Elevate Series

REVIEW ARTICLE

AI-based digital image dietary assessment methods compared to humans and ground truth: a systematic review

Eleanor Shonkoff[a], Kelly Copeland Cara[b], Xuechen (Anna) Pei[b], Mei Chung[b], Shreyas Kamath[c], Karen Panetta[c] and Erin Hennessy[b,d]

[a]School of Health Sciences, Merrimack College, North Andover, MA, USA; [b]Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA, USA; [c]School of Engineering, Tufts University, Medford, MA, USA; [d]ChildObesity180, Tufts University, Boston, MA, USA

ABSTRACT

Objective: Human error estimating food intake is a major source of bias in nutrition research. Artificial intelligence (AI) methods may reduce bias, but the overall accuracy of AI estimates is unknown. This study was a systematic review of peer-reviewed journal articles comparing fully automated AI-based (e.g. deep learning) methods of dietary assessment from digital images to human assessors and ground truth (e.g. doubly labelled water).

Materials and Methods: Literature was searched through May 2023 in four electronic databases plus reference mining. Eligible articles reported AI estimated volume, energy, or nutrients. Independent investigators screened articles and extracted data. Potential sources of bias were documented in absence of an applicable risk of bias assessment tool.

Results: Database and hand searches identified 14,059 unique publications; fifty-two papers (studies) published from 2010 to 2023 were retained. For food detection and classification, 79% of papers used a convolutional neural network. Common ground truth sources were calculation using nutrient tables (51%) and weighed food (27%). Included papers varied widely in food image databases and results reported, so meta-analytic synthesis could not be conducted. Relative errors were extracted or calculated from 69% of papers. Average overall relative errors (AI vs. ground truth) ranged from 0.10% to 38.3% for calories and 0.09% to 33% for volume, suggesting similar performance. Ranges of relative error were lower when images had single/simple foods.

Conclusions: Relative errors for volume and calorie estimations suggest that AI methods align with – and have the potential to exceed – accuracy of human estimations. However, variability in food image databases and results reported prevented meta-analytic synthesis. The field can advance by testing AI architectures on a limited number of large-scale food image and nutrition databases that the field determines to be adequate for training and testing and by reporting accuracy of at least absolute and relative error for volume or calorie estimations.

KEY MESSAGES

- These results suggest that AI methods are in line with – and have the potential to exceed – accuracy of human estimations of nutrient content based on digital food images.
- Variability in food image databases used and results reported prevented meta-analytic synthesis.
- The field can advance by testing AI architectures on a limited number of large-scale food image and nutrition databases that the field determines to be accurate and by reporting accuracy of at least absolute and relative error for volume or calorie estimations.
- Overall, the tools currently available need more development before deployment as stand-alone dietary assessment methods in nutrition research or clinical practice.

Introduction

Artificial intelligence (AI) methods such as deep learning [1] are now being used to further nutrition science [2–4], with applications to food classification, image- and non-image-based assessment of dietary intake, and identification of biomarkers of foods or nutrients [2,4], among others. Additionally, the nutrition field has also seen development and growth in investigating digital images (DIs) in dietary assessment[5]. DIs have been used successfully to supplement other methods of evaluating dietary intake – such as

enhancing written food records or image-assisted 24h dietary recall – and can be a stand-alone tool when images are high-quality and adequately capture all foods pre-consumption[5] .This review investigates literature at the intersection of AI and DI dietary assessment.

AI-based systems may revolutionize capabilities for accuracy, speed, or complex pattern recognition; their use in clinical healthcare settings for vision-based evaluation may be more accurate than human doctors for detection of cancerous pulmonary nodules or fractured wrists on radiology scans[6] .In clinical nutrition, AI techniques have been used to aid in diet optimization for health conditions, food image recognition, risk prediction, self-monitoring of dietary intake, precision nutrition, and analyzing links between diet patterns and health outcomes [1,2]. AI image-based processes have been tested for assessing nutrient intake in hospitalized patients [7], estimating protein content of supplement powders [8], fully-automating calorie intake estimation [9], estimating carbohydrate content of foods for diabetics [10], and estimating children's fruit and vegetable consumption [11]. There is a significant range of potential applications for this technology in the nutrition field.

One specific line of research has been to evaluate the potential of AI dietary assessment methods for self-monitoring dietary intake, which is consistently found to be associated with weight loss [12] but people often underestimate dietary intake (especially energy) [13,14], particularly those with obesity [15]. Digital image dietary assessment has been shown to reduce underreporting [16]. Technology-supported self-monitoring may improve dietary changes, adherence, and anthropometric outcomes compared to other methods [17,18]. Findings from the burgeoning field of AI-assisted dietary assessment suggest potential to estimate portion size [19,20], carbohydrate content [21], and calorie content of foods [22,23].

AI methods can be built to rely on some human input for algorithm completion (i.e. 'semi-automated' or 'semi-automatic'), or they could aim to go from digital images to estimated diet or nutrition-related information through a fully-automated or automatic process (i.e. independent of the user's input) [24,25]. In contrast to other recent work (e.g. Doulah et al. [26]), we focus on fully-automatic processes in this review. Steps of the fully automatic process include food segmentation, classification, volume and nutrient estimation. Recent reviews have found that convolutional neural networks (CNNs) are frequently used across steps of the fully automatic process in dietary assessment using images [27,28], and that they performed better on large publicly available datasets than other approaches (e.g. hand-extracted features fed to a traditional machine learning classifier) [27]. Three recent reviews have examined AI-based dietary assessment based on images [24,25,27]. Dalakleidi et al. [27] found that estimating volume, calorie, and nutrients were the least-researched portion of the process and noted challenges such as lack of depth information and annotated datasets. Wang et al. [24] reviewed and discussed many vision-based methods for automatic dietary analysis and found the need for a large-scale benchmark dataset. Hochsmann and Martin [25] reviewed image-based dietary assessment methods that included human and AI approaches; they concluded that human management of input from digital images seems necessary to ensure accurate results at this stage of development. Finally, Kaur et al.'s [29] systematic review summarized a variety of deep neural networks that have been used to analyze digital images for nutrient information and identified some common approaches and databases.

Despite these reviews, what remains unknown is the accuracy of the full array of fully automated AI-techniques at estimating energy consumed compared to best-practices (e.g. weighed plate waste). AI shows promise as a route for meeting the critical need for accurate, low-burden dietary assessment. AI-based digital image methods have the potential to reduce burden when they require fewer tools (e.g. smartphone instead of a heavy scale). Critical review of this array of techniques is needed, particularly as these types of technologies become increasingly deployed for use by the public. For instance, in some countries, smartphones are already on the market equipped with camera technology purported to provide feedback about calorie content of foods [30,31]. The objective of this study was to conduct a systematic review of the literature comparing fully automated AI-based methods of dietary assessment from digital images to human assessors and to ground truth.

Materials and methods

The key question of this systematic review is: How similar are AI techniques to human beings or ground truth when analyzing digital images of food for diet related features? Study methods followed the National Academy of Medicine's Standards for Systematic Reviews [32] and the Cochrane Handbook for Systematic Reviews of Interventions [33]. Results are reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [34]. A study protocol was developed prior

to data extraction (available on request). This review was not registered.

Data sources and searches

Search strategies limited results to articles reporting on the use of AI methods to process and analyze digital food images. Search terms included 'artificial intelligence' and related terms (e.g. deep learning, machine learning, neural network), terms related to image coding or classification, and common dietary assessment terminology (e.g. nutrition assessment, food analysis, food intake, volume, quantity, portion, calories) (see Supplemental eTable 1). Database searches were conducted in Ovid MEDLINE® (1946 to May 26, 2023) and Cochrane Central Register of Controlled Trials (1991 to May 26, 2023), Embase (1966 to May 26, 2023), and Web of Science Core Collection (1900 to May 27, 2023) plus reference mining in related review articles. Only articles published in peer-reviewed journals were included. For relevant conference presentations, we performed manual searches for peer-reviewed publications and affiliated labs.

Study selection

After removing duplicate citations, two independent investigators screened all titles, abstracts, and full-text articles in Covidence online software based on eligibility criteria outlined in Supplemental eTable 2. Disagreements were resolved by consensus or a third investigator. Articles were eligible if they reported comparisons of AI methods to digital food image assessment by human assessors (typically dietitians) or to ground truth as defined by the original study. Supplementary eTable 3 presents excluded articles and reasons.

