electronics

Article

Deep-Learning-Assisted Multi-Dish Food Recognition Application for Dietary Intake Reporting

Ying-Chieh Liu [1,2,3], Djeane Debora Onthoni [4], Sulagna Mohapatra [4], Denisa Irianti [1] and Prasan Kumar Sahoo [4,5,*]

[1] Department of Industrial Design, Chang Gung University, Guishan, Taoyuan 33302, Taiwan; ycl30@mail.cgu.edu.tw (Y.-C.L.); leonna.denisa@gmail.com (D.I.)
[2] Department of Industrial Design, College of Management and Design, Ming-Chi University of Technology, Taishan, New Taipei City 24301, Taiwan
[3] Department of Internal Medicine, Health Promotion Center, Chang Gung Memorial Hospital, Taoyuan 33302, Taiwan
[4] Department of Computer Science and Information Engineering, Chang Gung University, Guishan, Taoyuan 33302, Taiwan; d0421008@cgu.edu.tw (D.D.O.); d0521007@cgu.edu.tw (S.M.)
[5] Department of Neurology, Chang Gung Memorial Hospital, Linkou, Guishan Dist., Taoyuan 333423, Taiwan
* Correspondence: pksahoo@mail.cgu.edu.tw

Abstract: Artificial intelligence (AI) is among the major emerging research areas and industrial application fields. An important area of its application is in the preventive healthcare domain, in which appropriate dietary intake reporting is critical in assessing nutrient content. The traditional dietary assessment is cumbersome in terms of dish accuracy and time-consuming. The recent technology in computer vision with automatic recognition of dishes has the potential to support better dietary assessment. However, due to the wide variety of available foods, especially local dishes, improvements in food recognition are needed. In this research, we proposed an AI-based multiple-dish food recognition model using the EfficientDet deep learning (DL) model. The designed model was developed taking into consideration three types of meals, namely single-dish, mixed-dish, and multiple-dish, from local Taiwanese cuisine. The results demonstrate high mean average precision (mAP) = 0.92 considering 87 types of dishes. With high recognition performance, the proposed model has the potential for a promising solution to enhancing dish reporting. Our future work includes further improving the performance of the algorithms and integrating our system into a real-world mobile and cloud-computing-based system to enhance the accuracy of current dietary intake reporting tasks.

Keywords: EfficientDet; dietary assessment; multiple-dish; food image recognition; mHealth; deep learning; artificial intelligence

check for updates

Citation: Liu, Y.-C.; Onthoni, D.D.; Mohapatra, S.; Irianti, D.; Sahoo, P.K. Deep-Learning-Assisted Multi-Dish Food Recognition Application for Dietary Intake Reporting. *Electronics* **2022**, *11*, 1626. https://doi.org/10.3390/electronics11101626

Academic Editor: George A. Papakostas

Received: 23 April 2022
Accepted: 15 May 2022
Published: 19 May 2022

1. Introduction

By 2025, the increased healthcare costs from chronic diseases will become a significant financial burden [1–3]. Maintaining healthy eating is a crucial behavior change strategy [4,5] and is commonly associated with treating nutrition-related chronic diseases, such as obesity and diabetes [6]. Applying diet and nutrition information for the prevention of chronic diseases by the use of smart devices is becoming an effective approach to ameliorating the negative effects of chronic diseases [7–9]. Furthermore, self-management in mobile health (mHealth) applications have the potential to deliver effective and scalable dietary interventions at a low cost [7,8,10–12]. However, among the many challenges to food intake reporting via mobile devices are the high errors that have been addressed in many studies that are subject to each individual's memory and ability in reporting food ingredients and the estimation of food size [13,14]. Therefore, the burden on participants' intake reporting needs to be reduced to a large extent.

Recently, artificial intelligence (AI) techniques have been increasingly applied to food and nutrition applications [15,16]. With the increased use of smart devices and the rise of deep learning (DL) techniques [17,18], the opportunity of building AI-based food intake reporting is becoming possible, involving a number of integrated issues, e.g., mobile or wearable devices [19,20], food recognition technologies [21], food datasets [22,23], and the contexts of use [24]. Food-recognition-related techniques have been developed for supporting dietary assessment in identifying food items, food size, and features [24,25]. AI technology raises the possibility of facilitating tasks in recognition performance in food images [24]. However, major challenges remain unsolved.

Challenge 1: Low accuracy in recognition of local dishes. Research [25] compared the recognition accuracy of ten commercial paid platforms, and the varying results between different platforms demonstrate a wide range of differences from poor to excellent under some realistic settings. Furthermore, many local dishes were often shown to have less accurate results. For a wide variety of food types, such as local Asian cuisine, improvements in local dish recognition are required.

Challenge 2: Limited work in recognition of multiple dishes. Existing works focus on the images of a single dish with a single or mixed-dish ingredient. However, recognition of multiple dishes is important in health-related application for the following reasons. Using a dish-based service will require users to upload each dish photo of a meal which can be time-consuming and less efficient. Further, food reporting is frequent and needs to be done more than one time per day. A single-dish-based rather than multiple-dish-based upload would pose issues in the usability of the health application [26].

Challenge 3: Recognizing the same food from a variety of appearances. Food items suffer from inter- and intraclass variations. Interclass variation can be found when similar foods (e.g., stir-fried cabbage and stir-fried cabbage carrot) look similar in terms of intensity in single and mixed dishes. Intraclass variation can be found when the appearances of a particular food item are different due to factors in image angles (top view or side view), color intensity (caused by methods of cooking), occlusions from other food items, types of background (e.g., bowls or plates), etc., [27].

Challenge 4: Multiple dishes in a set-meal dish menu are changing and different. Thus, a set meal with a single dish and mixed dishes served on a single tray can have many possibilities. In addition, some dishes are not always available due to the seasons. New combinations of a set meal can cause frequent updates of the food dataset [28].

Goals

The overall research purpose was to improve the effectiveness of AI meal-based applications for use in food and nutrition services. We aimed to integrate ML innovations of a realistic mobile health application [29] using mobile ICT and AI technology to allow people to easily and accurately report their dietary intake under real conditions. To address the aforementioned challenges in multiple-dish recognition, the low accuracy of some Asian local foods, and the difficulties in recognizing mixed dishes, single dishes, and multiple dishes, we proposed AI-enhanced multiple-dish food recognition using the EfficientDet-D1 deep learning (DL) model. The goals of our work can be listed as follows:

- Deep-learning-assisted multiple-dish food recognition model was developed that can recognize food items automatically.
- The proposed DL model can work robustly in recognizing the single dishes, mixed dishes, and multiple dishes of local Taiwan cuisines, which could be helpful for deciding the appropriate healthy dietary intake.
- The performance evaluation and comparison with existing state-of-the-art recognition models were conducted.

In this paper, though a multi-dish food recognition model was designed, a hypothetical question could be asked: “How might we be able to efficiently recognize the multiple dishes of local foods?” We aimed to resolve the aforementioned challenges, i.e., Challenges 1 and 2, by developing an AI-based multiple-dish food recognition model using the EfficientDet deep learning (DL) model to improve the accuracy of the local dish recognition. This paper is organized as follows. Section 2 describes related works; Section 3 elucidates the methodology of our proposed food recognition model using DL that includes data collection, training, and evaluation procedures. Performance evaluation, results discussion, and conclusions are given in Section 4, Section 5, and Section 6, respectively.

2. Related Works

Applying convolutional neural network (CNN) with different approaches has been a main trend for dietary assessment. For example, the GoogLeNet CNN model was applied to develop a mobile app called "Im2Calories" [30]. The process of dish recognition includes food segmentation steps. Firstly, a Food101-Background dataset was used to determine the presence of a food object on the food image. After food objects were identified, a multi-stage classification process for image segmentation was implemented. Later, the system used semantic image segmentation to produce correct labels on every food object on a plate. The experiment trained and tested the obtained dataset of 2517 restaurants and used fixed-class food image datasets such as Food-101, Food-201, Gfood-3D, and Nfood-3D. By using over 250,000 images to train the classifier on multiple datasets, the system was comprehensive in detecting a large number of food categories with an accuracy of up to 76%. Moreover, the model was especially suited for the fine-grain distinctions between categories such as different hamburgers. This approach can be implemented to solve the challenges of recognizing similar types of food with various appearances.

