Public Health Nutrition: 23(15), 2700–2710 doi:10.1017/S1368980020000816

Validating an automated image identification process of a passive image-assisted dietary assessment method: proof of concept

Tsz-Kiu Chui[1], Jindong Tan[2], Yan Li[2] and Hollie A. Raynor[1,]*
[1]Department of Nutrition, University of Tennessee, Knoxville, TN 37996, USA: [2]Department fof Mechanical, Aerospace and Biomedical Engineering, University of Tennessee, Knoxville, TN, USA

Submitted 5 May 2019: Final revision received 31 January 2020: Accepted 5 March 2020: First published online 10 June 2020

Abstract

Objective: To validate an automated food image identification system, DietCam, which has not been validated, in identifying foods with different shapes and complexities from passively taken digital images.
Design: Participants wore Sony SmartEyeglass that automatically took three images per second, while two meals containing four foods, representing regular- (i.e., cookies) and irregular-shaped (i.e., chips) foods and single (i.e., grapes) and complex (i.e., chicken and rice) foods, were consumed. Non-blurry images from the meals' first 5 min were coded by human raters and compared with DietCam results. Comparisons produced four outcomes: true positive (rater/DietCam reports yes for food), false positive (rater reports no food; DietCam reports food), true negative (rater/DietCam reports no food) or false negative (rater reports food; DietCam reports no food).
Setting: Laboratory meal.
Participants: Thirty men and women (25·1 ± 6·6 years, 22·7 ± 1·6 kg/m^2, 46·7 % White).
Results: Identification accuracy was 81·2 and 79·7 % in meals A and B, respectively (food and non-food images) and 78·7 and 77·5 % in meals A and B, respectively (food images only). For food images only, no effect of food shape or complexity was found. When different types of images, such as 100 % food in the image and on the plate, <100 % food in the image and on the plate and food not on the plate, were analysed separately, images with food on the plate had a slightly higher accuracy.
Conclusions: DietCam shows promise in automated food image identification, and DietCam is most accurate when images show food on the plate.

Keywords
Image-assisted dietary assessment
Automated food identification
Digital food images
Food shape
Food complexity

Accurate dietary assessment is essential to understand how diet impacts health[1]. Commonly used self-reported dietary assessment methods (i.e., 24-h dietary recall, food record, FFQ) are prone to errors due to limited accuracy in capturing all items and portion sizes consumed[2,3]. These methods are labour intensive in data collection and/or analysis[4,5]. Self-reported dietary data appear to have systematic bias, in which populations with obesity are more likely to underreport intake[6–13].

The incorporation of technology via images, using active and passive methods, into dietary assessment is one way to improve dietary assessment accuracy[14]. Active image-assisted dietary assessment methods require individuals to manually capture images or videos with digital cameras, smartphones and other picture-capturing devices[14]. Studies on active methods show that these methods provide comparable accuracy of dietary information when compared with objective dietary assessment methods[15–17]. However, these active methods are not fully automated in identifying food or estimating portion sizes, still creating burden for participants and providers/staff in documenting information accurately (i.e., participants may need to provide additional information other than images to assist with accuracy, and images may need to be viewed by providers/staff to identify items and portions consumed). Most importantly, active methods still rely on humans to manually capture images; thus, if images are not taken or images are of poor quality, accuracy is diminished.

*Corresponding author: Email hraynor@utk.edu

Passive image-assisted dietary assessment methods, in which images or videos automatically capture dietary intake through the use of wearable devices/tools, are the next generation of image-assisted dietary assessment methods developed. It is believed that this method can reduce human errors as the process of collecting dietary information requires less effort and training than the active methods for both participants and providers/staff[14]. Research in this area is still in its infancy and has developed as technology has advanced, but results of reviewed passive methods showed promise in accuracy of assessing dietary information, with some methods automating portion size estimation[18–21].

To fully automate image-assisted dietary assessment methods, valid image identification methods need to be developed. Jia *et al.*[22] conducted the only validation study of an automatic food detection method using images collected from a wearable device (eButton), in which images were classified as containing food or non-food only. Investigators found the accuracy rates of 91·5 and 86·4 % when averaged between true positive (TP; correctly identifying food images) and true negative (TN; correctly identifying non-food images)[22]. Thus, currently none of the passive-assisted dietary assessment methods possess automatic food identification that can identify specific food items. One additional challenge of images collected passively with wearable devices is that because images are captured when the participant is moving and eating, the items being consumed are not fully captured (partial images) or fully clear (may be partially blocked by hands, hair, etc.) and may not be on a flat surface (i.e., being held in a hand or on a utencil), making identification more challenging. Therefore, the purpose of this proof-of-concept investigation was to validate an automated food image identification system, DietCam, using images pasively taken by Sony SmartEyeglass (wearable device) to identify food items with different shapes and complexities. Furthermore, the type of food images (i.e., full or partial) was examined for accuracy of identification. The specific aims were (i) to describe the agreement between human raters and DietCam in identifying foods and no food in images taken by Sony SmartEyeglass, (ii) to describe the agreement between human raters and DietCam in identifying foods in different shapes (regular *v.* irregular) and complexities (single food *v.* mixed food) and (iii) to describe the agreement between human raters and DietCam for types of coded food images (i.e., when 100 % of the food available is in the image and on the plate, when <100 % of the food available is in the image and on the plate and when the food is not on the plate (on an eating utensil or in a hand)) in identifying foods in different shapes and complexities.

Method

Study design

A mixed factorial design was used, with between-subject factor of meal orders (1 or 2) and within-subject factors of meals (meals A and B), food shapes (regular and irregular) and food complexities (single food and mixed food) (see Table 1). Participants were randomised into one of the two meal orders. In each meal, participants were given a meal that included a regular-shaped single food (i.e., cookie or grapes), an irregular-shaped single food (i.e., chips or ice cream), a regular-shaped mixed food (i.e., sandwich or wrap) and an irregular-shaped mixed food (i.e., pasta dish or chicken and rice dish). Dependent variables were the identification of foods (percentage of TP, false positive (FP), false negative (FN) and TN).

