INPLASY

MLLMs for Image-Based Food Analysis - Systematic Review

INPLASY202690039
doi: 10.37766/inplasy2026.9.0039
Received: 8 September 2026
Published: 9 September 2026

Cerro, L; Brown, J; Woodside, JV; Green, BD.

Corresponding author:
Brian Green

b.green@qub.ac.uk

Author Affiliation:
Queens University Belfast.

ADMINISTRATIVE INFORMATION

Support - UKRI AI CDT in Sustainable Understandable agri-food Systems Transformed by Artificial INtelligence (SUSTAIN).

Review Stage at time of this submission - Preliminary searches.

Conflicts of interest - None declared.

INPLASY registration number: INPLASY202690039

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 9 September 2026 and was last updated on 9 September 2026.

INTRODUCTION

Review question / Objective To systematically identify, evaluate. And synthesise evidence on the use of multimodal large language models (MLLMs) for dietary and nutritional assessment from food images.

Specifically, this review aims to:

1. Evaluate the accuracy of MLLM-derived estimates of energy and nutrient intake compared with established reference standards.

2. Assess the accuracy of food quantity, portion size, volume, and weight estimation from food images.

3. Examine the performance of MLLMs within clinical nutrition and dietetic applications.

4. Evaluate the reproducibility and transparency of reported MLLM based dietary assessment methods.

5. Identify methodological strengths, limitations, sources of bias, and evidence gaps within the current literature.

Rationale Accurate dietary assessment is important in nutrition research and clinical practice; however, established methods such as food diaries, dietary recalls, and food-frequency questionnaires can be time-consuming, burdensome for participants, and subject to recall and reporting bias. Image-based dietary assessment offers a potential means of reducing this burden by enabling dietary information to be derived directly from photographs of foods and meals.

Recent advances in multimodal large language models (MLLMs) have introduced new capabilities for food-image analysis. Unlike conventional

Cerro et al. INPLASY protocol 202690039. doi:10.37766/inplasy2026.9.0039 Downloaded from https://inplasy.com/inplasy-2026-9-0039/

computer-vision approaches, MLLMs can integrate visual information with natural-language reasoning and may provide detailed estimates of food identity, ingredients, portion size, energy content, and nutrient composition from food images. These capabilities have led to increasing interest in their potential use as dietary-assessment tools in research and clinical nutrition.

However, the evidence regarding the accuracy, reliability, and validity of MLLMs for dietary and nutritional assessment remains emerging and heterogeneous. Studies vary in the models evaluated, dietary-assessment tasks, reference standards, populations or datasets, and performance metrics used. A systematic synthesis of this evidence is therefore needed to establish the current performance of MLLMs for food-image-based dietary and nutritional assessment, identify methodological limitations and gaps in the evidence base, and inform future research and potential applications in nutrition and dietetics.

Condition being studied Dietary assessment; Nutritional assessment; Clinical nutrition; Digital health technologies: Artificial intelligence: Food-image analysis

This review focuses on the use of multimodal large language models (MLLMs) for dietary and nutritional assessment using food images. Specifically, it evaluates the ability of MLLMs to estimate dietary intake, identify food and ingredient identification, and estimate portion size, energy content, and macronutrient and micronutrient content from food-image inputs. The review encompasses applications in dietary assessment, clinical nutrition, and eating-behaviour assessment. Given the rapid emergence of vision-enabled generative AI systems, this review examines their potential as dietary-assessment instruments and explores their performance relative to established reference standards and health professionals in the field of nutrition and dietetics.

METHODS

Search strategy The literature search will be conducted across the following databases: PubMed, Web of Science Core Collection (WoSCC), Scopus, IEEE Xplore, Association for Computing Machinery Digital Library (ACM Digital Library), Embase, and arXiv.

The search strategy will be structured around three core conceptual blocks: (1) multimodal large language models (MLLMs), (2) food and nutrition, and (3) food images. Synonymous terms within each concept will be combined using the Boolean operator OR, while the three conceptual blocks will be combined using AND to identify studies relevant to all aspects of the review question.

