

机器学习在花果茶生产加工中的应用进展

庞文媛¹, 孙意岚², 王芹³, 陈杰博⁴, 沈朝增⁵, 周一鸣⁶, 庞杰^{1*}

(1. 福建农林大学食品科学学院, 福州 350002; 2. 福建农林大学生命科学学院, 福州 350002; 3. 沧州市物资储备和粮油质检中心, 沧州 061000; 4. 福建农林大学国家甘蔗工程技术研究中心, 福州 350002; 5. 宁德茗鼎茶业有限公司, 宁德 355300; 6. 上海应用技术大学香料香精技术与工程学院, 上海 201418)

摘要: 花果茶因其独特的口味和丰富的营养成分成为备受欢迎的新型饮品。然而, 在花果茶生产加工过程中, 仍存在原料溯源和品质控制方面的不足、质量安全控制标准化水平较低及加工工艺智能化不足等问题。传统分析手段对于花果茶有关数据的非线性信息挖掘和原料、成品传统分级方法的效率也较低。目前, 我国的花果茶产业正处于由传统加工向智能化转型的重要阶段, 机器学习在其中发挥着不可替代的重要作用。机器学习算法因其自主进行特征学习、强大的非线性拟合能力、端到端建模以及快速的特点, 成为花果茶加工中研究应用的热点。本文综述了机器学习算法在花果茶原料验收、原料加工、质量分级等方面的应用, 总结了机器学习算法在花果茶生产加工不同环节中的侧重点、优缺点和未来发展方向, 为花果茶加工的智能化发展提供了参考。

关键词: 花果茶; 智能化; 机器学习; 花果茶加工

Application progress of machine learning in the production and processing of flower and fruit tea

PANG Wen-Yuan¹, SUN Yi-Lan², WANG Qin³, CHEN Jie-Bo⁴,
SHEN Chao-Zeng⁵, ZHOU Yi-Ming⁶, PANG Jie^{1*}

(1. College of Food Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China; 2. College of Life Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China; 3. Cangzhou Strategic Reserves and Grain Oils Quality Inspection Center, Cangzhou 061000, China; 4. National Engineering Research Center of Sugarcane, Fujian Agriculture and Forestry University, Fuzhou 350002, China; 5. Ningde Mingding Tea Industry Co., Ltd., Ningde 355300, China; 6. School of Perfume and Aroma Technology, Shanghai Institute of Technology, Shanghai 201418, China)

ABSTRACT: Flower and fruit tea has become a popular new beverage due to its unique flavor and rich nutritional content. However, in the production and processing of flower and fruit tea, there are issues such as insufficient traceability and quality control for raw materials, low standardization of quality and safety control, and inadequate intelligent processing technology. Traditional analytical methods also have low efficiency in nonlinear information mining related to flower and fruit tea data and traditional grading methods for raw materials and finished products. Currently, China's flower and fruit tea industry is in an important stage of transformation from traditional processing to intelligent processing, in which machine learning plays an irreplaceable role. Due to its ability to autonomously

基金项目: 国家重点研发计划项目(2022YFD2101102)、福建省自然科学基金项目(2022J02021、2020J01132)

Fund: Supported by the National Key Research and Development Program of China (2022YFD2101102), and the Natural Science Foundation of Fujian Province (2022J02021, 2020J01132)

***通信作者:** 庞杰, 博士, 教授, 主要研究方向为天然植物多糖的结构与性能。E-mail: pang3721941@163.com

***Corresponding author:** PANG Jie, Ph.D, Professor, Fujian Agriculture and Forestry University, No.15, Shangxiadian Road, Cangshan District, Fuzhou 350002, China. E-mail: pang3721941@163.com

learn features, strong nonlinear fitting capability, end-to-end modeling, and fast computation, machine learning algorithms have become a research hotspot in flower and fruit tea processing. This article reviewed the applications of machine learning algorithms in flower and fruit tea raw material inspection, raw material processing, and quality grading, summarizes the focus, advantages, and limitations of machine learning algorithms in different stages of flower and fruit tea production and processing, so as to provide a reference for the intelligent development of flower and fruit tea processing.

KEY WORDS: flower and fruit tea; intelligence; machine learning; flower and fruit tea processing

0 引言

随着消费者对茶产品认识的提高和对茶产品需求的逐渐增多,花果茶作为一种新型茶产品,其市场份额也日益增加^[1]。花果茶是一种混合了花、干果和茶叶的茶类产品^[2],市场将这三类茶统称为花果茶。其中,花茶又分为窈制花茶和调香花茶,窈制花茶包括茉莉花茶、桂花茶等,调香花茶则包括玫瑰花茶、果味花茶等^[3];果茶分为果汁饮料以及由干果和茶叶拼配而成的复合果茶。

花果茶的生产加工包括原料验收、原料加工和质量检测 3 大过程。其中,花茶加工步骤包括原料验收、预处理、窈制、通花、起花、复火窈制、提花和包装;果茶加工步骤包括原料验收、预处理、干制、调配和包装。加工环节对成品茶品质影响最大,传统的花茶加工方式一般为窈制,窈制可分为传统窈制、增湿连窈和隔离窈制^[4]。在花茶的窈制过程中,花茶品质易受窈制时间、窈制温度、配花量的影响^[5]。果茶的原料加工方式包括直接打浆或干制加工(图 1),干制方式有自然晒干、真空冷冻干燥、热风干燥、远红外干燥等^[6],果茶的品质一般通过改变干燥温度^[7]和干燥时间^[8]来进行调控。传统分级方法通过感官审评和品质检验判定花茶和果茶品质优劣。然而,感官审评容易受个人嗜好的制约,品质检验存在检测时间长、操作烦琐、效率低等缺点^[9]。



图 1 果茶加工方式

Fig.1 Processing method of fruit tea

此外,我国暂无明确、统一的花果茶国家加工标准和果茶加工标准。花茶中,仅茉莉花茶制定了相应的国家生产加工标准,即 GB/T 34779—2017《茉莉花茶加工技术规范》,加工标准的欠缺限制了花果茶的规模化和标准化生

产。由于花果茶在原料验收和加工过程中受多种因素影响,同一批次的花果茶品质差异较大,难以准确地进行质量等级的划分,甚至可能出现以次充好的情况。因此,发展智能化的花果茶加工技术并制定更加完善的花果茶分级标准是花果茶研究领域亟待解决的问题。

