

SUN Yong, 王译霄, 宣晓婷, 等. 基于机器学习联合网络爬虫算法的果汁鉴伪技术的研究进展[J]. 华中农业大学学报, 2022, 41(1): 269-276.
DOI: 10.13300/j.cnki.hnlkxb.2022.01.028

基于机器学习联合网络爬虫算法的果汁鉴伪技术的研究进展

SUN Yong¹, 王译霄¹, 宣晓婷², 张熙悦¹, 邓文艺³, 靳欢¹, 凌建刚²

1. 宁波诺丁汉大学化工与环境工程学院, 宁波 315100;

2. 宁波市农业科学研究院/国家蔬菜加工技术研发专业中心/宁波市农产品保鲜工程重点实验室, 宁波 315400;

3. 湖南一品东方生物科技有限公司, 长沙 410000

摘要 果汁鉴伪技术正成为全球食品营养安全领域的研究热点,其发展对果汁质量与安全控制至关重要。结合机器学习联合网络爬虫算法检索文献技术,本文综述了近年来果汁鉴伪技术的研究进展与应用领域,并着重阐述了非靶向/靶向代谢组学鉴伪技术的研究进展。结合多学科交叉和计算机技术发展趋势,对检测方法耦合发展、监督型机器学习统计工具开发、无人监督鉴伪等趋势作了预测,以期对果汁鉴伪领域的深入研究提供参考。

关键词 果汁鉴伪; 果汁掺假; 网络爬虫; 机器学习; 非靶向/靶向标记法; 代谢组学

中图分类号 TS 275.5 **文献标识码** A **文章编号** 1000-2421(2022)01-0269-08

果汁是以水果为原料,经压榨、离心、萃取等物理方法得到原汁后再经加工制成,其富含维生素、矿物质、糖分和果胶等营养素^[1-2]。根据不同的加工方式,可将果汁分为非浓缩还原型和浓缩还原型果汁,目前市场上大多数纯果汁都是浓缩还原型果汁^[3-5]。此外,相较部分发达国家,我国纯果汁的市场份额占5%以下,因此,果汁市场具有巨大的发展潜力与前景。

受到西方饮食习惯以及健康理念的影响,在我国果汁产品逐渐受到人们喜爱。果汁行业较水果零售业具有更高的附加值,可以有效解决部分水果保鲜期短、运输成本高等缺点。由于经济利益驱动,果汁掺杂、掺假等现象层出不穷^[3,6]。这些掺假手段包括:兑水稀释、添加外源糖、果渣以及其他果汁勾兑、使用低劣果汁掺假^[7]。这些掺假手段除了使消费者的经济利益受损外,还可能对消费者造成严重的健康危害。因此,各国都在积极制定相应的政策法规以完善果汁生产工艺和品质控制。目前,果汁鉴伪技术主要有3个难点:一是水果种类的多样性;二是地域性差异;三是不同储存和处理技术的差异性等,导致果汁鉴伪的检测流程长、手段复杂、费用高昂等问题^[8]。随着代谢组学技术的发展,目前已

有研究将代谢组学应用于果汁鉴伪中,根据研究目的不同,可分为靶向代谢组学和非靶向代谢组学,其中靶向标记法是针对某个或某一组特定的靶向化合物进行检测,通常包括氨基酸、有机酸、寡糖、多糖以及单糖等^[9]。该方法最大优点在于能够相对准确和定量检测出小分子代谢产物,为果汁鉴伪以及确定果汁产地提供可靠的参考依据^[8,10]。而非靶向标记法主要是通过检测果汁内的代谢产物的变化来鉴定真伪,如以未掺假的果汁作为基准,掺假果汁中代谢成分的分解与转化呈现不同趋势,进而实现鉴伪。

本文借助计算机机器学习联合网络爬虫算法检索文献,总结近年来国内外主要的果汁鉴伪技术的研究与应用进展,分析其优势和不足,并着重阐述基于代谢组学的非靶向/靶向标记法在果汁鉴伪中的应用,以期对果汁鉴伪技术研究提供参考。

1 网络爬虫算法与数据统计

网络爬虫是模拟浏览器发送网络请求,按照特定规则自动抓取互联网信息的程序。利用爬虫技术对文献进行抓取,可得到互联网上所有公开学术文献信息,并可批量下载。与传统文献检索方式相比,

收稿日期: 2021-09-14

基金项目: 宁波市公益类科技计划项目(2019C10104; 2019C10033); 国家重点研发计划项目(2018YFC1903500)

