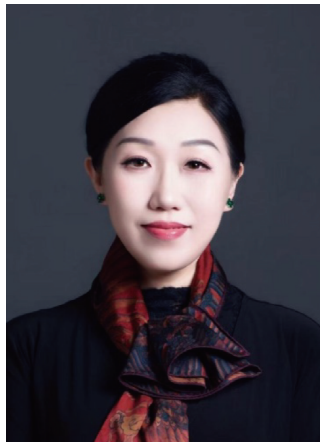


DOI: 10.3724/cbls.2026027

CSTR: 32203.14.cbls.2026027

文章编号: 1004-0374(2026)01-0001-17

· 特约综述 ·



王慧,上海交通大学二级教授、特聘教授、博士生导师,公共卫生学院院长、数字医学研究院执行院长、医学院单细胞组学与疾病研究中心中方主任,海南国际医学中心数智医学研究院院长,上海市黄浦区侨联主席,上海交通大学医学院侨联主席。国家杰出青年科学基金获得者、国家百千万人才工程入选者、中国青年女科学家奖获得者、国务院特殊津贴专家、吴杨奖获得者、中国科学院百人计划,上海市三八红旗手、五一劳动奖章获得者,上海市优秀学术带头人,上海市领军人才,上海交通大学校长奖获得者。任国务院食品安全委员会专家委员、国家健康科普专家、中国营养科学界首席专家,第一届食品安全国家标准审评委员会微生物分委员会主任委员、中国检验检测学会数智主动健康产业协同创新中心主任,上海市食品安全风险评估专家委员会膳食营养专业委员会主任委员,上海市毒理学会理事长,上海市大数据社会应用研究会数字健康专委会主任委员,上海市侨界专业人士联谊会首任轮值会长等。长期从事主动健康、慢病防控与营养食品安全研究,Elsevier高被引学者,科技部重点专项首席科学家。在*Nature*、*JAMA*等发表论文260余篇,主编专著9部,申请专利60项,成果转化7项,形成国家、行业标准25项,获全国优秀科普微视频作品、上海市健康科普推优选树“杰出人物”、上海市课程思政教学设计展示活动特等奖、上海市优秀教学成果一等奖等。

## 营养大模型的技术架构、应用进展与未来挑战

张成东<sup>1,2#</sup>, 孔浩楠<sup>2#</sup>, 杨元<sup>1</sup>, 闫媛媛<sup>1</sup>, 童天朗<sup>1,2\*</sup>, 王慧<sup>1,2\*</sup>

(1 上海交通大学公共卫生学院,上海 200025; 2 上海交通大学医学院,海南国际医学中心数智医学研究院,琼海 571400)

**摘要:** 营养信息学正由传统基于规则与常规机器学习范式,迈向以大语言模型(large language model, LLM)与多模态大模型(multimodal large language models, MLLM)为核心的新阶段。本文系统综述了2019–2025年间营养大模型领域的研究进展,归纳了视觉-语言对齐、领域知识注入、检索增强生成(retrieval-augmented generation, RAG)及可解释推理等关键架构与训练技术。在此基础上,本文详细梳理了模型在个性化膳食推荐、营养状态评估、疾病营养管理及膳食自动化记录等典型场景的应用现状。此外,本文总结了Nutrition5k、NutriBench等核心数据集与评测基准的演变历程。最后,针对模型可信度、数据隐私、跨文化泛化及临床循证支持等挑战,本文提出未来研究应深度融合临床证据,构建高质量多模态数据体系,并推进人机协同的精准营养服务落地,以提升临床转化价值。

**关键词:** 营养大模型;多模态学习;大语言模型;个性化营养;检索增强生成;评测基准

中图分类号: TP18; R151

文献标识码: A

### Technical architecture, application progress, and future challenges of nutrition foundation models

收稿日期: 2026-01-06; 修回日期: 2026-01-28

基金项目: 国家自然科学基金重点项目(82030099); 国家重点研发计划项目(2022YFD2101500)

#共同第一作者

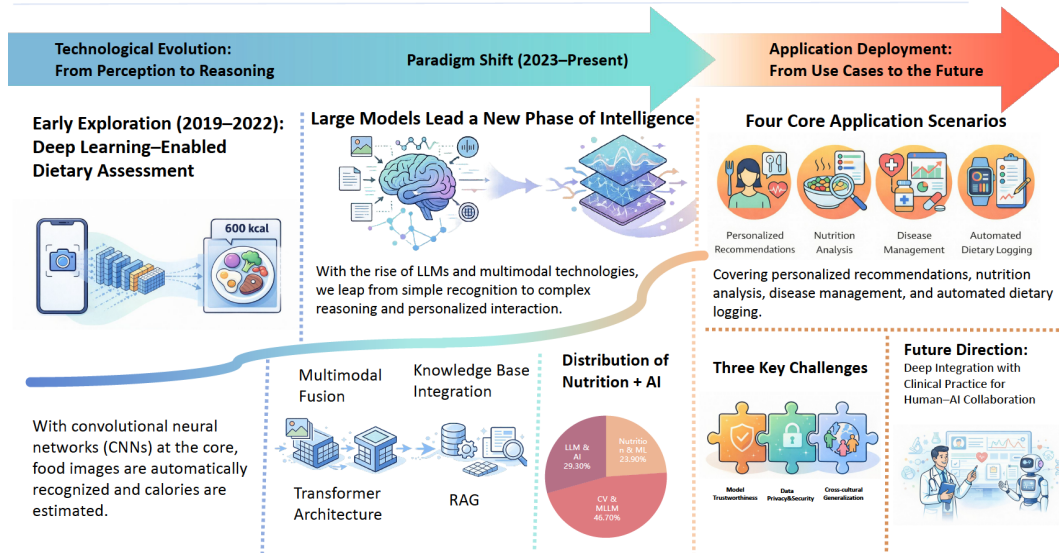
\*通信作者: E-mail: huiwang@shsmu.edu.cn(王慧); 183947@shsmu.edu.cn(童天朗)

ZHANG Cheng-Dong<sup>1, 2#</sup>, KONG Hao-Nan<sup>2#</sup>, YANG Yuan<sup>1</sup>, YAN Yuan-Yuan<sup>1</sup>,  
TONG Tian-Lang<sup>1, 2\*</sup>, WANG Hui<sup>1, 2\*</sup>

(1 School of Public Health, Shanghai Jiao Tong University, Shanghai 200025, China; 2 Institute of Digital and Intelligent Medicine, Hainan International Medical Center, Shanghai Jiao Tong University School of Medicine, Qionghai 571400, China)

**Abstract:** Nutrition informatics has undergone a significant paradigm shift in recent years. Approaches historically grounded in rule-based decision support and classical task-specific machine learning pipelines are increasingly being superseded by an ecosystem centered on large language models (LLMs) and multimodal vision-language foundation models. This review synthesizes researches published between 2019 and 2025, with the objectives of clarifying architectural patterns that enable nutrition-oriented perception and reasoning, summarizing advances and identifying gaps across major application scenarios, and outlining strategic directions for reliable translation research in clinical and public health practice. Based on a systematic analysis of 92 representative studies, we organize the current landscape into three interrelated research trajectories: (1) Vision and multimodal modeling for dietary perception, focusing on food recognition, ingredient parsing, portion estimation, and nutrient prediction from meal images and videos. Recent methodologies increasingly adopt Transformer-based encoders and explicit vision-language alignment, leveraging depth cues and scale calibration to improve robustness under complex real-world conditions. (2) LLM-based nutrition agents for interactive guidance, supporting dietary counseling, meal planning, and health coaching. To mitigate challenges such as hallucinations and numerical inconsistency, current research emphasizes domain adaptation, tool-augmented computation, and retrieval-augmented generation (RAG) to ground model responses in reliable nutrition databases and clinical guidelines. (3) Personalization-oriented hybrid systems, which combine foundation models with structured components—such as knowledge graphs and causal inference frameworks—while integrating individual-level multi-omics signals, biomarkers, and lifestyle data. These systems aim to generate and optimize meal plans under strict constraints of safety, clinical feasibility, and patient adherence. Across these trajectories, interpretability has transitioned from an optional feature to a core system requirement, driven by the needs of clinical accountability and risk auditing. Concurrently, evaluation protocols are expanding from image-centric datasets (e.g., Nutrition5k) to comprehensive benchmarking suites designed for multimodal reasoning. Despite rapid progress, limitations persist regarding model factuality, privacy preservation, and external validity across diverse cuisines and socioeconomic settings. We advocate for evidence-grounded pipelines, standardized multimodal datasets with clinical endpoints, and unified evaluation frameworks spanning accuracy, safety, and bias. Human-in-the-loop deployment remains essential to quantify benefit-risk profiles and facilitate the regulatory adoption of AI-driven nutrition services.

#### Technical Architecture, Application Progress, and Future Challenges of Nutrition Foundation Models



**Key words:** nutrition large language model; multimodal learning; large language model; personalized nutrition; retrieval-augmented generation; evaluation benchmark

营养信息学(nutrition informatics, NI)是利用信 息技术与数据驱动方法解决营养健康问题的交叉学

科领域。在传统范式下,营养研究主要依赖人工饮食记录与实验室生化分析<sup>[1]</sup>。过去十年,大数据与人工智能(artificial intelligence, AI)技术的飞速发展,为营养信息学注入了新的活力。尤其是深度学习技术的引入,使得从海量非结构化饮食数据中自动提取营养信息成为可能<sup>[2]</sup>。2015年,研究者开始尝试利用卷积神经网络(convolutional neural network, CNN)从餐食图像中推断能量摄入等关键营养指标<sup>[3]</sup>;2019年前后,更精细的能量估算方案与系统化建模方法逐渐成为主流探索方向,实现了对复杂菜肴热量的自动化评估<sup>[2,4]</sup>。这些早期工作验证了AI在营养领域的可行性,但同时也暴露了数据标注成本高昂、模型可解释性不足等局限。