Example

1. Pubmed

("multimodal large language model*"[tiab] OR "multi-modallarge language model*"[tiab] OR "multimodal LLM*"[tiab] OR "visionlanguage model*"[tiab] OR "vision-language model*"[tiab] OR"MLLM"[tiab] OR "MLLMs"[tiab] OR "VLM"[tiab] OR "VLMs"[tiab] OR"large multimodal model*"[tiab] OR "LMM"[tiab] OR "LMMs"[tiab] OR"ChatGPT"[tiab] OR "GPT-4"[tiab] OR "GPT-4V"[tiab] OR "GPT-4o"[tiab] OR "GPT-4 with vision"[tiab] OR "Gemini"[tiab] OR "Gemini Pro"[tiab] OR "Gemini Flash"[tiab] OR "Claude 3"[tiab] OR "Claude 3.5"[tiab] OR "Claude 4"[tiab] OR "Claude with vision"[tiab] OR "LLaVA"[tiab] OR "LLaVA-Med"[tiab] OR "MiniGPT-4"[tiab] OR "MiniGPT-v2"[tiab] OR "InstructBLIP"[tiab] OR "BLIP-2"[tiab] OR "Qwen-VL"[tiab]OR "CogVLM"[tiab] OR "InternVL"[tiab] OR "FoodLMM"[tiab] OR (("foundation model*"[tiab] OR "generative AI"[tiab] OR "generativeartificial intelligence"[tiab]) AND ("vision"[tiab] OR "visual"[tiab] OR"image*"[tiab] OR "multimodal"[tiab]))) AND ("Diet"[Mesh] OR"Diet, Food, and Nutrition"[Mesh] OR "Nutrition Assessment"[Mesh]OR "Food"[Mesh] OR "Nutritional Status"[Mesh] OR "Energy Intake"[Mesh] OR "Diet Records"[Mesh] OR "Food Analysis"[Mesh] OR"Meals"[Mesh] OR "diet*"[tiab] OR "dietary"[tiab] OR "dietary intake"[tiab] OR "dietary assessment"[tiab] OR "dietetic*"[tiab] OR"nutrition*"[tiab] OR "nutritional"[tiab] OR "food"[tiab] OR "foods"[tiab] OR "meal*"[tiab] OR "calorie*"[tiab] OR "caloric"[tiab] OR"energy intake"[tiab] OR "macronutrient*"[tiab] OR "micronutrient*"[tiab] OR "portion siz*"[tiab] OR "food recognition"[tiab] OR "foodclassification"[tiab] OR "food log*"[tiab] OR "food diar*"[tiab] OR"food frequency"[tiab] OR "FFQ"[tiab])

AND ("Diagnostic Imaging"[Mesh] OR "Photography"[Mesh] OR "image*"[tiab] OR "imaging"[tiab] OR "photo*"[tiab] OR "picture*"[tiab] OR "visual"[tiab] OR"vision"[tiab] OR "computer vision"[tiab] OR "Camera"[tiab]) AND ("2022/01/01"[DP] : "3000"[DP])

2. Web of Science Core Collection

TS=("multimodal large language model*" OR "multi-modal large language model*" OR

Cerro et al. INPLASY protocol 202690039. doi:10.37766/inplasy2026.9.0039 Downloaded from https://inplasy.com/inplasy-2026-9-0039/

"multimodal LLM*" OR "vision language model*" OR "vision-language model*" OR

"MLLM" OR "MLLMs" OR "VLM" OR "VLMs" OR "large multimodal model*" OR "LMM" OR "LMMs" OR

"ChatGPT" OR "GPT-4" OR "GPT-4V" OR "GPT-4o" OR "GPT-4 with vision" OR

"Gemini" OR "Gemini Pro" OR "Gemini Flash" OR

"Claude 3" OR "Claude 3.5" OR "Claude 4" OR "Claude with vision" OR

"LLaVA" OR "LLaVA-Med" OR "MiniGPT-4" OR "MiniGPT-v2" OR

"InstructBLIP" OR "BLIP-2" OR "Qwen-VL" OR "CogVLM" OR "InternVL" OR "FoodLMM" OR

("foundation model*" AND ("vision" OR "visual" OR "image*" OR "multimodal")) OR

("generative AI" AND ("vision" OR "visual" OR "image*" OR "multimodal")) OR

("generative artificial intelligence" AND ("vision" OR "visual" OR "image*" OR "multimodal")))

