Current Developments in Nutrition 8 (2024) 102154

American Society for Nutrition
Excellence in Nutrition Research and Practice

Current Developments in Nutrition

journal homepage: https://cdn.nutrition.org/

Original Research

Automated Artificial Intelligence–Based Thai Food Dietary Assessment System: Development and Validation

Phawinpon Chotwanvirat [1,2], Aree Prachansuwan [1], Pimnapanut Sridonpai [1], Wantanee Kriengsinyos [1,*]

[1] Human Nutrition Unit, Food and Nutrition Academic and Research Cluster, Institute of Nutrition, Mahidol University, Nakhon Pathom, Thailand;
[2] Diabetes and Thyroid Center, Theptarin Hospital, Khlong Toei, Bangkok, Thailand

A B S T R A C T

Background: Dietary assessment is a fundamental component of nutrition research and plays a pivotal role in managing chronic diseases. Traditional dietary assessment methods, particularly in the context of Thai cuisine, often require extensive training and may lead to estimation errors.
Objectives: To address these challenges, Institute of Nutrition, Mahidol University (INMU) iFood, an innovative artificial intelligence–based Thai food dietary assessment system, allows for estimating the nutritive values of dishes from food images.
Methods: INMU iFood leverages state-of-the-art technology and integrates a validated automated Thai food analysis system. Users can use 3 distinct input methods: food image recognition, manual input, and a convenient barcode scanner. This versatility simplifies the tracking of dietary intake while maximizing data quality at the individual level. The core improvement in INMU iFood can be attributed to 2 key factors, namely, the replacement of Yolov4-tiny with Yolov7 and the expansion of noncarbohydrate source foods in the training image data set.
Results: This combination significantly enhances the system's ability to identify food items, especially in scenarios with closely packed food images, thus improving accuracy. Validation results showcase the superior performance of the INMU iFood integrated V7-based system over its predecessor, V4-based, with notable improvements in protein and fat estimation. Furthermore, INMU iFood addresses limitations by offering users the option to import additional food products via a barcode scanner, thus providing access to a vast database of nutritional information through Open Food Facts. This integration ensures users can track their dietary intake effectively, with expanded access to over 3000 food items added to or updated in the Open Food Facts database covering a wide variety of dietary choices.
Conclusions: INMU iFood is a promising tool for researchers, health care professionals, and individuals seeking to monitor their dietary intake within the context of Thai cuisine and for ultimately promoting better health outcomes and facilitating nutrition-related research.

Keywords: image-assisted dietary assessment, artificial intelligence, dietary assessment, Thai food, validation, macronutrients

Introduction

Dietary assessment is a cornerstone of nutrition-related research and plays a pivotal role in managing chronic diseases. The acquisition of accurate dietary intake information, however, presents a formidable challenge. In particular, demand persists for innovative dietary assessment approaches that offer enhanced measurement precision, ease of use, reduced workload, and cost efficiency [1].

Conventionally, dietary assessment has been the domain of dietitians or nutritionists. In Thailand, dietitians usually calculate the nutritive values of dishes based on standard recipes, adjusting them to account for visible or readily known modified ingredients, often with the aid of food exchange lists [2,3]. This approach mitigates 2 significant issues, namely, the limited coverage of prepared foods in the latest Thai food composition tables as well as variability in ingredient usage in Thai cuisine, as driven by individual preferences [4,5]. An unintended consequence of this method is the lengthy training required for effective dietary assessment, particularly for individuals lacking dietary and mathematical proficiency. Additionally, inherent errors in portion size estimation are inevitable among untrained individuals [6–8].

* Corresponding author. *E-mail address:* wantanee.krieng@mahidol.ac.th (W. Kriengsinyos).

https://doi.org/10.1016/j.cdnut.2024.102154
Received 3 December 2023; Received in revised form 24 March 2024; Accepted 29 March 2024; Available online 4 April 2024

Abbreviations

API	An Application Programming Interface (API) is a set of protocols, tools, and definitions that allows different software applications to communicate with each other. It defines the methods and data formats that applications can use to request and exchange information
INMU	Institute of Nutrition, Mahidol University
mAP@0.5	Mean Average Precision at an Intersection over Union (IoU) threshold of 0.5 is a metric commonly used to evaluate the performance of object detection algorithms. For food recognition, when the system identifies food and predicts bounding boxes that overlap in the correct position by at least 50%, this is considered as a correct identification. The average precision is calculated across all categories in a test dataset
MS COCO	Microsoft Common Objects in Context is a large-scale dataset designed for object recognition and image understanding tasks in computer vision. It contains a diverse set of images with everyday scenes and objects including food
PWA	A Progressive Web Application (PWA) is an application built using web technologies that can be installed on a user's device through a compatible web browser. It operates dependently on its respective browser, similar to opening a website in Fullscreen mode, and creates a shortcut icon on the device home screen, akin to a native application. PWAs offer a seamless user experience, delivering enhanced functionality beyond traditional web applications, although not identical to native applications, based on the functionality supported by their respective browser. Currently, Chromium-based browsers provide the most comprehensive PWA support
RGB	RGB stands for Red, Green, Blue and is a standard color model in computer systems. In an RGB image, each pixel is defined by its intensity in these three colors. RGB images augmented with depth information are called RGBD images
RMSE	Root means square error
TOST	Two one-sided test
Yolo	Yolo (You Only Look Once) is an open-source object detection model originally developed by Joseph Redmon. It features a user-friendly training system with pre-train and fine-tune capabilities. Over time, newer models have incorporated modifications to the original algorithms, integrating advanced techniques to enhance performance and efficiency.

Taking food photographs before eating sessions could help reduce these types of errors. However, the portion size estimation from food images still relies on human estimation [9–12]. The advent of high-performance smartphones has ushered in the potential to estimate the nutritive values of food from images. A set of sequentially independent algorithms, namely, food recognition, food portion estimation, and computation of representative nutrients using food composition tables, were used to handle the human-dependent complexity task in estimating the nutritive values of food from images [13–17]. This norm has become the traditional approach for image-assisted dietary assessment. The system's overall accuracy relies on the precision and effectiveness of algorithms responsible for each step, which have undergone continual enhancement over time through the contributions of numerous researchers [18–22].

Within the contemporary landscape of dietary assessment and nutrition analysis, a notable void exists for Thai cuisine. Despite the richness and diversity of Thai culinary offerings, only a few dedicated systems are tailored exclusively to recognize Thai foods [23–27], with their focus limited primarily to food recognition rather than covering the entire dietary assessment process. Although a few systems provide estimated nutritive values, oftentimes, they do not include validation results in their reports [28–30].

This deficiency poses a significant challenge for individuals monitoring their nutritional intake, as well as for health care professionals and researchers striving for precision in dietary analysis within the Thai culinary realm. Recognizing this gap, our mission has been to develop a smartphone-based dietary assessment system as a means to address this challenge.

To meet this need, we have built on our previous work in carbohydrate estimation from Thai food images and are now introducing Institute of Nutrition, Mahidol University (INMU) iFood, which is an automated artificial intelligence (AI)-integrated Thai food dietary assessment system. This application represents a versatile solution capable of delivering an acceptable level of dietary assessment results covering a diverse array of Thai dishes. In this report, our primary focus is to demonstrate improvement in the automated Thai food dietary assessment system. We compared the estimated macronutrient results produced by the system—obtained from food images, not barcode or manual input—against those generated by the previous carbohydrate estimation system [31], using measured weight as the ground truth in this report.

Methods

System outline

INMU iFood has 2 main components: an automated Thai food analysis system running on a server as well as a mobile web frontend, which runs on the user's device after complete installation. The frontend is a progressive web application (PWA) responsible for user interaction, food image uploads, storage of the food composition database, and calculation of representative nutritive values, including protein, carbohydrates, and fat. The backend is an automated Thai food analysis system, which is an updated version of our previous work [31]. A system outline is given in Figure 1.