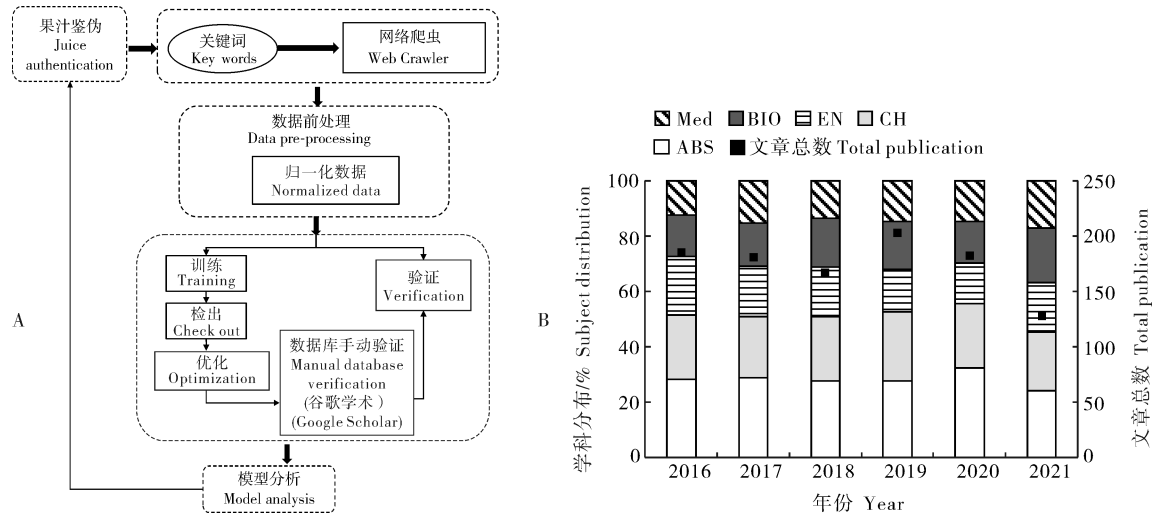
SUN Yong, E-mail: Yong.Sun@nottingham.edu.cn

通信作者: 凌建刚, E-mail: nbnjg@163.com

使用爬虫算法检索可以节省文献获取时间,显著提高工作效率。

本文主要对几大主流数据库网站(谷歌学术 Google academic、Scopus、Springer、MDPI 等)进行收集,由于数据量巨大,我们使用了独立研发的基于网络爬虫和数据分析算法的模型,该算法抓取数据、验证数据的流程如图 1A 所示。使用关键词包括 adulteration、authenticity、biomarker 等。所有文献都将通过自动和手动双验证以达到最大限度相关

性。目前数据相关性可达 90% 以上^[11-12]。根据文献内容,可分为五大类即农业生物类(ABS)、化学类(CH)、工程类(EN)、生物化学类(BIO)以及医药类(Med)。由图 1B 可见,2016 年至今,有关果汁鉴伪的学术文章基本保持每年 150 篇左右。从学科分类上看,果汁鉴伪技术主要出现在农业与生命科学领域,每年的分布都在 30% 以上,与农业食品安全有着紧密联系。因此,果汁鉴伪与安全保障仍将是近年的研究热点。



Med:医学 Medical;BIO:生物化学 Biochemistry;EN:工程学 Engineering;CH:化学 Chemistry;ABS:农业和生物科学 Agricultural and biological sciences.

图 1 数据收集和验证算法流程图(A)和文章数据统计(B)

Fig.1 Flowchart of data collection and validation algorithms(A)and article data statistics(B)

2 常用鉴伪方法

表 1 列举了七大类不同的检测方法。物理化学法是利用果汁中的特征组分来确定是否存在掺假,但存在一定局限性,如在加工过程中添加此类组分所导致的鉴伪精确度低,另外该方法也易受到品种、产地、成熟度及种植条件等因素的影响。光谱分析在果汁鉴伪中也有较多应用,其具有检测速度快、无损、测量方便和重复性佳等优点。Vardin 等^[13]研究发现 FTIR 技术结合 PCA 分析可以鉴别石榴浓缩汁中掺假葡萄浓缩汁。León 等^[14]借助近红外光谱检测苹果汁中掺糖情况,鉴伪度在 91%~100%。核磁技术虽然需要较长的分析时间、样品处理繁杂且成本高,但其通过¹⁸O/¹⁶O、²H/¹H、¹³C/¹²C 这些特征比率定量分析,所得的鉴伪精确度较高。此外,基于水果指纹 DNA 的检测方法通常具有较好的特异

性^[14],在进行分析之前,需要对指纹 DNA 进行 PCR 扩增,并利用水果物理化学性质的不同进行鉴伪,但通常该方法的鉴伪度不高,一般可以达到 25%。目前 PCR 技术已在苹果汁、橙汁、梨汁等果汁鉴伪中进行应用。Morton 等^[15]借助 PCR 技术和琼脂糖凝胶电泳分析苹果汁的特征基因图谱,实现了苹果汁的 PCR 鉴伪。同样有报道成功从橙汁中鉴别柑橘汁,其变异系数仅为 2.5%^[16]。研究表明,以 DNA 为基础的 PCR 技术具有灵敏度高、条件易优化等优点^[17],但其劣势在于果汁样品中 DNA 的含量较低,且加工方式会影响其含量,从而不利于 PCR 技术的应用。综上,每一种分析方法都各有利弊,相应的鉴伪度高低差别较大,未来将沿着不同检测方法的耦合方向发展,为果汁鉴伪开发出更高效、准确的检测技术和分析方法。