近年来,随着人工智能产业的发展,食品规模化和标准化的生产问题得到较好的解决^[10]。机器学习是一种可以针对高维数据进行快速处理并不断学习优化自身算力的一种计算机学科,现已广泛应用于医疗诊断、电子信息技术、材料、能源和食品等领域。在食品领域中,机器学习已被广泛应用于包括花果茶在内的茶叶原料分级识别、加工工艺优化、香气滋味识别判定、食品安全质量检验和质量分级等方面^[11]。其中,K-近邻算法(K-nearest neighbor, KNN)、支持向量机(support vector machine, SVM)、随机森林(random forest, RF)、朴素贝叶斯(naive bayes, NB)、卷积神经网络(convolutional neural network, CNN)等算法已被用于茶叶等原材料分级识别中^[12-14],使用相关分类算法能够进行原料溯源,实现原料采收的调整和补救工作^[15-16];在对食品加工工艺进行优化时,人工神经网络(artificial neural network, ANN)等算法可以针对各个花果茶加工工序进行过程动力学预测,优化工艺参数,确保成品品质^[17-19];在辅助花果茶香气滋味的识别判定时,机器学习通过无监督算法的应用能够自动识别出成品茶香气品质等级,为检测挥发物的不良特性提供科学依据,并且可以有效避免感官审评中主观因素的影响^[20-21],但在处理高噪声、小样本的茶叶代谢组学数据时,模型可能会出现过拟合,需要进行输入数据集筛选和模型调参以保证结果的准确性^[22-23];在花果茶等食品安全质量检验方面,机器学习可以根据食品的种类、生产年份、掺杂物、掺假点和检测点来判定食品是否为假冒伪劣产品^[24-25]。在产品质量分级时,机器学习有效避免了人工分级和化学检验分级的缺陷,使产品分级变得更加准确,也为食品标准化生产和规范化提供了技术保障。

目前,机器学习广泛应用于花果茶行业的原料识别分类、加工参数控制和质量等级分类等不同领域。当前尚未有完整的综述研究从原料验收到成品质量等级分类的应用进展。因此,本文旨在总结机器学习在花果茶行业从原料验收到质量等级分类的应用现状,分析其优缺点,为智

能化加工生产标准的制定提供参考。

1 机器学习算法在花果茶生产中的应用

不同的机器学习算法具有各自的优缺点(表 1), 根据研究问题和数据类型, 机器学习算法主要分为有监督学习算法、无监督学习算法和强化学习算法^[37]。有监督学习算法根据训练集标签类型的不同, 一般分为回归和分类问题, 常见的监督学习算法包括 NB、SVM、决策树(decision tree, DT)、线性回归、逻辑回归和深度学习算法等, 其中深度学习算法包括了 ANN、CNN 等不同类型^[38]。与监督学习算法不同, 无监督学习算法的训练集样本没有标签, 一般不能达到量化的效果。无监督学习算法可以分为聚类、降维和关联分析 3 种, 常见的无监督学习算法为 K-均值(K-means clustering algorithm, K-means)、PCA、独立成分分析(independent component analysis, ICA)等^[39]。强化学习算法涵盖更多离散型和随机型数学知识, 包括时间差分方法、蒙特卡洛法、动态规划、值函数逼近法等^[40]。机器学习算法具有快速、无损、可靠等特点, 在花果茶产业中的应用不断扩大。后文将分别阐述机器学习算法在花果茶生产中的原材料验收、加工和成分分析等不同阶段的应用进展。

1.1 花果茶原料验收

花果茶的原材料品质对产品质量具有直接影响, 因此原材料验收非常重要。花果茶原料验收包括重金属、农药残留检测和病虫害防治, 以及原料分级等 3 个方面。

在原料病虫害防治中, 陶国柱^[41]利用 CNN 对茉莉花病虫害进行图像识别, 将人工收集和检索的茉莉花病虫害图像进行预处理后, 利用 CNN 的 4 种网络模型训练并

对比, 通过 Momentum 和 RMSProp 算法对 MoblieNetV2 网络进行优化改进, 优化后的网络收敛更快, 识别准确率达到 94.34%。通过深度学习和图像识别算法能够方便快捷的识别出茉莉花病虫害, 但该研究未针对茉莉花不同组织的病虫害进行分析, 也未与其他机器学习算法进行比对。

在原料识别分级上, 机器学习结合计算机视觉系统进行分类, 能够获取和处理视觉信息, 并在不用人工干预的情况下做出决策, 已在大规模产品分级和质量检测中推广^[42]; LIU 等^[43]利用深度神经网络结合计算机视觉系统对 7 种商业菊花茶进行分类, 对开花期的菊花和成品茶的区分正确率为 96%和 89%, 但由于所用的菊花茶样本量较小, 而利用深度神经网络处理时所需要的样本数据量较大时效果更好, 因此该分类结果可能存在误差, 需要二次验证。该方法同样能够进行茶叶及其他相关花果茶原料的自动挑选和分类。CHEN 等^[44]使用表面解吸化学电离质谱法区分茶产品, 在质谱指纹图谱的基础上, 利用绿茶、乌龙茶和茉莉花茶的质谱原始数据进行 PCA 分析, 能够准确区分 3 种茶叶; 基于表面解吸化学电离质谱指纹的 PCA 分类结果能够区分感官方法无法区分的 4 种绿茶, 实验结果较人工分类更加精确, 未来能够在花果茶分类中进行应用。在对水果等花果茶原料的品质分类上, 赵凡^[45]利用 CNN 算法分别对苹果和梨进行分级识别, 正确率为 94.49%, 在保证准确率的同时, 有效解决了传统机械分级机对水果外观品质分级时可能存在的精度低和速度慢等问题, 极大节约人力物力。

机器学习尚未应用在花果茶原料的农药和重金属残留的识别和预测上。未来的研究可以收集花果茶原材料农药和重金属残留数据, 用于机器学习算法建模, 提高花果茶原料产量。

表 1 不同机器学习算法的优缺点
Table 1 Advantages and disadvantages of different machine learning algorithms