CNN-based Caffe framework [31] was applied on the proposed multiple-dish recognition model which was later implemented on a client–server architecture [27]. The system involves users' input to manually draw a bounding circle on a food item, which is time-consuming. Therefore, developing an automatic food recognition model to draw bounding boxes on each food item is necessary to reduce the food reporting time, which can enhance the user's experience.

Later, region mining algorithms were used to identify each food in a multiple-dish setting. These food recognition processes were performed in the cloud which facilitates the real-time integration between the mobile app and AI server. In the testing environment, the system used the FooDD dataset [28] with an achieved accuracy of up to 94% for a 30-class dataset.

Initially, wearable devices have been utilized to recognize multiple foods through physical activity such as eating [19,20]. In [19], spectrogram images were generated and fed into pre-trained CNN AlexNet DL architecture. The adopted architecture was used to develop a classification model that can recognize six food categories. For the same purpose, a simple drink-and-eat classification model using CNN on raw accelerometer data has been developed [20]. Using the classification approach, food recognition on ultrasonic Doppler signal has been developed [21]. The proposed smartphone app was developed using CNN and 30 food categories, which are mainly single-dish. However, Taiwanese local dishes contain several mixed dishes. Based on our knowledge, most of the existing works have only considered a small number of food categories. Meanwhile, some of the existing works have involved a large number of food categories, which were from open datasets such as Food-101, UECFood100, EUCFood256, etc., [22,23].

Another CNN-based approach, named faster region-based CNN (Faster R-CNN), was applied—a three-step algorithm to recognize multiple dishes [32]. These steps included region-based detection, food item detection, and food item classification. A diverse dataset has been tested by combining Asian-style cuisine food image datasets such as UEC-FOOD100 and UEC-FOOD256 and Western-style cuisines such as FOOD101. The achieved accuracy was up to 71% for a 100-class dataset. Although the food image datasets are of Asian cuisines, they do not cover the Taiwanese local dishes such as pork belly with soy sauce, sweet potato leaves, etc. Similarly, implemented in smartphone devices, food recognition in classification [22] and food types [23] has been proposed using FOOD101, UEC-FOOD100, and UEC-FOOD256. The food classification model using a CNN model had an accuracy of 94% with a total of 29 food types comprising general food and fruits [22]. In a food detection model, You Only Look Once (YOLO) has been adopted using 256 food types [23]. The food detection model achieved 76.36% mAP with an inference time of 15 ms.

The applications of CNN came out with various models. However, those studies were performed in an experiment setting with a fixed-class large food image dataset. It might not be suitable for the application in a real environment with frequent menu changes in a restaurant. Therefore, in this study, we used a parsing approach to enable automatic recognition of food items with multiple dishes. In addition, the fixed-class large food image dataset had only a few template images per dish. Some studies [33–36] applied the k-nearest neighbor (KNN) classifier as an alternative solution. For instance, a hierarchical fine-grained model was used to recognize buffet-style multiple dishes [37]. The JISS-22 dataset was used and obtained from one specific restaurant. The achieved accuracy was about 78% for a 50-class dataset.

The recent implementations of AI-based services created possibilities to support the preventive healthcare system. First, the main trend of self-management of chronic diseases or weight control was closely related to healthy eating that accounts for daily goal management of a targeted calorie and balanced intake of macronutrients in protein, fat, carbohydrates, etc. For example, in a real-time example, AVA (eatwithava.com; accessed date: 14 February 2022) and Calorie Mama (www.caloriemama.ai; accessed date: 14 February 2022) provided automatic calorie calculation based on AI-based recognition of uploaded food photos. Furthermore, "Foodvisor" served to recognize food items, estimate the serving sizes, and monitor users' eating habits [38]. "SnapCalorie" used Google Lens and Cloud Vision API to recognize the food images and automatically calculate calories, fat, carbs, and protein [39].

In [40], dietary intake reporting applications/systems involving food image recognition could be built into either an automated or semi-automated approach. Most recent research [27,30,41] built systems using an automated approach. However, a comparative performance study [25] performed an experiment with samples of plain foods, processed foods, drinks, and mixed dishes in a realistic setting using different containers, lighting, and angles of perspective. The varying results between different platforms demonstrate a wide range of differences from poor to excellent showing that applications following an automated approach have yet to be appropriately utilized in a real environment. The contributions of our work with respect to some of the related works are summarized in Table 1.

Table 1. Summary of related works' contributions.

Ref #	Food Type/Task	Contribution	# of Considered Food Items	Performance
[22]	Single-dish and fruits/ detection	Develop application for children with visual impairments	29	0.95
[23]	Mixed-dish/ detection	Display food nutrition factors	356	0.75
[30]	Mixed-dish/ segmentation	Estimate food size	201	0.25
[37]	Multiple-dish/ detection	Develop application for buffet-style food	50	0.78
Our work	Single/mixed/ multiple-dish/set meal detection	Improve dietary intake reporting	87	0.92

3. Methodology

The designed multiple-dish food recognition model is detailed and explained in this section. Three types of meals—single-dish, mixed-dish, and multiple-dish—were considered as input. These inputs were divided, labeled, trained, evaluated, and tested. The overall workflow of the proposed model is shown in Figure 1.

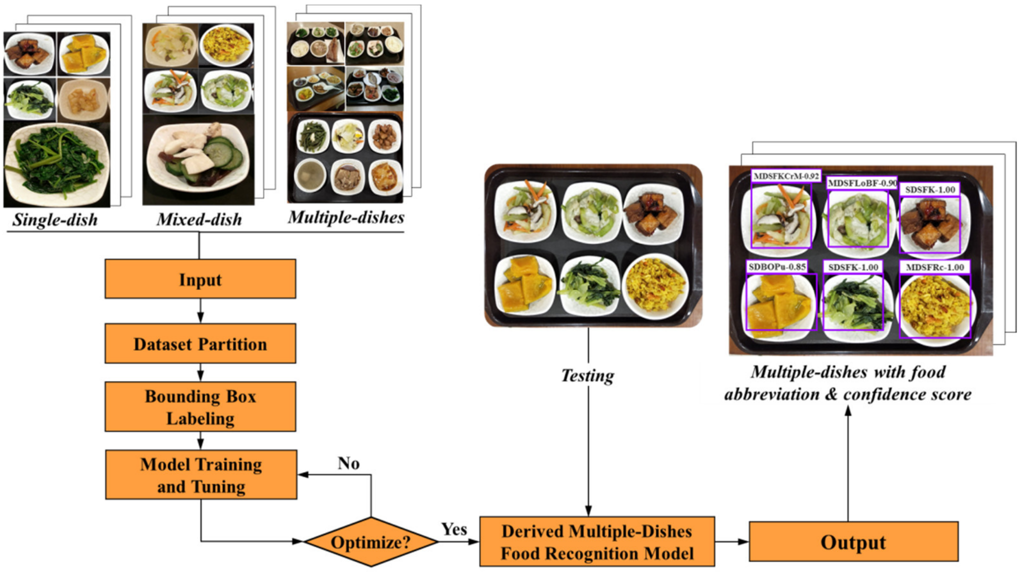

Figure 1. Proposed multiple-dish food recognition model workflow.

3.1. Data Acquisition and Labeling

For this study, the dataset of food images was taken with smartphone and collected in a real environment, i.e., at the cafeteria of Chang Gung Health and Culture Village. Following our previous study [29], we categorized the meals into mixed-dish, single-dish, and multiple-dish, as shown in Figure 2. Accordingly, we collected 31 mixed-dish food items, 27 sets of multiple-dish, and 56 single-dish food items with the corresponding images such as 447, 556, and 1326 images, respectively. We cropped manually each food item in each multiple-dish image and combined them with the set of mixed-dish and single-dish sets. Each multiple-dish meal contains 4 to 6 dishes, which are a random combination of single and/or mixed dishes. Image cropping on set meals was conducted to obtain additional single-dish and mixed-dish images. In addition, blurred or low-quality images were excluded. After cropping, the total number of images for mixed-dish and single-dish became 847 and 1557, respectively. Finally, a total number of 4733 food images were collected, which contained 87 food items, as shown in Figure 2.