For the frontend, the PWA was constructed using jQuery, with mobile-specific modern user interface components based on the OnsenUI library [32]. The browser's built-in IndexedDB [33] was used to store user-generated information and the food

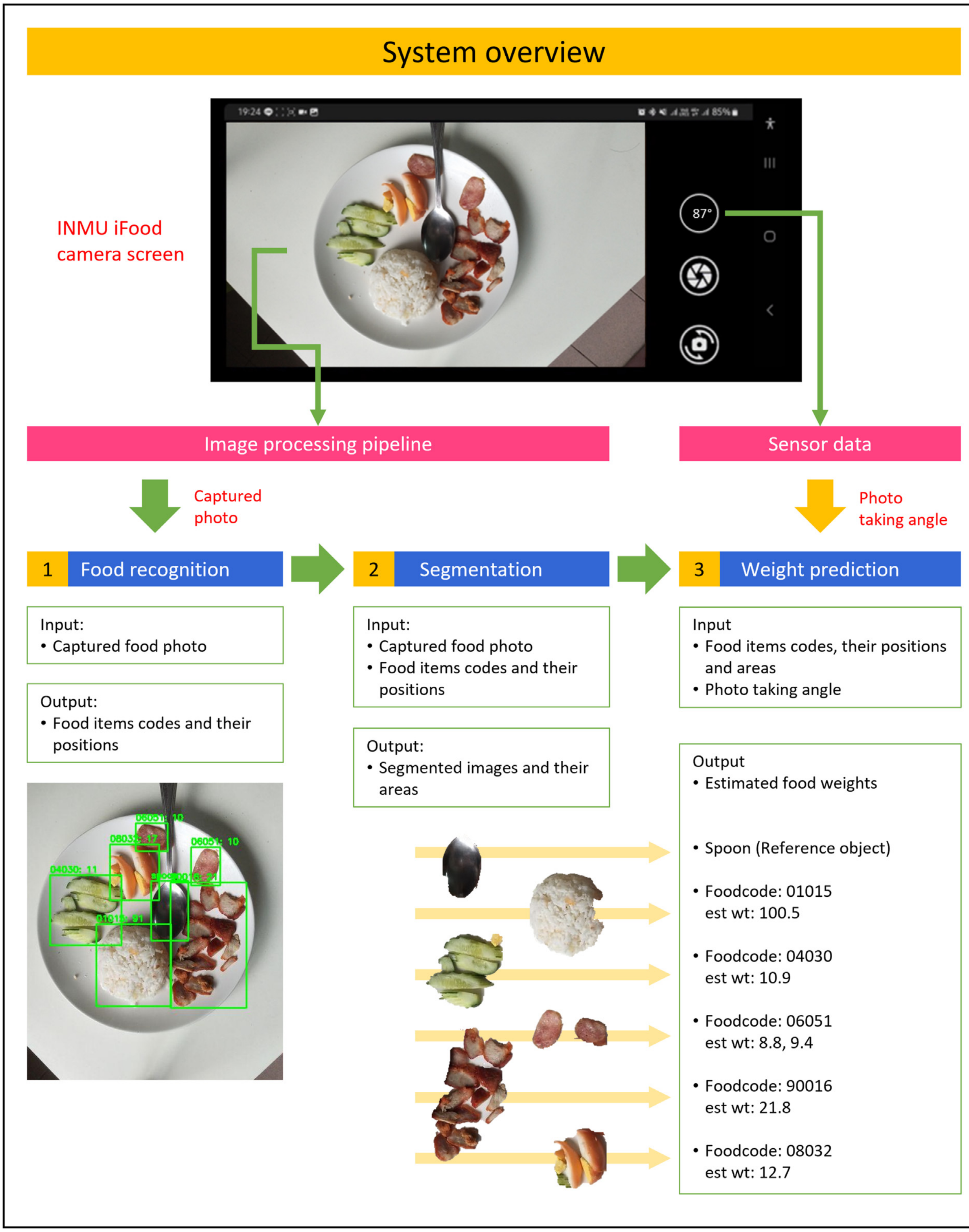

FIGURE 1. System overview of INMU iFood. The client-side camera interface is displayed at the top, aiding in capturing food photographs and measuring photograph-taking angle at the same time. Once the image is acquired successfully, it is transmitted to the Thai food analysis system running on the server for food image analysis. On completion of the analysis, identified food objects, their estimated weights, and locations are sent back to the client-side for the calculation of nutritive values. Est wt, estimated weight; INMU, Institute of Nutrition, Mahidol University.

composition database on a user's device. The INMU iFood frontend was designed and tested exclusively for chromium-based browsers (i.e., Google Chrome or Microsoft Edge for Android), which are available for Android smartphones only. No specific hardware is required for using the frontend beyond the requirements of its respective browser, functional camera resolution of >2 MP, a functional accelerometer or gyroscope, and sufficient internal memory for storing user-created content. Additionally, an always-on internet connection is required during usage. After successful installation on a supported device, the application icon appears on the device's home screen. On launching, the main interface promptly engages with the user and consists of only 2 tabs: an overall tab and a food search tab. Screenshots of the INMU iFood frontend are shown in Figure 2.

In the 'Overall' tab, there are three subsections: the food image section (Figure 2A), the food item list section (Figure 2B), and the macronutrients analysis section (Figure 2C). User can initiate food image capture by tapping the 'Add Food Image' button (Figure 2A). This action launches the built-in camera interface designed for simultaneous food image capture and measures the photo-taking angle. The photo-taking angle is displayed in a circular label above the shutter button to ensure that the user can capture an image at the correct degree, as shown in Figure 1. The interface guides the user to rotate the mobile device for capturing photos in landscape mode, aligning with the requirements of the automated Thai food analysis system. The built-in camera interface does not launch if the device remains in portrait position. The app provides initial instructions for taking photos to ensure optimal system usage, including maintaining distances of 30 cm from the food, ensuring clear presentation without items overlapping, placing the food item on a low-reflective material along with a tablespoon, and capturing the