算法	优点	缺点
CNN	输入图像和网络的拓扑结构吻合性好; 特征提取和分类同时进行, 适应性更强 ^[26-27] 。	较深层的卷积网络提取的更偏重局部信息, 并且深层卷积的计算量需求大, 输入图片大小固定, 在嵌入式设备应用方面有局限性 ^[28] 。
ANN	能够处理海量数据、协调多种非线性因素以及提高输出速度 ^[29] 。	网络层数较多时, 容易陷入局部最优解, 也容易产生过拟合 ^[30] 。
NB	分类准确率高, 速度快, 训练和查询大量数据时效率高; 可以处理小规模的数据 ^[31] 。	对输入数据的形式要求高; 在模型中进行分类决策时有错差率 ^[32] 。
PCA ^[33]	使得数据集更易使用; 降低算法的计算开销; 去除噪声; 使得结果容易理解; 完全无参数限制。	特征值分解存在局限性, 变换的矩阵必须是方阵; 在非高斯分布情况下, PCA 得出的主元可能非最优解。
RF	特征越多、能容忍高数据的噪音, 具有高预测精度, 不容易过拟合 ^[34] 。	对高维度特征筛选和选取效率较低、对动态数据聚类的泛化误差估值较大 ^[35] 。
SVM ^[36]	可解决高维度和局部极值问题, 局部最优解也是全局最优解。	对孤立点和噪声点敏感、运算成本大。

注: 主成分分析(principal component analysis, PCA)。

1.2 花果茶加工

花果茶的加工类型包括花茶加工、果茶果干加工及茶类加工。由表 2 可知,大多数果蔬干燥效果评估通常结合 ANN 算法进行。SAGLAM 等^[50]利用 5 种干燥方式对 3 种不同品种的苹果进行干燥,并通过多种机器学习算法(ANN、KNN、RF、高斯过程和 SVM)估计水分比、水分含量和干燥速率,研究结果表明,RF 评估苹果干水分比的相关系数(R)最高,RF、SVM 和 KNN 在水分含量和干燥率的评估上也有较高的 R 值。除了可以有效评估花果茶的干燥特性之外,机器学习算法还可以建立干燥模型,ZHANG 等^[51]通过高湿空气冲击和热空气辅助红外加热技术,对蓝莓采后热烫干燥工艺进行优化,研究人员以热烫时间、红外加热温度和空气流速为输入层,干燥时间、比能耗、抗坏血酸含量和再水化能力为输出层构建 ANN 模型,并利用非支配排序遗传算法优化蓝莓漂烫/干燥过程的操作条件,实现干燥时间和比能耗最小化,确定蓝莓抗坏血酸含量最大化和再水化参数。其中干燥时间、比能耗、抗坏血酸含量和再水化能力的实验数据与预测值的相对误差分别为 1.43%、2.06%、2.07%和 3.08%。由于该研究漂烫干燥的工艺复杂性,不同操作温度下的传热、传湿效率、蓝莓微观结构以及能耗的影响无法清晰量化,然而由于 ANN 具有协调多种非线性因素的特点能够进行数据拟合,因此蓝莓干燥时间、比能耗、抗坏血酸含量和再水化能力等数值被有效评估,这为花果茶干燥时间和抗坏血酸含量的智能调控提供了借鉴方法。

在花茶加工方面,茉莉花茶通常是通过窈制的方式进行加工,窈制过程的每个工序都需要不同参数进行调控^[52]。在窈制茉莉花茶的过程中,通常会依靠制茶师的经验来设置窈制参数,这使得加工过程相对主观。机器学习能够与质谱、光谱等技术结合,智能地筛选和设置花果茶加工过程中不同阶段的温度、湿度参数,以此来找到最佳生产条件。例如,茉莉花茶在不同层次的窈制过程中,所需的窈制温度不同,窈制次数越高,所需温度越低,香气越浓郁、鲜活^[53];枇杷花茶在真空冷冻干燥和微波干燥时,微波干燥

时间越长,枇杷花茶中的黄酮和总三萜含量越低,色泽、香味和形态也会受到破坏,从而降低品质^[54]。因此,机器学习可以参与花果茶不同加工方法制备出的成品茶数据分析,通过降维、最优参数筛选等步骤来优化加工方法,提高花果茶品质。尽管目前相关研究较少,但可以借鉴表 2 中果干原料常用的机器学习算法,例如在花茶干燥过程中,构建 ANN 干燥动力学模型预测花茶中的水分含量,并根据水分含量控制窈制温度和窈制时间。此外,还可以构建不同加工参数时花茶主要氧化活性物质的 ANN 模型,根据所需氧化活性物质含量制定花茶窈制最适加工参数。因此,机器学习技术能够构建花果茶的干燥动力学模型和主要氧化活性物质模型,预测花果茶干燥时的水分含量。未来研究可以进一步拓展其在花果茶加工过程中的影响因素筛选、智能参数设定,以及花果茶加工过程中数字化、智能化控制的研究。

茶叶是花果茶制作过程中不可或缺的原料之一。在茶叶制作第一步骤中,大部分茶类产品采用萎凋方法进行加工,包括冷冻萎凋、超高压萎凋、光源萎凋、做青萎凋等方法^[55],其最终目的均为控制酶活性,确保茶叶品质。然而,萎凋过程中酶活性的控制精确度难以提高,需要结合其他技术来辅助控制。使用机器学习算法可以找到茶叶最佳的萎凋条件,筛选出茶叶的最大酶活性和恰当的水分含量。经过预处理后的新鲜茶叶,通过自然干燥或者人工干燥使其水分活度降低,随着水分的减少,茶叶内会发生不同物理化学变化。其中,多酚氧化酶(polyphenoloxidase, PPO)和过氧化物酶(peroxidase, PO)的活性会发生显著改变。在有氧条件下,PPO 和 PO 催化氧化其他酚类化合物形成各种氧化多酚化合物,这些化合物对茶的香气滋味和色泽产生一定的影响。为最大化实现 PPO 和 PO 的活性,SHRILEKHA 等^[56]利用 ANN 对红茶在萎凋过程中的酶活性和水分含量变化进行拟合,并通过遗传算法进行优化^[57],结合两种算法成功筛选出生产红茶的最佳萎凋参数。这种利用酶活性最大化优化茶叶生产的方法为茶叶质量控制提供了新思路。

表 2 机器学习算法在不同原料加工方式中的应用比较

Table 2 Comparison of the application of machine learning algorithms in different raw material processing methods

