# Comparing the Accuracy and Efficiency of Existing AI Based Food Detection Tools

UNIVERSITY OF TURKU
Department of Computing
Master of Science (Tech) Thesis
Software Engineering
October 2025
Shyam Bhetuwal

UNIVERSITY OF TURKU
Department of Computing

SHYAM BHETUWAL: Comparing the Accuracy and Efficiency of Existing AI Based Food Detection Tools

Master of Science (Tech) Thesis, 122 p., 7 app. p.
Software Engineering
October 2025

---

The food industry is evolving more rapidly as well as dynamically and the consciousness of consumers on what they are eating is increasing. Current food detection and nutrients analysis platforms exhibit significant variability in accuracy, reliability, user experience and integration capabilities.

There is lack of adequate standardized evaluation and comprehensive comparative analysis, leading to limited insights and guidance for different use cases. This research has aimed to fill this gap by conducting a detailed analysis of the existing AI based models and platforms and integrating sensors and re-correction logic in the application. The accurate detection and evaluation of food items and their nutritional content is critical for applications in healthcare, dietary management and the entire food industry. Reliance on a single data modality like only on the (rgb image) limits the detection for similarly looking food items or the portion sizes. This can reduce the effectiveness of dietary monitoring tools. While general purpose AI models like OpenAI vision have multimodal capabilities even good in food detection shows varying levels of accuracy and efficiency with AI assisted food detection platforms like Calorie Mama AI,FoodSAM, LLaVA and LogMeal.

Despite advances in AI there is a lack of accurate evaluation and comparison of these algorithms, particularly regarding their performance with diverse datasets and real-time images. Additionally, integrating these AI algorithms into a user-friendly interface to process the live images and present data effectively can have significant software engineering challenges. Different users group (researchers, developers and end users) require a robust, efficient and intuitive system that can handle the real-time data processing ensuring accurate food detection and nutritional evaluation, re-correct the data and export the data for further analysis and even automated system for the data analysis process. Moreover, while individual algorithms can perform well either in detection, in segmentation or in analysis in the controlled environments, integrating them all into a user-friendly interface for the real application has many additional challenges.

This study has developed an AIoT based integrated system as a solution with camera, scale and APIs from different image detection models to display the most accurate result. Then the accuracies from different sources has been analyzed.

Keywords: Machine Vision, AI, AIoT, Machine Learning, React, API, Redux, Python, FastAPI, WebSocket, OpenAI Vision API, LogMeal API, FoodSAM, SAM, CNN, Deep Learning, R-CNN , YOLO, LLaVA, Data Analysis, MAE, RMSE, MAPE, Food Detection, Nutrient Analysis

TURUN YLIOPISTO
Tietotekniikan laitos

SHYAM BHETUWAL: Comparing the Accuracy and Efficiency of Existing AI Based Food Detection Tools