Participants

Thirty participants were recruited through flyers posted around the University of Tennessee Knoxville campus (see Fig. 1 for participant flow). Interested individuals were phone screened for eligibility. Eligibile participants were aged 18 and 65 years, had a BMI between 18·5 and 24·9 kg/m^2, had no food allergies/intolerance to study foods, had no dietary plan/restriction that prevented

Table 1 Description of study design

Meal order	Meal session 1	Meal session 2
1 (*n* 15)	Meal A	Meal B
	Turkey and Provolone Cheese Sandwich (regular-shaped mixed food)	Ham and Cheddar Cheese Wrap (regular-shaped mixed food)
	Chicken and Wild Rice (irregular-shaped mixed food)	Pasta with Broccoli in Alfredo Sauce (irregular-shaped mixed food)
	Chocolate Chip Cookie (regular-shaped single food)	Red Seedless Grapes (regular-shaped single food)
	Potato Chips Original (irregular-shaped single food)	Chocolate Ice Cream (irregular-shaped single food)
2 (*n* 15)	Meal B	Meal A
	Ham and Cheddar Cheese Wrap (regular-shaped mixed food)	Turkey and Provolone Cheese Sandwich (regular-shaped mixed food)
	Pasta with Broccoli in Alfredo Sauce (irregular-shaped mixed food)	Chicken and Wild Rice (irregular-shaped mixed food)
	Red Seedless Grapes (regular-shaped single food)	Chocolate Chip Cookie (regular-shaped single food)
	Chocolate Ice Cream (irregular-shaped single food)	Potato Chips Original (irregular-shaped single food)

https://doi.org/10.1017/S1368980020000816 Published online by Cambridge University Press

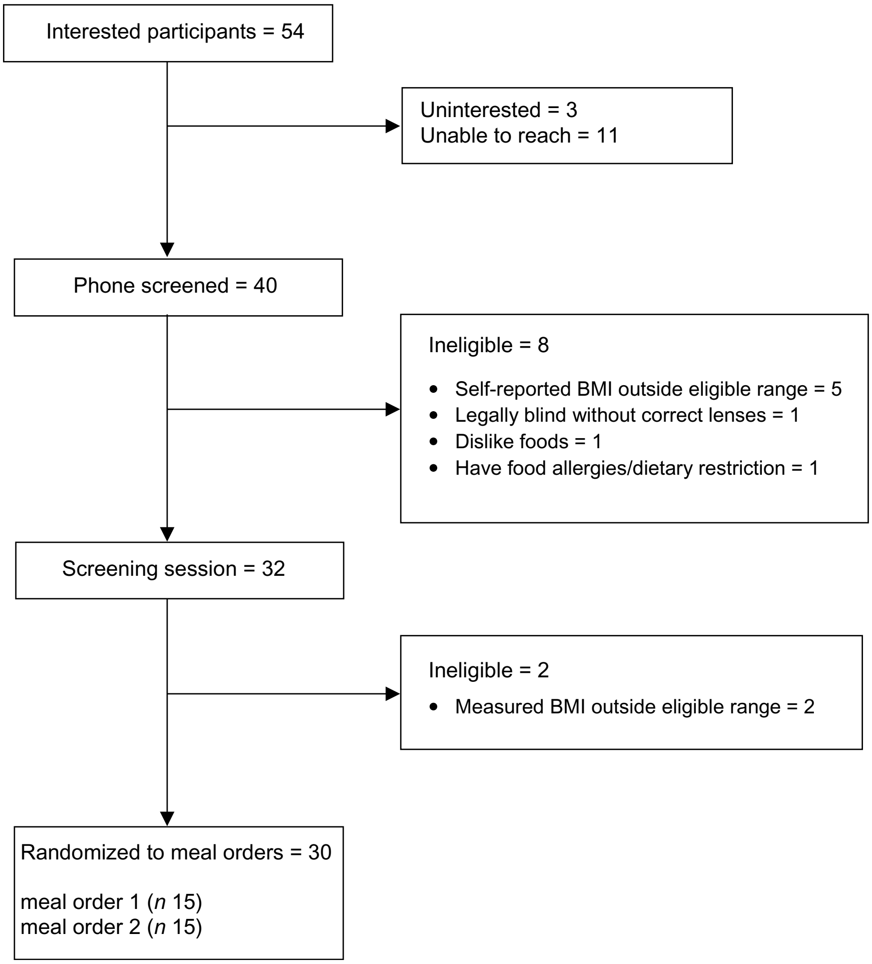

Interested participants = 54
Uninterested = 3
Unable to reach = 11
Phone screened = 40
Ineligible = 8
• Self-reported BMI outside eligible range = 5
• Legally blind without correct lenses = 1
• Dislike foods = 1
• Have food allergies/dietary restriction = 1
Screening session = 32
Ineligible = 2
• Measured BMI outside eligible range = 2
Randomized to meal orders = 30
meal order 1 (*n* 15)
meal order 2 (*n* 15)

Fig. 1 Flow of study participants

the consumption of study foods, reported a favourable preference for study foods (rated each food item ≥3 out of 5 on a Likert scale), were able to complete all meal sessions within 4 weeks of the screening session, were not legally blind without corrected lenses and were able to eat a meal while wearing Sony SmartEyeglass. Individuals were excluded if they wore electronic medical devices (pacemakers and implantable defibrillators) that would be effected by the controller of Sony SmartEyeglass(23).

Study procedure

Screening session

Following phone screening, eligibile participants were invited to a 30-min in-person screening in which participants signed consent forms and filled out demographic questionnaire. Height and weight measures were taken to confirm eligibility.

Meal sessions

Participants were scheduled for two 40-min meal sessions, with approximately 1 week between each session. Participants were asked to stop eating a minimium of 2 h prior to the scheduled meal sessions and only consume water during that period. During meal sessions, participants were instructed that after putting on the Sony SmartEyeglass to initiate the recording via the controller of the Sony SmartEyeglass. After the recording was initiated and prior to starting to eat, participants were instructed to look at each provided food at the table. Then, participants were also instructed to turn their head towards their left shoulder, look at each food from the side and then repeat the same step by turning their head towards the right shoulder. Participants were then asked to start the meal by taking one bite of each provided food. For the first bite of each food, participants were instructed to hold the food, either in their hand or on a fork or a spoon (depending on the food), approximately 12 inches in front of the Sony SmartEyeglass and to look at the food. Following taking the first bite of each provided food, participants were instructed to eat normally until satisfied. Participants were given 30 min to eat. The second meal session followed the same procedure as the first meal session. After the second meal session was completed, participants were given a $20 gift card.

Meal

The meals contained foods categorised into two food shapes (regular and irregular) and two food complexities (single food and mixed food). Each meal contained four foods (see Table 1), with the four foods representing the four potential food categories. Each meal provided approximately 50 % of daily estimated energy needs for each sex[24]. Thus, each meal provided different portions of foods to males and females, in which the overall amount provided was approximately 5125·4 kJ for males and 3974·8 kJ for females. Each food provided approximately 25 % of the energy for each meal.

Wearable device: Sony SmartEyeglass

Sony SmartEyeglass, developed by Sony Corporation, is an eyeglass that is intended to be operated as an Android system mobile device[25]. Sony SmartEyeglass has a display, built-in camera, sensors and a touch-sensitive controller and keys[25]. Sony SmartEyeglass is designed to be worn as usual eyeglasses, and the user is able to operate the eyeglasses via the touch-sensitive controller[25]. The controller can also be connected to an Android system device wirelessly[25]. In the current study, an application was developed for Sony SmartEyeglass to automatically take approximately three images every second.