近年来,人工智能领域迎来了以大语言模型(large language model, LLM)为核心的新范式<sup>[5]</sup>。LLM是指具有超大规模参数,并在海量通用语料上预训练的深层神经网络(如GPT系列),其具备强大的语义理解、知识整合与逻辑推理能力,在医疗健康领域展现出巨大潜力<sup>[5,6]</sup>。在营养领域引入LLM,旨在利用其卓越的推理能力解决个性化膳食规划、复杂营养咨询等知识密集型任务。例如,多项评估研究显示,ChatGPT在生成膳食计划与健康建议方面具有较高的可靠性与科学性<sup>[7,8]</sup>。与此同时,多模态学习(multimodal learning)的发展促使模型能够同时处理图像、文本、生理信号等多种模态的数据。视觉-语言预训练模型(如CLIP、ALIGN等<sup>[9-11]</sup>)及多模态大模型(如GPT-4o Vision)能够协同理解食物图像与专业营养知识,在食物识别与精准评估任务上表现优异<sup>[12,13]</sup>。受此技术浪潮驱动,自2023年以来,营养大模型相关研究呈现爆发式增长<sup>[14,15]</sup>。据统计,约75%的精准营养方向AI研究发表于2020年之后<sup>[16]</sup>,而ChatGPT等生成式AI的问世更是标志着营养信息学进入了智能化新阶段<sup>[14]</sup>。

本文系统综述了2019–2025年间营养大模型领域的研究进展,涵盖技术架构、应用实践、数据资源及评估体系等核心维度。本综述共深度分析了92篇领域内的高质量文献,并做了相应的统计分析。首先,介绍营养信息学的发展背景以及大模型范式兴起的驱动因素,并梳理传统方法与早期深度学习的应用局限;其次,详尽剖析营养大模型的技术架构分类与核心组件,归纳其主流应用场景与典型技术案例;随后,介绍该领域的核心数据集与统一评估基准,并总结

模型性能的评价指标体系;最后,针对当前面临的技术瓶颈与伦理挑战进行讨论,并展望未来研究方向。

## 1 相关工作与背景

### 1.1 基于规则与机器学习的传统营养信息学方法

在大模型范式确立之前,营养信息学主要依赖预定义规则库与传统机器学习模型处理相关数据。早期的营养决策支持系统多基于专家定义的知识库与参数化统计模型,对用户的膳食记录进行逻辑判断与评估。此类方法高度依赖专家经验,缺乏面对个体差异的泛化能力,且难以处理非结构化的数据源。在膳食评估方面,传统方法广泛采用24小时膳食回忆法、食物频率问卷(food frequency questionnaire, FFQ)等人工采集手段,通过匹配食物成分数据库计算摄入量。该流程不仅耗时费力,且准确性极大程度上取决于用户的配合度与主观认知<sup>[17]</sup>。在计算工具层面,多元回归、主成分分析(principal component analysis, PCA)及支持向量机(support vector machine, SVM)等统计与经典机器学习算法被用于预测健康指标,但受限于特征工程的复杂性,2010年末之前营养信息学的自动化与智能化水平相对有限。

### 1.2 深度学习推动食物识别与营养评估

随着计算机视觉技术的成熟,基于深度学习的食物识别于2016年后逐渐成为研究焦点<sup>[18,19]</sup>。研究者相继构建了包含数万级食物图像及配料标签的数据集,如Food-101<sup>[20]</sup>与Recipe1M+<sup>[21]</sup>。利用卷积神经网络CNN,模型能够自动识别菜肴种类并估算营养成分。例如,Fengqing Zhu课题组提出了端到端的食物图像能量估计算法,通过学习“能量分布图”预测卡路里,其精度显著优于传统手工特征分析<sup>[2,4]</sup>。此外,自然语言处理(natural language processing, NLP)技术亦开始应用于分析用户饮食日志与社交媒体记录,利用循环神经网络与长短期记忆网络(convolutional neural networks, RNN; long short-term memory, LSTM)提取营养信息。此阶段的标志性成果包括:goFOOD™系统将图像识别与营养数据库深度集成,实现了全天候膳食自动化评估<sup>[22]</sup>;Google发布的Nutrition5k数据集提供了5 000道菜品的RGB-D图像及高精度营养标注,极大地推动了多模态食物分析技术的发展<sup>[23]</sup>。总体而言,2019–2021年是深度学习赋能营养信息学的奠基期,各类专用模型的涌现为后续基础模型的研发奠定了技术基础<sup>[24]</sup>。

### 1.3 大语言模型与多模态模型的嵌入

大语言模型LLM通常指基于Transformer架构、在海量语料库上训练的深层神经网络<sup>[5]</sup>。这类模型(如GPT-3、BERT等)具备卓越的上下文理解与生成能力,可执行复杂的推理任务<sup>[6]</sup>。将LLM引入营养领域的内在驱动力在于:绝大部分营养知识与临床建议均以文本(如指南、教科书)形式存在,预训练模型有望通过大规模知识内化,在微调后胜任专业语境下的营养咨询与文献综述工作<sup>[5,25]</sup>。然而,模型固有的“幻觉”现象要求必须结合检索增强生成(retrieval-augmented generation, RAG)等技术以确保信息的真实性<sup>[26]</sup>。

另一方面,多模态模型旨在融合视觉、文本及传感器等多源异构数据,这对于处理涉及食物照片、饮食日志与生理指标(如连续血糖监测CGM)的营养问题至关重要。典型的演进路径包括视觉-语言双塔模型(如CLIP, contrastive language-image pre-training)与具备交叉注意力机制的Transformer。2023年后,具备多模态对话能力的生成式模型(如GPT-4o)能够根据图像直接生成营养分析报告,展现了极强的领域适配潜力<sup>[7]</sup>。此外,知识图谱(knowledge graph, KG)与图神经网络(graph neural networks, GNN)的引入也增强了系统在食物推荐中的逻辑推理能力与可解释性<sup>[26,27]</sup>。结合RAG方法的混合智能系统,通过连接外部权威数据库,既保证了生成内容的时效性与准

确性,又扩展了模型的知识覆盖广度<sup>[28]</sup>。

### 1.4 研究现状与热点趋势

纵观2019–2025年的相关研究(图1),可以发现一些显著的趋势。首先,研究主题上涵盖了从膳食图像识别、营养成分估算,到个性化营养推荐、慢性病膳食管理等多个方面。图像处理和 multimodal 营养估算始终是核心热点,占据将近一半的文献比例。与此同时,个性化营养(如结合个体基因、代谢信息进行饮食调控)也是重要方向,反映了精准医疗的大趋势<sup>[28-32]</sup>。另外,大模型技术自2023年爆发后,其应用迅速渗透到营养领域的各个角落,包括构建营养对话系统、辅助营养教育、优化食谱生成等<sup>[3,6,33,34]</sup>。图2中的关键词云直观表明,“营养”“膳食”“个性化”“多模态”“LLM”等是近年来文献中高频出现的词汇,体现了研究的关注焦点。可以预见,随着研究的深入,这些热点主题将继续发展,并产生更多交叉融合的新课题。

## 2 营养大模型技术架构

### 2.1 模型架构分类

营养领域的大模型架构可按技术范式分为三类<sup>[35,36]</sup>:(a)基于计算机视觉的多模态感知模型,即以图像分析为核心,整合文本等多源模态信息的架构;(b)以LLM为核心的智慧系统,侧重于处理营养文本

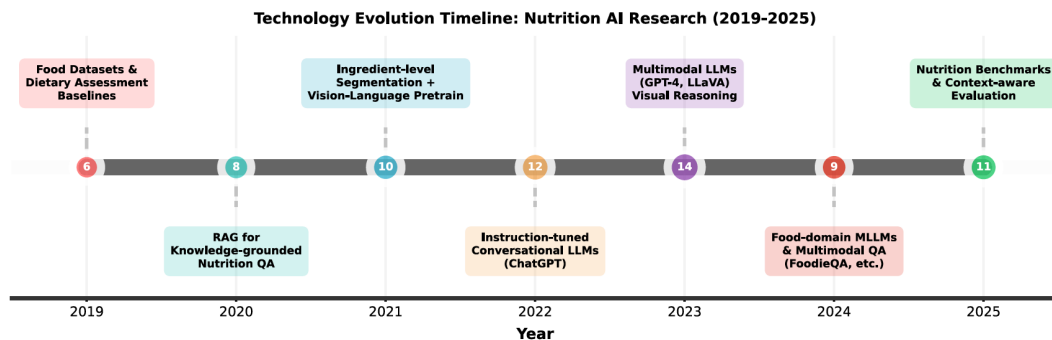


图1 2019–2025年营养AI方向代表性的技术总览

图中标注了代表性技术的里程碑,例如检索增强生成、视觉预训练、多模态大模型等,并用节点大小表示当年相关研究数量(基于本文的检索与统计口径)。总体上,该领域在2022年末至2023年期间(以ChatGPT等对话式大模型的广泛传播为标志)出现明显的研究爆发,技术关注点也由早期以深度学习为主的感知任务,逐步转向以Transformer为核心的生成与推理能力,并进一步发展到多模态融合与面向应用的系统化评测。

Figure 1 Technology evolution timeline: nutrition AI research (2019–2025)

This figure summarizes representative milestones in nutrition AI from 2019 to 2025, where node size indicates the number of related studies in each year (based on this work’s retrieval and counting protocol). Overall, a noticeable acceleration is observed around late 2022 to 2023, coinciding with the widespread adoption of instruction-tuned conversational LLMs such as ChatGPT. The technical focus evolves from early deep-learning-based dietary assessment and dataset-driven baselines to Transformer/LLM-centered generation and reasoning, followed by multimodal fusion and more systematic, application-oriented evaluation.