Diplomityö, 122 s., 7 liites.
Department of Computing
Lokakuu 2025

---

Elintarvikeala on muuttumassa yhä kilpailullisemmaksi ja dynaamisemmaksi, ja kuluttajien tietoisuus siitä, mitä he syövät, kasvaa jatkuvasti. Nykyiset ruoan tunnistus- ja ravinneanalyysialustat vaihtelevat huomattavasti tarkkuuden, luotettavuuden, käyttökokemuksen ja integrointikyvykkyyden osalta. Alan kattavan standardoidun arvioinnin sekä vertailevan analyysin puute rajoittaa käyttäjien saamaa tietoa ja ohjausta. Tämä tutkimus pyrkii täyttämään kyseisen aukon suorittamalla yksityiskohtaisen analyysin olemassa olevista tekoälypohjaisista alustoista, standardoimalla arviointimittarit ja tarjoamalla käytännön suosituksia näiden teknologioiden tehokkuuden ja käytettävyyden parantamiseksi. Ruoan tarkka tunnistus ja ravintosisällön arviointi ovat kriittisiä sovelluksissa, kuten terveydenhuollossa, ruokavalion hallinnassa ja elintarviketeollisuudessa.
Yhteen datamuotoon, kuten pelkkään RGB-kuvaan, tukeutuminen rajoittaa tunnistustarkkuutta erityisesti samannäköisten ruokien tai annoskokojen arvioinnissa. Vaikka yleiskäyttöisillä tekoälymalleilla, kuten OpenAI Visionilla, on multimodaalisia kykyjä, erityisesti ruoan tunnistukseen suunnatut tekoälyavusteiset alustat, kuten Calorie Mama AI, LLaVA ja LogMeal, eroavat merkittävästi tarkkuuden, suorituskyvyn ja datan kattavuuden osalta. Vaikka tekoälyn kehitys on ollut nopeaa, näiden algoritmien kattava arviointi ja vertailu, erityisesti monipuolisilla tietoaineistoilla ja reaaliaikaisilla kuvilla, on edelleen puutteellista.Lisäksi tekoälyalgoritmien integroiminen käyttäjäystävälliseen käyttöliittymään, joka kykenee käsittelemään live-kuvia ja esittämään tiedot tehokkaasti, aiheuttaa merkittäviä ohjelmistoteknisiä haasteita. Käyttäjät (tutkijat, kehittäjät ja loppukäyttäjät) tarvitsevat vankan, tehokkaan ja intuitiivisen järjestelmän, joka pystyy käsittelemään reaaliaikaista dataa, tunnistamaan ruoat tarkasti, arvioimaan ravintosisällön ja viemään tiedot jatkoanalyysiä varten. Vaikka yksittäiset algoritmit voivat toimia hyvin joko tunnistuksessa, segmentoinnissa tai analyysissä kontrolloiduissa olosuhteissa, niiden yhdistäminen käyttäjäystävälliseksi käyttöliittymäksi todellisia sovelluksia varten tuo mukanaan lisähaasteita.
Tässä tutkimuksessa kehitetään AIoT-pohjainen integroitu järjestelmä, joka hyödyntää kameraa, vaakaa ja eri kuvantunnistusmallien rajapintoja (API), tuottaen mahdollisimman tarkan tuloksen. Tämän jälkeen analysoidaan eri lähteistä saatujen tulosten tarkkuudet.

Asiasanat: Konenäkö, Tekoäly, AIoT, Koneoppiminen, React, API, Redux, Python, FastAPI, WebSocket, OpenAI Vision API, LogMeal API, FoodSAM, SAM, CNN, Syväoppiminen, R-CNN, YOLO, LLaVA, Data-analyysi, MAE, RMSE, MAPE, Ruoantunnistus, Ravinneanalyysi

# Contents

# Abbreviations

| | |
|---|---|
| **AI** | Artificial Intelligence |
| **ML** | Machine Learning |
| **SLR** | Systematic Literature Review |
| **PRISMA** | Preferred Reporting Items for Systematic Reviews and Meta-Analyses |
| **GPT** | Generative Pre-trained Transformer |
| **AIoT** | Artificial Intelligence of Things |
| **IoU** | Intersection over Union |
| **LLaVA** | Large Language and Vision Assistant |
| **IBDA** | Image-Based Dietary Assessment |
| **NLP** | Natural Language Processing |
| **NMS** | Non-Maximum Suppression |
| **CNN** | Convolutional Neural Network |
| **R-CNN** | Region-based Convolutional Neural Network |
| **Faster R-CNN** | Faster Region-based Convolutional Neural Network |
| **SAM** | Segment Anything Model |
| **FoodSAM** | Food Segment Anything Model |
| **YOLO** | You Only Look Once |
| **MobileNetV2** | Mobile Network Version 2 |
| **API** | Application Programming Interface |
| **IoT** | Internet of Things |
| **RMSE** | Root Mean Square Error |
| **MAE** | Mean Absolute Error |
| **MAP** | Mean Absolute Percentage |
| **MAPE** | Mean Absolute Percentage Error |

| | |
|---|---|
| **AWS** | Amazon Web Services |
| **UEC-256** | University of Electro-Communications Food Dataset with 256 Categories |
| **USDA** | United States Department of Agriculture |
| **WFR** | Weighed Food Records |
| **GPU** | Graphics Processing Unit |

# 1 Introduction

## 1.1 Background

Recent studies have provided significant insights into the nutrients present in food and their impact on human health. Latest technology-based dietary assessment methods offer rapid feedback, greater engagement especially for younger users and cost-effective scalability compared to traditional paper-based approaches [1]. Some notable areas of previous research includes the impact of quality of Nutrients on health, quality of nutrients based on versatile food dishes, food related environmental sustainability and AI based technologies in food recognition and detection.