Automatic analysis tool: DietCam

DietCam[26] is an application that has an algorithm called multi-view food recognition designed to automatically recognise foods from images. In the current paper, we use an updated DietCam algorithm with a deep learning technology[27]. The new DietCam is composed of automatically trained neural network features and detector[27]. New DietCam has the advantage of extending more categories of food detection and higher detection rate and accuracy[27]. DietCam was used to analyse images taken by the Sony SmartEyeglass.

Process of food identification of images

Training DietCam

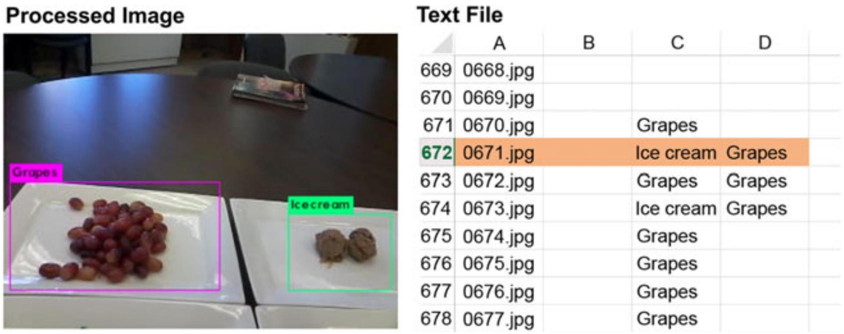
Fourteen randomly selected images from each meal, a total of twenty-eight images, from ten randomly selected participants were used as training images for DietCam. DietCam was trained for food identification for the general food categories (e.g., sandwich, cookie, wrap, grapes, etc.). Foods in the selected images were framed and annotated with food category or categories (each image could have a range of 0–4 food categories annotated) using MATLAB version R2017b with coded programme written by a research staff. Each framed and annotated food category was then cropped out into small image patches for data augmentation by adding additional external images for training and generalisation purpose. The version 2012 dataset from the PASCAL Visual Object Classes[28] with over 17 000 images was used for the data augmentation during training. The training achieved an average of 97 % accuracy.

Automatic image analysis by DietCam

All food images from the remaining twenty participants were input into DietCam for automatic image analysis. Processed food images were labelled with names of the food categories appearing in the image, with a rectangle frame around the identified foods, and provided in a text file with a list of foods identified in each image (e.g., see Fig. 2).

Reference coded by human raters

To determine the accuracy of food identification by DietCam, images captured in the first 5 min of each meal session, with the 5-min period starting when the first food image appeared in the meal, were selected. This period was selected because it captured the start of the meal when participants were instructed to capture images of the food from several angles before starting to eat and also captured images of the food while eating. As approximately three images were taken each second, over a 5-min period, it was anticipated that approximately 900 images would be collected per meal. As images were to be coded in twenty participants for two meals, it was anticipated that approximately 36 000 images would be coded. The selected images were coded by human raters (who were one of the investigators and research staff) into one of the three codes: (i) blurry image, (ii) no food or (iii) specific food



	A	B	C	D
669	0668.jpg			
670	0669.jpg			
671	0670.jpg		Grapes	
672	0671.jpg		Ice cream	Grapes
673	0672.jpg		Grapes	Grapes
674	0673.jpg		Ice cream	Grapes
675	0674.jpg		Grapes	
676	0675.jpg		Grapes	
677	0676.jpg		Grapes	
678	0677.jpg		Grapes	

Fig. 2 Results of DietCam food identification. On the left, a processed image by DietCam is shown, with each rectangle frame representing one food identification, which also appears on the associated text file showed on the right and is highlighted

https://doi.org/10.1017/S1368980020000816 Published online by Cambridge University Press

in an image. For images that contained a food, three codes for each food were used for further classification: (i) 100 % of the food visible on the serving plate (100 % food images), (ii) <100 % of the food visible on the serving plate (partial food images) and (iii) the food in the image but not on the plate (no plate food images), such as when food was on a fork or held in hand. When images contained food, they were coded into all possible food codes, meaning one image could include more than one food image code (e.g., see Fig. 3).

To determine inter-rater agreement for image coding, 33 % of all coded images were coded by two raters. Two raters coded one meal until 90 % agreement was achieved. Once 90 % agreement was achieved, the raters coded thirteen meals independently. Percentage agreement between the raters was determined by dividing the total numbers of images in which agreement occurred between raters by the total numbers of images, then multiplying by 100.

Comparison of results: human raters v. DietCam

Results from human raters were compared with results in the text file of food identification by DietCam. For each image, the comparison produced four outcomes shown in Table 2 for each potential food in the image: TP (rater and DietCam both identify the food), FP (rater does not identify the food while DietCam identifies the food), FN (rater identifies the food while DietCam does not identify the food) or TN (rater and DietCam both do not identify the food). The performance of DietCam in food identification was evaluated using measures of sensitivity, specificity and accuracy.

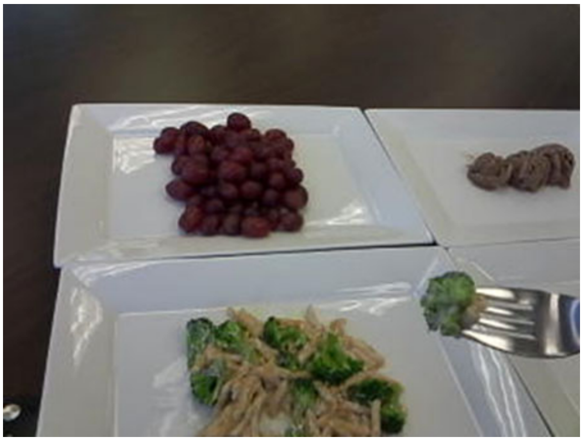


Fig. 3 Example image coded by raters. This image was coded by raters as grapes 100 % available and visible on the serving plate, ice cream 100 % available and visible on the serving plate, pasta dish less than 100 % available and visible on the serving plate and pasta dish in the image but not on the plate

Table 2 Definition of four comparison outcomes: true positive (TP), false positive (FP), true negative (TN) and false negative (FN)

	Predicted (DietCam)	
Actual (human raters)	Specific food present	Specific food not present
Specific food present	TP	FN
Specific food not present	FP	TN

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

and

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}}.$$

Statistical analyses

Data were analysed using SPSS version 24.0 (SPSS Inc.). Demographic information was collected using self-reported questionnaire, which included commonly used classification by the National Institutes of Health (https://grants.nih.gov/grants/guide/notice-files/not-od-15–089.html). For nominal/ordinal data, χ^2 tests, and for interval/ratio data, independent sample *t* tests, with the between-subject factor of meal orders, were conducted to examine the difference between meal orders on participant characteristics. Due to significance found between meal orders for race and ethnicity, these variables were used as covariates in subsequent analyses.

