Article

Not peer-reviewed version

A Hybrid CNN–MLLM Architecture for Image-Based Nutrition Estimation and Advisory Insulin Decision Support in Type 1 Diabetes

Jean C. Velombe , Sema Bayraktar , Adnan Kavak * , Muhammad Jamil , Alpaslan B. Inner , Gautam Srivastava , Hossein Fotouhi *

Posted Date: 11 May 2026

doi: 10.20944/preprints202605.0623.v1

Keywords: type 1 diabetes; food image recognition; nutrition estimation; insulin bolus decision support; mobile health

 doi:10.20944/preprints202605.0623.v1

Article

A Hybrid CNN–MLLM Architecture for Image-Based Nutrition Estimation and Advisory Insulin Decision Support in Type 1 Diabetes

Jean C. Velombe [1,2], Sema Bayraktar [2], Adnan Kavak [1,2,*], Muhammad Jamil [1,2], Alpaslan B. Inner [1,2], Gautam Srivastava [3,4] and Hossein Fotouhi [5,*]

[1] Department of Computer Engineering, Kocaeli University, İzmit, Türkiye

[2] Wireless Information and Intelligent Systems (WINS) Research Center, Kocaeli University, Kocaeli, Türkiye

[3] Department of Computer Science and Engineering, Sogang University, Seoul 04107, Republic of Korea

[4] Research Center for Interneural Computing, China Medical University, Taichung 40402, Taiwan

[5] Department of Computer Science and Computer Engineering, Mälardalen University (MDU), Västerås 72123, Sweden

* Correspondence: hossein.fotouhi@mdu.se (H.F.); akavak@kocaeli.edu.tr (A.K)

Abstract

Accurate estimation of meal composition from food images can support safer and more reliable insulin bolus decision-making for individuals with Type 1 diabetes. Existing food recognition and nutrition estimation systems are often designed for general dietary logging and do not directly integrate food analysis with personalized insulin therapy parameters. This study presents an image-based nutrition estimation and insulin decision-support module developed within the AI-assisted Diabetes Care (AIDCARE) platform. The proposed system uses a convolutional neural network (CNN) to classify food items from a single meal image and retrieves reference nutritional values from a food composition database. A separate multimodal large language model (MLLM)-based estimation component is then used to estimate portion size, allowing carbohydrate and nutrient values to be scaled according to the observed serving. A curated food image dataset containing 40 food categories was used to evaluate three CNN architectures: ResNet50, Inception V3, and EfficientNet-B0. EfficientNet-B0 achieved the best classification performance, with 94.91% validation accuracy, 95.55% precision, 94.87% recall, and 94.90% F1-score. The portion-estimation component achieved an MAE of 12.27 g and an RMSE of 15.11 g. The estimated carbohydrate value is combined with user-specific clinical parameters, including the insulin-to-carbohydrate ratio and insulin sensitivity factor, to generate advisory bolus guidance. To support safety, the system requires user confirmation or correction of the recognized food category and estimated portion before insulin guidance is displayed. The proposed system is intended for advisory decision support only and is not designed to replace clinical judgment or autonomous insulin delivery systems.

Keywords: type 1 diabetes; food image recognition; nutrition estimation; insulin bolus decision support; mobile health

---

1. Introduction

People with Type 1 diabetes must monitor daily food intake and estimate appropriate meal-related insulin doses to maintain stable blood glucose levels [1,2]. In real-world settings, manual approaches such as visual portion estimation or food weighing are commonly used. However, these methods are time-consuming and prone to error because they depend on subjective judgment. Such errors may reduce the accuracy of insulin dosing and make blood glucose control more difficult [3–5]. In

 doi:10.20944/preprints202605.0623.v1

recent years, advances in artificial intelligence (AI), deep learning (DL), and computer vision (CV) have provided practical approaches for reducing human error in dietary assessment. These methods can support food recognition, nutrient estimation, and semi-automated dietary tracking from food images [6,7]. Some studies have investigated 3D reconstruction approaches for estimating food volume and nutrient content from food images [8]. However, reconstruction-based methods often require specialized hardware, depth information, reference objects, or multiple images captured from different angles. These requirements limit their practical use in everyday mobile health settings, especially for individuals who need fast and simple meal logging using a standard smartphone camera. Several AI-powered applications have also been developed for general calorie counting and nutrition logging. However, few available systems directly connect image-based food analysis with personalized insulin bolus decision support for individuals with Type 1 diabetes. Therefore, there remains a need for a mobile-friendly and clinically constrained system that can estimate food category, portion size, nutrient content, and advisory insulin guidance from a single meal image.

In this study, we propose an image-based nutrition estimation and insulin decision-support module within the AIDCARE platform. The system allows users to upload a single meal image, predicts the food category, estimates the portion size, retrieves reference nutritional values, and computes advisory bolus guidance using user-specific clinical parameters. The system incorporates a human-in-the-loop mechanism that allows users and healthcare professionals to review, confirm, or correct food and portion estimates before insulin guidance is displayed. This design preserves the advisory role of the system and supports safer decision-making in diabetes self-management.

$$\min_{\theta,\phi} \frac{1}{N} \sum_{i=1}^{N} \Big(L_{\text{cls}}(y_i, f_\theta(I_i)) + \lambda \, L_{\text{wt}}(w_i, h_\phi(I_i)) \Big) \tag{1}$$

Equation (1) sets out the main training objective for the single-image method. In this scenario, every example contains one RGB meal photo I_i together with the actual food label y_i and the portion weight w_i. The classifier $f_\theta(\cdot)$ helps in predicting the category while the regressor $h_\phi(\cdot)$ estimates the weight; both depend on the same input image. The factor λ facilitates balance the focus between classification accuracy and food portion estimation. In this case $f_\theta(\cdot)$ serves as the CNN backbone for identifying food types with parameters θ and utilizes a typical classification loss $\mathcal{L}_{\text{cls}}$. In the same way, $h_\phi(\cdot)$ tackles the portion or weight regression with parameters ϕ and calculates differences through $\mathcal{L}_{\text{wt}}$. The objective function mainly leads our proposed methodology to system design level and model choice. In our AI-assisted diabetes care (AIDCARE) solution, insulin recommendation is a sub-module and it is designed under clinician-defined safety constraints. To implement this research, we selected the most suitable classifiers like ResNet50, EfficientNet-B0, and Inception V3 and trained them using images captured by actual users. After finding the category of the food, our system provides carbohydrate content and other nutritional information and transforms them according to the estimated portion. The modified values contribute to bolus calculations that incorporate the individual's Insulin-to-Carbohydrate Ratio (ICR) and Insulin Sensitivity Factor (ISF). The yield of the system provide help to guide the end users (patients) and clinicians in making insulin dosing decisions. To safeguard well-being of the patients must always verify and either accept or adjust the identified food category and nutrient estimates before any insulin advice is given. This human-in-the-loop feature maintains user control over essential aspects. It should be noted that Eq. (1) expresses the conceptual joint objective of the proposed methodology. In the deployed AIDCARE implementation, however, food recognition and portion estimation are not realized as a single jointly trained network. Instead, the practical system follows a hybrid two-stage pipeline in which a CNN-based classifier predicts the food category and a separate multimodal model estimates portion size. Therefore, Eq. (1) should be interpreted as a high-level formulation of the coupled tasks rather than the exact optimization routine used in deployment. Based on our experimentation, EfficientNet-B0 attained the uppermost classification accuracy compared to the other DL models. Our proposed hybrid approach shows a practicable process for conducting nutrition analysis from food images and assisting the patients with

 doi:10.20944/preprints202605.0623.v1

bolus decision. The rest of the paper is organized in 6 main sections. In Section 2 we have explored the related work. Section 3 provides the complete architecture of the AIDCARE system and its main modules for better understanding our AIDCARE system. Section 4 presents the proposed research methodology. Section 5 reports the experimental results and analyzes performance. It also discusses the research implications and limitations. The last Section 6 concludes the paper.

2. Related Work

The central prerequisite in the daily self-management of Type 1 diabetes is the exact estimation of carbohydrate because meal-related insulin dosing rely straightforwardly on the amount of carbohydrate consumed [9]. On daily basis, many patients use manual carbohydrate counting or visual estimation of meal portions. Although these approaches are extensively used but user experience the fundamental dependance. These approaches are also exposed to substantial estimation errors, particularly for mixed dishes, unbalanced serving sizes, and meals made outside the home. As a result, the imprecise meal calculation may lead to under- or overestimation of insulin boluses and, consequently, inferior glycemic control. To overcome the problem of manual meal assessment, many researchers explored automatic dietary analysis using CV and DL [10,11]. Preliminary work in this area focused mainly on food image classification, where the meal images are mapped to one or more food categories. The advent of the convolutional neural network (CNN) has led to an improvement in food recognition performance, with ResNet, Inception, and EfficientNet being common food recognition networks for dietary assessment tasks [12,13]. These studies demonstrated that CNN-based models can successfully learn robust visual representations for food categories under challenging, realistic conditions near real-world environments with distracting clutter, similar and different classes, and changing lighting. However, food classification alonen isn't sufficient for clinical nutritional support because knowing the dish doesn't give the spent amount nor nutrient consumption directly.