Data extraction

Two independent investigators extracted study characteristic and outcome data and resolved discrepancies by discussion, or a third investigator resolved them. Relative errors for AI-estimated energy or nutrients were extracted or calculated (|actual − estimated|/actual)*100) from published results when possible to allow for comparability across papers (in contrast to absolute error, which was paper-specific). Thus, the relative errors represented the difference between AI-estimated results and the ground truth. For example, if a paper defined 'ground truth' for calories in a hamburger as the weight (g) of the burger multiplied by the calories per gram, then a relative error of 1% would mean that the AI system's estimation of calories differed from the weighed plate waste measurement of calories by only 1% (indicating high accuracy). When multiple relative errors were presented, the result from the best-performing or main proposed version of the AI architecture was extracted (see Supplemental eTable 4). If available, we extracted reported average relative errors (over various food types or methods) and highest and lowest reported individual relative errors (from the best-performing architecture or iteration).
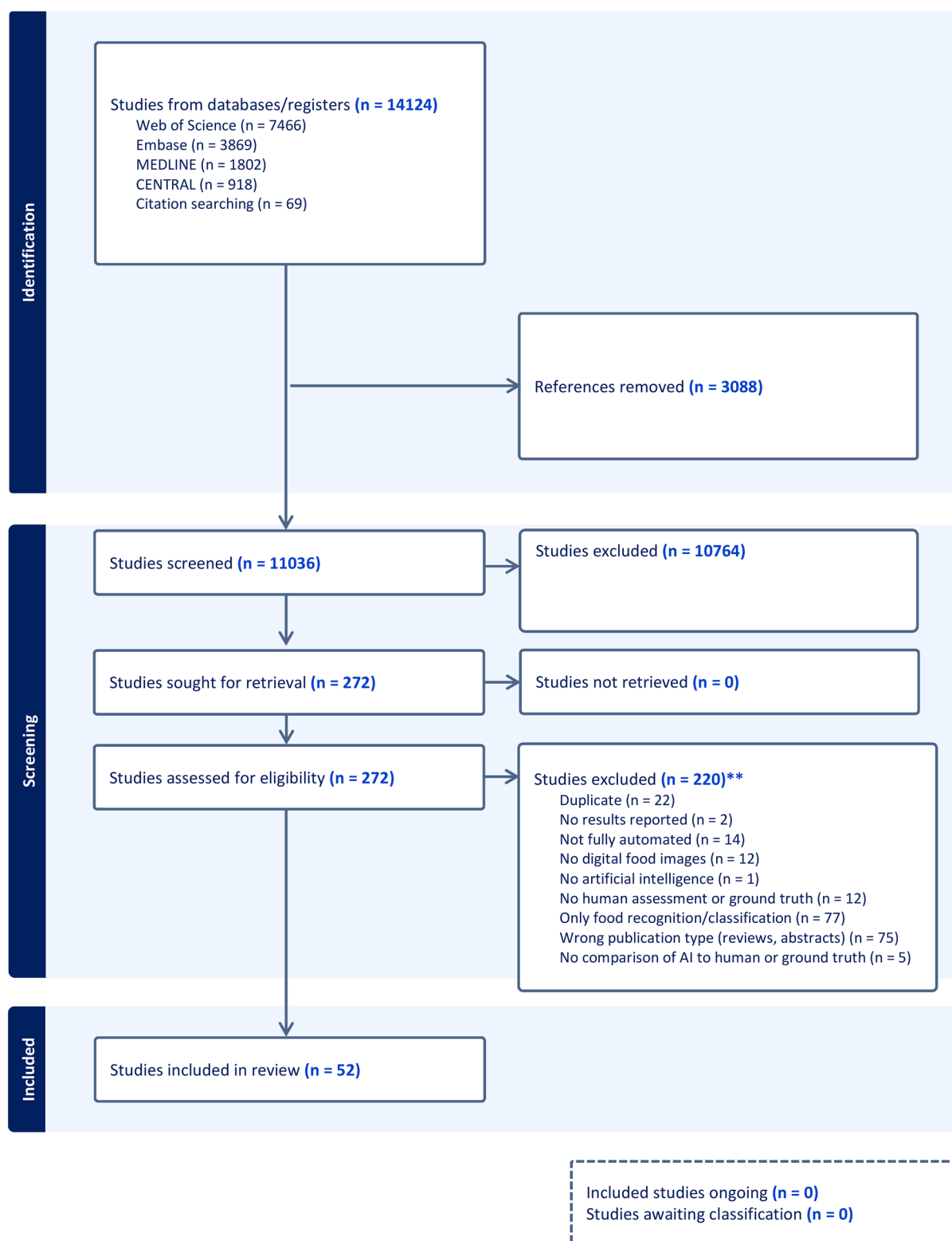
Risk of bias

To our knowledge, no tool exists to assess risk of bias for the types of AI studies included in this review where system input – not sample size or study design – are considered the main source of bias. Therefore, AI engineers were consulted regarding potential sources of bias at any stage of the image analysis or accuracy evaluation processes. Two independent investigators documented these potential biases for each study, and discrepancies were adjudicated by a third investigator (see Supplemental eTable 5).

Data synthesis

The main outcome was accuracy of AI methods to estimate dietary components. A narrative synthesis of all included studies and summary tables are presented. Due to high heterogeneity in reported effect measures and study methods, data were not appropriate to perform meta-analyses [33]. Instead of calculating an overall effect size, forest plots of individual study results are presented to facilitate qualitative synthesis for papers with relative error for calories or volume. Plots were created using Stata/SE 17.0 for Windows.

Results

The literature selection process is summarized [35] in Figure 1. Altogether, 52 papers published from 2010 to 2023 (44 papers published after 2015) were included [7–11,19–23,36–77]. Study authors were affiliated with institutions across 17 countries, with 35% in the United States (see Table 1). Studies received funding from government (44%), private (33%), education (25%) or other sources (2%), with 33% reporting no sources of external funding. The most commonly reported outcomes were estimated calories (52% of studies) and volume (40%). Most papers (81%) compared results from AI methods to ground truth alone. The two

Studies from databases/registers **(n = 14124)**
    Web of Science (n = 7466)
    Embase (n = 3869)
    MEDLINE (n = 1802)
    CENTRAL (n = 918)
    Citation searching (n = 69)

References removed **(n = 3088)**

Studies screened **(n = 11036)**

Studies excluded **(n = 10764)**

Studies sought for retrieval **(n = 272)**

Studies not retrieved **(n = 0)**

Studies assessed for eligibility **(n = 272)**

Studies excluded **(n = 220)****
    Duplicate (n = 22)
    No results reported (n = 2)
    Not fully automated (n = 14)
    No digital food images (n = 12)
    No artificial intelligence (n = 1)
    No human assessment or ground truth (n = 12)
    Only food recognition/classification (n = 77)
    Wrong publication type (reviews, abstracts) (n = 75)
    No comparison of AI to human or ground truth (n = 5)

Studies included in review **(n = 52)**

Included studies ongoing **(n = 0)**
Studies awaiting classification **(n = 0)**

Identification

Screening

Included

** Reasons for excluded papers can be found in Supplementary eTable

Figure 1. AI SR project.

common definitions of ground truth were calculation using nutrient information [7,9, 10,22,23,39,41,42, 45,46,49,51–53,55,57,60,62–64,67,71,73–76] (e.g. a database with nutrition information, such a a USDA database; 51%) and weighed food (27%) [7–10,38,39,42,44,48,52,56,58,66,72] (see Table 1 and Supplemental eTable 4). No studies directly measured energy content of foods (i.e. bomb calorimetry).

Table 1. Summary of study characteristics.

Characteristic	N (of studies)	% of 52 studies[a]
Authors' country affiliations		
USA	18	35%
China	10	19%
India	9	17%
Switzerland	8	15%
Japan	6	12%
Australia	2	4%
Indonesia	2	4%
Malaysia	2	4%
UK	2	4%
South Korea	2	4%
Italy	1	2%
Saudi Arabia	1	2%
Bangladesh	1	2%
Canada	1	2%
Kuwait	1	2%
New Zealand	1	2%
Thailand	1	2%
Source of funding		
Government	23	44%
Private (hospital, non-profit, corporate, etc.)	17	33%
Reported no external source or did not report a source	17	33%
Education (e.g. University)	13	25%
Other (e.g. crowdfunding)	1	2%
Food measure estimated by AI		
Calories	27	52%
Volume	21	40%
Carbohydrate	12	23%
Weight or Mass	11	21%
Protein	11	21%
Fat	8	15%
Area	5	10%
Other (e.g. nutrient, ingredient)	9	17%
Fiber	4	8%
Portion/serving/count	1	2%
Definitions of ground truth		
Calculation using nutrient information	26	51%
Weighed food	14	27%
Other	14	27%
Water displacement	9	17%
Manual measure / digital image annotation / 3-D model	9	17%
Methods of validation (comparators for AI)[b]		
Ground truth alone	42	81%
Ground truth and human assessor	8	15%
Ground truth and calorie app	1	2%
Human assessor alone	1	2%
Types of results/tests reported		
Relative error (i.e. accuracy %, error %)	32	62%
Absolute error	18	35%
Correlation coefficient / Coefficient of variation	16	31%
Other (e.g. actual and estimated calories but not error)	12	23%
Bland-Altman	8	15%
t-test	6	12%
Root Mean Square Error of Approximation or Mean Square Error	5	10%
Regression line	4	8%
Food or nutrient data source		
Health organization, hospital, or national/international food database	15	29%
United States Department of Agriculture database	13	25%
Not applicable (e.g. examined volume)	12	23%
Other (e.g. cooking website, corporate website)	12	23%
Not reported	6	12%
Tool used to capture images		
System included at least one digital camera or smartphone	30	59%
Not applicable	15	29%
Other (e.g. depth camera, wearable device)	6	12%
Not reported	1	2%
Source of images		
Newly captured images	25	48%
Pre-existing images (not newly captured)	16	31%
Both newly captured and pre-existing images	10	19%
Not reported	1	2%

(*Continued*)

Table 1. Continued.