AND

TS=("diet*" OR "dietary" OR "dietary intake" OR "dietary assessment" OR

"dietetic*" OR "nutrition*" OR "nutritional" OR "food" OR "foods" OR "meal*" OR

"calorie*" OR "caloric" OR "energy intake" OR "macronutrient*" OR "micronutrient*" OR

"portion siz*" OR "food recognition" OR "food classification" OR

"food log*" OR "food diar*" OR "food frequency" OR "FFQ")

AND

TS=("image*" OR "imaging" OR "photo*" OR "picture*" OR

"visual" OR "vision" OR "computer vision" OR "camera*")

AND PY=(2022-2026)

3. Scopus

TITLE-ABS-KEY("multimodal large language model*" OR "multi-modal large language model*" OR

"multimodal LLM*" OR "vision language model*" OR "vision-language model*" OR

"MLLM" OR "MLLMs" OR "VLM" OR "VLMs" OR "large multimodal model*" OR "LMM" OR "LMMs" OR

"ChatGPT" OR "GPT-4" OR "GPT-4V" OR "GPT-4o" OR "GPT-4 with vision" OR

"Gemini" OR "Gemini Pro" OR "Gemini Flash" OR

"Claude 3" OR "Claude 3.5" OR "Claude 4" OR "Claude with vision" OR

"LLaVA" OR "LLaVA-Med" OR "MiniGPT-4" OR "MiniGPT-v2" OR

"InstructBLIP" OR "BLIP-2" OR "Qwen-VL" OR "CogVLM" OR "InternVL" OR "FoodLMM" OR

("foundation model*" AND ("vision" OR "visual" OR "image*" OR "multimodal")) OR

("generative AI" AND ("vision" OR "visual" OR "image*" OR "multimodal")) OR

("generative artificial intelligence" AND ("vision" OR "visual" OR "image*" OR "multimodal")))

AND

TITLE-ABS-KEY("diet*" OR "dietary" OR "dietary intake" OR "dietary assessment" OR

"dietetic*" OR "nutrition*" OR "nutritional" OR "food" OR "foods" OR "meal*" OR

"calorie*" OR "caloric" OR "energy intake" OR "macronutrient*" OR "micronutrient*" OR

"portion siz*" OR "food recognition" OR "food classification" OR

"food log*" OR "food diar*" OR "food frequency" OR "FFQ")

AND

TITLE-ABS-KEY("image*" OR "imaging" OR "photo*" OR "picture*" OR

"visual" OR "vision" OR "computer vision" OR "camera*")

Cerro et al. INPLASY protocol 202690039. doi:10.37766/inplasy2026.9.0039 Downloaded from https://inplasy.com/inplasy-2026-9-0039/

AND PUBYEAR > 2021 AND PUBYEAR < 2027

4. IEEE Xplore
("Abstract":"multimodal large language model" OR "Abstract":"multimodal large language models"

OR "Abstract":"multi-modal large language model" OR "Abstract":"multi-modal large language models"

OR "Abstract":"vision language model" OR "Abstract":"vision language models"

OR "Abstract":"vision-language model" OR "Abstract":"vision-language models"

OR "Abstract":"large multimodal model" OR "Abstract":"large multimodal models"

OR "Abstract":MLLM OR "Abstract":MLLMs OR "Abstract":VLM OR "Abstract":VLMs OR "Abstract":LMM

OR "Abstract":LMMs OR "Abstract":"multimodal LLM" OR "Abstract":"multimodal LLMs"

OR "Abstract":ChatGPT OR "Abstract":"GPT-4" OR "Abstract":"GPT-4V" OR "Abstract":"GPT-4o"

OR "Abstract":"GPT-4 with vision" OR "Abstract":Gemini OR "Abstract":"Gemini Pro"

OR "Abstract":"Gemini Flash" OR "Abstract":"Claude 3" OR "Abstract":"Claude 3.5"

OR "Abstract":"Claude 4" OR "Abstract":"Claude with vision" OR "Abstract":LLaVA

OR "Abstract":"LLaVA-Med" OR "Abstract":"MiniGPT-4" OR "Abstract":"MiniGPT-v2"

OR "Abstract":InstructBLIP OR "Abstract":"BLIP-2" OR "Abstract":"Qwen-VL" OR "Abstract":CogVLM

OR "Abstract":InternVL OR "Abstract":FoodLMM)