For the first aim, to describe the agreement of identification between raters and DietCam for each specific food and no food images, a mixed ANCOVA, with within-subject factors of meals (meals A and B), food categories in each meal (meal A: cookies, chips, chicken and rice, sandwich and no food; meal B: grapes, ice cream, pasta dish, wrap and no food) and comparison outcomes (TP, FP, FN and TN), with covariates of ethnicity, race and meal orders, was conducted. For this overall analysis, the three codes for the types of food images (100 % food images, partial food images and no plate food images) coded by human raters were recoded and combined into one variable as an overall food code.

For aims 2 and 3, the images coded by raters as no food were excluded from the analyses; thus, the same numbers of images were used in both aims. For aim 2, similar to aim 1, all types of food images (100 % food images, partial food images and no plate food images) coded by human raters were combined into one variable as an overall food code. A mixed ANCOVA was conducted, with within-subject factors of meals (meals A and B), food shapes (regular and irregular), food complexities (single food and mixed food) and comparison outcomes (TP, FP, FN and TN), with covariates of ethnicity, race and meal orders. For aim 3, the same analyses were conducted, but each analysis only included one type of code for the image. For within-subject

https://doi.org/10.1017/S1368980020000816 Published online by Cambridge University Press

comparisons, Greenhouse–Geisser corrections were used to adjust for sphericity. *Post hoc* pairwise comparisons using Bonferroni corrections were used to determine-which groups differed in percentage agreement. For significant outcomes, α was <0·05.

Results

Participant characteristics

Participant characteristics by meal orders are presented in Table 3. Participants were aged 25·1 ± 6·6 years with a BMI of 22·7 ± 1·6 kg/m^2. Participants were 56·7 % female, 43·3 % male, with 96·6 % having some college education and 86·7 % never married. No significant difference occurred between meal orders 1 and 2 for age (P = 0·68), BMI (P = 0·59), sex (P = 0·27), education level (P = 0·15) and marital status (P = 1·00). For race, participants were predominately White (46·7 %) and Asian (46·7 %). Significance was found between meal orders for race (χ^2(3) = 13·7, P = 0·003) with 80·0 % of participants in meal order 1 identifying as Asian and 73·3 % of participants in meal order 2 identifying as White. Significance was also found between meal orders for ethnicity (χ^2(1) = 6·0, P = 0·01] with 100 % of participants identifying as non-Hispanic/Latino in meal order 1 and 66·7 % identifying as non-Hispanic/Latino in meal order 2.

All food and no food images

A total of 36 412 images were coded, in which 2106 images (5·8 %) were coded by raters as blurry. Thus, after excluding blurry images, a total of 34 306 images (17 279 in meal A and 17 027 in meal B) were included in analyses.

Table 3 Participant characteristics (mean and SD)

	Meal order 1 (*n* 15) *		Meal order 2 (*n* 15)*	
	Mean	SD	Mean	SD
Age (years)	25·3	6·2	24·8	7·1
Sex (%)				
Male	53·3		33·3	
Female	46·7		66·7	
BMI (kg/m^2)	22·6	1·6	22·8	1·7
Marital status (%)				
Married	13·3		13·3	
Never married	86·7		86·7	
Education status (%)				
High school (10–12 years)	6·7		0	
Some college (<4 years)	6·7		33·3	
College/University degree	40·0		46·7	
Graduate/professional education	46·7		20·0	
Race (%)				
American Indian/Alaskan Native	0		6·7	
Asian	80·0^a		13·3^b	
White	20·0^a		73·3^b	
Other	0		6·7	
Ethnic heritage (%)				
Hispanic/Latino	0^a		33·3^b	
Not Hispanic/Latino	100·0^a		66·7^b	

*See Table 1 for the description of meal orders.
[a,b]Values with different superscripts are significantly different (P < 0·05).

References coded by human raters

For images included in analyses, 31 617 images (92·2 %) were coded as having foods. A total of 2689 images (7·8 %) were coded as having no food. The total number of codes for both meals A and B identified by human raters was 64 040, with 51·0 % in meal A and 49·0 % in meal B. Rater results are presented in Table 4.

The mean meal percentage agreement between raters was 84·5 ± 3·7 % (images from thirteen meals), and the percentage agreement for meal A (images from six meals) was 85·3 ± 3·4 % and for meal B (images from seven meals) was 83·9 ± 4·0 %.

Identification by DietCam

DietCam identified 26 737 images (77·9 %) with food and 7569 images (22·1 %) with no food. The total number of codes for both meals A and B identified by DietCam was 40 401 codes, with 51·5 % in meal A and 48·5 % in meal B. The results of the identification by DietCam are presented in Table 4.

Food and no food identification

For the identification of each specific food and no food, the overall mean of TP was 22·8 ± 3·4 %, FP was 1·1 ± 0·3 %, TN was 56·7 ± 6·8 % and FN 19·4 ± 4·8 % (see Table 2 for the definition of outcomes). After adjusting for race, ethnicity and meal order, a statistical significance of comparison outcomes was found ($F_{3,48}$ = 4·608, P = 0·04). *Post hoc* tests using Bonferroni correction indicated statistical significance between all comparison outcomes (P < 0·01). The comparison of identification results between human raters and DietCam for each meal is shown in Table 5. The accuracy of identification was 81·2 % for meal A and 79·6 % for meal B. The results of sensitivity, specificity and accuracy are shown in Table 6.

All food images: food shape and complexitiy

For the identification of specific foods in images only containing foods, the overall mean of TP was 24·7 ± 3·1 %, FP was 1·0 ± 0·4 %, TN was 53·4 ± 5·8 % and FN was 20·9 ± 4·5 %. A main effect of comparison outcomes was found ($F_{3,48}$ = 4·9, P = 0·03). *Post hoc* tests using Bonferroni correction indicated statistical significance between all comparison outcomes (P < 0·05). No significant main effects or interactions were found with food shapes or complexities. The comparison of identification results between human raters and DietCam for each meal is shown in Table 5. The accuracy of identification was 78·7 % for meal A and 77·5 % for meal B. The results of sensitivity, specificity and accuracy are shown in Table 6.

Types of food images

The comparison of identification results between human raters and DietCam for different types of images is shown in Table 5.

