为 2.7956，并且变量从 750 个减少至 33 个，变量压缩率为 95.6%，既提高了模型的预测精度，也降低了模型的运算时间。基于色泽和纹理所建立的模型效果要

低于基于近红外光谱数据所建立的模型效果，其中 R_c 值为 0.9326， R_p 值为 0.9018，RPD 值为 2.1365。

对比非线性的 SVR 模型结果可知，全光谱-SVR 校正集 R_c 值为 0.9423， R_p 值为 0.9212，RPD 值为 2.3535，模型的预测性能较全光谱-PLSR 模型性能有所提高。分析比较 CARS-SVR、VCPA-SVR、VCPA-IRIV-SVR、RF-SVR 四种模型结果可知，基于 VCPA-IRIV 算法所建立的模型效果做好，其 R_c 值为 0.9727，RMSECV 值为 0.0087， R_p 值为 0.9545，RMSEP 值为 0.0098，RPD 值为 2.9506。基于颜特征所建立的 SVR 模型效果显著优于所建立的 PLSR 模型效果，但是相比基于光谱数据所建立的模型效果较差。其中 SVR 算法是基于 Matlab 平台下的 Li-SVMLAB 工具箱进行创建的，该工具箱选用 RBF 作为核函数，以 PCA 后的前 10 个 PCs 作为模型的输入，杀青叶样品水分含量作为输出，从而建立绿茶杀青叶水分含量预测模型。因惩罚因子 (c) 和 RBF 参数 (g) 对模型效果具有较大影响，因此采用交叉验证函数 `tunelsssvm` 进行参数优化^[30]。

基于近红外光谱数据所建立的模型效果相比图像数据实现了更好的预测精度，具有更高的 R_c 值、 R_p 值和 RPD 值，以及更低的 RMSECV 和 RMSEP 值。导致这一结果的原因可能是机器视觉可以直接捕获颜色相关成分信息，但是这是一种间接的测量，有些成分并不能完全捕获，相比之下，近红外光谱对 OH 基团、NH 基团和 CH 基团具有很好的响应，可以很好的捕获绿茶加工过程中的成分变化^[55]。

表 3-3 基于光谱和图像信息的模型性能指标

Table 3-3 Model performance based on spectrum and image information

数据	模型	筛选方法	变量数	PCs	建模集		预测集		RPD
					R_c	RMSECV	R_p	RMSEP	
	PLSR	全光谱	750	10	0.9385	0.0119	0.9071	0.0143	2.1804
		CARS	20	8	0.9486	0.0109	0.9315	0.0111	2.5376
		VCPA	11	10	0.9405	0.0121	0.9024	0.0139	2.2406
		VCPA-IRIV	33	8	0.9550	0.0102	0.9422	0.0103	2.7956
		RF	25	9	0.9653	0.0101	0.9208	0.0112	2.4035
光谱数据		全光谱	750	9	0.9423	0.01002	0.9212	0.0132	2.3535
		CARS	20	7	0.9523	0.0096	0.9437	0.0103	2.7428

		VCPA	11	7	0.9446	0.0115	0.9401	0.0107	2.6629
		VCPA-IRIV	33	6	0.9727	0.0087	0.9545	0.0098	2.9506
		RF	25	8	0.9770	0.0074	0.9382	0.0107	2.3863
颜色 特征	PLSR			8	0.9326	0.0126	0.9018	0.0147	2.1365
	SVR	Normalize +PCA	15	7	0.9593	0.0069	0.9374	0.0106	2.4902

3.5.5 基于数据融合的预测模型

表 3-4 显示了使用融合光谱和颜色特征数据进行绿茶杀青过程中水分含量预测的模型效果。基于数据级数据融合所建立的 PLSR 和 SVR 模型获得了较好的模型效果，其中所建立的 SVR 模型精度较高，其 Rc 值为 0.9550，Rp 值为 0.9026，RPD 值为 1.8573。然而这种预测效果是不可采用的，因为使用单个传感器获得的模型取得了更高的预测精度。与基于低级数据融合所建立的 PLSR 和 SVR 模型效果相比，基于光谱数据的所建立的模型具有更好的 Rc 值、Rp 值以及 RPD 值。这些结果表明，不合理的数据融合并不能有效的提高预测精度。相比之下，基于特征级数据融合所建立的 PLSR 和 SVR 模型获得了更好的预测精度，具有更高的 Rc、Rp 和 RPD 值以及更低的 RMSECV 和 RMSEP 值。其中基于 CARS 算法提取的特征波长融合 15 个颜色特征所建立的 SVR 模型效果最优，其 Rc 值为 0.9742，Rp 值为 0.9719，RMSEP 值为 0.0079，RPD 值为 4.1546，预测值和实测值的关系如图 3-6 所示。其中根据水分测定仪测得杀青叶含水率方差在 1%的量级，模型所获得的 RMSEP 分布在 0.7%~1.2%，而 RMSEP 值本身就是指误差值，因此模型预测结果满足水分测定仪测定结果。总体而言，通过近红外光谱和机器视觉的有效结合，实现了对绿茶加工过程中水分含量的准确预测，为低成本、快速、有效的绿茶质量评估提供了手段。

表 3-4 基于数据融合的模型性能指标

Table 3-4 Model performance indicators based on data fusion

模型	数据融合	变量筛选方法	PCs	建模集		预测集		
				Rc	RMSECV	Rp	RMSEP	RPD
PLSR	数据级融合	Normalize + PCA	10	0.9483	0.0110	0.8732	0.0119	1.6598

	CARS	9	0.9715	0.0081	0.9634	0.0084	3.8056
	VCPA	9	0.9709	0.0082	0.9636	0.0084	3.8139
特征级融合	VCPA-IRIV	10	0.9716	0.0081	0.9583	0.0090	3.5492
	RF	9	0.9656	0.0083	0.9568	0.0100	3.4798
数据级融合	Normalize + PCA	9	0.9550	0.0104	0.9026	0.0112	1.8573
	CARS	7	0.9742	0.0075	0.9719	0.0079	4.1546
SVR	VCPA	7	0.9734	0.0080	0.9707	0.0081	4.0237
特征级融合	VCPA-IRIV	8	0.9773	0.0073	0.9661	0.0083	3.8210
	RF	8	0.9725	0.0081	0.9660	0.0079	3.6713

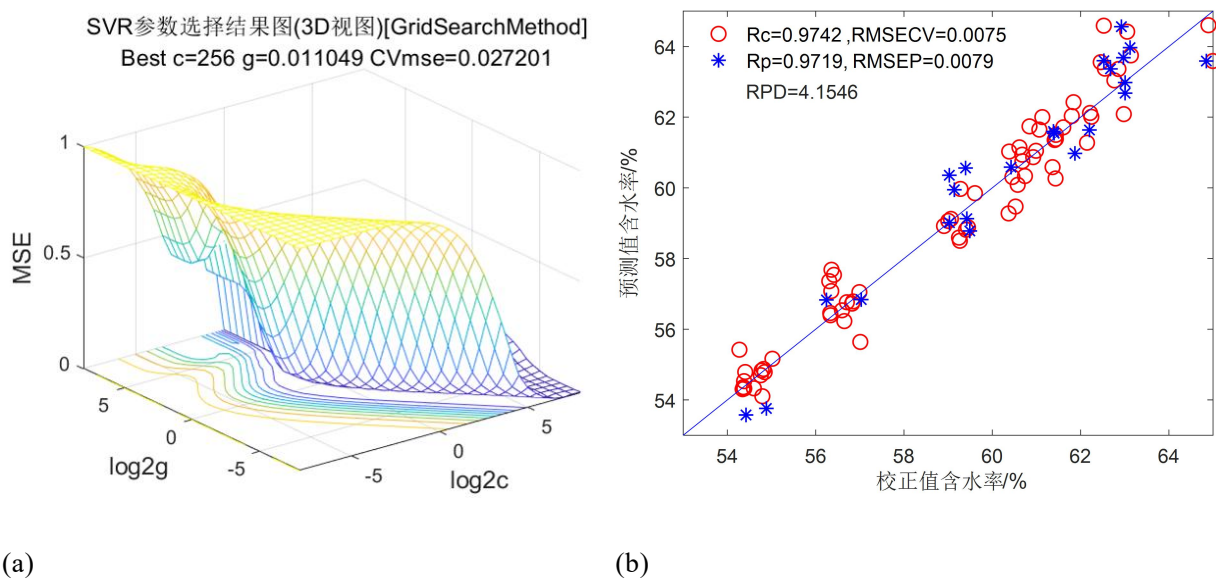


图 3-6 基于 SVR 的模型优化

Fig.3-6 Model optimization based on SVR

3.6 本章小结

本研究证明了融合近红外光谱和的机器视觉技术来预测绿茶杀青过程中水分含

量的可行性，通过采用数据级和特征级两种数据融合的方式，并建立线性 PLSR 和非线性 SVR 定量预测模型进行对比。结果表明，基于单一传感器数据所建立的模型效果要优于数据级数据融合所建立的模型效果，表明无效的数据融合并不能提高模型的预测效果。相比之下，基于特征级数据融合所建立的模型效果明显优于单一传感器和低级数据融合所建立的模型效果。其中基于 CARS 算法提取光谱特征波长融合图像的 15 个颜色特征，然后采用 Normalize 和 PCA 分析所建立的 SVR 模型效果最优，其中 R_c 值为 0.9742， R_p 值为 0.9719，RPD 值为 4.1546，且 RMSEC 和 RMSEP 差值仅为 0.004。因此，本研究证明了融合光谱和成像系统对绿茶加工过程中水分含量的预测能力，克服了单一传感器预测能力不足的问题。

第4章 基于近红外光谱的绿茶干燥程度判别模型建立

4.1 引言

干燥作为绿茶加工过程中的最后一道工序，其主要目的是进一步去除茶叶中的水分，使成品茶的含水率达到标准要求。成品绿茶水分含量越高，在储存的过程中，越容易发生茶叶品质的变化，当成品绿茶水分含量超过12%时，即使在密封条件下茶叶内部的各种化学也可以继续进行，但是茶叶水分过低，容易造成成品茶过于干燥易碎。因此在绿茶生产过程中，含水量要控制在7%以下，而经过精制后的成品茶叶含水率要控制在4%-7%之间。因此对绿茶干燥程度的适度性判别对茶叶品质的提高尤为重要。

4.2 材料和方法

4.2.1 试验材料

实验材料选用“高桥银峰茶”，嫩度为一芽一叶和一芽二叶。采摘地点为湖南省长沙市高桥镇采摘时间为2020年4月7日。

4.2.2 试验设备及工艺流程

1、摊青、杀青设备均采用第二章和第三章所述的摊青机和杀青机设备。经摊青后的茶叶水分含量为65.05%，经杀青过后的茶叶水分为54.24%。

2、揉捻设备及加工工艺

将杀青完成后的茶叶摊凉1.5小时后放入6CR-55型（长沙湘丰智能装备有限公司）揉捻机内，由制茶工艺大师通过调节揉捻机旋转轮来控制茶叶在揉捻过程中所受的的压力大小，揉捻按照空揉10min→轻揉10min→中揉10min→轻揉5min，总计35min。揉捻结束后，叶片揉破变轻，卷转成条，体积缩小，同时部分茶汁溢附在表面，茶叶成条率在80%以上，揉捻后茶叶含水率为50.15%。

3、干燥设备及加工工艺

干燥实验在6CHZ-9型茶叶烘焙机内进行，由于揉捻过后的茶叶水分含量过高，因此干燥过程采用初烘、复烘和足烘的加工工艺。将揉捻过后的茶叶样品放入至

6CHZ-9 型烘焙机进行初烘，烘干温度设置为 100℃，时间为 45min，水分含量为 12.87%；初烘结束后将茶叶样品摊凉 1h 后进行复烘，复烘温度为 90℃，烘干时间为 60min，水分含量降至 6.02%；复烘结束后摊凉 60min 进行足烘，足烘温度设置为 80℃，烘干时间为 60min，水分含量降至 4%以下。

4.2.3 光谱采集

本研究采用 IAS-3100 近红外光谱仪器对绿茶干燥过程中水分含量进行光谱信息采集，水分检测方式与第二章和第三章保持一致。在近红外光谱采集时，初烘水分含量降至 12.87%的过程中，每隔 15 分钟取一次茶叶样品，每次取 10 组进行光谱信息采集；复烘水分含量降至 6.02%的过程中，每隔 10min 去一次样品，每次取 10 组进行光谱信息采集；足烘水分含量降至 4%的以下的过程中，每隔 10 分钟取一次样品，每次取 10 组进行光谱信息采集，因此共获得 120 条光谱数据，其中水分含量大于 7%的光谱数据值为 60 条，水分含量在 4%-7%之间的光谱数据为 30 条，水分含量低于 4%的光谱数据为 30 条。其中，水分含量大于 7%的光谱数据为干燥不足数据，在 4%-7%之间的为干燥适度数据，低于 4%的为干燥过度数据。

4.2.4 数据分析方法

1、光谱数据预处理与主成分分析

在近红外光谱数据采集的过程中，由于石英培养皿中茶叶样品表面不平整、仪器的基线漂移、光散射等影响使得原始光谱中包含无用的信息导致建模精度降低。为了提高模型的精度，降低干扰信息的影响，本章节采用标准正态变换（SNV）、最大最小归一化（Max-Min）、均值中心化（MC）和多元散射校正（MSC）法对原始光谱进行预处理。由于提取的近红外光谱波段较多，可能存在大量的冗余信息，因此需要对光谱数据进行特征提取，以达到简化模型，提高模型精度的目的。主成分分析法（PCA）是一种常用的简化数据集的技术，可以在保证光谱主要特征不丢失的情况下，降低数据集的维数，将多特征转化为少量主要特征的方法。

2、特征波长选择方法

试验中，采集的每条近红外光谱数据有 750 个波段，导致整体的数据矩阵非常大，存在大量的冗余信息，降低了模型的精度。因此为了减少原始光谱数据中的数据冗余，剔除无关信息，提高模型预测精度和运算速度，本研究采用随机蛙跳算法（RF），竞争性自适应加权法（CARS），迭代保留信息变量（IRIV）和结合迭代保留信息变量的

变量组合集群分析法（VCPA-IRIV）4种算法提取特征波长，并选出最优的特征变量筛选方法。

3、模型建立和性能评估

本研究利用 Kennard-Stone（K-S）算法对 126 个红茶样本数据集进行划分，其中校正集 84 个，预测集 42，校正集和预测集比例为 2:1。分别采用偏最小判别分析法（Partial least squares discrimination analysis, PLS-DA）、k-邻近（K-Nearest Neighbor, KNN）分类算法和极限学习机（Extreme Learning Machines）算法建立绿茶不同干燥程度判别模型。其中校正集识别率（Calibration set Identification rate）和预测集识别率（Prediction set Identification rate）为模型的评价指标，校正集识别率和预测集识别率越高，则模型的效果越好。

PLS-DA 是一种基于 PLS 回归的线性判别分析方法。PLS-DA 算法在特征提取过程中获取样本的分类信息，并在自变量和分类变量之间建立回归模型。然后有效地提取与分类相关的特征变量，实现对数据的分类和识别^[56]。

KNN 是一种以模式识别的基本假设-同类样本在模式空间相互靠近-为依据的分类算法，是一种基于记忆的分类器，不需要模型拟合，其核心思想是根据不同特征值之间的距离进行分类。它根据最近的一个或多个样本的类别来确定整个样本所属的类别^[57]。KNN 是一种理论上成熟的算法，也是机器学习中最简单的算法之一。