AND

("Abstract":diet OR "Abstract":dietary OR "Abstract":"dietary assessment" OR "Abstract":"dietary intake"

OR "Abstract":nutrition OR "Abstract":nutritional OR "Abstract":food OR "Abstract":foods

OR "Abstract":meal OR "Abstract":meals OR "Abstract":calorie OR "Abstract":calories

OR "Abstract":"energy intake" OR "Abstract":macronutrient OR "Abstract":macronutrients

OR "Abstract":"portion size" OR "Abstract":"food recognition" OR "Abstract":"food classification"

OR "Abstract":"food logging" OR "Abstract":"food diary" OR "Abstract":FFQ)
AND
("Abstract":image OR "Abstract":images OR "Abstract":imaging OR "Abstract":photo
OR "Abstract":photograph OR "Abstract":visual OR "Abstract":vision OR "Abstract":"computer vision")
AND ("Publication Year":2022 OR "Publication Year":2023 OR "Publication Year":2024
OR "Publication Year":2025 OR "Publication Year":2026)

Search strings have also been devised for:
5. Association of Computing Machinery Digital Library (ACM DL)
6. Embase (Ovid)
7. ArXiv
There is insufficient space here to outline these.

Participant or population Included
1. Human participants in studies where MLLM-derived dietary or nutritional information is generated, evaluated against a reference standard, validated, or applied using food images.
2. Clinical and non-clinical populations in nutrition, dietetics, clinical nutrition, or eating-behaviour contexts where food-images baased assessment is undertaken using MLLMs.
3. Curated food-image datasets (e.g., food-101, Nutrition 5K, FoodSeg103, Recipe1M, Periodic Table of Food Initiative) used to evaluate MLLMs for dietary assessment, nutritional estimation, clinical nutrition or eating-behaviour related tasks.
4. Studies involving food images as the source of dietary or nutritional assessment outputs.

Excluded
1. Studies not involving human participants or curated food-image datasets relevant to dietary, nutritional, dietetic, clinical nutrition, or eating-behaviour applications.
2. Studies using non-food-image datasets only.
3. Studies involving food images but without a dietary, nutritional, dietetic, clinical nutrition or eating-behaviour application.
4. Pure benchmark studies lacking nutrition or dietetic context.

Cerro et al. INPLASY protocol 202690039. doi:10.37766/inplasy2026.9.0039 Downloaded from https://inplasy.com/inplasy-2026-9-0039/

5. Animal studies will be excluded.

Intervention Included
1. Studies applying a multimodal large language model (MLLM) to food images for dietary assessment, nutritional assessment, clinical nutrition or eating-behaviour related tasks.
2. MLLMs operationalised as models integrating a transformer-based large language model backbone with visual input and producing natural-language outputs.
3. Commercial, proprietary, and open-source MLLMs capable of processing food images and generating dietary, nutritional or dietetic outputs (e.g., GPT-4V/4o, Gemini, Claude Vision, LLaVA, MiniGPT-4, Qwen-VL, FoodLMM).
4. Hybrid systems in wich outputs from computer vision, image segmentation, ood recognition, or related methods are incorporated into an MLLM reasoning step to generate dietary or nutritional outputs.
5. Studies involving development, adaptation, fine-tuning, evaluation, validation, or application of eligible MLLMs.
6. Studies published from 1st January 2022 onwards.

Excluded

1. Studies using traditional computer vision, machine learning, or deep learning approaches without an MLLM component.
2. Studies using text-only large language models without visual food-image input.
3. Studies using dual-encoder vision-language models (e.g., CLIP-only approaches) without a generative LLM reasoning component.
4. Studies where food images are not used as an input to the model.
5. Studies involving MLLMs for non-dietary or non-nutritional applications.
6. Studies published before 1st January 2022.

Comparator Included
1. Established dietary assessment methods, including weighed food records, food diaries, food records, 24-hour dietary recalls, validated FFQs, and other recognized dietary assessment instruments.
2. Registered dietician assessment, nutritionist assessment, expert annotators, or otherqualified human raters.
3. Laboratory-measured nutrient analysis, chemical composition analysis, or validated nutritional databases used as reference standards.
4. Alternative computational approaches, including CNN-based systems, classical computer vision methods, food-recognition systems, or non-generative vision-language models.
5. Labelled datasets, benchmark datasets, ground-truth annotations, predefined reference values, or gold-standard classifications.
6. Studies without an explicit comparator, provided objective performance metrics are reported against labelled datasets, ground truth, expert annotations, benchmark datasets, or predefined reference standards.

