

机器学习与多组学结合推动精准营养的研究进展

王璐^{1,2,3} 王宇^{2,3,4} 曾俊^{2,3,4} 陈伟^{4,5} 江华^{2,3,4,5}

【摘要】 在大数据的时代背景下,精准营养已逐渐成为临床营养研究的一个重要趋势。精准营养综合了基因组学、代谢组学、表型组学等多组学的研究理论和方法,并引入机器学习等计算科学技术,使得人们对各种疾病/病理生理条件下营养代谢紊乱的认识达到了新高度,也为发现新的营养干预靶点以及干预的模式提供了全新范式。本文就精准营养中机器学习模型建立的数据准备及算法选择方法,结合机器学习在多组学研究中的应用成果进行综述。在大数据背景下,应用机器学习的精准营养研究不仅能够为每例患者营养支持方案制定策略提供科学依据,同时还可以探索各个营养素之间的联系甚至因果关系,因此机器学习是精准营养研究中必不可少的技术手段。然而,本文也提出目前机器学习在精准营养中的应用受限于数据的异质性高、算法的特异性低等挑战,还有很多未知领域需要探索。

【关键词】 营养基因组学; 精准医学; 机器学习; 深度学习; 代谢组学

Research progress of precision nutrition promoted by the combination of machine learning and multi-omics Wang Lu^{1,2,3}, Wang Yu^{2,3,4}, Zeng Jun^{2,3,4}, Chen Wei^{4,5}, Jiang Hua^{2,3,4,5}. ¹ Graduate School of North Sichuan Medical College, Nanchong 637000, China; ² Institute of Emergency and Disaster Medicine, ³ Department of First-Aid Centre, Sichuan Academy of Medicinal Sciences, Sichuan Provincial People's Hospital, Chengdu 610072, China; ⁴ Sichuan Provincial Emergency and Critical Care Clinical Research Center, Chengdu 610072, China; ⁵ Department of Parenteral and Enteral Nutrition, Peking Union Medical College Hospital, Beijing 100730, China

Corresponding author: Jiang Hua, Email: cdjianghua@qq.com

【Abstract】 With the development of clinical big data, precision nutrition has gradually become an important trend in clinical nutrition research. Precision nutrition integrates the research theories and methods of genomics, metabolomics, phenomics and other omics, and introduces computational science and technology such as machine learning, which makes people's understanding of nutritional metabolic disorders under various diseases/pathophysiological conditions reach a new height, and also provides a new paradigm for the discovery of new nutritional intervention targets and intervention modes. This paper reviews the data preparation and algorithm selection methods of machine learning model building in precision nutrition, combined with the application of machine learning in multi-omics research. Under the background of big data, precision nutrition research with the application of machine learning can not only provide scientific basis for the formulation of nutrition support plan for each patient, but also explore the relationship and even causal relationship between various nutrients. Therefore, machine learning is an essential technical means in precision nutrition research. However, this paper also points out that the current application of machine learning in precision nutrition is limited by high heterogeneity of data and low specificity of algorithms, and there are still many unknown areas to explore.

【Key words】 Nutrigenomics; Precision medicine; Machine learning; Deep learning; Metabolomics

经典营养学的发展为精准营养理论的提出奠定了扎实

DOI:10.3877/cma.j.issn.1673-9450.2022.06.014

基金项目:四川省重点研发项目(2019YFS0534,2020YFS0392,2021YFS0378)

作者单位:637000 南充,川北医学院研究生院¹;610072 成都,四川省医学科学院·四川省人民医院急诊医学与灾难医学研究所²,急救中心³;610072 成都,四川省急危重症临床研究中心⁴;100730 北京协和医院临床营养科⁵