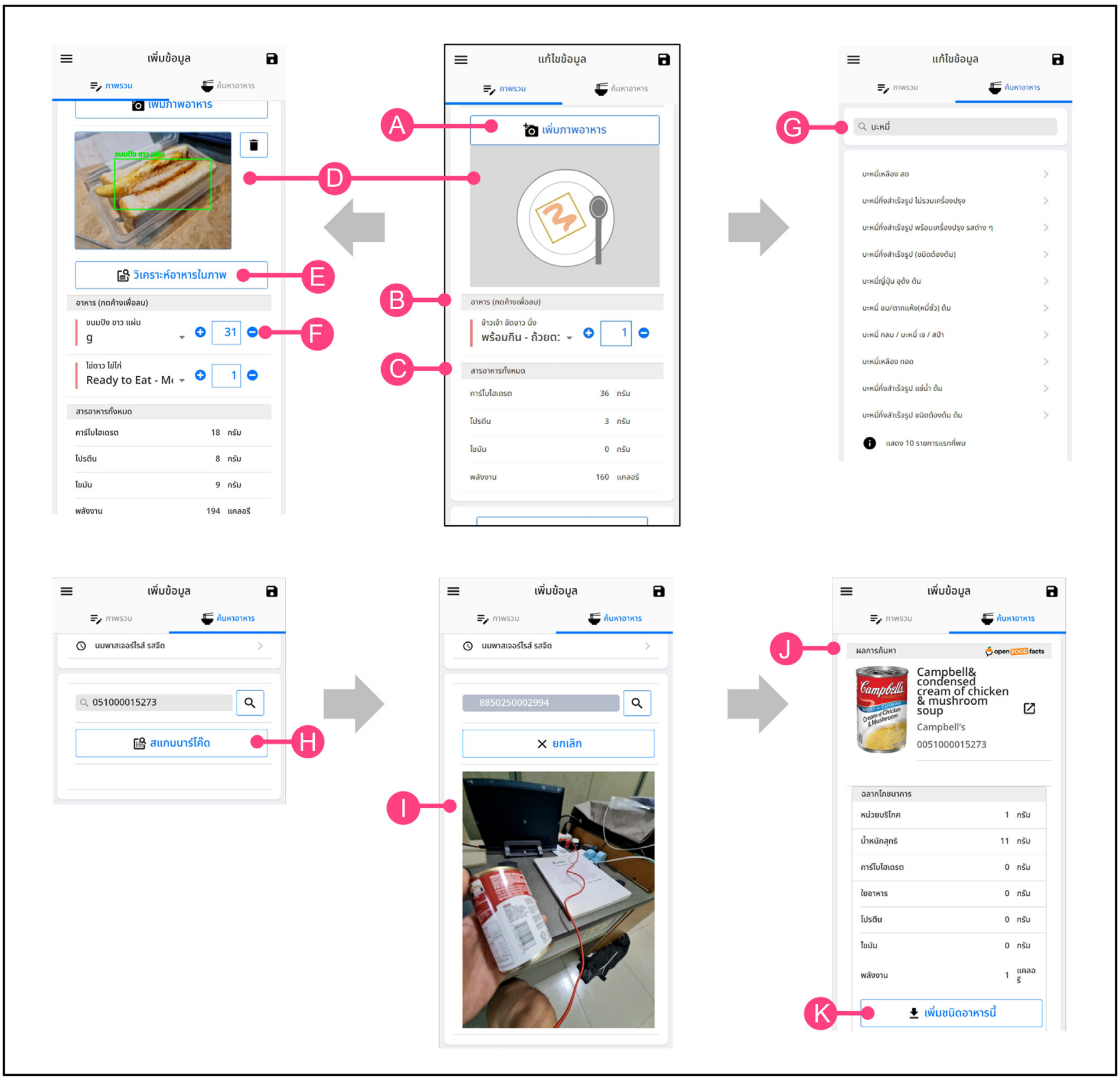

FIGURE 2. The Institute of Nutrition, Mahidol University iFood frontend displayed in Thai, with the main interface positioned at the top-center of the figure. Key components include the following: (A) launching the built-in camera interface, (B) selecting the food items listed in this area for nutrient analysis, (C) displaying a summary table of calculated macronutrients, (D) displaying the captured food image, (E) submitting the captured food image to the Thai food analysis system, (F) presenting identified food items in the analyzed food image and their estimated weights, (G) providing a search box for manual input and the 10 related food items listed underneath for selection, and (H–K) demonstrating the import of food item information via a barcode scanner, which retrieves nutrition data from OpenFoodFacts.org.

photo within a 30° to 90° angle from the tabletop. Once the food image is successfully captured, it is displayed in Figure 2D. The 'analyze food image' button is promptly available for the user to decide to send the captured image to the Thai food analysis system. After successful analysis, the detected food items, along with their estimated weights, are automatically added to the food item list section, as displayed in Figure 2F, and green boxes are drawn to display the detected food items to the user for rechecking purposes.The 'Food Search' tab offers two additional features: a manual search box (Figure 2G) and the ability to import additional food product information via a barcode scanner. By selecting the barcode scanner option (Figure 2H), the built-in camera screen is displayed, guiding the user to position the food item barcode within the camera's view (Figure 2I). Once the barcode is successfully captured, the product information, along with its nutrients, is retrieved from the Open Food Facts database and displayed on the screen for the user to review (Figure 2J). Users can add the retrieved product to their personal food database by selecting the "import to database" button (Figure 2K).

For the backend, the automated Thai food analysis system retains the traditional approach for image-assisted dietary assessment that entails 3 sequential processes: food recognition, segmentation, and food weight prediction. These are managed by 3 independent algorithms—namely, the state-of-the-art object detector named Yolov7 [34], GrabCut [35], and feedforward neural networks, respectively. Nutrient calculations are managed on the frontend. The system used only 2 input factors: a single RGB image (representing red, green, and blue, a standard color image format in computer systems) and a measured photograph-taking angle to ensure convenience and broader device support.

To achieve broader food recognition coverage, the food image data set was expanded from 175 classes, primarily focusing on foods contributing carbohydrates, to 400 classes of Thai food image components. The food items were chosen based on their frequent usage, covering ~80% of food items used in the analysis of the previous report [36] from a Thai diabetes population and including all solid food items in the official Thai food exchange list [2]. The selected food items were purchased from 3 to 4 local vendors and prepared as ready-to-eat forms. Some foods, such as fresh vegetables and cooked meats, underwent modifications like dicing, chopping, or grinding to alter their physical appearance and portion size.

The food ingredients were separated following common practices. In cases where separation was not possible, such as fried rice with egg, it was treated as a new item (plain fried rice and fried rice with egg). In contrast, dried tiny prawns or chopped coriander in noodle were considered a visual variation of the noodle, resulting in a single item (i.e., only the noodle was annotated) owing to their providing small amounts of nutrients. These decisions were applied during the food object annotation in the later stage. Various styles of tablespoons were randomly positioned alongside the foods, but only the bowl part of the tablespoon was used as a reference object owing to its low variation in size.

Containers were placed on an electric-powered rotating plate. Although the rotating plate was in motion, burst shots were taken simultaneously at 30°, 60°, and 90° angles from the tabletop, resulting in a series of multiangle food images. These methods were applied collectively to increase the variability of food ingredients' appearance and portion size, which is required for training both the food recognition and portion size estimation algorithms. Example images from the expanded data set are shown in Figure 3.

In the new system, we used the state-of-the-art object detector named You Only Look Once (Yolo) but switched from Yolov4-tiny in the former system to the Yolov7 [34] standard model, which was released 2 y later and gained improvements in detection accuracy and computational efficiency. A Microsoft Common Objects in Context (MS COCO) pretrained model was retrained with the expanded food image dataset together with 1800 nonfood images. The new recognition system achieved a mean average precision at an intersection over union threshold of 0.5 (mAP@0.5) score of 88.3%. This metric means that, on average, the system achieved 88.3% accuracy in recognizing food objects across all test images in the data set. Predictions were considered correct when the system predicted the boundaries of food objects overlapping the true location by ≥50%.

Although Yolov7 offers both object detection and segmentation, only object detection tasks were used because there was no segmentation data set available for custom segmentation training. The identified objects were subsequently cropped based on their boundary boxes, which were obtained from the results of Yolov7, whereas GrabCut [35] was used for background subtraction. It extracted foreground objects based on the color distribution within specific boundary boxes without requiring any training step—a simple, yet fast and sufficient technique for images with few objects in monotonous backgrounds. Based on the fact that segmentation ability directly impacts food weight prediction, it is indirectly measured through the overall performance of the system in this report.

For food weight prediction, the same approach as in the previous work was used with some modifications. In essence, simple feedforward neural networks were trained to predict food weight using input features including shooting angle, food object areas and locations, as well as reference object area and location when available. In the previous work, the training regimens were considered complete when they achieved a calculated carbohydrate estimation error of <10 g, as measured by root mean square error (RMSE). In this work, owing to the inclusion of more foods with nonsignificant carbohydrate content, the training regimens were completed when the RMSE was <0.5 units of its exchange weight. For example, one exchange of steamed white rice weighed 60 g, thus an estimation error of <30 g was deemed acceptable. Although this error may seem high, it is meaningful enough to provide clinical benefits. This level of error in carbohydrate estimation, for instance, does not affect blood glucose concentrations in children with type 1 diabetes [37].

Statistical analysis

To demonstrate overall improvement in estimation accuracy of the new system (V7 based), we compared it with the previous carbohydrate estimation system (V4 based) using the same set of 20 known-weight food images and performed a ground truth comparison with measured weight. We randomly selected 20 food menus from the image data set, which included 7 mixed dishes (comprising 2 types of rice, 2 types of noodles, pork, chicken, omelets, boiled eggs, shrimp, meatballs, and 9 vegetable components), 4 fruits, and 9 Thai traditional desserts.