For model training and testing purposes, we divided the total food image dataset into 80:20 ratios. Accordingly, we had 3786 and 947 images for training and testing, respectively. The data did not overlap between the training and testing sets. Each food image on our dataset was labeled following our previous study [29] which described a food category i.e., single- or mixed-dish, with cooking method i.e., stir-fried, followed by food ingredient name. For instance, “pan-fried pork” is in the category of “single-dish”, the cooking method of “pan-fried”, and the food ingredient of “pork”. Next, we applied abbreviation for each food image with format food category followed by cooking method and food ingredient name. For instance, “Single-Dish Boiled Okra” was abbreviated as “Single-Dish” to “SD”, “Boiled” to “BO”, and “Okra” to “Ok” which results in “SDBOOk”, and “Mixed-Dish Stir Fried Cabbage with Carrot” was abbreviated as “Mixed-Dish” to “MD”, “Stir Fried” to “SF”, and “Cabbage with Carrot” to “CCr” which results in “MDSFCCr”. A summary of types of dishes is given in Table 2. For detection purposes, we labeled each food item using bounding box techniques. We drew the coordinates of top left and bottom right using open-source software LabelImg (v1.8.4) and annotated them with the defined class abbreviations.

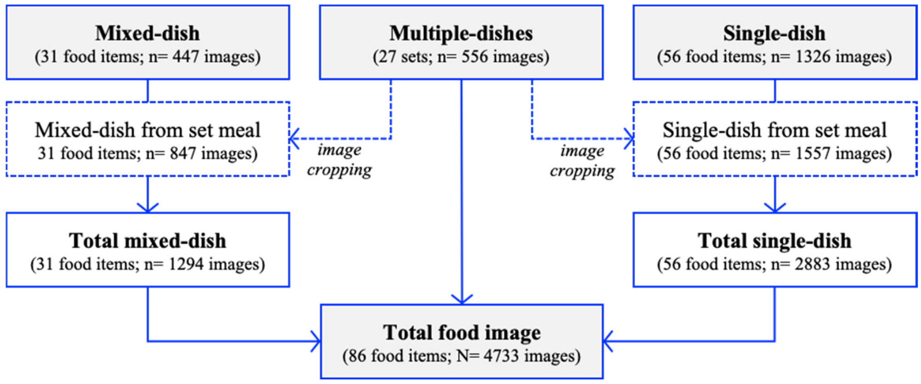
Mixed-dish
(31 food items; n= 447 images)
Multiple-dishes
(27 sets; n= 556 images)
Single-dish
(56 food items; n= 1326 images)
Mixed-dish from set meal
31 food items; n= 847 images)
image cropping
image cropping
Single-dish from set meal
(56 food items; n= 1557 images)
Total mixed-dish
(31 food items; n= 1294 images)
Total single-dish
(56 food items; n= 2883 images)
Total food image
(86 food items; N= 4733 images)

Figure 2. The collection process of data samples using local food image dataset.

Table 2. Type of dish information summary.

Food Type	Cooking Type	Abbreviations	Total Items
Single-dish	Stir-Fried	SDSF	24
	Boiled	SDBO	4
	Braised	SDBR	16
	Steamed	SDST	4
	Pan-Fried	SDPF	8
Mixed-dish	Stir-Fried	MDSF	20
	Braised	MDBO	7
	Boiled	MDBO	4

3.2. *Deep-Learning-Based Food Recognition Model*

To design AI-based multiple-dish food recognition model that can work in a real-time environment, a model is required to work fast with better accuracy of result under the circumstance of heterogeneous data found in single dishes, mixed dishes, and multiple dishes. Therefore, we adopted EfficientDet-D1 with EfficientNet-B1 as the backbone. EfficientDet detector architecture with EfficientNet was selected because previous research has shown better performance in comparison to other state-of-the-art object detection architectures [42,43]. Essentially, adopted EfficientDet comprised three steps in feature extraction, bidirectional feature pyramid network (BiFPN), and detection classification heads. For feature extraction, we adopted EfficientNet-B1 network from pre-trained ImageNet. The backbone comprised seven blocks {p1, p2, p3, p4, p5, p6, and p7}. Table 3 shows the details of each block.

Table 3. EfficientNet-B1 block details.

Block #	Input Filter	Kernel Size	Stride #	# of Repetitions	Output Filter
p1	32	3	1	1	16
p2	16	3	2	2	24
p3	24	5	2	2	40
p4	40	3	2	3	80
p5	80	5	1	3	112
p6	112	5	2	4	192
p7	192	3	1	1	320

After extracting the features from seven different scales, those features were passed to BiFPN blocks for the scaling. For backbones B1, with the scaling configurations including $\varnothing = 1$, the configurations were input size = 640, number of channels = 88 and number of layers = 4 for BiFPN, and number of layers for box or class = 3, where $\varnothing$ was compounded coefficient. In formal style, the BiFPN new width scale and depth were calculated as given in Equations (1) and (2).

$$\text{New width scale } = 88.\left(1.35^{\varnothing}\right), \tag{1}$$

$$\text{New depth scale } = 4 + \varnothing, \tag{2}$$

Moreover, for the box or class prediction and input resolution, the formal way of scaling was defined as given in Equations (3) and (4).

$$\text{New box or class prediction scale } = 3 + \varnothing/3, \tag{3}$$

$$\text{New input resolution } = 640 + \varnothing.128, \tag{4}$$

3.3. Procedures of DL Model

Taking the training set with a total of 3786 images, we designed the training and tuning model using *k*-fold cross-validation which was used in [44]. The training dataset was divided approximately equal to *k* = 5 subsets, {s1, s2, . . . , s*k*}. In each round of training, *k-1* total number of subsets was selected as a training set, where a subset s*k* was selected as testing. The tuning was performed in each round of training. After finding the optimized model in each round of training, we verified the robustness of the designed model using the testing set. The optimized model was achieved by finding the optimal values of image dimension = 640 × 640, total number of epochs = 300, total number of steps = 867, learning rate = 0.001, and batch = 20. Moreover, the final optimized model was used further for testing. The 947 images of testing set were used to evaluate the robustness of our derived model.

4. Experimental Results

The training, tuning, and testing were experimented on Ubuntu v18.04.3. For the multiple-dish food recognition model, we used TensorFlow framework v1.14. The experiments were performed by using Python programming language along with some Python libraries such as Pandas v0.24.2, OpenCV v4.5.1.48, Numpy v1.16.2, etc. All of these configurations were operated on GPU TITAN RTX 24 GB × 4 with 256 GB memory.

4.1. Evaluation Metrics

In this section, the performance of our adopted AI-based food recognition model is presented by comparing it with two existing recognition architectures, i.e., single-shot detector (SSD) one-stage detector with Inception V2 and Faster R-CNN two-stage detector with Inception ResNet V2 backbones. For the real-time experiment, SSD was reported to be powerful and outperformed You Only Look Once (YOLO) V4 tiny architecture in detecting green and reddish tomatoes [45]. Moreover, Faster R-CNN had better performance in comparison to SSD [46]. Therefore, these two architectures were selected for comparison with our proposed model. The factors for performance comparison were chosen in terms of accuracy, precision, recall, F1-score, and then mAP. The comparison was performed for all rounds of training. Thus, the standard deviation (SD) values were also derived. Furthermore, for a more in-depth analysis, a precision–recall curve of our model was drawn. Additionally, we analyzed the AP value considering different types of dishes, localization loss, and inference time.