### Nutritional Quality and Health Benefits

Research has highlighted the nutritional quality of various food groups, including fruits, vegetables, grains, and dairy products and the health benefits associated with these foods, such as reduced risk of chronic diseases, improved mental health, and overall well-being [2], [3]. Consumption of balanced Nutrition as Harvard's plate model has direct positive impact on overall health.Over the past few decades, diet quality has worsened, impacting health inspired by Harvard's balanced plate model, this paper presents an innovative image-based system to assess meal healthiness [2]. Long-term adherence to healthy dietary patterns are associated with significantly greater odds of healthy aging, including better cognitive, physical and mental health outcomes [3].

### Diet Quality and Nutritional Status

Specific research has focused on the diet quality and nutritional status of different dietary groups, such as vegetarians, vegans, and omnivores [4]. These studies explore how different diets affect nutrient intake and health outcomes.Food authentication verifies the composition and origin of foods, ensuring accurate assessment of diet quality and nutritional status and Analytical methods confirm what is truly in the food we consume, directly enhancing the reliability of nutritional evaluations [5]. Higher diet-quality scores (like HEI, AHEI, Mediterranean scores) were significantly associated with Lower BMI, waist circumference, and waist-to-height ratio (i.e., less adiposity). Higher blood levels of carotenoids and omega-3 fatty acids (nutritional biomarkers) [6]. While vegetarian diets may lead to deficiencies in certain nutrients (e.g., vitamin B12, calcium), they also offer potential health benefits such as lower risks of obesity, cardiovascular diseases, and certain cancers [3], [7].

### Environmental Sustainability

There is growing interest in the environmental sustainability of food production and consumption. Research in this area examines how dietary choices impact the environment and explores strategies to promote sustainable eating practices [7]. Environmental sustainability in cafeteria food management involves minimizing food waste through smart tracking and consumer feedback systems, which can reduce bio-waste and lessen the environmental impact of food overproduction and disposal [8]. For instance, studies have shown that animal-based foods, particularly red meat, contribute disproportionately to greenhouse gas emissions, land use, and water consumption compared to plant-based alternatives [9], [10]. Furthermore, the selection of life cycle assessment (LCA) databases used to evaluate food footprints can significantly affect the estimation of environmental impacts, which challenges the comparability and validity of sustainability assessments across regions [10]. Despite these differences, consistent findings across studies indicate that reducing meat intake is a key lever for promoting environmentally sustainable diets [11].

### AI in Food detection and analysis

Researching the nutrients present in food is one of the most researched topic in recent days. It addresses critical aspects of public health, dietetics and environmental sustainability. The topic's relevance is enhanced by increasing public interest in healthy eating, sustainability, and the rise of artificial intelligence technology in nutrition research. Recent articles [12], [13] reviews how AI improves food safety using tools like spectroscopy, imaging, and sensors. It also explores combining AI with IoT and blockchain to boost traceability and transparency in the food supply chain. This paper reviews the integration of AI into nutritional assessment, highlighting its ability to analyze complex dietary data while addressing challenges such as data privacy, ethical considerations, and the need for robust validation [12], [14].

## 1.2 Problem Statement

The food industry is becoming more competitive and dynamic, and consumers' are more aware of how nutrients in food impacts overall health. Current food detection platforms are not able to solve the challenges in nutrient science completely lacking standard evaluation and comparative analysis. This research aims to fill this gap by conducting a detailed analysis of existing AI-based platforms, standardizing the evaluation metrics, and providing actionable insights to enhance the effectiveness and usability of these technologies.