通信作者:江华, Email: cdjianghua@qq.com

的基础,也为当代大数据背景下精准营养研究的发展创造了条件。20世纪初维生素的发现大大促进了人们对代谢物质与人类健康关系的认知。在过去的几十年中,营养学界的科学家们在了解营养如何与健康相互作用方面取得了显著进展。近年来,随着经济和科技的共同发展,公众对个人健康状况日益关注,传统“一刀切”的营养定制方案已经无法满足大众的需求。早在2019年的《食品营养与健康的十大主要趋势(10 Key Trends in Food, Nutrition & Health 2019)》中就指出,营养膳食方案的“碎片化和个性化”位居第六,是目前关注度最高的趋势^[1]。2020年以来,在新型冠状病毒肺炎

炎疫情的冲击下,公众对自身健康的重视程度于短时间内剧增,精准营养也迅速成为了营养研究中的顶流。

精准营养又称为个性化营养,是指综合考虑遗传因素、饮食习惯和膳食模式、昼夜节律、健康状况、社会经济和社会心理特征、饮食环境、体力活动和微生物组等因素,制定与个体和群体健康相关的全面、动态营养建议的一套研究和实践范式^[2]。近十年以来,在大数据时代背景下,真实世界临床研究的提出使观察性研究逐渐上升至证据金字塔的顶端,为精准营养的进一步发展提供了契机^[3]。与此同时,研究者们也逐渐意识到,精准营养研究需要采集大量数据资料。因此,在研究中需要依赖于规范的数据科学手段,实现完善的数据采集,并进行数据清洗以建立可应用于临床研究的数据集。建立这样的数据集后,还需要对数据进行降维处理。进而实现对数据进行分类、聚类、减噪等以实现特定用途的最终目的。如特定营养方式干预后对特定疾病患者的临床结局是否有所改善(大多数研究方式是对患者结局进行预测,并观察干预手段对预测结果的影响)。例如,2015年Cell发表了一篇重要的研究,该研究将糖尿病患者的肠道菌群数据、个性化膳食干预方案数据相结合,采用机器学习算法对具有不同肠道菌群表型的患者采用不同膳食干预后的血糖反应进行了预测,研究表明,基于机器学习算法和微生物组学的干预效果优于传统统计学方式;该研究已被公认为是实现疾病精准营养干预的一个成功案例^[4]。

由此,人们认识到,欲实现精准营养研究,必须采用以机器学习为基础的数据科学,结合以基因组学、微生物组学、代谢组学等多组学方式探索各种病理生理条件下特定膳食模式/营养素对个体患者的干预效果,最终方能达到个体化营养支持治疗的良好疗效^[5]。

本文首先介绍机器学习的数据准备与算法选择,随后具体阐述机器学习在多组学研究中的应用现状及推动作用。

一、机器学习的数据准备

我国的医疗信息化建设始于20世纪90年代,经过近30年的发展,各类临床资料都实现了在互联网环境下以多种数字化格式进行存储、传输的信息化革命。目前临床数据的总量可称海量,且数据量仍呈指数级增长。医疗信息化的持续发展,产生并不断积累大量的数据,庞大的数据中汇聚各类信息,蕴含着极大的社会价值。海量数据只有通过有效地采集及专业化处理,才能为其应用提供有价值的参考和佐证。从医疗大数据中提取信息的能力迅速成为战略性发展方向和要求^[6-7]。目前临床医疗大数据的采集面临以下3个问题,即病历文本问题、数据挖掘流程问题以及数据质量问题。

(一)病历文本问题

病历文本是患者病史、家族史、症状等信息的重要来源,也是医师对患者进行初步诊断的重要依据。然而病例文本复杂多变,难以被计算机直接识别。自然语言处理(natural language processing, NLP)是指利用人类交流所使用的自然语言与机器进行交互通讯的技术,是计算机科学领域与人工智能领域中的一个重要方向。它能够将文本结构化以便于研究者进行数据采集。目前,研究显示,结合深度学习方法的NLP在近年中逐年增加,其中临床医疗领域是NLP应用的主要方向,占比89.2%^[8]。同年,Chen等^[9]开发了基于电子病历的