Characteristic	N (of studies)	% of 52 studies[a]
Use of a fiducial marker or indicator of a known size		
No	22	42%
Yes	14	27%
Not reported	9	17%
Other	5	10%
Mixed (yes and no, depending on experiment or dataset)	2	4%
Process of AI estimation		
Volume to nutrients/energy	20	38%
Not applicable	9	17%
Other	9	17%
Direct estimate of nutrients from image	5	10%
Portion to nutrients (e.g. 4 fries)	5	10%
Mass to nutrients	3	6%
Weight to nutrients	3	6%
Type of Convolutional Neural Network		
Did not specify type of CNN	17	33%
Not applicable (i.e. did not report using a CNN)	11	21%
Other	10	19%
ResNet or DenseNet	6	12%
Mask R-CNN or other R-CNN	6	12%
Inception or Xception	6	12%
Multi-task CNN	3	6%
VGG	3	6%
Deep CNN	2	4%

AI: artificial intelligence.
[a]Totals sum to greater than 100% if more than one characteristic appeared in one study, such as authors from one paper affiliated with research centers in more than one country.
[b]In two studies, dietitian's assessment were the ground truth definition [11, 76].

Databases for nutrient information and food images

Twenty-five percent (25%) of studies used a USDA database to determine nutrient information, and a health organization, hospital or national/international food database was the source for another twenty-nine percent (29%). Approximately half (48%) of studies captured new food images,; thirty-one percent (31%) used pre-existing images (such as downloaded from the internet), and nineteen percent (19%) used a combination of new and pre-existing images. Twenty-two (22) papers reported using named food image databases, with some papers reporting more than one. Databases included: Angles-13 [43], Food-101 [60,67,73], Plates-18 [43], American Calorie Annotated [23], CALO Mama [70], Calorie-Annotated Food Photo Dataset [22,23], CFNet-34 (ChineseFoodNet+new images) [66], ChinaFood-100 [60], ChinaMarketFood-109 [60,67], FLD-DET [49], FLD-469 [49], Food2k [76], Food-101 [60,67,73], FoodLog [49,72], Fruit 360 [73], FRUITS [69], ImageNet [7,9, 51,67,68,76], ImageNet-1000 [23,46,60,67,68,76], Inselspital [10,41–43], Japanese Calorie-Annotated Food Photo Dataset [23], JISS-22 [49], JISS-DET [49], Korea Food Image database [64], MADiMA [55], fast food database [55], Meals-14 [43], Meals-45 [43], Nasco Life/form Food replica [19], Nutrient Intake Assessment Database [7], Nutrition5K [76], Rakuten18 [51], UECFood-100 [60,67], UECFood-256 [60,67], UNIMIB2016 dataset [46,68,71], VFDL-15 [62], VFDS-15 [62], and Yale-CMU-Berkeley object set [47]. Very few AI systems were tested using the same image datasets: six (6) papers used ImageNet or Image-Net-1000; Inselspital was used in four (4); UNIMIB2016, Food-101, UECFood-100/UECFood-256, Calorie-Annotated Food Photo Dataset, ChinaMarketFood-109 and FoodLog were used in two (2).

Notably, the most frequently reported tools for capturing images were smartphones or digital cameras, with 59% of studies using them either alone or in combination with other tools such as structured light systems, depth sensors, infrared projectors, or wearable devices. Some food images included a fiducial marker (27%) or an indicator of known size (10%) as a reference to aid the AI system in estimating depth for calculating volume in real-world images. This reference scale is established using objects of known dimensions, such as fiducial markers, a photographer's thumb [39,57], a credit card [10,21,37], the serving container [20], the width of the mobile phone [19], a Rubik's Cube [54], or superimposing a square grid on the image [45]. The procedure involves physically placing these items on the dining table or in the image. Many studies reported not using a fiducial marker (42%), some varied across iterations within the study (4%) and some did not report either way (17%) (also see Supplemental eTable 6).

Artificial intelligence methods

Broadly, AI systems aimed to take a digital image as input (e.g. an image of plate with a burger and fries) and provide diet- or nutrition-related information as output (e.g. energy contained in the foods imaged, volume of the burger in the image). They used a variety of methods to achieve this aim. Many had stepwise approach: processing the image, identifying and classifying foods, segmenting the image into specific areas containing specific foods, and linking that information to a nutrition database to determine nutrition facts about the foods contained in the image. Most papers took unique approaches (see Supplemental eTable 6).

In many papers, the AI system estimated nutrient or energy content by first estimating the volume of food contained in the image and then calculating nutrients/energy (38%; see Table 1). A majority of papers (79%) reported using some type of deep learning based Convolutional Neural Network (CNN) (see Table 1 and Supplemental eTable6). Deep-learning methods are those that can take raw data (such as an image, as an array of pixel values) and identify how to represent it at successively higher levels of abstraction [78]. Neural networks 'learn to map a fixed-size input (for example, an image) to a fixed-size output (for example, a probability for each of several categories)' and convolutional neural networks can process information from grid like data, such as image and videos. They have the capability to automatically learn and extract hierarchical features from the data, thus making them more powerful in computer vision [78]. Various deep CNN architectures were used for food detection and classification, such as deep neural network [55,62], feed-forward NN classifier [45], mask R-CNN [47,55], MTCnet with and without CRF (Conditional Random Field) [7], multi-task CNN or contextual network [7,22,51], adaptive neural networks [11], Whale Levenberg Marquardt Neural Network classifier [46], ResNet50 [49], and VGG-16 [22,23].Twelve studies (34%) used a 3-D volume or model reconstruction for food volume estimation [7,10,20,37,40–43,47,50,55,56], in which the AI architecture creates an internal representation of a three dimensional model that can be output and viewed by the human (appearing like a picture or video). Other approaches included rotating a contour around a central axis to construct a ball-shaped model [21], using stereo matching or stereo image [41,43,48,55], random sample consensus (RANSAC) [41,43,52], Speeded Up Robust Features (SURF) [41,43,44], surrounding box [48], structured light systems [50], or virtual reality technology to aid in estimating size compared to a cube [19].

A substantial dataset is typically necessary for conducting AI training with CNNs. As summarized above, some researchers used existing datasets to train and evaluate their AI models, while others created entirely new and extensive datasets. These datasets were designed for the purpose of training and testing various AI methods. To enhance dataset diversity and address the data-intensive nature of AI training, some researchers created smaller datasets and merged them with existing ones. In fact, a majority of the studies (67%) expanded their training datasets for deep learning methods that require substantial data by either developing entirely new extensive datasets or combining new ones with pre-existing ones. As described, very few named databases were used for more than one study. The most frequently used were ImageNet or ImageNet-1000, which was used for six tests (across all studies/papers), and Inselspital which was used for four. Importantly, this lack of database overlap prohibited meta-analysis, which requires close alignment across studies in order to pool data.

Types of results reported

Supplemental eFigure 1 shows the percent of studies reporting the most common accuracy results by diet-related component. The types of analyses conducted and results reported varied widely (e.g. averaging across multiple types of foods vs. within food types only; a small number of images from a limited set of foods such as an apple and banana vs. numerous images of complex mixed dishes, such as curry). Because studies defined ground truth differently, the same effect measure (e.g. relative error) did not consistently represent the same comparison. As shown in Table 1, the most commonly reported diet-related outcomes were volume, calories/energy, carbohydrates, weight/mass and protein. The most frequently reported results for these were relative error for volume (40%) and calories/energy (31%), and absolute error for volume (21%) and calories/energy (25%) (see Supplementary eFigure 1). Relative error was reported for a lower percentage of studies reporting on carbohydrates (12%), weight/mass (6%) or protein (6%). Other reported measures included Bland-Altman analysis; actual and estimated values (e.g. volumes, nutrients); and tests of mean differences.