ELM 是一种单隐含层前馈型神经网络学习算法，具有高效处理非线性数据回归问题等优点，克服了一般神经网络学习速度慢，易产生过拟合的问题^[58]。

4.3 数据分析与处理

4.3.1 光谱特征与预处理方式

本实验采集的是 900-1700nm 范围内 800 个波段的近红外光谱数据，由于光谱后端有明显的噪声，因此去除后端噪声明显的部分，采用 900-1650nm 之间的光谱进行分析，原始光谱如图 4-1（a）所示，从图中可以看出三种不同干燥程度的绿茶样本光谱曲线有较为明显的差异，3 组不同干燥程度的绿茶样本分别赋值 1、2、3。将采集的不同干燥程度的茶叶光谱数据分别进行平均处理，图 4-1（b）为获取的一组（干燥不足、干燥适度、干燥过度）平均光谱曲线。通过平均光谱曲线可以看出，当茶叶干燥不足水分含量过高时，光谱图像存在明显的吸收峰和反射谷，在 980nm、1095nm、1200nm 和 1450nm 处吸收峰和反射谷较为明显，这与水分子中 O-H 基团的伸缩振动有关；当水分含量低于 7%时，可以看出在 900-1150nm 处的光谱曲线逐

渐平缓，吸收峰和反射谷逐渐消失，这主要受水分子中 O-H 基团逐渐消失的影响。

由于原始光谱中包含了除样品本身之外的其它噪声信息，为了有效抑制和消除光谱数据中的噪声，提高所建立的模型性能，本研究采用 SNV、Max-Min、MC 和 MSC 四种预处理方法减少原始光谱中噪声的影响，预处理后的光谱图如图 4-1 (c)，图 4-1 (d)，图 4-1 (e)，图 4-1 (f) 所示。

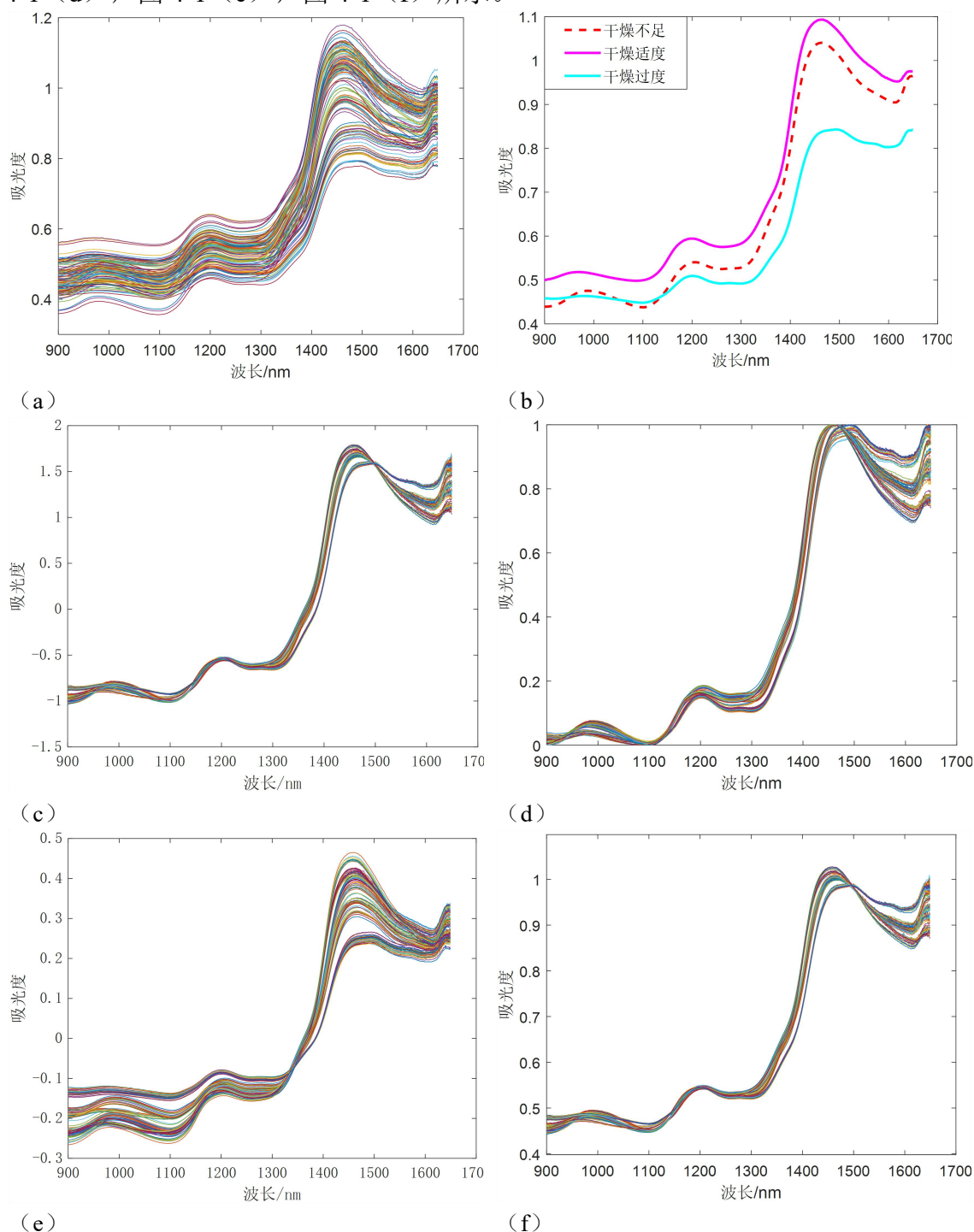


图 4-1 不同预处理方法对光谱的影响

Fig.4-1 Effects of different pretreatment methods on spectra

4.3.2 主成分分析

将不同干燥程度的原始光谱分别进行 SNV、Max-Min、MC 和 MSC 预处理，随后进行主成分分析，图 4-2 为对原始光谱依次预处理后进行主成分分析得到的前 3 个主成分投影图。对 SNV 预处理后的光谱数据进行主成分分析，其中第一主成分（PC1）贡献率为 93.19%，PC2 贡献率 5.17%，PC3 贡献率为 1.21%，前三个主成分累计贡献率为 99.57%，仅有 0.43% 的信息损失，表明前三个主成分可以代表原始光谱的绝大部分信息，三维载荷图如图 4-2（a）所示。对 Max-Min 预处理后的光谱数据进行主成分分析，其中 PC1 贡献率为 89.45%，PC2 贡献率为 7.10%，PC3 贡献率为 2.26%，前三个主成分累计贡献率为 98.81%，三维载荷图如图 4-2（b）所示。对 MC 预处理后的光谱数据进行主成分分析，其中 PC1 贡献率为 97.19%，PC2 贡献率 2.06%，PC3 贡献率为 0.57%，前三个主成分累计贡献率为 99.82%，三维载荷图如图 4-2（c）所示。对 MSC 预处理后的光谱进行主成分分析，其中 PC1 贡献率 93.33%，PC2 贡献率为 5.11%，PC3 贡献率为 1.20%，前三个主成分累计贡献率为 99.64%，三维载荷图如图 4-2（d）所示。

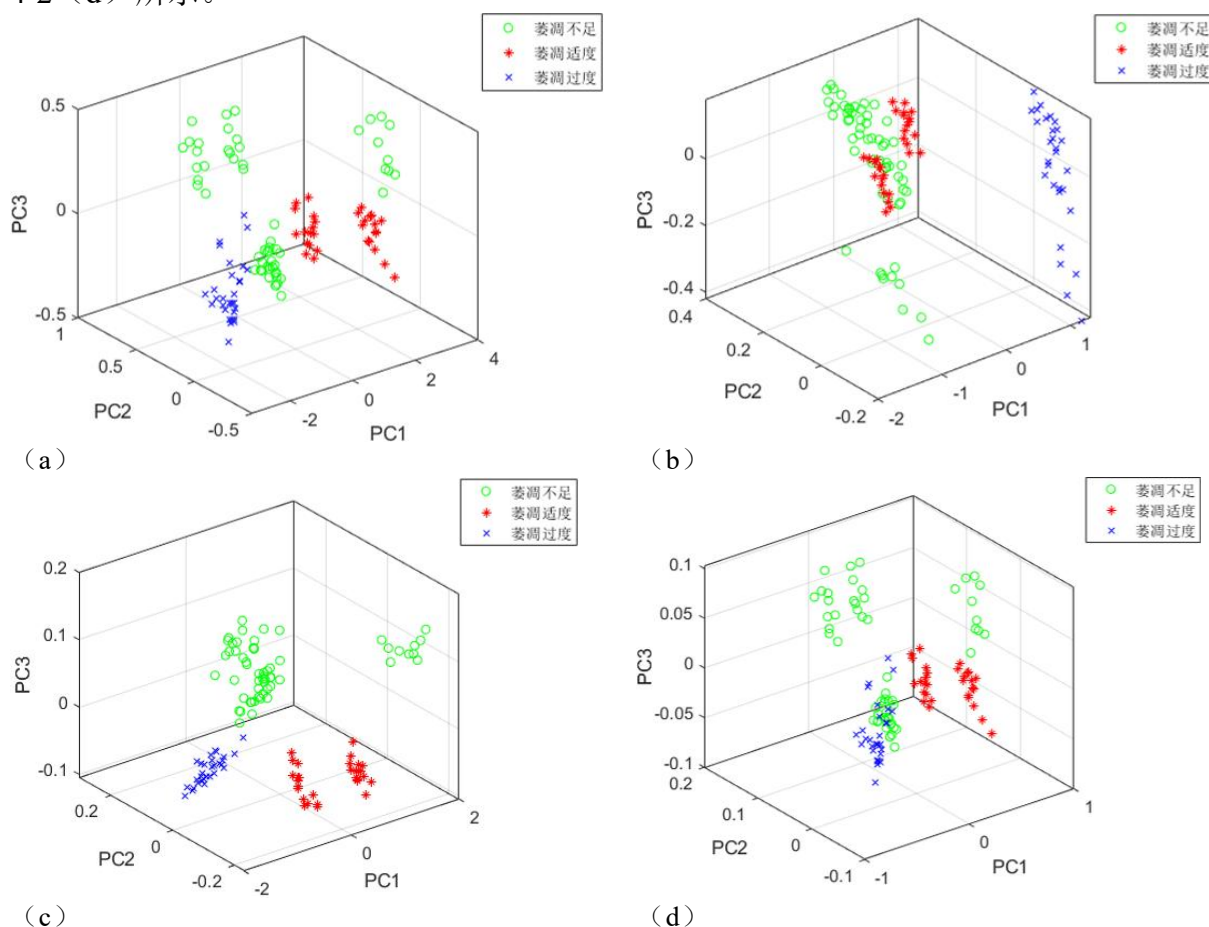


图 4-2 不同预处理后前三个主成分 3D 载荷分布图

Fig.4-2 Load graphs for the top three principal component scores of different preprocessing methods

4.3.3 基于不同预处理方法的线性判别

将 4 种不同预处理后的光谱数据作为 PLS-DA 模型的输入，建立基于不同干燥程度的绿茶样品判别分析模型，判别结果如表 4-1 所示。由表可知，基于原始光谱所建立的判别模型效果可以实现粗略的判别，其校正集识别率为 86.67%，预测集识别率为 83.33%，识别率均低于 90%，识别效果不佳。与原始光谱相比，所有预处理后的 PLS-DA 模型的判别性能均有所增加，其中，基于 MC 预处理后的光谱识别率效果最好，校正集识别正确率为 94.44%，预测集识别率为 93.33%，校正集识别率和预测集识别率均高于 90%，因此选取 MC 方法作为最终的预处理方法。

表 4-1 使用 PLS-DA 算法所建模型的判别分析结果

Table 4-1 Discriminant results of PLS-DA models											
预处理方法	干燥品质	NC	校正集判别结果				NP	预测集判别结果			
			1	2	3	识别率 /%		1	2	3	识别率/%
RAW	1	45	38	0	0	86.67	15	11	0	0	83.33
	2	20	0	20	0		10	0	10	0	
	3	25	0	0	24		5	0	0	4	
SNV	1	40	36	0	0	93.33	20	18	0	0	90
	2	26	0	25	0		4	0	4	0	
	3	24	0	0	23		6	0	0	5	
Max-Min	1	39	35	0	0	92.22	21	19	0	0	86.67
	2	25	0	24	0		5	0	4	0	
	3	26	0	0	24		4	0	0	3	
MC	1	47	44	0	0	94.44	13	12	0	0	93.33
	2	22	0	22	0		8	0	8	0	
	3	21	0	0	19		9	0	0	8	
MSC	1	40	37	0	0	93.33	20	18	0	0	90
	2	26	0	25	0		4	0	4	0	
	3	24	0	0	22		6	0	0	5	

4.3.4 特征波长筛选

4.3.4.1 基于 CARS 方法提取特征波长

CARS 提取的最大因子数由蒙特卡罗交叉验证决定, 采样次数设置为 50 次, 运行次数设置为 50 次, 所选的波长变量采用 10 折 PLS 交叉验证建模。基于 CARS 筛选特征波长的过程如图 4-3 所示, 选取 RMSECV 最小的一次为最优采样点, 当采用次数为 29 时 RMSECV 最小, 因此共筛选出 25 个特征波长, 特征波长为 913、932、951、964、965、970、971、972、1001、1002、1007、1009、1010、1165、1361、1427、1430、1467、1499、1504、1505、1533、1578、1587、1590。

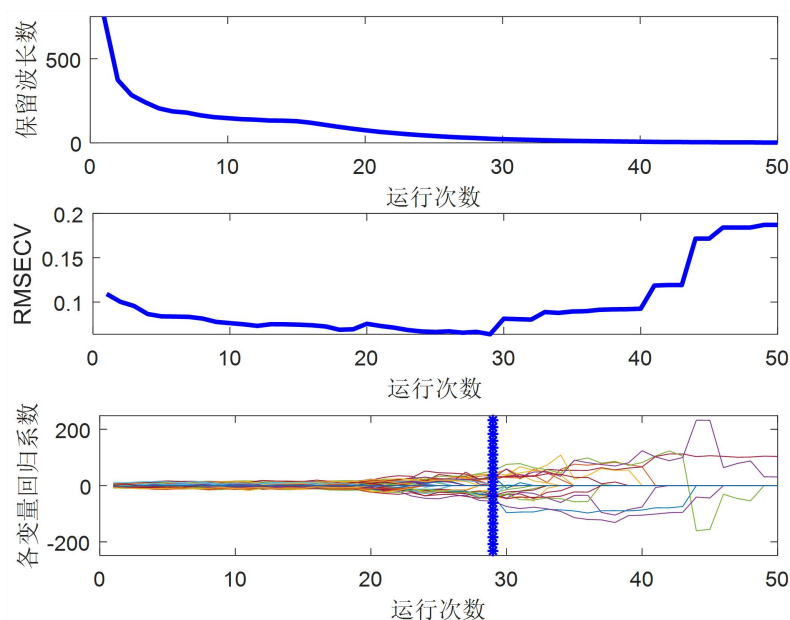


图 4-3 基于 CARS 算法的特征变量优选

Fig.4-3 Optimization of characteristic variables based on CARS algorithm

4.3.4.2 基于 VCPA 方法提取特征波长

VCPA 是一种新型变量筛选方法, 该算法基于基于模型集群分析 (model population analysis, MPA) 思想, 采用二进制矩阵采用法 (binary matrix sampling, BMS), 光谱中每个变量都存在被选中的几率, 利用指数衰减函数 (EDF) 确定最终被保留的波长变量, 缩小变量空间, 计算出剩余变量所有可能组合的 RMSECV 值, 并选取最小 RMSECV 的变量组合作为最终的建模数据。本章选择 BMS 运行的次数设为 1000, 最佳子集占比为 0.1, EDF 运行的次数设为 50, 剩余变量数目设为 10, 在 EDF 运行过程中, RMSECV 变化如图 5-1 (a) 所示, 最终选取 10 个特征波长: 933、938、963、964、1002、1008、1108、1429、14999、1556, 变量压缩率为