Table 4 Distribution of codes: DietCam and human raters

	DietCam		Human raters	
	Numbers of codes	%	Numbers of codes (%)	%
Meal A = 17 279 images				
Chocolate chip cookies (total)	6801	32·7	**8990***	**27·5**
Item 100 % on the plate	–		4624†	51·4
Item partially on the plate	–		3987†	44·3
Item in the image but not on the plate	–		379†	4·2
Potato chips (total)	3814	18·3	**9830***	**30·1**
Item 100 % on the plate	–		4703†	47·8
Item partially on the plate	–		4752†	48·3
Item in the image but not on the plate	–		375†	3·8
Chicken and wild rice (total)	3364	18·3	**6335***	**19·4**
Item 100 % on the plate	–		1496†	23·6
Item partially on the plate	–		3985†	62·9
Item in the image but not on the plate	–		854†	13·5
Turkey and Provolone Cheese Sandwich (total)	3090	14·9	**5950***	**18·2**
Item 100 % on the plate	–		1605†	27·0
Item partially on the plate	–		3214†	54·0
Item in the image but not on the plate	–		1131†	19·0
No food	3724	17·9	1539	4·7
Total numbers of codes	20 793		32 644	
Meal B = 17 027 images				
Chocolate Ice Cream (total)	2119	10·8	**9306***	**29·6**
Item 100 % on the plate	–		6238†	67·0
Item partially on the plate	–		2762†	29·7
Item in the image but not on the plate	–		306†	3·3
Grapes (total)	7583	38·7	**10 209***	**32·5**
Item 100 % on the plate	–		4380†	42·9
Item partially on the plate	–		5524†	54·1
Item in the image but not on the plate	–		305†	3·0
Pasta with broccoli and Alfredo sauce (total)	3259	16·6	**6256***	**19·9**
Item 100 % on the plate	–		1259†	20·1
Item partially on the plate	–		3853†	61·6
Item in the image but not on the plate	–		1144†	18·3
Ham and Cheddar Cheese Wrap (total)	2802	14·3	**4475***	**14·3**
Item 100 % on the plate	–		1928†	43·1
Item partially on the plate	–		1828†	40·8
Item in the image but not on the plate	–		719†	16·1
No food	3845	19·6	1150	3·7
Total numbers of codes	19 608		31 396	

*Percentage calculated from the total numbers of codes.
†Percentage calculated from the total numbers of codes under each food.

Hundred percentage food images

For 100 % food images, the overall mean of TP was 11·5 ± 2·9 %, FP was 14·2 ± 2·8 %, TN was 65·1 ± 4·9 % and FN was 9·2 ± 2·7 %. A main effect of comparison outcomes was found ($F_{3,48} = 11{\cdot}3$, $P < 0{\cdot}0001$). *Post hoc* tests using Bonferroni correction indicated statistical significance between TP and TN ($P < 0{\cdot}0001$), TP and FN ($P = 0{\cdot}033$), FP and TN ($P < 0{\cdot}0001$), FP and FN ($P < 0{\cdot}0001$) and TN and FN ($P < 0{\cdot}0001$). No statistical significance was found between TP and FP ($P = 0{\cdot}11$).

A significant interaction of food shapes × comparison outcomes occurred ($F_{3,48} = 4{\cdot}4$, $P = 0{\cdot}022$). The pairwise comparisons indicated statistical significance between regular and irregular food shapes for all comparison outcomes ($P < 0{\cdot}05$). For regular-shaped foods, the overall mean was 14·8 ± 5·6 % for TP, 16·9 ± 4·5 % for FP, 62·9 ± 6·2 % for TN and 5·3 ± 3·5 % for FN. For irregular-shaped food, the overall mean was 8·2 ± 3·7 % for TP, 11·5 ± 3·4 % for FP, 67·2 ± 7·2 % for TN and 13·2 ± 4·1 % for FN. No statistical significance was found for interactions with food complexities or for a main effect of food shapes or complexities. The accuracy of identification was 77·6 % for meal A and 75·2 % for meal B, as shown in Table 6. The accuracy of identification was 77·7 % and 75·1 % for regular-shaped foods and irregular-shaped foods, respectively.

Partial food images

For partial food images, the overall mean of TP was 13·0 ± 2·9 %, FP was 12·6 ± 2·9 %, TN was 63·8 ± 4·4 % and FN was 10·5 ± 3·3 %. A main effect of comparison outcomes was found ($F_{3,48} = 9{\cdot}8$, $P < 0{\cdot}0001$). *Post hoc* tests using Bonferroni correction indicated statistical significance between TP and TN ($P < 0{\cdot}0001$), FP and TN ($P < 0{\cdot}0001$) and TN and FN ($P < 0{\cdot}0001$). No significant difference was found between TP and FP ($P = 1{\cdot}00$), TP and FN ($P = 0{\cdot}16$) and FP and FN ($P = 0{\cdot}42$).

A significant interaction of food shapes × comparison outcomes occurred ($F_{3,48} = 4{\cdot}9$, $P = 0{\cdot}005$). The pairwise

https://doi.org/10.1017/S1368980020000816 Published online by Cambridge University Press

https://doi.org/10.1017/S1368980020000816 Published online by Cambridge University Press