To measure the performance of the proposed model, we used the intersection over union (IoU) metric. This metric is a commonly used metric for object detection or recognition models. Based on this metric, three parameters could be derived: true positive (TP), false positive (FP), and false negative (FN). Based on these parameters, we calculated four basic metrics: accuracy, precision, recall, and F1-score. The formal calculation of these metrics is addressed in Equations (5)–(8).

$$Accuracy = \frac{TP}{TP + FP + FN} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{8}$$

Finally, we calculated the average precision (AP) for all 87 classes and then derived the mean AP (mAP) as the overall performance which can be also plotted in a precision–recall curve by taking the IoU threshold = 0.5. As shown in Equation (9), we analyzed the localization loss using focal loss (*FocalL*), where p_t is the probability of predicted class and optimal value of gamma γ = 1.5. In addition, the inference time of the designed model is observed.

$$FocalL = -(1 - p_t)^{\gamma} \log(p_t) \tag{9}$$

4.2. Results of Performance Metrics

In Table 4, our model in the defined configuration of hyper-parameters has the accuracy and precision achieved > 0.80, whereas SSD Inception V2 and Faster R-CNN Inception ResNet V2 could only achieve maximal performance < 0.60. The result in the table shows that our model detected dishes more accurately than the other two under the 87 classes. This result is also supported by 0.97 of recall and 0.93 of F1-score, which were higher than other object detectors of SSD and Faster R-NN. In addition, the average of SD performance of the model from all the metrics was 0.01 in comparison to SSD Inception V2 and Faster R-CNN Inception ResNet V2, which were 0.035 and 0.05, respectively.

Table 4. Comparison in terms of accuracy, precision, recall, and F1-score metrics.

Architecture	Evaluation Metrics ± SD			
	Accuracy	Precision	Recall	F1-Score
Our Model	0.87 (±0.01)	0.88 (±0.01)	0.97 (±0.01)	0.93 (±0.01)
SSD Inception V2	0.53 (±0.03)	0.6 (±0.03)	0.56 (±0.04)	0.56 (±0.04)
Faster R-CNN Inception ResNet V2	0.54 (±0.02)	0.64 (±0.04)	0.53 (±0.09)	0.57(±0.06)

4.3. Results of Mean Average Precision

To evaluate how precisely the predicted bounding box located and classified the dishes in comparison to ground truth, we further compared the model by using mAP for all rounds of training. As shown in Table 5, our model recognized the food items with an mAP = 0.90 and lower value of SD = 0.01 from all the rounds. As shown in Table 5, overall mAP performance for all rounds was > 0.80 which was the same performance we observed in accuracy, precision, recall, and F1-Score. SSD had the lowest performance with mAP = 0.57 followed by Faster R-CNN with mAP = 0.64. It was observed that SSD and Faster R-CNN failed in detecting the multiple dishes.

Table 5. Comparison of mAP metrics.

Architectures	R1	R2	R3	R4	R5	mAP	±SD
Our Model	0.88	0.90	0.91	0.91	0.89	0.90	±0.01
SSD Inception V2	0.63	0.55	0.56	0.59	0.53	0.577	±0.04
Faster R-CNN Inception ResNet V2	0.51	0.75	0.77	0.63	0.55	0.646	±0.1

For a more in-depth analysis, we also used precision–recall curve, AP, and localization loss metrics on the model after all rounds. Figure 3 depicts the precision–recall curve of all 87 classes. The curve shows the trade-off between precision and recall for different confidence values predicted by our designed model. As a result, the model had a higher precision of up to 0.7 of recall which gradually decreased with the increasing value of recall.

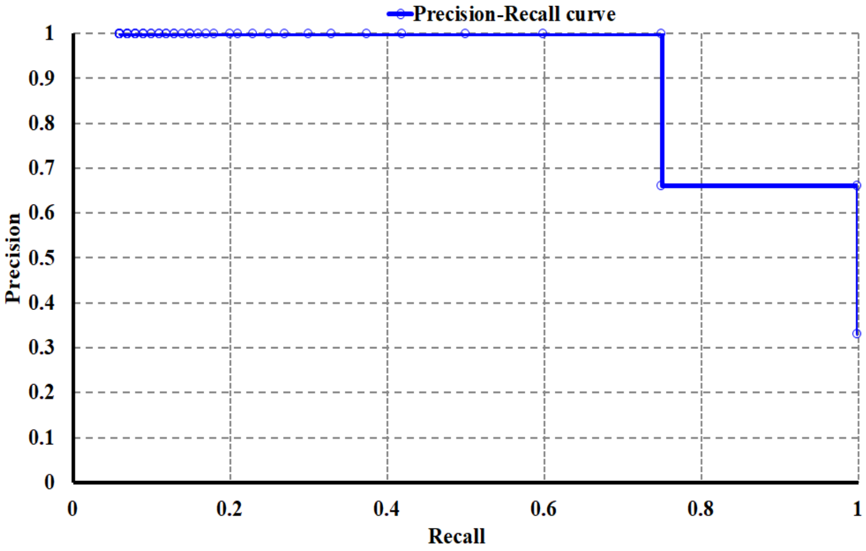

Figure 3. Precision–recall curve of our designed model.

Based on the precision–recall curve, we tried to estimate the area under the curve (AUC) using AP. The AP was measured for all 87 classes. Due to the large number of classes, we calculated the average AP for all types of dishes, as shown in Table 6. Due to the high number of classes or number of food items, the single-dish category still had high AP = 0.88 for all 56 classes, whereas mixed-dish had much better AP = 0.96. Moreover, the localization loss of the designed model was calculated. As shown in Figure 4, our model loss calculation was the lowest <0.1 in comparison to SSD Inception V2 and Faster R-CNN Inception ResNet V2.

Table 6. AP value of all types of dishes.

Type of Dishes	Total Number of Food Items	Avg AP	mAP
Single-dish + included in multiple-dish	56	0.88	0.92
Mixed-dish + included in multiple-dish	31	0.96	

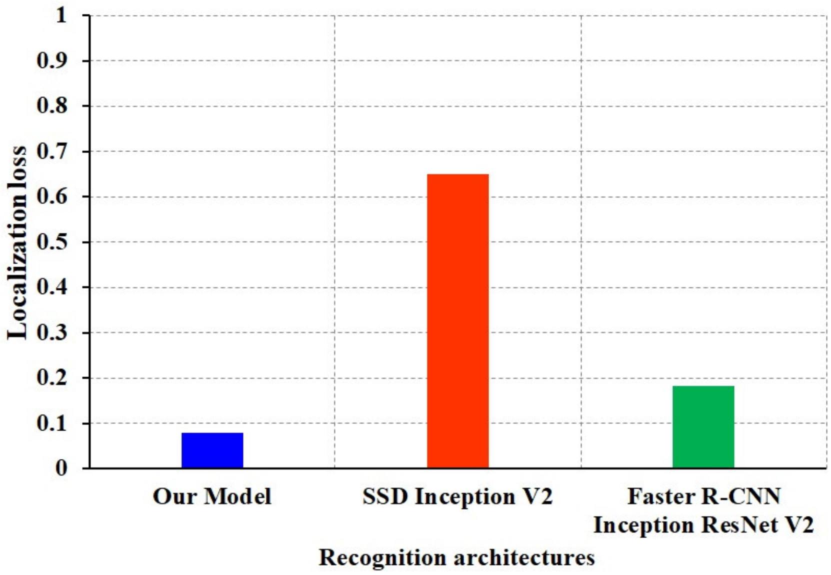

Figure 4. Localization loss comparison.

4.4. Results on Model Speed

To implement the function of food recognition, e.g., using a front-end mobile app and a back-end server, the derived model had to be of better time efficiency for a real-time application. As shown in Figure 5, the model had the lowest value of inference time, i.e., 4 s/image. The value in SSD was =5 s/image, and in Faster R-CNN, it was =21 s/image. With this relatively low inference time, EfficientDet could be more efficient to be implemented in a real-time mobile application.