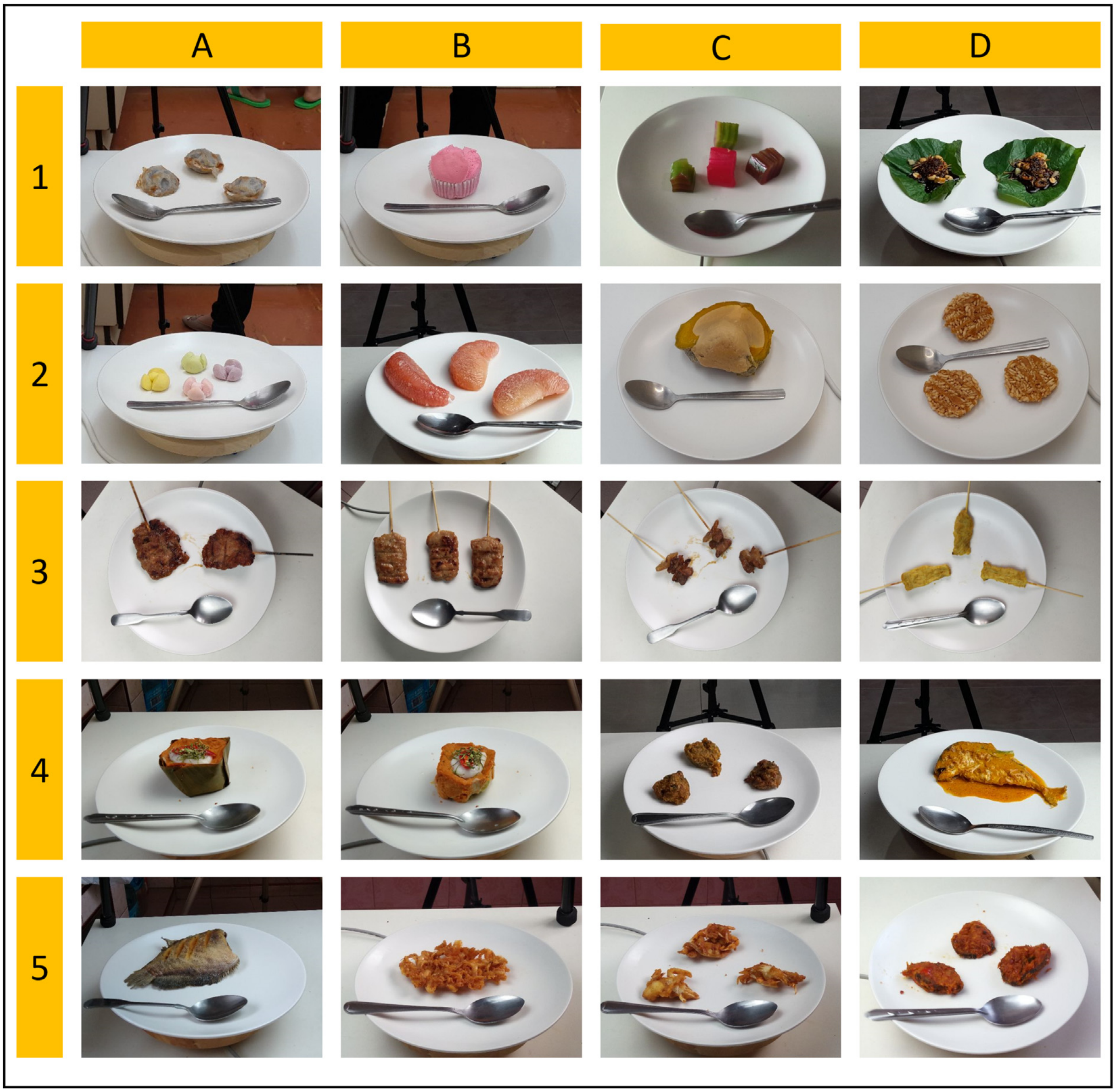

FIGURE 3. Variations in food, visual style, and portion sizes for an expanded dataset. The first two rows (A1:D2) feature Thai traditional desserts, including coconut milk pudding (Krok), fluffy cupcakes (Pui Fai), pandan layer cake (Chan), and savoury leaf wraps (Miang Kham). Additionally, the second row includes shortbread cookies (Kleeb Lum Duen), pomelo, pumpkin custard, and rice crispy (Nang Led). In the third row (A3:D3), variations in portion sizes of grilled pork are showcased. In the fourth row, A4 and B4 depict steamed fish curry (Homok) with and without a leaf container, C4 represents fried fish cake (Tod Mun), and D4 illustrates mackerel in dried curry (Chuu Chee Pla Too). Finally, A5 to D5 feature deep-fried snake-skin gourami, deep-fried battered shrimp, soft-shell crispy crab, and catfish in dried curry (Chuu Chee Pla Duk).

These foods were obtained from local vendors and modified into 2–3 variations in portion sizes and separated into individual components. These foods were prepared in ready-to-eat forms, and their ingredients were arranged according to the common practices of Thai dietitians. Most of the food was placed on a plate, and only noodle dishes were placed in a 6-inch bowl without the addition of soup liquid. Food photographs were captured with a smartphone held in landscape position, within a range of 30 cm from the food objects, at angles of 30° to 90° from the tabletop, using a generic Android smartphone (Samsung SM-N770F), with 10–15 images taken per food. From this data set, 20 images were chosen randomly, encompassing a total of 48 food items. The macronutrient contents of each image could be determined by summarizing the macronutrient contents of each detected food ingredient in that image.

Statistical analyses were conducted to assess the accuracy of the estimated macronutrients from both systems and compare them with the ground truth. The accuracy of estimated macronutrients from each system was evaluated using the RMSE.

The means of estimated macronutrients from each system were compared against the ground truth using a two 1-side test for difference in dependent means [38] to assess equivalence. In line with the system design, equivalence bounds were defined based on 0.5 exchange unit, resulting in ±3.5 g of protein, ±8 g of carbohydrate, and ±2.5 g of fat. These values

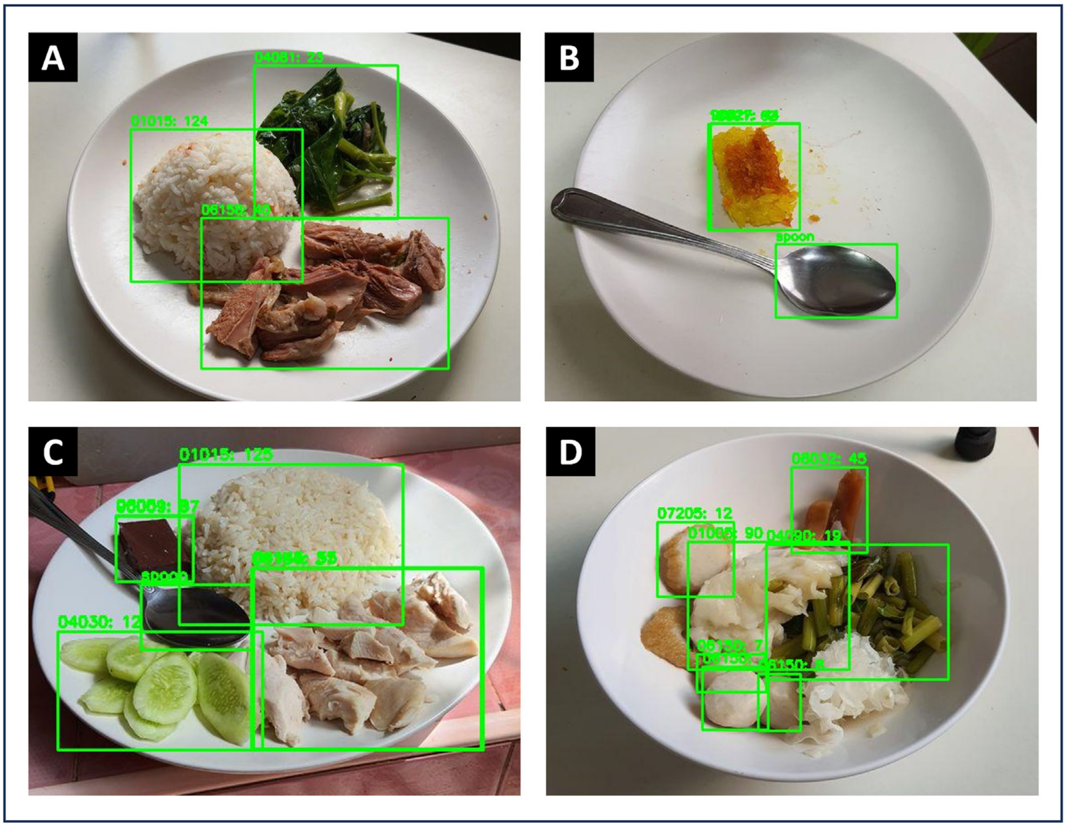

FIGURE 4. Example of the accuracy of detection of the V7-based system, in which 2 items were undetected, and the boiled morning glory was misclassified as fresh morning glory in Figure 4D.

approximately correspond to Cohen d_z of ±0.6, indicating a medium effect size.