表 1 果汁鉴伪主要方法

Table 1 Comparison of different detection methods

类别 Category	检测技术 Detection technology	鉴别度/% Degree of accuracy	文献 References
基于 DNA-检测 Based on DNA-detection	PCR, PCR-LOC, SCAR	1~25	[18-19]
物理化学法 Physical-chemical method	导电能力、等电点 Conductivity, isoelectric points	10~30	[20]
元素分析 Elemental analysis	电感耦合等离子体光发射光谱、电感耦合等离子质谱 Inductively coupled plasma photoemission spectra, inductively coupled plasma mass spectra	1~10	[21]
同位素分析 Isotopic analysis	IRMS, SNIF-NMR	0.1~10	[22-23]
光谱分析 Spectrum analysis	紫外线吸收光谱、红外光谱、拉曼光谱、核磁共振 UV absorption, infrared, Raman, NMR	0.1~10	[24]
分离法 Isolation method	GC, LC, TLC, CE	0.1~20	[25]
质谱 Mass spectra	GC-MS, LC-MS	1~25	[26]

注 Note: 聚合酶链式反应(PCR) Polymerase chain reaction; 芯片实验室(LOC) Laboratory on the chip; 序列特征扩增区(SCAR) Sequence characterized amplified regions, 同位素比值质谱(IRMS) Isotope ratio mass spectrometry; 定点天然同位素分馏核磁共振 Site-specific natural isotope fractionation-nuclear magnetic resonance(SNIF-NMR); 薄层色谱(TLC) Thin layer chromatography, 毛细管电泳(CE) Capillary electrophoresis; 气相色谱-质谱法(GC-MS) Gas chromatography-mass spectrometer; 液相色谱-质谱法(LC-MS) Liquid chromatography-mass spectrometer.

3 代谢组学在果汁鉴伪中的应用

代谢组学作为现代食品科学研究的重要技术手段,在食品鉴伪中已逐渐开展应用研究。利用核磁共振(NMR)、质谱(MS)、色谱(HPLC、GC)及联用技术,对分子质量 <1 ku 的代谢物进行定量定性分析以达到物种鉴定的目的^[27]。根据研究目的不同,代谢组学分为靶向代谢组学(靶向标记法)和非靶向代谢组学(非靶向标记法)。靶向标记法是对特定代谢物进行针对性、特异性的检测与分析,而非靶向标记法则是针对有机体内源性代谢物进行全面、系统的分析。两者有机结合可以更好地发现差异代谢物并进行定量分析。

3.1 靶向标记法

因水果种类、地域、品种和代谢途径等不同,水果内部的物理化学性质也有所不同,使其呈现出独特的性质,从而赋予水果不同的特性。在果汁鉴伪过程中,精准、快速地分辨其代谢成分的变化是至关重要的。靶向标记法能够对果汁化学成分进行深入分析,是其中应用最为广泛的检测方法之一。但是该方法的分析过程较为复杂,同时要求找到足够的生物活性物质作为标记,这些不利因素也从不同程度上限制了其在品控和检测上的大规模应用。

表 2 总结了不同果汁中的生物活性物质。从表 2 可以看出,不同果汁呈现出不同的生物标记物,以 2-甲基丙基-乙酸为例,其广泛存在于苹果汁中,但

在其他类型的果汁中含量较低,可用于苹果汁的鉴别。然而仅以 2-甲基丙基-乙酸作为特征生物标记物会影响准确度。因此,增加生物标记物的数量和种类可最大限度降低误差,但也导致分析方法不可避免的复杂化,从而难以统一标准化。因此,利用通用方法和特异性方法联合鉴伪是未来发展趋势。首先使用相对快速、经济的通用方法进行初筛,之后再使用特异性方法对初筛样品进行针对性的特异性分析。

3.2 非靶向标记法

非靶向标记法是指采用 LC-MS、GC-MS、NMR 技术,无偏向性地检测受体体内受到刺激或扰动前后所有小分子代谢物(分子质量 1 ku 以内)的动态变化,并通过生物信息学分析筛选差异代谢物,对差异代谢物进行通路分析,揭示其变化的生理机制。相较于靶向标记法,非靶向标记法重在发现差异代谢物,可以定性及相对定量分析,操作简便、成本较低。开发非靶向标记方法是靶向标记方法的重要补充。由于靶向标记方法的诸多限制,开发更为有效和简便的分析方法显得尤为重要。现代高分辨率质谱以及分离技术如超高压液相色谱等技术的发展,使非靶向标记方法得以长足的发展。结合数据挖掘以及多维化学计量分析技术,非靶向标记方法可以更准确而高效地分析果汁化学成分。Jandric 等^[41]研究发现采用超高效液相色谱-四级杆串联飞行时间质谱(UPLC-QTOF/MS)联用技术可快速检测果汁中掺杂低劣果汁。刘晗璐等^[42]同样利用 UPLC-

QTOF/MS 技术实现了 NFC 和 FC 橙汁的区别,为 NFC 果汁的品质鉴别提供了新思路。

表 3 统计了非靶向标记法在不同果汁鉴别中的应用。以花青素为例,由于花青素是细胞内重要的抗氧化剂,在果汁加工和贮藏中,花青素将出现不同程度的代谢与降解^[43-44]。如果出现人工掺杂,花青

素含量也会出现异常水平,从而为鉴别提供很好的标记^[45]。此外,非靶向标记方法能够找到果汁种类以及产地的相关性,因为不同产地的水果其所处生长环境不同,其代谢产物也呈现出细微的变化。除了主要代谢产物,水果中的次级代谢产物(醛、阿魏酸等)也可作为标记物。实践表明,使用非靶向标记