98.67%，特征波长筛选结果如图 4-4。

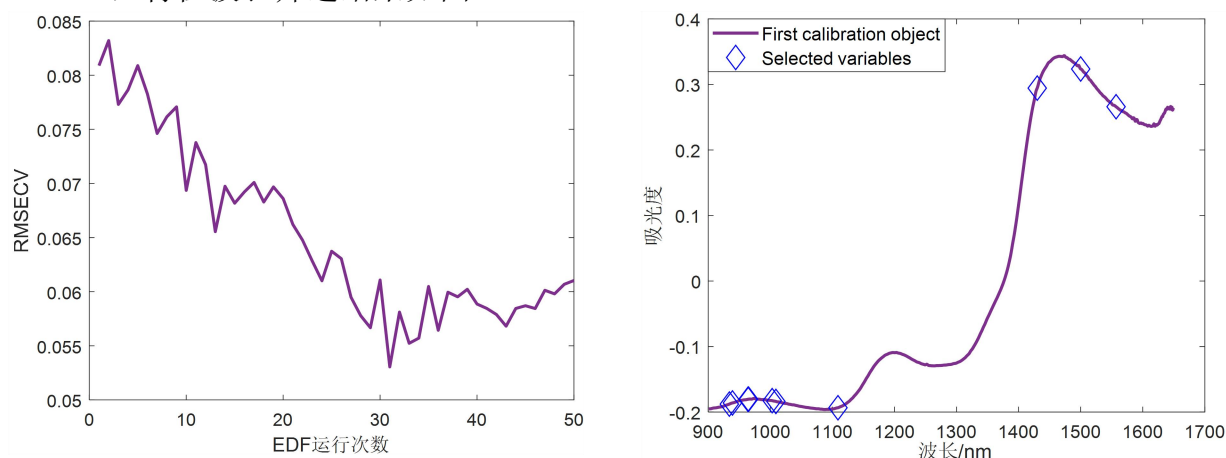


图 4-4 基于 VCPA 算法的特征变量优选

Fig.4-4 Optimization of characteristic variables based on VCPA algorithm

4.3.4.3 基于 VCPA-IRIV 方法提取特征波长

VCPA-IRIV 作为一种新型特征波长筛选算法，可以准确地筛选出特征波长，减少冗余信息。设置二进制矩阵采样法（BMS）采样次数为 1000 次，指数衰减函数（EDF）设置为 50，最优子模型的比例为 10%，最终筛选出 29 个特征变量，特征波长为 932、938、951、952、964、965、970、1003、1008、1026、1027、1056、1073、1139、1140、1165、1311、1338、1412、1415、1428、1429、1434、1451、1458、1498、1531、1555、1607，筛选结果如图 4-5 所示。

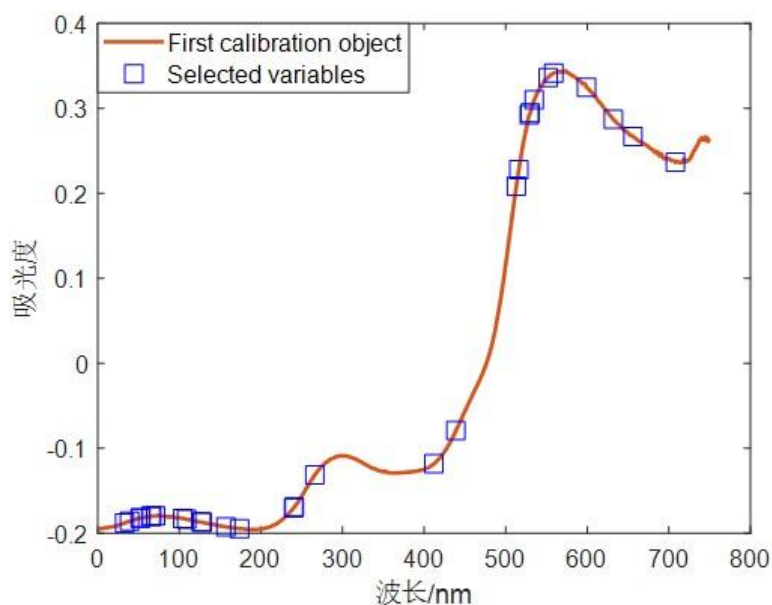


图 4-5 基于 VCPA-IRIV 算法的特征变量优选

Fig.4-5 Optimization of characteristic variables based on VCPA- IRIV algorithm

4.3.4.4 基于 VCPA-GA 方法提取特征波长

本研究设置二进制矩阵采样法 (BMS) 采样次数为 1000 次, 指数衰减函数 (exponentially decreasing function, EDF) 设置为 100, 最优子模型的比例为 10%, EDF 最后一次运行的左变量个数为 100, 最终筛选出 33 个特征波长, 特征波长为: 1499、1531、964、1067、1165、1556、1532、962、1003、1008、963、1498、1504、1001、970、1072、1428、1429、965、1505、1512、1002、1157、951、1135、1422、1513、1555、915、931、936、966、1361, 筛选结果如图 4-6 所示。

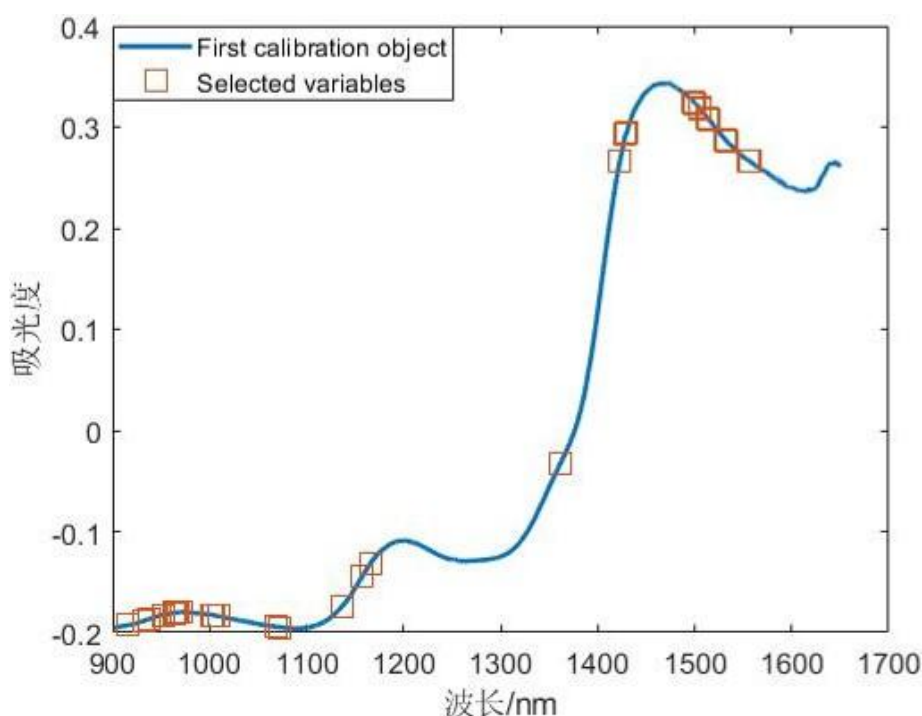


图 4-6 基于 VCPA-GA 算法的特征变量优选

Fig.4-6 Optimization of characteristic variables based on VCPA-GA algorithm

4.3.5 基于不同变量筛选的 KNN 判别模型

基于不同特征波长所建立 KNN 模型判别分析结果如表 4-2 所示。从表 4-2 可知通过不同特征波长筛选算法并结合主成分分析所建立的 KNN 判别模型效果相比于基于全光谱所建里的判别模型效果明显更优, 其中基于 VCPA-IRIV 算法结合 PCA 分析所建立的 KNN 判别模型的识别效果最好, 校正集识别率从 95.56%提高至 98.89%, 预测集识别率从 93.33%提高至 100%, 并且变量个数从 750 个减少至 29 个, 变量压缩率高达 96.13%, 识别结果如图 4-7 所示。而采用 VCPA 算法提取特征波长所建立的 KNN 模型识别效果最差, 但是校正集和预测集识别率虽然均高于 90%, 但是

识别效果低级基于全光谱所获得的模型效果，造成这一结果的原因可能是 VCPA 算法变量压缩率高达 98.66%，仅保留 10 个特征变量，剔除了一些相关变量，造成判别率下降。而基于 CARS 和 VCPA-GA 算法所建立的 KNN 判别模型效果相对于基于全波段所建立的 PLS-DA 判别模型效果校正集识别率和预测集识别率均高于 95%，识别效果较好。

表 4-2 使用 k 邻近算法所建模型的判别分析结果

Table 4-2 Results of discriminant analyses performed utilizing the k-nearest neighbors model

变量筛选方法	变量数	干燥品质	NC	校正集判别结果				NP	预测集判别结果			
				1	2	3	识别率/%		1	2	3	识别率/%
全光谱	750	1	47	44	3	0		13	11	2	0	
		2	22	1	21	0	95.56	8	0	8	0	93.33
		3	21	0	0	21		9	0	0	9	
CARS	25	1	47	46	1	0		13	12	1	0	
		2	22	0	22	0	98.88	8	0	8	0	96.66
		3	21	0	0	21		9	0	0	9	
VCPA	10	1	47	44	3	0		13	12	1	0	
		2	22	0	22	0	95.56	8	0	8	0	93.33
		3	21	0	0	20		9	0	0	8	
VCPA-IRIV	29	1	47	46	1	0		13	13	0	0	
		2	22	0	22	21	98.89	8	0	8	0	100
		3	21	0	0	21		9	0	0	9	
VCPA-GA	33	1	47	45	2	0		13	12	1	0	
		2	22	0	22	0	96.67	8	0	8	0	96.67
		3	21	0	0	20		9	0	0	9	

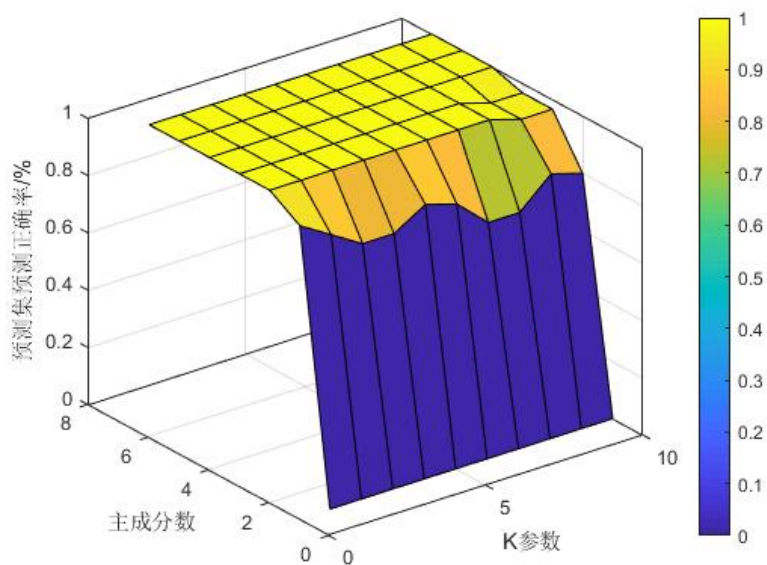


图 4-7 不同 K 值和 PCs 对应 KNN 模型的结果示意图

Fig.4-7 The result of KNN model at the different K value and principal component

4. 3. 6 基于不同变量筛选的 ELM 判别模型

基于不同特征波长筛选方法所建立的 ELM 模型判别效果如表 4-3 所示，从表 4-3 的结果可知，基于特征波长所建立的模型较全光谱-ELM 模型判别效果均有所提高，其中基于 VCPA-IRIV 方法提取特征波长结合 PCA 分析建立的 ELM 模型判别效果最好，校正集识别正确率和预测集识别正确率均达到了 100%，识别结果如图 4-7 所示。采用 VCPA 和 CARS 算提取特征波长识别效果较差，但校正集识别正确率和预测集识别正确率均达到 95%以上。

表 4-3 使用 ELM 算法所建模型的判别分析结果

Table 4-3 Results of discriminant analyses performed utilizing the ELM model

变量筛选方法	变量数	干燥品质	NC	校正集判别结果				NP	预测集判别结果			
				1	2	3	识别率 /%		1	2	3	识别率 /%
全光谱	750	1	47	44	3	0		13	12	1	0	
		2	22	0	22	0	94.44	8	0	8	0	93.33
		3	21	0	2	19		9	0	0	9	
		1	47	45	2	0		13	13	0	0	
		2	23	0	23	0		7	0	7	0	

CARS	25	3	20	0	0	20	97.77	10	0	0	10	100
		1	46	44	2	0		14	13	1	0	
VCPA	10	2	24	1	23	0	96.66	6	0	6	0	96.67
		3	20	0	0	20		10	0	0	10	
		1	46	46	0	0		14	14	0	0	
VCPA-IRIV	29	2	22	0	22	0	100	8	0	8	0	100
		3	22	0	0	22		8	0	0	8	
		1	46	46	0	0		14	14	0	0	
VCPA-GA	33	2	24	0	24	0	100	6	1	5	0	96.66
		3	20	0	0	20		10	0	0	10	

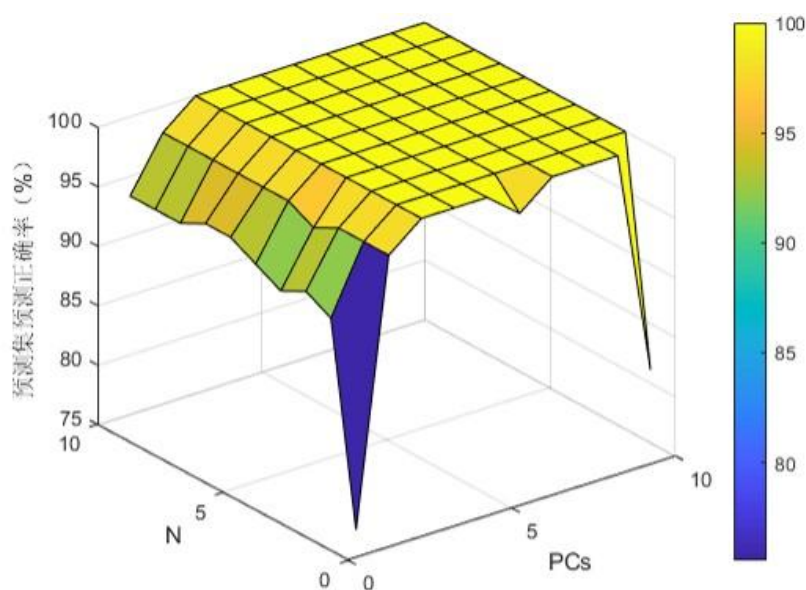


图 4-8 不同 PCs 和 N 对应的 ELM 模型判别结果

Fig.4-8 Discriminant results of ELM model corresponding to different PCs and N

4.3.7 PLS-DA、KNN 和 ELM 模型的比较

综合对比基于全波段建立的 PLS-DA 模型，基于特征波长和主成分分析方法所建立的 KNN 和 ELM 模型中，PLS-DA 模型的判别效果最差，ELM 模型判别效果最好。在建立的 PLS-DA 模型中基于 MC 预处理后判别效果最好，得的校正集识别正确率和预测集识别率均高于 90%，识别效果较好，因此选取 MC 为最佳预处理

