Automatic Image Recognition Meal Reporting among Young Adults: A Randomized Controlled Trial

Prasan Kumar Sahoo, Sherry Yueh-Hsia Chiu, Yu-Sheng Lin, Chien-Hung Chen, Denisa Irianti, Hsin-Yun Chen, Mekhla Sarkar, Ying-Chieh Liu

Submitted to: JMIR mHealth and uHealth
on: May 21, 2024

Table of Contents

Automatic Image Recognition Meal Reporting among Young Adults: A Randomized Controlled Trial

Prasan Kumar Sahoo[1] PhD; Sherry Yueh-Hsia Chiu[2] PhD; Yu-Sheng Lin[3] MD; Chien-Hung Chen[4] MSC; Denisa Irianti[1] MSC; Hsin-Yun Chen[5] MSC; Mekhla Sarkar[6] PhD; Ying-Chieh Liu[3, 1] PhD

[1]Department of Industrial Design, College of Management, Chang Gung University Taoyuan TW
[2]Department of Health Care Management Taoyuan TW
[3]Health Promotion Center, Department of Internal Medicine, Chang Gung Memorial Hospital Taoyuan TW
[4]Cyber Security Technology Institute, Institute for Information Industry Taipei TW
[5]Department of Nutrition Therapy, Chang Gung Memorial Hospital Taoyuan TW
[6]Department of Computer Science and Information Engineering Taoyuan TW

Corresponding Author:
Ying-Chieh Liu PhD
Department of Industrial Design, College of Management, Chang Gung University
No.259, Wenhua 1st Rd., Guishan Dist.
Taoyuan
TW

Abstract

Background: Advances in artificial intelligence (AI) technology have raised new possibilities for the effective evaluation of daily dietary intake, but more empirical study is needed for the use of such technologies under realistic meal scenarios. This study developed an automated food recognition technology, which was then integrated into its previous design to improve usability for meal reporting. The newly developed app allowed for the automatic detection and recognition of multiple dishes within a single real-time food image as input. Application performance was tested using young adults in authentic dining conditions.

Objective: A two-group comparative study was conducted to assess app performance using metrics including accuracy, efficiency, and user perception. The experimental group, named Automatic Image-based Reporting (AIR) group, was compared against a control group using the previous version, named the Voice Input Reporting (VIR) group. Each application is primarily designed to facilitate a distinct method to food intake reporting. AIR users capture and upload images of their selected dishes, supplemented with voice commands where appropriate. VIR users supplement the uploaded image with verbal inputs for food names and attributes.

Methods: The two mobile apps were subjected to a head-to-head parallel randomized evaluation. A cohort of 42 young adults aged 20-25 years (9 male and 34 female) was recruited from a university in Taiwan and randomly assigned to two groups, i.e., AIR (n=22) and VIR (n=20). Both groups were assessed using the same menu of seventeen dishes. Each meal was designed to represent a typical lunch or dinner setting, with one stable, one main course, and three side dishes. All participants used the app on the same type of smartphone, with the interfaces of both using uniform user interactions, icons, and layouts. Analysis of the gathered data focused on assessing reporting accuracy, time efficiency, and user perception.

Results: For the AIR group, 86% dishes were correctly identified, whereas 68% dishes were accurately reported. The AIR group exhibited a significantly higher degree of identification accuracy compared to the VIR group ($p<.001$). The AIR group also required significantly less time to complete food reporting ($p<.001$). SUS scores showed both apps were perceived as having high usability and learnability ($P=.20$).

Conclusions: The AIR group outperformed the VIR group in terms of accuracy and time efficiency for overall dish reporting within the meal testing scenario. While further technological enhancement may be required, the integration of AI vision technology into existing mobile applications holds promise. Our results provide an evidence-based contributing for the integration of automatic image recognition technology into existing apps in terms of user interaction efficacy and overall ease of use. Further empirical work is required including full-scale randomized controlled trials and assessments of user perception under a range of dining conditions. Clinical Trial: International Standard Randomized Trial Registry ISRCTN27511195; https://doi.org/10.1186/ISRCTN27511195

(JMIR Preprints 21/05/2024:60070)
DOI: https://doi.org/10.2196/preprints.60070

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.
2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Automatic Image Recognition Meal Reporting among Young Adults: A Randomized Controlled Trial

Abstract

Background: Advances in artificial intelligence (AI) technology have raised new possibilities for the effective evaluation of daily dietary intake, but more empirical study is needed for the use of such technologies under realistic meal scenarios. This study developed an automated food recognition technology, which was then integrated into its previous design to improve usability for meal reporting. The newly developed app allowed for the automatic detection and recognition of multiple dishes within a single real-time food image as input. Application performance was tested using young adults in authentic dining conditions.

Objective:-A two-group comparative study was conducted to assess app performance using metrics including accuracy, efficiency, and user perception. The experimental group, named Automatic Image-based Reporting (AIR) group, was compared against a control group using the previous version, named the Voice Input Reporting (VIR) group. Each application is primarily designed to facilitate a distinct method to food intake reporting. AIR users capture and upload images of their selected dishes, supplemented with voice commands where appropriate. VIR users supplement the uploaded image with verbal inputs for food names and attributes.

Methods: The two mobile apps were subjected to a head-to-head parallel randomized evaluation. A cohort of 42 young adults aged 20-25 years (9 male and 34 female) was recruited from a university in Taiwan and randomly assigned to two groups, i.e., AIR (n=22) and VIR (n=20). Both groups were assessed using the same menu of seventeen dishes. Each meal was designed to represent a typical lunch or dinner setting, with one stable, one main course, and three side dishes. All participants used the app on the same type of smartphone, with the interfaces of both using uniform user interactions, icons, and layouts. Analysis of the gathered data focused on assessing reporting accuracy, time efficiency, and user perception.

Results: For the AIR group, 86% dishes were correctly identified, whereas 68% dishes were accurately reported. The AIR group exhibited a significantly higher degree of identification accuracy compared to the VIR group (p<.001). The AIR group also required significantly less time to complete food reporting (p<.001). SUS scores showed both apps were perceived as having high usability and learnability (P=.20).