The other key area of the research has thus concentrated on the estimation of the size of the portion or amount of food. Examiners used geometric modelling, fiducial markers, depth sensors, multi-view reconstruction or reference objects to obtain the quantity of food present in an image [14–16]. Many of these methods require controlled acquisition conditions, specialized hardware, and/or multiple images taken from various angles, but can also enhance quantitative estimation. These requirements slow them down in normal use for mobile health apps where users typically opt to take a one-time snapshot of a meal with their phone's camera. That is why, there remains solid attention in lightweight, single-image methodologies that can deliver appropriately precise portion estimates without particular sensing apparatus. More recently, multimodal models have been investigated for food-related reasoning works, including ingredient understanding, nutritional description, and approximate portion assessment from camera images [17]. Compared with purely CNN-based pipelines, multimodal large language models (MLLMs) provide superior flexibility in interpretation meal context and in generation of human-readable estimations. However, most existing applications of MLLMs in nutrition remain intensive on general dietary logging, calorie estimation, or explanatory meal analysis. Their role in clinically meaningful decision support, specifically for insulin dosing, has not yet been appropriately recognized. Parallel to advances in CV, a number of mobile and web-based nutrition apps have been designed to support dietary monitoring, calorie counting, meal logging, and personalized nutrition assistance [18,19]. These systems have shown worth in refining adherence, increasing nutrition awareness, and streamlining self-tracking. However, most of them are developed for general wellness, weight management, or broad diet management rather than for the requirements of people with Type 1 diabetes. Most of the systems stop at food recognition or nutrient estimation and do not incorporate meal analysis with user-specific insulin therapy parameters such as the insulin-to-carbohydrate ratio (ICR) and insulin sensitivity factor (ISF). So, they deliver valuable nutritional advice but do not directly support the meal-bolus decision procedure. There is also a significant difference between general dietary valuation systems and diabetes-oriented decision support systems.

 doi:10.20944/preprints202605.0623.v1

In the context of diabetes care, our goal is not just simply to estimate food calories or macronutrients present in the food, but also to provide a detailed outlook of meal-related information with personalized, and clinically explained guidance. For this purposed we linked food recognition and portion estimation with carbohydrate calculation and then connected these estimation to customized treatment parameters. Additionally, because errors in meal estimation can spread into insulin suggestions, such systems must incorporate suitable precautions, transparency, and user mistake. Existing techniques have only incompletely tackled this end-to-end issue, and very few researches have presented a practical mobile-friendly pipeline that combines image-based meal analysis with advisory insulin support under a human-in-the-loop technique. Against this context, our research work addresses a gap at the intersection of food image analysis, multimodal estimation, and Type 1 diabetes self-management. The proposed AIDCARE module does not treat food recognition as an isolated computer vision problem; instead, it implants image understanding into a broader decision support workflow. Particularly, our approach joins CNN-based food classification with single-image portion estimation using a multimodal model, fetches reference nutrient values from a food composition database, scales these calculations according to the estimated serving size, and finally computes an advisory bolus suggestion using user-specific clinical parameters. In contrast to the systems that rely on 3D reconstruction or multiple images, our approach is designed for Type 1 diabetes patients use with a single meal photo uploading freedom. At the same time, unlike traditional calorie-counting applications, our system extends beyond nutrition logging to support insulin-related decision making. Another unique feature of our system is its explicit human-in-the-loop feedback design. Rather than fully automating therapy decisions, the system needs nutrition validation or correction of the recognized food category and estimated portion before showing insulin guidance. The wider AIDCARE ecosystem includes clinician oversight through a dedicated professional dashboard, enabling healthcare professionals to describe individualized parameters and monitor usage. This makes the proposed approach more suitable for real-world adoption for the patients having diabetes Type 1.

3. System Overview and Hybrid AI Architecture

We developed the decision support system for image-based nutrition analysis as a central module in the AIDCARE platform; this broader digital health tool is designed to assist individuals in managing diabetes in their everyday lives. Within this methodology, our research emphasizes the assessment of dietary intake and guidance for insulin boluses. The system integrates food recognition from images with nutritional information tailored to users and relevant clinical parameters. Figure 1 illustrates the full architecture of this hybrid AI method along with its integration into the overall AIDCARE ecosystem.

3.1. Platform-Level System Architecture

The proposed system has four major modules including mobile app, web panel for healthcare professional, backend server and AI processing engine. All these modules are connected using REST APIs and exchange data with a central database for persistent storage and access. The mobile app is the developed for the patients having Type 1 diabetes and it is the point of interaction for end users. The user can upload single food image and this image is processed by AI engine for classification and nutrient estimation. Depending on the input image the application shows estimated nutrient breakdowns. The results are shared with the patients and once the user approve they can ask for bolus recommendation. The backend server acts as the main module between the other modules. It is responsible for user authentication, secure data exchange between modules, request routing, and for sending storage requests to the central database. The AI engine is responsible for core computational tasks such as food image category classification, portion size estimation, and calculating essential carbohydrate values. The results are used for bolus calculations. Mathematically:

$$\widehat{\pi}(I) = \mathrm{softmax}\big(f_\theta(I)\big) \tag{2}$$

 doi:10.20944/preprints202605.0623.v1

here I is the input image, $f_\theta(\cdot)$ represents the CNN class prediction function with parameters θ, and $\widehat{\boldsymbol{\pi}}(I)$ is the predicted probability distribution among all food categories. Here K denotes the number of food categories in the dataset, in our case, $K = 40$).

$$\widehat{y} = \arg\max_k \widehat{\pi}_k(I) \tag{3}$$

here $\widehat{y}$ is the predicted category with the highest probability among all possible classes.

$$\widehat{w} = h_\phi(I) \tag{4}$$

where $h_\phi(\cdot)$ denotes the portion-size estimation model and $\widehat{w}$ is the raw estimated food weight in grams derived from the input image.

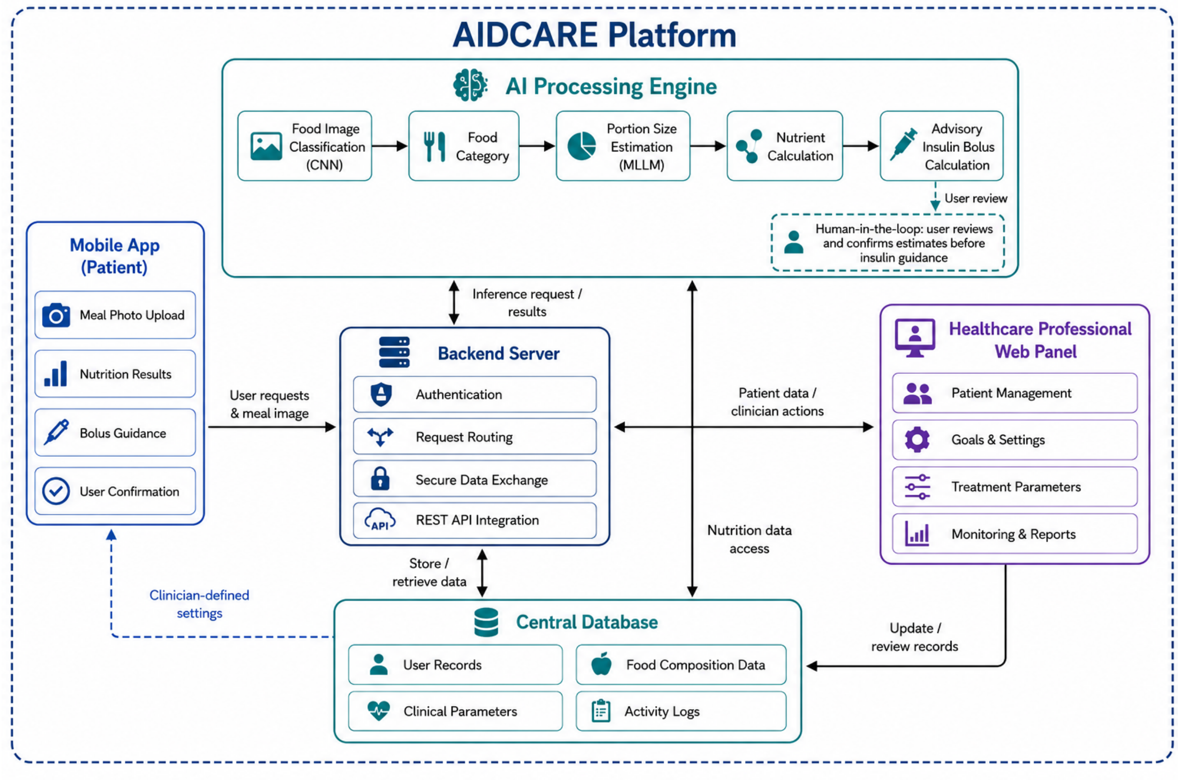

Figure 1. Overall architecture of the proposed system within the AIDCARE platform.

The component-based architecture makes it simple to connect with other parts of AIDCARE solution. The AIDCARE platform offers additional features including body measurements, exercise planing, diabetes-related education, gamification module and HbA1c risk prediction.