果干原料	机器学习算法	干燥方式	用途	特点
柚子果皮 ^[46]	ANN	冷冻干燥、微波干燥、强制对流	预测干燥系数、样品品质、水分含量、水分比率	快速简单的预测指定参数
西红柿 ^[47]	ANN	烘箱干燥	构建植物化学成分氧化活性模型	ANN 可以根据植物化学参数的数据对抗氧化活性进行建模,且建模准确率高
脐橙 ^[48]	ANN、RF、SVM、KNN、高斯过程	空气对流干燥	预测干燥水分和干燥速率	多算法结合,模型准确率高,最佳干燥条件准确率高。
柿子 ^[49]	ANN、SVM、KNN	真空干燥、热风干燥	创建柿子干燥动力学模型,预测柿子片水分含量	在建立柿子的干燥模型中,人工神经网络非常接近理论数学模型中的最高模型。

部分花果茶的茶基需要进行发酵加工, 茶叶的发酵可以影响茶的汤色、口感、香气、滋味和外形^[58], 是影响成品茶品质的重要因素之一。目前, 茶叶发酵程度的判别方法包括感官法、电化学传感技术法、成分传感法、叶温法、电子鼻技术、发酵液水溶液分光光度法等^[59]。这些方法快速、简单, 但茶叶感官法、叶温法等会存在主观认知差异, 精确度有限。引入机器学习算法辅助分析可以有效避免个体认知差异, 从海量数据中学习并获取新的模型, 不断优化模型性能以提高结果的准确性。基于儿茶素和茶色素含量与茶叶发酵程度的关系, ZHU 等^[60]利用高效液相色谱(high performance liquid chromatography, HPLC)和紫外-可见光分光光度计(ultraviolet-visible spectrophotometer, UV-Vis)测定儿茶素和茶色素含量数据, 采用 LCR 仪识别茶叶在发酵过程中的电学特性数据, 通过 PCA 和层次聚类对茶叶发酵程度进行分类, 并利用多层感知器神经网络、SVM 和 RF 构建茶叶发酵度判别模型。结果显示, 儿茶素和茶色素含量与电性能存在高度相关性, 该模型的测试集平均准确率分别为 88.9%、100%和 78.92%。由此可见, 机器学习算法可为区分发酵茶叶样品提供客观支撑, 所建立的模型不仅相对稳定, 而且具有较强的可重复性。

综上所述, 对于花果茶不同干燥阶段湿度的判定和发酵程度的控制, RF 算法表现出了较好的分类效果和高预测精度; 在花果茶干燥阶段, ANN 算法可协调多种非线性因素, 具有快速、准确的分类建模效果。然而, 目前花茶的窖制步骤尚未建立完整的自动化流程, 同时在花果茶拼配方面缺乏机器学习算法参与的智能化花果茶配伍研究, 需要后续研究完善。

1.3 花果茶成品质量分级

1.3.1 花果茶的成分分析

花果茶中含有多种内源性代谢物, 对其化合物成分进行检测, 能够分析花果茶的口感、气味、营养价值并进行分级加工。目前, 检测花果茶成分的方法有电子鼻、固相微萃取-气相色谱-质谱法(solid phase microextraction-gas chromatography-mass spectrometry, SPME-GC-MS)、气相色谱-质谱法(gas chromatography-mass spectrometry, GC-MS)等。这些技术虽然在一定程度上可实现高效检测, 但并不能直接判断花果茶品质。机器学习算法可以辅助分析花果茶中的挥发性成分和非挥发性成分, 确定生物标志物和特征代谢物, 辅助完成花果茶质量检测中的等级鉴定。

结合 GC-MS 和 PCA 技术可以用于分析鉴定花果茶中的挥发性成分。PCA 可以消除 GC-MS 冗余数据, 并消除 GC-MS 原始数据之间的相互影响, 找到数据中的重要特征, 分析出挥发性物质的主要成分。LUO 等^[61]采用气相色谱-嗅觉-质谱法、电子鼻结合 PCA 对不同产地菊花茶的挥发性代谢物进行聚类, 确定菊花茶样品差异的关键挥发性(香气活性)成分, 并利用挥发性成分的差异评估菊花茶质

量。该组合技术在质量评估和产地识别上较为高效, 取代了人工在感官上未能辨别出样品挥发性成分的细微差异的缺陷, 虽然该实验处理数据规模较小, 但 PCA 技术可以处理海量数据, 且不需要进行降噪处理就能分析出主要成分, 为区分大量花果茶种类及花果茶智能化发展提供了技术保障。

在利用茶叶香气成分进行等级区分时, BORAH 等^[62]通过电子鼻测定不同等级茶叶挥发性成分的数据, 利用 PCA 对不同类别的香气特征进行可视化, K-means 和自组织映射(self-organizing maps, SOM)将数据分为 8 个聚类。接着, 通过多层感知器网络、径向基函数网络和构造概率神经网络将茶叶香气分类, 3 个神经网络的分类准确率分别为 90.77%、92.31%和 93.85%, 这说明基于神经网络的电子鼻系统可以对不同等级茶的香气成分组成进行分析, 这也标志着数字术语标准化茶叶香气成为可能。同样, 香气成分分级同样适用于花果茶分级, 但不同种类的花果茶香气成分有所差异, 未来研究需要收集花果茶等级标准品进行香气成分定性参考。

花果茶非挥发性成分中, 茶多酚含量是评判茶叶品质的重要指标之一。LUO 等^[63]的研究表明, 高光谱图像的光谱-空间深层特征可以估算茶多酚含量。该报道首先测得绿茶茶多酚的含量, 然后利用近红外光高光谱成像获取绿茶的高光谱图像, 再利用 CNN 对茶叶高光谱数据进行光谱深度特征和空间深度特征提取。接着, 基于光谱深度特征、空间深度特征和光谱-空间深度特征, 利用 RF、支持向量回归、偏最小二乘回归等算法对绿茶的茶多酚含量进行模型预测。结果表明, 高光谱图像的光谱-空间深度特征预测性能最好, 训练集 $R^2=0.949$, 测试集 $R^2=0.938$, 具有较好的评估能力。与此同时, RF 模型对茶多酚含量预测效果最好。相关研究为机器学习算法结合高光谱技术提取光谱空间深层特征估算茶叶、花果茶中的主要成分奠定了基础, 也为未来利用花果茶中的关键物质结合机器学习进行等级分类和产地识别提供支持。