Conclusion: The AIR group outperformed the VIR group in terms of accuracy and time efficiency for overall dish reporting within the meal testing scenario. While further technological enhancement may be required, the integration of AI vision technology into existing mobile applications holds promise. Our results provide an evidence-based contributing for the integration of automatic image recognition technology into existing apps in terms of user interaction efficacy and overall ease of use. Further empirical work is required including full-scale randomized controlled trials and assessments of user perception under a range of dining conditions.

Trial Registration: International Standard Randomized Trial Registry ISRCTN27511195; https://doi.org/10.1186/ISRCTN27511195

Keywords:
Automatic food image recognition, speech recognition, artificial intelligence, usability evaluation,

mHealth, randomized trial

Introduction

Background

In 2022, the World Health Organization [1,2] classified 2.5 billion adults (> age 18 and older) as overweight and thus susceptible to chronic diseases associated with obesity. Failure to maintain appropriate nutrition among young people leads to a range of health issues later in life [3,4]. Young people also increasingly integrate smartphones into their daily lives, raising a growing interest in using this technology platform to delivering health-improving behavioral interventions, including healthy eating among young adults [5,6]. Such mobile health (mHealth) interventions are increasingly used to encourage healthy eating behaviors [7-10], and are increasingly popular among young users [11,12]. mHealth technologies already play a significant role in reshaping healthcare access [13,14] and allow for broad access to scalable and cost-effective solutions [8].

AI-based services are an emerging trend [15,16], with an increasingly substantial impact on various healthcare domains, providing enhanced accuracy, improved outcomes, and cost-effectiveness [17-19], and healthcare professionals have been found to hold favorable views toward AI [20, 21]. Machine learning, an increasingly mature AI application, has the potential to revolutionize the mHealth domain [11,14].

Challenges in dietary intake input

An ideal mHealth app should prioritize ease of use, reliability, and long-term engagement [13, 22,23]. However, manually entering dietary intake information poses significant usability challenges for mHealth app users [24], potentially resulting in inaccurate or incomplete reporting, and thus undermining the efficacy of managing healthy eating habits [16]. This raises an urgent need to minimize the operational loading of users [25,26] as the ease and effectiveness of food data entry methods has a direct and significant impact on the usability of dietary tracking applications [23]. Recent advancements in computer vision and deep learning show potential for replacing traditional input methods [27,28]. This study integrates automatic image recognition technology [29-31] into an mHealth app which originally used voice-based inputs, seeking to improve accuracy and time efficiency in meal reporting.

Objectives

This study assesses the potential advantages of integrating the developed automatic food recognition technology into the existing voice recording app [32]. The key technological components use newly developed AI features to integrate a set of food images representing both single and mixed dishes [31].

Methods

General Overview of the Approach

The original app was designed to aid individuals in improving their dietary habits [33]. The previous iteration included voice-inputs to enhance meal reporting, using AI services, specifically Google AI, to transcribe speech into text [32]. Users could use this feature to vocally report food ingredients, portion sizes, cooking methods, and other attributes for individual dishes. While the existing design was shown to be positive in terms of accuracy and was generally well-received by users, there were concerns regarding the accuracy and time-consuming nature of completing meal reporting for an entire meal. Furthermore, in authentic dietary intake scenarios, voice reporting during meal consumption was not always convenient. Consequently, we developed the latest version to enhance the existing design.

The iterative development process was rooted in a user-centered design model [34], and included research, ideation, and implementation stages. We reviewed the relevant literature and commercially available apps, along with extensive brainstorming sessions among team members to generate diverse design ideas. AI techniques have been increasingly applied to food identification and nutrition related applications, thus one idea raised in the ideation stage was to allow users take a single photo of an entire meal for analysis using AI-based recognition, rather than to process individual dishes in sequence. The initial automated meal recognition system was developed and validated using of our developed AI recognition model extracting features from convolutional networks (CNN) [31] tested under a laboratory setting. The results offered relatively high mean average precision for a range of dish types. The newly developed feature allowed users to upload meal image for automatic recognition using the AI engine located on a remote server. This functionality was then integrated into the previous app version to improve convenience, accuracy, and time-efficiency for meal reporting. The previous version, the voice-only reporting (VIR) app, allowed users to simultaneously use verbal reporting of food names and attributes [32].

App Implementation

The two apps were implemented in a 6.8-inch smartphone using the Android operating system. The AI server using in our previous research was improved to simultaneously recognize a set of multiple dishes and become a relatively high accuracy of food-image recognition under a laboratory setting [31]. Both apps used the Google Speech Cloud service (Google, Inc.) for continuous speech recognition. The developed interfaces included user-friendly design elements, such as large-sized buttons and text, a simple layout, and high-contrast colors. Based on recommended design guidelines [32], the two apps shared a common interface design including the placement of buttons, text, and icons. Clear and intuitive visual cues were used to facilitate user interaction.

App Operation Overview

Figure 1 summarizes the four major stages in user interaction for the AIR design. In the first stage, users take and upload a photo of a meal (see Fig. 1A-1C), followed by modification of the food ingredients and cooking method for each of the dish and the provision of additional information, e.g. portion size (Fig. 1I-1J). Users reviewed, revised and confirmed the image, calorie content, and macronutrient information for each dish (see Fig. 1J and 1K) and were given the opportunity to add missing dishes (Fig. 1J and 1L).

Automatic Image-Based Reporting

To activate the app, users click the "start" button (Fig. 1A) and then activate the smartphone's built-in camera (Fig. 1B). To capture the meal photo, the user clicks the icon in the lower-right corner (Fig. 1C), and can retake the photo using the icon on the lower-left. Once confirmed by the user using the upload button (lower-right in Fig. 1D), the meal photo is uploaded to the remote AI server for analysis. Analysis results are then sent to the app, with white dots appearing above each dish which was successfully recognized. In Fig. 1, four dishes were successfully recognized (see Fig. 1D). Figures 1E-1H illustrate the steps by which users select the correct dish name.