3.2. Diet Management Module and User Interaction Workflow

The diet management module forms the heart of our contribution and it is a module of AIDCARE solution. It permits the patients with Type 1 diabetes to document meals and obtain personalized guidance on insulin boluses derived from estimated carbohydrate intake. After uploading the image using smartphone camera, the AI engine first applies a CNN model (EfficientNet) to determine the food category. That classification then pulls corresponding reference nutritional data from an established food composition database.

$$y^* = C_y(\widehat{y}), \qquad w^* = C_w(\widehat{w}) \tag{5}$$

 doi:10.20944/preprints202605.0623.v1

where $C_y(\cdot)$ and $C_w(\cdot)$ denote the mandatory user confirmation operators for food category and portion size, respectively. During review, the user may either accept the predicted values or revise them, resulting in the final approved outputs y^* and w^*.

Next a multimodal analysis component processes the same image to predict portion size. Carbohydrate and other nutrient values are computed by adjusting the database entries according to this estimate.

The total carbohydrate estimate $\widehat{C}$ is computed by scaling database carbohydrate values using the confirmed portion size as defined in Eq. (8).

Finally these figures combine with the user's personal parameters notably the Insulin-to-Carbohydrate Ratio (ICR) and Insulin Sensitivity Factor (ISF) to produce bolus dosage suggestions. These remain strictly advisory to assist decision-making.

$$u_{\text{food}} = \frac{\widehat{C}}{\text{ICR}} \tag{6}$$

u_{food} is the insulin amount required to cover meal carbohydrates, $\widehat{C}$ is the estimated carbohydrate intake, and ICR is the insulin-to-carbohydrate ratio defined by clinicians.

For safety we built in a mandatory validation step: before any bolus advice appears users must review and either confirm or revise the detected food category and nutrient calculations. This human-in-the-loop design keeps users accountable for the most important details and reinforces the system's role as supportive rather than autonomous. The full sequence of interactions in the diet management module is shown in Figure 2.

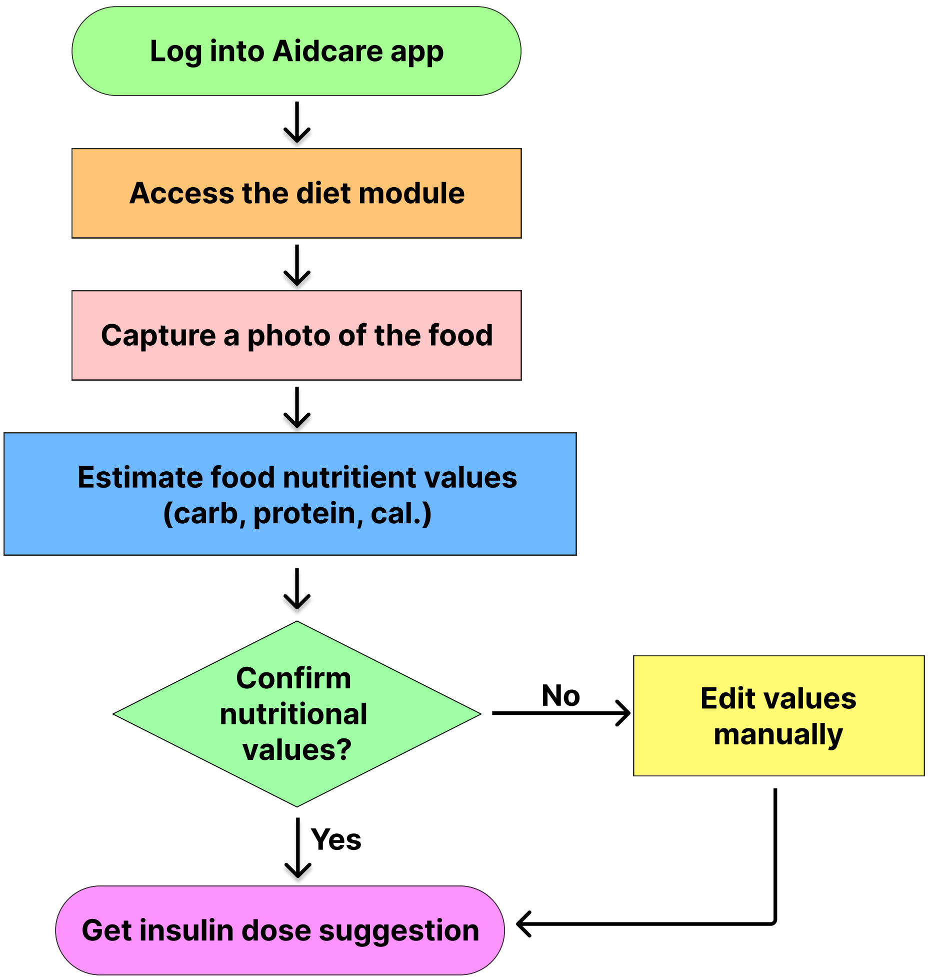

Figure 2. Workflow of the diet management module for image-based nutrition analysis and insulin bolus decision support.

doi:10.20944/preprints202605.0623.v1

3.3. Healthcare Professional Web Panel

In our AIDCARE solution, we made a separate web-based panel for healthcare professionals (dietitians) so that they can manage their patients, provide them some personalized treatment parameters, set blood glucose goals settings, and can track the overall progress of the patients. The web panel's main dashboard provides an outlook of key statistics for dietitian including critical alerts and upcoming appointments and online meetings with the patients. In the sidebar menu we added sections like My Patients, Appointments, Messages, AI-assisted Suggestions, Reports, and Online Meetings with the patients having Type 1 diabetes. The Goals & Settings pages (Figures 3, 4) allow healthcare professionals to list down personalized nutrition goals and treatment targets.

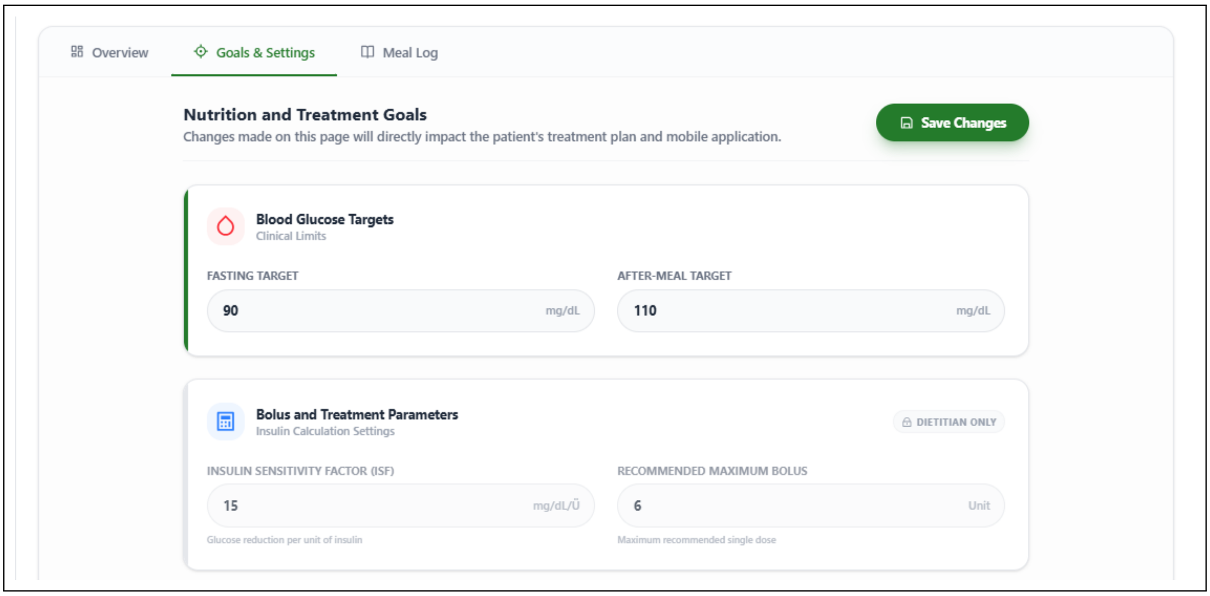

Figure 3. Blood glucose targets and bolus/treatment parameters.

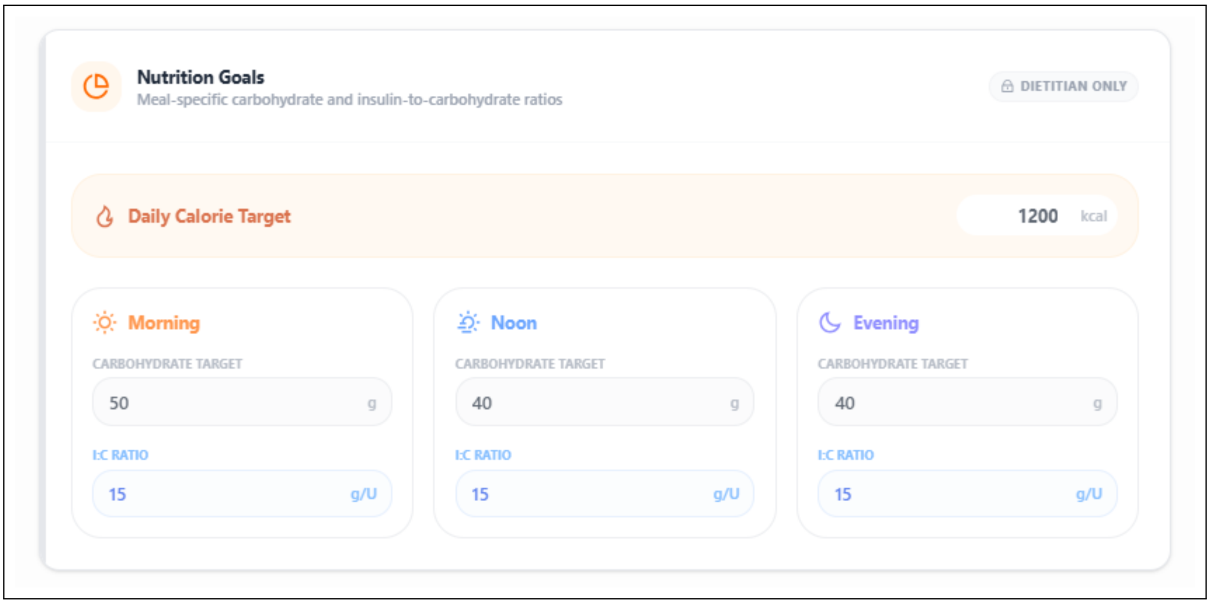

Figure 4. Meal-specific nutrition goals and insulin-to-carbohydrate ratios on the Goals & Settings page.