表 2 不同果汁中的生物标记物

Table 2 Biomarkers in different fruit juices

果汁 Juices	生物标记物 Biomarkers	文献 References
葡萄汁 Grape juice	腺嘌呤、丙氨酸、咖啡因、果糖、肌醇、络氨酸 Adenine, alanine, caffeine, fructose, inositol, tyrosine	[28]
苹果汁 Apple juice	丙氨酸、柠檬苹果酸、奎宁酸、木糖、2-甲基丙基-乙酸 Alanine, citric malic acid, quinic acid, xylose, 2-methylpropyl-acetic acid	[29]
苹果汁 Apple juice	1-己醇、2-甲基丙基-乙酸 1-Hexanol, 2-methyl propyl-acetic acid	[30]
苹果汁 Apple juice	丁醇、1-己醇、乙醇、2-甲基丙基-乙酸 1-Butanol, 1-hexanol, ethanol, 2-methylpropyl-acetic acid	[31]
柠檬汁 Lemon juice	落叶酸、阿魏酸、环内酯、柚皮苷 Dideciduous acid, ferulic acid, cyclosome, naringen	[32]
圣洁莓汁 Holy berry juice	3-氧基-反式阿魏酰基-苦味酸 3-Oxygen-trans-ferulic acid-picric acid	[33]
枣汁 Jujube juice	2,3-丁二醇、己醇、肉桂醛 2,3-Butanediol, hexanol, cinnamaldehyde	[34]
百香果汁 Passion fruit juice	庚酮、戊酸苄酯、己酸丁酯、乙基-巴豆酸 2-Heptanone, phenylmethyl ester, butyl hexanoate, ethyl-putaric acid	[35]
肉桂汁 Cinnamomum cassia juice	肉桂醛、丙酸正丁酯、苯甲酸、萜品醇 Cinnamaldehyde, propanoic acid, benzoic acid, terpineol	[36]
西红柿汁 Tomato juice	谷氨酰-苯丙氨酸 Glutamyl-phenylalanine	[37]
香蕉汁 Banana juice	2-甲基-2-甲基丙烷-丁酸、2-戊酯-乙酸 2-Methyl-2-methylpropane-butyrate, 2-pentyl-acetic acid	[38]
灯笼果汁 Lantern juice	己醛、D-藻糖蛋白、3-糖苷 Hexanal, D-trehalose protein, 3-glycoside	[39]
蓝莓汁 Blueberry juice	翠雀花素-3-阿糖甙、翠雀花素-3-半乳糖苷、翠雀花素-糖苷 Triafinin-3-glycoside, trafinin-3-galactoside, trafinin-glycoside	[40]

表 3 非靶向标记法在不同果汁鉴别中的应用

Table 3 Application of non-targeted labeling method in the authentication of different fruit juice

生物标记物 Biomarkers	果汁 Juices	检测方法 Detection method	统计学方法 Statistical method	文献 References
类黄酮、生物碱、有机酸 Flavonoids, alkaloids, organic acid	橙汁 Orange juice	HPLC-DAD-MS/MS	OPLS-DA	[42]
香草草甙、异构菜茵菌素、柚皮苷 Didymin, heterogeneous Rheintin, naringen	柑橘汁 Citrus juice	UPLC-QTOF/MS	PCA	[46]
绿原酸、表儿茶素 Chlorogenic acid, L-epicatechin	苹果汁 Apple juice	UHPLC-QTOF	OPLS-DA, HCA	[47]
脱落酸、柠檬苦素、阿魏酸 Abscisic acid, limonin, ferulic acid	柑橘汁 Citrus juice	UPLC-QTOF/MS	HCA, PLSDA	[48]
类黄酮、脂肪酸 Flavonoid, fatty acids	橙汁 Orange juice	HPLC-HR-MS	HCA, PLSDA	[49]
樟脑萜、乙酰基呋喃、柠檬烯 Camphene, acetyl furan, cinene	橙汁 Orange juice	HS-SPME-GC-MS	PCA, PLSDA	[50]
葡萄糖胺、芥酸酰胺、三肽 Glucosamine, erucylamide, tripeptide	橙汁 Orange juice	DART-QTOF-MS	PLSDA	[51]

注 Note: GC-MS: 气相色谱-质谱联用仪 Gas chromatograph-mass spectrometer-computer; HPLC: 超高效液相色谱仪 High performance liquid chromatography; HPLC-DAD: 高效液相色谱仪-连接二级阵列检测器 High performance liquid chromatography-diode-array detector; UPLC: 超高效液相色谱仪 Ultrahigh performance liquid chromatography; HS-SPME-GC-MS: 顶空固相微萃取-气相色谱-质谱联用技术 Headspace solid phase micro-extraction gas chromatography mass spectrometry; QTOF-MS: 四极杆飞行时间质谱仪 Quadrupole time-of-flight mass spectrometry; DART-QTOF-MS: 实时直接分析-四极杆飞行时间质谱 Direct analysis in real time-quadrupole time-of-flight mass spectrometry; HCA: 层次聚类分析 Hierarchical cluster analysis; PLSDA: 偏最小二乘判别分析 Partial least squares discriminant analysis.