方法。基于不同特征波长筛选方法结合 PCA 所建立的 KNN 模型中, 判别效果要优于 PLS-DA, 其中基于 VCPA-IRIV 特征波长筛选方法所建立的模型效果最优, 校正集识别率为 98.89, 预测集识别率为 100%, 且变量从 750 个压缩至 29 个, 变量压缩率达到 96.13%。经不同变量筛选方法所建立的所有 ELM 模型中, 校正集识别正确率和预测集识别正确率均高于 95%, 其中基于 VCPA-IRIV 特征波长筛选方法所建的模型效果最优, 校正集识别正确率和预测集识别正确率均为 100%。其中 ELM 取得最优的效果可能是因为 ELM 属于非线性单隐层前馈神经网络(Single-hidden layer feedforward neural network, SLFN)学习训练的模型, 初始化参数少、学习训练速度快, 建模过程不需要调整隐含层参数, 相当于一个线性参数模式, 训练过程简单, 能获得全局最优解并且泛化能力强。

4.4 本章小结

本研究采用近红外光谱分析技术鉴别绿茶的干燥程度, 基于不同的预处理方法、多种变量筛选方法和主成分分析, 尝试建立判别性能较高的近红外光谱模型, 为利用近红外光谱技术对红茶中外源蔗糖的判别和品质控制提供参考。主要结论如下:

1) 为了降低光谱噪声对模型性能的影响, 分析了 SNV、Max-Min、MC、MSC 四种不同的光谱降噪方法, 通过建立 PLS-DA 模型性能对比得出, 经过 MC 预处理方法建立的判别模型性能较好, 校正集判别正确率和预测集判别正确率分别为 94.44%和 93.33%, 表明经过 MC 预处理后的模型判别效果较好, 可以显著提高模型的精度, 其被确定为最终的预处理方法。

2) 探讨了通过 CARS、VCPA、VCPA-IRIV 和 VCPA-GA 特征波长筛选方法结合 PCA 分析, 建立的 KNN 和 ELM 模型性能, 发现基于 VCPA-IRIV 特征波长筛选和 PCA 后建立的 ELM 模型性能最优, 校正集识别率和预测集识别率均达到了 100%, 并且变量压缩率达到了 96.13%, 极大地提高了建模的效率。

第5章 基于机器视觉和近红外光谱的成品绿茶水分含量预测模型建立

5.1 引言

本研究第2, 3, 4章验证了绿茶每道加工工艺水分含量检测的可行性, 本章节将从绿茶整个加工过程出发, 以相同的加工工艺条件和实验设备, 以不同于第2, 3, 4章的茶叶品种种类验证绿茶整个加工过程中水分含量进行预测的可行性。

5.2 材料和方法

5.2.1 样品制备

试验于2021年4月22日在湖南省长沙智能装备股份有限公司实施, 茶鲜叶原料采集于湖南省岳阳市平江县, 茶叶品种采用楮叶齐茶, 嫩度为一芽一叶和一芽二叶。

将鲜叶原料(含水率为76.75%)自然摊放在萎凋槽中, 摊叶厚度为3-4cm, 室内环境温度保持在20℃左右, 相对湿度为60-65%, 为了保持茶叶水分散失均匀, 每隔一个小时茶叶翻拌一次, 在茶叶摊青8h之后, 采用水分测定仪测量茶叶水分含量(测量三次取平均值), 含水率为66.90%; 杀青在6CS-60型滚筒式杀青机内进行, 杀青温度为270+250+230℃, 滚筒转速为22r/min, 杀青时间为5分30秒, 水分降到54.21%; 杀青叶摊凉1小时后进行揉捻, 揉捻按照空揉10min→轻揉10min→中揉10min→轻揉5min, 总计35min, 含水率降至49.23%; 毛火和足火工序均在6CHZ-9型茶叶烘焙机内进行, 毛火温度设为110℃, 摊叶厚度2~3cm, 时间为15min, 含水率降到32.43%; 理条机温度为100℃, 理条时间为25min, 水分降到19.36%; 足烘温度为90℃, 足烘时间为30min, 水分降至6.02%。加工过程如图5-1所示。

5.2.2 光谱数据获取与水分含量测定

采用IAS-3100光谱仪(Intelligent Analysis Service Co. LTD)对绿茶加工过程中的摊青、杀青、毛火、理条和足火样品进行光谱采集。波长范围为900-1700nm; 光谱精度为±1nm, 分辨率为4cm⁻¹。试验前, 为了保证采集光谱数据的准确性, 首先将近红外光谱仪预热30min, 从而维持光谱仪器内部系统的稳定, 仪器预热结束后, 对仪器

进行自检，查看仪器运行是否正常，随后进行近红外光谱数据的采集。每个时刻取 14 个样本，将每个样品取约 100g 均匀平铺在规格为 $\Phi 70\text{mm} \times 10\text{mm}$ 的石英培养皿中，每个茶叶加工过程样进行 3 次光谱数据采集，取 3 次采集光谱数据的平均值作为该时刻样品的光谱信息。理条过程中采集的 50 条光谱、足火过程中采集的 42 条光谱数据，总共 253 条光谱数据用于含水率定量预测模型的建立。

在每个时序下采集完光谱数据后，从中取 3g 样品放入水分测定仪（MA35M-000230V1,Sartorius）中测定该时刻样品的水分值，该过程做 3 次重复，取 3 次重复结果的平均值作为该时刻下的真实水分含量。

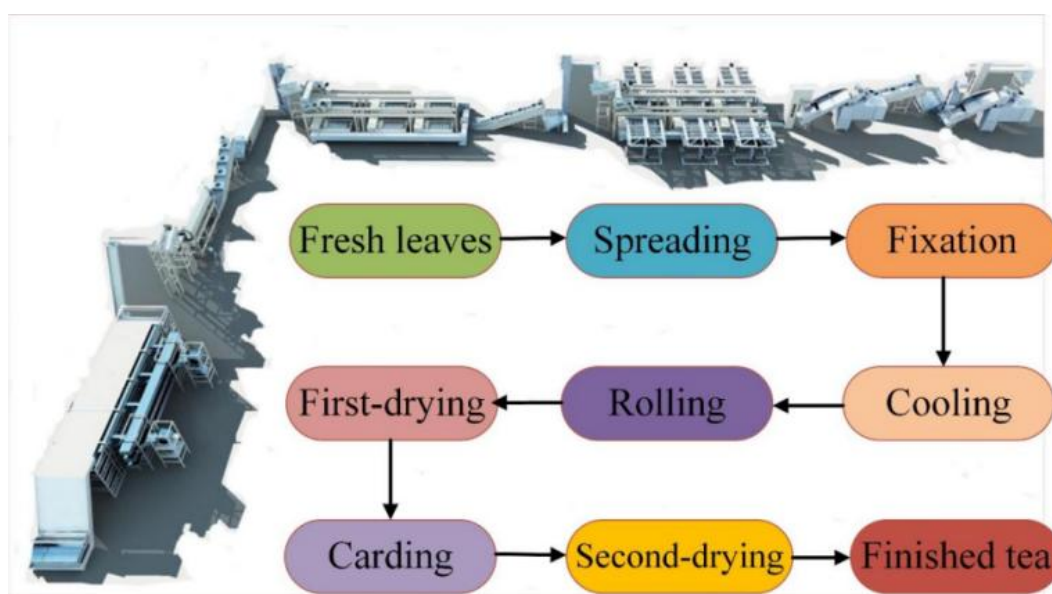


图 5-1 针叶绿茶加工生产线

Fig.5-1 Processing of needle green tea automatic production line

5.2.3 机器视觉系统的搭建和光谱、图像采集

在本研究中，所用的机器视觉系统和 GUI 软件处理系统与第三章保持一致。图像传感器采用专业工业相机（FI-S200C-G），其中镜头为 4mm 低畸变镜头，传感器为 1/2.8CMOS 图像传感器，分辨率为（1080*1080），曝光时间设为 0.09ms；光源采用 DOME 单色纯白弧形光源，强度为 100 勒克斯（lx）；图像的类型为 bmp 格式。相机镜头被固定在暗箱顶部，以确保拍摄角度和距离恒定。在图像采集前，为了保持暗箱中光源的稳定，提前 30 分钟打开弧形光源，将每个采集完光谱数据的样品放置于相机的视野中，采集样品图像，该研究共获得 253 张图像用于数据的分析和模型的建立。光谱数据的采集和图像的获取如图 5-2 所示。

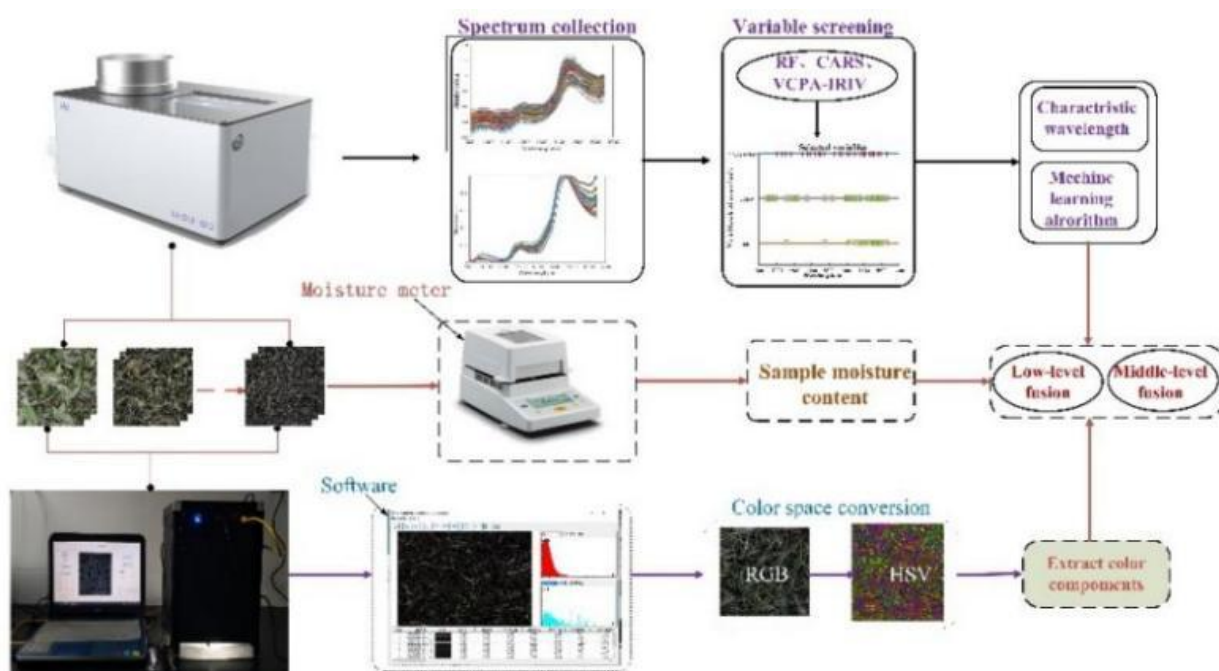


图 5-2 数据融合流程图

Fig.5-2 Flowchart of the data fusion used in this study

5.2.4 NIR 数据预处理与图像特征提取

近红外光谱采集过程中，由于仪器系统、样品和环境的影响，原始光谱中包含了基线漂移、光散射和随机噪声等无用光谱信息。为提高模型的精度，减少无用信息对所建模型的影响，本研究采用 4 中预处理方法来改善原始光谱信息，包括标准正态变换（SNV）、去趋势（DT）、均值中心化（MC）和标准差归一化（Z-Score）。

基于 Matlab GUI 模块开发图像处理模块（软件著作权号：2014SR149549）以图像中像素点（1728,1152）为中心，自动分割出 1000×1000 像素区域，提取该兴趣区域的色泽和纹理特征。根据 Dong 等^[45]应用的公式将 RGB 颜色空间转换为 HSV 空间。最后，获得了样品图样的 9 个颜色特征以及 6 个纹理特征其中红色通道均值（ R ）、绿茶通道均值（ G ）、蓝色通道均值（ B ）、色点均值（ H ）、饱和度均值（ S ）、亮度均值（ V ）、超绿变换（ $2G-R-B$ ）、红色通道均值与绿色通道均值的比值（ R/G ）和颜色（ hab^* ）；6 个纹理特征分别为：平均灰度值（ m ）、标准差（ δ ）、平滑度（ r ）、三阶距（ μ ）、一致性（ U ）、熵（ e ）。

5.2.5 NIR 特征波长筛选与特征提取

本研究采集光谱波段为 900-1700 nm，共有 800 个波长点，原始数据量庞大、存在

大量的冗余、无用信息，在建立模型的过程中容易降低模型的精度，因此需要对波长变量进行优选，从全光谱中获取最有效的特征光谱信息，从而提高检测速度和模型的预测精度。本研究采用随机蛙跳（RF）、竞争自适应重加权采样（CARS）以及结合变量组成集群分析的迭代保留信息变方法（VCPA-IRIV）3种算法提取特征波长。

变量筛选和特征提取都是对原始数据进行降维和压缩，剔除与待测对象无关的信息，简化建模过程，但不同的是，变量筛选方法是从原始光谱数据中筛选出有代表性的数据，而特征提取方法是通过映的方法对原始数据进行重组，从中提取新的特征供下一步处理分析。通常先经过变量筛选方法再进行特征提取，最后利用新提取的特征进行建模。本研究采用的特征提取方法为主成分分析法（PCA），在保证光谱主要光谱特征不丢失的情况下，降低数据集的维数，将多特征转化为少量主要特征^[59]。

5.2.6 光谱与图像特征信息数据融合

数据融合主要是通过使用互补的输入数据来增强结合技术之间的协同工作，以实现更好的模型预测^[60]。将提取的光谱特征波长、9个颜色特征和6个纹理特征集成到数据矩阵中进行下一步的分析，以避免潜在的信息丢失，既提高模型的精度也可以提高模型的稳定性。数据融合方式和第三章数据融合方式保持一致。

5.2.7 模型的建立和性能评估

本研究采用 Kennard-Stone (K-S) 算法对绿茶加工过程图像和光谱数据进行划分，其中校正集与预测集比例为 3:1^[61]，其中校正集有 190 个光谱和图像数据，预测集有 63 个光谱和图像数据。分别利用光谱和图像特征信息建立 PLSR、SVR 水分含量定量预测模型。本章节所采用的，模型性能评估指标和评价标准与第二章和第三章保持一致。本章节中所有的数据分析与处理均在 Origin 2018 和 Matlab 2019a (the MathWorks Inc., Massachusetts, USA) 软件中进行。