The patients can view on their mobile application once added by the healthcare professionals. Changes here will update the treatment plan in real time to ensure it matches the clinician's intent and the patient's experience. The app is directly linked to the web panel of the same healthcare

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 May 2026 doi:10.20944/preprints202605.0623.v1

professional, which makes it easy to adjust its parameters according to patient-specific data, activity level and glycemic variations obtained from the web panel. Figure 5 is the dashboard overview.

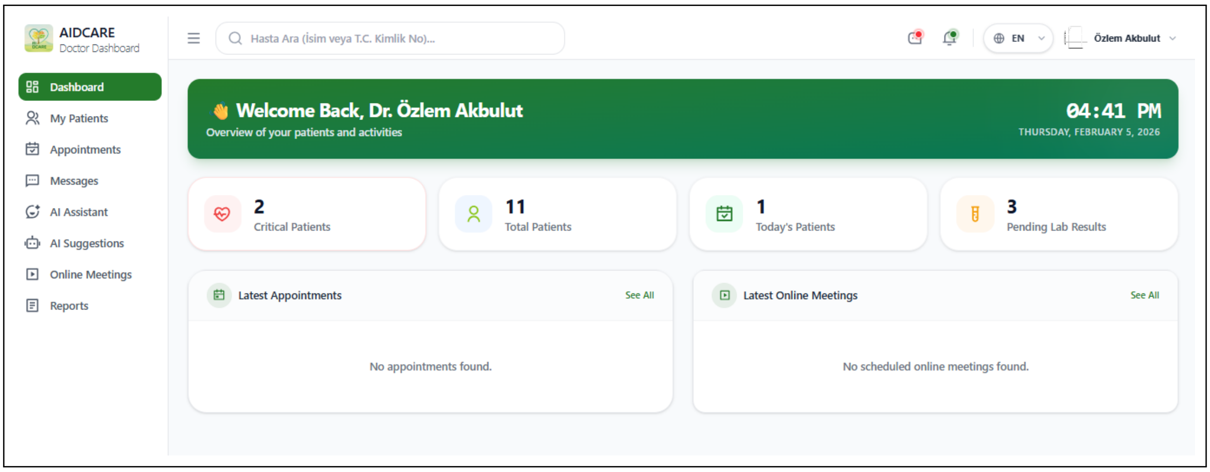

Figure 5. Overview of the AIDCARE clinician dashboard, showing sidebar navigation, patient statistics, and activity summaries.

4. Materials and Methods

4.1. Dataset and Pre-Processing

We assembled a custom dataset of high-resolution food images to train and evaluate the food classification models effectively in a Turkish food context. The dataset sources included publicly available recipe websites and real-world images captured by users to capture dishes commonly eaten in everyday Turkish diets. After careful curation and annotation, the final collection contains 40 distinct food categories that represent typical daily intake patterns. After data cleaning a total of 25,578 images were retained for training purposes. To address class imbalance a class balancing procedure was applied to create a more uniform training dataset. Each class was adjusted to approximately 350 images with a tolerance of $\pm$50 images.

The balanced dataset was randomly partitioned into training (80%) and validation (20%) sets using stratified sampling to preserve class distribution. Sample images from the dataset are illustrated in Figure 6.

Figure 6. Sample images from the food image dataset used for classification.

 doi:10.20944/preprints202605.0623.v1

To improve model generalization and robustness, we applied data augmentation technique with the help of Python's torchvision transforms library. This technique provided random horizontal and vertical flips, random rotations within $\pm 20^{\circ}$ angles, color jittering (contrast, brightness and saturation), and random resizing and cropping to an input resolution of 224×224 pixels.

All images present in the dataset were normalized with the help of ImageNet [20] statistics with mean values of 0.485, 0.456, 0.406 and standard deviations of 0.229, 0.224, 0.225 respectively. We applied this normalization step to make the input images compatible with the pre-trained CNN architectures (EfficientNet-B0, ResNet50, and Inception V3); all of which were originally trained on ImageNet. Matching the expected input distribution in this way enables more effective transfer learning; it allows the models to leverage learned features without significant distribution shift issues.

4.2. Food Image Classification Models

To determine the most suitable model for food category recognition in our Turkish-context dataset, we evaluated three well-established convolutional neural network (CNN) architectures: ResNet50 [21], EfficientNet-B0 [22], and Inception V3 [23]. These were chosen because they have repeatedly demonstrated strong performance on large-scale image classification benchmarks and have already seen successful use in various food image analysis tasks [24]. We applied transfer learning starting from weights pre-trained on ImageNet1K. During fine-tuning, the convolutional feature extraction layers remained frozen to preserve the general visual representations learned from ImageNet. Only the final fully connected classification layer was replaced and trained from scratch to fit our 40-class food recognition problem.

$$\mathcal{L}_{\text{cls}} = -\sum_{k=1}^{K} y_k \log \hat{\pi}_k \tag{7}$$

$\mathcal{L}_{\text{cls}}$ denotes the cross-entropy classification loss, K is the number of food classes, y_k is the ground-truth label, and $\hat{\pi}_k$ is the predicted probability for class k. All models used cross-entropy loss and were optimized with the Adam algorithm. We trained consistently with a starting learning rate of 0.001, batch size of 32, and 10 epochs. To prevent overfitting and improve convergence, a ReduceLROnPlateau scheduler dynamically reduced the learning rate whenever validation loss stopped improving.

4.2.1. ResNet50

ResNet50 [21] is a 50-layer deep convolutional network that relies on residual (shortcut) connections to address the challenges of training very deep architectures.

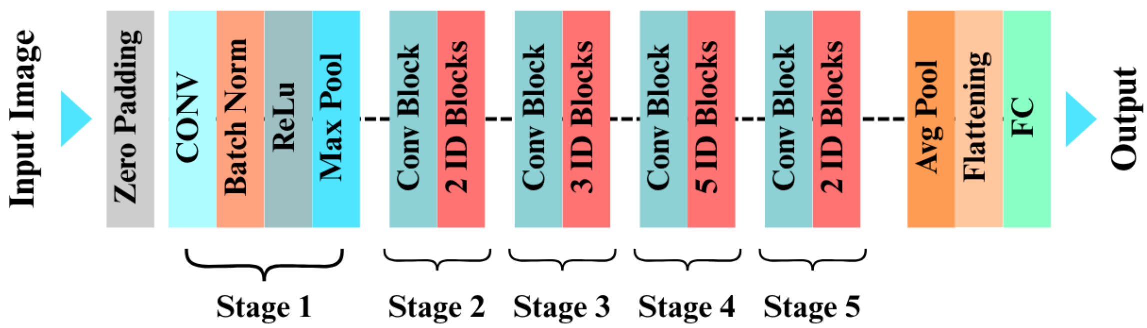

Figure 7. Architecture of the ResNet50 model.

4.2.2. EfficientNet-B0

EfficientNet-B0 [22] employs compound scaling to balance depth, width, and resolution, resulting in strong performance with low computational cost.

 doi:10.20944/preprints202605.0623.v1

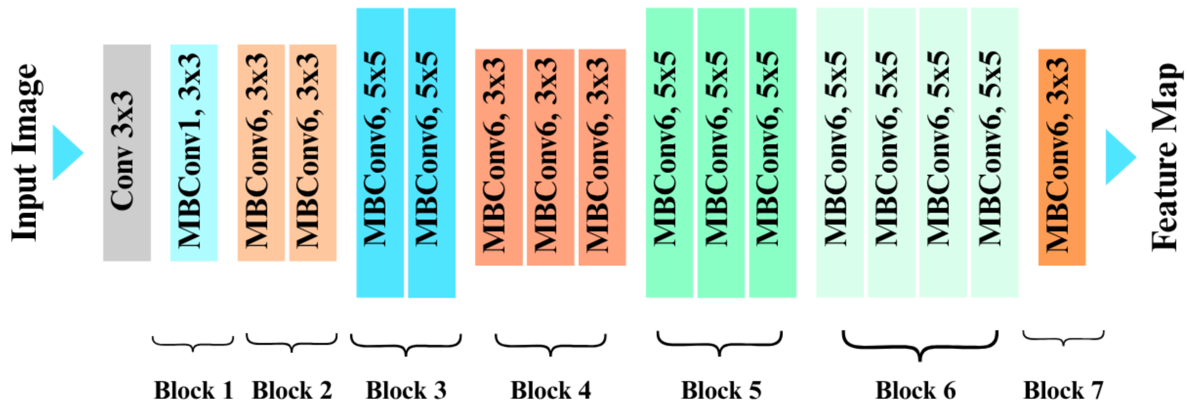

Figure 8. Architecture of the EfficientNet-B0 model.