For each recognized dish, up to three possible names with the highest confidence scores are provided (see Fig. 1E). The user then selects the correct name. Upon confirmation, the color of the small circle switches from white to green, indicating that the dish had been completed (Figs. 1E-1H).

If the dish item is not recognized, or if the selection provided does not include an appropriate answer, the user resorts to an alternative input method, i.e., voice input. To use the voice input method, the user drags a boundary square to surround a specific dish, then clicks the "voice" button and recites the dish name aloud (see Figs. 1I-1K). For example, Fig. 1J shows an unrecognized dish with one to three possible answers, prompting the user to respond select an option using the “check” button or to return to the previous step using the “undo” button. Once all the dishes are reported, the user proceeds to the Food Diary Page to input related information such as portion size and cooking method (see Fig. 1J). Once dish reporting is complete, detailed information is provided for the entire meal (Fig. 1L above), and for each individual item (Fig. 1L below), including calorie count, food item name, and food attributes.

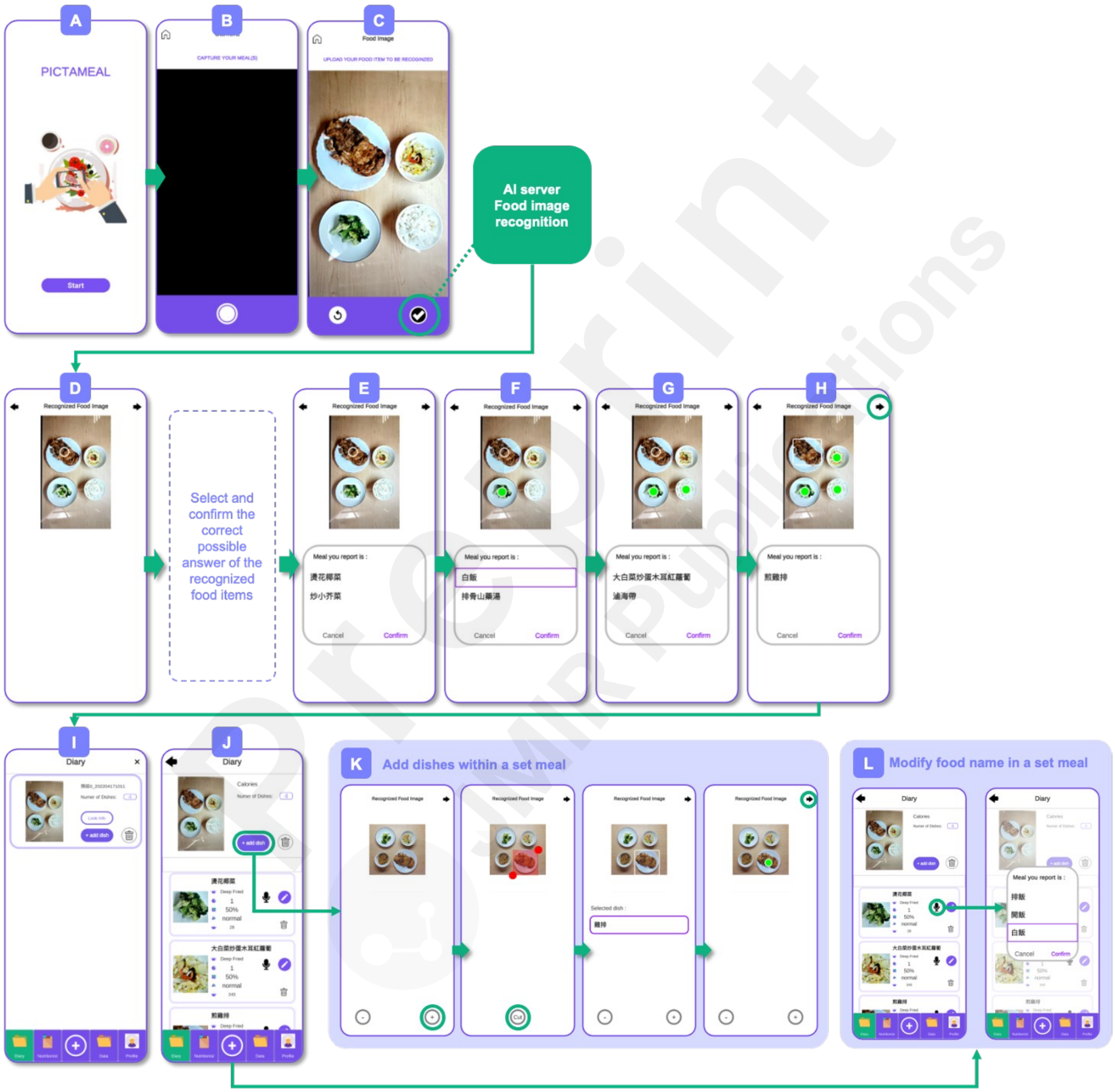

Figure 1. AI image-based food reporting operation, using steamed rice as an example

Voice-Input Reporting

The voice-input reporting process initially follows the procedure depicted in Figs. 1A-1C. However, rather than uploading the image to the server for automatic recognition (as depicted in Figs. 1E-1H), users manually manipulate the cursor (depicted as a red square in Fig. 1K) to scroll down and adjust the size of the area displaying the dish image. The scrolled image is then stored for later visualization on the subsequent page (Fig. 1H). Once the desired dish is identified, the user clicks the microphone icon and records the dish name. Then the user uses the voice-input function to report the dish ingredients and cooking method, prompting the system to provide up to three options for selection and confirmation by clicking the check button, thereby concluding the reporting process. If none of the supplied answers are correct, the user clicks the microphone icon and repeats the process until the correct answer is determined. Once each dish reporting is finalized, a green circle appears above the dish. Once all dish reporting is complete, the user proceeds to the Food Diary Page to input related food information using the voice reporting process previously discussed.