Excluded
1. Studies reporting MLLM outputs without any objective evaluation, reference standard, ground truth, benchmark, expert assessment, or performance metric.
2. Studies where performance cannot be assessed due to the absence of sufficient evaluation data.

Study designs to be included 1. Diagnostic-accuracy studies; 2. Validation studies; 3. Comparative evaluation studies; 4. Cross-sectional evaluation studies; 5. Real-world deployment or implementation studies; 6. Development-and-validation studies; 7. Primary research studies; 8. Conference papers where the full paper is available; 9. Peer-reviewed publications and eligible preprints where no peer-reviewed version exists.

Eligibility criteria Verbatim protocol definition
Studies are eligible if the food-image analysis component uses a multimodal large language model (MLLM), defined as a model that integrates a transformer-based language model backbone with visual input and produces generative natural-language outputs (e.g., GPT-4V/4o, Gemini 1.x/2.x, Claude 3/3.5/4 with vision, LLaVA family, MiniGPT-4, InstructBLIP, Qwen-VL, CogVLM, InternVL, food-specific fine-tunes).
Studies using only CNN-based food recognition or dual-encoder VLMs (e.g., CLIP) without a generative LLM reasoning step are excluded. Hybrid pipelines routing CNN/segmentation output into an MLLM reasoning step are included.

Date boundary
Single boundary: 2022-01-01 to present.

Rationale: 2022-01-01 captures the immediate pre-MLLM vision-language wave (Flamingo, early BLIP-2) where some studies qualify as MLLMs under the operational definition above and others do not - eligibility is determined by the operational definition (transformer LLM backbone + visual input + generative NL output), not by date. Pure dual-encoder VLMs (CLIP) and CNN-only food recognition remain excluded regardless of

Cerro et al. INPLASY protocol 202690039. doi:10.37766/inplasy2026.9.0039 Downloaded from https://inplasy.com/inplasy-2026-9-0039/

publication year. No sensitivity-analysis arm at 2023-01-01.

Pre-adjudicated borderline examples
CLIP + LLM-decoder pipelines - include if decoder is true LLM (≥1B parameters, generative)
goFOOD / specialist food systems - include if integrates an MLLM component
Text-only LLM evaluated on food images via captioning - exclude
MLLM-as-judge studies - include only if a food-image task is being evaluated
CNN + caption generator with template output - exclude
BLIP-2 with frozen LLM - include (borderline, justify)
LLaVA-Med fine-tuned on food - include
GPT-4 receiving only OCR text from a food package - exclude.

Information sources The literature search will be conducted across the following databases: PubMed, Web of Science Core Collection (WoSCC), Scopus, IEEE Xplore, Association for Computing Machinery Digital Library (ACM Digital Library), Embase, and arXiv.
PubMed; Purpose: Master search development and validation
Scopus; Purpose:Multidisciplinary coverage
Web of Science Core Collection; Purpose: Confirmatory multidisciplinary coverage
IEEE Xplore ; Purpose: Computer science and engineering literature
ACM Digital Library; Purpose: Computer science literature
Embase; Purpose: Clinical and biomedical literature
arXiv; Purpose: Recent preprints and non-indexed literature.

Main outcome(s) Included
Primary Outcome
1. Measurement accuracy of MLLM-estimated energy and macronutrient content (protein, fat, carbohydrate) derived from food images, compared against an established reference standard.
2. Accuracy reported using recognised quantitative metrics, including but not limited to:
- Mean Absolute Error (MAE)
- Mean Absolute Percentage Error (MAPE)
- Root Mean Squared Error (RMSE)
- Other comparable measures of agreement, validity, or error.
3. Reference standards may include weighed food records, registered dietician assessment, laboratory nutrient analysis, validated dietary assessment methods, and where reported, doubly labelled water biomarkers.