4.2.3. Inception V3

Inception V3 [23] utilizes parallel convolutional branches to capture multi-scale spatial features efficiently.

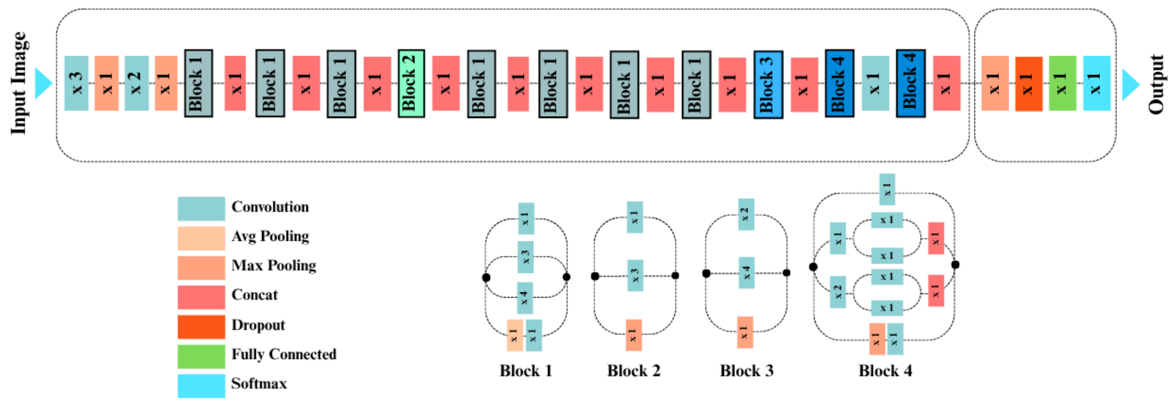

Figure 9. Architecture of the Inception V3 model.

4.3. Hybrid AI-Based Portion and Nutrient Estimation

Our hybrid approach combines CNN-driven food classification with a multimodal AI component to derive both portion size and nutritional estimates directly from a single food photograph, as illustrated in Figure 10.

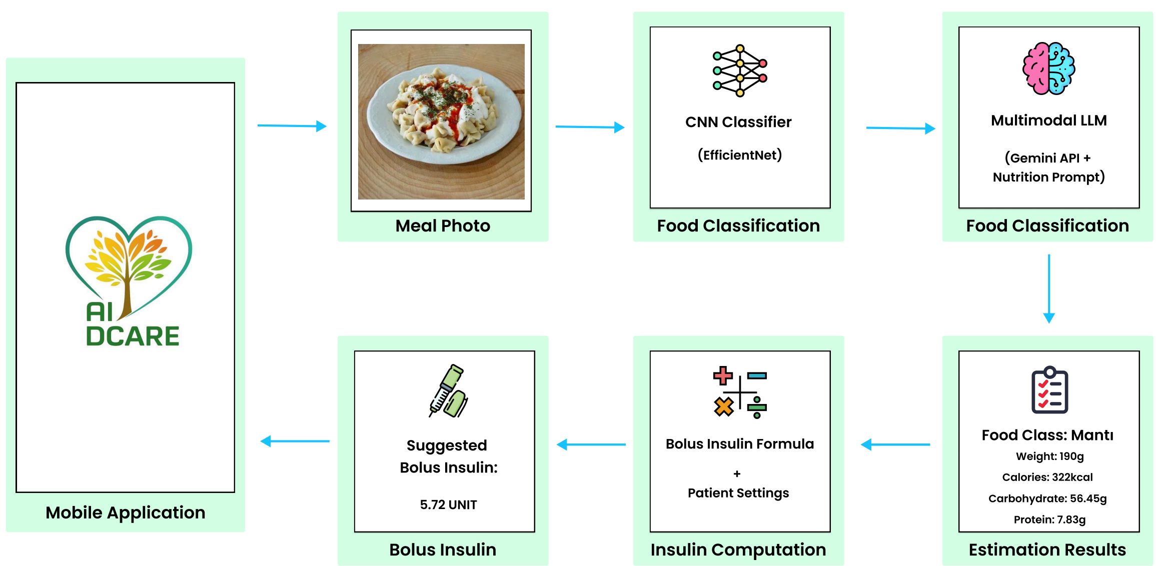

Figure 10. Pipeline of the hybrid nutrient estimation and insulin dosage guidance system.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 May 2026 doi:10.20944/preprints202605.0623.v1

Once the food category is identified, the portion size is estimated and used to scale reference nutritional values obtained from the food database.

$$\widehat{C} = \frac{w^*}{100} r_{\text{carb}}(y^*) \tag{8}$$

$$\Delta C_b = \frac{r_{\text{carb}}(y^*)}{100} \Delta w, \qquad |\Delta C_b| \leq \frac{r_{\text{carb}}(y^*)}{100} |\Delta w| \tag{9}$$

To make the effect of portion-estimation uncertainty explicit, we additionally consider the first-order propagation of weight error into carbohydrate estimation. The above relation shows that the carbohydrate error grows linearly with the food-specific carbohydrate density and the portion-estimation error. $\widehat{C}$ is the estimated total carbohydrate amount in grams, w^* is the confirmed portion weight, and $r_{\text{carb}}(y^*)$ is the carbohydrate value per 100 g for the confirmed food category.

Table 1. Example carbohydrate calculations using standard portion sizes for selected food categories.

Food category	Portion size (g)	Carb/100 g (g)	Total Carb/portion (g)
Mantı	190	29.71	56.45
Lahmacun	150	21.51	32.27
Dry beans	70	29.42	20.59
Yaprak sarma	130	17.81	23.16
Mercimek çorbası	300	8.28	24.85
Tarhana soup	250	4.99	12.47
Baklava	160	49.36	78.97
Künefe	120	45.75	54.90
Adana kebabı	151	1.06	1.60
Börek	200	35.83	71.67

4.4. Insulin Bolus Decision Support Calculations

Meal-related insulin is computed from the estimated carbohydrate intake as follows:

$$u_{\text{food}} = \frac{\widehat{C}}{\text{ICR}} \tag{10}$$

$$\Delta u_{\text{food}} = \frac{\Delta C_b}{ICR} = \frac{r_{\text{carb}}(y^*)}{100\, ICR} \Delta w \tag{11}$$

$$|\Delta u_{\text{food}}| \leq \frac{r_{\text{carb}}(y^*)}{100\, ICR} |\Delta w| \tag{12}$$

Combining Eq. (8) with the meal-bolus rule yields a direct sensitivity relation between portion-estimation error and meal-related insulin uncertainty. This formulation makes the safety interpretation of portion-estimation accuracy more transparent. Here u_{food} is the insulin dose for meal carbohydrates, $\widehat{C}$ is total carbohydrates in grams, and ICR is the insulin-to-carbohydrate ratio. Correction insulin is calculated to adjust for deviations from target glucose:

$$u_{\text{corr}} = \max\left(0, \frac{\text{CBG} - \text{TBG}}{\text{ISF}}\right) \tag{13}$$

where CBG is the current blood glucose, TBG is the target blood glucose, and ISF is the insulin sensitivity factor. The $\max(0, \cdot)$ term prevents negative correction insulin values when CBG $<$ TBG.

$$\mathcal{W}(CBG) = \begin{cases} 1, & CBG < TBG \\ 0, & CBG \geq TBG \end{cases} \tag{14}$$

Since Eq. (13) suppresses negative correction insulin, it is useful to define an explicit low-glucose warning indicator. This preserves the advisory character of the system while making the safety logic mathematically explicit. The total advisory insulin bolus is then:

$$u_{\text{total}} = u_{\text{food}} + u_{\text{corr}} \tag{15}$$

$$u_{\text{safe}} = \max\Big(0, \min\{u_{\max}, u_{\text{food}} + u_{\text{corr}}\}\Big) \tag{16}$$

To reflect clinician-defined treatment limits more explicitly, the advisory bolus can also be written in constrained form, where $u_{\max}$ denotes the maximum recommended bolus configured in the clinical settings panel.

u_{total} represents the total insulin bolus suggested by the system before user confirmation.

4.5. Evaluation Protocol

Food classification performance was evaluated using accuracy, precision, recall, and macro-averaged F1-score. The balanced dataset was split into training and validation sets using stratified sampling with an 80%/20% ratio to preserve class distribution. For the multi-class classification setting, precision, recall, and F1-score were computed using a one-vs-rest strategy and then macro-averaged across all food categories. Portion-weight estimation performance was evaluated using mean absolute error (MAE) and root mean squared error (RMSE). All experiments were conducted in a GPU-accelerated environment using the same training configuration across the evaluated CNN architectures to ensure a consistent comparison.