Study Design and Participant Recruitment

A parallel two-group randomized trial was designed to evaluate the relative effectiveness of the two apps. The study protocol was reviewed by the Ethics Committee of Chang Gung Memorial Hospital and received Institutional Review Board approval (202101985B0C501). Study subject recruitment was conducted through notices placed on bulletin boards in Chang Gung University in Taoyuan City, Taiwan. Registration, schedule arrangement, and collection of background information were conducted through an online form. Biographic data were used to allocate participants into the AIR and VIR groups. Self-reported baseline information included gender, age, BMI, experience in nutrition education, use of nutrition-related apps, cooking experience, and experience using mobile phones/tablets. Eligible participants were (1) aged from 20 to 25 years and (2) capable of operating the app on their mobile phones. Participants currently under any form of dietary control, currently engaged in deliberate weight loss, or following a vegetarian diet were excluded. The assessment was conducted in a cafeteria at Chang Gung University.

Dishes for the experiment were selected under the supervision of a senior nutritionist and consisted of typical local foods familiar to study participants. Foods were presented in terms of set meals involving 17 food items were used to represent lunch and dinner. Each set meal contained five food items (i.e., a staple food; a main course; a dish with one ingredient, such as stir-fried broccoli; a dish with two ingredients, such as stir-fried egg with tomato; and a dish with three ingredients).

Sample Size Estimation

Our previous study of customized dietary recording among young subjects [35] found that the mean duration difference for the assessment of a fried pork chop was 0.7464 (2.60 versus 3.34 seconds) with SD=0.93. The sample size of that study provided a statistical power of 80% and a two-tailed alpha level of 5%, indicating the minimum sample size required was 26 subjects each for the AIR and VIR groups.

Randomization

A total of 42 young adults were recruited and provided informed consent. SAS was used to generate randomized lists of equal size with a 1:1 ratio for the two study, with 22 and 20 participants respectively assigned to the AIR and VIR groups.

Evaluation Outcomes

Three outcome measures were evaluated to assess the respective performance of the two mobile apps. Data were collected automatically from each participant's interaction, including the tapping of function buttons, and each interaction was logged with a timestamp. In addition, all suggested results were recorded, along with the user’s subsequent actions.

Accuracy

Accuracy was defined as the degree to which the response provided by the participant during app operation matched the predefined answer. Conversely, an incorrect response occurred if the app output did not match the predefined answer. The accuracy rate for a certain dish, e.g., fried rice, was calculated as the overall number of correct responses divided by the total number. Furthermore, following our previous study [32], four error types were defined as follows: "Missing cooking method(s)", where the app provided an incorrect answer due to the absence of a cooking method (e.g., "stir-fry"); "Incorrect cooking method", where the app provided an incorrect answer due to the use of the wrong cooking method; "Irrelevant food name", where the incorrect answer was due to a food name not matching the desired one; and "Missing food ingredient(s)", defined as when essential food ingredients (e.g., rice) were not included in the reported answer.

Task Duration

The operating duration began when the participant started to report a food dish until the reporting task was completed. For the AIR group, having uploaded the image, the task duration was calculated from the time the participant clicked on the dish displayed in the screen to when the participant clicked the “complete” button. For the VIR group, the task duration was calculated from the time the participant clicked the “voice” button to begin recording to when the participant clicked the “complete” button.

Perception

Brooke’s System Usability Scale (SUS) [36] measures participant perception using a questionnaire comprising 10 items on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Consistent with the approach outlined by Bangor et al. [37], the mean SUS score was used alongside an adjective rating scale in which mean scores of 35.7, 50.9, 71.4, and 85.5 respectively corresponded to the adjective scales "poor," "ok," "good," and "excellent".

Assessment Procedures

The experiment was conducted by two research assistants who obtained informed consent from all participants. All participants used an identical 6.8-inch Android smartphone, and all trials took place in the same university cafeteria, with individual participant sessions scheduled by appointment. Prior to the experiment, each participant underwent training by watching an instructional video illustrating how to use the food reporting app. Each participant was allowed to interact with the app for several minutes to ensure familiarity prior to the trial, and was informed that task completion time was included as a performance metric.

Each individual trial included two consecutive sessions, each requiring the reporting of a set meal. Each set meal included a staple food, a main course, a dish with one ingredient, a dish with two ingredients, and a dish with three ingredients. Participants were allowed to rest for up to three minutes between each test. The total test duration for each participant was approximately one hour, and all participants successfully completed the assessment. Figure 3 depicts the experimental process.