法所找到的标记物与已知标准物联合对果汁进行鉴伪,通常准确率都较高。由表 3 可见,在标记物中,依据水果的种类、产地以及因呼吸作用引起的次级代谢,非靶向分析方法可以相对快速和准确地确定不同果汁的指纹图谱。与靶向标记法相比,目前非靶向标记法的研究也呈现逐步上升的趋势。2020—2021 年,从 Scopus 发表的论文数量上看,平均每年的发表数量在 30 篇(<https://www.scopus.com/>)。非靶向标记方法因其分析过程相对快速和高效,正在逐步成为靶向标记法的有力辅助手段。

此外,随着现代高分辨率质谱以及分离技术的快速发展,在果汁鉴伪中将产生大量的数据,因而统计学方法显得尤为重要。在众多的统计分析方法中,ANOVA 是在不同领域应用最为广泛的数理统计方法,通过 ANOVA 分析,可以初步找到参量变化对于响应结果是否具有统计显著性^[11]。偏最小二乘回归(PLS)法通过投影预测变量和观测变量得到一个新空间来建立一个线性回归模型,目前在靶向和非靶向代谢组学中得到广泛应用^[52]。基于非相似性的偏最小二乘回归(DPLS)是在偏最小二乘回归的基础上发展起来的统计分析方法,其最大特点是能够达到对主成分的回归^[53]。此外,主成分分析(PCA)作为一种统计分析、简化数据集的方法而广泛应用于数据处理^[37]。俞邱豪^[54]在小浆果类果汁鉴伪过程中,通过对果汁中化学成分进行 PCA 分析得到 32 种特征标记化合物可用于区别浆果类果汁。HCA 和 PCA 最大的优势在于无监督学习。这对于机器学习和数据挖掘具有非常重要的意义^[12]。综上所述,基于果汁样品的自身复杂性,使用单一的统计数据方法往往会带来较大误差,目前通用的方法是联合使用这些统计分析方法,从而避免单一统计方法带来的系统性误差。此外,随着人工智能技术的不断迭代,基于非监督型的机器学习统计工具将可能成为未来全自动化果汁鉴伪技术开发的核心,这也对果汁鉴伪的整体算法建立和优化提出了新的要求。随着计算机算力和分析设备检测限的不断提高,全自动无人监督的果汁鉴伪技术将成为未来的发展趋势。

4 结论与展望

随着社会和科技的快速发展,人们对绿色、健康、方便的理念越来越重视,在巨大的经济利益驱动下,果汁掺杂、掺假等问题严重,亟需高效、可靠、简

单的果汁鉴伪技术。相较于传统鉴伪技术,代谢组学是针对生物代谢物对样品进行检测,具有精确、可定量等优势,适合果汁等深加工产品的鉴伪。其中非靶向标记法因其相对快速和高效的分析手法,正在逐步成为靶向标记法的有力辅助,在果汁鉴伪领域的应用具有巨大潜力。基于果汁样品的产地、品种、加工和保存方式的多样性、复杂性,使用单一的鉴伪技术往往会出现假阳性和假阴性现象,需要深入研究以加强代谢组学技术在复合果汁检测中的适用性。此外,随着数据挖掘和计算机技术的不断更新迭代,基于非监督型的机器学习统计工具如 PCA、HCA 等将可能成为未来智能化果汁鉴伪开发的核心。这也对果汁鉴伪技术整体算法的建立和优化提出了新的要求。未来应沿着不同检测方法的耦合方向,为果汁鉴伪提供准确率高的分析方法。

致 谢 衷心感谢伦敦大学学院化学工程系刘奕杨教授协助算法编程与数据收集,感谢宁波诺丁汉大学英语语言教育中心杨善如教授协助论文撰写,感谢宁波诺丁汉大学化工与环境工程学院何俊教授、中国科学院过程工程研究所王云山和杨刚教授对本文提出的宝贵修改意见。

参考文献 References

- [1] WADOOD S A, GUO B L, ZHANG X W, et al. Recent development in the application of analytical techniques for the traceability and authenticity of food of plant origin[J/OL]. *Microchemical journal*, 2020, 152: 104295[2021-09-14]. <https://doi.org/10.1016/j.microc.2019.104295>.
- [2] 孙勇, 张金平, 杨刚, 等. 二氧化氯氧化云杉木质素的光谱研究[J]. *光谱学与光谱分析*, 2007, 27(8): 1551-1554. SUN Y, ZHANG J P, YANG G, et al. Study on the spectra of spruce lignin with chlorine dioxide oxidation[J]. *Spectroscopy and spectral analysis*, 2007, 27(8): 1551-1554(in Chinese with English abstract).
- [3] DASENAKI M, THOMAIDIS N. Quality and authenticity control of fruit juices: a review[J/OL]. *Molecules*, 2019, 24(6): 1014[2021-09-14]. <https://doi.org/10.3390/molecules24061014>.
- [4] 康孟利, 崔燕, 尚海涛, 等. 非热杀菌在 NFC 果汁上的应用前景[J]. *北方园艺*, 2016(18): 190-193. KANG M L, CUI Y, SHANG H T, et al. Application prospect of non-thermal sterilization in NFC juice[J]. *Northern horticulture*, 2016(18): 190-193(in Chinese with English abstract).
- [5] AKYLLDLZ A, KARACA E, AĞÇAM E, et al. Changes in quality