5. Results and Discussion

5.1. Food Classification Performance

For the food classification task, we evaluated three convolutional neural network (CNN) architectures: ResNet50 [21], EfficientNet-B0 [22], and Inception V3 [23]. The exact category recognition is an essential step towards reliable carbohydrate estimation. We fine-tuned all the models using transfer learning. To quantitatively evaluate classification performance, we used standard metrics; for the multi-class setting ($K = 40$), precision, recall, and F1-score are computed in a one-vs-rest manner and reported as macro-averages across classes.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively, and Accuracy measures the overall proportion of correctly classified samples.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{18}$$

Precision indicates how many of the predicted food categories are correct.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{19}$$

Recall measures how many of the true food categories are successfully identified by the model.

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{20}$$

F1-score is the harmonic mean of precision and recall, and is a balanced performance measure. The best overall performance was achieved by EfficientNet-B0, which achieved a validation accuracy of 94.91%. It is competitive with others because of the joint optimization of the depth, width and resolution of the compound, which enables a very efficient detection of the discriminative features even for a relatively small number of parameters. The model was quite reliable for many categories such as *Lahmacun* that are clearly visible. When compared to the following, Inception V3 was the next closest with 91.67% validation accuracy. The Inception modules in its multi-scale approach were found to be effective in extracting detailed textures and the spatial context of the image. It performed well on the small number of food items with distinctive appearance, such as *Adana kebabı* and poorly on items with similar appearance such as *Mantı*. ResNet50 achieved 87.96% validation accuracy and it was only comparatively lower in its recognition of visually homogeneous items like *Tarhana soup* as it were showed stable performance. The comparison of accuracy, precision, recall and F1-score between the two are shown in Table 2. The alternatives were always outperformed by efficientNet-B0, which was the backbone of choice for downstream nutrient estimation and bolus support.

Table 2. Food classification performance of different CNN models.

Model	**Accuracy** ↑	**Precision** ↑	**Recall** ↑	**F1-score** ↑
ResNet50	87.96%	85.48%	84.20%	84.27%
Inception V3	91.67%	91.92%	91.62%	91.63%
EfficientNet-B0	**94.91%**	**95.55%**	**94.87%**	**94.90%**

5.2. Gemini-Based Food Weight Estimation

We evaluated the Gemini multimodal large language model (MM-LLM) for food portion weight estimation using mean absolute error (MAE) and root mean squared error (RMSE), defined as:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|w_i - \widehat{w}_i| \tag{21}$$

w_i is the true food weight, $\widehat{w}_i$ is the estimated weight, and N is the number of test samples.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(w_i - \widehat{w}_i)^2} \tag{22}$$

RMSE penalizes larger estimation errors more strongly by squaring the difference between true and estimated weights. Across the test set, the model achieved an MAE of 12.27 g and an RMSE of 15.11 g, indicating that most estimates were within a practically acceptable range for real-world smartphone images. Figure 11 shows that estimation accuracy remains stable across varying dataset sizes, while Figure 12 illustrates a symmetric error distribution centered near zero. These results highlight the inherent uncertainty of single-image portion estimation and justify the system's mandatory user confirmation step.

doi:10.20944/preprints202605.0623.v1

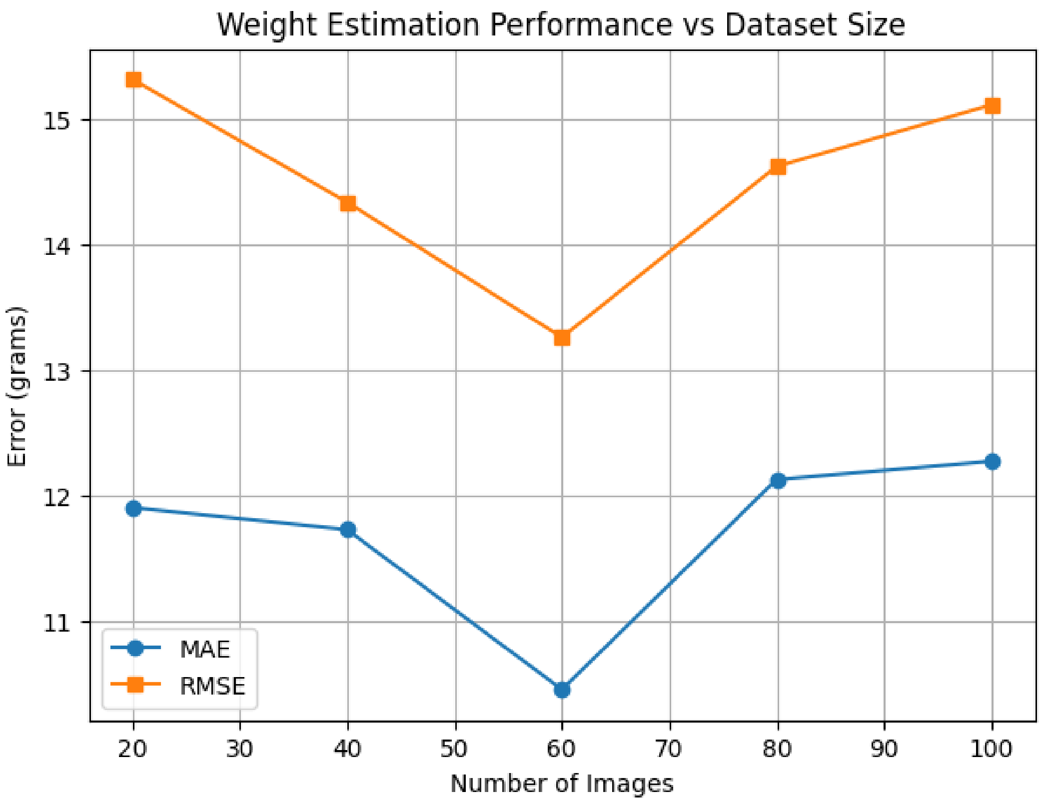

Figure 11. Weight estimation performance versus dataset size.

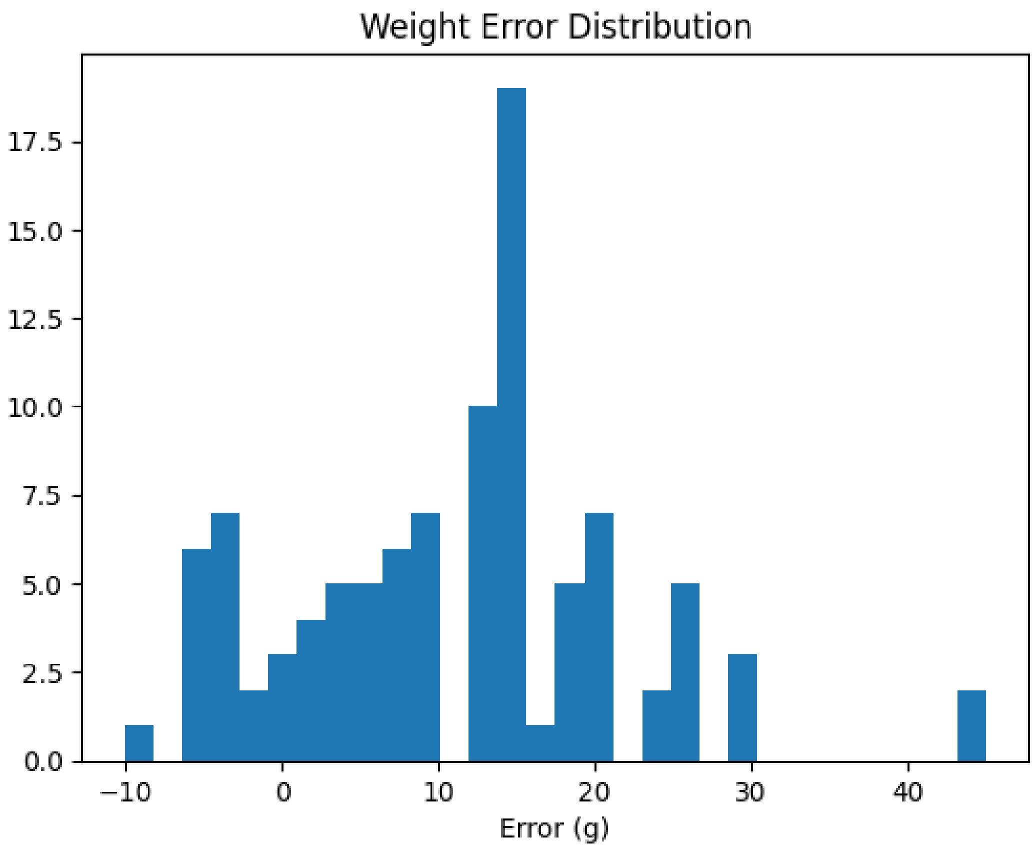

Figure 12. Weight estimation error distribution.

 doi:10.20944/preprints202605.0623.v1

5.3. System Deployment and End-to-End Evaluation on the AIDCARE Platform

The complete AI-assisted pipeline was integrated into the AIDCARE platform and evaluated under realistic usage contexts. From an end-to-end perspective, the final insulin recommendation can be explained as a function of estimated carbohydrates and clinical parameters:

$$u_{\text{total}} = \frac{\widehat{C}}{\text{ICR}} + \max\left(0, \frac{\text{CBG} - \text{TBG}}{\text{ISF}}\right) \tag{23}$$

$\widehat{C}$ is the estimated carbohydrate amount, ICR is the insulin-to-carbohydrate ratio, CBG is current blood glucose, TBG is target blood glucose, and ISF is the insulin sensitivity factor. Figures 13 and 14 demonstrate the deployed workflow. In the illustrated example, a meal of *Mantı* was correctly classified, its portion estimated at 190 g, resulting in 56.45 g of carbohydrates and a suggested bolus of 5.72 units. The mandatory user confirmation step ensures that all insulin guidance remains advisory and under user and clinician control.