Table 5 Comparison of identification results: DietCam *v.* human raters

		Numbers of codes (%)									
		Food and no food images (*n* 17 279)		All types of food images (*n* 15 740)		100 % food images (*n* 15 740)		Partial food images (*n* 15 740)		No plate food images (*n* 15 740)	
Meal A		*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Chocolate chips cookies	True positive	6456^{a}	37·4	6456^{a}	41·0	3373^{a}	21·4	3069^{a}	19·5	212^{a}	1·3
	False positive	345^{b}	2·0	255^{b}	1·6	3338^{a}	21·2	3642^{a}	23·1	6499^{b}	41·3
	True negative	8244^{c}	47·7	6795^{c}	43·2	7778^{b}	49·4	8111^{b}	51·5	8862^{c}	56·3
	False negative	2234^{d}	12·9	2234^{d}	14·2	1251^{c}	7·9	918^{a}	5·8	167^{d}	1·1
Potato chips	True positive	3727^{a}	21·6	3727^{a}	23·7	2069^{a}	13·1	1653^{a}	10·5	182^{a}	1·2
	False positive	87^{b}	0·5	55^{b}	0·3	1713^{a}	10·9	2129^{a}	13·5	3600^{b}	22·9
	True negative	7688^{c}	44·5	6181^{c}	39·3	9324^{b}	59·2	8859^{b}	56·3	11 765^{c}	74·7
	False negative	5777^{d}	33·4	5777^{d}	36·7	2634^{c}	16·7	3099^{a}	19·7	193^{d}	1·2
Chicken and wild rice	True positive	3215^{a}	18·6	3215^{a}	20·4	866^{a}	5·5	2348^{a}	14·9	420^{a}	2·7
	False positive	149^{b}	0·9	116^{b}	0·7	2465^{a}	15·7	983^{a}	6·2	2911^{b}	18·5
	True negative	11 499^{c}	66·5	9993^{c}	63·5	11 779^{b}	74·8	10 772^{b}	68·4	11 975^{c}	76·1
	False negative	2416^{d}	14·0	2416^{d}	15·4	630^{c}	4·0	1637^{a}	10·4	434^{d}	2·8
Turkey and Provolone Cheese Sandwich	True positive	2939^{a}	17·0	2939^{a}	18·7	1292^{a}	8·2	1582^{a}	10·1	293^{a}	1·9
	False positive	151^{b}	0·9	110^{b}	0·7	1757^{a}	11·2	1467^{a}	9·3	2756^{b}	17·5
	True negative	11 715^{c}	67·8	10 217^{c}	64·9	12 378^{b}	78·6	11 059^{b}	70·3	11 853^{c}	75·3
	False negative	2474^{d}	14·3	2474^{d}	15·7	313^{c}	2·0	1632^{a}	10·4	838^{d}	5·3
No food	True positive	1332^{a}	7·7	–	–	–	–				
	False positive	2392^{b}	13·8	–	–	–	–				
	True negative	13 348^{c}	77·2	–	–	–	–				
	False negative	207^{d}	1·2	–	–	–	–				
Meal B		Food and no food Images (*n* 17 027)		All types of food images (*n* 15 877)		100 % Food images (*n* 15 877)		Partial food images (*n* 15 877)		No plate food images (*n* 15 877)	
Chocolate Ice Cream	True positive	2000^{a}	11·7	2000^{a}	12·6	1407^{a}	8·9	594^{a}	3·7	47^{a}	0·3
	False positive	119^{b}	0·7	101^{b}	0·6	694^{a}	4·4	1507^{a}	9·5	2054^{b}	12·9
	True negative	7883^{c}	46·3	6751^{c}	42·5	8945^{b}	56·3	11 608^{b}	73·1	13 517^{c}	85·1
	False negative	7025^{d}	41·3	7025^{d}	44·2	4831^{c}	30·4	2168^{a}	13·7	259^{d}	1·6
Grapes	True positive	7312^{a}	42·9	7312^{a}	46·1	3247^{a}	20·5	4065^{a}	25·6	232^{a}	1·5
	False positive	271^{b}	1·6	216^{b}	1·4	4281^{a}	27·0	3463^{a}	21·8	7296^{b}	46·0
	True negative	6848^{c}	40·2	5753^{c}	36·2	7216^{b}	45·4	6890^{b}	43·4	8276^{c}	52·1
	False negative	2596^{d}	15·2	2596^{d}	16·4	1133^{c}	7·1	1459^{a}	9·2	73^{d}	0·5
Pasta with Broccoli and Alfredo Sauce	True positive	3139^{a}	18·4	3139^{a}	19·8	854^{a}	5·4	2284^{a}	14·4	506^{a}	3·2
	False positive	120^{b}	0·7	108^{b}	0·7	2393^{a}	15·1	963^{a}	6·1	2741^{b}	17·3
	True negative	11 556^{c}	67·9	10 418^{c}	65·6	12 225^{b}	77·0	11 061^{b}	69·7	11 992^{c}	75·5
	False negative	2212^{d}	13·0	2212^{d}	13·9	405^{c}	2·6	1569^{a}	9·9	638^{d}	4·0
Ham and Cheddar Cheese Wrap	True positive	2485^{a}	14·6	2485^{a}	15·7	1346^{a}	8·5	992^{a}	6·3	244^{a}	1·5
	False positive	317^{b}	1·9	265^{b}	1·7	1404^{a}	8·8	1758^{a}	11·1	2506^{b}	15·8
	True negative	12 449^{c}	73·1	11 351^{c}	71·5	12 545^{b}	79·0	12 291^{b}	77·4	12 652^{c}	79·7
	False negative	1776^{d}	10·4	1776^{d}	11·2	582^{c}	3·7	836^{a}	5·3	475^{d}	3·0
No food	True positive	1015^{a}	6·0	–	–	–	–				
	False positive	2830^{b}	16·6	–	–	–	–				
	True negative	13 047^{c}	76·6	–	–	–	–				
	False negative	135^{d}	0·8	–	–	–	–				

Mean values within a column with different superscripts are significantly different ($P < 0.05$).

Table 6 Comparison results of sensitivity, specificity and accuracy

	Sensitivity (%)	Specificity (%)	Accuracy (%)
Meal A			
Food and no food images	85·0	80·0	81·2
All types of food images	96·8	72·0	78·7
100 % food images	45·0	89·5	77·6
Partial food images	51·3	84·2	75·4
No plate food images	6·7	96·5	72·4
Meal B			
Food and no food images	81·4	79·0	79·7
All types of food images	95·6	71·6	77·5
100 % food images	43·9	85·5	75·2
Partial food images	50·8	87·4	78·4
No plate food images	6·6	97·0	74·7

comparisons indicated statistical significance between regular and irregular food shapes for all comparison outcomes ($P < 0·01$). For regular-shaped food, the overall mean was 15·2 ± 4·5 % for TP, 16·5 ± 5·7 % for FP, 60·7 ± 4·4 % for TN and 7·6 ± 3·4 % for FN. For irregular-shaped food, the overall mean was 10·9 ± 3·4 % for TP, 8·8 ± 3·8 % for FP, 66·9 ± 5·4 % for TN and 13·5 ± 4·9 % for FN. No statistical significance was found for interactions with food complexities or for a main effect of food shapes or complexities. The accuracy of identification was 75·4 % for meal A and 78·4 % for meal B, as shown in Table 6. The accuracy of identification was 76·0 and 77·8 % for regular-shaped foods and irregular-shaped foods, respectively.

No plate food images

For no plate food images, the overall mean of TP was 1·6 ± 0·9 %, FP was 24·1 ± 3·4 %, TN was 71·9 ± 2·9 % and FN was 2·4 ± 1·1 %. A main effect of comparison outcomes was found ($F_{3,48} = 36·3$, $P < 0·0001$). *Post hoc* tests using Bonferroni correction indicated statistical significance between all comparison outcomes ($P < 0·05$).

A significant interaction of food shapes × comparison outcomes occurred ($F_{3,48} = 11·5$, $P = 0·001$). The pairwise comparisons indicated statistical significance between regular- and irregular-shaped foods for FP ($P < 0·0001$) and TN ($P < 0·0001$). For regular-shaped food, the overall mean was 1·5 ± 1·0 % for TP, 30·2 ± 4·4 % for FP, 65·9 ± 4·0 % for TN and 2·4 ± 1·7 % for FN. For irregular-shaped food, the overall mean was 1·7 ± 1·4 % for TP, 17·9 ± 4·2 % for FP, 78·0 ± 4·3 % for TN and 2·4 ± 1·2 % for FN. No statistical significance was found for interactions with food complexities or for a main effect of food shapes or complexities. The accuracy of identification was 72·4 % for meal A and 74·7 % for meal B, as shown in Table 6. The accuracy of identification was 67·4 and 79·7 % for regular-shaped foods and irregular-shaped foods, respectively.