1.3.2 花果茶无损分级

花果茶的分级需要实现优价优级、低价低级的目的, 既保证了消费者的权益, 又提升其商品化程度。传统的花果茶和茶叶分级一般通过人工筛选, 按照产品的色、香、味、形来区分定级, 但这种方式存在主观性, 无法避免出现误判。相比之下, 机器学习在食品质量分级中的效果更加精确, 同时能够大大提高效率。部分机器学习结合光谱技术在茶叶等级分类中的应用研究见表 3。由表 3 可知, SVM 算法在针对高维、小样本的茶叶分析数据时表现出优秀的分类性能。对于花果茶等级分类, 由于花果茶种类不同, 其挥发性物质和拼配材料也有所差异, 导致花果茶等级鉴定难度增大。因此后续进行花果茶成品等级区分时, 可以同时结合不同检测方法、SVM 算法以及图像识别技术, 对花果茶的色、香、形进行区分, 从而提高分类准确性和效率。

表 3 机器学习在茶叶等级分类中的应用
Table 3 Application of machine learning in tea grade classification

项目	仪器	算法	用途	特点
春米茶 (绿茶) ^[64]	傅里叶近红 外光谱仪	模糊最大不确定度线性 判别分析、KNN	鉴定茶叶等级	傅里叶近红外光谱仪成本低、操作简单、耗时短, 模糊最大不确定度线性判别分析可以识别一些难以区分的数据获取更多特征信息且与 K 值结合稳定性和分类精度高
绿茶 ^[65]	近红外 光谱仪	PCA、多维缩放 MDS、局 部保持投影 LPP、SVM、 索尔普 SOLPP	识别不同等级、产 地和品种的绿茶	索尔普 SOLPP 可以从原始光谱中提供有效的低维特征, 提高模型性能, 所有的分类精度为 100%
祁门 红茶 ^[66]	可见光红外 光谱仪	SVM	茶叶品质分类	可见光红外光谱仪成本低、高效, SVM 训练误差模型复杂度小, 对红茶品质分类模型高
铁观音 ^[67]	荧光高光 谱图技术	SVM	铁观音等级鉴定	荧光高光谱图可以快速、无损、准确地检测样品信息, 经过优化后的 SVM 模型对铁观音等级识别精度高

贝叶斯算法可以通过先验概率和所抽取的样本来推测计算后验概率, 并且通过调整后验概率来调整所有参数的分布。结合神经网络算法, 它能够增强网络泛化能力, 有效解决过拟合的问题。例如, 贝叶斯算法结合 CNN 算法可以有效解决茶叶同类不同级的分类问题, 同样适用于不同等级花果茶的分类。柳泽瑞^[68]使用高光谱数据, 结合贝叶斯算法和 CNN 算法对 7 种茶叶(每种茶叶各有 700 个样品)进行分类, 分级准确率为 96.88%; 在对 3 种不同等级龙井茶进行分类时, 准确率达到了 99.06%, 有效避免了人工分级中存在的主观性问题, 并避免了化学分析方法中可能存在的操作失误给样本带来的损坏和浪费。

对于果干的分级, AHMAD 等^[69]采用了计算机视觉系统对无花果干进行图像识别, 并利用无花果干的颜色、尺寸、形状和纹理等特征进行数据挖掘。随后, 利用 3 种不同的 DT 算法(C4.5 决策树学习器、错误率降低剪枝、逻辑模型树)对无花果干的特征进行分类模型的构建, 其中最佳分类模型为错误率降低剪枝算法, 准确率为 91.74%, 该方法不仅效率高、成本低, 而且可以为其他果干的质量分级提供借鉴。

2 总结与展望

目前, 机器学习已经在花果茶生产加工的多个环节得到应用, 包括原料验收、加工调控、质量分级等方面, 但仍然存在许多需要解决的问题: 其一, 花果茶种类繁多, 不同花果茶的加工方式存在一定的差异, 需要提高机器学习算法在不同花果茶加工中关键参数设定的准确性; 其二, 不同种类的花果茶所含活性成分有所差异, 需要推进机器学习算法在花果茶加工过程中的改进和迭代, 建立一套通用的活性物质分析算法或者融合型机器学习算法; 其三, 花果茶原料及部分花果茶产品存在农药残留和重金属残留超标问题, 机器学习在相关领域的辅助研究不足, 需要继续推进。

机器学习技术在花果茶生产加工中的后续应用主要体现在两个方面: 一是与无损检测技术结合, 例如: 红外

光谱技术、荧光高光谱技术等, 推动机器学习技术在光谱技术和处理方法上的创新, 应用到花果茶产品原料采摘和分级、成品质量分级等环节; 二是与广泛靶向代谢组学检测技术结合, 如: 质子转移反应飞行时间质谱法、超高效液相色谱-四极杆-静电场轨道阱质谱法等, 发挥机器学习技术在花果茶加工数据中的分类和预测能力, 提高花果茶加工的精细化、科学化水平。

参考文献

- [1] 杨莉. 新冠疫情(COVID-19)对全球茶叶市场的影响分析[J]. 茶叶通讯, 2022, 49(1): 123-128.
YANG L. Analysis of the impact of COVID-19 epidemic on global tea market [J]. J Tea Commun, 2022, 49(1): 123-128.
- [2] ZIENIEWSKA I, ZALEWSKA A, ZENDZIAN-PIOTROWSKA M, et al. Antioxidant and antiglycation properties of seventeen fruit teas obtained from one manufacturer [J]. Appl Sci, 2020, 10(15): 5195.
- [3] 黄友谊, 冀志霞. 茶分类方法初探[J]. 茶叶机械杂志, 2001, (1): 24-26.
HUANG YY, JI ZX. Preliminary study on tea classification method [J]. J Tea Mach, 2001, (1): 24-26.
- [4] 李先玉. 高品质茉莉花茶加工技术及品质影响关键因素的研究现状[J]. 福建轻纺, 2019, (3): 30-34.
LI XY. Research status of processing technology and key factors affecting quality of high-quality jasmine tea [J]. Light Text Ind Fujian, 2019, (3): 30-34.
- [5] 庞晓莉, 司辉清, 李永菊. 窈制技术因子对蜡梅花茶品质的影响[J]. 茶叶科学, 2007, (4): 316-322.
PANG XL, SI HQ, LI YJ. Effects of scenting technical factors on the quality of *Chimonanthus praecox* scented tea [J]. J Tea Sci, 2007, (4): 316-322.
- [6] 秦丹丹. 干制方式对无花果品质的影响及无花果茶的研制[D]. 太原: 山西大学, 2018.
QIN DD. Effect of drying methods on fig quality and development of fig tea [D]. Taiyuan: Shanxi University, 2018.
- [7] 王玉婷, 陈奕, 李雨波. 干燥方式对香蕉片总多酚含量及其抗氧化性的影响[J]. 食品科学, 2013, 34(23): 113-117.
WANG YT, CHEN Y, LI YB. Effects of drying methods on total