临床信息抽取和肝细胞癌的NLP系统,结果表明中文NLP系统在临床研究中具有潜在的应用价值。目前,已有研究证实NLP可对各类影像的综合信息的数据进行分析,并且在经过验证后,可得到预测模型,形成适合本地情况的临床决策支持系统,以此应用到计算机医嘱系统中去^[10]。

(二)数据挖掘流程

医院的信息系统十分复杂,就一家三级甲等医院而言,院级系统通常10~20个。如医院信息系统、临床信息系统、实验室信息系统、放射信息管理系统、影像归档和通信系统、病案系统等。若以临床各专科而论,则信息系统可多达数百个。因此数据挖掘过程面临的首要挑战就是从上述纷繁复杂的系统中获取符合临床研究需求的数据^[11]。目前仍是未突破的难点。

(三)数据质量

1. 数据质量评估:数据质量是使用临床数据的重要先决条件^[12]。许多研究表明,电子病历数据包含可能影响研究结果的错误,且电子病例数据普遍存在缺失、错误、无效、不完整、不一致等质量问题^[13]。因此,数据采集的过程中,必须首先对数据进行评估。目前,人们一般认为数据评估过程可分为6个步骤:(1)收集评估需求;(2)确定和采集评估数据;(3)建立评估需求与评估数据之间的映射关系;(4)提出质量评估标准;(5)执行数据质量评估;(6)分析评估结果,最终判定该数据集是否适用于研究目的。

2. 数据清洗:数据清洗是保证机器学习的数据质量的重要步骤,指清洗混杂数据,其包括将重复数据剔除,对错误的数据予以纠正,补全数据并且提供其一致性的过程。而机器学习技术的应用主要是在数据补全方面。临床数据收集过程繁琐且复杂,难免会出现无效值以及缺失值。对于缺失 $\geq 30\%$ 的特征或无效值,予以排除。而对于缺失比例小于30%的特征或样本,目前常用的插值方法很多,如均值插入法、朗格拉日插值法、牛顿插值法、三次样条插值法、概率主成分分析法及多重插值法等。对于静态数据而言,王宇^[14]研究发现,概率主成分分析法在数据补全的过程中相较其他几种方法具有更好的稳定性。而对于动态变化的数据(如重症监护室的患者生命体征监测数据常以分钟或秒记),研究者们常需要进行连续采集。对于这类数据,采用三次样条插值法能够很好地避免插值处数据将连续数据离断的可能性,因其插值的本质就是先用一条光滑的曲线将数据点串联,再在缺失数据点投影到拟合的曲线上进行取值。Lei等^[15]于2020年发表基于深度神经网络的人白蛋白预测模型研究中就采用三次样条插值法恢复数据连续性,并证明基于深度神经网络的动态系统模型可以用来精准预测时间序列数据变化。

3. 数据降维:使用合适的降维方法能够去除对模型结果产生混淆的数据,有效提高模型的精度,是数据准备中较重要的步骤,也是时间成本最高的部分。而对降维方法的选择对于机器学习模型建立研究极其重要,不当的降维方法不仅无法排除产生偏倚的混淆因素,还可能失去数据矩阵中的重要信息^[16]。根据降维方式的不同,数据降维方法分为特征提取和特征选择。从已有的临床研究经验看,营养代谢类研究中使用的数据之间往往存在较强的相关性,因此使用特征

选择的方法可能造成重要信息的丢失。在这种情况下,较好的选择是使用特征提取方法。潜在结构判别分析和偏最小二乘判别分析均属于有监督的特征提取方法,常用于识别2组之间的差异和具有较大鉴别力的变量,这2种方法在面临高危灾难挑战的营养基因组学以及营养代谢组学研究中较为常用,而在数据维度相对较低的营养流行病学以及营养素研究中很少用到^[17]。相对而言,主成分分析是一种因子分析方法,作为目前最常用的无监督数据降维方法,常用于研究膳食营养研究的模式^[18]。