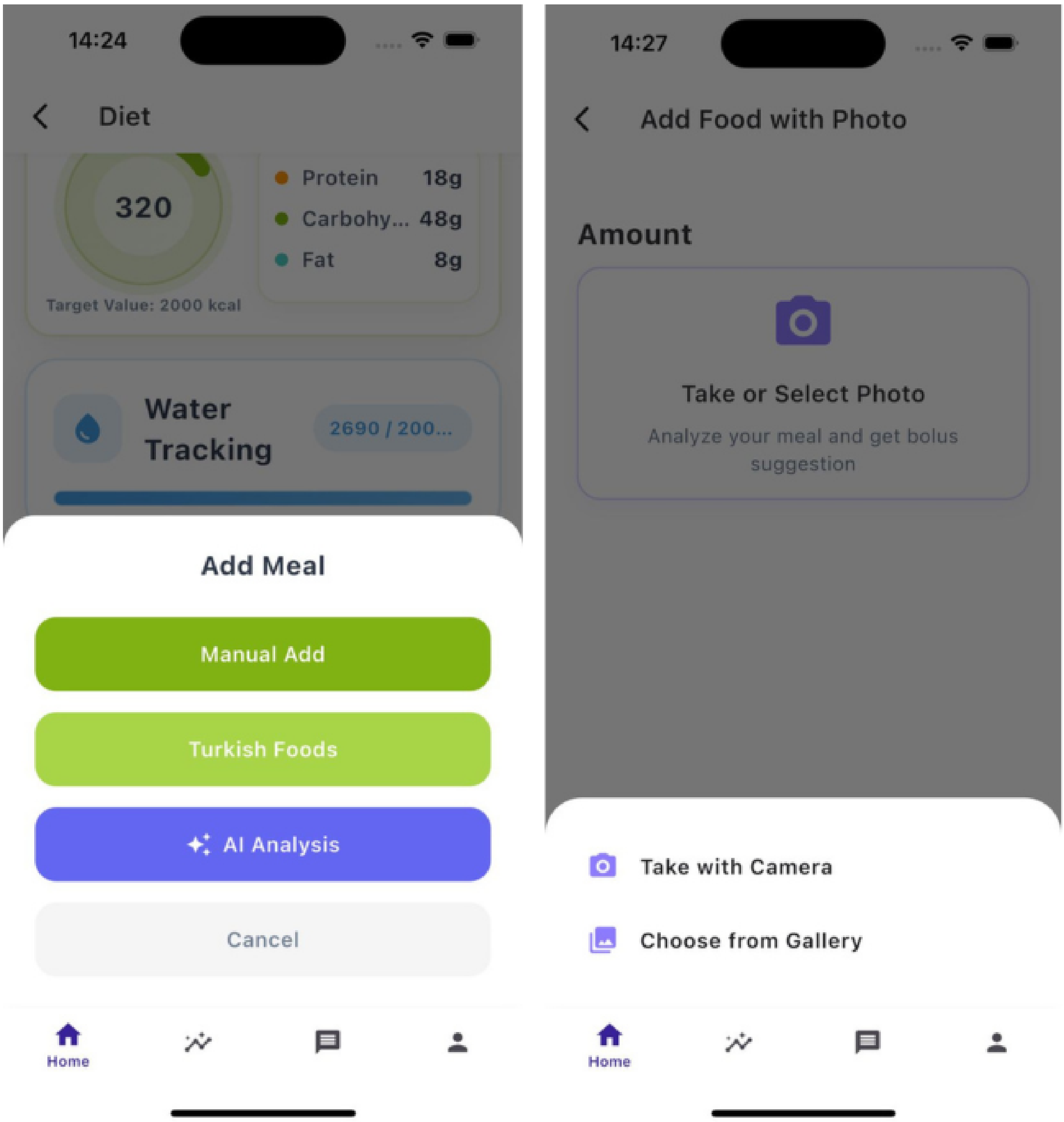

Figure 13. Food photo upload interface: manual or AI-based entry options and camera- or gallery-based food image input.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 May 2026 doi:10.20944/preprints202605.0623.v1

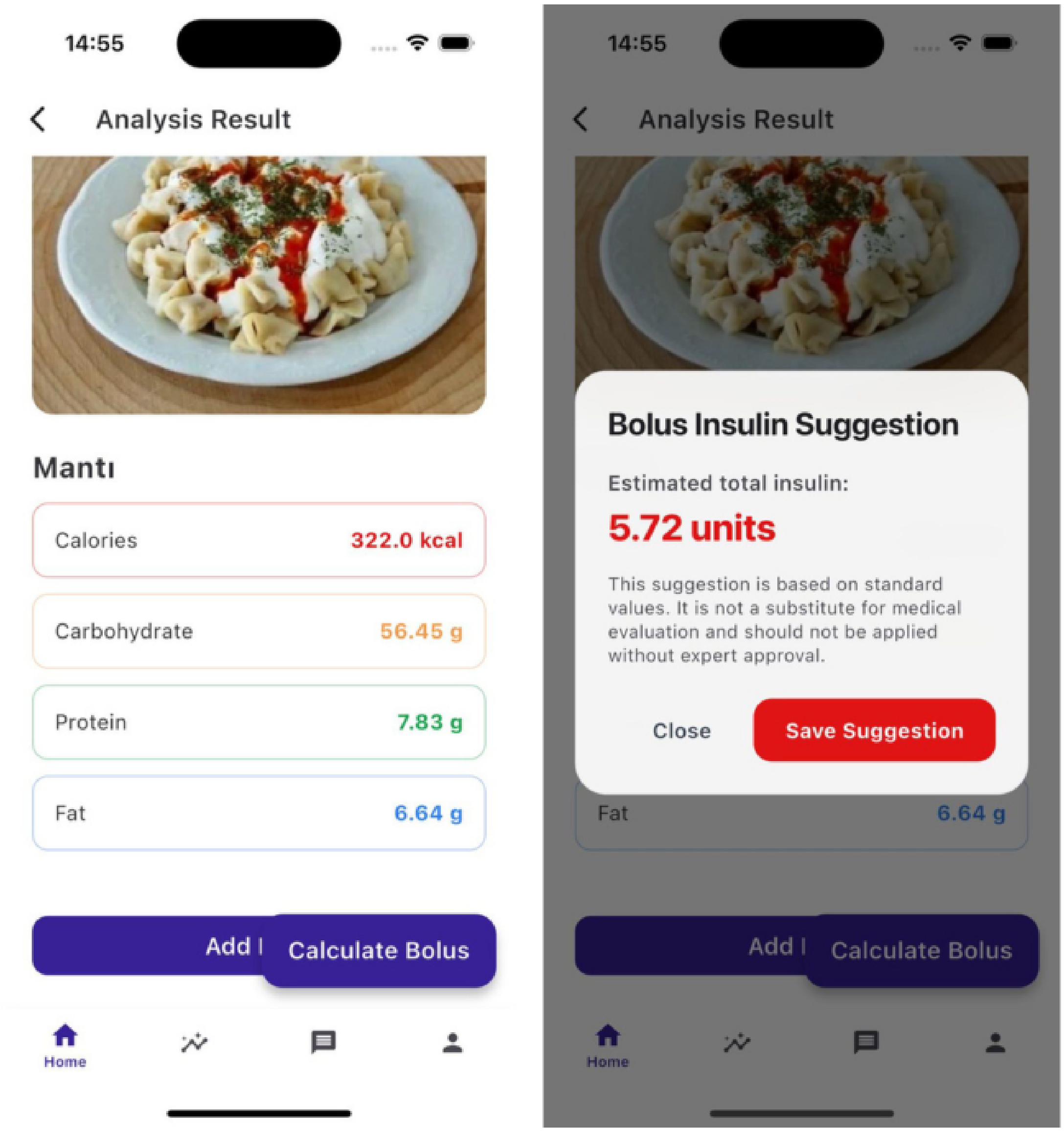

Figure 14. Food recognition, nutrient estimation, and insulin bolus decision-support results.