Secondary Outcomes
4. Portion, volume, or weight estimation accuracy, including error in estimated food quantity (grams, milliliters, servings, or portions) compared with reference standard.
5. Performance within clinical populations where dietary assessment informs clinical decision-making, including but not limited to:
- Type 1 diabetes (e.g., carbohydrate counting)
- Type 2 diabetes
- Chronic kidney disease
- Other nutrition-related clinical populations
6. Reproducibility and transparency of MLLM performance, including reporting of prompts, model identifiers and version, decoding parameters, retrieval augmentation, fine-tuning procedures, and other information required to reproduce results.
7. Measures of agreement, reliability, robustness, or consistency were reported.

Excluded
1. Studies reporting only food identification, food classification, food detection, segmentation or recognition performance without assessing dietary or nutritional outcomes.
2. Studies focused solely on methodological comparisons (e.g., zero-shot versus fine-tuned models, or architectural comparisons) without reporting dietary or nutritional assessment outcomes.
3. Studies reporting usability, acceptability, user satisfaction, feasibility, or clinician perceptions as the primary outcome without evaluation of dietary or nutritional assessment performance.
4. Studies reporting only computational performance metrics (e.g., inference speed, model size, computation efficiency, token usage, or hardware requirements) without assessment of dietary or nutritional outcomes.
5. Studies lacking objective dietary, nutritional, or assessment-related outcome measures.

Additional outcome(s) Additional outcomes will include accuracy of food identification and ingredient recognition; estimation of food quantity, portion size, volume, and weight; performance within clinical nutrition and dietetic applications; and comparisons with health professionals or alternative computational approaches where reported.

Additional methodological outcomes will include reproducibility and transparency of MLLM-based dietary assessment methods, including reporting

Cerro et al. INPLASY protocol 202690039. doi:10.37766/inplasy2026.9.0039 Downloaded from https://inplasy.com/inplasy-2026-9-0039/

of model and version, prompting procedures, image-processing methods, repeated measurements, and other relevant implementation details. Reported model errors, failure modes, and hallucinations will also be extracted where available.

Data management Zenodo will be used to log relevant files at the relevant times they arise.
DOI addresses will be provided in any subsequent peer-review publications when they will be placed in the public domain.
For example deposits could include
1. MLLM Food SR - Protocol
2. MLLM Food SR - Search Strategies
3. MLLM Food SR - Data
4. MLLM Food SR - Analysis
5. MLLM Food SR - Manuscript.

Quality assessment / Risk of bias analysis Risk of bias will be assessed using an appropriate study-design-specific framework. PROBAST+AI will be used where applicable for studies evaluating MLLM-based prediction or estimation methods, with alternative validated risk-of-bias tools applied and justified where appropriate according to study design.

Relevant AI-specific methodological considerations, including dataset selection, reference-standard quality, model and version specification, prompting procedures, image preprocessing, repeated testing, and independence of model evaluation, will also be considered. Risk-of-bias judgements will inform the interpretation and synthesis of the review findings.

Strategy of data synthesis A narrative synthesis will be conducted using a pre-specified dual-axis framework (application sub-domain × MLLM methodology) with reporting informed by relevant SWiM principles. Application sub-domains will include dietary assessment, nutritional estimation, clinical nutrition, and behaviour/adherence. MLLM methodologies will include zero-shot and few-shot prompting, fine-tuning, retrieval-augmented generation (RAG), hybrid approaches, and agentic/tool-use approaches.

Included studies will be mapped across this framework, with study characteristics, reference standards, and performance outcomes summarised descriptively and narratively. Areas for which no eligible evidence is identified will be highlighted as research gaps.

No formal meta-analysis is planned due to anticipated heterogeneity in study designs, MLLMs, datasets, reference standards, dietary-assessment tasks, and reported outcome measures.

Subgroup analysis Dietary assessment
Studies using MLLMs to estimate or characterise individual dietary intake (what was eaten, how much, when). Identifying foods eaten, eating occasions, meal composition, and dietary intake. GPT-4V estimating meals from food-diary photos vs weighed records Nutritional estimation Studies using MLLMs to estimate nutrient content (calories, macros, micronutrients) of foods. Estimating calories, macronutrients, micronutrients, or nutrient density. MLLM + RAG over USDA FDC to estimate kcal/macros from meal photos
Clinical nutrition
Studies applying MLLMs in clinical-dietetic contexts (hospital, outpatient, condition-specific) Renal-diet adherence monitoring via meal photos with MLLM analysis
Behaviour & adherence
Studies using MLLMs to support behaviour change, eating-behaviour assessment, intervention adherence tracking MLLM-coded eating-occasion classification for behaviour-change interventions.