Additionally, the linear association and agreement between estimated results were measured using Pearson correlation coefficient (r) and Lin concordance correlation coefficient (Rc). The agreement was visualized using Bland–Altman plots, and the limits of agreement were calculated to demonstrate the variability of these estimated results. These statistical methods allowed us to assess accuracy and agreement between the 2 systems. A 2-sided P value of <0.05 was considered statistically significant.

Results

For detection performance, the comparison was calculated based on all 48 food objects contained in the validated data set. The V7-based system demonstrated a 14.6% improvement in correctly detecting food objects compared with the V4-based system (V7-based: 43 of 48 items compared with V4-based: 36 of 48 items). Only 2 items were undetected, namely, white jelly fungus and fish strips, whereas 3 items were misclassified in their cooking variations: fresh morning glory instead of boiled morning glory, stir-fried kale instead of boiled kale, and blanched mung bean sprout instead of fresh mung bean sprout. Figure 4 presents visual examples of food images along with the corresponding system detection results from the validation set.

For the validation data set, the V7-based system could estimate the sum of nutrients more accurately than the V4-based system for protein (132 g compared with 94 g, ground truth 152 g) and fat (111 g compared with 86 g, ground truth 112 g), with comparable results for carbohydrate estimation compared with the V4-based system (622 g compared with 584 g, ground truth 605 g).

When comparing the group means ± SD of macronutrients with the ground truth, the V7-based system demonstrated significant equivalence for protein (6.7 ± 7.2 compared with 7.6 ± 8.6; P = 0.002 and 0.031, respectively), carbohydrate (31.1 ± 14.8 compared with 30.3 ± 14.8; P = 0.015 and 0.004, respectively), and fat (5.6 ± 5.3 compared with 5.6 ± 5.1; P = 0.007 and 0.007, respectively). The 90% confidence interval for all macronutrients fell within the lower and upper equivalence bounds as predefined for each macronutrient. On the contrary, the V4-based system achieved equivalent results only for carbohydrate (29.2 ± 15.3 compared with 30.0 ± 14.8; P = 0.003 and 0.019, respectively) but not for protein (4.7 ± 4.6 compared with 7.6 ± 8.6; P < 0.0001 and P = 0.302, respectively) and fat (4.3 ± 4.7 compared with 5.6 ± 5.1; P < 0.0001 and P = 0.158, respectively).

Moreover, the V7-based system outperformed the V4-based system in several quality aspects. It exhibited lower estimated variation, as evidenced by reduced RMSE values for protein (2.7 compared with 6.5), carbohydrate (7.9 compared with 9.4), and fat (1.9 compared with 3.2). In addition, it demonstrated higher agreement, as indicated by increased concordance correlation coefficients (Rc), for protein (0.94 compared with 0.57), carbohydrate (0.91 compared with 0.79), and fat (0.83 compared with 0.78). Furthermore, it showed stronger correlation, as determined by Pearson correlation coefficient (r), for protein (0.96

TABLE 1
Statistical results for the estimation of error, agreement, and relationship as measured by RMSE, *Rc*, and *r* between V7-based and V4-based models against measured weight as ground truth, respectively

Nutrient	Mean ± SD	TOST *P*	TOST 90% CI	RMSE	*Rc*	*r*
Protein						
Ground truth	7.6 ± 8.6					
V7-based system	6.7 ± 7.2	0.002, 0.031[1]	−1.139, 1.339	2.7	0.94	0.96
V4-based system	4.7 ± 4.6	<0.001, 0.302	−5.226, −0.574	6.5	0.57	0.74
Carbohydrate						
Ground truth	30.3 ± 14.8					
V7-based system	31.1 ± 14.8	0.015, 0.004[1]	−3.287, 4.887	7.9	0.91	0.85
V4-based system	29.2 ± 15.3	0.003, 0.019[1]	−5.260, 3.060	9.4	0.79	0.80
Fat						
Ground truth	5.6 ± 5.1					
V7-based system	5.6 ± 5.3	0.007, 0.007[1]	−1.438, 1.438	1.9	0.83	0.93
V4-based system	4.3 ± 4.7	<0.001, 0.158	−2.661, 0.061	3.2	0.78	0.81

CI, confidence interval; *Rc*, Lin concordance correlation coefficient; RMSE, root mean square error; TOST, two 1-sided test.
[1] A significant difference when both lower and upper *P* values fall within predefined equivalence bounds.

compared with 0.74), carbohydrate (0.85 compared with 0.80), and fat (0.93 compared with 0.81). Detailed results of all statistical analyses are presented in Table 1.

To visualize the agreement between the ground truth and the 2 system versions, Bland–Altman plots revealed that the V7-based system demonstrated narrower limits of agreement than the V4-based system for all macronutrients. This indicates that the V7-based system had lower variability than the V4-based system. Additionally, no significant outliers were observed in the estimated protein values from the V7-based system, whereas some outliers were found in the results from the V4-based system. The plots are displayed in Figure 5.

Discussion

INMU iFood embodies state-of-the-art dietary assessment technology integrated with a validated automated Thai food image analysis system that has been tested for robustness and accuracy in estimating nutritive values from Thai food images. It offers 3 distinct input methods: food image recognition, manual input, and the convenience of a barcode scanner. This versatility ensures that users can track their dietary intake with less effort and provides adequate quality at the individual level.

The INMU iFood integrated V7-based system outperformed the previous V4-based system across all measured nutrients based on the comparison of results. The most significant improvement was in the estimated results for protein and fat, where the V4-based system's outcomes were deemed unacceptable. Even in the case of carbohydrates, the V7-based system demonstrated a nonsignificant modest improvement.

The primary contributors to this improvement were 2-fold: *1*) the replacement of Yolov4-tiny with Yolov7, primarily because Yolov7 excels in detecting objects placed in closer proximity; and *2*) the expansion of noncarbohydrate food images in the expanded data set, mostly entailing a wide variety of meat and leafy vegetables. This improvement is evident in Figure 4C, D, which present the most challenging images in the validation set. These images featured small food items closely packed within the same frame. The V4-based system could only detect 2 of 4 items in Figure 4C and 3 of 9 items in Figure 4D. In contrast, the