5.4. Discussion

Our AI hybrid decision support system combined image-based food category recognition, estimation of nutrients, and personalized insulin bolus guidance for self-management of patients having diabetes. The investigational outlook indicate that DL models can classify meal categories more accurately using a single image of the food. This method provided a concrete initial point for precise carbohydrate count and insulin guidance. Among the classification DL models of we tried, EfficientNet-B0 provided the finest scores with 94.91% validation accuracy, 95.55% precision, 94.87% recall, and 94.90% F1-score. As compared to ResNet50 and Inception V3, EfficientNet-B0 also provided better results at lower computational cost. After recognizing food categories, portion sizes were estimated using MLLM (Gemini Image API). A significant attention for any image-based system in diabetes care is how carbohydrate estimate errors spread to insulin dosing. In our dataset, a 15g weight error characteristically interpreted to 1—5g of carbohydrate deviance changing by food thickness. For food substances with moderate carb (around 20g carbohydrates per 100g), a 15g portion miscalculate leads to approximately 3g extra or missing carbs. Supposing a normal Insulin-to-Carbohydrate Ratio (ICR) of 1:10g/U, this would cause an insulin adjustment of about 0.3 units. In a clinical environment, such small differences typically fall well within the boundaries of safe margins for bolus. To additional protect users, AIDCARE system entails active review and possible correction of the detected food and portion estimation before displaying any insulin recommendation. This human-in-the-loop module

 doi:10.20944/preprints202605.0623.v1

shadows well-known unsurpassed practices for AI-assisted clinical decision support and support to mitigate the impact of infrequent errors. Notwithstanding the strengths presented by these procedures, there are some limitations. The portion approximation presently uses a general-purpose multimodal model and task-specific fine-tuning on annotated portion datasets. While technical feasibility and low error rates are promising, the true test will come from larger-scale clinical studies examining long-term effects on HbA1c, hypoglycemia rates, and sustained user engagement. From a mathematical modeling perspective, the present bolus formulation should be interpreted as a first-order advisory model. It does not explicitly account for insulin-on-board, physical activity, delayed fat/protein effects, glucose trend information, or other dynamic physiological factors. Therefore, the proposed equations are intended to provide an explainable and clinically constrained decision-support rule rather than a fully autonomous treatment model.

6. Conclusion

We proposed an applied AI decision support system developed particularly for image-based nutrition inquiry and insulin bolus management to help peoples with routine diabetes self-monitoring and self-management. Our proposed hybrid system brought together CNN-based food recognition, and vision-based volume estimation. The system provides dependable carbohydrate calculations based on image uploaded using mobile phone. It doesn't depend on resource-intensive 3D reconstruction or on multiple images. Such features make the system ideal for smart healthcare mobile applications specially in diverse nutritional environments such as Turkish cuisine. We tested multiple DL models and among the classification models we tested, EfficientNet-B0 provided the best results with 94.91% validation accuracy, 95.55% precision, 94.87% recall, and 94.90% F1-score. After recognizing food categories, portion sizes were estimated using MLLM (Gemini Image API). This method produced carbohydrate estimations that were precise enough with MAE around 10–15g. We send all the information to the nutritionist dashboard to cope with human-in-the-loop mechanism to keep the bolus guidance purely advisory and to ensure it compliance with clinical standards for decision support rather than fully automated processes.

Funding: This research was funded by the Excellence in Production Research Framework through XPRES (Excellence in Production Research), and by R2Microgrid project under the RESILIENT competence center, financed by the Swedish Energy Agency and co-financed by Mälardalen University and industrial partners.

Data Availability Statement: AI tools (Gemini Pro and Grammarly) were used only to improve grammar, spelling, and language clarity. No part of the research design, analysis, results, or scientific content was generated by these tools. All technical and intellectual contributions in this manuscript were made by the authors.

Conflicts of Interest: The authors declare that they have no conflicts of interest related to this work.

Abbreviations

The following abbreviations are used in this manuscript:

AIDCARE	AI-assisted Diabetes Care
AI	Artificial Intelligence
CNN	Convolutional Neural Network
CV	Computer Vision
DL	Deep Learning
ICR	Insulin-to-Carbohydrate Ratio
ISF	Insulin Sensitivity Factor
MAE	Mean Absolute Error
MLLM	Multimodal Large Language Model
MoH	Ministry of Health
RMSE	Root Mean Squared Error

 doi:10.20944/preprints202605.0623.v1

References

1. Wallace, T.; Heath, J.; Koebbel, C. The impact of flash glucose monitoring on adults with type 1 diabetes' eating habits and relationship with food. *Diabetes Research and Clinical Practice* **2023**, *196*, 110230.
2. Schneider, S.; Barber, T.M.; Saemann, M.; Thurston, J. Is there anything left to eat? A lived experience insight of following a restrictive type 1 diabetes and dialysis diet. *Expert Review of Endocrinology & Metabolism* **2025**, pp. 1–14.
3. Noaro, G.; Zhu, T.; Cappon, G.; Facchinetti, A.; Georgiou, P. A personalized and adaptive insulin bolus calculator based on double deep q-learning to improve type 1 diabetes management. *IEEE Journal of Biomedical and Health Informatics* **2023**, *27*, 2536–2544.
4. Annuzzi, G.; Apicella, A.; Arpaia, P.; Bozzetto, L.; Criscuolo, S.; De Benedetto, E.; Pesola, M.; Prevete, R. Exploring nutritional influence on blood glucose forecasting for type 1 diabetes using explainable AI. *IEEE journal of biomedical and health informatics* **2023**, *28*, 3123–3133.
5. Krishnamoorthy, D.; Doyle III, F.J. Safe and personalized meal bolus calculator for type-1 diabetes using bayesian optimization. *IEEE Transactions on Biomedical Engineering* **2022**, *70*, 1481–1492.
6. Bidyalakshmi, T.; Jyoti, B.; Mansuri, S.M.; Srivastava, A.; Mohapatra, D.; Kalnar, Y.B.; Narsaiah, K.; Indore, N. Application of artificial intelligence in food processing: Current status and future prospects. *Food Engineering Reviews* **2025**, *17*, 27–54.
7. Quan, W.; Zhou, J.; Wang, J.; Huang, J.; Du, L. Machine Learning-Driven Precision Nutrition: A Paradigm Evolution in Dietary Assessment and Intervention. *Nutrients* **2025**, *18*, 45.
8. Ma, P.; Hong, H.C.; Jia, X.; Chi, C.J.; Xiao, N.; Fan, B.; Wang, F.; Wei, C.I. Structure from motion-convolutional neural network model (SfM-CNN) achieved accurate portable Chinese dietary chemical composition estimation for dietary recall. *Food Chemistry* **2025**, *489*, 144908.
9. Sterner Isaksson, S.; Hellman, J.; Wijkman, M.; Imberg, H.; Ólafsdóttir, A.F.; Axelsen, M.; Lind, M. Diet and diet-related challenges with specific focus on carbohydrates and carbohydrate counting in adults with type 1 diabetes: a cross-sectional study. *BMJ open* **2025**, *15*, e101619.
10. Khawate, S.; Gaikwad, S.; Davda, Y.; Shirbhate, R.; Gham, P.; Borate, V. Dietary Monitoring with Deep Learning and Computer Vision. In Proceedings of the 2025 International Conference on Computing Technologies & Data Communication (ICCTDC). IEEE, 2025, pp. 1–5.
11. Saad, A.M.; Rahi, M.R.H.; Islam, M.M.; Rabbani, G. Diet engine: A real-time food nutrition assistant system for personalized dietary guidance. *Food Chemistry Advances* **2025**, *7*, 100978.
12. Alam, I.N.; Junior, F.A.; Susanto, R. Transfer learning approach with EfficientNet to enhance food recognition systems for health monitoring. *Commun. Math. Biol. Neurosci.* **2025**, *2025*, Article–ID.
13. Motghare, M.; Joshi, V.K.; Jamal, M.; Shanmugam, M.; Patil, R.V.; Rawool, N.J. Leveraging Deep Learning for Nutritional Profiling: A Comparative Analysis of EfficientNetB0, MobileNetV2, and ResNet50 for Healthy and Unhealthy Food Classification. In Proceedings of the 2025 IEEE International Conference on Recent Advances in Computing and Systems (REACS). IEEE, 2025, pp. 1–10.
14. Raju, V.B.; Sazonov, E. FOODCAM: a novel structured light-stereo imaging system for food portion size estimation. *Sensors* **2022**, *22*, 3300.
15. Li, B. Food Portion Analysis for Nutritional Monitoring and Dietary Assessment. PhD thesis, University of Pittsburgh, 2025.
16. Chopra, M.; Purwar, A. A synergistic hybrid approach for calorie estimation using multi-view food images and deep learning-based segmentation and classification. *Multimedia Tools and Applications* **2026**, *85*, 143.
17. Qi, H.; Zhu, B.; Ngo, C.W.; Chen, J.; Lim, E.P. Advancing food nutrition estimation via visual-ingredient feature fusion. In Proceedings of the Proceedings of the 2025 International Conference on Multimedia Retrieval, 2025, pp. 1091–1099.
18. Phalle, A.; Gokhale, D. Navigating next-gen nutrition care using artificial intelligence-assisted dietary assessment tools—a scoping review of potential applications. *Frontiers in Nutrition* **2025**, *12*, 1518466.
19. Zhao, M. Enabling Better Healthy Food Recommendation By Incorporating Contextual Knowledge. PhD thesis, University of Sheffield, 2025.
20. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009, pp. 248–255.
21. Senapati, B.; Talburt, J.R.; Naeem, A.B.; Batthula, V.J.R. Transfer learning based models for food detection using ResNet-50. In Proceedings of the 2023 IEEE International Conference on Electro Information Technology (eIT). IEEE, 2023, pp. 224–229.

22. Tan, M.; Le, Q.V. EfficientNet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning (ICML), 2019.
23. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception architecture for computer vision. *arXiv preprint arXiv:1512.00567* **2015**.
24. Chen, J.; Zhu, B.; Ngo, C.W.; Chua, T.S.; Jiang, Y.G. A Study of Multi-Task and Region-Wise Deep Learning for Food Ingredient Recognition. *IEEE Transactions on Image Processing* **2021**, *30*, 1514–1526. https://doi.org/10.1109/TIP.2020.3045639.