- polyphenol content and antioxidant activity of banana slices [J]. *Food Sci*, 2013, 34(23): 113–117.
- [8] 郑晓冬, 宋焯, 潘少香, 等. 干制方法对红枣加工过程产品质量控制研究进展[J]. *食品科技*, 2020, 45(7): 64–68.
ZHENG XD, SONG Y, PAN SX, *et al.* Research progress on quality control of red jujube processing by drying methods [J]. *Food Sci Technol*, 2020, 45(7): 64–68.
- [9] 温立香, 张芬, 何梅珍, 等. 茶叶品质评价技术的研究现状[J]. *食品研究与开发*, 2018, 39(15): 197–204.
WEN LX, ZHANG F, HE MZ, *et al.* Research status of tea quality evaluation technology [J]. *Food Res Dev*, 2018, 39(15): 197–204.
- [10] KUHL N, SCHEMMER M, GOUTIER M, *et al.* Artificial intelligence and machine learning [J]. *Mach Learn Know Extr*, 2022, 32: 2235–2244.
- [11] LIN YD, MA J, WANG QJ, *et al.* Applications of machine learning techniques for enhancing nondestructive food quality and safety detection [J]. *Crit Rev Food Sci*, 2022, 32: 2235–2244.
- [12] 浦宇文. 基于近红外光谱的茶鲜叶分级研究与试验[D]. 合肥: 安徽农业大学, 2022.
PU YW. Research and experiment on tea fresh leaf grading based on near-infrared spectroscopy [D]. Hefei: Anhui Agricultural University, 2022.
- [13] 万广, 陈忠辉, 方洪波, 等. 基于特征融合的随机森林模型茶鲜叶分类[J]. *华南农业大学学报*, 2021, 42(4): 125–132.
WAN G, CHEN CH, FANG HB, *et al.* Classification of tea fresh leaves in a random forest model based on feature fusion [J]. *J South China Agric Univ*, 2021, 42(4): 125–132.
- [14] FUENTES A, YOON S, KIM SC, *et al.* A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition [J]. *Sensors*, 2017; 17(9): 2022.
- [15] ZHANG LZ, DAI HM, ZHANG JL, *et al.* A study on origin traceability of white tea (white peony) based on near-infrared spectroscopy and machine learning algorithms [J]. *Foods*, 2023, 12(3): 10.
- [16] 颜丽. 食品安全风险过程控制的贝叶斯统计与知识发现[D]. 广州: 华南理工大学, 2011.
YAN L. Bayesian statistics and knowledge discovery of food safety risk process control [D]. Guangzhou: South China University of Technology, 2011.
- [17] SAIKIA D, SARMA DK, BORUAH PK, *et al.* An ANN model to estimate the impact of tea process parameters on tea quality [J]. *J Circuit Syst Comput*, 2015, 24(9): 1550139.
- [18] 王胜鹏, 滕靖, 郑鹏程, 等. 青砖茶压制压力优化及 GCG 近红外快速检测模型建立[J]. *农业工程学报*, 2020, 36(8): 271–277.
WANG SP, TENG J, ZHENG PC, *et al.* Optimization of pressing pressure of green brick tea and establishment of GCG near-infrared rapid detection model [J]. *Trans Chin Soc Agric Eng*, 2020, 36(8): 271–277.
- [19] KHAN MIH, SABLANI SS, NAYAK R, *et al.* Machine learning-based modeling in food processing applications: State of the art [J]. *Compr Rev Food Sci F*, 2022, 21(2): 1409–1438.
- [20] GONZALEZ VC, TONGSON E, FUENTES S. Integrating a low-cost electronic nose and machine learning modelling to assess coffee aroma profile and intensity [J]. *Sensors*, 2021, 21(6): 2016.
- [21] LU WH, CHEN J, LI XB, *et al.* Flavor components detection and discrimination of isomers in Huaguo tea using headspace-gas chromatography-ion mobility spectrometry and multivariate statistical analysis [J]. *Anal Chim Acta*, 2023, 1243: 340842
- [22] NITISH S, GEOFFREY EH, ALEX K, *et al.* Dropout: A simple way to prevent neural networks from overfitting [J]. *J Mach Learn Res*, 2014, 15(1): 1929–1958.
- [23] PENG CY, REN FY, YE ZH, *et al.* A comparative UHPLC-Q/TOF-MS-based metabolomics approach coupled with machine learning algorithms to differentiate Keemun black teas from narrow-geographic origins [J]. *Food Res Int*, 2022, 158: 111512
- [24] SOON JM, ABDUL WIR. A bayesian approach to predict food fraud type and point of adulteration [J]. *Foods*, 2022, 11(3): 328
- [25] ZHAO J, LIU WW, CHEN Y, *et al.* Identification of markers for tea authenticity assessment: Non-targeted metabolomics of highly similar oolong tea cultivars (*Camellia sinensis* var. *sinensis*) [J]. *Food Control*, 2022, 142: 109223
- [26] 许虞俊. 基于嵌入式平台的卷积神经网络结构设计与优化[D]. 南京: 东南大学, 2021.
XU YJ. Structure design and optimization of convolution neural network based on embedded platform [D]. Nanjing: Southeast University, 2021.
- [27] 姚正锟. 基于深度神经网络的表征学习及其精算应用[D]. 天津: 南开大学, 2022.
YAO ZK. Representation learning based on deep neural network and its actuarial application [D]. Tianjin: Nankai University, 2022.
- [28] 张驰, 郭媛, 黎明. 人工神经网络模型发展及应用综述[J]. *计算机工程与应用*, 2021, 57(11): 57–69.
ZHANG C, GUO Y, LI M. Review on the development and application of artificial neural network models [J]. *Comput Eng Appl*, 2021, 57(11): 57–69.
- [29] 张所任. 基于人工神经网络的园林植物配置研究[J]. *林业调查规划*, 2022, 47(5): 196–200.
ZHANG SR. Study on landscape plant arrangement based on artificial neural network [J]. *For Surv Plan*, 2022, 47(5): 196–200.
- [30] 王祎, 贾文雅, 尹雪婷, 等. 人工神经网络的发展及展望[J]. *智能城市*, 2021, 7(8): 12–13.
WANG W, JIA WY, YIN XT, *et al.* Development and prospect of artificial neural network [J]. *Intell City*, 2021, 7(8): 12–13.
- [31] 王旭辉. 基于朴素贝叶斯算法的属性混合大数据分类提取方法[C]. *Proceedings of 2022 the 6th International Conference on Scientific and Technological Innovation and Educational Development*, 2022.
WANG XH. Attribute hybrid big data classification extraction method based on naive bayes algorithm [C]. *Proceedings of 2022 the 6th International Conference on Scientific and Technological Innovation and Educational Development*, 2022.
- [32] 徐玲玲, 迟冬祥. 面向不平衡数据集的机器学习分类策略[J]. *计算机工程与应用*, 2020, 56(24): 12–27.