- attributes during production steps and frozen-storage of pomegranate juice concentrate[J/OL]. Journal of food composition and analysis, 2020, 92: 103548 [2021-09-14]. <https://doi.org/10.1016/j.jfca.2020.103548>.
- [6] SUN Y, YANG G, ZHANG L, et al. Preparation of high performance H₂S removal biochar by direct fluidized bed carbonization using potato peel waste[J]. Process safety and environmental protection, 2017, 107: 281-288.
- [7] 牛灿杰, 张慧, 陈小珍. 果汁掺假鉴别检测技术研究进展[J]. 江苏农业科学, 2015, 43(6): 292-294. NIU C J, ZHANG H, CHEN X Z. Advances in the detection technology of fruit juice adulteration identification [J]. Jiangsu agricultural sciences, 2015, 43(6): 292-294 (in Chinese).
- [8] WILLEMS J L, LOW N H. Structural identification of compounds for use in the detection of juice-to-juice debasing between apple and pear juices[J]. Food chemistry, 2018, 241: 346-352.
- [9] UZHEL A S, BORODINA A N, GORBOVSKAYA A V, et al. Determination of full organic acid profiles in fruit juices and alcoholic beverages using novel chemically derivatized hyperbranched anion exchanger[J/OL]. Journal of food composition and analysis, 2021, 95: 103674 [2021-09-14]. <https://doi.org/10.1016/j.jfca.2020.103674>.
- [10] WOZNIAK Ł, SZCZEPANSKA J, ROSZKO M, et al. Occurrence of maltose in apple juices; improved method of analysis, typical levels, and factors affecting it[J/OL]. LWT, 2020, 124: 109154 [2021-09-14]. <https://doi.org/10.1016/j.lwt.2020.109154>.
- [11] SUN Y, YANG G, WEN C, et al. Artificial neural networks with response surface methodology for optimization of selective CO₂ hydrogenation using K-promoted iron catalyst in a micro-channel reactor[J]. Journal of CO₂ utilization, 2018, 24: 10-21.
- [12] WANG Y S, YANG G, SAGE V, et al. Optimization of dark fermentation for biohydrogen production using a hybrid artificial neural network (ANN) and response surface methodology (RSM) approach[J/OL]. Environmental progress & sustainable energy, 2021, 40(2): 13485 [2021-09-14]. <https://doi.org/10.1002/ep.13485>.
- [13] VARDIN H, TAY A, OZEN B, et al. Authentication of pomegranate juice concentrate using FTIR spectroscopy and chemometrics[J]. Food chemistry, 2008, 108(2): 742-748.
- [14] LEÓN L, KELLY J D, DOWNEY G. Detection of apple juice adulteration using near-infrared transmittance spectroscopy[J]. Applied spectroscopy, 2005, 59(5): 593-599.
- [15] MORTON A, ADAMS A N, BARBARA D J. Rapid PCR-based identification of apple (*Malus*) cultivars and species[J]. BCPC monograph, 1993, 54: 289-294.
- [16] KNIGHT A. Development and validation of a PCR-based heteroduplex assay for the quantitative detection of mandarin juice in processed orange juices[J]. Agro food industry Hi-Tech, 2000, 11(2): 7-8.
- [17] HAN J X, WU Y J, HUANG W S, et al. PCR and DHPLC methods used to detect juice ingredient from 7 fruits[J]. Food control, 2012, 25(2): 696-703.
- [18] WOOLFE M, GURUNG T, WALKER M J. Can analytical chemists do molecular biology? A survey of the up-skilling of the UK official food control system in DNA food authenticity techniques[J]. Food control, 2013, 33(2): 385-392.
- [19] 韩建勋, 黄文胜, 吴亚君, 等. 果汁中梨成分分子生物学鉴别-实时荧光 PCR 方法研究[J]. 中国食品学报, 2010, 10(1): 207-213. HAN J X, HUANG W S, WU Y J, et al. Study on the biotechnological adulteration detection of pear component in fruit juice by real-time fluorescence PCR method[J]. Journal of Chinese institute of food science and technology, 2010, 10(1): 207-213 (in Chinese with English abstract).
- [20] DZUGAN M, WESOŁOWSKA M, ZAGUŁA G, et al. The comparison of the physicochemical parameters and antioxidant activity of homemade and commercial pomegranate juices[J]. Acta scientiarum polonorum technologia alimentaria, 2018, 17(1): 59-68.
- [21] DE ANDRADE M F, DA SILVA I J S, PIMENTEL M F, et al. Ultrasonic nebulization inductively coupled plasma optical emission spectrometry method for wine analysis[J/OL]. Spectrochimica acta part B: atomic spectroscopy, 2020, 170: 105924 [2021-09-14]. <https://doi.org/10.1016/j.sab.2020.105924>.
- [22] ABRAHIM A, CANNANAVAN A, KELLY S D. Stable isotope analysis of non-exchangeable hydrogen in carbohydrates derivatised with N-methyl-bis-trifluoroacetamide by gas chromatography. Chromium silver reduction/high temperature conversion-isotope ratio mass spectrometry (GC-CrAg/HTC-IRMS) [J/OL]. Food chemistry, 2020, 318: 126413 [2021-09-14]. <https://doi.org/10.1016/j.foodchem.2020.126413>.
- [23] 唐翠娥, 张莉, 李涛, 等. 果汁中添加外源糖检测技术的研究进展[J]. 食品科学, 2014, 35(9): 306-311. TANG C E, ZHANG L, LI T, et al. Progress in analytical techniques for added sugar in fruit juice[J]. Food science, 2014, 35(9): 306-311 (in Chinese with English abstract).
- [24] REID L M, O' DONNELL C P, DOWNEY G. Recent technological advances for the determination of food authenticity[J]. Trends in food science & technology, 2006, 17(7): 344-353.
- [25] BROOKS C, PARR L, SMITH J M, et al. A review of food fraud and food authenticity across the food supply chain, with an examination of the impact of the COVID-19 pandemic and Brexit on food industry [J/OL]. Food control, 2021, 130: 108171 [2021-09-14]. <https://doi.org/10.1016/j.foodcont.2021.108171>.
- [26] LOHUMI S, LEE S, LEE H, et al. A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration[J]. Trends in food science & technology, 2015, 46(1): 85-98.
- [27] 杨慧, 步雨珊, 易华西. 代谢组学在乳酸菌研究中的应用[J]. 天然产物研究与开发, 2019, 31(8): 1474-1479, 1349. YANG H,