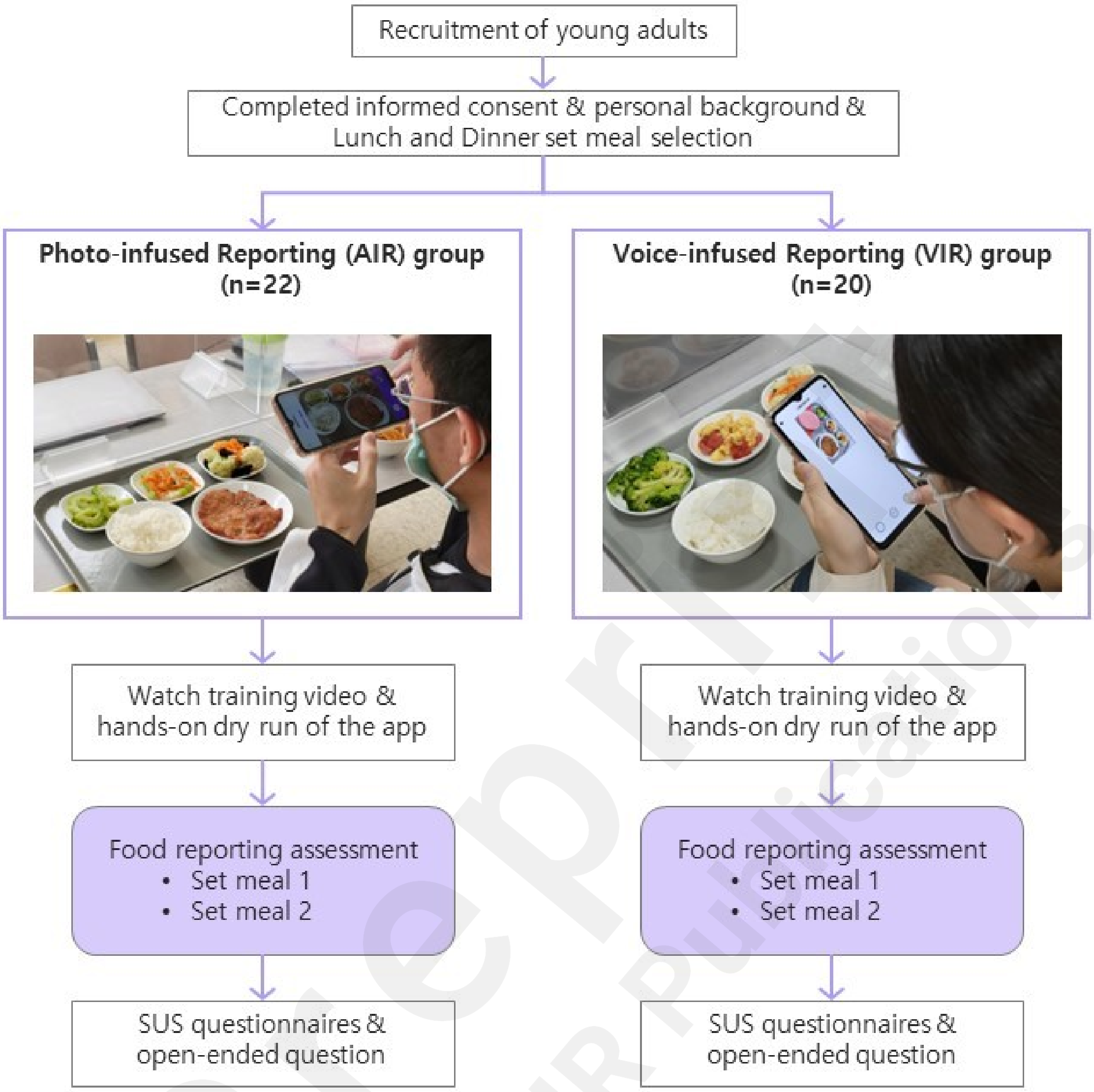

Figure 3. App evaluation flow using a randomized design.

Statistical Analysis

Chi-square and t-test were applied to respectively examine the baseline characteristics of participants for categorical and continuous variables. The accuracy between different groups was reported as the proportion of error, calculated as the number of errors/total answer items. The time duration for operating assessment was also used to evaluate efficiency. As the time duration is a continuous variable, a t-test was used to assess and compare the difference between the two groups. This comparison was also applied for dishes with different ingredients. SAS version 9.1.4 software (SAS Institute) was used to conduct all statistical analyses. In all two-tailed statistical tests, a *p*-value below 0.05 was considered as having statistical significance.

Results

Participant Characteristics

All 42 participants completed the experiment (9 male and 34 female, with a mean age of 21). As shown in Table 1, 22 and 20 respondents were randomly assigned to the AIR and VIR groups.

Table 1. Participant characteristics of AIR and VIR reporting group [a]

Variables	Total (n=42)	AIR (n=22)	VIR (n=20)	*P-value*
Gender, n (%)				
Female	33 (79)	18 (82)	15 (76)	.601
Male	9 (21)	4 (18)	5 (24)	
Age (years)[a], mean (SD)	21.21 (1.39)	21.32 (1.29)	21.10 (1.52)	.617
BMI (kg/m2)[a], mean (SD)	21.10 (3.17)	20.50 (2.77)	21.80 (3.51)	.210
Education level, n (%)				.601
Bachelor’s degree	34 (81)	17 (77)	17 (85)	
Master’s degree	8 (19)	5 (23)	3 (15)	
Q1. Experience with nutrition-related courses, n (%)				.591
Yes	27 (64)	15 (68)	12 (60)	
No	15 (36)	7 (32)	8 (40)	
Q2. Experience with health education, n (%)				.977
Yes	23 (55)	12 (55)	11 (55)	
No	19 (45)	10 (45)	9 (45)	
Q3. Experience in cooking, n (%)				.505
Yes	39 (93)	21 (95)	18 (90)	
No	3 (7)	1 (5)	2 (10)	
Q4. Experience using nutrition-related apps, n (%)				.753
Yes	22 (52)	11 (50)	11 (55)	
No	20 (48)	11 (50)	9 (45)	

[a]Age and BMI data were analyzed with analysis of variance.

Overall Accuracy

The AIR and VIR groups achieved respective overall accuracy levels of 86% (189/220) and 68% (136/200) for all 17 food dishes. Within the food categories, the AIR group was significantly more accurate than the VIR group for the staple food ($p < .05$), main course ($p < .05$) and sides dish with three ingredients ($p < .05$) dishes. No significant differences were found for the other food categories (i.e., sides dish with one or two ingredients).

The AIR group achieved accuracy rates exceeding 95% for individual dishes including stir-fried noodle, streamed rice, fried rice, stir-fried bitter melon, stir-fried bitter melon with carrot, and stir-fried bean sprout with carrot and black fungus. The VIR group achieved accuracy rates exceeding 95% for stir-fried noodle, steamed rice, fried rice, and stir-fried bitter melon with carrot. The lowest accuracy for both groups was for stir-fried broccoli (less than 30% accuracy).