- XU LL, CHI DX. Machine learning classification strategy for unbalanced data sets [J]. *Comput Eng Appl*, 2020, 56(24): 12–27.
- [33] 石峰源, 张春明, 姜丽辉, 等. 采用主成分分析的迭代最近点算法优化与验证[J]. *激光与光电子学进展*, 2022, 59(22): 189–197.
- SHI FY, ZHANG CM, JIANG LH, *et al.* Optimization and verification of iterative nearest point algorithm based on principal component analysis [J]. *Las Optoelect Prog*, 2022, 59(22): 189–197.
- [34] 王博, 庄暨军, 熊军, 等. 基于高维特征聚类优化的随机森林算法研究[J]. *井冈山大学学报(自然科学版)*, 2022, 43(5): 52–56.
- WANG B, ZHUANG JJ, XIONG J, *et al.* Research on stochastic forest algorithm based on high-dimensional feature clustering optimization [J]. *J Jinggangshan Univ (Nat Sci)*, 2022, 43(5): 52–56.
- [35] 刘卫明, 陈伟达, 毛伊敏, 等. 结合增益率与堆叠自编码器的并行随机森林算法[J/OL]. *计算机应用研究*: 1–12. [2022-12-12]. DOI: 10.19734/j.issn.1001-3695.2022.08.0374.
- LIU WM, CHEN WD, MAO YM, *et al.* A parallel random forest algorithm combining gain rate and stacked self-encoder [J/OL]. *Appl Res Comput*: 1–12. [2022-12-12]. DOI: 10.19734/j.issn.1001-3695.2022.08.0374
- [36] 李福祥, 王雪, 张驰, 等. 基于边界点的支持向量机分类算法[J]. *陕西理工大学学报(自然科学版)*, 2022, 38(3): 30–38.
- LI FX, WANG X, ZHANG C, *et al.* Support vector machine classification algorithm based on boundary points [J] *J Shaanxi Univ Technol (Nat Sci Ed)*, 2022, 38(3): 30–38.
- [37] 徐洪学, 孙万有, 杜英魁, 等. 机器学习经典算法及其应用研究综述[J]. *电脑知识与技术*, 2020, 16(33): 17–19.
- XU HX, SUN WY, DU YK, *et al.* Overview of classical machine learning algorithms and their applications [J]. *Comput Knowl Technol*, 2020, 16(33): 17–19.
- [38] ZHANG GS, LIU B, ZHU TQ, *et al.* Visual privacy attacks and defenses in deep learning: A survey [J]. *Artif Intell Rev*, 2022, 55: 4347–4401.
- [39] 夏杰桢, 曹蓉, 吴琪. 机器学习结合密度泛函理论计算在材料科学中的研究进展[J]. *化学通报*, 2022, 85(10): 1224–1232.
- XIA JZ, CAO R, WU Q. Research progress of machine learning combined with density functional theory calculation in materials science [J]. *Chem Bull*, 2022, 85(10): 1224–1232.
- [40] 马骋乾, 谢伟, 孙伟杰. 强化学习研究综述[J]. *指挥控制与仿真*, 2018, 40(6): 68–72.
- MA CQ, XIE W, SUN WJ. A review of reinforcement learning research [J]. *Command Control Simul*, 2018, 40(6): 68–72.
- [41] 陶国柱. 基于卷积神经网络的茉莉花病虫害识别算法研究[D]. 南宁: 广西民族大学, 2021.
- TAO GZ. Research on identification algorithm of jasmine diseases and insect pests based on convolution neural network [D]. Nanning: Guangxi University for Nationalities, 2021.
- [42] VITHU P, MOSES JA. Machine vision system for food grain quality evaluation: A review [J]. *Trends Food Sci Technol*, 2016, 56: 13–20.
- [43] LIU CL, LU WY, GAO BY, *et al.* Rapid identification of chrysanthemum teas by computer vision and deep learning [J]. *Food Sci Nuir*, 2020, 8(4): 1968–1977.
- [44] CHEN HW, LIANG HZ, DING JH, *et al.* Rapid differentiation of tea products by surface desorption atmospheric pressure chemical ionization mass spectrometry [J]. *J Agric Food Chem*, 2007, 55(25): 10093–10100.
- [45] 赵凡. 基于 FPGA 的卷积神经网络水果分级系统的设计与实现[D]. 镇江: 江苏大学, 2021.
- ZHAO F. Design and implementation of convolution neural network fruit grading system based on FPGA [D]. Zhenjiang: Jiangsu University, 2021.
- [46] İSMAIL K, AZIM DT, CEYLİN Ş, *et al.* Modeling and developing a smart interface for various drying methods of pomelo fruit (*Citrus maxima*) peel using machine learning approaches [J]. *Comput Electron Agric*, 2019, 165(C): 104928.
- [47] LUQMON A, ABDULRASAQ O, SEGUN A, *et al.* Chemical components retention and modelling of antioxidant activity using neural networks in oven dried tomato slices with and without osmotic dehydration pre-treatment [J]. *J Food Meas Charact*, 2017, 11(4): 2247–2258.
- [48] ÇETİN N. Prediction of moisture ratio and drying rate of orange slices using machine learning approaches [J]. *J Food Process Pres*, 2022, 46(11): e17011.
- [49] KHALED AY, KABUTEY A, SELVI KÇ. *et al.* Application of computational intelligence in describing the drying kinetics of persimmon fruit (*Diospyros kaki*) during vacuum and hot air drying process [J]. *Processes*, 2020, 8(5): 544.
- [50] SAGLAM C, CETIN N. Machine learning algorithms to estimate drying characteristics of apples slices dried with different methods [J]. *J Food Process Pres*, 2022, 46(10): e16496.
- [51] ZHANG WP, WANG K, CHEN C. Artificial neural network assisted multiobjective optimization of postharvest blanching and drying of blueberries [J]. *Foods*, 2022, 11(21): 3347.
- [52] 卢健, 王东, 汪群, 等. 传统窰制与隔离窰制中茉莉花茶香气成分的比较分析[J]. *中国农学通报*, 2022, 38(25): 158–164.
- LU J, WANG D, WANG Q, *et al.* Comparative analysis of aroma components of jasmine tea in traditional scenting system and isolated scenting system [J]. *Chin Agric Sci Bull*, 2022, 38(25): 158–164.
- [53] ZHANG C, ZHOU CZ, TIAN CY, *et al.* Volatilomics analysis of jasmine tea during multiple rounds of scenting processes [J]. *Foods*, 2023, 12(4): 812.
- [54] 石勇, 石训, 孙晓瑞, 等. 冷冻干燥联合微波干燥对枇杷花茶品质的影响[J]. *保鲜与加工*, 2023, 23(1): 52–56.
- SHI Y, SHI X, SUN XR, *et al.* Effect of freeze drying combined with microwave drying on the quality of loquat flower tea [J]. *Storage Process*, 2023, 23(1): 52–56.
- [55] 姜兴旭, 李玉娇, 王晓, 等. 夏秋茶萎凋新技术研究进展[J]. *现代食品*, 2018, (7): 184–187.
- JIANG XX, LI YJ, WANG X, *et al.* Research progress on new withering technology of summer and autumn tea [J]. *Mod food*, 2018, (7): 184–187.
- [56] SHRILEKHA D, TANMOY S, ASHIS KD. Improving black tea quality through optimization of withering conditions using artificial neural network and genetic algorithm [J]. *J Food Process Pres*, 2021, 45(3):