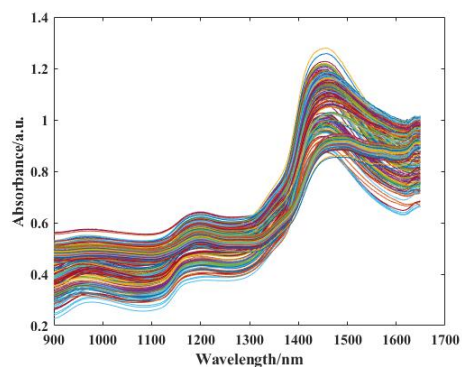
5.3 结果和讨论

5.3.1 光谱特征和预处理结果

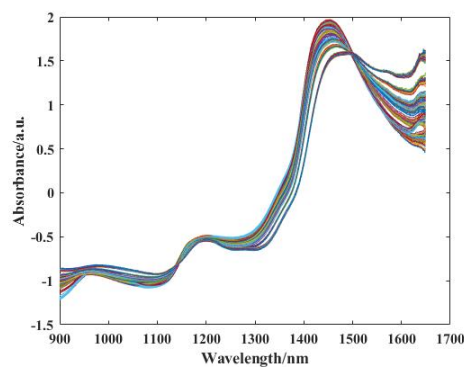
所有样本的原始光谱、预处理以及每个加工过程的平均光谱如图 5-3 所示。从平均光谱中可以看出摊青和杀青主要吸收峰位于 960 nm (10530 cm^{-1})、1440 nm (6944 cm^{-1})

和 1220 nm (8197 cm^{-1})。因为在摊青和杀青的过程中水分含量较高,而 OH 键是水分的主要标志性基团,960 nm (10530 cm^{-1})处的吸收峰主要与水分子中 OH 基团二级倍频有关,1440 nm (6944 cm^{-1})处的吸收峰是由于水分 OH 基团一级倍频在 1440 nm (6944 cm^{-1})伸缩振动引起的,1220 nm (8197 cm^{-1})处的吸收峰是由于蛋白质 CH 基团的二级倍频以及水分子 OH 的合频吸收带所引起的。毛火、理条和足火的水分含量较低,从图中可以看出吸收峰主要位于 1220 nm (8197 cm^{-1})和 1440 nm (6944 cm^{-1})处,在 960 nm 处的吸收峰逐渐消失,这是因为水分子 OH 基团二级倍频的消失,以及 CH 键伸缩振动引起的。毛火、理条和足火在 1620 nm 处光谱反射率明显高于摊青和杀青的光谱反射率,这也与 CH 键在 1600-1800 nm 处的一级倍频有关。

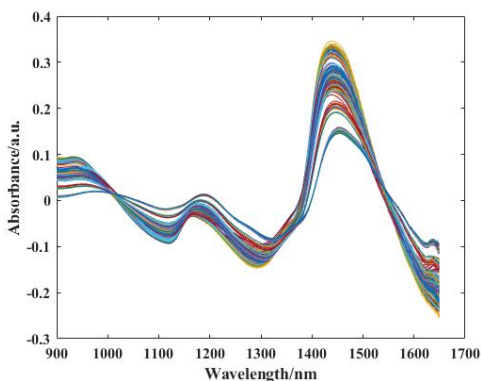
基于原始光谱和预处理后的数据建立 PLSR 模型,以获得最佳的预处理结果,结果如表 5-1 所示。基于原始光谱所建立的 PLSR 模型校正集相关系数 R_c 值为 0.9678, RMSEC 为 0.0523, 预测集相关系数 R_p 值为 0.9542, RMSEP 为 0.0564, RPD 值 3.0977。与原始光谱相比,预处理后的模型效果均有所提高,其中经过 Z-score 预处理所建立的 PLSR 模型的预测效果最好,校正集相关系数为 0.9715,预测集相关系数为 0.9602, RPD 值为 3.4836,表明 Z-score 可以有效的表面散射,仪器噪声引起的光谱误差,可以有效的提高模型的精度。因此选择 Z-score 作为最终的预处理方法,以进行后续建模分析。



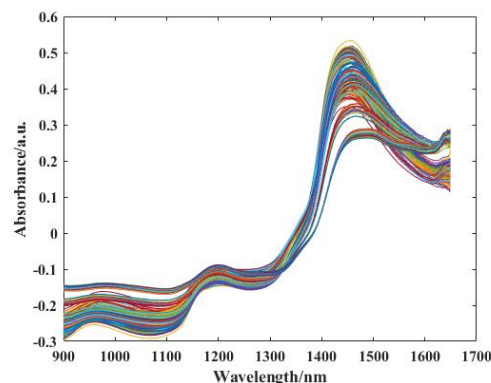
(a) 原始光谱



(b) SNV 预处理后的光谱



(c) Detend 预处理后的光谱



(d) MC 预处理后的光谱

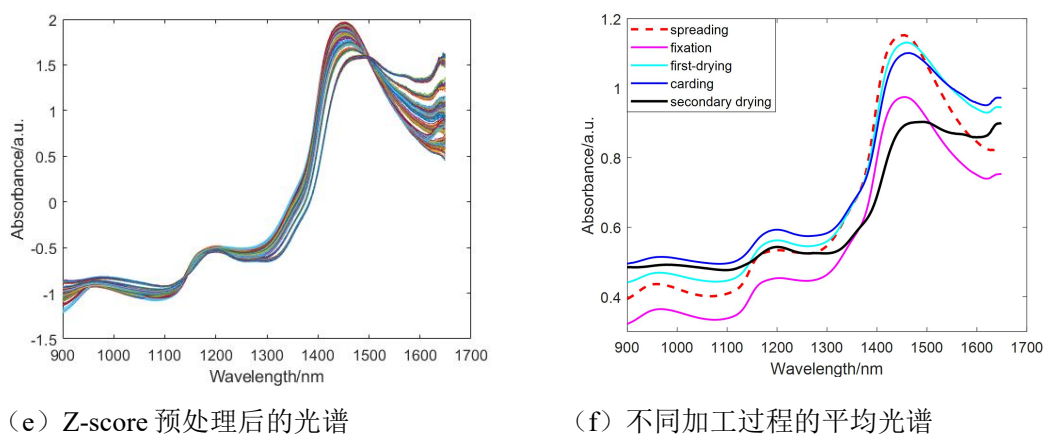


图 5-3 不同加工过程中近红外原始光谱、预处理光谱和平均光谱图

Fig.5-3 NIR raw spectrum、 various preprocessed spectrum and average spectra in different processes

表 5-1 不同预处理方法含水量预测模型结果

Table 5-1 Result of moisture content prediction model for different pretreatment methods

Pretreatment methods	PCs	Calibration set		Prediction set		RPD
		Rc	RMSEC	Rp	RMSEP	
RWA	8	0.9678	0.0523	0.9542	0.0564	3.0977
SNV	6	0.9712	0.0510	0.9586	0.0586	3.2620
Detrend	8	0.9657	0.0512	0.9610	0.0581	3.2525
MC	6	0.9688	0.0511	0.9600	0.0578	3.2635
Z-score	6	0.9715	0.0510	0.9602	0.0561	3.4836

5. 3. 2 加工过程中颜色变量响应

为了更好的研究绿茶加工过程中叶片颜色变化规律，该研究对整体的视觉变化进行了分析。根据加工顺序排列图像，然后提取图像的平均颜色。结果如图 5-3(A)和 5-3(B)所示。为了更好的区分每个加工过程中图像的差异性，对绿茶加工过程中的平均

颜色进行增强,结果如图 5-4 (C) 所示。在不同的加工过程中,颜色会有一定差异,为了进一步地区分差异,增强了图像的饱和度和亮度。之后,研究将 RGB 图像转换为 HSV 颜色模型,然后提取图像的平均颜色。结果如图 5-4 (D) 和 5-4 (E) 所示。

总体上,鲜叶的颜色在加工后发生了显著的变化:从茶鲜叶样品的鲜绿色到卷曲和颜色变暗、失去光泽,至理条和烘干阶段基本形成针状、深绿带灰的外关特征,这种由相机镜头捕捉到的颜色变化过程通过图像处理算法进行量化。不同颜色变量空间,包括 RGB 和 HSV 以及纹理特征值的变化过程如图 5-5 (a) ~ 5-5 (o) 所示。

若从图像中提取颜色和纹理特征组成数据矩阵,则该矩阵属于高维数组,且各个变量之间的维数不同,因此有必要对该数组进行归一化 (Z-score) 预处理,以消除数组维数不同对模型精度和稳定性的影响。此外,各个变量之间仍然存在一定的相关性,这表明各个变量之间存在冗余信息,若将所有信息变量作为模型的输入,则会造成模型过拟合现象。因此,在建立定量预测模型之前,应对其进行降维处理。选用 PCA 消除各个变量之间的共线性,避免过拟合现象的发生。

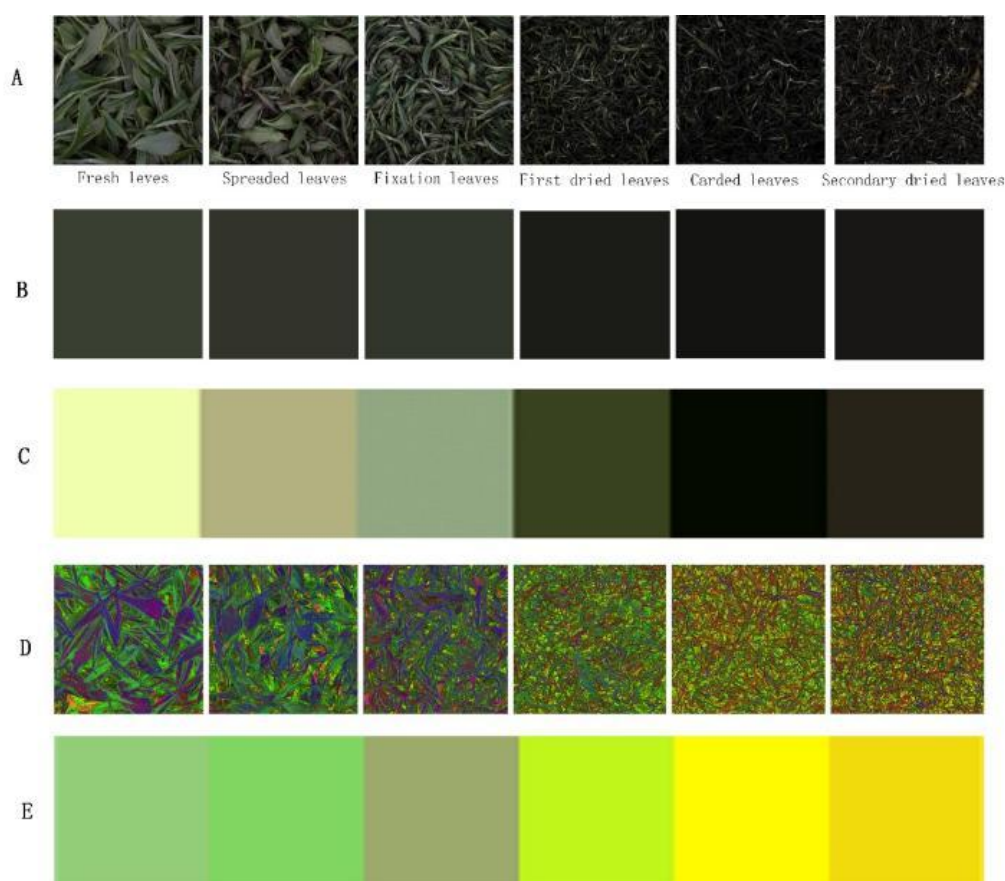


图 5-4 (A) 绿茶加工的 RGB 图像; (B) RGB 图像的平均颜色; (C) 增强 RGB 平均彩色图像;
(D) 绿茶不同加工工艺下的 HSV 图像; (E) HSV 图像的平均颜色

Fig.5-4 (A) RGB images of green tea processing; (B) Average color of RGB images; (C) Strengthened RGB average color images; (D) HSV images of green tea processing; (E) Average color of HSV images

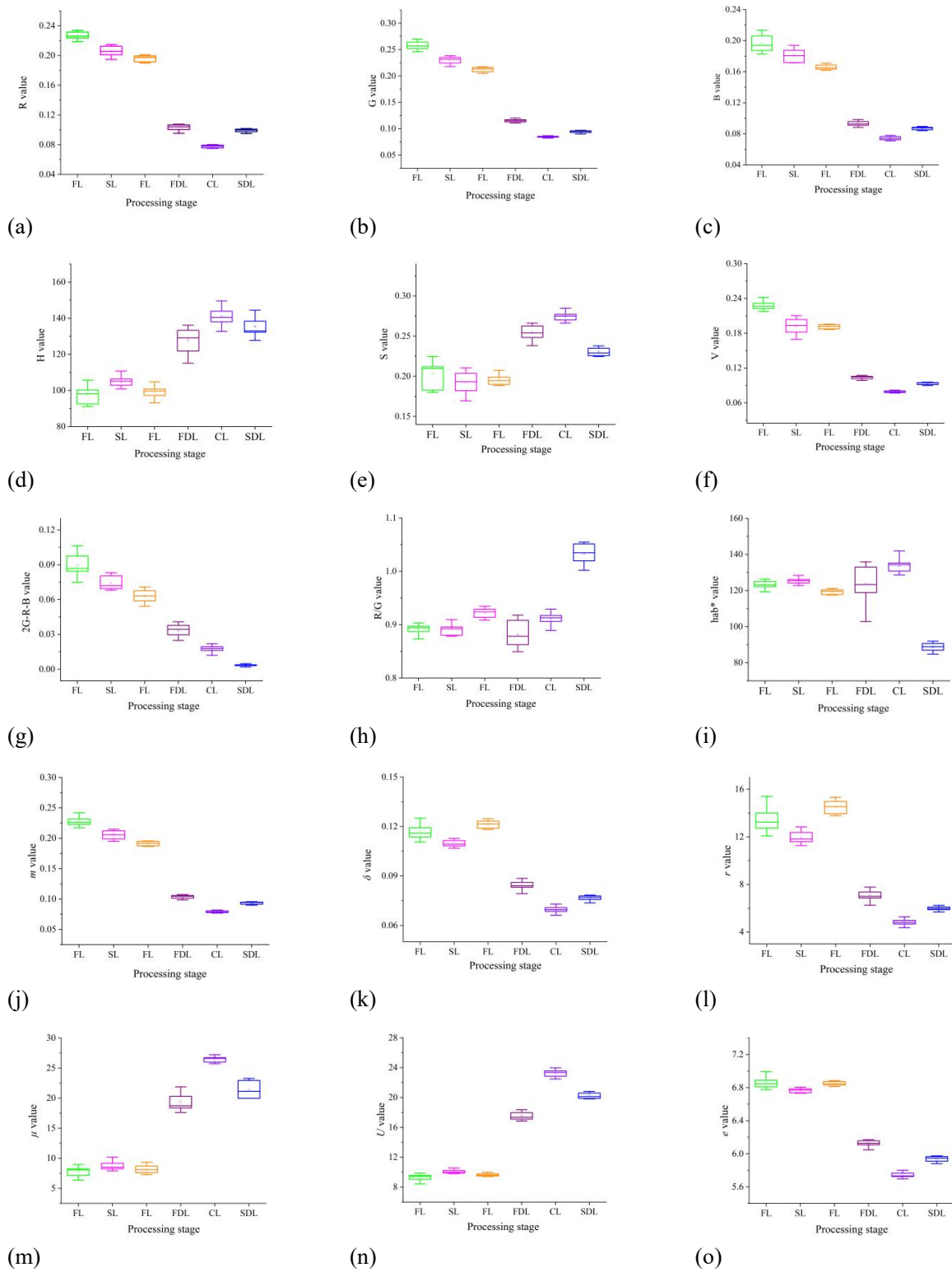


图 5-5 图像颜色和纹理特征；（a-i）颜色变量的箱线图：(a) R;(b) G; (c) B; (d) H; (e) S; (f) V; (g)