二、机器学习模型算法选择

Kirk等^[19]于2021年发表的一篇系统评价显示,目前营养相关机器学习研究中89%都使用了有监督学习方法。Jones等^[20]的研究是为数不多的采用无监督机器学习方法对数据进行分类的研究,其采用K-means从腕带数据中对患者活动进行分类,判断患者进行的运动类型,该研究证明了应用无监督机器学习方法在建立可用于独立数据集的聚类模型中的可行性。理论上来说,无监督机器学习方法能够很好地节省人力,且其结果相比于有监督学习更加客观。但现有的研究结果并没有显示出无监督机器学习降维的效果优于有监督学习方法。Kirk等^[19]的系统评价还发现分类模型在精准营养研究中占比66%,其中最常用的分类方法有随机森林、支持向量机、K近邻查询等,其中随机森林、支持向量机在所有机器学习方法中使用频率最高。此外,深度学习方法作为更接近于机器学习的一种建模方法,在营养研究中的应用却很少,仅在食物图像识别的研究中显示出较好的结果。这可能是源于简单的机器学习模型相较于深度学习模型具有更强的可解释性,能够提高机器学习工具的可用性。

需要着重指出一点的是,目前对于机器学习算法的研究,尚缺乏很好的规范性。2021年,Deng等^[21]发表了一个系统评价,对于脓毒症领域的机器学习/深度学习研究进行了全面评估,该研究发现即使研究目的相同,但是数据来源、数据预处理方法、特征工程和机器学习算法不同会对预测结果产生极大的影响,使得很多研究结果实际上无法外推应用。简而言之,机器学习的算法很多,但并不存在普适性的“最优算法”,而是应该根据不同的疾病特征、研究目的和数据集,进行学习算法的个性化设计。

三、机器学习在精准营养研究中的应用

目前机器学习在精准营养中的应用主要是在营养基因组学、营养代谢组学以及营养临床研究3个方面。其中营养基因组学是以分子生物学技术为基础,应用DNA芯片、蛋白质组学等技术来研究营养素对机体基因的转录、翻译表达及代谢机制的科学,是基因层面的研究。营养代谢组学是一门系统地研究不同状态下人群机体代谢能力与各类营养素或营养支持方案相互影响的学科,也是营养学研究中机器学习应用最早的领域。

(一)营养基因组学研究

2002年,第一届国际营养基因组学会议推动了营养基因组学的发展。营养基因组学研究通过揭示营养素作用机制、探索发现更多用来评价营养状况的分子标志物,最终实现个体精细化营养方案的制定。2020年,Khorraminezhad等^[22]在考察了16项营养与多组学学科的联合研究后提出,

由于目前的营养研究大多结合了多组学的数据,因此更需要使用机器学习技术等最新的数据分析手段来补充传统的统计分析的不足,以充分解释营养对健康和疾病的影响。

营养基因组学研究早在10余年前就受到了广泛关注,如今随着大量的基因位点被发现,机器学习在营养基因组学研究中逐渐显示出其优越性。但目前营养基因组学相关的研究较少。2022年,Lee等^[23]使用机器学习基于全基因组和表观基因组探究基因-基因和基因-饮食之间的相互作用,以预测肥胖发生的风险,该研究对体重指数和402 793个单核苷酸多态性、415 202个DNA甲基化位点和397个膳食之间的相互作用进行了全基因组和表观基因组范围的联合扫描,采用1 573个样本组成的训练集以及394个样本组成的测试集建立预测模型。在比较几种机器学习算法后,该研究发现随机梯度提升模型提供了最佳的肥胖预测准确度(测试样本中特征曲线下面积为0.72)。

此外,现有研究中很多基因位点存在重复性差的问题,即便是能够完全重复的位点也仅能解释目前一半左右的性状。至今为止,全基因组关联研究的基因位点检测方法依然无法满足机器学习的要求,在寻求机器学习更优算法的同时,仍需更加先进的基因位点检测手段加持。