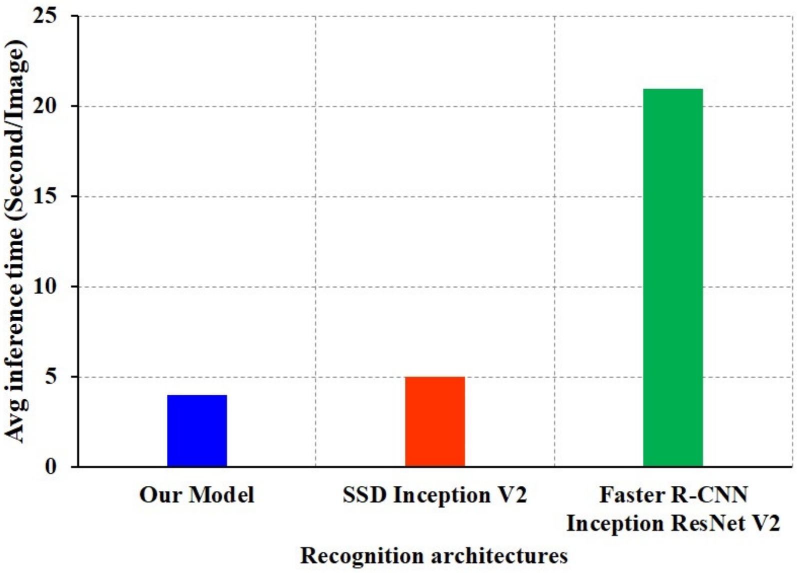

Figure 5. Inference time comparison.

4.5. Results on Different Dataset

In order to justify the robustness of our proposed detection model, we analyzed the performance by using the open dataset. Since we considered the local Taiwanese cuisine for detection, local Indian cuisine [47] was considered to implement and compare our model. We implemented our food recognition model on the single-dish images obtained from the Kaggle dataset [48]. This dataset comprises 10 food items with 100 numbers of images. Those 10 items are Rasgulla, Modak, Misi_roti, Kalakand, Imarti, Gulab_jamun, Chikki, Bandar_laddu, Aloo_tikki, and Adhirasam. Training and testing data division followed the same procedure as described in Section 3.1. As shown in Figure 6, it is observed that the performance of our model using our dataset has similar performance with the open dataset. Based on the test set, our model implemented on the images of the open dataset can achieve an accuracy (ACC) = 0.8, precision (PRE) = 0.84, recall (REC) = 0.941, F1-score (F1-S) = 0.88, and mAP = 0.8. These results prove that our model can be implemented to recognize a variety of single-dish images of other countries' cuisines.

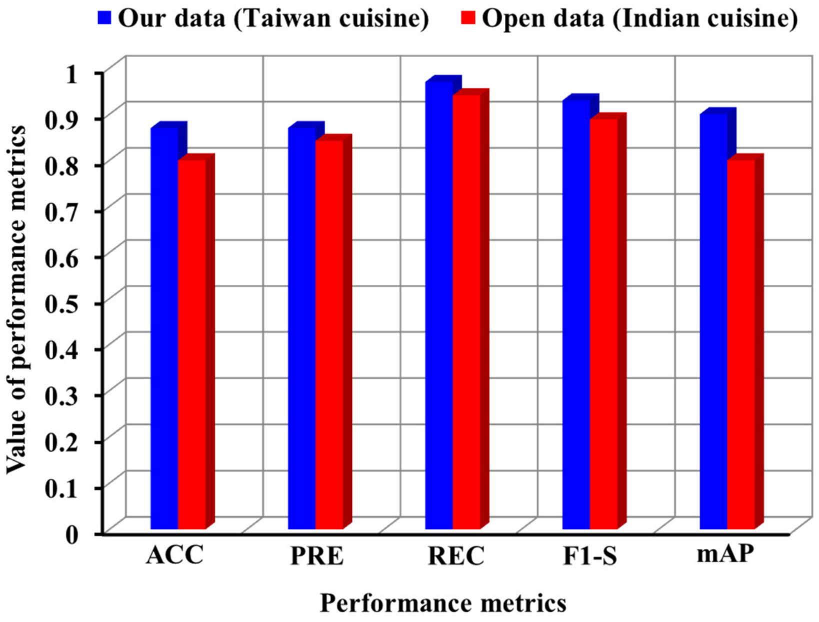

Figure 6. Comparison of performance metrics between our data with open dataset.

Considering the testing set of images from the open dataset, detection of some of the single-dish food types of Indian cuisine is shown in Figure 7. It is observed that our food detection model can accurately recognize the single dishes Imarti, Chikki, and Rasgulla with a higher confidence score.

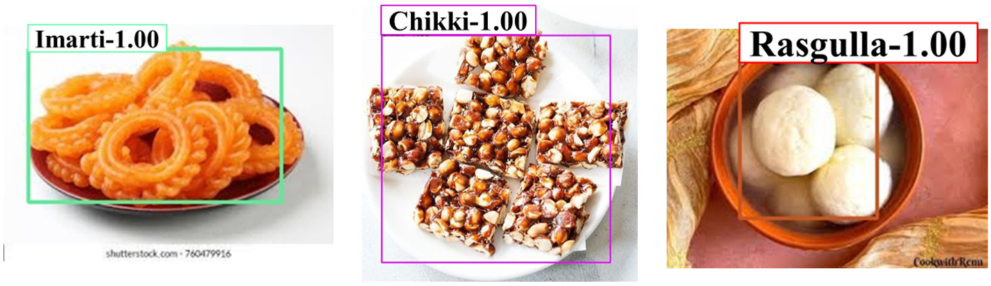

Figure 7. Recognition results on Indian cuisine open data.

4.6. DL-Based Food Recognition

Based on the model, the predicted results are presented in the testing set. The results are presented in three different dish types, i.e., single-dish, mixed-dish, and multiple-dish meals as shown in Figure 8a–c, respectively. Each dish in Figure 8 has a bounding box with a colored square and the top-left label on the box describing the abbreviation of the dish name with a specific confidence score ranging from 0.00 to 1.00.

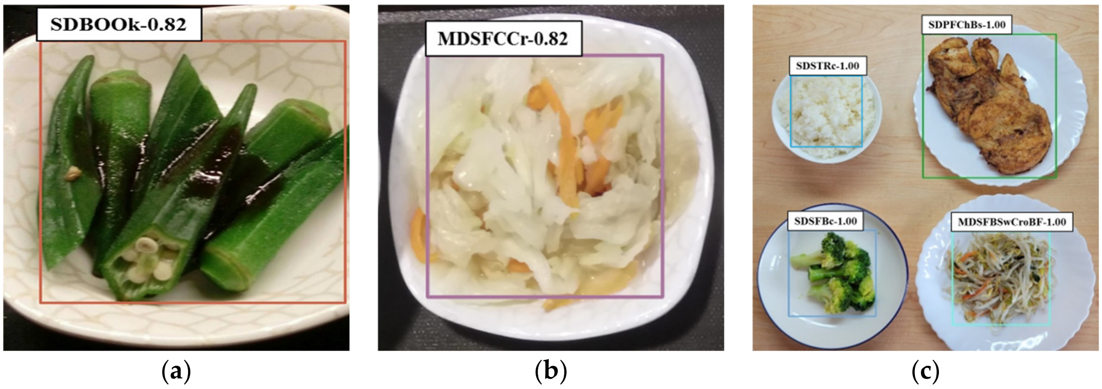

Figure 8. Recognition results of the three dish types: (**a**) single-dish; (**b**) mixed-dish; (**c**) multiple-dish.

Figure 9 shows three different food items in the type of single-dish. The model is able to recognize correctly the dish name with the information on the type of dish, e.g., single-dish, way of cooking, e.g., stir-fried, and food item, e.g., kelp buds. For instance, "SDSFKBu" was "Single Dish Stir Fried Kelp Buds", "SDBRPTss" was "Single Dish Braised Pig Trotters with Soy Sauce", and "SDSTRc" was "Single Dish Steamed Rice". The confidence score was 1.00.