2G-R-B; (h) R/G; (i) hab*; and(j-o) box plots of texture (j) m; (k) δ ; (l) r; (m) μ ; (n) U; (o) e

Fig.5-5 Image color and texture characteristics; (a-i) box plots of color variables:(a) R;(b) G; (c) B; (d) H;

(e) S; (f) V; (g) 2G-R-B; (h) R/G; (i) hab*; and(j-o) box plots of texture (j) m; (k) δ ; (l) r; (m) μ ;

(n) U; (o) e

5.3.3 NIR 特征波长提取

基于最优的 Z-score 预处理方法, 通过不同的变量筛选方法提取特征波长, 随后进行主成分分析。其中经过 RF 算法共筛选出 25 个特征长, 经过 CARS 算法筛选出 20 个特征波长, 而经过 VCPA-IRIV 算法筛选出了 32 个特征波长, 筛选结果如图 5 所示。从图 5-6 中可以看出特征波长主要分布于 1400-1650nm 范围内, 因为绿茶加工完成后水分子 OH 伸缩振动一级倍频在 1440nm 处左右, 故特征波长 1440 范围内较为集中。而且绿茶加工过程中也会伴随着氨基酸、茶多酚和咖啡碱等物质的变化, NH 基团存在于氨基酸中, 1500nm 处为 NH 基团一级倍频吸收谱带位置。CH 基团存在于咖啡因之中, 一级倍频为 1600-1800nm 范围内, 故在 1400-1650nm 范围内特征变量较为集中。

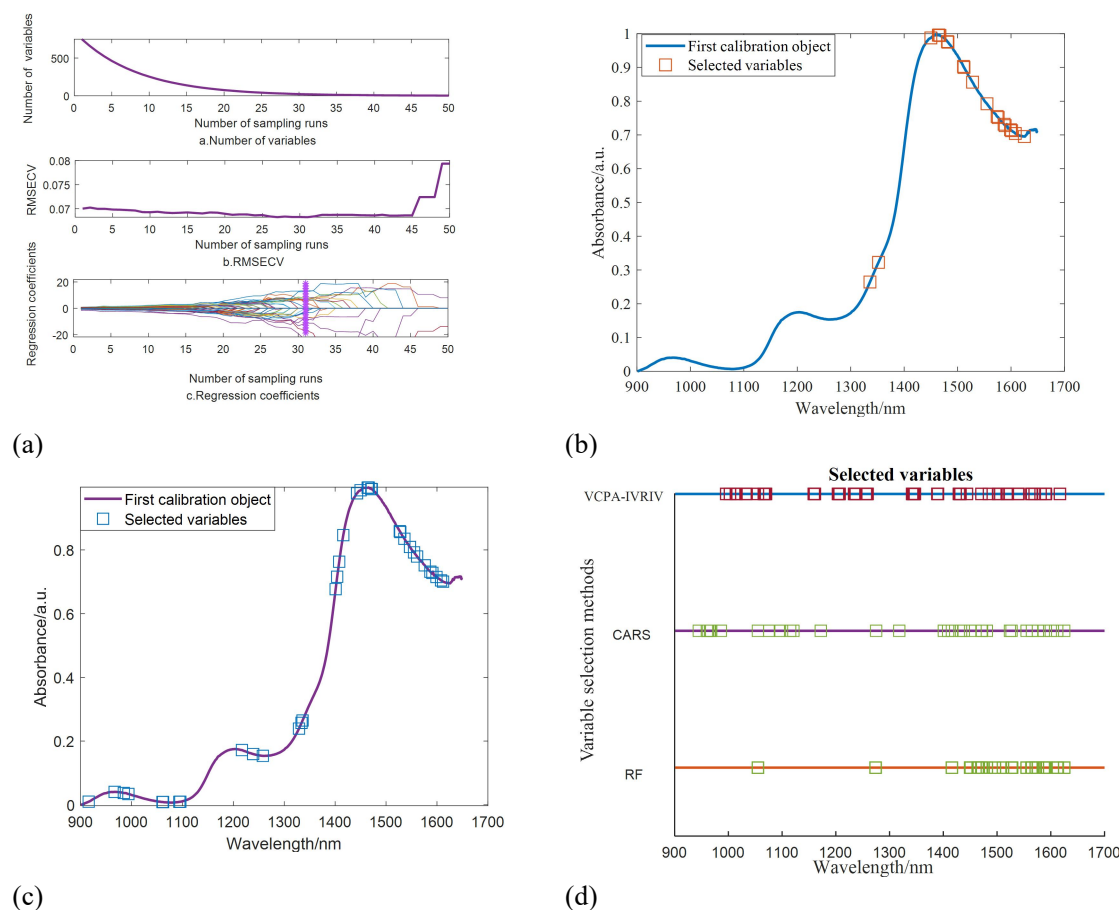


图 5-6 不同变量筛选方法选择的特征波长变量: (a) 随机青蛙; (b) 竞争性自适应重加权抽样法; (c) 结合变量组成集群分析的迭代保留信息变方法; (d) 通过不同的变量选择方法选择特征波长变量

Fig.5-6 Characteristic wavelength variables selected by different variable screening methods:(a) Random frog; (b) Competitive adaptive reweighted sampling; (c) Variable combination population analysis combined with genetic algorithm; (d) Selection of characteristic wavelength variables by different variable selection methods

5.3.4 基于单个传感器数据的预测模型

基于光谱特征波段所建立的 PLSR 和 SVR 模型预测结果如表 5-2 所示。对比线性的 PLSR 预测模型结果可知，采用 VCPA-IRIV+PCA 算法所建立的预测模型效果最优，其 Rc 值从 0.9715 提高至 0.9725，Rp 值从 0.9602 提高至 0.9689，RPD 从 3.4836 提高至 4.0035，并且特征波长从 750 减少至 37，变量压缩率高达 95.07%。采用 CARS 结合 PCA 所建立的模型效果相比预处理后的模型效果较差，可能 CARS 在提取特征波长的过程中可能剔除了一些重要的特征波长，因此建立的 PLSR 模型中 Rp 值较全波段模型精度降低。为了获得较优模型以实现绿茶加工过程中水分含量的定量预测，建立线性 PLSR 模型的同时，建立 SVR 非线性预测模型进行比较。从表 2 中可知，相比于 PLSR 模型结果，基于不同变量筛选方法和主成分分析后建立的 SVR 模型结果均有所提高，其中所建立的 VCPA-IRIV+PCA+SVR 模型效果最优，Rc 值从 0.9745 提高至 0.9796，Rp 值从 0.9689 提高至 0.9736，RPD 值从 4.0035 提高至 4.4154，均优于其他模型效果，参数寻优结果以及预测值和实测值的关系如图 5-7（a）和 5-7（b）所示。

基于图像色泽和纹理特征所建立的 PLSR 和 SVR 模型中，没有对色泽和纹理特征组成的数据矩阵归一化处理所建立的模型效果较差，Rp 值高于 Rc 值，这表明各个变量之间存在较多的冗余信息，模型出现过拟合现象。因此在基于图像特征建立模型时，应首先对组成的数据矩阵归一化处理。从表 5-2 中可以看出，对图像色泽和纹理特征组成的数据矩阵归一化处理后所建立的 PLSR 和 SVR 模型中 Rc 值分别达到了 0.9581 和 0.9743，Rp 值分别为 0.9505 和 0.9669，RPD 值分别为 3.5388 和 4.08553，表明模型具有很好的预测性能，其中所建立的 SVR 模型参数寻优结果以及预测值和实测值的关系如图 5-7（c）和 5-7（d）所示。但是相比于近红外光谱预测模型，基于图像色泽和纹理特征所建立的模型具有更低的 Rc 和 Rp 值以及更高的 RMSEC 和 RMSRP 所示。导致这一结果的原因可能是机器视觉可以直接捕获颜色相关成分信息，但是这是一种间接的测量，有些成分并不能完全捕获，相比之下，近红外光谱对 OH 基、NH 基和 CH 基团具有很好的相应，可以很好的捕获绿茶加工过程中的内质成分。

表 5-2 利用光谱和图像特征数据所建立的 PLSR 和 SVR 绿茶加工过程中水分含量预测模型

Table 5-2 Performance of PLSR and SVR models for moisture content prediction during green tea processing by using single spectral and image feature data.

Prediction method	Models	Data dimensionality reduction	PCs	Calibration set		Prediction set		RPD
				Rc	RMSEC	Rp	RMSEP	

Near-infrared spectroscopy	PLSR	RF+PCA	8	0.9734	0.0510	0.9679	0.0505	3.9175
		CARS+PCA	9	0.9740	0.0508	0.9556	0.0585	3.3921
		VCPA-IRIV+PCA	8	0.9745	0.0508	0.9689	0.0504	4.0035
	SVR	RF+PCA+PCA	5	0.9791	0.0454	0.9680	0.0513	3.9798
		CARS+PCA	6	0.9831	0.0412	0.9566	0.0608	3.4202
		VCPA-IRIV+PCA	5	0.9796	0.0443	0.9736	0.0460	4.4154
	PLSR		10	0.9576	0.0623	0.9702	0.0575	4.0211
		PCA	8	0.9722	0.0564	0.9760	0.0507	3.9958
			10	0.9581	0.0610	0.9505	0.0656	3.5388
	SVR	Z-score + PCA	7	0.9743	0.0481	0.9669	0.0512	4.0855

5.3.5 基于数据融合的预测模型

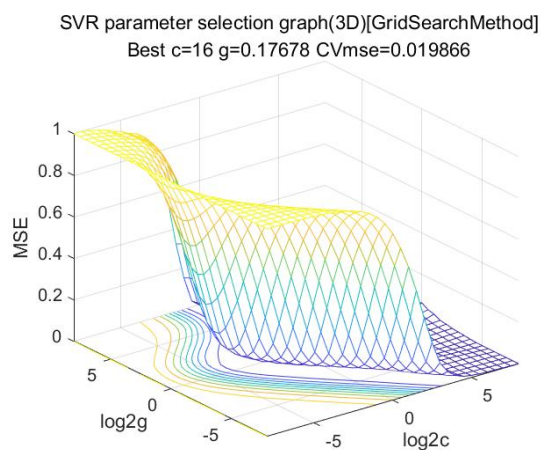
表 5-3 显示使用光谱和图像特征数据融合进行绿茶加工过程中水分含量预测的模型性能。基于低级数据融合所建立的 PLSR 模型得到了很好的预测效果，其 R_c 值为 0.9736， R_p 值为 0.9756，RPD 值为 4.1740，但是这种预测精度是不可接受的，因为相比于单一传感器所建立的模型效果并没有提高，而且预测模型的 R_p 值低于 R_c 值，表明各变量信息之间存在较多的冗余信息。与低级数据融合所建立的 PLSR 型相比，基于特征级融合的 PLSR 模型获得了更好的预测精度，因为在特征级融合的过程中剔除了很多无用的光谱信息，仅用有代表性的特征波段和图像特征进行数据融合。其中基于 VCPA-IRIV 算法筛选光谱特征波段融合图像色泽和纹理特征所建立的模型下效果最优，其 R_c 值为 0.9785， R_p 值为 0.9742，RPD 值为 4.5678，建立的 PLSR 模型效果优于基于单个传感器数据所建立的模型。

对比建立的 SVR 模型效果，基于数据级数据融合所建立的模型得到了很好的效果， R_c 值为 0.9783， R_p 值为 0.9756，RPD 值为 4.0764，但是相比于基于光谱所建立模型效果较差，因此此种模型也是不适用的。对于特征级数据融合所建立的 SVR 模型效果均优于数据级融合，其中基于 CARS 算法提取光谱特征波长然后融合图像色泽和纹理特征所建立的模型效果最优， R_c 、 R_p 以及 RPD 值分别为 0.9804、0.9777 和 4.5002，模型参数寻优结果以及预测值值和实测值的关系如图 5-7 (e) 和 5-7 (f) 所示。

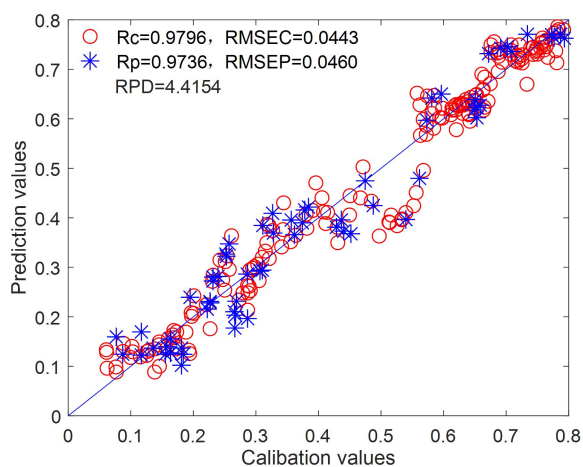
表 5-2 利用合数据融合数据所建立的 PLSR 和 SVR 水分含量预测模型

Table 5-3 Performance of PLS and SVR models for moisture content prediction during green tea processing by using fused data.