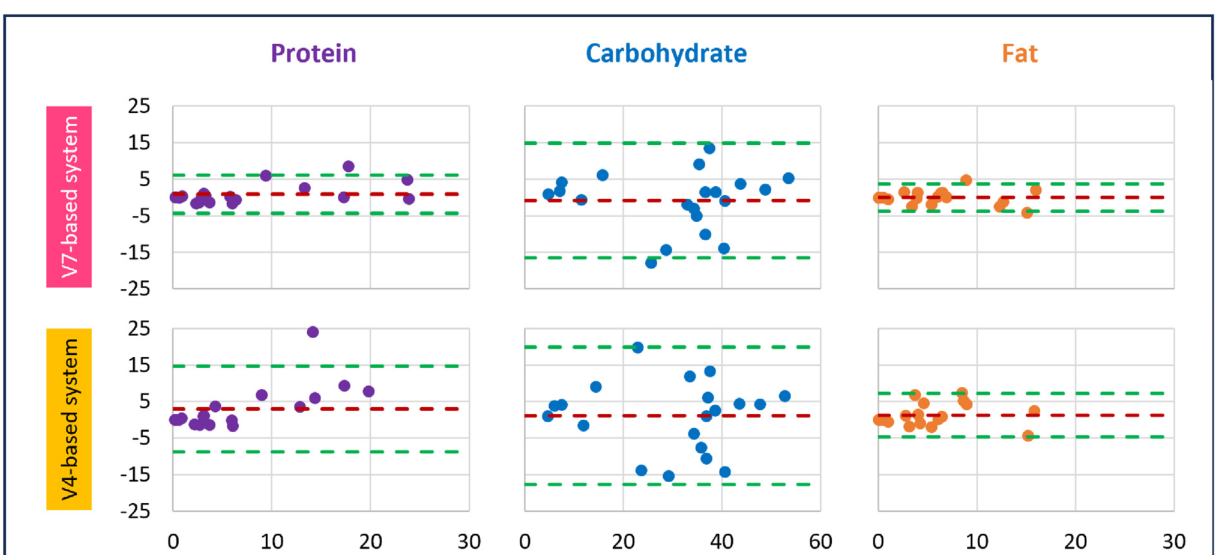

FIGURE 5. Bland–Altman plots demonstrate the agreement between the calculated macronutrient content of ground truth, the V7-based system, and the previous V4-based system. The green dotted lines represent the limits of agreement between the other methods and the ground truth.

V7-based system successfully detected all items in Figure 4C and 7 of 9 items in Figure 4D. Our approach is primarily reliant on detecting individual food ingredients within a food image, rather than classifying the entire recipe. Thus, a higher number of detected food ingredients results in a closer estimation.

The food recognition component can use any state-of-the-art object detector, which works by identifying the location of its known objects in an input image and drawing bounding boxes that cover the identified objects as accurately as possible. However, we opted for the Yolov7 standard model owing to its popularity among developers, ample documentation, and a wealth of reported issues and solutions available on its GitHub repository. The Yolov7 family applies new techniques including Extended Efficient Layer Aggregation Networks, Model Scaling, and Trainable bag-of-freebies, which result in a 1.5% increase in average precision, a 75% reduction in usage parameters, and a 36% reduction of usage computation, compared with the Yolov4 family [34]. Furthermore, we noticed that although Yolov8 [39] had been released at the time of this research study, Yolov7 was licensed under GPL v3.0, whereas Yolov8 used AGPL-3.0. Consequently, we made the decision to use Yolov7 based on this consideration.

For food portion estimation, the system aimed to use only a single RGB image in order to strike a balance between estimation accuracy—which requires more information, that is, $\geq$2 images or depth information from a depth sensor—and convenience based on broader device support. The pixel-counting approach was introduced in the early stages of image-assisted dietary assessment owing to its straightforwardness [15,40,41]. However, this approach is susceptible to variations in photograph-taking angle and distance between objects and capturing devices. To address these issues, 1 common approach involves using a physical reference object with a known exact size to calibrate the real size of a food object. Various physical objects have been introduced for this purpose, including a special physical card [40,41], specific circular plate [42] or bowl [43], and common objects found in daily life such as chopsticks [44], a 1-yuan coin [45], a wallet [46], a user's thumb [13], or a grain of rice [47]. This approach uses multiple advanced equations to correct the distortion in food images and leads to a reported relative error in portion estimation of approximately <30% (ranging between 6.65% and 27.60%), which appears to be acceptable.

This approach was selected and reimplemented through a newer method, namely the feedforward regression model. In the early development phase, data exploration demonstrated the possibility of applying this method by incorporating both suitable intervals of food portion sizes and the use of a tablespoon for calibration. The feedforward models were trained in terms of shooting angles, food object areas and locations, as well as reference object area and location. Some information for exploration was provided in Supplemental Figure 1. The systems were tested to ensure that they could produce food weight estimation errors of <0.5 units of their exchange weight when measured by RMSE. This approach might exhibit sensitivity to several parameters, including the distance between food objects and a user's phone, as well as the photograph-taking angles. Especially when these parameters exceed the recommended range, this error can be mitigated by providing comprehensive instructions to the user about the system's limitations.

We acknowledge that the system's validation was conducted using specific data sets, which might not be publicly available, making it challenging to compare with other available systems. Although there are currently public data sets containing Thai food images [23,48,49], unfortunately, they lack information on food weight or calorie content. Consequently, 1 possible approach would be to rely on human estimation of the food weight presented in these public data sets, which makes it difficult to define ground truth. However, at present, we have not implemented this method. Instead, the equivalent test and other statistical analyses were used to assess the accuracy of macronutrient estimation against measured weight as ground truth for the validation data set, which comprised 48 food items representing 20% of cumulative energy in the previous report [36]. This approach could strengthen the comparison of results.

When compared with other systems in the era of deep learning, where AI algorithms tend to be larger in size and more complex in their operations, our system still adheres to the traditional approach of image-assisted dietary assessment. In this system, all algorithms work independently, and fine-tuning is limited to each algorithm rather than the entire process.

Recently, numerous advanced techniques have been explored for food portion estimation, including deep learning–based volume estimation from single RGB images. The deep convolutional neural network—considered one of the key techniques in deep learning that is powerful in extracting features from images—has become a fundamental component in image classification and object detection algorithms. This technique was trained using RGB images with depth information (RGBD images). Subsequently, the system was able to predict food volume directly from RGB images. Although this method provides an error of ~50–100 mL for most foods, in some cases, the error exceeds 300 mL. Another carbohydrate counting system was introduced [50] that reported food-specific portions known as bread units (BUs), each defined as containing 12–15 g carbohydrates. This closely resembles the widely used carb unit in the diabetes field and is similar to the 1 exchange of starch group in Thai exchange system. The system was trained on an RGBD image data set containing human-annotated BU information. The state-of-the-art algorithm achieved an RMSE of 1.53 BU, whereas humans achieved an RMSE of 0.89 BU compared with the ground truth.

In addition to estimating the visible portion of food in an image, deep learning can also reconstruct the hidden back side of a food object that is not visible in the image owing to limited viewing angles [51,52]. This is achieved through a combination of deep learning and 3-dimensional reconstruction, using a developed point completion network. This system achieved a mean volume estimation error of 15.3% on real food. However, a limitation in applying this system is its difficulty in obtaining both image and 3-dimensional information of foods. Furthermore, deep learning techniques such as generative adversarial network, a backbone of image generator AIs like DALL·E, can be applied to directly estimate the energy content of food when trained with food images together with energy distribution mapping. This novel system has yielded a mean error of 209 kcal in the estimated energy per eating occasion [53].

However, implementation of these techniques may not be suitable in the Thai context for 2 interconnected reasons. First, deep learning requires a large data set for effective training. Second, acquiring a substantial image data set with additional information, such as food weight, object depth information, or energy distribution mapping, is costly and labor-intensive work.

Given these constraints, our current approach remains useful when users acknowledge system limitations, including taking food photographs in a landscape view, positioning the camera ~30 cm from the food, and capturing the image at an angle ranging from 30° to 90° from the tabletop. We also recommend separating food ingredients when possible. Although the data set includes several shapes of containers, we still recommend using a plate if possible. Other containers, such as bowls or foam boxes, resulting in food items being densely clustered in the same visual space, caused problems even for our human annotators in indicating the true area of the food objects, thus worsening the system's estimation accuracy.