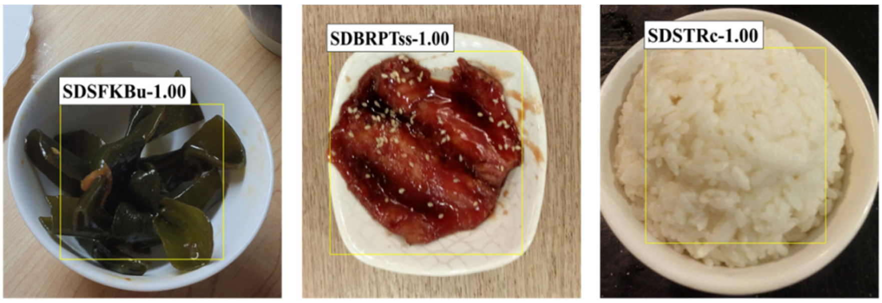

Figure 9. Recognition results for different types of single dishes.

As for the mixed dishes shown in Figure 10, the three mixed-dish items were "MDSFCCwCraBF" standing for "Mixed Dish Chinese Cabbage with Carrot and Black Fungus", "MDBOSolChSo" for "Mixed Dish Boiled Sesame Oil Chicken Soup", and "MDSFBSwCraBF" for "Mixed Dish Stir Fried Bean Sprout with Carrot and Black Fungus". Each dish had a confidence score of 0.99, 1.00, and 1.00, respectively.

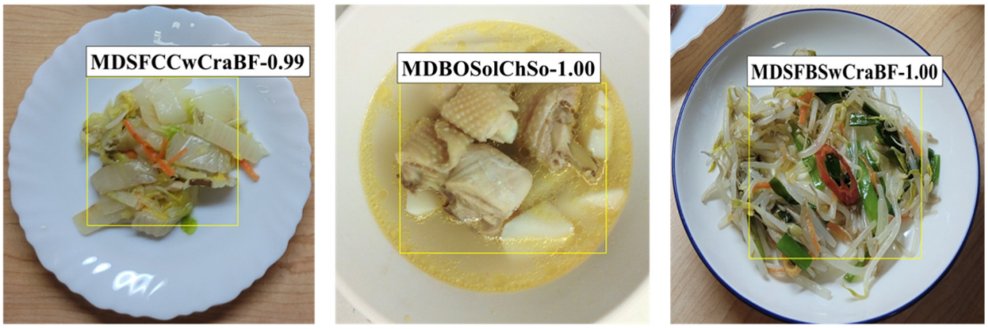

Figure 10. Recognition results for different types of mixed dishes.

For multiple dishes, we tested different sets that had different numbers of dishes (T). The model recognized multiple dishes with a total number of T = 3, T = 4, T = 5, and T = 6 dishes, as shown in Figure 11a–d, respectively. As a result, all the prediction results have a confidence score larger than >0.5. For instance, as shown in Figure 11b, the designed model recognized correctly “SDSTRc” and “SDPFChBs” for single-dish and “MDSFCCwCraBF” and “MDSFBwCraBF” for mixed-dish with an average confidence score of = 0.99. In Figure 11, prediction results appear for each dish where the same prediction class did not repeat for the same dish.

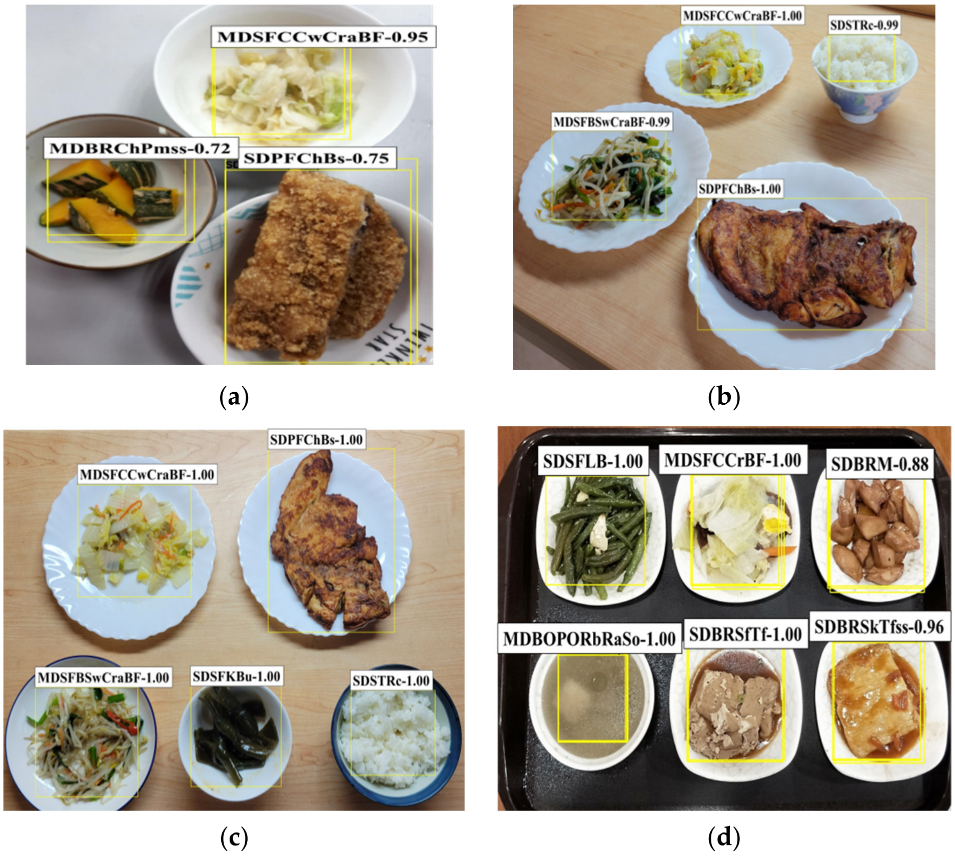

Figure 11. Recognition results for multiple dishes with different number of dishes (T): (**a**) T = 3, (**b**) T = 4; (**c**) T = 5; (**d**) T = 6.

5. Discussion

The contribution of this paper was to propose a model for recognizing various types of meals, such as single-food dishes, mixed-food dishes, and multiple-dish meals in homes, restaurants, and school cafeterias. Especially multiple dishes on a plate or tray [49] were common in many food reporting cases. Further, food reporting was often frequent. However, the existing applications of image recognition did not directly support the recognition of multiple-dish images. Therefore, an identification scheme was desirable in detecting and recognizing multiple dishes from a single image. The accuracy and timely response of food intake were critical and closely related to health-related applications in calorie calculation that are a key sub-function for health self-management application. The following addresses preformation of the proposed model, the potential application of the model, and the limitation.

5.1. Evaluation of the Model

The proposed model was shown to positively enhance its efficiency and accuracy in comparison to two major architectures, i.e., SSD Inception V2 and Faster R-CNN Inception ResNet V2. Moreover, the input from a human in data producing and preprocessing created human–AI interaction which helps AI to learn in a sustainable way. AI technology could gradually improve the richness of information and accuracy. Next, the proposed model extracted features with convolutional networks (CNNs) from a pre-trained EfficientNet-B1 for the backbones. Then, the extracted features were used further for the multiple-dish food recognition model by adopting EfficientDet-D1. The use of a three-step parsing method accelerated training and produced efficient results, as shown in Table 4. Although the trained food images have a limited number, the system is able to recognize three dish categories with relatively high accuracy. In overall performance, the design model can recognize all types of meals, namely single-dish, mixed-dish, and multiple-dish meals, with an accuracy = 0.87, precision = 0.88, recall = 0.97, and F1-score = 0.93.

The improved accuracy and confidence in our model could be explained by the following two reasons. Firstly, in the parsing step, we had the system identify its major attributes, i.e., in our research, we trained the model to identify the meal classified as single-dish, mixed-dish, or multiple-dish. Furthermore, the selection of EfficientDet architecture as the detector showed high-speed and highly accurate detection in comparison to SSD and Faster R-CN. This can be supported by the value of mAP = 0.92 with the lowest average inference time < 5 s. However, the result might be changed under a wide variety of tested dishes. Therefore, further investigation is needed to test with more types of food. Moreover, further research is needed to account for all possible meal reports, such as reporting bento- and buffet-like meals containing a range of single and mixed dishes.