- BU Y S, YI H X. Application of metabolomics in the study of lactic acid bacteria[J]. Natural product research and development, 2019, 31(8): 1474-1479, 1349 (in Chinese with English abstract).
- [28] EISENMANN P, EHLERS M, WEINERT C, et al. Untargeted NMR spectroscopic analysis of the metabolic variety of new apple cultivars[J/OL]. Metabolites, 2016, 6(3): 29 [2021-09-14]. <https://doi.org/10.3390/metabo6030029>.
- [29] ALI K, MALTESE F, ZYPRIAN E, et al. NMR metabolic fingerprinting based identification of grapevine metabolites associated with downy mildew resistance[J]. Journal of agricultural and food chemistry, 2009, 57(20): 9599-9606.
- [30] GUO J, YUE T L, YUAN Y H. Feature selection and recognition from nonspecific volatile profiles for discrimination of apple juices according to variety and geographical origin[J]. Journal of food science, 2012, 77(10): C1090-C1096.
- [31] GAN H H, SOUKOULIS C, FISK I. Atmospheric pressure chemical ionisation mass spectrometry analysis linked with chemometrics for food classification. A case study: geographical provenance and cultivar classification of monovarietal clarified apple juices[J]. Food chemistry, 2014, 146: 149-156.
- [32] RIEDEL T, CLAEYS M, SCHULZ H, et al. Comparative study of Fischer-Tropsch synthesis with H₂/CO and H₂/CO₂ syngas using Fe- and Co-based catalysts[J]. Applied catalysis A: general, 1999, 186(1/2): 201-213.
- [33] YAHAGI T, MASADA S, OSHIMA N, et al. Determination and identification of a specific marker compound for discriminating shrub chaste tree fruit from *Agnus castus* fruit based on LC/MS metabolic analysis[J]. Chemical & pharmaceutical bulletin, 2016, 64(4): 305-310.
- [34] KHALIL M N A, FEKRY M I, FARAG M A. Metabolome based volatiles profiling in 13 date palm fruit varieties from Egypt via SPME GC-MS and chemometrics[J]. Food chemistry, 2017, 217: 171-181.
- [35] PORTO-FIGUEIRA P, FREITAS A, CRUZ C J, et al. Profiling of passion fruit volatiles: an effective tool to discriminate between species and varieties[J]. Food research international, 2015, 77: 408-418.
- [36] FERREIRA L, PERESTRELO R, CÂMARA J S. Comparative analysis of the volatile fraction from *Annona cherimola* Mill. cultivars by solid-phase microextraction and gas chromatography-quadrupole mass spectrometry detection[J]. Talanta, 2009, 77(3): 1087-1096.
- [37] CAPRIOLI G, LOGRIFFO S, CAHILL M G, et al. High-performance liquid chromatography LTQ-Orbitrap mass spectrometry method for tomatidine and non-target metabolites quantification in organic and normal tomatoes[J]. International journal of food sciences and nutrition, 2014, 65(8): 942-947.
- [38] PONTES M, PEREIRA J, CÂMARA J S. Dynamic headspace solid-phase microextraction combined with one-dimensional gas chromatography-mass spectrometry as a powerful tool to differentiate banana cultivars based on their volatile metabolite profile[J]. Food chemistry, 2012, 134(4): 2509-2520.
- [39] LLANO S M, MUÑOZ-JIMÉNEZ A M, JIMÉNEZ-CARTAGENA C, et al. Untargeted metabolomics reveals specific withanolides and fatty acyl glycoside as tentative metabolites to differentiate organic and conventional *Physalis peruviana* fruits[J]. Food chemistry, 2018, 244: 120-127.
- [40] WANG S Y, CHEN C T, SCIARAPPA W, et al. Fruit quality, antioxidant capacity, and flavonoid content of organically and conventionally grown blueberries[J]. Journal of agricultural and food chemistry, 2008, 56(14): 5788-5794.
- [41] JANDRIĆ Z, ROBERTS D, RATHOR M N, et al. Assessment of fruit juice authenticity using UPLC-QTOF MS: a metabolomics approach[J]. Food chemistry, 2014, 148: 7-17.
- [42] 刘哈璐, 张九凯, 韩建勋, 等. 基于 UPLC-QTOF-MS 代谢组学技术的 NFC 和 FC 橙汁差异成分比较[J]. 食品科学, 2021, 42(6): 229-237. LIU H L, ZHANG J K, HAN J X, et al. Analysis of differential composition between not from concentrate and from concentrate orange juices using UPLC-QTOF-MS-based metabolomics[J]. Food science, 2021, 42(6): 229-237 (in Chinese with English abstract).
- [43] EKER M E, AABY K, BUDIC-LETO I, et al. A review of factors affecting anthocyanin bioavailability: possible implications for the inter-individual variability[J/OL]. Foods, 2019, 9(1): 2 [2021-09-14]. <https://doi.org/10.3390/foods9010002>.
- [44] 崔燕, 谭卓, 宣晓婷, 等. 天然复合水蜜桃果汁配方优化及超高压对果汁品质的影响[J]. 食品工业科技, 2021, 42(6): 151-158, 165. CUI Y, TAN Z, XUAN X T, et al. Formula optimization of natural compound honey peach juice and the effects of high hydrostatic pressure on its quality[J]. Science and technology of food industry, 2021, 42(6): 151-158, 165 (in Chinese with English abstract).
- [45] WISTAFF E A, BELLER S, SCHMID A, et al. Chemometric analysis of amino acid profiles for detection of fruit juice adulterations: application to verify authenticity of blood orange juice[J/OL]. Food chemistry, 2021, 343: 128452 [2021-09-14]. <https://doi.org/10.1016/j.foodchem.2020.128452>.
- [46] JANDRIĆ Z, ISLAM M, SINGH D K, et al. Authentication of Indian citrus fruit/fruit juices by untargeted and targeted metabolomics[J]. Food control, 2017, 72: 181-188.
- [47] XU L, XU Z Z, WANG X, et al. The application of pseudotargeted metabolomics method for fruit juices discrimination[J/OL]. Food chemistry, 2020, 316: 126278 [2021-09-14]. <https://doi.org/10.1016/j.foodchem.2020.126278>.
- [48] ARBONA V, IGLESIAS D J, GÓMEZ-CADENAS A. Non-targeted metabolite profiling of citrus juices as a tool for variety discrimination and metabolite flow analysis[J]. BMC plant biology, 2015, 15(1): 1-16.
- [49] CUEVAS F J, PEREIRA-CARO G, MORENO-ROJAS J M, et al. Assessment of premium organic orange juices authenticity using HPLC-HR-MS and HS-SPME-GC-MS combining data