The accurate detection and evaluation of food items and their nutritional content are critical for applications in healthcare, dietary management, and the food industry. A systematic review by Amoutzopoulos et al. [15] of current existing tool showed that only few studies were done with quality. The study identified 542 portion size estimation elements (PSEEs) and found that the accuracy of dietary data is heavily influenced by the type of tool used. Reliance on a single data modality (e.g., only RGB images) limits detection for similarly looking food items or portion sizes. This can reduce the effectiveness of dietary monitoring tools. Existing food detection algorithms provided by AI tools such as OpenAI Vision API, Calorie Mama AI, LLaVA, FoodSAM, LogMeal and similar other vary significantly in terms of accuracy, efficiency and data comprehensiveness. Despite advances in AI there is a lack of comprehensive evaluation and comparison of these algorithms regarding their performance with diverse datasets and real-time images. [15]

Furthermore, integrating heterogeneous AI models developed using different architectures and training datasets into a unified, integrated user-friendly interface for real-time image's metadata analysis poses major challenges in data integration, data mapping, and error-correction mechanisms. End users, researchers, and developers need a system that functions effectively in real-world conditions, ensures accurate recognition and detection and automates information correction and analysis to deliver the most reliable results. However, while the individual AI models perform well either in detection, segmentation or analysis under controlled conditions and datasets, integrating these models with different architectures into a unified, real-time and user-friendly interface is a big challenge. This research addresses these integration challenges and aims to develop a framework that enhances interoperability, computational efficiency, and accuracy in AI-based food detection systems.

However, different models exhibit varying strengths across different domains some perform better in recognition, some in detection, some in pixel-level segmentation under controlled datasets and conditions. Integrating these models with different architectures into a unified, real-time and user-friendly interface remains as a significant challenge. This research addresses these integration challenges and aims to develop a system that enhances interoperability, computational efficiency and accuracy in AI-based food detection systems.

An Artificial Intelligence of Things (AIoT) based system is being developed using a camera, digital scale and APIs from various image detection and segmentation tools to identify and analyze food items in the University Of Turku Software Engineering Lab as a part of Flavoria [8] project known as Flavoria Flex. Part of the work presented in this thesis has been **submitted and accepted** as a peer-reviewed conference paper at the **2025 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT)**, titled as **"Automated Image Recognition System for Determining Energy Composition of Meals by AI-Powered Detection and Identification of Food Items - A Study Utilizing Flavoria Flex"**. I started this Research and Development project from the beginning days in February 2024 and I am the first author of the published paper. I was primarily responsible for researching and developing the AI-based detection pipelines, re-correction pipelines, data extraction, conducting the data analysis and preparing the manuscript of the paper by the support of my supervisors and team. The methodology and results presented in the paper are also elaborated in this thesis, as they form a core component of my research. The contributions of the co-authors are gratefully acknowledged, particularly in the areas such as scalable modular system development, nutritional science concepts and supervisory support from project researchers and senior researchers. [16]

## 1.3 Research Questions

This thesis has aimed to address the problem statement and answer the below research questions (RQs):

1. **Comparative Accuracy of Food Recognition Algorithms**

RQ 1.1: How do the accuracies of image recognition algorithms, including food-specific models such as LogMeal API and FoodSAM and general-purpose image recognition models such as the OpenAI Vision compare in identifying various dishes against the ground truth names?

RQ 1.2: How accurate are the total weights of the dishes estimated by each algorithm and estimated by the system compared to the ground truth weight?

RQ 1.3: How accurate are the total energy and macro nutrients of the dish compared to the ground truth?

RQ 1.4: How consistent are the measurements across different weights of the food dishes?

RQ 1.5: What is the possibility of conducting individual component based analysis based on the available test data?

2. **Data Integration and Enhanced Accuracy**

RQ 2.1: How effective is the integration of a machine vision system with a weighing scale in improving nutrient estimation accuracy compared to direct image-based methods (e.g., LogMeal, OpenAI Vision)?

RQ 2.2: How accurate are the nutrient estimates from the machine vision system integrated with a weighing scale when re-calibrated against the ground truth menu-based information?

3. **Applications and Implications**

RQ 3.1: How can the findings of this research be applied to develop more accurate and user-friendly food tracking applications?

RQ 3.2: What implications do these findings have for nutrition research and AI-assisted dietary assessment?

## 1.4 Significance of the Study

This research has contributed to both the nutrition science and software engineering research by providing a framework to integrate several AI algorithms into a single application. It has also offered the re-correction logic for direct AI based detection data. This system has provided real time data extraction and analysis. Valuable insights are obtained about the usability and accuracy of food detection supporting the developers, researchers and professionals in AI, Nutrition and healthcare.