Discussion

This validation study was designed to describe the agreement between human raters and DietCam in identifying specific foods and no food in images, examining foods of different shapes and complexities for all food images and foods of different shapes and complexities in various types of food images. When identification examined the presence of specific foods or no food in an image, DietCam showed an averaged accuracy of 81·2 % for meal A and 79·6 % for meal B. Jia *et al.*[22] examined an automated system identifying the presence of food or no food in images and found an accuracy rate of 91·5 % when analysing images (3900 total images) collected by eButton in thirty participants. When analysing more images (29 515 total images) collected in 1 week by a single participant, the results showed an accuracy rate of 86·4 % in food and no food identification[22]. When analysing a comparable amount of images, the current study found that DietCam is similar, but slightly lower, in accuracy for identifying food and no food images. This difference may be a consequence of DietCam identifying specific foods and no food being present in images (i.e., cookie, ice cream, no food), while the previous investigation was only trying to identify if food was present or not (i.e., food and no food). As DietCam was trained to perform a more specific identification task, this may create more opportunity for error.

The results of analyses for overall food images in shapes and complexities indicate that DietCam shows promise in automated food image identification as over 77 % of images were identified correctly. DietCam also has a low false identification percentage (identifying a specific food when it is not in the image), 1·0 ± 0·4 %. The findings also suggest that there was no difference in DietCam's ability in identifying regular- and irregular-shaped foods and single and mixed foods.

When analysing the results of specific types of food images, there was an interaction of food shapes by comparison outcomes, with regular-shaped foods having a slightly higher accuracy percentage for the 100 % food images and the irregular-shaped foods having a higher accuracy percentage for the partial and no plate food images. This difference may be a consequence of the partial and no plate food images showing incomplete images of the foods, thus showing all foods as irregular shaped, making it more challenging for DietCam to identify the regular-shaped foods. Overall, DietCam appears to more accurately identify foods in both 100 % and partial food images. For no plate food images, DietCam has higher amounts of FP, over 20 %, as compared with 100 % and partial food images (<20 %). This would also mean that when analysing no plate food images, DietCam may identify a specific food when the food is not present in the image. These results suggest that only analysing 100 % and partial food images may optimise food identification outcomes.

https://doi.org/10.1017/S1368980020000816 Published online by Cambridge University Press

The findings in the current study are novel since none of the previously investigated passive image-assisted dietary assessment methods possess an automated food identification system that has the ability to automatically identify specific food items. Previous studies(18–21) validating different passive image-assisted dietary assessment methods relied on participants or human raters to recognise specific food items consumed. The automated food identification through DietCam completely eliminated human effort in the food identification process. Boushey *et al.*(17) validated an active image-assisted dietary assessment method, mobile food record, with an automated system to classify food items. However, the results of the automated identification required participants to review and confirm while also giving the options for participants to make changes as needed(17). The investigators also did not provide any information on how frequently a participant needed to correct or change the automated identification conducted by mobile food record. In addition, mobile food record required a specific colour fiducial marker to facilitate the food identification process(29), while DietCam does not require any reference objects to facilitate the process.

For strengths, the current study included a larger and more diverse sample compared with most previous studies investigating passive image-assisted dietary assessment(18–22). Furthermore, the current study included thousands of pictures in the analyses and analysed pictures collected during an actual eating situation. Most importantly, this was the first study to examine if food characteristics and types of images influence accuracy of automated identification.

The study has several limitations. First, the text files did not specify if the rectangle frame was correctly placed on the identified foods or not. For example, the text file may identify that cookies were in the image, but the actual image may have a rectangle frame around the sandwich and label the frame cookies (i.e., frame around the wrong food). Thus, there may be additional errors in identification than what could be determined from the text file. Second, the current DietCam system was trained on a limited number of images (*n* 28). Since the collected images captured a wide variety of different angles of the foods, the small number of training images might not have captured all the angles required to completely train the DietCam system to identify each food item. Third, the current study included limited types of foods to test the automated food identification of DietCam. Thus, it is unclear on DietCam's ability in correctly identifying foods that are consumed in eating occasions with greater variety of foods or across several eating occasions in a day. This limits the generalisability of the results of the investigation. Future investigations will need to examine DietCam's ability to identify foods in several types of eating occasions, across several days in free-living situations, which will allow multiple types of foods to be consumed in highly variable settings to increase the generalisability of the findings. As this research is conducted, ideally the capacity is enhanced regarding food identification in DietCam that uses a standardised data system for coding food, such as FoodEx2(30), to increase ability to link collected image data to food composition data.

To better enhance the understanding of the accuracy of DietCam in the automated food identification, future studies should investigate number of images that are needed for the food identification in dietary assessment. Passive image-assisted dietary assessment collects more images than active image-assisted dietary assessment (i.e., thousands of images *v.* two images (one before and one after food consumption) during an eating occasion). Thus, from a dietary assessment standpoint, the food identification potentially may only need to be performed until no new additional foods are identified in an eating occasion. For this type of process, a specific food, once it was identified in at least one image, does not need to be correctly identified in every image since that food would be considered a consumed item. From this perspective, as all foods consumed were identified in at least one image, DietCam has 100 % accurate identification, but there is not 100 % accuracy in all images taken. For accuracy of dietary assessment, it is most important that the automated food identification system would not identify a food in an image when that food was not actually there, and thus not consumed.

Due to the number of images that passive image-assisted dietary assessment collects, it also has the capability to provide enhanced information about dietary intake, such as speed of eating, introduction of other items consumed that were not planned to be consumed at the start of the eating occasion and the environment in which eating is occurring. This additional information may be important, particularly for interventions changing dietary intake. Thus, this method of assessment has the capacity to collect more detailed information in real time than active image-assisted dietary assessment. Future research is needed to examine this aspect of passive image-assisted dietary assessment.

In conclusion, DietCam shows promise in accurately identifying specific food items with different shapes and complexities. When the types of images are examined, DietCam is most accurate when 100 % and partial food images are analysed. Future research should focus on enhancing DietCam's ability to identify in greater detail, beyond broad categories of food, components of foods consumed and examining the feasibility of this system in analysing images collected in free-living situations.