- fusion and chemometrics[J]. Food control, 2017, 82: 203-211.
- [50] LI S J, HU Y X, LIU W, et al. Untargeted volatile metabolomics using comprehensive two-dimensional gas chromatography-mass spectrometry : a solution for orange juice authentication[J/OL]. Talanta, 2020, 217: 121038[2021-09-14]. <https://doi.org/10.1016/j.talanta.2020.121038>.
- [51] XU L, XU Z Z, KELLY S, et al. Integrating untargeted metabolomics and targeted analysis for not from concentrate and from concentrate orange juices discrimination and authentication[J/OL]. Food chemistry, 2020, 329: 127130[2021-09-14]. <https://doi.org/10.1016/j.foodchem.2020.127130>.
- [52] KAMILOGLU S. Authenticity and traceability in beverages[J]. Food chemistry, 2019, 277: 12-24.
- [53] SANTOS P M, PEREIRA-FILHO E R, RODRIGUEZ-SAONA L E. Rapid detection and quantification of milk adulteration using infrared microspectroscopy and chemometrics analysis[J]. Food chemistry, 2013, 138(1): 19-24.
- [54] 俞邱豪. 基于代谢组学的小浆果汁真伪鉴别研究[D]. 杭州: 浙江大学, 2017: 85. YU Q H. Research on the authentication of berry fruit juice using metabonomics technology[D]. Hangzhou: Zhejiang University, 2017: 85 (in Chinese with English abstract).

Progress of fruit juices authenticity technology based on machine learning and web crawler algorithm

SUN Yong¹, WANG Yixiao¹, XUAN Xiaoting², ZHANG Xiyue¹,
DENG Wenyi³, JIN Huan¹, LING Jiangang²

1. Department of Chemical and Environmental Engineering,
University of Nottingham Ningbo, Ningbo 315100, China;

2. Academy of Agricultural Sciences of Ningbo City / National Professional
Research and Development Center of Vegetable Processing Technology/
Key Laboratory of Agricultural Products Preservation Engineering of Ningbo City,
Ningbo 315400, China;

3. Hunan Yipin-dongfang Biotechnology Co. Ltd, Changsha 410000, China

Abstract Fruit juices authenticity technology is becoming a study hotspot in the field of global food nutrition and safety, and its development is very important to control the quality and safety of fruit juice. Combining machine learning with web crawler algorithm retrieval technology, this article reviews the progress and application of fruit juices authenticity technology. It emphasized the progress of fruit juices authenticity technology based on the non-targeted/targeted metabolomics. Combining the development trend of multidisciplinary and computer technology, the trend of coupling development of detection methods, development of supervised machine learning statistical tools, and unsupervised authenticity was predicted. It will provide a reference for further in-depth study in the field of fruit juice authentication.

Keywords fruit juice authenticity; fruit juice adulteration; web crawler; machine learning; non-targeted/targeted biomarker method; metabolomics

(责任编辑:赵琳琳)