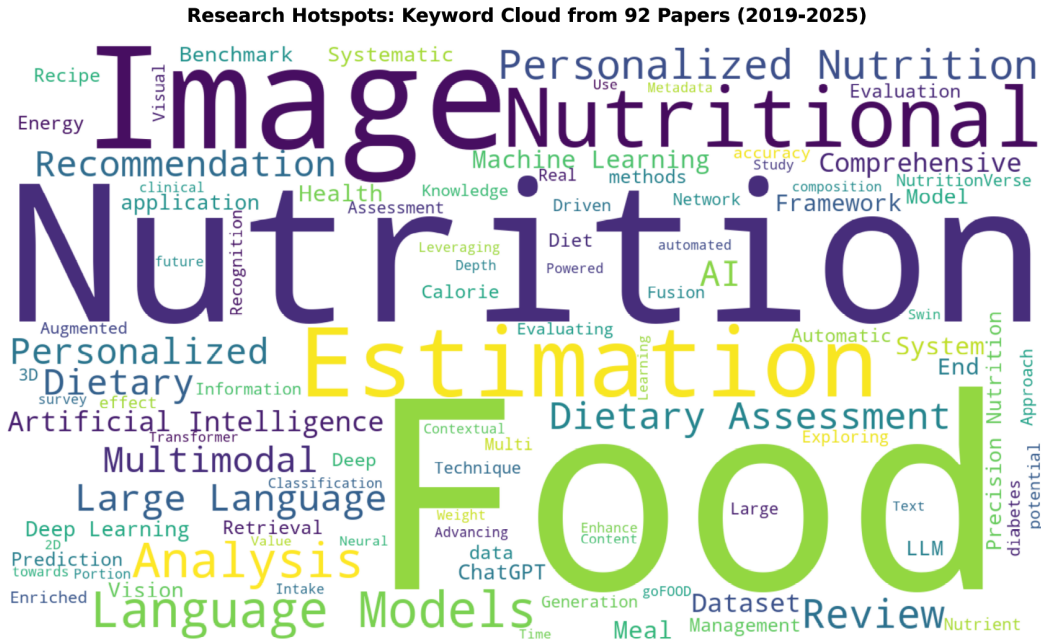


图2 研究热点:营养 AI 关键词词云(2019–2025)

该图统计了92篇相关文献的标题关键词,突出反映了该领域的核心主题与高频关键词(词频大小与字号正相关)。“Nutrition”(营养)、“Food”(食物)、“Dietary”(膳食)、“Personalized”(个性化)、“Multimodal”(多模态)、“LLM”等词非常突出,表明营养、膳食、个性化、多模态基础模型是近年来研究的焦点。

Figure 2 Keyword cloud of research hotspots in nutrition AI (2019–2025)

This figure is generated from high-frequency terms in the titles and abstracts of 92 relevant papers. Persistent themes include “Nutrition”, “Food”, and “Dietary”, while “Personalized”, “Multimodal”, and “LLM” increase markedly in recent years, indicating that nutrition, dietary, personalized, and multimodal foundational models have been the focus of research in recent years.

理解、交互式问答与逻辑决策;(c)融合传统机器学习与领域知识的个性化模型,包括利用临床数据、知识图谱构建的混合驱动范式。根据92篇核心文献的统计显示,计算机视觉与多模态感知方向的研究最为活跃(占比46.7%),LLM及其系统集成研究占比29.3%,其余23.9%属于个性化营养与常规机器学习范畴(图3)。这一分布表明,视觉感知能力仍是营养AI的基础,而具备复杂推理能力的LLM正成为前沿趋势。在具体架构层面,研究广泛涵盖了改进型CNN(主要用于图像任务)、视觉与语言协同Transformer(用于序列建模和跨模态对齐)、图神经网络(用于营养知识图谱推理)以融合多种子模块的复杂系统架构<sup>[22,37]</sup>。

## 2.2 关键技术组件

不同范式的营养AI模型共享若干核心技术组件。首先,注意力机制(attention mechanism)的广泛应用是模型性能跃升的关键。基于Transformer架构的自注意力机制能够有效捕捉长程序列依赖,已成为LLM与视觉Transformer模型的核心支撑<sup>[38]</sup>。在营养领

域,注意力机制大幅提升了细粒度食物识别与营养预测的鲁棒性。例如,Hui Wang课题组<sup>[38]</sup>通过Swin Transformer提取食物特征,相比传统CNN方法,其热量估算的平均绝对百分比误差(MAPE)降低至约15%。其次,多模态对齐与融合技术(multimodal fusion)在处理饮食相关的多源数据中发挥了枢纽作用。常见策略包括:早期融合(在输入层直接合并图像和文本特征)、中期融合(通过跨模态注意力在模型中融合)以及后期融合(将分别预测的结果再整合)。如CalorieLLaVA模型通过视觉编码器与LLM的显式对齐,实现了“以图导文”的联合推理,能够基于餐食快照生成翔实的营养评估报告<sup>[39,40]</sup>。再者,领域知识注入(knowledge injection)与检索增强技术亦是提升系统专业性的重要组件。通过构建营养知识图谱,并结合GNN或RAG技术,模型在回答复杂营养问题时展现出更高的循证准确性与逻辑可解释性<sup>[26,41,42]</sup>。最后,模型优化和训练技巧在营养大模型中亦有所体现。例如,小样本场景下常采用迁移学习(如使用在ImageNet上预训练的CNN,再微调于食

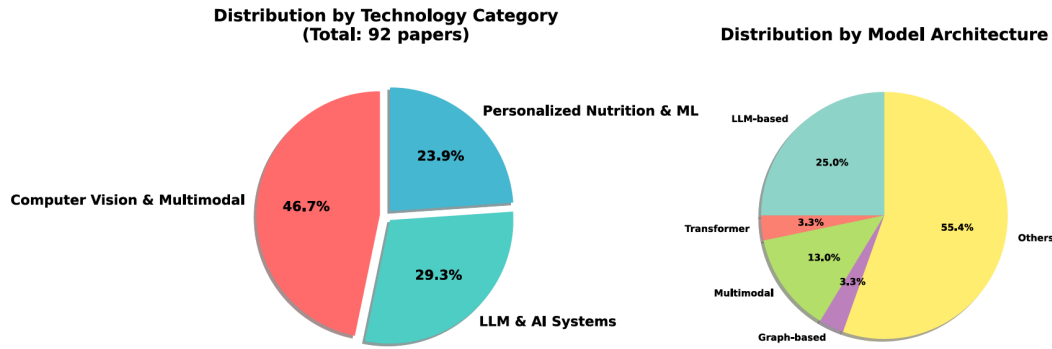


图3 营养AI研究模型类别与架构分布(2019–2025年)

(a)按技术类别划分的文献占比:计算机视觉与多模态占46.7%,大语言模型与AI系统占29.3%,个性化营养与机器学习占23.9%。(b)按具体模型架构划分的占比:包括基于LLM的模型、Transformer模型、CNN模型、多模态融合模型、图模型等<sup>[35,36]</sup>。可以看出,以计算机视觉为核心的多模态方法占主导地位,其次是LLM驱动的系统方法,表明多模态融合和大模型技术正引领营养AI的发展趋势。

Figure 3 Distribution of research directions and model architectures for nutrition AI papers from 2019 to 2025 ( $n=92$ )

(a) By technology category, computer vision and multimodal studies account for the largest share, followed by LLM-driven AI systems and personalized nutrition-related research. (b) By model architecture, “LLM-based” and “Multimodal” paradigms constitute a major portion, suggesting an ongoing shift from conventional deep-learning models toward large-model and multimodal-fusion-centric approaches. Percentages are computed under the retrieval and labeling protocol of this work, primarily to reflect structural changes in research emphasis.

物数据集)以弥补数据不足;针对多任务需求引入多任务学习框架,让一个模型同时预测多种营养指标<sup>[41]</sup>;在交互式系统中尝试强化学习(如让AI根据用户反馈调整推荐,以优化长期健康收益);以及利用大量未标注数据进行自监督预训练等。这些技术手段丰富了营养AI模型的能力边界。例如,Nutrition5k研究中构建的多任务模型能同时预测卡路里、脂肪、蛋白质等多项指标,相比单任务模型效率更高<sup>[23]</sup>;又如有研究利用对比学习从食谱-图像对中学习跨模态表示,提升了图像菜谱检索的效果<sup>[22]</sup>。可以预见,随着基础模型能力的提升和领域适应技术的发展,这些关键组件将在未来营养AI系统中得到更加成熟的应用。

### 2.3 代表性模型实例

为系统展现技术架构在实际业务中的演进,表1对比了2020–2025年间具有代表性的营养AI系统及其核心特征<sup>[43,44]</sup>。

表1展示了早期以计算机视觉为主的膳食评估模型(goFOOD<sup>TM</sup>、Nutrition5k),以及近期融合LLM、多模态、知识图谱的新型系统(CalorieLLaVA、HealthGenie、RAG-based等)。从中可以看出,不同系统在架构类型、输入模态、主要功能任务等方面各具

特色。例如,goFOOD<sup>TM</sup>是2020年提出的较早的AI膳食评估系统,采用传统CNN和规则的方法完成饮食记录的图像识别与营养计算,并在临床上进行了验证;2021年的Nutrition5k主要贡献在于发布大规模标准数据集,带动了后续模型的开发;2023年兴起的ChatGPT等LLM则能实现零样本的营养问答与建议生成;2024年推出的NutriBench成为评估LLM营养能力的基准数据集;2025年的CalorieLLaVA将视觉-语言大模型用于热量估算,HealthGenie将知识图谱与LLM结合用于健康饮食指导,而RAG技术开始被用于提升营养问答的准确性。

从表1的对比中还可以发现几个值得关注的共性趋势:其一,目前多数系统尚缺乏严格的临床验证,列举的7个系统更多停留在离线实验、用户研究或技术验证阶段;其二,实时性和多模态能力在新近系统中已成为标配,说明营养AI应用正朝着即时交互和数据多源融合方向发展;其三,可解释性和知识融合逐渐受到重视,2025年的几个系统通过引入知识图谱、RAG等方式提高了模型决策的透明度和可信度<sup>[26,27]</sup>。总的来说,技术架构的演进与应用需求息息相关,早期注重功能实现和精度提升,近期则开始更多关注实用性和可靠性,为营养AI走向真实世

表1 2020–2025年代表性营养AI系统的技术特征对比  
Table 1 Technology comparison: representative nutrition AI systems

System	Architecture	Input	Task	Accuracy	Clinical	Real time	Explain	Contribution
goFOOD (2020)	CNN Multi-task	RGB-D	Assessment	15%~20%	√	√	Low	Early integrated AI system for dietary assessment
Nutrition5k (2021)	ResNet	RGB-D	Estimation	10%~15%	√	√	Low	Large-scale, high-quality dataset for benchmarking
ChatGPT (2023)	GPT-3.5/4	Text	Q&A	50%~80%	×	√	Med	Zero-shot capability in open-domain nutrition advice
NutriBench (2024)	LLM Benchmark	Text	Evaluation	Benchmark	×	N/A	N/A	Comprehensive benchmark for nutrition-focused LLMs
CalorieLLaVA (2025)	LLaVA	Image +Text	Calorie	12%~18%	×	√	Med	Multimodal approach for calorieestimation
HealthGenie (2025)	KG+LLM	Text+KG	Guidance	-	×	√	High	Integration of structured knowledge for reasoning
RAG Systems (2025)	RAG+LLM	Text+Docs	Advice	Improved over base LLM	×	-	High	Enhanced answer reliability with retrieved evidence