(二)营养代谢组学研究

由于个体对同一营养物质的代谢能力相差迥异,其代谢组也会出现与之相对应的差异。如2017年,Li等^[24]基于液相色谱-质谱联用方法对胃癌术后患者代谢进行分析,结合随机对照试验,发现对于胃癌术后患者补充复合维生素可缓解胃癌患者创伤性应激,改善从分解代谢向合成代谢的代谢转变。此外,Tan等^[25]基于核磁共振谱的代谢组学研究发现,对于重症患儿,在肠外营养中增加多种微量元素补充,可以改善肝肾功能。因此,可以通过机器学习方法对其代谢产物进行分类,从而建立不同人群与其对应的代谢能力或途径之间的联系,实现个体化营养支持^[26-27]。

近年来有更多研究关注到营养支持与患者肠道内菌群的种类和分布之间的相关性。早在1999年,就有研究显示营养支持有助于机体维护肠黏膜的免疫屏障功能^[28]。随后,更多研究结果提示合理的营养支持方案能够刺激肠上皮再生,改善肠道局部血流灌注^[29-31]。2019年一篇关于食物和肠道菌群对于健康人群和患者相互影响的综述中明确提出,由于饮食、宿主和微生物群之间的相互作用是因人而异的,因此了解这种复杂的相互作用网络需要借助机器学习算法和人工智能等新颖的计算机技术,对大型数据集进行创新型地分析研究^[32]。同年,一项采用多变量分析和机器学习微生物组模型的横断面研究探索肠道微生物群与2型糖尿病之间的相互影响,结果表明糖尿病患者胰岛素抵抗与微生物变异密切相关,同时也进一步证明了在开发预防或延迟2型糖尿病精准医学方法时,肠道微生物群是一个重要的可改变因素^[33]。随后,2021年发表的一篇队列研究就根据2型糖尿病相关肠道微生物组特征,构建了具有已识别特征的微生物组风险评分,并且通过动物实验进一步证实了微生物组风险评分与2型糖尿病之间的关系^[34]。2022年,Wang等^[35]发表的一篇系统评价也发现使用益生菌调节肠道菌群是一种有效且安全的营养疗法,可减少重症患者的脓毒症并

发病。此外,吴桐等^[36]基于美国肠道计划公开数据,应用机器学习方法进行了临床验证,对肥胖和健康人群肠道菌群进行比较分析,研究筛选得出1 655个体重指数为18.5~25.0 kg/m²的健康成年人,和898个体重指数>30 kg/m²肥胖成年人的肠道菌群数据,结果显示2组人群在肠道菌群结构上差异无统计学意义,但在菌群的门属水平上差异有统计学意义;研究中基于肥胖人群与健康人群间菌群差异建立肥胖人群的预测模型,其中梯度提升回归树预测效果最好,测试集特征曲线下面积值达到0.725。2021年也有研究使用机器学习方法证实个体肠道微生物群配置可能调节宿主代谢并改变个体对特定食物干预的反应^[37]。

(三)营养临床研究

机器学习在临床研究中的应用方面很广,包括营养流行病学以及营养素等方面。为了保证模型建立的准确性,研究者往往需要更多的时间完成数据清洗和数据预处理的工作。因此目前这部分的研究仍处于探索阶段。2021年Sak和Suchodolska^[38]发表了一篇关于营养科学研究中的人工智能应用的综述,纳入了1987年至2020年间的55项研究,这55项研究中有42项是营养素研究,13项是营养流行病学研究。

营养流行病学是指应用流行病学的方法研究人群营养与健康及疾病关系的科学。Morgenstern等^[39]于2021年在*Advances in Nutrition*发表了一篇综述,强调了大数据和机器学习在营养流行病学中的重要性。同年,Oliveira Chaves等^[40]发表了一篇研究机器学习技术用于评估人群食物摄入情况的系统评价,纳入了36项研究,系统评价结果显示,目前在营养流行病学的研究中最常用的机器学习算法类型是有监督机器学习方法,且大部分研究使用的是分类算法,其中最常用的模型是决策树。