## 1.5 Structure of the Thesis

This thesis is structured into seven chapters. Chapter 1, introduces the study by outlining the background, problem statement, research questions, and the significance of the research. Chapter 2, presents a comprehensive review of the relevant literature, focusing on recent technological advancements related to the research topic, and provides insights into how these developments can help address existing challenges. Chapter 3, discusses the case studies formulated to investigate the research questions and objectives, and also reviews prior work on integrating sustainability into the agile development process. Chapter 4 details the methodologies and systems developed to conduct the experimental studies. Chapter 5, presents the results of the experiments, offering both qualitative and quantitative analyses. Chapter 6 draws conclusions from the study, linking the findings to the existing literature and research context. Finally, Chapter 7 serves as the discussion chapter, addressing the research questions, identifying potential sources of error, and recommendations for future research.

## 1.6 Declaration of Generative AI

Artificial intelligence tools has been integrated in our system for data collection and re-correction to support the writing of this thesis.

**OpenAI Vision** (GPT-4 with image input capabilities) was directly used for detecting, identifying, and re-correcting dish items in the images as part of the food detection system. It also helped to classify food items.

**ChatGPT** has been used to assist with sentence refinement, idea clarification as a learning tool, translation and understanding the language to English(for example translation of Finnish food names, translation of English into Finnish as well), grammar and writing style improvements. It also served to simplify the understanding of advanced technical concepts throughout the project, validated from highly relevant academic publications.

Example prompts used with these AI tools are documented in Appendix C. AI tools were used to refine my own writing and sentence structures. The refined outputs from the AI tools were reviewed, validated, and edited by me. All research plans, implementation methodologies, analysis, results, conclusions, and further research ideas presented in this thesis are entirely my own.

# 2 Literature Review

Food recognition algorithms have evolved significantly over the years, with advancements in digital technologies, health-related consumer behaviors, machine learning and computer vision. Zhou et al. [17] proposed a knowledge-enhanced feature synthesizer to improve zero-shot food detection, enabling models to identify unseen food categories by using information from multiple sources like ingredients relationships, food attributes or visual features about food characteristics to detect food items better. Existing algorithms in food detection and nutrition analysis are based on models trained on large data sets of preprocessed images. Understanding the complexity of food systems and their impact on dietary health highlights the importance of accurate food detection technologies to support nutrition monitoring and policy development. [7], [17]

## 2.1 Overview of Food Recognition Algorithms

The algorithms are used in applications for identifying food, getting nutritional information for dietary assessment, personalized nutrition tracking and health monitoring. Current techniques used in food recognition include object detection, image segmentation and analysis based on trained algorithms [18]. Izbassar and Shamoi [2] developed an image-based system using segmentation and classification to access meals against Harvard's Healthy Eating Plate. The framework offers personalized recommendations and helped in healthy heating.

### 2.1.1 Deep Learning Techniques for Food Detection

In food detection and nutritional analysis several advanced deep learning models are used with distinct features to optimize the accuracy and efficiency [19]. Recent advancements in deep learning and machine vision technologies have empowered intelligent systems to accurately monitor food consumption throughout the entire eating process. These algorithms are used in applications like dietary tracking to offer advanced features for real-time food recognition and detailed nutritional analysis [20]. A Systematic review [21] titled "Applications of Artificial Intelligence, Machine Learning and Deep Learning in Nutrition" has explored how the AI technologies are being integrated into the field of nutrition, particularly for personalized dietary recommendations, dietary assessment and disease prediction. Using a hybrid approach based on Systematic Literature Review (SLR) and Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, the review has systematically analyzed the scientific literature, ensuring the robust study selection and quality assessment. The findings has highlighted AI's versatility in handling the complex nutritional data and its promise to enhance dietary recommendations for individual health outcomes [18], [21].

#### 2.1.1.1 CNN Based Models

Convolutional Neural Network (CNN) is a deep learning model popular for image recognition and classification tasks as CNN has the abilities to efficiently learn and recognize the complex patterns in visual data [22]. The CNN based algorithms capture the patterns, shapes, textures and edges in input images irrespective of variations in position, orientation, scale or translation. Integration of CNN with other traditional method known as multi modal approaches, where image data is combined with other information sources, such as ingredient lists, cooking methods, nutritional labels has enhanced the accuracy of food recognition and nutritional analysis [23].