Acknowledgements

Acknowledgements: The authors would like to acknowledge all participants and research staff for their significant contributions to the current study. *Financial support:* The current research received no specific grant from any funding agency, commercial or not-for-profit sectors. *Conflict of interest:* There are no conflicts of interest. *Authorship:*

https://doi.org/10.1017/S1368980020000816 Published online by Cambridge University Press

T.-K.C. and H.A.R. were responsible for designing the study. T.-K.C. was responsible for collecting the data, analysing the data and writing the manuscript. H.A.R. contributed to analysing the data and the writing of the manuscript. T.-K.C. and H.A.R. interpreted the results of analyses. J.T. advised on the image collection and analysis process. Y.L. trained the DietCam system and processed the image analysis using DietCam. All authors approved the final draft of the manuscript. *Ethics of human subject participation:* The current study was conducted according to the guidelines laid down in the Declaration of Helsinki, and all procedures involving human subjects/patients were approved by the Institutional Review Board at the University of Tennessee Knoxville. Written informed consent was obtained from all subjects/patients. The current study was registered at ClinicalTrials.gov (NCT03267004).

References

1. Kirkpatrick SI & Collins CE (2016) Assessment of nutrient intakes: introduction to the special issue. *Nutrients* **8**, 184.
2. Thompson FE & Subar AF (2013) Dietary assessment methodology. In *Nutrition in the Prevention and Treatment of Disease*, 3rd ed., pp. 5–46 [A Coulston, C Boushey, M Ferruzzi *et al.*, editors]. Bethesda, MD: National Cancer Institute.
3. Thompson FE, Subar AF, Loria CM *et al.* (2010) Need for technological innovation in dietary assessment. *J Am Diet Assoc* **110**, 48–51.
4. Lam YY & Ravussin E (2016) Analysis of energy metabolism in humans: a review of methodologies. *Mol Metab* **5**, 1057–1071.
5. Johnson RK, Yon BA & Hankin JH (2008) Dietary assessment and validation. In *Research Successful Approaches*, 3rd ed., pp. 187–204 [ER Monsen and L Van Horn, editors]. Chicago, IL: Diana Faulbaber.
6. Black AE, Prentice AM, Goldberg GR *et al.* (1993) Measurements of total energy expenditure provide insights into the validity of dietary measurements of energy intake. *J Am Diet Assoc* **93**, 572–579.
7. Buhl KM, Gallagher D, Hoy K *et al.* (1995) Unexplained disturbance in body weight regulation: diagnostic outcome assessed by doubly labeled water and body composition analyses in obese patients reporting low energy intakes. *J Am Diet Assoc* **95**, 1393–1400; quiz 1401–1392.
8. Prentice AM, Black AE, Coward WA *et al.* (1986) High levels of energy expenditure in obese women. *Br Med J (Clin Res Ed)* **292**, 983–987.
9. Moshfegh AJ, Rhodes DG, Baer DJ *et al.* (2008) The US Department of Agriculture Automated Multiple-Pass Method reduces bias in the collection of energy intakes. *Am J Clin Nutr* **88**, 324–332.
10. Pikholz C, Swinburn B & Metcalf P (2004) Under-reporting of energy intake in the 1997 National Nutrition Survey. *N Z Med J* **117**, U1079.
11. Scagliusi FB, Ferriolli E, Pfrimer K *et al.* (2008) Underreporting of energy intake in Brazilian women varies according to dietary assessment: a cross-sectional study using doubly labeled water. *J Am Diet Assoc* **108**, 2031–2040.
12. Subar AF, Kipnis V, Troiano RP *et al.* (2003) Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study. *Am J Epidemiol* **158**, 1–13.
13. Bandini LG, Schoeller DA, Cyr HN *et al.* (1990) Validity of reported energy intake in obese and nonobese adolescents. *Am J Clin Nutr* **52**, 421–425.
14. Gemming L, Utter J & Ni Mhurchu C (2015) Image-assisted dietary assessment: a systematic review of the evidence. *J Acad Nutr Diet* **115**, 64–77.
15. Martin CK, Correa JB, Han H *et al.* (2012) Validity of the Remote Food Photography Method (RFPM) for estimating energy and nutrient intake in near real-time. *Obesity (Silver Spring)* **20**, 891–899.
16. Rollo ME, Ash S, Lyons-Wall P *et al.* (2015) Evaluation of a mobile phone image-based dietary assessment method in adults with type 2 diabetes. *Nutrients* **7**, 4897–4910.
17. Boushey CJ, Spoden M, Delp EJ *et al.* (2017) Reported energy intake accuracy compared to doubly labeled water and usability of the mobile food record among community dwelling adults. *Nutrients* **9**, 312.
18. Arab L, Estrin D, Kim DH *et al.* (2011) Feasibility testing of an automated image-capture method to aid dietary recall. *Eur J Clin Nutr* **65**, 1156–1162.
19. Gemming L, Rush E, Maddison R *et al.* (2015) Wearable cameras can reduce dietary under-reporting: doubly labelled water validation of a camera-assisted 24 h recall. *Br J Nutr* **113**, 284–291.
20. Pettitt C, Liu J, Kwasnicki RM *et al.* (2016) A pilot study to determine whether using a lightweight, wearable micro-camera improves dietary assessment accuracy and offers information on macronutrients and eating rate. *Br J Nutr* **115**, 160–167.
21. Jia W, Chen HC, Yue Y *et al.* (2014) Accuracy of food portion size estimation from digital pictures acquired by a chest-worn camera. *Public Health Nutr* **17**, 1671–1681.
22. Jia W, Li Y, Qu R *et al.* (2018) Automatic food detection in egocentric images using artificial intelligence technology. *Public Health Nutr* **22**, 1168–1179.
23. Sony Coporation (2015) SmartEyeglass developer edition reference guide.
24. U.S. Department of Health and Human Services & U.S. Department of Agriculture (2015) Appendix 2: Estimated calorie needs per day, by age, sex, and physical activity level. In *2015–2020 Dietary Guidelines for Americans*, 8th ed., pp. 77–78. Washington DC: USDDH and USDA.
25. Sony Coporation (2017) SmartEyeglass: API Overview. https://developer.sony.com/develop/wearables/smarteyeglass-sdk/api-overview/ (accessed February 2018).
26. Kong F, He H, Raynor HA *et al.* (2015) DietCam: multi-view regular shape food recognition with a camera phone. *Pervasive Mob Comput* **19**, 108–121.
27. Redmon J, Divvala S, Girshick R *et al.* (2016) You only look once: unified, real-time object detection. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788. Las Vegas, NV: IEEE.
28. Everingham M, Van Gool L, Williams CKI *et al.* (2012) The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html (accessed February 2018).
29. Boushey CJ, Spoden M, Zhu FM *et al.* (2017) New mobile methods for dietary assessment: review of image-assisted and image-based dietary assessment methods. *Proc Nutr Soc* **76**, 283–294.
30. European Food Safety Authority (2015) The food classification and description system FoodEx 2 (revision 2). *EFSA Support Publ* **12**, 804E.

https://doi.org/10.1017/S1368980020000816 Published online by Cambridge University Press