Model	Fused data	Variable Compression	PCs	Calibration set		Prediction set		RPD
				Rc	RMSEC	Rp	RMSEP	
PLSR	Low-level	Z-srore+ PCA	9	0.9736	0.0491	0.9756	0.0499	4.1740
		RF+PCA	9	0.9779	0.0479	0.9747	0.0479	4.5051
	Middle-level	CARS+PCA	7	0.9761	0.0491	0.9737	0.0497	4.3161
		VCPA-IRIV	8	0.9785	0.0484	0.9742	0.0478	4.5678
SVR	Low-level	Z-srore+PCA	7	0.9783	0.0453	0.9756	0.0514	4.0764
		RF+PCA	6	0.9794	0.0437	0.9773	0.0499	4.2040
	Middle-level	CARS+PCA	6	0.9804	0.0425	0.9777	0.0490	4.5002
		VCPA-IRIV+PCA	6	0.9795	0.0433	0.9775	0.0505	4.2156



(a) 基于近红外光谱数据的模型参数优化



(b) 基于近红外的预测集与校正集之间关系

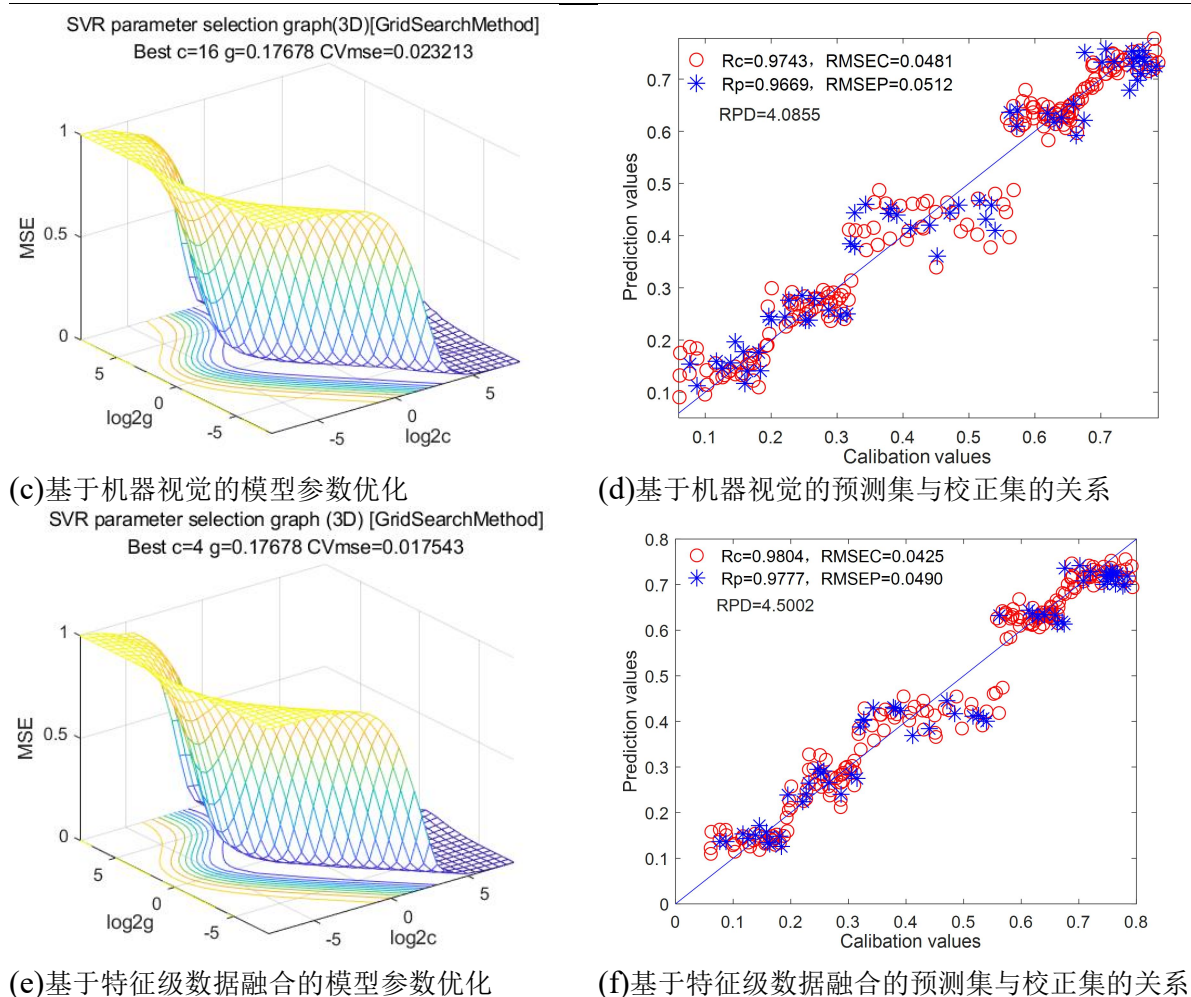


图 5-7 基于 SVR 的模型优化

Fig.5-7 Model optimization based on SVR

5.4 本章小结

本研究证明了使用近红外光谱结合机器视觉技术对绿茶加工过程中水分含量进行定量预测的可行性。分别采用单一传感器获得的数据以及融合两个传感器之间的数据进行水分含量的定量预测。结果表明：

(1) 基于近红外光谱技术，可以有效地对绿茶加工过程中水分含量进行预测，其中基于 Z-score 光谱预处理，VCPA-IRIV 特征波长筛选方法所建立的 SVR 模型效果最优，其 R_c 值为 0.9796，RMSEC 为 0.0443， R_p 值为 0.9736，RMSEP 为 0.0460，RPD 为 4.4154。

(2) 基于机器视觉技术，所建立的模型效果要低于基于光谱数据所建立的模型效果。其中对提取的 9 个颜色特征 6 个纹理特征所组成的数据矩阵没有归一化处理所建立的模型效果较差， R_c 值要低于 R_p 值，存在较多冗余的信息，模型并不可用；经 Z-

score 预处理和 PCA 后的图像信息所建立的模型获得了很好的预测效果, 其中建立的 SVR 模型效果最好, R_c 值为 0.9743, RMSEC 为 0.0481, R_p 值为 0.9669, RMSEP 为 0.0512, RPD 为 4.0855.

(3) 基于机器视觉和近红外光谱技术的数据级数据融合并不能有效的提高模型效果, 这是因为存在较多的冗余信息导致的。而通过 CARS 特征波长筛选方法筛选出的 20 光谱特征波段融合图像的 15 个色泽和纹理特征所建立 SVR 水分含量预测模型获得了最优的模型效果, 优于单一传感器以及数据级数据融合所获得的模型效果。其中 R_c 值为 0.9804, R_p 值为 0.9777, RPD 值为 4.5002, 表明机器视觉结合近红外光谱能够实现绿茶加工过程中水分含量的定量预测, 有效的提高单个传感器预测精度不足的问题。

第 6 章 结论与展望

6.1 结论

本研究以国家重点研发子课题：茶叶精制智能化装备研发为依托，针对绿茶加工过程中水分含量主要依靠制茶师傅经验判断水分含量，受主观因素影响较大，水分含量无法实现精准、实时检测等问题，基于近红外光谱技术，同时结合机器视觉技术，以不同加工工序下绿茶在制品为研究对象，使用机器学习及化学计量学方法，构建绿茶不同加工工序下的水分含量预测模型和干燥程度判别模型，实现绿茶水分含量有效、快速评估。主要的研究成果如下：

(1) 探讨了采用近红外光谱技术对绿茶摊青过程中水分含量进行实时预测的可行性，分别建立了基于近红外光谱检测技术的绿茶摊青水分含量预测模型。在模型建立的过程中，探讨了多种预处理方法、变量筛选方法对模型性能的影响。

研究表明，最小二乘平滑滤波（SG FIR）原始光谱降噪效果最优，SPA、CARS、RF、IRIV、VCPA、VCPA-GA 变量筛选方法可以有效地压缩光谱数据变量，并筛选出于水分含量相关的最佳特征波长。采用优选的特征波长构建 PLSR 和 SVR 预测模型，结果表明，基于 VCPA-GA 特征波长筛选方法所建立的 SVR 模型效果最优， R_p 值为 0.9658，RPD 为 3.6787，表明模型具有很好的预测性能，近红外光谱技术可以实现绿茶摊青过程中含水率的快速检测。

(2) 探讨了采用近红外光谱结合机器视觉技术对绿茶杀青过程中水分含量预测的可行性，建立了一种基于数据融合的绿茶杀青过程中水分含量的定量评价模型，实现了对绿茶杀青在制品茶叶的水分含量的快速表征。

基于团队研制的机器视觉采集系统和 GUI 分析软件，采集不同杀青时刻下茶叶样品图像信息并提取图像中的色泽和纹理特征，在采集图像信息的同时采集茶叶光谱数据，分析不同杀青时刻下图像色泽和光谱特征的变化规律和相关性。通过不同光谱预处理方法和特征提取方法筛选出于杀青叶水分含量相关的特征波长，将光谱特征信息与图像色泽和纹理信息相融合，从而建立 PLSR 和 SVR 预测模型。结果表明，基于 CARS 算法提取光谱特征波长融合图像的 15 个颜色特征，然后采用 Normalize 和 PCA 分析所建立的 SVR 模型效果最优，其中 R_c 值为 0.9742， R_p 值为 0.9719，RPD 值为 4.1546，且 RMSEC 和 RMSEP 差值仅为 0.004。因此，本研究证明了融合光谱和成像系统对绿茶加工过程中水分含量的预测能力，克服了单一传感器预测能力不

足的问题。

(3) 建立绿茶干燥程度判别模型时, 使用的预处理方法可以有效消除光谱中的噪声信息, 然后通过建立 PLS-DA 判别模型, 优选出 MC 方法为最优预处理方法。通过特征波长筛选方法去除光谱中冗余信息的干扰, 分别建立 KNN 和 ELM 判别模型, 最终确定 VCPA-IRIV-ELM 为绿茶干燥程度最佳判别模型, 校正集和预测集判别率均为 100%, 判别效果极好。研究成果为绿茶干燥过程中水分含量控制提供了新的思路和技术途径。

(4) 验证了另一茶叶品种在相似加工工艺下整个加工过程水分含量的预测模型效果, 实验设备和试验方案与第 2, 3, 4 章所用一致, 图像信息和光谱数据提取时, 茶叶样品来源于实际的茶叶生产线之中。基于图像和光谱信息融合建立的预测模型中, 采用 CARS 特征波长筛选方法筛选出的 20 光谱特征波段融合图像的 15 个色泽和纹理特征所建立 SVR 水分含量预测模型获得了最优的模型效果, 优于单一传感器以及数据级数据融合所获得的模型效果。其中 R_c 值为 0.9804, R_p 值为 0.9777, RPD 值为 4.5002, 表明机器视觉结合近红外光谱能够实现绿茶从鲜叶至成品茶整个加工过程中水分含量的预测。

6.2 展望

本研究利用近红外光谱技术和机器视觉对绿茶不同加工工艺下的水分含量信息进行解析, 通过经典机器学习算法和化学计量学方法建立了不同绿茶加工过程的水分含量定量预测模型, 旨在提高名优绿茶品质, 对促进绿茶加工智能化、数字化的加工过程, 具有重要的理论价值和应用前景。结合本研究的工作, 笔者认为, 本研究仍有一些不足之处, 值得进一步研究和分析:

(1) 绿茶杀青、理条和干燥过程中水分含量的检测还无法应用于实际生产中, 还没有做水分检测与加工机械的集成, 后期研究重点可以放在研究更具有工业应用价值且能应用于实时监测的设备上, 实现绿茶整个加工过程中水分含量实时监控。

(2) 本研究采用的茶叶品种为高桥银峰茶和诸叶齐茶, 没有分析其他茶叶品种下对于模型预测精度的影响, 后续的研究可以探究绿茶含水率检测模型迁移性能的研究, 通过迁移学习算法, 建立鲁棒性较强的统一化模型。

(3) 本研究主要采用经典的预处理方法、变量筛选方式和机器学习算法, 今后的研究方向利用近红外光谱技术结合深度学习算法, 探讨水分含量预测的可行性。

参考文献

- [1] 潘蓉, 赵学尽, 杜建斌, et al. 2021 年中国茶叶进出口贸易情况简析 [J]. 中国茶叶, 2022, 44(03): 25-30.
- [2] 段会杰. 茶文化在思政教育中的渗透路径分析 [J]. 福建茶叶, 2022, 44(03): 249-251.
- [3] FU Y-Q, WANG J-Q, CHEN J-X, et al. Effect of baking on the flavor stability of green tea beverages [J]. Food Chemistry, 2020, 331.
- [4] MURUGESH C S, MANOJ J B, HAWARE D J, et al. Influence of water quality on nutritional and sensory characteristics of green tea infusion [J]. Journal of Food Process Engineering, 2017, 40(5).
- [5] ALICJA K, TOMASZ K, CLAUDIA M, et al. Green Tea Catechins Induce Inhibition of PTP1B Phosphatase in Breast Cancer Cells with Potent Anti-Cancer Properties: In Vitro Assay, Molecular Docking, and Dynamics Studies [J]. Antioxidants (Basel, Switzerland), 2020, 9(12).
- [6] TAN H R, LAU H, LIU S Q, et al. Characterisation of key odourants in Japanese green tea using gas chromatography-olfactometry and gas chromatography-mass spectrometry [J]. Lwt-Food Science and Technology, 2019, 108: 221-232.
- [7] 黄镇雄. 应用光谱技术的茶叶含水率在线检测设备研发及其试验研究 [D]; 福建农林大学, 2020.
- [8] 虞昕磊. 鲜叶摊放方式对绿茶色、香、味品质成分代谢的影响研究 [D]; 华中农业大学, 2020.
- [9] YI L, XIAOLONG L, WEIJIAO L, et al. Detection of chlorpyrifos and carbendazim residues in the cabbage using visible/near-infrared spectroscopy combined with chemometrics [J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2021, 257.
- [10] ZHIMING G, OSEI B A, LIMEI Y, et al. Intelligent evaluation of taste constituents and polyphenols-to-amino acids ratio in matcha tea powder using near infrared spectroscopy [J]. Food Chemistry, 2021, 353.
- [11] 魏玉震. 基于光谱和光谱成像技术的茶叶含水率检测机理和方法研究 [D]; 浙江大学, 2019.
- [12] 陈应会, 陈浩. 手工制茶与茶文化发展探究 [J]. 山西农经, 2018, (14): 55.
- [13] 章恩宽, 徐彬, 王娟娟. 浅谈信阳毛尖茶传统手工炒制技艺 [J]. 中国茶叶, 2013, 35(07): 22-23.

- [14] 陈寿松, 占杨, 郑功宇, et al. 中国茶叶加工, 2013, 茶叶含水率常用测定方法及比较分析(03): 33-36.
- [15] 刘颖, 曹佳佳, 章浩伟, et al. 低场核磁共振技术快速检测鲜乳水分方法研究 [J]. 食品科学, 2014, 35(14): 93-96.
- [16] 李然, 陈珊珊, 俞捷, et al. 应用低场核磁共振技术测定茶叶含水量 [J]. 茶叶科学, 2010, 30(06): 440-4.
- [17] 永田忠博, 酒井慎介, 陈荣冰. 用核磁共振法分析茶树叶片中水分的状态 [J]. 茶叶科学简报, 1992, (01): 42-44.
- [18] YAN S, XIAOZHONG W, HANLEI X, et al. Quality evaluation of Keemun black tea by fusing data obtained from near-infrared reflectance spectroscopy and computer vision sensors [J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2021, 252.
- [19] LI C, ZONG B, GUO H, et al. Discrimination of white teas produced from fresh leaves with different maturity by near-infrared spectroscopy [J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2020, 227(C).
- [20] 王文霞, 马本学, 罗秀芝, et al. 近红外光谱结合变量优选和 GA-ELM 模型的干制哈密大枣水分含量研究 [J]. 光谱学与光谱分析, 2020, 40(02): 543-549.
- [21] DUTHEN S, LEVASSEUR-GARCIA C, KLEIBER D, et al. Using near-infrared spectroscopy to determine moisture content, gel strength, and viscosity of gelatin [J]. Food Hydrocolloids, 2021, 115.
- [22] LU B, WANG X, LIU N, et al. Feasibility of NIR spectroscopy detection of moisture content in coco- peat substrate based on the optimization characteristic variables [J]. Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy, 2020, 239.
- [23] 陈琳, 董春旺, 高明珠, et al. 基于近红外光谱的红茶干燥中含水率无损检测方法 [J]. 茶叶科学, 2016, 36(02): 184-190.
- [24] 白晓丽, 郭卫华, 孔俊豪, et al. 速溶普洱茶中水分、咖啡碱和茶多酚含量近红外光谱快速测定方法的建立 [J]. 食品工业科技, 2019, 40(01): 234-238+245.
- [25] 王巧华, 梅璐, 马美湖, et al. 利用机器视觉与近红外光谱技术的皮蛋无损检测与分级 [J]. 农业工程学报, 2019, 35(24): 314-321.
- [26] SONG Y, WANG X, XIE H, et al. Quality evaluation of Keemun black tea by fusing data obtained from near-infrared reflectance spectroscopy and computer vision sensors [J]. Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy, 2021, 252.
- [27] WANG Y, LI L, LIU Y, et al. Enhanced quality monitoring during black tea processing by the fusion of NIRS and computer vision [J]. Journal of Food Engineering, 2021, 304.
- [28] 董春旺, 梁高震, 安霆, et al. 红茶感官品质及成分近红外光谱快速检测模型建立 [J]. 农业工程学报, 2018, 34(24): 306-313.