Sensitivity analysis Seed-paper recall validation was undertaken to assess the sensitivity of the master PubMed search strategy v1.0 before it was locked. A sample of landmark studies known to satisfy the review eligibility criteria was identified from scoping
searches. PubMed-indexed seed papers were expected to be retrieved by the final search strategy. Failure to retrieve an eligible seed paper triggered iterative refinement of the search strategy until satisfactory recall was achieved. Papers that were not indexed in PubMed (e.g., arXiv preprints) were recorded but were not considered search failures, as they were expected to be identified through subsequent database-specific searches.

Seed paper: Lo et al. 2024, “Dietary Assessment with Multimodal ChatGPT: A Systematic Analysis” (IEEE J Biomed Health Inform 2024)
Identifier: PMID 38900623
Expected in PubMed?: Yes
Retrieved by v1.0?: Yes
Search Concept(s) responsible/Notes: MLLM block: ChatGPT[tiab] (added at iteration 1). Nutrition block: dietary assessment[tiab]. Image block: Camera*[tiab] (added at v1.0 - the abstract uses “wearable cameras”, not “image”).

Cerro et al. INPLASY protocol 202690039. doi:10.37766/inplasy2026.9.0039 Downloaded from https://inplasy.com/inplasy-2026-9-0039/

Seed paper: Romero-Tapiador et al. 2025, “Are Vision-Language Models Ready for Dietary Assessment?” (arXiv, 2024)
Identifier: arXiv:2504.06925
Expected in PubMed?: No (arXiv only)
Retrieved by v1.0?: Not applicable
Search Concept(s) responsible/Notes: Expected absence from PubMed; will be captured by subsequent arXiv/IEEE searches

Seed paper: Yan et al. 2025 “DietAI24 as a framework for comprehensive nutrition estimation using multimodal large language models” (Communications Medicine 2025)
Identifier: PMID 41193610
Expected in PubMed?: Yes
Retrieved by v1.0?: Yes
Search Concept(s) responsible/Notes: Retrieved already; no iteration required

Seed paper: Yin et al. 2024 “FoodLMM: A Versatile Food Assistant using Large Multi-modal Model” (arXiv, 2024)
Identifier: arXiv:2312.14991
Expected in PubMed?: No
Retrieved by v1.0?:Not applicable
Search Concept(s) responsible/Notes: Expected absence; will be captured by subsequent arXiv/IEEE searches

Seed paper: O’Hara et al. 2025 “An Evaluation of ChatGPT for Nutrient Content Estimation from Meal Photographs” (Nutrients 2025)
Identifier: O’Hara et al. 2025 “An Evaluation of ChatGPT for Nutrient Content Estimation from Meal Photographs” (Nutrients
2025)
Expected in PubMed?: Yes
Retrieved by v1.0?: Yes
Search Concept(s) responsible/Notes: MLLM block: “language and vision models”[tiab]. Nutrition block: “meal” “dietary assessment”[tiab]. Image block:"photo*"[tiab].

Language restriction English-language publications only.

Country(ies) involved United Kingdom.

Other relevant information Two independent reviewers will screen titles, abstracts, and full texts in two phases. Any disagreements will be resolved through discussion or consultation with a third reviewer. Screening procedures will follow standard systematic review methods.

Keywords Nutrition; Dietetics; Dietary assessment; Nutritional assessment; Clinical nutrition; Digital health; Artificial intelligence; Generative AI; Multimodal large language models; Food-image analysis.

Dissemination plans Submission to a leading international Human Nutrition or Computer Science peer-reviewed Journal.

Contributions of each author

Author 1 - Laura Cerro.
Email: lcerro01@qub.ac.uk
Author 2 - James Brown.
Email: jamesbrown@lincoln.ac.uk
Author 3 - Jayne Woodside.
Email: j.woodside@qub.ac.uk
Author 4 - Brian Green.
Email: b.green@qub.ac.uk

Cerro et al. INPLASY protocol 202690039. doi:10.37766/inplasy2026.9.0039 Downloaded from https://inplasy.com/inplasy-2026-9-0039/