Relative error of calorie and volume estimations from AI architectures

Forty-two relative errors could be extracted or calculated from published results for calories, volume, or

both in 36 studies (69%). Some studies planned *a priori* to train the AI system until a certain percent accuracy was reached, so extracted relative errors do not necessarily represent the upper limits of a system, just accuracy when results were published. As noted, relative errors could not be directly compared or synthesized because they were based on different types of food images as input into the AI architectures (e.g. a single apple vs. a plate of mashed potatoes with meatloaf). Thus, Figures 2 and 3 display average relative errors (for calories and volume, respectively), ranges of individual relative errors, and standard deviations (where reported) grouped by type of food in images (single vs. multiple foods), but no overall pooled relative error.

For calories, reported average relative errors (e.g. over different types of foods tested within the same study) ranged from 0.10% to 38.3%; the lowest individual relative error (from the range over all papers) was 0.00% and the highest was 79.6%. Six of these 20 relative errors (30%) were from AI estimates using single food images (vs. multiple foods). For volume, reported average relative errors ranged from 0.09% to 33%; individual relative errors had a low of 0.00% and a high of 34.20%; fifteen of these 22 (68%) were from single food images. Average relative errors and ranges tended to be smaller for the single food images compared to multiple food images, suggesting that AI architectures more closely approximated ground truth when images contained simpler food. The range of reported average relative errors for calories and volume were comparable, if slightly lower for volume, which had a higher proportion of single food images. This suggests relatively similar accuracy for calories and volume.

Food image characteristics as potential sources of bias

Table 2 reports descriptive statistics about food image characteristics that we identified as potential sources of bias (see Supplemental eTable 5). Methods of reducing bias included having a trained person take the images (50% of studies); controlling the setting (33%

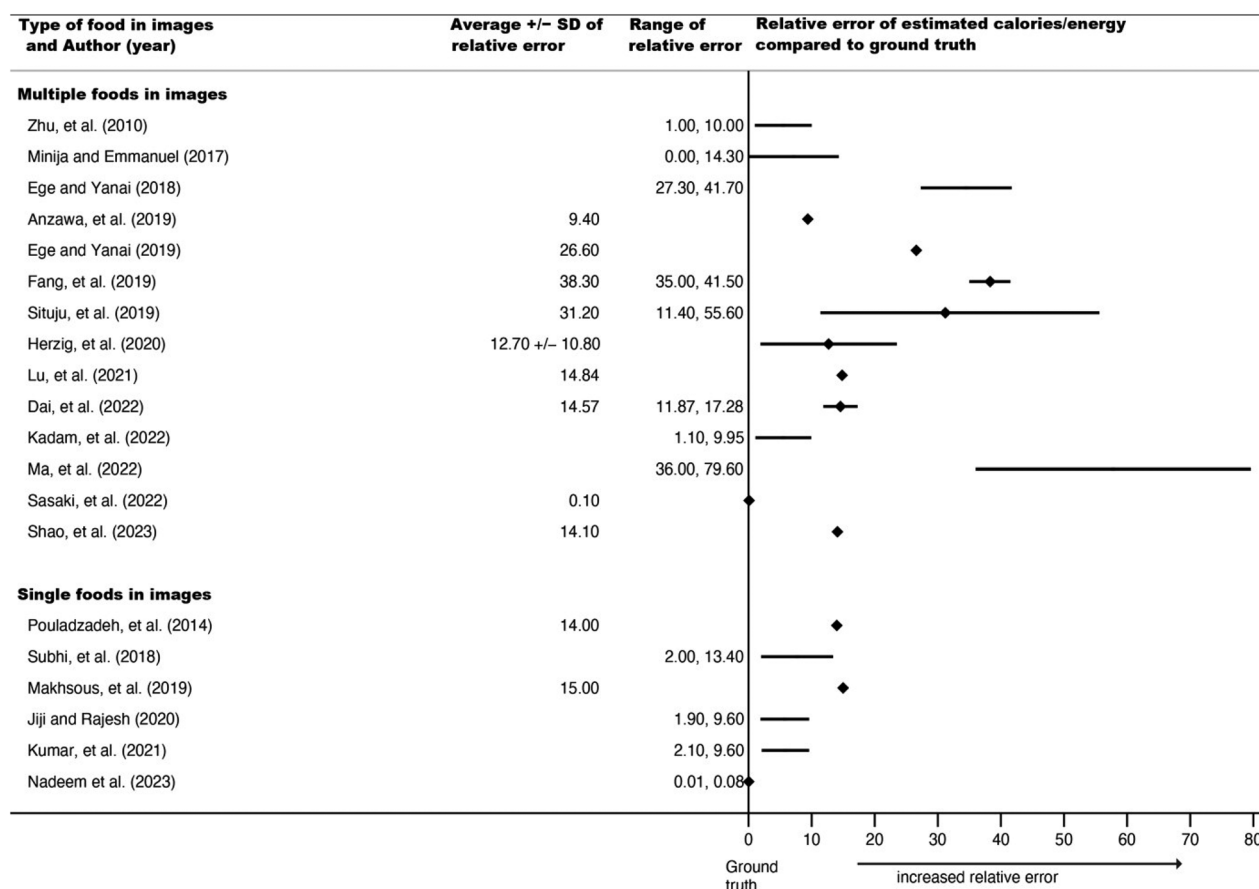**Figure 2.** Forest plot of average relative errors and range of relative errors for AI-estimated calories by year and type of food items in image (*n* = 20 papers). ♦ indicates a reported or calculated average relative error, except for Nadeem et al, 2023, for which the range was too small to appear using a line, so it indicates the range. – indicates the range between the lowest and highest individual relative errors for the best-performing or proposed AI architecture
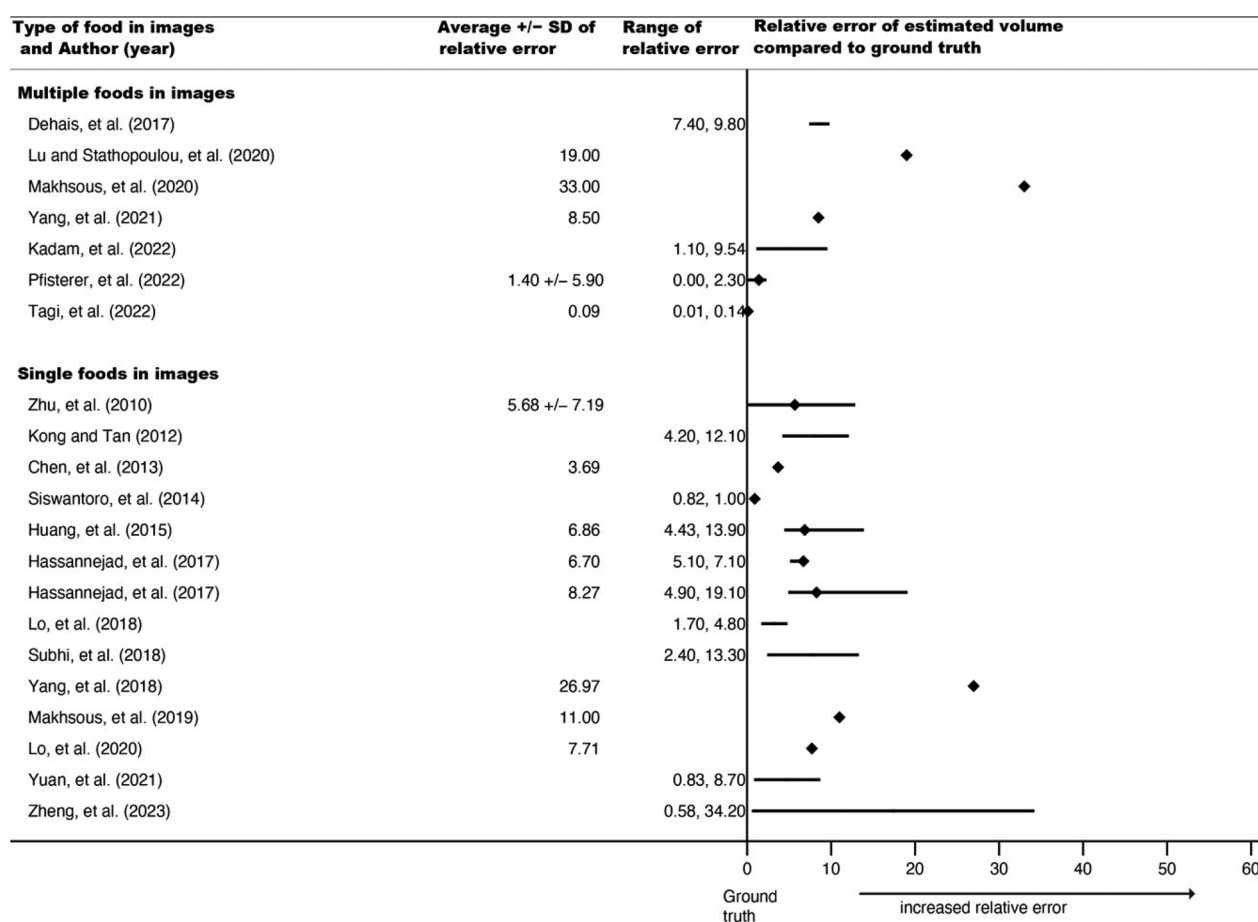
Type of food in images and Author (year)	Average +/− SD of relative error	Range of relative error	Relative error of estimated volume compared to ground truth
Multiple foods in images			
Dehais, et al. (2017)		7.40, 9.80	
Lu and Stathopoulou, et al. (2020)	19.00		
Makhsous, et al. (2020)	33.00		
Yang, et al. (2021)	8.50		
Kadam, et al. (2022)		1.10, 9.54	
Pfisterer, et al. (2022)	1.40 +/− 5.90	0.00, 2.30	
Tagi, et al. (2022)	0.09	0.01, 0.14	
Single foods in images			
Zhu, et al. (2010)	5.68 +/− 7.19		
Kong and Tan (2012)		4.20, 12.10	
Chen, et al. (2013)	3.69		
Siswantoro, et al. (2014)		0.82, 1.00	
Huang, et al. (2015)	6.86	4.43, 13.90	
Hassannejad, et al. (2017)	6.70	5.10, 7.10	
Hassannejad, et al. (2017)	8.27	4.90, 19.10	
Lo, et al. (2018)		1.70, 4.80	
Subhi, et al. (2018)		2.40, 13.30	
Yang, et al. (2018)	26.97		
Makhsous, et al. (2019)	11.00		
Lo, et al. (2020)	7.71		
Yuan, et al. (2021)		0.83, 8.70	
Zheng, et al. (2023)		0.58, 34.20	