- [29] 田世杰. 基于透射光谱的苹果霉心病判别影响因素及其修正方法研究 [D]; 西北农林科技大学, 2020.
- [30] YANG C, ZHAO Y, AN T, et al. Quantitative prediction and visualization of key physical and chemical components in black tea fermentation using hyperspectral imaging [J]. *Lwt-Food Science and Technology*, 2021, 141.
- [31] 刘云宏, 王庆庆, 石晓微, et al. 金银花贮藏过程中绿原酸含量的高光谱无损检测模型研究 [J]. *农业工程学报*, 2019, 35(13): 291-299.
- [32] AN T, YU H, YANG C S, et al. Black tea withering moisture detection method based on convolution neural network confidence [J]. *Journal of Food Process Engineering*, 2020, 43(7).
- [33] PANG L, WANG J H, MEN S, et al. Hyperspectral imaging coupled with multivariate methods for seed vitality estimation and forecast for *Quercus variabilis* [J]. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy*, 2021, 245.
- [34] XIE S G, LI Y H, WANG X, et al. Research on estimation models of the spectral characteristics of soil organic matter based on the soil particle size [J]. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy*, 2021, 260.
- [35] GUANGXIN R, YING L, JINGMING N, et al. Assessing black tea quality based on visible–near infrared spectra and kernel-based methods [J]. *Journal of Food Composition and Analysis*, 2021, 98(publish).
- [36] JIANG H, XU W D, DING Y H, et al. Quantitative analysis of yeast fermentation process using Raman spectroscopy: Comparison of CARS and VCPA for variable selection [J]. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy*, 2020, 228.
- [37] YUN Y H, WANG W T, TAN M L, et al. A strategy that iteratively retains informative variables for selecting optimal variable subset in multivariate calibration [J]. *Anal Chim Acta*, 2014, 807: 36-43.
- [38] 黄俊仕, 王冬欣, 熊爱华, et al. 基于混合变量选择的绿茶酚氨比近红外光谱检测方法 [J]. *江西农业大学学报*, 2020, 42(06): 1270-1276.
- [39] WANG Y J, JIN S S, LI M H, et al. Onsite nutritional diagnosis of tea plants using micro near-infrared spectrometer coupled with chemometrics [J]. *Computers and Electronics in Agriculture*, 2020, 175.
- [40] 尹军峰, 闵航, 许勇泉, et al. 摊放环境对名优绿茶鲜叶茶多酚及儿茶素组成的影响 [J]. *茶叶科学*, 2008, (01): 22-27.
- [41] BOKUCHAVA M A, SKOBELEVA N I. The biochemistry and technology of tea manufacture [J]. *Critical reviews in food science and nutrition*, 1980, 12(4): 303-70.

- [42] DONG C, YE Y, YANG C, et al. Rapid detection of catechins during black tea fermentation based on electrical properties and chemometrics [J]. Food Bioscience, 2021, 40.
- [43] 张娟, 原帅, 张骏. 基于小波变换-遗传算法-偏最小二乘的草莓糖度检测研究 [J]. 分析科学学报, 2020, 36(01): 111-115.
- [44] 董春旺, 朱宏凯, 周小芬, et al. 基于机器视觉和工艺参数的针芽形绿茶外形品质评价 [J]. 农业机械学报, 2017, 48(09): 38-45.
- [45] DONG C-W, ZHU H-K, ZHAO J-W, et al. 基于机器视觉和非线性的芽形绿茶外形感官品质评价 (英文) [J]. Journal of Zhejiang University-Science B(Biomedicine & Biotechnology), 2017, 18(06): 544-548.
- [46] WANG Y-J, JIN G, LI L-Q, et al. NIR hyperspectral imaging coupled with chemometrics for nondestructive assessment of phosphorus and potassium contents in tea leaves [J]. Infrared Physics & Technology, 2020, 108.
- [47] 孙俊, 金夏明, 毛罕平, et al. 高光谱图像技术在掺假大米检测中的应用 [J]. 农业工程学报, 2014, 30(21): 301-307.
- [48] 穆炳宇, 张淑娟, 李泽珍, et al. 基于近红外光谱特征提取的花椰菜灰霉病早期检测 [J]. 光谱学与光谱分析, 2021, 41(08): 2543-2548.
- [49] 李春霖. 基于化学计量学和近红外光谱技术的龙井茶感官及化学品质评价研究 [D]; 浙江大学, 2019.
- [50] 邱瑞瑾, 王登良, 徐文贤, et al. 绿茶色泽的形成机理及调控 [J]. 广东茶业, 2007, (05): 14-17.
- [51] 刘宗岸, 毛志方, 施海根, et al. 做青过程中叶绿素的变化研究 [J]. 中国茶叶加工, 2007, (03): 21-23.
- [52] 王华杰, 滑金杰, 余勤艳, et al. 基于 IRAE-HS-SPME/GC-MS 分析杀青方式对绿茶栗香形成的影响 [J]. 食品科学, 2021, 42(14): 209-217.
- [53] 程丽娟, 刘贵珊, 何建国, et al. 基于高光谱开发滩羊肉中高铁肌红蛋白含量的定量函数 [J]. 光谱学与光谱分析, 2020, 40(04): 1263-1269.
- [54] WANG J J, ZAREEF M, HE P H, et al. Evaluation of matcha tea quality index using portable NIR spectroscopy coupled with chemometric algorithms [J]. J Sci Food Agr, 2019, 99(11): 5019-5027.
- [55] DONG C W, ZHU H K, WANG J J, et al. Prediction of black tea fermentation quality indices using NIRS and nonlinear tools [J]. Food Science and Biotechnology, 2017, 26(4): 853-860.
- [56] YUAN R R, LIU G S, HE J G, et al. Classification of Lingwu long jujube internal bruise over time based on visible near-infrared hyperspectral imaging combined with partial least squares-discriminant analysis [J]. Computers and Electronics in Agriculture, 2021, 182.

- [57] ZHANG L, JI H Y. Identification of wheat grain in different states based on hyperspectral imaging technology [J]. Spectrosc Lett, 2019, 52(6): 356-366.
- [58] YU Z Y, FANG H, ZHANGJIN Q N, et al. Hyperspectral imaging technology combined with deep learning for hybrid okra seed identification [J]. Biosystems Engineering, 2021, 212: 46-61.
- [59] CEBI N, YILMAZ M T, SAGDIC O. A rapid ATR-FTIR spectroscopic method for detection of sibutramine adulteration in tea and coffee based on hierarchical cluster and principal component analyses [J]. Food Chemistry, 2017, 229.
- [60] WANG Y, LI L, LIU Y, et al. Enhanced quality monitoring during black tea processing by the fusion of NIRS and computer vision [J]. Journal of Food Engineering, 2021, 304: 110599.
- [61] LUO X, XU L J, HUANG P, et al. Nondestructive Testing Model of Tea Polyphenols Based on Hyperspectral Technology Combined with Chemometric Methods [J]. Agriculture-Basel, 2021, 11(7).

致 谢

时光飞逝，三年的硕士求学生涯即将结束。值此毕业论文完成之际，我要向三年来给予我关心、支持和帮助的亲人、老师和同学们表示最诚挚的感谢。

首先，我要感谢石河子大学和中国农业科学院茶叶研究所为我提供了良好的科研平台。其次我要感谢指导我论文的导师们。论文的选题、实验方案设计、数据分析与处理，甚至是论文的撰写都离不开罗昕老师、胡斌老师和董春旺老师的精心指导，每一个环节都倾注了导师的心血。罗昕老师治学严谨，学识渊博，对待学生和藹亲近，在我考研失败时，是罗昕老师愿意将我招收为课题组的一员，让我非常感激；胡斌老师学识渊博，为我营造了一种良好的精神氛围，您的博学让我崇敬，您的治学态度让我钦佩；董春旺老师有着广阔的视野和丰富的创新思维，在杭州中国农业科学院茶叶研究所进行实验过程中，给予我很多帮助，为我提供了良好的实验平台、学习氛围和生活环境，对于我论文的试验、撰写和修改都倾注了很大的心血，在我申博期间董春旺老师给予我的帮助，我铭记于心。

感谢石河子大学的梁高震师哥和何浩猛师哥在我研究生入学之前对我无私的帮助，感谢安霆师哥、陈永师哥、靳奉奎师哥、付磊师哥、郭孟宇师哥、潘峰师哥、毛自斌师哥和杨崇山师哥，感谢蔡会、刘策、薛星星、周小周同学，感谢张人天、陈智磊、蔡一全、解颜宇、王建和徐洛川师弟的陪伴，感谢我的室友郑雄飞、王国梁、譙超凡在生活上的帮助；感谢中茶所的汪洁琼师姐、薛金金师姐、王华杰师姐在我科研和生活上的指导与关心，祝愿你们博士顺利毕业；感谢沈帅老师、陈佳瑜、吴仕敏和陈丹同学闲暇之余的聚餐活动，为我排解了很多烦恼；感谢姜嘉胤师弟、程亦凡师弟、盛旭峰师弟、余勤艳师妹，许继业师妹、王哲师妹、刘钰懿师妹和徐丽丽师妹对我的科研和生活上的帮助。

感谢我的父母和姐姐，感谢你们一直以来对我的包容与无私奉献，感谢你们在我生活遇到困难时给与我的支持和关怀。感谢曹金翠同学三年来对我的包容和忍让，在我遇到挫折和困难时的不离不弃。

最后，我想再次感谢所有关心、鼓励和帮助我的老师、同学、朋友和家人，祝你们身体健康，万事胜意！

刘中原

2022年6月

于石河子

作者简介

刘中原，男，生于1995年6月，籍贯江苏省徐州市人。2019年毕业于北京理工大学珠海学院工业自动化学院机械工程专业，获工学学士学位。同年9月起在石河子大学机械电气工程学院机械工程专业就读硕士。

在学期间主要参与的研究项目

1. 国家重点研发计划项目：茶叶精制智能化装备研发（2018YFG0700500）。
2. 国家自然科学基金：小叶种工夫红茶发酵品质信息多维时空分布感知研究（31972466）。
3. 中央级科研基本业务专项：名优红茶标准化精准加工关键技术装备及集成示范（1610212016018）。
4. 中央级科研基本业务专项：茶叶数字化、智能化生产管控关键技术研究及集成示范（1610212021004）。

在研期间撰写的论文

1. **Zhongyuan Liu**, Chongshan Yang, Xin Luo, Bin Hu*, Chunwang Dong *. Research on the online rapid sensing method of moisture content in famous green tea spreading [J]. Journal of Food Process Engineering, doi: 10.1111/jfpe.13853.(中科院三区, **SCI, IF=2.356**)
2. Chunwang Dong¹, **Zhongyuan Liu**¹, Chongshan Yang, Ting An, Bin Hu, Jing Jin*, Yang Li*. Rapid detection of exogenous sucrose in black tea samples based on near-infrared spectroscopy [J]. Infrared Physics & Technology, doi:10.1016/j.infrared.2021.103934. (中科院二区, **SCI, IF=2.638**)
3. **Zhongyuan Liu**, Rentian Zhang, Chongshan Yang, Bin Hu, Xin Luo, Yang Li*, Dong Chunwang*. Research on moisture content detection method during green tea processing based on machine vision and near infrared spectroscopy technology[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, doi:10.1016/J.SAA.2022.120921 (中科院二区, **SCI, IF=4.098**)
4. 董春旺, 刘中原, 杨明, 王梅, 李杨, 张人天, 林智*. 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, doi: 10.7506/spkx1002-6630-20211213-151 (**EI, IF=2.187**)

5. Jiangming Jia, Xiaofen Zhou, Yang Li, Mei Wang, **Zhongyuan Liu*** and Chunwang Dong*. Establishment of a rapid detection model for the sensory quality and components of Yuezhou Longjing tea using near-infrared spectroscopy[J]. LWT - Food Science and Technology (Accept, 中科院一区, SCI, IF=4.952)

在研期间参与的论文

1. Dong Chunwang, Yang Chongshan, **Liu Zhongyuan**, Zhang Rentian ,Yan Peng, An Ting, Zhao Yan*, Li Yang*. Nondestructive Testing and Visualization of Catechin Content in Black Tea Fermentation Using Hyperspectral [J]. Sensors, doi:10.3390/S21238051 (中科院三区, SCI, IF=3.576)

2. Yang Chongshan, Zhao Yan, An Ting, **Liu Zhongyuan**, Jiang Yongwen, Li Yaqi*, Dong Chunwang*. Quantitative prediction and visualization of key physical and chemical components in black tea fermentation using hyperspectral imaging [J]. LWT - Food Science and Technology, doi:10.1016/J.LWT.2021.110975 (中科院一区, SCI, IF=4.952)

3. 王盛琳, 杨崇山, **刘中原**, 柳善建*, 董春旺*. 基于电特性的红茶发酵中茶多酚含量快速检测方法[J]. 茶叶科学, doi:10.13305/j.cnki.jts.2021.02.007 (中文核心, IF=1.782)

在研期间申报的专利

1. 罗昕, **刘中原**, 郭孟宇, 等. 一种筛筒式茶叶切割器[P].申请公布号: CN111587659A (发明专利, 实审, 第二作者主笔)。

获奖情况

1. 机收膜杆混合物剪切破结装置, 全国三维数字创新设计大赛龙鼎奖--新疆赛区一等奖;
2. 立筒式自动防堵泵前过滤器, “中联重工”杯第五届全国大学生智能农业装备创新大赛优秀奖;
3. 基于感应熔覆工艺的防腐耐磨管道制造装备, 石河子大学第六届“互联网+”大学生创业大赛三等奖;
4. 获得石河子大学研究生一等学业奖学金;

石河子大学硕士研究生学位论文 导师评阅表

研究生姓名	刘中原	学制	三年
专业	机械工程	研究方向	机械工程

学术评语:

论文针对绿茶加工数字化和智能化加工水平偏低, 主要依靠制茶师傅经验判断水分含量, 从而确定加工工艺, 受主观因素影响较大, 水分含量无法实现精准、实时检测等问题, 基于近红外光谱技术, 同时结合机器视觉技术和化学计量学等方法, 以不同加工工序下绿茶在制品为研究对象, 使用机器学习算法, 构建了绿茶不同加工工序下的水分含量定量预测模型和定性判别模型, 实现绿茶水分含量有效、快速评估。论文选题、基本论点和结论具有一定的理论意义或实用价值, 论文的研究背景和思路是否清晰。论文对课题所涉及的国内外学术动态、重要文献资料有较全面地了解和评述, 研究方法或设计方案恰当, 研究步骤和过程科学规范, 研究内容的有一定的难度和深度, 工作量饱满。

论文研究目标明确, 方案及方法合理可行, 论述层次清晰, 撰写认真, 表达清楚, 图表运用规范。

论文和答辩表明作者已掌握了机械工程专业的基础理论和专业知识, 具备了从事科学研究或独立承担专门技术工作的能力, 达到了硕士论文的要求。

同意参加硕士论文答辩。

指导教师签字:

罗昕

2022年6月14日