5.2. Potential Implementation

The proposed model could be implemented into an automatic or semi-automatic application framework. Given the limited dishes to be recognized, we built an app prototype to include the automatic multiple-dish food recognition model into an existing framework. The implementation was to build an AI-enhanced version of our previous dietary intake app [24]. The user flow of the AI-enhanced application for multiple-dish food recognition architecture was followed. Firstly, the user took the photo, and the system obtained multiple-dish data under two-tier architecture. The server API was implemented in Python. The architecture allowed photos to be sent from the app to the AI server and received the responded data from the AI multiple-dish server. The feedback from the photo image was presented to the user. The responded data are the recognized results in which possible answers are provided for each dish of a meal. The user is presented with a possible dish name that requires the user's selection of the proper dish name. For a new dish unable to be identified from the AI server, manual input by means of voice or text input is required, which additionally allows the user to add or edit its food attribute (e.g., cooking method, portion, sugar level, and salt level).

Based on our initial trial in the use of the application, one of the challenges in uploading the meal-based photo was in users' intake behavior. Some families might start to eat before all dishes have finished being cooked. Therefore, it might not be possible to photograph all the dishes of a meal at one time. Therefore, the application should also provide partial meal or even single-dish reporting functions to accommodate different users' needs. For correct reporting of type and quantity of intake, further enhancement in AI-image detection should include left-over food images. Moreover, the volume of the food needs to be taken into consideration in calculating the amounts of macronutrients.

6. Conclusions

This paper presented an AI-based multiple-dish food recognition model to support meal-based image recognition. The proposed algorithm accurately recognized multiple dishes with 0.92 mAP where the recognition was performed faster with inference time <5 s. As the uploading of intake photos would be conducted for each meal daily, the usability of the application needs to be considered. This paper also attempted to demonstrate the possible application in mHealth. The model proposed in this paper attempted to reflect a realistic user scenario that has potential to support appropriate dietary intake in terms of efficiency in one-time meal reporting. In addition, the model showed potential for a promising solution for the recognition of multiple dishes of local cuisines. It was observed that the current approach has potential in recognizing the multiple dishes of single- and mixed-dish types of food. An automatic food recognition model was developed, which can recognize the food with bounding boxes on each food item to reduce the food reporting time and enhance the user experience. Our future work includes further improving the performance of the model and integrating it with the mobile app to facilitate the real-time integration between the mobile app and AI server and to enhance the accuracy of current measurements of dietary intake.

Author Contributions: Y.-C.L. and P.K.S. developed the concept and data audit of this study, particularly for the analysis checking, manuscript writing, and final critical revision. D.D.O. and S.M. made a significant contribution through the building, training, and analyzing of the food image recognition engine. D.I. supported our data collection, analysis, and administration. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Ministry of Science and Technology, Taiwan, under grant number MOST-NARPD3J0111-3 and in part by the Ministry of Science and Technology, Taiwan, under grant number 110-2221-E-182-008-MY3. Also, the Research Fund of Chang Gung Memorial Hospital, under grant number BMRPD67.

Acknowledgments: We acknowledge all members of Artificial Intelligence and Big Data Computing (ABC) Lab, Department of Computer Science and Information Engineering, Chang Gung University, Taiwan, who participated in this project and Food Interaction Design Lab, Department of Industrial Design, College of Management, Chang Gung University, Taiwan.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hajat, C.; Stein, E. The global burden of multiple chronic conditions: A narrative review. *Prev. Med. Rep.* **2018**, *12*, 284–293. [CrossRef] [PubMed]
2. Afshin, A.; Sur, P.J.; Fay, K.A.; Cornaby, L.; Ferrara, G.; Salama, J.S.; Mullany, E.C.; Abate, K.H.; Abbafati, C.; Abebe, Z.; et al. Health effects of dietary risks in 195 countries, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **2019**, *393*, 1958–1972. [CrossRef]
3. Springmann, M.; Wiebe, K.; Mason-D'Croz, D.; Sulser, T.B.; Rayner, M.; Scarborough, P. Health and nutritional aspects of sustainable diet strategies and their association with environmental impacts: A global modelling analysis with country-level detail. *Lancet Planet. Health* **2018**, *2*, e451–e461. [CrossRef]
4. Neuhouser, M.L. The importance of healthy dietary patterns in chronic disease prevention. *Nutr. Res.* **2019**, *70*, 3–6. [CrossRef]
5. Debon, R.; Coleone, J.D.; Bellei, E.A.; De Marchi, A.C.B. Mobile health applications for chronic diseases: A systematic review of features for lifestyle improvement. *Diabetes Metab.Syndr. Clin. Res. Rev.* **2019**, *13*, 2507–2512. [CrossRef]
6. Yannakoulia, M.; Mamalaki, E.; Anastasiou, C.A.; Mourtzi, N.; Lambrinoudaki, I.; Scarmeas, N. Eating habits and behaviors of older people: Where are we now and where should we go? *Maturitas* **2018**, *114*, 14–21. [CrossRef]
7. Wang, Y.; Min, J.; Khuri, J.; Xue, H.; Xie, B.; Kaminsky, L.A.; Cheskin, L.J. Effectiveness of mobile health interventions on diabetes and obesity treatment and management: Systematic review of systematic reviews. *JMIR mHealth uHealth* **2020**, *8*, e15400. [CrossRef]
8. Vandellanote, C.; Muller, A.M.; Short, C.E.; Hingle, M.; Nathan, N.; Williams, S.L.; Lopez, M.L.; Parekh, S.; Maher, C.A. Past, present, and future of eHealth and mHealth research to improve physical activity and dietary behaviors. *J. Nutr. Educ. Behav.* **2016**, *48*, 219–228.e1. [CrossRef]
9. Faiola, A.; Papautsky, E.L.; Isola, M. Empowering the aging with mobile health: A mHealth framework for supporting sustainable healthy lifestyle behavior. *Curr. Probl. Cardiol.* **2019**, *44*, 232–266. [CrossRef]
10. Lee, J.A.; Choi, M.; Lee, S.A.; Jiang, N. Effective behavioral intervention strategies using mobile health applications for chronic disease management: A systematic review. *BMC Med. Inform. Decis. Mak.* **2018**, *18*, 12. [CrossRef]
11. Lunde, P.; Nilsson, B.B.; Bergland, A.; Kværner, K.J.; Bye, A. The effectiveness of smartphone apps for lifestyle improvement in noncommunicable diseases: Systematic review and meta-analyses. *J. Med. Internet Res.* **2018**, *20*, e9751. [CrossRef] [PubMed]
12. Messner, E.M.; Probst, T.; O'Rourke, T.; Stoyanov, S. mHealth applications: Potentials, limitations, current quality and future directions. *Digit. Phenotyping Mob. Sens.* **2019**, 235–248. [CrossRef]
13. Cade, J.E. Measuring diet in the 21st century: Use of new technologies. *Proc. Nutr. Soc.* **2017**, *76*, 276–282. [CrossRef] [PubMed]
14. Eldridge, A.L.; Piernas, C.; Illner, A.K.; Gibney, M.J.; Gurinović, M.A.; De Vries, J.H.; Cade, J.E. Evaluation of new technology-based tools for dietary intake assessment—An ILSI Europe Dietary Intake and Exposure Task Force evaluation. *Nutrients* **2019**, *11*, 55. [CrossRef]
15. Angra, S.; Ahuja, S. Machine learning and its applications: A review. In Proceedings of the International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), Chirala, India, 23–25 March 2017; pp. 57–60.