INMU iFood primarily targets individuals concerned with a healthy lifestyle, with a primary focus on energy, protein, carbohydrate, fat, and sodium. However, in the current version, sodium content is not included owing to several reasons. Our approach primarily relies on detecting individual food ingredients within a food image rather than classifying the entire recipe. This has led to another important problem: all estimated nutrients must be derived from visible ingredients only. Although sodium is a significant concern for overconsumption by individuals, it mainly comes from fish sauce, soy sauce, shrimp paste, or salt added during cooking, which are invisible in an image. In contrast, other micronutrients are less concerning for individuals. Furthermore, in the current version of the Thai food composition database, only sodium, calcium, iron, thiamin, and riboflavin were analyzed for >80% of all food items. This necessitates caution in interpreting the results of other micronutrients. Consequently, we have decided to limit the results from INMU iFood to only macronutrients. When users require a comprehensive analysis of micronutrients, we recommend using the latest version of INMUCAL-Nutrients, our standard nutrient analysis system, instead of INMU iFood.

Although INMU iFood can recognize 400 food items from food images, this number may not suffice for everyday scenarios without the inclusion of substitution items. Hence, the iFood frontend incorporates the ability to import additional food products via a barcode scanner, with nutrition information sourced from Open Food Facts [54], a free, crowdsourced, open-source collaborative service that maintains a comprehensive food product database.

In Thailand, the exact number of food products that have consumption information contained in their nutritional labels consumptions is unknown. However, based on the report from the National Statistical Office, Ministry of Digital Economy and Society, the consumption rates for manufactured snacks, milks and their products, and functional food and nutritional supplements are 48.3%, 70.6%, and 21.6%, respectively [55]. Consequently, these products constitute a substantial portion of Thai people's diets. Chemically analyzed information on nutrition labels are required for snack foods, bakery products, semiprocessed foods (i.e., instant noodles, instant congees, and instant soups), chilled and frozen ready-to-eat meals, and beverages, including all milk products and seasoning sauces. Producers of other foods are only encouraged to provide this nutrition information [56]. It seems to be reasonable to use product-specific nutrition labeling information for tracking personal intake. Although macronutrients, saturated fat, dietary fibers, added sugar, and sodium are usually present in food products, the current version of INMU iFood supports only macronutrient content. We have manually added and edited food items available in Thailand continuously since 2019 until now. More than 3000 food products that are available in the Thai market are ready to use [57]. The labeling information is stored within this service, and the complete information can be accessed through the Open Food Facts Application Programming Interface (API). However, this method causes another issue as the system allows anyone to read, retrieve, or edit information on any products. To ensure data quality, the stored information undergoes bimonthly rechecks and updates on a regular basis.

INMU iFood represents an advancement in the field of dietary assessment. It offers a versatile and user-friendly tool for estimating the nutritional content of Thai cuisine from food images. Through the integration of state-of-the-art technology and a validated automated Thai food analysis system, it also offers a reliable means for users to monitor their dietary intake. Although it outperforms our previous system, it cannot work flawlessly without limitations. Nevertheless, INMU iFood has the potential to enhance dietary monitoring in the Thai context and contribute to improved health outcomes. INMU iFood also has the potential to serve as a model for other countries in developing similar tools covering their specific culinary food systems.

Acknowledgments

We thank the Institute of Nutrition, Mahidol University, for the support and use of their facilities.

Author contributions

The authors' responsibilities were as follows – PC, WK: designed the research; PC, AP, PS: conducted the research; PC, PS: analyzed the data; PC, AP, WK: wrote the paper; WK: had primary responsibility for final content; and all authors: read and approved the final manuscript.

Conflict of Interest

The authors report no conflicts of interest.

Funding

This work was funded by the Program Management Unit for Human Resources & Institutional Development, Research and Innovation (PMU-B) under contract number B04G640044. It is important to note that the supporting source, PMU-B, had no involvement in the research design, data collection, analysis, interpretation, or any restrictions concerning publication.

Data availability

Data described in the manuscript will be made available on request pending.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cdnut.2024.102154.

References

[1] F.E. Thompson, A.F. Subar, C.M. Loria, J.L. Reedy, T. Baranowski, Need for technological innovation in dietary assessment, J. Am. Diet. Assoc. 110 (2010) 48–51.

[2] R. Samasut, Thai food exchange list, Thai J. Parenter, Enter. Nutr. 15 (2004) 33–45.

[3] S. Likitmaskul, S. Reutrakul, S. Kongsomboonvech, P. Rawdaree (Eds.), Carbohydrate counting for health and diabetes control, 1st ed. Diabetes Association of Thailand, Bangkok, 2017, p. 75.

[4] P. Van Esterik, Food culture in Southeast Asia, Greenwood Publishing Group, 2008.

[5] D. Wahyudi, K. Daniel, Food culture of Southeast Asia: perspective of social science and food science, Kassel University Press, 2017, p. 124.

[6] D. Lansky, K.D. Brownell, Estimates of food quantity and calories: errors in self-report among obese patients, Am. J. Clin. Nutr. 35 (1982) 727–732.

[7] M.A. Zegman, Errors in food recording and calorie estimation: clinical and theoretical implications for obesity, Addict. Behav. 9 (1984) 347–350.

[8] M.B.E. Livingstone, P.J. Robson, J.M.W. Wallace, Issues in dietary intake assessment of children and adolescents, Br. J. Nutr. 92 (Suppl 2) (2004) S213–S222.

[9] B.L. Six, T.E. Schap, F.M. Zhu, A. Mariappan, M. Bosch, E.J. Delp, et al., Evidence-based development of a mobile telephone food record, J. Am. Diet, Assoc. 110 (2010) 74–79.

[10] T.A. Nicklas, C.E. O'Neil, J. Stuff, L.S. Goodell, Y. Liu, C.K. Martin, Validity and feasibility of a digital diet estimation method for use with preschool children: a pilot study, J. Nutr. Educ. Behav. 44 (2012) 618–623.

[11] C.K. Martin, T. Nicklas, B. Gunturk, J.B. Correa, H.R. Allen, C. Champagne, Measuring food intake with digital photography, J. Hum Nutr. Diet. 27 (2014) 72–81.

[12] D.-H. Wang, M. Kogashiwa, S. Kira, Development of a new instrument for evaluating individuals' dietary intakes, J. Am. Diet. Assoc. 106 (2006) 1588–1593.

[13] P. Pouladzadeh, S. Shirmohammadi, R. Almaghrabi, Measuring calorie and nutrition from food image, IEEE Trans. Instrum. Meas. 63 (2014) 1947–1956.

[14] F. Zhu, M. Bosch, I. Woo, S. Kim, C.J. Boushey, D.S. Ebert, et al., The use of mobile devices in aiding dietary assessment and evaluation, IEEE J. Sel. Top. Signal Process. 4 (2010) 756–766.

[15] W. Zhang, Q. Yu, B. Siddiquie, A. Divakaran, H. Sawhney, "Snap-n-Eat": food recognition and nutrition estimation on a smartphone, J. Diabetes Sci. Technol. 9 (2015) 525–533.

[16] M. Anthimopoulos, J. Dehais, S. Shevchik, B.H. Ransford, D. Duke, P. Diem, et al., Computer vision-based carbohydrate estimation for type 1 patients with diabetes using smartphones, J. Diabetes Sci. Technol. 9 (2015) 507–515.

[17] F. Zhu, M. Bosch, C.J. Boushey, E.J. Delp, An image analysis system for dietary assessment and evaluation, Proc. Int. Conf. Image Proc. (2010) 1853–1856.

[18] A. Doulah, M.A. Mccrory, J.A. Higgins, E. Sazonov, A Systematic review of technology-driven methodologies for estimation of energy intake, IEEE Access 7 (2019) 49653–49668.

[19] M.A. Subhi, S.H. Ali, M.A. Mohammed, Vision-based approaches for automatic food recognition and dietary assessment: a survey, IEEE Access 7 (2019) 35370–35381.

[20] W. Tay, B. Kaur, R. Quek, J. Lim, C.J. Henry, Current developments in digital quantitative volume estimation for the optimisation of dietary assessment, Nutrients 12 (2020) 1167.

[21] M. Wang, W. Leelapatra, A review of object detection based on convolutional neural networks and deep learning, Int. Sci. J. Eng. Technol. 6 (2022) 1–7.