Table 2. Overall accuracy comparison in the AIR and VIR groups

Type	Dish	Result	Method AIR (n=22)	%	VIR (n=20)	%	p-value
	Overall	Correct	189	86%	136	68%	<0.001
		Incorrect	31	14%	64	32%	
Staple food	Overall	Correct	43	98%	32	80%	0.012
		Incorrect	1	2%	8	20%	
	Stir-fried noodle	Correct	20	95%	11	58%	0.007
		Incorrect	1	5%	8	42%	
	Steamed rice	Correct	19	100%	17	100%	--
		Incorrect	0	0%	0	0%	
	Fried rice	Correct	4	100%	4	100%	--
		Incorrect	0	0%	0	0%	
Main course	Overall	Correct	40	91%	29	73%	0.044
		Incorrect	4	9%	11	28%	
	Pan-fried chicken breast	Correct	24	92%	13	68%	0.055
		Incorrect	2	8%	6	32%	
	Braised pork chop	Correct	16	89%	16	76%	0.418
		Incorrect	2	11%	5	24%	
Side dish with 1 ingredient	Overall	Correct	31	70%	20	50%	0.055
		Incorrect	13	30%	20	50%	
	Stir-fried eggplant	Correct	14	82%	5	50%	0.102
		Incorrect	3	18%	5	50%	
	Stir-fried cauliflower	Correct	4	50%	6	46%	1.000
		Incorrect	4	50%	7	54%	
	Stir-fried broccoli	Correct	0	0%	1	25%	1.000
		Incorrect	4	100%	3	75%	
	Stir-fried bitter melon	Correct	2	100%	0	0%	0.333

		Incorrect	0	0%	2	100%	
Side dish with 2 ingredients	Overall	Correct	37	84%	30	75%	0.300
		Incorrect	7	16%	10	25%	
	Stir-fried egg with tomato	Correct	15	94%	14	93%	1.000
		Incorrect	1	6%	1	7%	
	Stir fried loofah with carrot	Correct	9	69%	8	67%	1.000
		Incorrect	4	31%	4	33%	
	Stir-fried pork with bell pepper	Correct	10	100%	5	50%	0.033
		Incorrect	0	0%	5	50%	
	Stir-fried bitter melon with carrot	Correct	3	60%	3	100%	0.464
		Incorrect	2	40%	0	0%	
Side dish with 3 ingredients	Overall	Correct	38	86%	25	63%	0.022
		Incorrect	6	14%	15	38%	
	Stir-fried Chinese cabbage with carrot and black fungus	Correct	18	90%	11	65%	0.109
		Incorrect	2	10%	6	35%	
	Stir-fried cauliflower with carrot and black fungus	Correct	16	80%	11	65%	0.460
		Incorrect	4	20%	6	35%	
	Stir-fried bean sprout with carrot and black fungus	Correct	4	100%	3	50%	0.200
		Incorrect	0	0%	3	50%	

Time Efficiency

Compared to the VIR group, the AIR group required significantly less time for task completion ($p < 0.001$), requiring around 2 to 25 seconds per task, with “steamed rice” being the fastest (mean 2.05, SD 1.43 s) and “stir-fried cauliflower with carrot and black fungus” the slowest (mean 24.90, SD 20.60 s). In the VIR group, the operation time ranged from around 12 to 22 seconds per task, with “stir-fried kelp” being the fastest (mean 12.76, SD 4.12 s) and “stir-fried cauliflower with carrot and black fungus” the slowest (mean 21.65, SD 5.25 s).

In the three food categories, i.e., the staple food, main course, and dishes with one ingredient, the

AIR group required significantly less time to the VIR group. Reporting performance for the two food categories did not differ significantly for dishes with two or three ingredients.
Within the six dishes in the categories of staple food and main course, except “braised pork chop”, the rest five dishes in the AIR required significantly less time to the VIR group. Within the five 1-ingredient dishes, the AIR group required significantly less time to report “stir-fried cauliflower” ($p < 0.001$) and was faster for “stir-fried bitter melon” ($p < 0.05$). For dishes with two ingredients the AIR group required significantly less time in “stir-fried egg with tomato” ($p < 0.05$). For three dishes with 3 ingredients, the two groups did not reveal significantly different.

Table 3. Dish reporting time in the AIR and VIR groups.

Dish	Reporting time (seconds), mean (SD)				p-value
	AIR (n=22)		VIR (n=20)		
Overall	12.43 (12.42)		16.25 (5.22)		.000
Staple food	44	5.61 (6.21)	40	14.53 (5.29)	.000
Stir-fried noodle	21	8.76 (7.42)	19	15.40 (5.75)	.003
Steamed rice	19	2.05 (1.43)	17	13.85 (5.37)	.000
Fried rice	4	6.00 (4.90)	4	13.50 (1.76)	.028
Main course	44	9.32 (6.87)	40	15.53 (4.19)	.000
Pan-fried chicken breast	26	6.27 (5.33)	19	16.37 (5.03)	.000
Braised pork chop	18	13.72 (6.56)	21	14.76 (3.19)	.524
Side dish with 1 ingredient	44	12.16 (9.55)	40	15.93(5.38)	.031
Stir-fried eggplant	17	15.06 (11.51)	10	20.40 (6.73)	.196
Stir-fried kelp	13	9.08 (6.03)	11	12.76 (4.12)	.101
Stir-fried cauliflower	8	8.00 (6.18)	13	14.54 (3.49)	.000
Stir-fried broccoli	4	20.33 (14.64)	4	16.65 (4.25)	.566
Stir-fried bitter melon	2	14.00 (1.41)	2	18.70 (0.42)	.046
Side dish with 2 ingredients	44	15.27 (13.81)	40	15.48 (3.58)	.927
Stir-fried egg with tomato	16	12.44 (4.49)	15	15.36 (3.21)	.047
Stir fried loofah with carrot	13	20.62 (22.68)	12	16.92 (3.63)	.583
Stir-fried pork with bell pepper	10	11.00 (4.52)	10	14.58 (3.91)	.074
Stir-fried bitter melon with carrot	5	19.00 (14.09)	3	13.33 (3.56)	.531
Side dish with 3 ingredients	44	19.80 (17.20)	40	19.76 (5.88)	.989
Stir-fried Chinese cabbage with carrot and black fungus	20	16.85 (13.35)	17	18.18 (5.52)	.705
Stir-fried cauliflower with carrot and black fungus	20	24.90 (20.60)	17	21.65 (5.25)	.531
Stir-fried bean sprout with carrot and black fungus	4	9.00 (6.22)	6	18.87 (7.91)	.070

System Usability Scale and Subjective Perception

Table 4 summarizes the SUS score and its two divisions in terms of usability and learnability. Overall scores showed no significant differences between the AIR and VIR groups, but both groups had the overall score at 84.72 and 83.00 respectively, indicating that the participants in both groups considered the app to be generally easy-to-use and easy-to-learn.