界应用铺平道路。

### 3 应用场景与案例

营养大模型的最终价值在于其落地应用的实际成效。学术界的研究应用主要聚焦于前后协同的四大研究领域:自动化食物识别与记录、膳食成分分析与营养评估、个性化精准营养推荐及临床疾病营养干预。其中,营养评估领域进一步细分为饮食评估与追踪等具体研究方向,而个性化精准营养推荐则包括了食谱与餐食规划等细化的应用场景。图4展示了各应用场景相关研究在该领域的文献占比分布<sup>[45]</sup>。

#### 3.1 食物识别与膳食记录

食物识别可以说是营养AI最基础也最重要的应用之一<sup>[18]</sup>,其识别准确性直接关系到后续营养摄入估算的有效性。传统上,食物识别依靠用户自行记录或拍照,然后由专家进行人工判读<sup>[17]</sup>。而计算机视觉技术的发展让这一过程实现自动化成为可能<sup>[47]</sup>。从早期的CNN分类模型到现在的多模态大模型,食物识别的准确率不断提高。Asbroeck等<sup>[48]</sup>的研究比较了多种食物图像识别平台在膳食评估中的表现,发现结合深度学习的系统识别常见食品的准确率超过90%,显著优于旧有的方法,但是对于食品数量的估计不太准确。除了识别食物种类和食品数量,模型还需要估计食物的分量(例如一碗饭多少克)。为此,一些算法利用参考物大小校准、三维重建等方法推断食物体积<sup>[49,50]</sup>。近年来,多模态模型通过结合图像和文本(如用户用语音描述食物或菜单信息),在更加复杂的场景下实现了鲁棒的识别。例

如,有研究构建了一个名为“NutrifyAI”的系统,利用实时目标检测模型识别视频中的食物,并结合营养数据库实时显示摄入的食品信息<sup>[51]</sup>。在外卖点餐和餐厅环境中,研究者也尝试部署视觉识别模型自动识别餐盘中的食物种类和分量,帮助用户记录就餐情况。总体而言,食物识别技术已经相对成熟,许多移动应用和可穿戴设备(如智能眼镜)开始集成此类功能<sup>[52]</sup>。然而,挑战主要在于食物多样性和复杂用餐场景:模型需要识别世界各地成千上万种食物,且要应对混合菜肴、遮挡、光照变化等现实条件<sup>[4]</sup>。为此,持续扩充训练数据集、利用合成数据增强(如NutritionVerse生成的大量虚拟食物图像<sup>[53]</sup>)以及开发更强大的视觉模型(例如融合Transformer的视觉架构<sup>[54]</sup>)都是当前的研究重点。可以预见,随着这些问题逐步解决,全自动的膳食记录将成为可能,从而极大地方便个人和临床对饮食数据的获取。

#### 3.2 营养分析与膳食评估

在获得可靠的食物图像与食品信息之后,营养与膳食评估的定量分析成为了连接食品数据与营养干预的桥梁。典型任务包括:从食物图像中估计其热量和营养成分、根据饮食记录计算宏量和微量营养素摄入、评估膳食模式和营养均衡性等。在这方面,前述的食品识别发挥了核心作用<sup>[2]</sup>。许多研究构建了能够识别食物种类并量化营养的图像分析模型。例如,Google的Im2Calories项目探索了通过手机拍照自动获取膳食热量的方法。近年来的多模态大模型进一步提升了营养分析的自动化水平<sup>[55]</sup>。alorieLLaVA等模型利用图像-文本双通道,一方面识

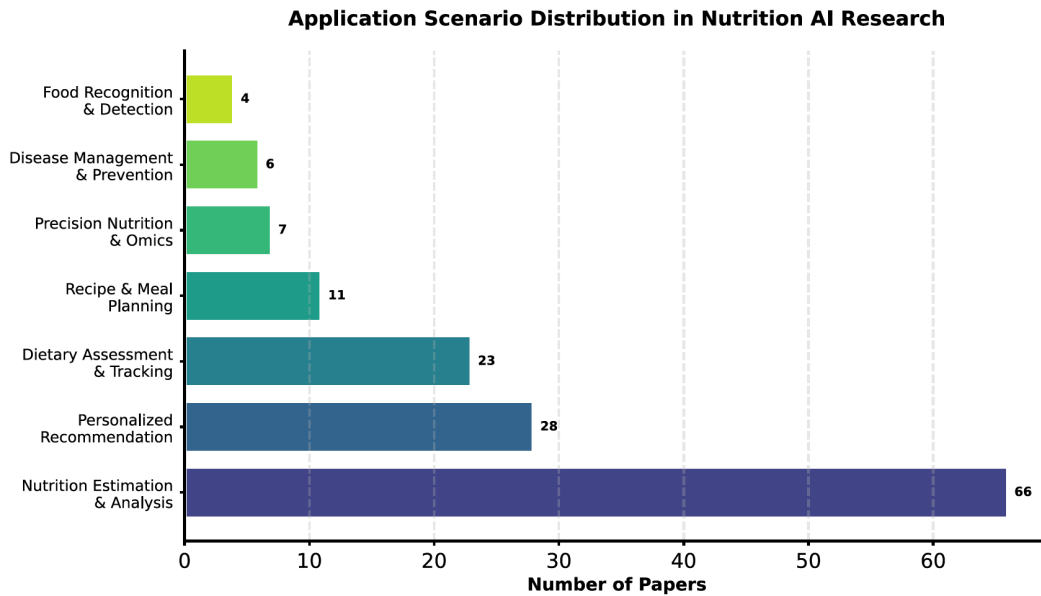


图4 营养大模型应用场景的研究分布(按论文数量统计)

图中共有四大研究领域的七个具体研究方向,此处按频次高低排序。“营养成分估算与分析”和“个性化营养推荐”是研究最多的方向,其次是“饮食评估与追踪”“食谱与餐食规划”。相对而言,“精准营养与多组学”属于新兴前沿领域,相关研究较少<sup>[33,46]</sup>。这一分布反映了学界对营养评估和个性化干预的高度关注,同时表明将多组学数据融入营养干预的研究尚处起步阶段。需要注意的是,一篇研究可能涉及多个应用场景,此统计并非严格互斥。

**Figure 4 The distribution of papers across application scenarios in nutrition AI (sorted in descending order by count)**

The figure shows seven specific research directions across four major research areas, sorted by frequency. “Nutrition Estimation & Analysis” and “Personalized Nutrition Recommendation” are the most researched directions, followed by “Dietary Assessment & Tracking” and “Recipe & Meal Planning”. In contrast, “Precision Nutrition & Omics” is an emerging frontier field with relatively few studies<sup>[33,46]</sup>. This distribution reflects the academic community’s strong focus on nutrition assessment and personalized interventions, while also indicating that research on integrating multi-omics data into nutrition interventions is still in its early stages. It is important to note that a single study may involve multiple application scenarios, and therefore, this statistic is not mutually exclusive.

别食物及分量,另一方面结合LLM生成详细的营养报告<sup>[39,40]</sup>。此外,一些研究专注于提高食物体积和重量估计的准确性,例如利用深度传感或3D重建技术从单张照片推断食物体积<sup>[49,56]</sup>。营养分析模型的评估指标通常包括识别准确率(对食物种类的判断)、重量/体积估计误差以及热量和养分的预测误差等<sup>[55]</sup>。当前最先进的模型在理想条件下对常见菜肴的热量估算平均误差可降至20%左右<sup>[55]</sup>,但在复杂菜肴或遮挡情况下误差仍较高。除了个人膳食分析,大模型也开始用于群体营养监测。例如,有研究利用多模态模型分析社交媒体食品图片,以了解大众饮食趋势和营养状况,从而为公共卫生政策提供依据。总之,营养分析场景下的大模型应用正逐步从实验室走向实际应用,一些智能手机App已开始内置食物识别和营养估算功能。然而,这一领域也面临挑战,包括模型对新食物的识别能力、在真实环境下的鲁棒性以及如何获取高质量的标注数据等<sup>[4,54]</sup>。

### 3.3 个性化营养推荐

在营养摄入分析研究不断深化的基础上,该领域的研究前沿正逐步拓展至更具现实指导意义的个性化营养干预方向。个性化营养旨在整合个体的健康状况、膳食习惯、基因组学特征等多维度信息,以量身定制饮食方案,促进个人的健康管理发展。AI大模型在此场景中扮演重要角色:通过综合分析个人数据和海量营养知识,生成定制化的膳食建议。例如,ChatGPT等LLM已被用于生成个性化的每日食谱和营养建议。Ruopeng An课题组<sup>[57]</sup>的系统综述表明,ChatGPT能够给出与注册营养师相当质量的膳食计划建议,在膳食多样性和营养均衡性方面表现良好。又如Tsolakidis等<sup>[28]</sup>研究了大型语言模型在减重食谱生成中的应用,结果显示结合个体特征信息的LLM可以提供具有针对性的减肥食谱方案。除了LLM,传统机器学习结合传感器数据也被用于个性化营养:一些移动应用通过设备获取用户

的运动和生理数据,并利用算法推荐实时的饮食调整。总体而言,大模型使得个性化营养从“专家手工制作”走向“智能动态生成”成为可能。然而,需要强调的是,模型生成的建议需要经过营养师审核,确保其安全性和可行性。例如,有研究对ChatGPT生成的饮食建议进行质量评估后指出,尽管整体可靠,但在处理复杂健康状况时偶有错误,需要辅以人工校验<sup>[58]</sup>。因此,当前更现实的模式是“AI辅助的营养师”:由大模型提供初步方案,再由专业人员把关修改。这种人机协作有望大幅提高个性化营养服务的效率和可及性。