(X-axis: 0, 10, 20, 30, 40, 50, 60; 0 = Ground truth; → increased relative error)

Figure 3. Forest plot of average relative errors and range of relative errors for AI-estimated volume by year and type of food items in image ($n = 22$ papers). ♦ indicates a reported or calculated average relative error; except for Siswantoro et al., 2014, for which the range was too small to appear using a line, so it indicates the range. — indicates the range between the lowest and highest individual relative errors for the best-performing or proposed AI architecture

'mostly controlled'), food layout ('somewhat controlled' 37%) or lighting ('somewhat controlled', 31%); and reporting on image features (31% on pixels or resolution; 31% on some aspect of image quality). Types of images were roughly equally divided among standardized foods (38%; e.g. chain restaurant), non-standardized (27%; e.g. home-prepared), and other (35%). Images for most papers contained multiple food items, either solely (35%) or in addition to images of single food items (38%).

Discussion

In recent years, significant interest has emerged in investigating use of AI methods for conducting accurate dietary assessment using digital food images [3]. This review found 35 papers (studies), most published since 2015, investigating AI approaches using a wide variety of tools and techniques. Calories, food volume, and carbohydrate content were the most frequently reported dietary components, and absolute error and relative error were the most common accuracy indicators. AI estimations had a wide range of average relative errors compared to ground truth, with relatively comparable ranges between calories and volume (approximately 0.10% to 38% for calories and 0.09% to 33% for volume). This was not surprising, as calories and other nutrients were often calculated from AI estimated volumes. No papers used bomb-calorimetry, a gold standard ground truth method to measure energy in nutrition science; but weighed plate waste and nutrient information from posted restaurant menus (where some papers got their calorie information) have been found to be strongly correlated with bomb-calorimetry methods [79,80]. Most papers used a convolutional neural network. USDA databases (https://fdc.nal.usda.gov/index.html) were the source of nutrient information in a fourth of studies, but only one food image database was used by more than one research group (ImageNet/ImageNet-1000). Image quality – a potential source of bias – was discussed in about a third of papers. Overall, the breadth of work

Table 2. Summary statistics for food image database characteristics.

Characteristic	N (of studies)	% of 52 studies
Image acquisition		
Active: Taken by researcher/study staff member or trained study participant	26	50%
Passive: Automatically detected by camera	9	17%
Not applicable	9	17%
Active: Not clear who took the image	7	13%
Not reported	1	2%
Image setting[a]		
Mostly controlled (e.g. Real-world with clear table and solid background)	17	33%
Somewhat controlled (e.g. Real-world with some objects on table OR background noise)	11	21%
Highly controlled (e.g. Laboratory with grid-paper)	9	17%
Other / indeterminate or mixed	9	17%
Not at all controlled (e.g. Real-world with some objects on table AND background noise)	6	12%
Image lighting[a]		
Somewhat controlled (e.g. Real-world with consistently good lighting)	16	31%
Highly controlled (e.g. Laboratory with staged lighting)	9	17%
Not at all controlled (e.g. Real-world with inconsistent lighting)	8	15%
Mixed or indeterminate lighting (e.g. Two or more lighting environments included)	7	13%
Mostly controlled (e.g. Real-world with professional lighting)	7	13%
Was image quality mentioned in the paper?		
Reported on some aspect of image quality	16	31%
Reported standard image pixels, or resolution	16	31%
Did not report on image quality	7	13%
Reported using only high-quality images	6	12%
Reported excluding low-quality images from analyses	2	4%
Type of foods in image		
Standardized (e.g. chain restaurant, cafeteria, packaged foods)	20	38%
Other (e.g. mix of standardized & non-standardized)	18	35%
Non-standardized (e.g. home-prepared, non-chain restaurant, buffet)	14	27%
Single vs. multiple food items in the image		
Single and multiple food item images	20	38%
Multiple separate food items (e.g. plate with burger and fries)	18	35%
Single food	13	25%
Unclear	1	2%
Food layout		
Somewhat controlled (e.g. arranged, but with some item overlap/occlusion)	19	37%
Highly controlled (e.g. laid out with clear separation between multiple items)	13	25%
Mostly controlled (e.g. arranged, but items are close together on a plate/tray)	10	19%
Other / indeterminate / mixed	7	13%
Not at all controlled (e.g. not arranged and/or much item overlap/occlusion)	3	6%

[a]If image conditions were not described, investigators from this study interpreted setting and/or lighting from images included in the publication.

is impressive, and AI tools are reaching a high degree of accuracy; however, the published literature contains substantial variability in food image databases used and types of analytical results reported, so findings at this stage of development cannot be synthesized.

AI architecture estimations ranged from 62% to 99% accurate (calculated: 1 - % error) over calories and volume measures when examining results that were averaged across images of multiple foods or food types (Figures 1 and 2). Images of simpler foods (i.e. single item, like an apple) appeared to co-occur with lower relative error rates. The ranges for highest error reported were slightly higher for calories (vs. volume), but a larger proportion of those images were of multiple food images, which may have been harder for AI systems than single food images and not indicative of greater accuracy for volume. Images with only one, simple food (like an apple) may have fewer 'distractions' for computer vision processes, reducing the likelihood that the AI system misclassifies areas in the image. In contrast, images with mixed dishes (such as curries) or plates with a variety of overlapping foods of varying heights or with unclear boundaries present higher risk for classification or segmentation errors.

For many studies, absolute or relative errors were directly reported or could be calculated from reported data. The results for AI appear to be within the range of accuracy levels from human coders of digital images in other studies [25]. As noted, in contrast to expectations, only one paper directly compared accuracy of AI methods to humans alone; and only 8 papers used human assessors at all, indicating an area open for research. One recent review of digital image dietary assessment methods reported that percent error of human methods ranged between 30% underestimation to 1% overestimation, indicating accuracy of 70% to 99% across various outcomes and ground truth methods (e.g. energy intake, servings, weighed food, doubly labeled water) [25]. Studies using human coders tend to report correlations between ground truth and human estimations for weight (g) and

calories (kcal); and they tend to use Bland-Altman plots to display associations across varying levels of leftover food. While all papers included in this review reported some type of analytic comparison between AI and ground truth, only some reported correlations. For results from AI architectures to be more directly comparable to results from human coders, AI researchers could report the correlation; and they could examine the AI tool's ability to identify nutrients consumed by estimating both full portions and leftover unconsumed food, back-calculating actual consumption; further, they could display results with Bland-Altman plots.