Table 4. System usability scale of the AIR and VIR groups

Score[a,b,c]	Assessment score, mean (SD)		*P*-*value*
	AIR (n=22)	VIR (n=20)	
Overall score	84.72 (12.66)	83.00 (11.95)	.655
Usability score	84.43 (11.34)	83.50 (9.30)	.774
Learnability score	85.00 (15.66)	82.50 (16.18)	.614

[a] Questionnaire was presented in Chinese.
[b] Mean scores for system usability with adjective ratings are as follows: 35.7 (“poor”), 50.9 (“ok”), 71.4 (“good”), and 85.5 (“excellent”).
[c] The questionnaire’s Cronbach α for AIR (α=0.91) and VIR (α=0.89) exceeded .70, indicating good internal consistency

Discussion

Principal Findings

While automatic image recognition is increasingly integrated into mobile meal reporting [28, 38,39], few comparative studies have examined how this technology creates additional value for existing apps. The results of the present study indicate that combining automatic image and voice recognition is not only feasible but also provides improvements over voice-only versions in terms of accuracy and time-efficiency. Our evidence-based findings provide insight for researchers or practitioners developing the next generation of dietary intake reporting applications, with implications as follows.

Accuracy in AIR Versus VIR

Correctly identifying and reporting the dish names and cooking methods is crucial to accurate daily intake management. Table 2 shows that user performance with AIR is significantly more accurate than VIR due to the integration of automatic image recognition. However, in the AIR group, the dishes that had been correctly recognized rate remain limited. Furthermore, the recognition accuracy rate of each dish for the food five categories fluctuated significantly within the AIR group. Image recognition errors in the trial may have resulted from the AI server failing to adequately recognize the dish or providing incorrect possible answers. In addition, errors could occur when participant required utilizing voice input to complete the reporting task when the image recognition did not work properly for some dishes. Failed image recognition could also occur due to poor image quality in the uploaded file, inappropriate lighting, the technical algorithm, food recognition technologies [31, 40], limited food datasets [25, 26], and the contexts of use [41].

Time Efficiency in AIR Versus VIR

Task completion for the AIR group was significantly faster than for the VIR group, suggesting that the integration of automatic image recognition can effectively reduce reporting time, though such performance improvements were inconsistent across food categories, and further enhancements are needed.

Participant Perception

Overall SUS results indicated participants found both apps to be easy to use. Regardless the differences on accuracy rate and time difference achieved by the two groups, the app in each of the groups demonstrated high overall SUS scores.

Limitations and Future Research

The experiments were conducted using a relatively small number of (disproportionately female) university students and a predetermined list of dishes. The sample size was restricted due to campus COVID-prevention measures at the time, but inclusion of 22 students in each group provided statistical power up to 74%. Generalizing the findings would require additional confirmation with a broader participant sample, using a wider range of food containers (e.g., bowls and plates). Future research should also use a broader range of authentic dishes and longer reporting periods.

Conclusion

Integrating AI image recognition in a voice-based meal reporting application was found to significantly improve reporting accuracy and time efficiency among young adult users. Further design improvements are required, as is testing in a broader range of authentic dining environments and a more varied array of food items.

Acknowledgements

The authors are thankful for all the research assistants at the Food Interaction Design Lab and the ABC Lab of the Chang Gung University. This research was funded by the Research Fund of Chang Gung Memorial Hospital and Chang Gung University (BMRPD 067) and the National Science and Technology Council, Taiwan (NARPD3N0121 and NSTC 110-2221-E-182-008-MY3).

References

1. Boutari C, Mantzoros CS. A 2022 update on the epidemiology of obesity and a call to action: as its twin COVID-19 pandemic appears to be receding, the obesity and dysmetabolism pandemic continues to rage on. Metabolism. 2022 Aug;133:155217.
2. Haththotuwa RN, Wijeyaratne CN, Senarath U. Worldwide epidemic of obesity. In: Obesity and Obstetrics. Elsevier; 2020. p. 3–8.
3. Chaudhary A, Sudzina F, Mikkelsen BE. Promoting healthy eating among young people-A review of the evidence of the impact of school-based interventions. Nutrients 2020;12(9).
4. Winpenny EM, van Sluijs EMF, White M, Klepp K-I, Wold B, Lien N. Changes in diet through adolescence and early adulthood: longitudinal trajectories and association with key life transitions. Int J Behav Nutr Phys Act. 2018;15(1).
5. Montero K, Kelly P. Young people and the aesthetics of health promotion: Beyond reason, rationality and risk. Milton Park, Abingdon, Oxon ; New York, NY : Routledge; 2016.
6. Wickman ME, Anderson NLR, Smith Greenberg C. The adolescent perception of invincibility and its influence on teen acceptance of health promotion strategies. J Pediatr Nurs 2008;23(6):460–8.
7. Mattei J, Alfonso C. Strategies for healthy eating promotion and behavioral change perceived as effective by nutrition professionals: A mixed-methods study. Front Nutr 2020;7:114.
8. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (integrate, design, assess, and share): A framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016;18(12):e317.
9. West JH, Hall PC, Hanson CL, Barnes MD, Giraud-Carrier C, Barrett J. There's an app for that: Content analysis of paid health and fitness apps. J Med Internet Res 2012;14(3):e72.
10. Chen Y, Perez-Cueto FJA, Giboreau A, Mavridis I, Hartwell H. The promotion of eating behaviour change through digital interventions. Int J Environ Res Public Health 2020;17(20).
11. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019;6(2):94–8.
12. Nikolaou CK, Tay Z, Leu J, Rebello SA, Te Morenga L, Van Dam RM, et al. Young people's attitudes and motivations toward social media and mobile apps for weight control: Mixed methods study. JMIR MHealth UHealth 2019;7(10):e11205.

13. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013;15(4):e86.
14. Bhatt P, Liu J, Gong Y, Wang J, Guo Y. Emerging Artificial intelligence–empowered mHealth: Scoping review. JMIR MHealth UHealth. 2022;10(6):e35053.
15. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. In: Artificial Intelligence in Healthcare. Elsevier; 2020. p. 25–60.
16. Torbjørnsen A, Ribu L, Rønnevig M, Grøttland A, Helseth S. Users' acceptability of a mobile application for persons with type 2 diabetes: a qualitative study. BMC Health Serv Res 2019;19(1).
17. Deniz-Garcia A, Fabelo H, Rodriguez-Almeida AJ, Zamora-Zamorano G, Castro-Fernandez M, Alberiche Ruano MDP, et al. Quality, usability, and effectiveness of mHealth apps and the role of artificial intelligence: Current scenario and challenges. J Med Internet Res 2023;25:e44030.
18. Krishnan G, Singh S, Pathania M, Gosavi S, Abhishek S, Parchani A, et al. Artificial intelligence in clinical medicine: catalyzing a sustainable global healthcare paradigm. Front Artif Intell 2023;6.
19. Alowais SA, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya AI, Almohareb SN, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Educ 2023;23(1).
20. Thomas LB, Mastorides SM, Viswanadhan NA, Jakey CE, Borkowski AA. Artificial intelligence: Review of current and future applications in medicine. Fed Pract 2021;38(11):527–38.
21. Oh S, Kim JH, Choi S-W, Lee HJ, Hong J, Kwon SH. Physician confidence in artificial intelligence: An online mobile survey. J Med Internet Res 2019;21(3):e12422.
22. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016;16(1).
23. Bergevi J, Andermo S, Woldamanuel Y, Johansson U-B, Hagströmer M, Rossen J. User perceptions of eHealth and mHealth services promoting physical activity and healthy diets: Systematic review. JMIR Hum Factors 2022;9(2):e34278.
24. Almiron-Roig E, Navas-Carretero S, Emery P, Martínez JA. Research into food portion size: methodological aspects and applications. Food Funct 2018;9(2):715–39.
25. Howes E, Boushey CJ, Kerr DA, Tomayko EJ, Cluskey M. Image-based dietary assessment ability of dietetics students and interns. Nutrients 2017;9(2):114.
26. Kong NA, Moy FM, Ong SH, Tahir GA, Loo CK. MyDietCam: Development and usability study of a food recognition integrated dietary monitoring smartphone application. Digit Health. 2023;9:205520762211493.
27. Kaushal S, Tammineni DK, Rana P, Sharma M, Sridhar K, Chen H-H. Computer vision and deep learning-based approaches for detection of food nutrients/nutrition: New insights and advances. Trends Food Sci Technol 2024;146(104408):104408.
28. Amugongo LM, Kriebitz A, Boch A, Lütge C. Mobile computer vision-based applications for food recognition and volume and calorific estimation: A systematic review. Healthcare 2022;11(1):59.
29. Van Asbroeck S, Matthys C. Use of different food image recognition platforms in dietary assessment: Comparison study. JMIR Form Res 2020;4(12):e15602.
30. Rantala E, Balatsas-Lekkas A, Sozer N, Pennanen K. Overview of objective measurement technologies for nutrition research, food-related consumer and marketing research. Trends Food Sci Technol 2022;125:100–
31. Liu Y-C, Onthoni DD, Mohapatra S, Irianti D, Sahoo PK. Deep-learning-assisted multi-dish food recognition application for dietary intake reporting. Electronics 2022;11(10):1626.
32. Liu Y-C, Chen C-H, Lin Y-S, Chen H-Y, Irianti D, Jen T-N, et al. Design and usability evaluation of mobile voice-added food reporting for elderly people: Randomized controlled trial. JMIR MHealth UHealth 2020;8(9):e20317.
33. Liu Y-C, Wu S-T, Lin S-J, Chen C-H, Lin Y-S, Chen H-Y. Usability of food size aids in mobile dietary reporting apps for young adults: Randomized controlled trial. JMIR MHealth UHealth. 2020;8(4):e14543.
34. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251.
35. Liu YC, Chen CH, Tsou YC, Lin YS, Chen HY, Yeh JY, Chiu SY. Evaluating Mobile Health Apps for Customized Dietary Recording for Young Adults and Seniors: Randomized Controlled Trial. JMIR mHealth and uHealth, 2019 Feb; 7(2), e10931.
36. Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry (pp.189-194)Publisher: Taylor & FrancisEditors. :189–94.
37. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. Journal of Usability Studies. 2009 May 1;4(3):114–23.
38. Hussain G, Maheshwari MK, Memon ML, Jabbar MS, Javed K. A CNN based automated activity and food recognition using wearable sensor for preventive healthcare. Electronics 2019;8(12):1425.
39. Fakhrou A, Kunhoth J. Al Maadeed, S. Smartphone-based food recognition system using multiple deep CNN models. Multimed Tools Appl. 2021;80:33011–32.
40. Boushey CJ, Spoden M, Zhu FM, Delp EJ, Kerr DA. New mobile methods for dietary assessment: review of

image-assisted and image-based dietary assessment methods. Proc Nutr Soc 2017;76(3):283–94.

41. Papathanail I, Abdur Rahman L, Brigato L, Bez NS, Vasiloglou MF, van der Horst K, Mougiakakou S. The Nutritional Content of Meal Images in Free-Living Conditions-Automatic Assessment with goFOODTM. Nutrients. 2023 Sep 2;15(17):3835. doi: 10.3390/nu15173835. PMID: 37686866; PMCID: PMC10490087.