### 3.4 疾病营养管理

在个性化营养研究不断深入的基础上,该领域的技术落地进一步聚焦于慢性病的膳食改善与防控。营养干预在慢性疾病的管理与预防中发挥关键作用,因此,越来越多研究致力于利用人工智能技术,辅助构建和优化与特定疾病相关的营养干预方案。典型应用包括:针对肥胖、糖尿病、高血压等慢性病的个性化饮食干预建议<sup>[59]</sup>,肿瘤患者营养支持方案优化,以及临床营养咨询等。大模型通过结合患者的医疗数据(如血糖监测记录、化验指标)与营养知识,可以为临床提供决策支持。例如,Gavai等<sup>[60]</sup>开发了一个针对肥胖和2型糖尿病的营养AI系统,采用RAG架构,实时从医学营养数据库中检索信息以辅佐LLM产生个性化的减重和控糖建议。该系统在初步试验中表现出较高的建议合理性。又如Lara Nasreddine课题组<sup>[61]</sup>研究了营养聊天机器人在糖尿病及代谢综合征管理中的应用,结果表明AI聊天机器人可以提供持续的饮食教育和监控,在临床上显著改善了患者的饮食依从性和代谢指标。对于临床疾病营养管理而言,安全性和循证性是大模型应用必须严格考量的。模型的建议必须符合临床营养指南,否则可能对患者造成危害。因此,一些研究致力于评估LLM给出营养建议的可靠程度。例如,Parameswaran等<sup>[27]</sup>在一项研究中比较了ChatGPT是否能提供符合心血管疾病预防膳食指南的建议,发现单纯LLM时有遗漏,引入检索相关指南内容后,建议的准确度和一致性有明显提升。这说明将知识检索与LLM相结合可以提高模型在专业领域的可靠性。总体而言,大模型在疾病营养管理中大有用武之地,但需要与临床流程紧密结合。未来的方向包括:开发受监管的临床决策支持系统,

与电子健康记录对接,实现对患者营养状况的自动监测和干预提醒<sup>[62]</sup>,开展大规模随机对照试验验证AI营养干预的疗效。

### 3.5 国内相关研究进展

2019–2025年间,中国人工智能在营养学领域的研究呈现出多点突破、协同推进的发展态势。在食物识别与膳食记录方面,基于ResNet、Inception V3及YOLOv5s等改进深度学习网络的研究显著提升了中餐复杂场景下的图像识别精度,实现了对菜品类别的自动化识别与特征提取<sup>[63–65]</sup>。在营养分析与膳食评估方面,通过引入多尺度拼图重构网络与多模态数据分析方法,视觉识别结果得以进一步映射至能量与矿物质摄入的自动化计算,有效降低了传统膳食调查对人工记录与主观回忆的依赖<sup>[30,66]</sup>,并在Transformer架构与自监督学习范式的推动下,逐步向细粒度食物识别及隐含营养成分精准感知拓展<sup>[67]</sup>。在个性化营养推荐方面,精准营养系统开始整合用户饮食偏好、组学数据与机器学习聚类算法,构建面向个体差异的饮食推荐模型与数字营养教育方案。在疾病营养管理领域,人工智能已应用于2型糖尿病、慢性肾脏病及炎症性肠病等代谢性与慢性疾病,通过营养风险预测模型与数字孪生模型,实现了从营养风险筛查、状态评估到辅助医学营养干预的全流程数字化管理<sup>[68,69]</sup>。

这些技术进展为商业化奠定了基础,并快速转化为两类主要落地路径:一是以AI驱动的产品创新,如飞鹤、仙乐健康等企业利用AI进行功能营养品研发与配方智能生成,实现了从需求洞察到产品定制的升级;二是以数智化服务解决方案的落地为核心,覆盖临床与消费两大场景。在临床侧,楚动科技、京东健康等推出的AI营养管理系统,已实现从院内营养风险筛查、干预到居家管理的全流程数字化服务;在消费侧,“AI餐桌”等平台则通过算法直接连接个性化推荐与健康餐配送,完成了推荐与评估的商业闭环。

## 4 数据集与评估基准

### 4.1 核心营养数据集演进

高质量的数据集对于训练和评估营养AI模型至关重要。2019–2025年间,一系列公开数据集的推出推动了本领域技术的长足进步。其中具有里程碑意义的是Nutrition5k<sup>[23]</sup>。该数据集由Google Research

在CVPR 2021上发布,包含5 000道多样菜肴的RGB-D(彩色+深度)图像,以及对应的主成分营养标注(热量、三大产能营养素等)。Nutrition5k是首个规模较大的营养标注数据集,为模型学习提供了标准平台,许多后来者都以此为训练或测试基准。此后,数据集的规模和多样性不断提升。2023年,一系列NutritionVerse数据集发布<sup>[53,70]</sup>;其中NutritionVerse-Synth利用高精度3D食物模型合成了84 984张多视角2D食物图像<sup>[53]</sup>,极大丰富了模型的训练样本;NutritionVerse-Real则收集了真实世界中的889张带准确称重的菜肴照片<sup>[71]</sup>,提供了真实场景下的评估基准;NutritionVerse-3D包含高质量的3D食物模型库<sup>[72,73]</sup>。这些数据集着重于多视角、多模态信息,对于解决食物体积估计具有重要意义。另一重要数据集是NutriBench,于2024年提出,专注于评估大语言模型在营养领域的能力<sup>[74]</sup>。NutriBench由一系列膳食描述及其对应营养答案组成,可用于量化LLM在营养问答和估算任务上的表现<sup>[75]</sup>。此外,传统的图像食物数据集仍然发挥作用,如Food-101(101类西式菜肴图片)、Recipe1M+(包含100万食谱-图片对,用于跨模态检索)等在不少研究中被用来预训练或辅助模型提升对常见食物的识别能力<sup>[20,22]</sup>。还有一些专用数据集针对细分问题:如SNAPMe数据集专注于日常餐食照片的营养评估<sup>[74]</sup>;EPIC-Kitchens扩展用于视频中的饮食行为分析等。总的来看,数据集建设呈现出规模扩大(从几千到数万以上样本)、模态丰富(RGB-D、3D、文本描述等多源信息)、场景真实化(合成数据转向真实生活数据)三大趋势。这为训练更健壮、更全面的营养AI模型奠定了数据基础。

#### 4.2 评估指标与基准测试

对于营养AI模型的性能评估,需要针对不同任务采用相应指标和基准。对于食物分类任务,常用指标为Top-1准确率、Top-5准确率等;对于营养成分回归任务,则采用均方误差(MSE)、平均绝对百分比误差(MAPE)等衡量模型对热量、营养素的预测精度<sup>[55]</sup>;在膳食模式分析中,则可能使用饮食质量评分(如Healthy Eating Index)或与健康结局的相关性作为评价标准(表2)。

对于LLM营养问答或建议生成,则需要设计特定的基准,例如NutriBench通过比较模型产出与标准答案计算准确率,还可以由人类专家对答案的相关

性和可靠性打分<sup>[75]</sup>。此外,一些研究采用专业考试作为评估方式:例如Youlin Li课题组<sup>[34]</sup>让ChatGPT参与营养师执业资格考试,结果GPT-4正确率可达70%以上,尤其在结合知识检索后正确率进一步提高。这种以人类水平为参照的评估可以直观体现模型是否达到实用要求。为了推动模型间的公平比较,公开基准测试变得日益重要。例如,前述的NutriBench为LLM提供了统一测试题库;ImageCLEF餐食分析挑战提供了多模态食物分析的竞赛平台,每年汇集不同模型在同一数据上的表现进行比拼<sup>[83,84]</sup>。这些基准有助于发现当前模型的短板并指引研究方向。一个普遍的发现是,领域适应对模型性能至关重要:通用模型直接用于营养任务时表现有限,在微调Nutrition5k等领域数据后性能显著提升<sup>[83]</sup>。因此,评估时也需考虑模型是否接受过领域微调。

#### 4.3 不同研究方向代表模型性能比较

不少研究在各自的数据集上报告了最新模型的优异表现。例如,在Nutrition5k数据集上,2022年的Swin-Nutrition模型将热量误差降低到约15%量级,在宏量营养素部分指标上的误差降低到约20%量级,相比2019年的基线算法(误差>30%)有大幅改进。又如在食物分类任务上,Transformer类模型(ViT、Swin等)在Food-101上已将Top-1准确率提高到88%以上,而早期CNN的精度约85%。这些提升并不只是模型结构更复杂的简单结果,而是反映了模型在营养任务关键瓶颈上的适应性增强:一方面,基于注意力机制的表征更擅长同时整合局部纹理与全局结构,使其在光照变化、遮挡、混合餐盘、摆盘差异等真实场景扰动下更稳健,从而对能量与宏量营养素估计中的细节部分更不易失真;另一方面,Transformer在预训练与迁移过程中更容易学习到可复用的食物属性特征(如加工方式、配料形态),因此面对不同菜系或同名菜品的多样化呈现时,往往比仅依赖局部纹理的传统CNN具有更好的跨域泛化潜力。

与视觉任务相比,基于个性化营养推荐的优良性则更集中体现在对个体差异的刻画与持续校准能力上。在个性化营养推荐方面,一系列基于机器学习预测的个性化膳食方案较通用膳食指南能带来显著的血糖控制改善<sup>[85-88]</sup>,这些模型能够学习并利用个体对同一食物的差异化代谢反应,从而在实践中实现对人群平均数值规范的超越。在LLM方面,2025年的一项研究比较了ChatGPT、Bard和BioGPT

表2 AI营养评估任务类型与数据汇总表  
Table2 Summary of AI nutrition assessment task types and data

Task	Typical Input Modality	Output Type	Representative Test Datasets	Core Metrics
Food Classification (single-label)	Single food image	One class label	Food-101 <sup>[20]</sup> ; UEC-Food100; Food-11; Fruit-360	Top-1 accuracy; Top-5 accuracy; Macro F1
Food Classification (large-scale, long-tail)	Single food image	One class label	Food2k <sup>[76]</sup> ; Vireo-172 <sup>[77]</sup> ; ChineseFoodNet <sup>[78]</sup>	Top-1 accuracy; Top-5 accuracy; Macro F1
Food Classification (noisy labels, robustness to noise)	Single food image	One class label	ETHZ Food-101N <sup>[20]</sup>	Top-1 accuracy; Top-5 accuracy; Macro F1
Food Detection (multi-item meals)	Multi-item meal image	Multiple bounding boxes	ImageCLEF meal analysis challenge	Mean average precision at IoU thresholds (e.g., mAP@0.5)
Food Segmentation (semantic)	Image	Pixel-level class map	FoodSeg103 <sup>[79]</sup>	Mean IoU
Portion Size and Weight Estimation	Image with plate reference; multi-view when available	Relative portion size; weight in grams	ImageCLEF meal analysis challenge	Mean absolute error; mean absolute percentage error
Volume Estimation (3D)	Multi-view images; monocular with priors	Volume (mL or cm <sup>3</sup> )	ImageCLEF meal analysis challenge	Mean absolute error; mean absolute percentage error
Calorie and Nutrient Estimation (regression)	Image; image plus recipe text (multimodal)	Calories (kcal); nutrients (gram, mg)	Nutrition5k	Mean absolute error; mean absolute percentage error, RMSE
Menu Matching and Food Database Linking	Menu text; item names; descriptions	Matched menu item; database entry	Open Food Facts	Recall@K; MRR; nDCG@K
Recipe Retrieval and Recipe Text Generation	Recipe text; ingredient lists; titles	Retrieved recipe; generated recipe text	recipe1M; RecipeNLG <sup>[80]</sup>	Recall@K; MRR (retrieval); ROUGE (generation, optional)
Knowledge Base Grounding and RAG Resources	Queries plus retrieved documents; knowledge graph triples	Grounded answer; linked entities; citations	CookBook-KG; Open Food Facts	Citation correctness rate; faithfulness score
Food and Recipe Question Answering	Natural language questions	Short factual answer; multi-hop reasoning answer	FoodieQA <sup>[81]</sup> ; RecipeQA <sup>[82]</sup>	Exact match; token-level F1; tolerance accuracy for numeric answers