- e15273.
- [57] 王雨芹. 遗传算法在辐射防护中的应用[J]. 电子技术与软件工程, 2022, (14): 143–146.
WANG YQ. Application of genetic algorithm in radiation protection [J]. Electron Technol Softw Eng, 2022, (14): 143–146.
- [58] 赵苗苗, 严亮, 张文杰, 等. 红紫芽熟茶适制性及渥堆发酵过程品质变化研究[J]. 食品安全质量检测学报, 2022, 13(19): 6212–6220.
ZHAO MM, YAN L, ZHANG WJ, *et al.* Study on suitability and quality change of red-purple bud ripe tea during pile fermentation [J]. J Food Saf Qual, 2022, 13(19): 6212–6220.
- [59] 雷攀登, 周汉琛, 田娟, 等. 高湿高雾环境下红茶发酵图像实时采集系统研制[J]. 农业工程学报, 2022, 38(12): 207–215.
LEI PD, ZHOU HC, TIAN J, *et al.* Development of real-time image acquisition system for black tea fermentation in high humidity and high fog environment [J]. Chin Soc Agric Eng, 2022, 38(12): 207–215.
- [60] ZHU HK, LIU F, YE Y, *et al.* Application of machine learning algorithms in quality assurance of fermentation process of black tea-Based on electrical properties [J]. J Food Eng, 2019, 263: 165–172.
- [61] LUO DS, CHEN J, GAO L, *et al.* Geographical origin identification and quality control of Chinese chrysanthemum flower teas using gas chromatography–mass spectrometry and olfactometry and electronic nose combined with principal component analysis [J]. Int J Food Sci Technol, 2017, 52(3): 714–723.
- [62] BORAH S, HINES EL, LEESON MS, *et al.* Neural network based electronic nose for classification of tea aroma [J]. Sens Instr Food Qual, 2008, 2(1): 7–14.
- [63] LUO N, LI YL, YANG BH, *et al.* Prediction model for tea polyphenol content with deep features extracted using 1D and 2D convolutional neural network [J]. Agriculture, 2022, 12(9): 1299.
- [64] WU X, HE F, WU B, *et al.* Accurate classification of chunmee tea grade using NIR spectroscopy and fuzzy maximum uncertainty linear discriminant analysis [J]. Foods, 2023, 12: 541.
- [65] LIU P, WEN YP, HUANG JS, *et al.* A novel strategy of near-infrared spectroscopy dimensionality reduction for discrimination of grades, varieties and origins of green tea [J]. Vib Spectrosc, 2019, 105(C): 102984.
- [66] REN GX, LIU Y, NING JM, *et al.* Assessing black tea quality based on visible–near infrared spectra and kernel-based methods [J]. J Food Compos Anal, 2021, 98: 103810.
- [67] LI Y, SUN J, WU X, *et al.* Grade identification of tieguanyin tea using fluorescence hyperspectra and different statistical algorithms [J]. J Food Sci, 2019, 84(8): 2234–2241.
- [68] 柳泽瑞. 基于贝叶斯卷积神经网络的茶叶分类方法研究[D]. 天津: 天津商业大学, 2021.
LIU ZR. Study on tea classification method based on bayesian convolution neural network [D]. Tianjin: Tianjin University of Commerce, 2021.
- [69] AHMAD B, HEMAD Z, MEHRDAD B, *et al.* Combined application of decision tree and fuzzy logic techniques for intelligent grading of dried figs [J]. J Food Process Eng, 2017, 40(3): e12456.

(责任编辑: 郑丽于梦娇)

作者简介



庞文媛, 硕士研究生, 主要研究方向为天然产物化学, 多学科交叉融合下的食品组分应用基础研究。

E-mail: 1093655593@qq.com



庞杰, 博士, 教授, 主要研究方向为天然植物多糖的结构与性能。

E-mail: pang3721941@163.com