在传统营养学研究的阶段,众多专家学者已针对单个营养素与疾病的关系进行了系统的研究。但伴随精准医学的提出,研究者们开始意识到,营养素和食物可以以协同或对抗的方式相互作用。2022年2月,一篇由加拿大研究团队发表的机器学习研究探索了营养素与心血管疾病风险之间的关系,同时建立模型并评估其预测性能,该研究纳入了12 130例患者,预测变量包括61个营养变量和14个社会经济、人口、心理和行为变量,应用随机森林算法建立模型,预测准确率达到0.821,该研究显示维生素和矿物质补充剂、咖啡因和酒精的使用是预测心血管疾病最重要的营养变量。其中补充剂的使用与风险降低相关,而咖啡因与风险增加相关,酒精与风险呈U型关联,该研究揭示了饮食与心血管疾病之间的不同关联,并且成功实现了多个营养素预测模型的建立^[41]。因此,食物、食物组或饮食模式的研究也许比单个营养素的研究更具有临床意义。

营养流行病学以及临床研究中由于数据复杂多变,常常面临多组不同类型数据的处理,而机器学习方法在此类数据分析中极具优势。从目前的研究看来,机器学习研究中关于数据清洗和数据降维部分的方法学显得尤其重要。“干净”的数据能够大幅提高机器学习模型建立的准确性,使模型更稳健、更具外推性。但目前机器学习在临床营养学的应用仍在发展的初级阶段,从现有研究看来其还有更多发展潜力。

四、总结和展望

美国卫生和公众服务部2020年5月份出台的《美国国立卫生研究院2020-2030营养研究战略计划》明确提出了精准营养的概念。同年12月我国提出了《“健康中国2030”规划纲要》和《国民营养计划(2017-2030年)》,将国民营养问题上升到国家战略层面。伴随着公众对身体素质健康的重视程度逐年上升,“个性化膳食”的呼声高涨,精准营养已是大势所趋。在大数据背景下,应用机器学习的精准营养研究不仅能够为每例患者营养支持方案制定策略提供科学依据,同时还可以探索各个营养素之间的联系甚至因果关系。但目前机器学习在精准营养中的应用受限于数据的异质性高、算法的特异性低等挑战,还有很多未知领域需要探索。

参 考 文 献

- [1] Businesswire. 10-Key-Trends-Food-Nutrition-Health-2019 [EB/OL]. (2019-07-25) [2022-05-13]. <https://www.businesswire.com/news/home/20190725005730/en/>.
- [2] 刘琰,陈伟. 精准营养新定义:理念与落实[J]. 中华预防医学杂志, 2022, 56(2): 151-153.
- [3] 何俏,时景璞. 临床真实世界研究中的实验性研究设计[J]. 中华流行病学杂志, 2018, 39(4): 519-523.
- [4] Zeevi D, Korem T, Zmora N, et al. Personalized Nutrition by Prediction of Glycemic Responses [J]. *Cell*, 2015, 163(5): 1079-1094.
- [5] Ferguson LR, De Caterina R, Görman U, et al. Guide and Position of the International Society of Nutrigenetics/Nutrigenomics on Personalised Nutrition: Part 1 -Fields of Precision Nutrition [J]. *J Nutrigenet Nutrigenomics*, 2016, 9(1): 12-27.
- [6] 马国耀,孙勇韬,马玉玲. 数据采集模板化技术在医疗大数据集成建设中的应用[J]. 中国卫生信息管理杂志, 2016, 13(4): 414-416.
- [7] 周光华,李岳峰. 数据挖掘技术在卫生统计信息工作中的应用研究[J]. 中国卫生信息管理杂志, 2012, 9(6): 82-86.
- [8] Wu S, Roberts K, Datta S, et al. Deep learning in clinical natural language processing: a methodical review[J]. *J Am Med Inform Assoc*, 2020, 27(3): 457-470.
- [9] Chen L, Song L, Shao Y, et al. Using natural language processing to extract clinically useful information from Chinese electronic medical records[J]. *Int J Med Inform*, 2019, 124: 6-12.
- [10] Patel TA, Puppala M, Ogunti RO, et al. Correlating mammographic and pathologic findings in clinical decision support using natural language processing and data mining methods[J]. *Cancer*, 2017, 123(1): 114-121.
- [11] 阮彤,高炬,冯东雷,等. 基于电子病历的临床医疗大数据挖掘流程与方法[J]. 大数据, 2017, 3(5): 83-98.
- [12] Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research[J]. *J Am Med Inform Assoc*, 2013, 20(1): 144-151.
- [13] 熊兴江,马敬东,徐承中. 电子健康档案数据质量评估与治理的综述研究[J]. 中国卫生信息管理杂志, 2018, 15(6): 637-642.
- [14] 王宇. 基于机器学习建立创伤代谢组学快速精准分析技术的研究[D]. 成都: 电子科技大学, 2019.