注:本表包含每类任务对应的典型输入模式、输出形式、代表性公开基准以及核心评价指标。该表旨在为不同研究工作提供统一的评估参照,从而支持模型可复现与比较。

Note: The summarizes of common evaluation task types and datasets used in nutrition AI research. This table includes the typical input modalities, output formats, representative public benchmarks, and core evaluation metrics corresponding to each task type. The table aims to provide a unified evaluation reference for different research works, thereby supporting model reproducibility and comparability.

在营养问答上的表现,发现ChatGPT在准确性和连贯性上略胜一筹,但在需要精确数值时经常出错,需要借助工具检索<sup>[89]</sup>,这反映了基于大模型的营养问答的核心挑战向着更可靠迈进。

总的来说,不同模型各有所长,目前尚不存在“通吃”所有营养任务的单一模型。这也正是本领域有趣和复杂之处:研究者需要针对具体问题选择或打造最合适的模型,并不断以数据为驱动进行改进迭代。今后,随着数据集和评价基准的完善,模型性能的量化对比将更加系统明确,也有望出现兼具鲁棒感知能力与深层医学推理能力的通用型营养基础模型。

## 5 挑战与未来方向

尽管营养大模型取得了令人瞩目的进展,但要实现其在真实世界的广泛应用,仍有诸多挑战亟待解决,同时也孕育着新的研究机会。

### 5.1 模型可解释性与可信性

营养干预直接关系个体健康结局,因此模型输出的可靠性与可解释性至关重要。目前多数深度学习模型(尤其是生成式LLM)仍存在“黑箱”属性,难以在医学严谨性要求极高的场景下获得完全信任<sup>[25]</sup>。未来需要加强模型的可解释性,例如结合知识图谱或因果模型,使模型的决策逻辑具备证据溯源能力<sup>[13]</sup>。可解释性提升不仅有助于营养师和用户信任模型,也能帮助开发者发现模型错误。近期已有工作尝试通过引入显著性可视化、生成解释文本等方式提高营养AI的透明度,但仍处于早期阶段。如何在保证模型精度的同时,让其决策过程达到人类可理解,是一项长期课题。

### 5.2 个性化营养的数据隐私与伦理

个性化营养不可避免地涉及个人的健康数据(如疾病史、基因信息)和日常生活数据(如饮食习惯、地理位置)。在收集和利用这些数据训练大模型时,必须严格保护用户隐私<sup>[90]</sup>。例如,很多饮食应用需要上传用户的饮食照片或记录,这就要求在数据传输和存储中采用加密、匿名化等措施,使其符合相关法律法规(如GDPR、HIPAA等)的要求。联邦学习等隐私计算技术可在一定程度上缓解这一问题,即在不集中用户数据的情况下训练全局模型。此外,营养AI的推广也涉及伦理与安全考量,例如算法可能存在偏见(对某些文化的食物识别率低、给

出的建议不适合某些族群)以及上下文夹带的信息存在隐私风险<sup>[83]</sup>。因此,未来研究应注重构建多元化的数据集和公平的算法,确保模型在识别少数族裔菜系或为不同社会经济背景的人群提供建议时,具备一致的高质量表现<sup>[83,90]</sup>。

### 5.3 文化多样性与精准化适配

全球饮食文化具有极强的地域特异性与多样性,这要求大模型具备卓越的跨文化泛化能力。这对大模型提出了跨文化适应的要求。目前许多模型在西方饮食数据上训练良好,但对东方饮食或其他少数族裔的食物认识不足,影响了其泛化能力<sup>[83]</sup>。未来需要持续扩充涵盖不同文化饮食的数据,并针对性地对模型进行微调适应。此外,不同个体的营养需求千差万别,理想的营养AI应能针对个人情况进行定制。这可能需要模型能够处理个体基因、多组学数据,实现“精确到个人”的营养推荐<sup>[91,92]</sup>。如何将群体水平的营养知识转化为个体化应用,考验着模型的灵活性和精细化程度,也是未来的重要研究方向。

### 5.4 临床验证与实用转化

目前绝大多数营养大模型研究仍停留在实验室或原型验证阶段,真正进入临床或大众市场的并不多。其中一个原因是缺乏严格的临床验证。医疗领域对新技术的要求是经过充分的循证证明其有效且安全。然而,让AI参与临床营养决策,需要考虑复杂的人体代谢和行为因素,评估其长期效果。这往往意味着需要开展耗时耗资的临床试验。在这方面已有少数开创性尝试:例如2024年发表的个性化营养RCT证明了机器学习指导的饮食干预优于传统方法<sup>[85]</sup>。未来,需要更多针对AI营养干预的临床研究,以积累证据、赢得医生和公众的信任。同时,在实用转化上,还需解决AI系统与现有医疗工作流程的整合问题,比如如何与营养门诊、健康管理App结合,谁来负责监督AI建议的执行等。此外,营养AI应用的商业模式、监管审批也需探索明确,以便优秀的研究成果能够顺利地转化为产品服务。可以预见,随着技术成熟和法规框架的完善,未来5~10年内会涌现出经过验证的AI营养师助手、智能膳食管理系统等,为大众健康提供有益的支持。

### 5.5 新技术融合与发展

面向未来,营养AI有望与更多新技术结合,催生新的研究方向。

例如,多组学与营养的融合:将个人基因组、肠道菌群、代谢组数据与饮食数据结合,利用大模型寻找更精细的饮食-健康关联,这是精准营养的重要前沿<sup>[91,93]</sup>。又如强化学习在饮食行为干预中的应用:通过让AI与用户反复交互,并根据健康指标反馈不断调整建议,实现动态优化的饮食干预策略。再如数字孪生与虚拟试验:未来可能构建每个人的“数字身体”模型,模拟不同营养干预的效果,从而为个体制定最优饮食方案。在技术层面,随着更强大的基础模型出现(参数量达万亿级、具备更强的推理能力),营养AI的能力上限将被不断刷新。在此之后,还可根据虚拟试验结果预测不同营养干预对身体的影响。通过模拟不同饮食方案的效果,数字孪生不仅可以帮助个体制定最优饮食方案,还可以避免传统临床试验中的高成本和高风险。在虚拟环境中,研究者能够探索营养干预的长期效果,并预测不同饮食模式对健康的潜在影响,从而为个体提供更加精准的健康管理方案。

随着越来越多垂类数据集制作的深入,面向营养领域的专属知识增强的大模型会是一大趋势,即在通用大模型中融入医学营养专业知识,使其既有强大的语言能力又具备可靠的专业知识库<sup>[13]</sup>。这种模型或许能回答复杂的营养问题,甚至发现新的饮食与健康规律。

另一个可以期待的方向是,随着可穿戴设备的普及,如智能手表、健康追踪器和智能餐具等,AI在营养干预中的实时监控能力将进一步增强。通过收集用户的生理数据、运动数据和睡眠数据,AI可以为个体提供实时、动态的饮食建议。这些设备的实时反馈将使得AI能够及时调整建议策略,确保个体始终在最佳健康状态下进行饮食管理,进而促进长期健康的实现。从更进一步的生态搭建来看,还可以进一步建立数字健康社区,从公共卫生健康的角度加速AI在健康道路上实现更精准的辅助管理。

总之,营养大模型作为一个新兴交叉领域,未来的技术演进将取决于AI和营养科学这两个领域的发展动态。可以预见的是,随着相关挑战逐步被克服,营养大模型将在公共营养、临床营养和个体健康管理等方面发挥越来越重要的作用。

## 6 结论

2019年以来,营养信息学领域见证了从深度学

习基础应用到大模型引领创新的跨越式发展。本综述回顾的92篇文献表明,大型语言模型和多模态模型正迅速成为营养研究的重要工具,引领技术范式的转变。特别是2023年ChatGPT的出现显著推动了营养AI的快速发展,相关研究产出在短时间内成倍增长<sup>[15]</sup>。在技术架构上,模型经历了从卷积神经网络、注意力机制,到当下多模态Transformer和知识增强LLM的演进,每一步都带来了性能和能力的跃升<sup>[38]</sup>。当前,多模态营养大模型能够在图像、文本等多源数据的支持下,实现对食物的精准识别和对个体营养的全面分析;LLM则让营养知识的提取和个性化建议生成变得前所未有的高效便捷。在应用层面,营养AI正在从实验室走向现实,在个性化营养指导、饮食监测、慢病管理等方面展现出积极作用<sup>[85]</sup>。可以预见,随着模型可靠性的提升和临床循证的积累,智能营养干预将成为未来健康管理的重要组成部分。

展望未来,我们认为营养大模型的研究与应用应着重把握以下几点。首先,以人为本,强化解释。营养干预关系到行为改变和长期依从,模型应能输出让用户易于理解和信服的建议,这需要融合人因研究和提高模型的可解释性<sup>[13]</sup>。其次,多学科融合,确保安全。营养学、大数据、临床医学、伦理学等领域专家应加强协作,从数据采集、模型训练到结果解读各环节共同把关,确保AI建议科学有效并符合伦理规范<sup>[90]</sup>。第三,建立标准,验证实效。建议制定营养AI模型的评价标准和指南,开展多中心的临床试验评估其健康效益,以获得监管机构和医疗界的认可。最后,关注弱势与多样群体。应确保模型对不同文化、不同健康状况的人群均适用,避免“算法鸿沟”,真正实现营养干预的普惠<sup>[89]</sup>。总之,营养大模型为提升人类营养健康水平提供了前所未有的机遇,也伴随着挑战和责任。只要学术界和业界携手努力,稳步推进,我们有理由期待一个由智能技术赋能的营养健康新时代的到来。