16. Van Erp, M.; Reynolds, C.; Maynard, D.; Starke, A.; Ibáñez Martín, R.; Andres, F.; Leite, M.C.; Alvarez de Toledo, D.; Schmidt Rivera, X.; Trattner, C.; et al. Using natural language processing and artificial intelligence to explore the nutrition and sustainability of recipes and food. *Front. Artif. Intell.* **2020**, *3*, 621577. [CrossRef]
17. De Moraes Lopes, M.H.B.; Ferreira, D.D.; Ferreira, A.C.B.H.; da Silva, G.R.; Caetano, A.S.; Braz, V.N. Use of artificial intelligence in precision nutrition and fitness. In *Artificial Intelligence in Precision Health*; Academic Press: Cambridge, MA, USA, 2020; pp. 465–496.
18. Zhao, X.; Xu, X.; Li, X.; He, X.; Yang, Y.; Zhu, S. Emerging trends of technology-based dietary assessment: A perspective study. *Eur. J. Clin. Nutr.* **2021**, *75*, 582–587. [CrossRef]
19. Hussain, G.; Maheshwari, M.K.; Memon, M.L.; Jabbar, M.S.; Javed, K. A CNN based automated activity and food recognition using wearable sensor for preventive healthcare. *Electronics* **2019**, *8*, 1425. [CrossRef]
20. Ortega Anderez, D.; Lotfi, A.; Pourabdollah, A. A deep learning based wearable system for food and drink intake recognition. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *12*, 9435–9447. [CrossRef]
21. Lee, K.S. Automatic Estimation of Food Intake Amount Using Visual and Ultrasonic Signals. *Electronics* **2021**, *10*, 2153. [CrossRef]
22. Fakhrou, A.; Kunhoth, J.; Al Maadeed, S. Smartphone-based food recognition system using multiple deep CNN models. *Multimed. Tools Appl.* **2021**, *80*, 33011–33032. [CrossRef]
23. Sun, J.; Radecka, K.; Zilic, Z. FoodTracker: A Real-time Food Detection Mobile Application by Deep Convolutional Neural Networks. *arXiv* **2019**, *preprint*. arXiv:1909.05994.
24. Boushey, C.; Spoden, M.; Zhu, F.; Delp, E.; Kerr, D. New mobile methods for dietary assessment: Review of image-assisted and image-based dietary assessment methods. *Proc. Nutr. Soc.* **2017**, *76*, 283–294. [CrossRef] [PubMed]
25. Van Asbroeck, S.; Matthys, C. Use of Different Food Image Recognition Platforms in Dietary Assessment: Comparison Study. *JMIR Form. Res.* **2020**, *4*, e15602. [CrossRef] [PubMed]
26. Allegra, D.; Battiato, S.; Ortis, A.; Urso, S.; Polosa, R. A review on food recognition technology for health applications. *Health Psychol. Res.* **2020**, *30*, 8. [CrossRef] [PubMed]
27. Jiang, L.; Qiu, B.; Liu, X.; Huang, C.; Lin, K. DeepFood: Food image analysis and dietary assessment via deep model. *IEEE Access* **2020**, *8*, 47477–47489. [CrossRef]
28. Min, W.; Jiang, S.; Liu, L.; Rui, Y.; Jain, R. A survey on food computing. *ACM Comput. Surv.* **2019**, *52*, 1–36. [CrossRef]
29. Liu, Y.C.; Chen, C.H.; Lin, Y.S.; Chen, H.Y.; Irianti, D.; Jen, T.N.; Yeh, J.Y.; Chiu, S.Y.H. Design and usability evaluation of mobile voice-added food reporting for elderly people: Randomized controlled trial. *JMIR mHealth uHealth* **2020**, *8*, e20317. [CrossRef]
30. Meyers, A.; Johnston, N.; Rathod, V.; Korattikara, A.; Gorban, A.; Silberman, N.; Guadarrama, S.; Papandreou, G.; Huang, J.; Murphy, K.P. Im2Calories: Towards an automated mobile vision food diary. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1233–1241.
31. Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R.; Guadarrama, S.; Darrell, T. Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the 22nd ACM international conference on Multimedia, Orlando, FL, USA, 3–7 November 2014; pp. 675–678.
32. Pouladzadeh, P.; Yassine, A.; Shirmohammadi, S. FooDD: An image-based food detection dataset for calorie measurement. In Proceedings of the International Conference on Multimedia Assisted Dietary Management, Genova, Italy, 7–8 September 2015.
33. Aizawa, K.; Maruyama, Y.; Li, H.; Morikawa, C. Food balance estimation by using personal dietary tendencies in a multimedia food log. *IEEE Trans. Multimed.* **2013**, *15*, 2176–2185. [CrossRef]
34. Aizawa, K.; Ogawa, M. Foodlog: Multimedia tool for healthcare applications. *IEEE MultiMedia* **2015**, *22*, 4–8. [CrossRef]
35. Horiguchi, S.; Amano, S.; Ogawa, M.; Aizawa, K. Personalized classifier for food image recognition. *IEEE Trans. Multimed.* **2018**, *20*, 2836–2848. [CrossRef]
36. Yu, Q.; Anzawa, M.; Amano, S.; Ogawa, M.; Aizawa, K. Food image recognition by personalized classifier. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 171–175.
37. Anzawa, M.; Amano, S.; Yamakata, Y.; Motonaga, K.; Kamei, A.; Aizawa, K. Recognition of multiple food items in a single photo for use in a buffet-style restaurant. *IEICE Trans. Inf. Syst.* **2019**, *102*, 410–414. [CrossRef]
38. Foodvisor. Available online: www.foodvisor.io (accessed on 14 February 2022).
39. SnapCalorie. Available online: www.snapcalorie.com (accessed on 14 February 2022).
40. Knez, S.; Šajn, L. Food object recognition using a mobile device: Evaluation of currently implemented systems. *Trends Food Sci. Technol.* **2020**, *99*, 460–471. [CrossRef]
41. Pouladzadeh, P.; Shirmohammadi, S. Mobile multi-food recognition using deep learning. *ACM Trans. Multimed. Comput. Commun. Appl.* **2017**, *13*, 1–21. [CrossRef]
42. Tan, M.; Pang, R.; Le, Q.V. Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10781–10790.
43. Wang, S.; Liu, Y.; Qing, Y.; Wang, C.; Lan, T.; Yao, R. Detection of insulator defects with improved ResNest and region proposal network. *IEEE Access* **2020**, *8*, 184841–184850. [CrossRef]
44. Onthoni, D.D.; Sheng, T.W.; Sahoo, P.K.; Wang, L.J.; Gupta, P. Deep learning assisted localization of polycystic kidney on contrast-enhanced CT images. *Diagnostics* **2020**, *10*, 1113. [CrossRef] [PubMed]

45. Magalhães, S.A.; Castro, L.; Moreira, G.; Dos Santos, F.N.; Cunha, M.; Dias, J.; Moreira, A.P. Evaluating the single-shot multibox detector and YOLO deep learning models for the detection of tomatoes in a greenhouse. *Sensors* **2021**, *21*, 3569. [CrossRef] [PubMed]
46. Tsai, M.F.; Lin, P.C.; Huang, Z.H.; Lin, C.H. Multiple Feature Dependency Detection for Deep Learning Technology—Smart Pet Surveillance System Implementation. *Electronics* **2020**, *9*, 1387. [CrossRef]
47. Ramesh, A.; Sivakumar, A.; Angel, S.S. Real-time Food-Object Detection and Localization for Indian Cuisines using Deep Neural Networks. In Proceedings of the 2020 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT), Hyderabad, India, 20–21 December 2020; pp. 1–6.
48. Kaggle. Indian Food Image Dataset. Available online: https://www.kaggle.com/datasets/iamsouravbanerjee/indian-food-images-dataset (accessed on 6 May 2022).
49. Liu, Y.C.; Chen, C.H.; Tsou, Y.C.; Lin, Y.S.; Chen, H.Y.; Yeh, J.Y.; Chiu, S.Y.H. Evaluating mobile health apps for customized dietary recording for young adults and seniors: Randomized controlled trial. *JMIR mHealth uHealth* **2019**, *7*, e10931. [CrossRef]