- [15] Lei C, Wang Y, Zhao J, et al. A patient specific forecasting model for human albumin based on deep neural networks[J]. *Compute Methods Programs Biomed*, 2020, 196: 105555.
- [16] Tchagna Kouanou A, Mih Attia T, Feudjio C, et al. An Overview of Supervised Machine Learning Methods and Data Analysis for COVID-19 Detection[J]. *J Healthc Eng*, 2021, 2021: 4733167.
- [17] Westerhuis JA, van Velzen EJ, Hoefsloot HC, et al. Multivariate paired data analysis: multilevel PLSDA versus OPLSDA[J]. *Metabolomics*, 2010, 6(1): 119-128.
- [18] Zhang F, Tapera TM, Gou J. Application of a new dietary pattern analysis method in nutritional epidemiology[J]. *BMC Med Res Methodol*, 2018, 18(1): 119.
- [19] Kirk D, Catal C, Tekinerdogan B. Precision nutrition: A systematic literature review[J]. *Comput Biol Med*, 2021, 133: 104365.
- [20] Jones P, Mirkes EM, Yates T, et al. Towards a Portable Model to Discriminate Activity Clusters from Accelerometer Data[J]. *Sensors (Basel)*, 2019, 19(20): 4504.
- [21] Deng HF, Sun MW, Wang Y, et al. Evaluating machine learning models for sepsis prediction: A systematic review of methodologies[J]. *iScience*, 2021, 25(1): 103651.
- [22] Khorraminezhad L, Leclercq M, Droit A, et al. Statistical and Machine-Learning Analyses in Nutritional Genomics Studies[J]. *Nutrients*, 2020, 12(10): 3140.
- [23] Lee YC, Christensen JJ, Parnell LD, et al. Using Machine Learning to Predict Obesity Based on Genome-Wide and Epigenome-Wide Gene-Gene and Gene-Diet Interactions[J]. *Front Genet*, 2022, 12: 783845.
- [24] Li ZJ, Chen W, Jiang H, et al. Effects of Postoperative Parenteral Nutrition Enhanced by Multivitamin on Metabolic Phenotype in Postoperative Gastric Cancer Patients[J]. *Mol Nutr Food Res*, 2018, 62(12): e1700757.
- [25] Tan Q, Wang Y, Zhang G, et al. The metabolic effects of multi-trace elements on parenteral nutrition for critically ill pediatric patients: a randomized controlled trial and metabolomic research[J]. *Transl Pediatr*, 2021, 10(10): 2579-2593.
- [26] Shen X, Wang C, Liang N, et al. Serum Metabolomics Identifies Dysregulated Pathways and Potential Metabolic Biomarkers for Hyperuricemia and Gout[J]. *Arthritis Rheumatol*, 2021, 73(9): 1738-1748.
- [27] Ulaszewska MM, Weinert CH, Trimigno A, et al. Nutrimetabolomics: An Integrative Action for Metabolomic Analyses in Human Nutritional Studies[J]. *Mol Nutr Food Res*, 2019, 63(1): e1800384.
- [28] Fan J, Meng Q, Guo G, et al. Effects of early enteral nutrition supplemented with arginine on intestinal mucosal immunity in severely burned mice[J]. *Clin Nutr*, 2010, 29(1): 124-130.