As noted, the relative errors reported in the included papers reflected a variety of AI methods and comparisons. Researchers had varying aims or needs, such as optimizing for computing space or processing time vs. optimizing for the lowest relative error. Extracted results represent the peer-reviewed literature but not necessarily the upper limits of accuracy. But to contextualize the current findings, a 99% accurate AI-based calorie management tool would lead to under- or over-estimation of calorie intake by only about 20 calories per day (based on a 2,000-calorie diet). If consistent over a year, those errors would equate to 7,300 calories per year or about 2 pounds. A tool that was 62% accurate would be off by 760 calories per day, or 277,400 calories (79 pounds) per year. These numbers drastically oversimplify the complex processes of appetite and food intake self-regulation, and omit important components of energy balance (such as physical activity), but they convey a sense of the scope of the impact of the current ranges of relative error. As AI-based dietary assessment tools develop, it will be important to determine what features of the food images, AI architecture, and end-users lead to over- and under-estimation of nutrient intake so that user-facing tools can be adjusted for accurate self-management of dietary goals or nutrition research. There are likely trade-offs between end-user burden (e.g. not requiring a fiducial marker, working in inconsistent lighting with overlapping food) and accuracy. It is important to note that this review included only published literature; the accuracy of any applications or camera technology currently on the market [30,31] were not evaluated (unless validity tests had been published).

Convolutional Neural Networks were the most common method of AI estimation, which aligns with recent research finding that CNN-based models are most frequently used [27]. We also found that most studies are testing AI systems on different food image datasets, highlighting the need for a large-scale benchmark food image dataset [24]. The current

review specifies characteristics that would be important for such a dataset and recommends analytic techniques and reporting guidelines that would allow comparisons across nutrition and engineering fields, and meta-analytic synthesis. Recent reviews envision AI as being on the precipice of tremendous leaps in accuracy, timeliness, and clinically meaningful nutrition outcomes [1,2, 4]. But they urge the nascent science to communicate in broader terms to connect across audiences, catch up to the AI advances made in medicine, and identify those areas where AI can accelerate progress vs. those that will always rely on traditional (human) judgement [1,2, 4]. As AI tools become sufficiently developed, coordination will need to occur with clinical fields as well (e.g. physicians, dietitians) for management of nutrition-related chronic disease.

The heterogeneity we found in results reported precluded meta-analytic synthesis for two reasons. First, for a valid comparison between two AI architectures both would need to have the same sets of databases input into the system. Yet, few databases were used in more than one study. This emerging field would benefit from agreement on the characteristics needed for a valid food image database for training and testing and from selecting a limited number of 'gold standard' databases. Potential valuable features might include a wide variety of food types (e.g. separate foods like individual fruits, mixed foods like curries, cultural variety); high image quality (including pixels, resolution); lighting condition; a variety of settings (e.g. chain restaurant, home table, mall food court); a variety of camera angles; and potentially the inclusion of a consistent fiducial marker or size referent (e.g. photographer's thumb). One recent review suggested that MADiMA [55] and Nutrition5K [81] seemed the most comprehensive datasets currently available for this type of use but noted that they did not have 'eastern style' foods [24]. (Authors described 'western style' foods tending to be served separately, while 'eastern' foods were mixed; the study was conducted in China). Researchers should also consider making their own databases publicly available so other researchers can compare AI performance.

Second, even if the inputs to the AI systems identified in this review had been the same, the outputs differed. The types of analyses conducted to determine system 'accuracy' varied across papers in at least two ways. First, the statistical methods differed. Absolute error was the most frequently reported numeric result, but over 10 types of indicators were reported. Second, the definition of ground truth varied. So, the absolute

error in one paper may have represented a comparison between AI and a dietitian coding the digital image, another may have compared AI to plate waste weighed on a household food scale, and a third may have compared it to 'one serving' from a nutrition database. An argument could be made that pooling these absolute values would be conceptually similar to pooling effect sizes across different populations (e.g. adolescents, seniors), and heterogeneity could be calculated and examined. In this case, that could not be done because of different system databases; but beyond that, there is no field-wide 'gold standard' for the type of analysis or definition of ground truth. Moving forward, researchers could strive to use a limited set of high-quality food image databases, test AI architectures on food images containing both single and multiple foods, and report overall relative errors (and standard deviations) averaged over many different food types.

This review also revealed a gap in the literature regarding validated tools for assessing risk of bias in these types of studies. As described above, bias could come from inputs into the system: a food image database with low-quality images, a limited set of 'easy' images, or a definition of ground truth that was not accurate. We addressed this challenge by evaluating studies on features that may have biased the food classification, segmentation, or dietary estimation process. We selected features such as image setting, content, lighting, and resolution because poor quality would create missing information potentially leading to error in the AI estimation. Results showed variety in the types of environments represented in images (e.g. chain restaurant, home tables) and some degree of control over the image setting and lighting. The majority (65%) reported something related to image quality (e.g. pixels, resolution, blurriness), though definitions of quality varied. We offer this list of characteristics to future researchers as a framework for reporting results in published literature.

In addition to bias resulting from food image database input, recent research raises the possibility that the processing approach itself has the potential to create or reveal other forms of bias, such as a recent finding that an AI system could identify patients' races from x-ray data [82]. That particular finding has not been replicated, and the way the identification happens is unclear (e.g. geographical history, previously undetectable melanin differences). But the development of these types of technology warrants awareness of ethical, legal, and medical considerations, particularly when the AI tools would be used to manage nutrition-related conditions affecting lifespan or life quality, such as diabetes, obesity, or heart disease

[6,83]. In particular, attention will need to be paid to: (a) how AI architectures are trained so they do not learn or magnify unconscious bias from human coders [84,85] and can separate user food choices from historical, social, cultural, or political confounders that could exacerbate health inequities; (b) expanding the universe of populations who engage in user testing and counteracting a reliance on subgroups (such as white females for self-monitoring and weight loss12]; (c) being able to explain the processes through which results are obtained [86] while being mindful of potential sources of bias for supervised vs. unsupervised learning; and (d) whether federal approval processes are needed for products released on the market to ensure health claims meet a threshold for demonstrating efficacy. These are only cautions. Overall, there is tremendous potential with AI-based dietary assessment but a significant need to develop and validate tools for assessing bias and comparing study quality.

Strengths of this systematic review included the process (i.e. strict inclusion/exclusion criteria, multiple independent coders) and synthesis of literature addressing an important and timely question. In the Engineering field, scientific results may be reported in preprint form or outside of peer-reviewed journal articles (e.g. conference proceedings). The current systematic review was limited to the published literature to match conventions in nutrition and health sciences. While some preprint reports go through a peer review process, the streamlined descriptions required of abstract submission prohibit detailed explanations of study methods or results for review. We acknowledge that the technology industry moves rapidly, particularly in a market space where commercial products are being created and released for sale, such as cell phones equipped with AI dietary assessment methods [30,31]. But for management of chronic conditions, the speed of innovation should be tempered by the use of testing and validation procedures that protect end user safety. Because this review was limited to peer-reviewed published articles, meaningful results may have been omitted. One recent review [27] relied on an almost entirely different set of papers due to different exclusion criteria. Publication bias is also more difficult to evaluate when substantial literature is in the preprint form. In this review, the procedure for that type of literature was to search for authors' publications subsequent to the abstract. Articles identified were included in the review. If no follow-up was published, it remains unknown whether the results were later found to be null or invalidated in some way or whether the conference proceedings were considered the end result with no additional need for publishing.

Conclusions

Significant interest has emerged in recent years in investigating the ability of AI methods to conduct accurate dietary assessment using digital food images. Most studies included in this review reported that their AI system was accurate according to their definition of ground truth. This systematic review found a wide range of food image databases, nutrient information, and types of results reported in the literature. Future researchers should consider using a limited number of databases that the field has determined to be adequate for training and testing. The field should come to consensus on what characteristics are required for a high-quality database (e.g. range of types of foods, number of items in images, lighting conditions and food layouts; linked to nutrient databases known to be accurate). Researchers should also consider reporting at least absolute error and average relative error for AI-based volume and calorie estimations. A validated risk of bias tool needs to be developed so study quality and results can be compared across these types of studies. As the field grows, we recommend focusing on consistent definitions of accuracy and ground truth and conducting cross-field collaboration with end users to identify best practices for using the technology.

Authors contributions

EH, ES, KCC, and XP contributed equally to the design of the research; KCC and XP contributed to the acquisition of the data; MC, KCC and ES analyzed and interpreted the data; ES and KCC drafted the manuscript; SK and KP contributed to study conceptualization and writing. All authors critically revised the manuscript, and all authors read and approved the final manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

The data that support the findings of this study are available from the corresponding author, ES, upon reasonable request.

References

[1] Limketkai BN, Mauldin K, Manitius N, et al. The age of artificial intelligence: use of digital technology in clinical nutrition. Curr Surg Rep. 2021;9(7):1. doi:10.1007/s40137-021-00297-3.

[2] Côté M, Lamarche B. Artificial intelligence in nutrition research: perspectives on current and future applications. Appl Physiol Nutr Metab. 2021;47(1):1–8. doi:10.1139/apnm-2021-0448.