## 参考文献

- [1] Bailey RL. Overview of dietary assessment methods for measuring intakes of foods, beverages, and dietary supplements in research studies. *Curr Opin Biotechnol*, 2021, 70: 91-6.
- [2] Fang S, Shao Z, Kerr DA, et al. An end-to-end image-

- based automatic food energy estimation technique based on learned energy distribution images: protocol and methodology. *Nutrients*, 2019, 11: 877.
- [3] Myers A, Johnston N, Rathod V, et al. Im2Calories: towards an automated mobile vision food diary[C]// Santiago, Chile: 2015 IEEE International Conference on Computer Vision (ICCV), 2015: 1233–41.
- [4] Mezgec S, Koroušić Seljak B. Deep neural networks for image-based dietary assessment. *J Vis Exp*, 2021: 61906.
- [5] Bergling K, Wang LC, Shivakumar O, et al. From bytes to bites: application of large language models to enhance nutritional recommendations. *Clin Kidney J*, 2025, 18: sfaf082.
- [6] Yang Z, Khatibi E, Nagesh N, et al. ChatDiet: empowering personalized nutrition-oriented food recommender chatbots through an LLM-augmented framework. *Smart Health*, 2024, 32: 100465.
- [7] Lo FPW, Qiu J, Wang Z, et al. Dietary assessment with multimodal ChatGPT: a systematic analysis. *arXiv*, 2023, <https://doi.org/10.48550/arXiv.2312.08592>.
- [8] Papastratis I, Konstantinidis D, Daras P, et al. AI nutrition recommendation using a deep generative model and ChatGPT. *Sci Rep*, 2024, 14: 14620.
- [9] Alayrac JB, Donahue J, Luc P, et al. Flamingo: a visual language model for few-shot learning. *arXiv*, 2022, <https://doi.org/10.48550/arXiv.2204.14198>.
- [10] Radford A, Kim JW, Hallacy C, et al. Learning transferable visual models from natural language supervision. *arXiv*, 2021, <https://doi.org/10.48550/arXiv.2103.00020>.
- [11] Jia C, Yang Y, Xia Y, et al. Scaling up visual and vision-language representation learning with noisy text supervision. *arXiv*, 2021, <https://doi.org/10.48550/arXiv.2102.05918>.
- [12] Ma P, Wu Y, Yu N, et al. UMDFood: Vision-language models boost food composition compilation. *arXiv*, 2023, <https://doi.org/10.48550/arXiv.2306.01747>.
- [13] RomeroTapiador S, Tolosana R, LacruzPleguezuelos B, et al. Are vision-language models ready for dietary assessment? Exploring the next frontier in AI-powered food image recognition. *arXiv*, 2025, <https://doi.org/10.48550/arXiv.2504.06925>
- [14] Kirk D, Kok E, Tufano M, et al. Machine learning in nutrition research. *Adv Nutr*, 2022, 13: 2573–89.
- [15] Panayotova GG. Artificial intelligence in nutrition and dietetics: a comprehensive review of current research. *Healthcare*, 2025, 13: 2579.
- [16] Kirk D, Catal C, Tekinerdogan B. Precision nutrition: a systematic literature review. *Comput Biol Med*, 2021, 133: 104365.
- [17] Zhang S, Callaghan V, Che Y. Image-based methods for dietary assessment: a survey. *J Food Meas Charact*, 2024, 18: 727–43.
- [18] Tahir GA, Loo CK. A comprehensive survey of image-based food recognition and volume estimation methods for dietary assessment. *Healthcare*, 2021, 9: 1676.
- [19] Agrawal K, Goktas P, Kumar N, et al. Artificial intelligence in personalized nutrition and food manufacturing: a comprehensive review of methods, applications, and future directions. *Front Nutr*, 2025, 12: 1636980.
- [20] Lee KH, He X, Zhang L, et al. CleanNet: transfer learning for scalable image classifier training with label noise[C]// Salt Lake City, USA: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 5447–56.
- [21] Marin J, Biswas A, Ofli F, et al. Recipe1M+: a dataset for learning cross-modal embeddings for cooking recipes and food images. *arXiv*, 2019, <https://doi.org/10.48550/arXiv.1810.06553>.
- [22] Papathanail I, Abdur Rahman L, Brigato L, et al. The nutritional content of meal images in free-living conditions—automatic assessment with goFOODTM. *Nutrients*, 2023, 15: 3835.
- [23] Thames Q, Karpur A, Norris W, et al. Nutrition5k: towards automatic nutritional understanding of generic food. *arXiv*, 2021, <https://doi.org/10.48550/arXiv.2103.03375>.
- [24] Wang W, Min W, Li T, et al. A review on vision-based analysis for automatic dietary assessment. *Trends Food Sci Technol*, 2022, 122: 223–37.
- [25] Belkhouribchia J, Pen JJ. Large language models in clinical nutrition: an overview of its applications, capabilities, limitations, and potential future prospects. *Front Nutr*, 2025, 12: 1635682.
- [26] Tsampos I, Marakakis E. DietQA: a comprehensive framework for personalized multi-diet recipe retrieval using knowledge graphs, retrieval-augmented generation, and large language models. *Computers*, 2025, 14: 412.
- [27] Parameswaran V, Bernard J, Bernard A, et al. Evaluating large language models and retrieval-augmented generation enhancement for delivering guideline-adherent nutrition information for cardiovascular disease prevention: cross-sectional study. *J Med Internet Res*, 2025, 27: e78625.
- [28] Tsolakidis D, Gymnopoulos LP, Dimitropoulos K. Artificial intelligence and machine learning technologies for personalized nutrition: a review. *Informatics*, 2024, 11: 62.
- [29] Sultana J, Ahmed BM, Masud MM, et al. A study on food

- value estimation from images: taxonomies, datasets, and techniques. *IEEE Access*, 2023, 11: 45910–35.
- [30] Cheng ST, Lyu YJ, Teng C. Image-based nutritional advisory system: employing multimodal deep learning for food classification and nutritional analysis. *Appl Sci*, 2025, 15: 4911.
- [31] Linseisen J, Renner B, Gedrich K, et al. Data in personalized nutrition: bridging biomedical, psycho-behavioral, and food environment approaches for population-wide Impact. *Adv Nutr*, 2025, 16: 100377.
- [32] Samy K, Osheba A, El-Din HS, et al. Personalized nutrition recommendations using machine learning[C]// Suez, Egypt: 2024 International Conference on Smart-Digital-Green Technologies and Artificial Intelligence Sciences (CSDGAIS), 2024: 1–5.
- [33] Wimalasiri C, Sahoo PK. Vision-based approach for food weight estimation from 2D Images. *arXiv*, 2024, <https://doi.org/10.48550/arXiv.2405.16478>.
- [34] Azimi I, Qi M, Wang L, et al. Evaluation of LLMs accuracy and consistency in the registered dietitian exam through prompt engineering and knowledge retrieval. *Sci Rep*, 2025, 15: 1506.
- [35] Niszczota P, Rybicka I. The credibility of dietary advice formulated by ChatGPT: robo-diets for people with food allergies. *Nutrition*, 2023, 112: 112076.
- [36] Li K JH, Balloccu S, Dusek O, et al. When LLMs can't help: real-world evaluation of LLMs in nutrition. *arXiv*, 2025, <https://doi.org/10.48550/arXiv.2511.20652>.
- [37] Yang E, Garcia T, Williams H, et al. From barriers to tactics: a behavioral science-informed agentic workflow for personalized nutrition coaching. *arXiv*, 2024, <https://doi.org/10.48550/arXiv.2410.14041>.
- [38] Wang H, Tian H, Ju R, et al. Nutritional composition analysis in food images: an innovative Swin Transformer approach. *Front Nutr*, 2024, 11: 1454466.
- [39] Yao D, Yao K, Zhou J, et al. CaLoRAify: calorie estimation with visual-text pairing and LoRA-Driven visual language models. *arXiv*, 2024, <https://doi.org/10.48550/arXiv.2412.09936>.
- [40] Tanabe H, Yanai K. CalorieLLaVA: image-based calorie estimation with multimodal large language models[C]// Palaiahnakote S, Schuckers S, Ogier JM, et al. Pattern recognition. *ICPR 2024 International Workshops and Challenges*. Cham, Switzerland: Springer, 2025: 63–75.
- [41] Yan R, Luo H, Lu J, et al. DietAI24 as a framework for comprehensive nutrition estimation using multimodal large language models. *Commun Med*, 2025, 5: 458.
- [42] Wang X, Sun Z, Xue H, et al. Artificial intelligence applications to personalized dietary recommendations: a systematic review. *Healthcare*, 2025, 13: 1417.
- [43] Ghosh T, McCrory MA, Marden T, et al. I2N: image to nutrients, a sensor guided semi-automated tool for annotation of images for nutrition analysis of eating episodes. *Front Nutr*, 2023, 10: 1191962.
- [44] Kalpakoglou K, Calderón-Pérez L, Boqué N, et al. An AI-based nutrition recommendation system: technical validation with insights from Mediterranean cuisine. *Front Nutr*, 2025, 12: 1546107.
- [45] Héritier H, Allémann C, Balakiriev O, et al. Food & You: a digital cohort on personalized nutrition. *medRxiv*, 2023, <https://doi.org/10.1371/journal.pdig.0000389>.
- [46] Eetemadi A, Rai N, Pereira BMP, et al. The computational diet: a review of computational methods across diet, microbiome, and health. *Front Microbiol*, 2020, 11: 393.
- [47] Qi H, Zhu B, Ngo CW, et al. Advancing food nutrition estimation via visual-ingredient feature fusion. *arXiv*, 2025, <https://doi.org/10.48550/arXiv.2505.08747>.
- [48] Asbroeck SV, Matthys C. Use of different food image recognition platforms in dietary assessment: comparison study. *JMIR Form Res*, 2020, 4: e15602.
- [49] Jiang L, Qiu B, Liu X, et al. DeepFood: food image analysis and dietary assessment via deep model. *IEEE Access*, 2020, 8: 47477–89.
- [50] Shao Z, Vinod G, He J, et al. An end-to-end food portion estimation framework based on shape reconstruction from monocular image. *arXiv*, 2023, <https://doi.org/10.48550/arXiv.2308.01810>.
- [51] Sudharson S, Priyanka K, Annamalai R, et al. Enhanced food classification system using YOLO models for object detection algorithm[C]//2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT). Delhi, India: IEEE, 2023: 1–6.
- [52] Romero-Tapiador S, Lacruz-Pleguezuelos B, Tolosana R, et al. AI4FoodDB: a database for personalized e-Health nutrition and lifestyle through wearable devices and artificial intelligence. *Database J Biol Databases Curation*, 2023, 2023: baad049.
- [53] Nair S, Tai CA, Chen Y, et al. NutritionVerse-Synth: an open access synthetically generated 2D food scene dataset for dietary Intake estimation. *arXiv*, 2023, <https://doi.org/10.48550/arXiv.2312.06192>.
- [54] Kwan Z, Zhang W, Wang Z, et al. Nutrition estimation for dietary management: a transformer approach with depth sensing. *arXiv*, 2024, <https://doi.org/10.48550/arXiv.2406.01938>.
- [55] Shao W, Hou S, Jia W, et al. Rapid non-destructive analysis of food nutrient content using Swin-Nutrition. *Foods*, 2022, 11: 3429.