- [29] Xu D, Lu Q, Deitch EA. Elemental diet-induced bacterial translocation associated with systemic and intestinal immune suppression[J]. *JPEN J Parenter Enteral Nutr*, 1998, 22(1): 37-41.
- [30] Okamoto K, Fukatsu K, Hashiguchi Y, et al. Lack of preoperative enteral nutrition reduces gut-associated lymphoid cell numbers in colon cancer patients: a possible mechanism underlying increased postoperative infectious complications during parenteral nutrition[J]. *Ann Surg*, 2013, 258(6): 1059-1064.
- [31] Heneghan AF, Pierre JF, Tandee K, et al. Parenteral nutrition decreases paneth cell function and intestinal bactericidal activity while increasing susceptibility to bacterial enteroinvasion[J]. *JPEN J Parenter Enteral Nutr*, 2014, 38(7): 817-824.
- [32] Danneskiold-Samsøe NB, Dias de Freitas Queiroz Barros H, Santos R, et al. Interplay between food and gut microbiota in health and disease[J]. *Food Res Int*, 2019, 115: 23-31.
- [33] Wu H, Tremaroli V, Schmidt C, et al. The Gut Microbiota in Prediabetes and Diabetes: A Population-Based Cross-Sectional Study[J]. *Cell Metab*, 2020, 32(3): 379-390. e3.
- [34] Gou W, Ling CW, He Y, et al. Interpretable Machine Learning Framework Reveals Robust Gut Microbiome Features Associated With Type 2 Diabetes[J]. *Diabetes Care*, 2021, 44(2): 358-366.
- [35] Wang K, Zeng Q, Li KX, et al. Efficacy of probiotics or synbiotics for critically ill adult patients: a systematic review and meta-analysis of randomized controlled trials[J]. *Burns Trauma*, 2022, 10: tkac004.
- [36] 吴桐, 王鸿超, 陆文伟, 等. 肥胖人群肠道菌群特征分析及机器学习模型[J]. *微生物学通报*, 2020, 47(12): 4328-4337.
- [37] Wang S, Zhang L, Wang D, et al. Gut Microbiota Composition is Associated with Responses to Peanut Intervenion in Multiple Parameters Among Adults with Metabolic Syndrome Risk[J]. *Molr Nutr Food Res*, 2021: 2001051.
- [38] Sak J, Suchodolska M. Artificial Intelligence in Nutrients Science Research: A Review[J]. *Nutrients*, 2021, 13(2): 322.
- [39] Morgenstern JD, Rosella LC, Costa AP, et al. Perspective: Big Data and Machine Learning Could Help Advance Nutritional Epidemiology[J]. *Adv Nutr*, 2021, 12(3): 621-631.
- [40] Oliveira Chaves L, Gomes Domingos AL, Louzada Fernandes D, et al. Applicability of machine learning techniques in food intake assessment: A systematic review[J]. *Crit Rev Food Sci Nutr*, 2021, 29: 1-18.
- [41] Morgenstern JD, Rosella LC, Costa AP, et al. Development of Machine Learning Prediction Models to Explore Nutrients Predictive of Cardiovascular Disease Using Canadian Linked Population-Based Data[J]. *Appl Physiol Nutr Metab*, 2022, 47(5): 529-546.

(收稿日期:2022-09-05)

(本文编辑:蒋妍妍)

王璐, 王宇, 曾俊, 等. 机器学习与多组学结合推动精准营养的研究进展[J/CD]. 中华损伤与修复杂志(电子版), 2022, 17(6): 540-544.