[22] F.P.W. Lo, Y. Sun, J. Qiu, B. Lo, Image-based food classification and volume estimation for dietary assessment: a review, IEEE J. Biomed. Health Inform 24 (2020) 1926–1939.

[23] C. Termritthikun, P. Muneesawang, S. Kanprachar, NU-InNet: Thai food image recognition using convolutional neural networks on smartphone, J. Telecommun. Electron. Comput. Eng. 9 (2017) 63–67.

[24] C. Termritthikun, S. Kanprachar, Accuracy improvement of Thai food image recognition using deep convolutional neural networks, in: 2017 International Electrical Engineering Congress (iEECON), 2017, pp. 1–4. Krabi, Thailand.

[25] C. Termritthikun, S. Kanprachar, NU-ResNet: deep residual networks for thai food image recognition, J. Telecommun. Electron. Comput. Eng. 10 (2018) 29–33.

[26] N. Hnoohom, S. Yuenyong, Thai fast food image classification using deep learning, in: 2018 International ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI-NCON), 2018, pp. 116–119.

[27] B. Vishwakarma, M. Shinde, S. Oka, S. Nighut, D.P. Salapurkar, Categorization of Thai fast food based on image classification, IJIRMPS 7 (2019) 19–20.

[28] N. Tammachat, N. Pantuwong, Calories analysis of food intake using image recognition, in: 2014 6th International Conference on Information Technology and Electrical Engineering (ICITEE), 2014, pp. 1–4.

[29] P. Temdee, S. Uttama, Food recognition on smartphone using transfer learning of convolution neural network, in: 2017 Global Wireless Summit (GWS), 2017, pp. 132–135.

[30] K. Sukvichai, W. Muknumporn, Development of a food categories and calories estimation full stack system based on multi-CNNs structures, TNI J. Eng. Technol. 8 (2020) 38–49.

[31] P. Chotwanvirat, N. Hnoohom, N. Rojroongwasinkul, W. Kriengsinyos, Feasibility study of an automated carbohydrate estimation system using Thai food images in comparison with estimation by dietitians, Front. Nutr. 8 (2021) 732449.

[32] Onsen UI 2: Beautiful HTML5 Hybrid Mobile App Framework and Tools [Internet]. Onsen UI. [cited June 17, 2023]. Available from: https://onsen.io/.

[33] IndexedDB API—Web APIs [Internet], MDN (2023) [cited September 27, 2023]. Available from: https://developer.mozilla.org/en-US/docs/Web/API/IndexedDB_API.

[34] C.-Y. Wang, A. Bochkovskiy, H.-Y.M. Liao [Internet], YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors (2022) [cited September 27, 2023]. Available from: https://arxiv.org/abs/2207.02696v1.

[35] C. Rother, V. Kolmogorov, A. Blake, GrabCut—interactive foreground extraction using iterated graph cuts, ACM Trans. Graph. 23 (2004) 309–314, https://doi.org/10.1145/1015706.1015720.

[36] Y. Thewjitcharoen, P. Chotwanvirat, A. Jantawan, N. Siwasaranond, S. Saetung, H. Nimitphong, et al., Evaluation of dietary intakes and nutritional knowledge in Thai patients with type 2 diabetes mellitus, J. Diabetes Res. 2018 (2018) 9152910.

[37] C.E. Smart, B.R. King, P. McElduff, C.E. Collins, In children using intensive insulin therapy, a 20-g variation in carbohydrate amount significantly impacts on postprandial glycaemia, Diabet. Med. 29 (2012) e21–e24.

[38] D. Lakens, Equivalence tests, Soc. Psychol. Personal, Sci. 8 (2017) 355–362.

[39] GitHub—ultralytics/ultralytics: NEW—YOLOv8 🚀 in PyTorch > ONNX > OpenVINO > CoreML > TFLite [Internet]. [cited September 27, 2023]. Available from: https://github.com/ultralytics/ultralytics.

[40] C.K. Martin, H. Han, S.M. Coulon, H.R. Allen, C.M. Champagne, S.D. Anton, A novel method to remotely measure food intake of free-living individuals in real time: the remote food photography method, Br. J. Nutr. 101 (2009) 446–456.

[41] Y. He, C. Xu, N. Khanna, C.J. Boushey, E.J. Delp, Food image analysis: segmentation, identification and weight estimation, Proc (IEEE Int Conf Multimed Expo) 2013 (2013) 1–6.

[42] W. Jia, Y. Yue, J.D. Fernstrom, N. Yao, R.J. Sclabassi, M.H. Fernstrom, et al., Imaged based estimation of food volume using circular referents in dietary assessment, J. Food Eng. 109 (2012) 76–86.

[43] Y. Yue, W. Jia, M. Sun, Measurement of food volume based on single 2-D image without conventional camera calibration, Annu. Int. Conf. IEEE. Eng. Med. Biol. Soc. 2012 (2012) 2166–2169.

[44] E.H. Akpa, H. Suwa, Y. Arakawa, K. Yasumoto, Smartphone-based food weight and calorie estimation method for effective food journaling, SICE J. Control Meas. Syst. Integr. 10 (2017) 360–369.

[45] Y. Liang, J. Li, Deep learning-based food calorie estimation method in dietary assessment [Internet], arXiv:170604062 [cs] (2017) [cited August 23, 2018]. Available from: http://arxiv.org/abs/1706.04062.

[46] K. Okamoto, K. Yanai, An automatic calorie estimation system of food images on a smartphone [Internet], in: Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management—MADiMa '16, ACM Press, Amsterdam, Netherlands, 2016 [cited August 23, 2018]. Available from: http://dl.acm.org/citation.cfm?doid=2986035.2986040.

[47] T. Ege, W. Shimoda, K. Yanai, A new large-scale food image segmentation dataset and its application to food calorie estimation

based on grains of rice, in: Proceedings of the 5th International Workshop on Multimedia Assisted Dietary Management, Association for Computing Machinery, New York, NY, 2019, https://doi.org/10.1145/3347448.3357162.
[48] GemmyTheGeek, GemmyTheGeek/FoodyDudy: First Release of Foody Dudy [Internet], Zenodo, 2021 [cited September 29, 2023]. Available from: https://zenodo.org/record/5069090.
[49] Thai Food 20 Class x25 Dataset [Internet]. Roboflow. [cited September 29, 2023]. Available from: https://universe.roboflow.com/supapong-sakulkoo-dl3ze/thai-food-20-class-x25.
[50] P.F. Christ, S. Schlecht, F. Ettlinger, F. Grün, C. Heinle, S. Tatavatry, et al., Diabetes60—inferring bread units from food images using fully convolutional neural networks, in: 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), 2017, pp. 1526–1535.
[51] F.P.-W. Lo, Y. Sun, J. Qiu, B. Lo, A novel vision-based approach for dietary assessment using deep learning view synthesis, 2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN), Chicago, IL, USA (2019) 1–4. https://doi.org/10.1109/BSN.2019.8771089.
[52] F.P.-W. Lo, Y. Sun, J. Qiu, B. Lo, Food volume estimation based on deep learning view synthesis from a single depth map, Nutrients 10 (2018) 2005.
[53] S. Fang, Z. Shao, D.A. Kerr, C.J. Boushey, F. Zhu, An end-to-end image-based automatic food energy estimation technique based on learned energy distribution images: protocol and methodology, Nutrients 11 (2019) 877.
[54] Open Food Facts [Internet]. [cited June 15, 2023]. Available from: https://world.openfoodfacts.org/.
[55] National Statistical Office, Ministry of Digital Economy and Society, The 2017 food consumption behaviour survey, National Statistical Office, Ministry of Digital Economy and Society, Bangkok, 2018, p. 96.
[56] S. Ratanakorn, Food regulations and enforcement in Thailand, Reference Module in, Food Science (2016), https://doi.org/10.1016/B978-0-08-100596-5.21147-5.
[57] Open Food Facts—Thailand: Nutrition facts completed [Internet]. [cited September 28, 2023]. Available from: https://th-en.openfoodfacts.org/state/en:nutrition-facts-completed.