- [56] Vinod G, He J, Shao Z, et al. Food portion estimation via 3D object scaling. arXiv, 2024, <https://doi.org/10.48550/arXiv.2404.12257>.
- [57] Guo P, Liu G, Xiang X, et al. From AI to the table: a systematic review of ChatGPT's potential and performance in meal planning and dietary recommendations. *Dietetics*, 2025, 4: 7.
- [58] Haman M, Školník M, Lošťák M. AI dietician: unveiling the accuracy of ChatGPT's nutritional estimations. *Nutrition*, 2024, 119: 112325.
- [59] Wang T, Holscher HD, Maslov S, et al. Predicting metabolite response to dietary intervention using deep learning. *Nat Commun*, 2025, 16: 815.
- [60] Gavai AK, van Hillegersberg J. AI-driven personalized nutrition: RAG-based digital health solution for obesity and type 2 diabetes. *PLOS Digit Health*, 2025, 4: e0000758.
- [61] Naja F, Taktouk M, Matbouli D, et al. Artificial intelligence chatbots for the nutrition management of diabetes and the metabolic syndrome. *Eur J Clin Nutr*, 2024, 78: 887–96.
- [62] Kalivaraprasad B, Prasad MVD, Reddy BH. Developing a multimodal deep learning system for comprehensive nutritional analysis of meals for diabetes management. *Int J Intell Syst Appl Eng*, 2024, 12: 780–8.
- [63] 刘宇昕, 闵巍庆, 蒋树强, 等. 多尺度拼图重构网络的食物图像识别. *软件学报*, 2022, 33: 4379–95.  
Liu YX, Min WQ, Jiang SQ, et al. Multi-scale jigsaw reconstruction network for food image recognition. *J Software*, 2022, 33: 4379–95.
- [64] 王海燕, 张渺, 刘虎林, 等. 基于改进的ResNet网络的中餐图像识别方法. *陕西科技大学学报*, 2022, 40: 154–60.  
Wang HY, Zhang M, Liu HL, et al. Chinese food image recognition method based on improved ResNet network. *J Shaanxi Univ Sci Tech*, 2022, 40: 154–60.
- [65] 姜超, 付杰, 甄小琼. 基于YOLOv5的食物识别系统设计与实现. *计算机科学与应用*, 2025, 15: 129–44.  
Jiang C, Fu J, Zhen XQ. Design and implementation of a food recognition system based on YOLOv5. *CSA*, 2025, 15: 129–44.
- [66] 黄惠, 王倩, 骆雅咏, 等. 慢性肾脏病患者的膳食评估. *解放军医学杂志*, 2024, 49: 946–51.  
Huang H, Wang Q, Luo YY, et al. Dietary assessment for patients with chronic kidney disease. *Med J Chin People's Lib Army*, 2024, 49: 946–51.
- [67] 宋静茹, 闵巍庆, 周鹏飞, 等. 基于Transformer的零样本食品图像检测. *食品工业科技*, 2024, 45: 18–26.  
Song JR, Min WQ, Zhou PF, et al. Transformer-based zero-shot food image detection. *Sci Technol Food Indust*, 2024, 45: 18–26.
- [68] 金亦周, 周利, 徐明智. 人工智能在2型糖尿病健康管理中的应用研究进展. *现代医药卫生*, 2023, 39: 4242–7.  
Jin YZ, Zhou L, Xu MZ. Research progress on the application of artificial intelligence in health management of type 2 diabetes. *J Mod Med Health*, 2023, 39: 4242–7.
- [69] 袁祎, 张红梅, 常立阳, 等. 人工智能在慢性肾脏病营养管理中的应用进展. *护理研究*, 2025, 39: 1033–7.  
Yuan Y, Zhang HM, Chang LY, et al. Progress in the application of artificial intelligence in nutritional management of chronic kidney disease. *Chin Nurs Res*, 2025, 39: 1033–7.
- [70] Keller M, Tai CA, Chen Y, et al. NutritionVerse-Direct: exploring deep neural networks for multitask nutrition prediction from food images. arXiv, 2024, <https://doi.org/10.48550/arXiv.2405.07814>.
- [71] Yunus R, Arif O, Afzal H, et al. A framework to estimate the nutritional value of food in real time using deep learning techniques. *IEEE Access*, 2019, 7: 2643–52.
- [72] Tai CA, Keller M, Kerrigan M, et al. NutritionVerse-3D: a 3D food model dataset for nutritional intake estimation. arXiv, 2023, <https://doi.org/10.48550/arXiv.2304.05619>.
- [73] Han Y, Cheng Q, Wu W, et al. DPF-Nutrition: food nutrition estimation via depth prediction and fusion. arXiv, 2023, <https://doi.org/10.48550/arXiv.2310.11702>.
- [74] Larke JA, Chin EL, Bouzid YY, et al. Surveying nutrient assessment with photographs of meals (SNAPMe): a benchmark dataset of food photos for dietary assessment. *Nutrients*, 2023, 15: 4972.
- [75] Hua A, Dhaliwal MP, Burke R, et al. NutriBench: a dataset for evaluating large language models on nutrition estimation from meal descriptions. arXiv, 2025, <https://doi.org/10.48550/arXiv.2407.12843>.
- [76] Min W, Wang Z, Liu Y, et al. Large scale visual food recognition. arXiv, 2023, <https://doi.org/10.48550/arXiv.2103.16107>.
- [77] Chen J, Ngo C. Deep-based ingredient recognition for cooking recipe retrieval[C]//Hanjalic A, Snoek C, Worring M, et al. Proceedings of the 24th ACM International Conference on Multimedia. New York, USA: Association for Computing Machinery, 2016: 32–41.
- [78] Chen X, Zhu Y, Zhou H, et al. ChineseFoodNet: a large-scale image dataset for Chinese food recognition. arXiv, 2017, <https://doi.org/10.48550/arXiv.1705.02743>.
- [79] Wu X, Fu X, Liu Y, et al. A large-scale benchmark for food image segmentation. arXiv, 2021, <https://doi.org/>

- 10.48550/arXiv.2105.05409.
- [80] Bień M, Gilski M, Maciejewska M, et al. RecipeNLG: a cooking recipes dataset for semi-structured text generation [C]//Davis B, Graham Y, Kelleher J, et al. Proceedings of the 13th International Conference on Natural Language Generation. Dublin, Ireland: Association for Computational Linguistics, 2020: 22–8.
- [81] Li W, Zhang X, Li J, et al. FoodieQA: a multimodal dataset for fine-grained understanding of Chinese food culture. arXiv, 2024, <https://doi.org/10.48550/arXiv.2406.11030>.
- [82] Yagcioglu S, Erdem A, Erdem E, et al. RecipeQA: a challenge dataset for multimodal comprehension of cooking recipes. arXiv, 2018, <https://doi.org/10.48550/arXiv.1809.00812>.
- [83] Coburn B, He J, Rollo ME, et al. Comprehensive evaluation of large multimodal models for nutrition analysis: a new benchmark enriched with contextual metadata. arXiv, 2025, <https://doi.org/10.48550/arXiv.2507.07048>.
- [84] Allegra D, Battiato S, Ortis A, et al. A review on food recognition technology for health applications. Health Psychol Res, 2020, 8: 1.
- [85] Bermingham KM, Linenberg I, Polidori L, et al. Effects of a personalized nutrition program on cardiometabolic health: a randomized controlled trial. Nat Med, 2024, 30: 1888–97.
- [86] Rojanaphan P. Precision nutrition: leveraging machine learning for personalized dietary recommendations and health outcomes. Int J Sci Res Manag IJSRM, 2024, 12: 1727–45.
- [87] Ferrario PG, Gedrich K. Machine learning and personalized nutrition: a promising liaison? Eur J Clin Nutr, 2024, 78: 74–6.
- [88] Varshney N, Jadhav N, Gupta K, et al. Personalized dietary recommendations using machine learning: a comprehensive review [C]//Raipur, India: 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI), 2023: 1–6.
- [89] Hosseinian A, Zahedani AD, Mansoor U, et al. January Food Benchmark (JFB): a public benchmark dataset and evaluation suite for multimodal food analysis. arXiv, 2025, <https://doi.org/10.48550/arXiv.2508.09966>.
- [90] Khamesian S, Arefeen A, Carpenter SM, et al. NutriGen: Personalized meal plan generator leveraging large language models to enhance dietary and nutritional adherence [C]//Copenhagen, Denmark: 2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2025: 1–7.
- [91] Gao F, Zhao X, Xia D, et al. HealthGenie: empowering users with healthy dietary guidance through knowledge graph and large language models. arXiv, 2025, <https://doi.org/10.48550/arXiv.2504.14594>.
- [92] Kassem H, Beevi AA, Basheer S, et al. Investigation and assessment of AI's role in nutrition—an updated narrative review of the evidence. Nutrients, 2025, 17: 190.
- [93] Da Silva BR, Brennan L, Horst MA, et al. Advancing precision nutrition: bridging mechanistic insight and clinical implementation. Nat Rev Endocrinol, 2025, 21: 515–7.