[3] Zhao X, Xu X, Li X, et al. Emerging trends of technology-based dietary assessment: a perspective study. Eur J Clin Nutr. 2021;75(4):582–17. doi:10.1038/s41430-020-00779-0.

[4] Oliveira Chaves L, Gomes Domingos AL, Louzada Fernandes D, et al. Applicability of machine learning techniques in food intake assessment: a systematic review. Crit Rev Food Sci Nutr. 2023;63(7):902–919. doi:10.1080/10408398.2021.1956425.

[5] Gemming L, Utter J, Mhurchu CN. Image-assisted dietary assessment: a systematic review of the evidence. J Acad Nutr Diet. 2015;115(1):64–77. doi:10.1016/j.jand.2014.09.015.

[6] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44–56. doi:10.1038/s41591-018-0300-7.

[7] Lu Y, Stathopoulou T, Vasiloglou MF, et al. An artificial intelligence-based system to assess nutrient intake for hospitalised patients. IEEE Trans Multimedia. 2021;23:1136–1147. doi:10.1109/TMM.2020.2993948.

[8] Shermila PJ, Milton A. Estimation of protein from the images of health drink powders. J Food Sci Technol. 2020;57(5):1887–1895. doi:10.1007/s13197-019-04224-4.

[9] 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(4):877. doi:10.3390/nu11040877.

[10] Vasiloglou MF, Mougiakakou S, Aubry E, et al. A comparative study on carbohydrate estimation: goCARB vs. Dietitians. Nutrients. 2018;10(6):741. doi:10.3390/nu10060741.

[11] Todd LE, Wells NM, Wilkins JL, et al. Digital food image analysis as a measure of children's fruit and vegetable consumption in the elementary school cafeteria: a description and critique. J Hunger Environ Nutr. 2017;12(4):516–528. doi:10.1080/19320248.2016.1275996.

[12] Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc. 2011;111(1):92–102. doi:10.1016/j.jada.2010.10.008.

[13] Subar AF, Freedman LS, Tooze JA, et al. Addressing current criticism regarding the value of self-report dietary data. J Nutr. 2015;145(12):2639–2645. doi:10.3945/jn.115.219634.

[14] Lissner L, Troiano RP, Midthune D, et al. OPEN about obesity: recovery biomarkers, dietary reporting errors and BMI. Int J Obes. 2007;31(6):956–961. doi:10.1038/sj.ijo.0803527.

[15] Wehling H, Lusher J. People with a body mass index ≥ 30 under-report their dietary intake: a systematic review. J Health Psychol. 2019;24(14):2042–2059. doi:10.1177/1359105317714318.

[16] Boushey CJ, Spoden M, Zhu FM, et al. New mobile methods for dietary assessment: review of image-assisted and image-based dietary assessment methods. Proc Nutr Soc. 2017;76(3):283–294. doi:10.1017/S0029665116002913.

[17] Lieffers JRL, Hanning RM. Dietary assessment and self-monitoring: with nutrition applications for mobile devices. Can J Diet Pract Res. 2012;73(3):e253–e260. doi:10.3148/73.3.2012.e253.

[18] Wang J, Sereika SM, Chasens ER, et al. Effect of adherence to self-monitoring of diet and physical activity on weight loss in a technology-supported behavioral intervention. Patient Prefer Adherence. 2012;6:221–226. doi:10.2147/PPA.S28889.

[19] Yang Y, Jia W, Bucher T, et al. Image-based food portion size estimation using a smartphone without a fiducial marker. Public Health Nutr. 2019;22(7):1180–1192.

[20] Chen H-C, Jia W, Yue Y, et al. Model-based measurement of food portion size for image-based dietary assessment using 3D/2D registration. Meas Sci Technol. 2013;24(10):105701. doi:10.1088/0957-0233/24/10/105701.

[21] Huang J, Ding H, McBride S, et al. Use of smartphones to estimate carbohydrates in foods for diabetes management. Stud Health Technol Inform. 2015;214:121–127.

[22] Ege T, Yanai K. Simultaneous estimation of dish locations and calories with multi-task learning. IEICE Trans Inf & Syst. 2019;E102.D(7):1240–1246. doi:10.1587/transinf.2018CEP0004.

[23] Ege T, Yanai K. Image-based food calorie estimation using recipe information. IEICE Trans Inf & Syst. 2018;E101.D(5):1333–1341. doi:10.1587/transinf.2017MVP0027.

[24] Wang W, Min W, Li T, et al. A review on vision-based analysis for automatic dietary assessment. Trends Food Sci Technol. 2022;122:223–237. doi:10.1016/j.tifs.2022.02.017.

[25] Höchsmann C, Martin CK. Review of the validity and feasibility of image-assisted methods for dietary assessment. Int J Obes. 2020;44(12):2358–2371. doi:10.1038/s41366-020-00693-2.

[26] Doulah A, McCrory MA, Higgins JA, et al. A systematic review of technology-driven methodologies for estimation of energy intake. IEEE Access. 2019;7:49653–49668. doi:10.1109/access.2019.2910308.

[27] Dalakleidi KV, Papadelli M, Kapolos I, et al. Applying image-based food recognition systems on dietary assessment: a systematic review. Adv Nutr. 2022;13(6):2590–2619. doi:10.1093/advances/nmac078.

[28] Tahir GA, Loo CK. A comprehensive survey of image-based food recognition and volume estimation methods for dietary assessment. Healthcare. 2021 2021;9(12):1676. doi:10.3390/healthcare9121676.

[29] Kaur R, Kumar R, Gupta M. Deep neural network for food image classification and nutrient identification: a systematic review. Rev Endocr Metab Disord. 2023;24(4):633–653. doi:10.1007/s11154-023-09795-4.

[30] Huawei Mate 20. 20 Pro unveiled: more AI, better cameras, wireless charging. Egypt Today. 2022.

[31] Farooqui A. Samsung's Bixby Assistant Can Count Food Calories. Ubergizmo. 2022. https://www.ubergizmo.com/2018/01/samsungs-bixby-assistant-can-count-food-calories/

[32] Finding what works in health care: standards for systematic reviews. 2011.

[33] Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated 2023 Aug). Cochrane; 2023. Available from: www.training.cochrane.org/handbook.

[34] Moher D, Liberati A, Tetzlaff J, PRISMA Group, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6(7):e1000097. doi:10.1371/journal.pmed.1000097.

[35] Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev. 2021;10(1):89. doi:10.1186/s13643-021-01626-4.

[36] Zhu F, Bosch M, Woo I, et al. The use of mobile devices in aiding dietary assessment and evaluation. IEEE J Sel Top Signal Process. 2010;4(4):756–766. doi:10.1109/JSTSP.2010.2051471.

[37] Kong F, Tan J. DietCam: automatic dietary assessment with mobile camera phones. Pervasive Mob Comput. 2012;8(1):147–163. doi:10.1016/j.pmcj.2011.07.003.

[38] Lee CD, Chae J, Schap TE, et al. Comparison of known food weights with image-based portion-size automated estimation and adolescents' self-reported portion size. J Diabetes Sci Technol. 2012;6(2):428–434. doi:10.1177/193229681200600231.

[39] Pouladzadeh P, Shirmohammadi S, Al-Maghrabi R. Measuring calorie and nutrition from food image. IEEE Trans Instrum Meas. 2014;63(8):1947–1956. doi:10.1109/TIM.2014.2303533.

[40] Siswantoro J, Prabuwono AS, Abdullah A, et al. Monte carlo method with heuristic adjustment for irregularly shaped food product volume measurement. Sci World J. 2014;2014:1–10. doi:10.1155/2014/683048.

[41] Anthimopoulos M, Dehais J, Shevchik S, et al. Computer vision-based carbohydrate estimation for type 1 patients with diabetes using smartphones. J Diabetes Sci Technol. 2015;9(3):507–515. doi:10.1177/1932296815580159.

[42] Rhyner D, Loher H, Dehais J, et al. Carbohydrate estimation by a mobile phone-based system versus self-estimations of individuals with type 1 diabetes mellitus: a comparative study. J Med Internet Res. 2016;18(5):e5567. doi:10.2196/jmir.5567.

[43] Dehais J, Anthimopoulos M, Shevchik S, et al. Two-view 3D reconstruction for food volume estimation. IEEE Trans Multimedia. 2017;19(5):1090–1099. doi:10.1109/TMM.2016.2642792.

[44] Hassannejad H, Matrella G, Ciampolini P, et al. A new approach to image-based estimation of food volume. Algorithms. 2017;10(2):66. doi:10.3390/a10020066.

[45] Minija SJ, Emmanuel WRS. Neural network classifier and multiple hypothesis image segmentation for dietary assessment using calorie calculator. Imag Sci J.

2017;65(7):379–392. doi:10.1080/13682199.2017.1356 610.

[46] Emmanuel WS, Minija SJ. Fuzzy clustering and whale-based neural network to food recognition and calorie estimation for daily dietary assessment. Sādhanā. 2018;43(5):1–19. doi:10.1007/s12046-018-0865-3.

[47] Lo FPW, Sun Y, Qiu J, et al. Food volume estimation based on deep learning view synthesis from a single depth map. Nutrients. 2018;10(12):2005. doi:10.3390/nu10122005.

[48] Subhi MA, Ali SHM, Ismail AG, et al. Food volume estimation based on stereo image analysis. IEEE Instrum Meas Mag. 2018;21(6):36–43. doi:10.1109/MIM.2018.8573592.

[49] Anzawa M, Amano S, Yamakata Y, et al. Recognition of multiple food items in a single photo for use in a buffet-style restaurant. IEICE Trans Inf Syst. 2019;E102 .D(2):410–414. doi:10.1587/transinf.2018EDL8183.

[50] Makhsous S, Mohammad HM, Schenk JM, et al. A novel mobile structured light system in food 3D reconstruction and volume estimation. Sensors. 2019;19(3):564. doi:10.3390/s19030564.

[51] Situju SF, Takimoto H, Sato S, et al. Food constituent estimation for lifestyle disease prevention by multi-task CNN. Appl Artif Intel. 2019;33(8):732–746. doi:10.1080/08839514.2019.1602318.

[52] Herzig D, Nakas CT, Stalder J, et al. Volumetric food quantification using computer vision on a depth-sensing smartphone: preclinical study. JMIR Mhealth Uhealth. 2020;8(3):e15294. doi:10.2196/15294.

[53] Jiji GW, Rajesh A. Food sustenance estimation using food image. Int J Image Grap. 2020;20(04):2050034. doi:10.1142/S0219467820500345.

[54] Lo FPW, Sun Y, Qiu J, et al. Point2Volume: a vision-based dietary assessment approach using view synthesis. IEEE Trans Ind Inf. 2020;16(1):577–586. doi:10.1109/TII.2019.2942831.

[55] Lu Y, Stathopoulou T, Vasiloglou MF, et al. goFOODTM: an artificial intelligence system for dietary assessment. Sensors. 2020;20(15):4283. doi:10.3390/s20154283.

[56] Makhsous S, Bharadwaj M, Atkinson BE, et al. DietSensor: automatic dietary intake measurement using mobile 3d scanning sensor for diabetic patients. Sensors. 2020;20(12):3380. doi:10.3390/s20123380.

[57] Pillai SM, Mohideen K. Food calorie measurement and classification of food images. Inter J Pharmaceut Res. 2020; doi:10.31838/ijpr/2020.12.04.262.

[58] Chotwanvirat P, Hnoohom N, Rojroongwasinkul N, et al. Feasibility study of an automated carbohydrate estimation system using thai food images in comparison with estimation by dietitians. Front Nutr. 2021;8:732449. doi:10.3389/fnut.2021.732449.

[59] Kumar RD, Julie EG, Robinson YH, et al. Recognition of food type and calorie estimation using neural network. J Supercomput. 2021;77(8):8172–8193. doi:10.1007/s11227-021-03622-w.

[60] Ma P, Lau CP, Yu N, et al. Image-based nutrient estimation for chinese dishes using deep learning. Food Res Int. 2021;147:110437. doi:10.1016/j.foodres.2021.110437.

[61] Papathanail I, Brühlmann J, Vasiloglou MF, et al. Evaluation of a novel artificial intelligence system to monitor and assess energy and macronutrient intake in hospitalised older patients. Nutrients. 2021;13(12):4539. doi:10.3390/nu13124539.

[62] Yang Z, Yu H, Cao S, et al. Human-Mimetic estimation of food volume from a single-View RGB image using an AI system. Electronics. 2021;10(13):1556. doi:10.3390/electronics10131556.

[63] Yuan D, Hu X, Zhang H, et al. An automatic electronic instrument for accurate measurements of food volume and density. Public Health Nutr. 2021;24(6):1248–1255. doi:10.1017/S136898002000275X.

[64] Dai Y, Park S, Lee K. Utilizing mask R-CNN for Solid-Volume food instance segmentation and calorie estimation. Appl Sci. 2022;12(21):10938. doi:10.3390/app122110938.

[65] Kadam P, Pandya S, Phansalkar S, et al. FVEstimator: a novel food volume estimator wellness model for calorie measurement and healthy living. Measurement. 2022;198:111294. doi:10.1016/j.measurement.2022.111294.

[66] Li H, Yang G. Dietary nutritional information autonomous perception method based on machine vision in smart homes. Entropy. 2022;24(7):868. doi:10.3390/e24070868.

[67] Ma P, Lau CP, Yu N, et al. Application of deep learning for image-based chinese market food nutrients estimation. Food Chem. 2022;373(Pt B):130994. doi:10.1016/j.foodchem.2021.130994.

[68] Pfisterer KJ, Amelard R, Chung AG, et al. Automated food intake tracking requires depth-refined semantic segmentation to rectify visual-volume discordance in long-term care homes. Sci Rep. 2022;12(1):83. doi:10.1038/s41598-021-03972-8.

[69] Prakash ASJ, Sriramya P. Accuracy analysis for image classification and identification of nutritional values using convolutional neural networks in comparison with logistic regression model. J Pharmac Negative Results. 2022;

[70] Sasaki Y, Sato K, Kobayashi S, et al. Nutrient and food group prediction as orchestrated by an automated image recognition system in a smartphone app (CALO mama): validation study. JMIR Form Res. 2022;6(1):e31875. doi:10.2196/31875.

[71] Minija SJ, Sam Emmanuel WR. Imperialist competitive algorithm-based deep belief network for food recognition and calorie estimation. Evol Intel. 2022;15(2):955–970. doi:10.1007/s12065-019-00265-y.

[72] Tagi M, Tajiri M, Hamada Y, et al. Accuracy of an artificial intelligence–based model for estimating leftover liquid food in hospitals: validation study. JMIR Form Res. 2022;6(5):e35991. doi:10.2196/35991.

[73] Haque RU, Khan RH, Shihavuddin ASM, et al. Lightweight and Parameter-Optimized Real-Time food calorie estimation from images using CNN-Based approach. Appl Sci. 2022;12(19):9733. doi:10.3390/app12199733.

[74] Zhang Q, He C, Qin W, et al. Eliminate the hardware: mobile terminals-oriented food recognition and weight estimation system. Front Nutr. 2022;9:965801. doi:10.3389/fnut.2022.965801.

[75] Nadeem M, Shen H, Choy L, et al. Smart diet diary: real-Time mobile application for food recognition. ASI. 2023;6(2):53. doi:10.3390/asi6020053.

[76] Shao W, Min W, Hou S, et al. Vision-based food nutrition estimation via RGB-D fusion network. Food Chem. 2023;424:136309. doi:10.1016/j.foodchem.2023.136309.

[77] Zheng X, Liu C, Gong Y, et al. Food volume estimation by multi-layer superpixel. Math Biosci Eng. 2023;20(4):6294–6311. doi:10.3934/mbe.2023271.

[78] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–444. doi:10.1038/nature14539.

[79] Gervis JE, Hennessy E, Shonkoff ET, et al. Weighed plate waste can accurately measure children's energy consumption from food in Quick-Service restaurants. J Nutr. 2019;150(2):404–410. doi:10.1093/jn/nxz222.

[80] Urban LE, McCrory MA, Dallal GE, et al. Accuracy of stated energy contents of restaurant foods. Jama. 2011;306(3):287–293. doi:10.1001/jama.2011.993.

[81] Thames Q, Karpur A, Norris W, et al. Nutrition5k: towards automatic nutritional understanding of generic food.

[82] Bray H. MIT, Harvard scientists find AI can recognize race from X-rays—and nobody knows how. Boston Globe. 2022. https://www.bostonglobe.com/2022/05/13/business/mit-harvard-scientists-find-ai-can-recognize-race-x-rays-nobody-knows-how/#:~:text=The%20study%20found%20that%20an, race%20with%2090%20percent%20accuracy

[83] Yu K-H, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng. 2018;2(10):719–731. doi:10.1038/s41551-018-0305-z.

[84] FitzGerald C, Hurst S. Implicit bias in healthcare professionals: a systematic review. BMC Med Ethics. 2017;18(1):19. doi:10.1186/s12910-017-0179-8.

[85] Marcelin JR, Siraj DS, Victor R, et al. The impact of unconscious bias in healthcare: how to recognize and mitigate it. J Infect Dis. 2019;220(220 Suppl 2):S62–S73. doi:10.1093/infdis/jiz214.

[86] Amann J, Blasimme A, Vayena E, et al. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak. 2020;20(1):310. doi:10.1186/s12911-020-01332-6.