Popular models use CNN which is more feasible in the estimation of food volume and nutritional contents and a great feature extraction tool which has gradually replaced the traditional machine learning algorithms [22]. Methods that combine visual data with scaled weight and other modalities with more authentic sources like nutritional information data bases, menus list will give more accurate predictions on food names and nutritional information [24]. CNN's hierarchical structure extracts the simple features in early layers and more complex features in deeper layers inspired by the human visual cortex. Powerful frameworks like TensorFlow, Pytorch, and Keras are the backbone to train the deep learning models [25].

**Features of CNN**

CNN architecture consists of multiple layers known as convolutional layers, pooling layers and fully connected layers. Features from the images are extracted by convolutional layers. The ReLu activation function helps network to learn the non-linear relationships between features in image making network more robust for identifying different patterns [25]. Set of filters known as kernels applied by the convolutional layer helps in detecting the edges, textures and patterns in image. To reduce the spatial dimensions of the feature maps and making the computation more effective to solve the overfitting is handled by the pooling layers. Combining these features and classification of image into different food categories is handled by fully connected layers. CNN models are trained on the labeled food images data sets learn to associate the specific features and are capable to enhance the model's performance and prevent overfitting using data augmentation, regularization and dropout techniques. [23], [25], [26]

**Pros of CNN based models**

CNN based models has several advantages including accuracy, automation friendly and robustness which are explained below.

**High accuracy and scalability:** CNN based models are high in the accuracy outperforming the traditional models as they are trained on complex data sets and are able to handle the diverse food items. [25], [27]

**Automation and Robustness:** CNN based models automate the extraction and calculations of nutrition reducing the speed for manual calculations. CNN based models are also robust to various lightning conditions. [22], [25], [27]

**Cons of CNN based models:**

CNN also has disadvantages like requirement of data, computational challenges, complexity and similar which are explained below.

**Data requirements and computational challenges:** CNN based models require large amount of data sets for training this may be challenging to obtain. Significant computation time and powerful GPUs are required as these are computationally intensive. [22], [25], [27]

**Overfitting and Complexity:** CNNs can overfit the training data reducing the ability for generalization of new images. Developing the model itself require advance knowledge in deep learning concepts. Designing and using the model requires good knowledge and skills in software. [22], [25], [27]

Deep Food introduced by Chang Liu et al. Deepfood [25], is one of the most prominent examples of CNN-based food recognition. Convolutional Neural Networks (CNNs) has revolutionized food recognition by automatically learning and extracting the features from large data sets of images that has leaded to the significant improvements in accuracy. For instance, the Deep Food algorithm uses CNNS to analyze the food images captured by mobile devices, achieving high accuracy on datasets like UEC-256 and food-101.

A systematic review on deep learning for food image recognition and nutrition analysis by Merieme Mansouri et al. [23] presents a comprehensive overview of the application of deep learning in this domain. The review focuses on three main areas: food image classification, food image segmentation, and volume estimation of food items to provide nutritional information. A total of 57 original articles were analyzed based on their use cases, the models employed, datasets used, experimental processes, and the main results. The review highlights the advantages of deep learning over traditional methods, while also discussing existing challenges, limitations and future research directions to enhance food recognition systems for dietary assessment [23].

#### 2.1.1.2 YOLO (You Only Look Once) Based Models

YOLO (You Only Look Once) as described and implemented in [28] is a popular real time object detection algorithm which has been effectively applied in food recognition. Single layered CNN is used by YOLO to recognize the food because it has real time processing and can detect many objects in single pass. The image is divided into grid so each grid cell is responsible for prediction of certain number of bounding boxes with their confidence scores or class probabilities and the integration of anchor boxes allows detection of objects of different sizes more accurately [29]. Recent versions of YOLO has the ability to detect small objects for distinguishing the similarly sized food items like fruits and desserts [30]. Latest versions of YOLO like YOLOv7, and YOLOv8 comes with more refined features, faster inference times and the ability to generalize across multiple food categories enhancing high accuracy in detection [31]. YOLOv8 as described in [32], the latest in the YOLO series, is highly accurate for real-time applications. Its grid-based architecture identifies and classifies multiple objects in a single image pass. By training on specialized food datasets (like Food Recognition 2022), YOLO can recognize varied items such as salads with dressings and soups within the same plate. Using YOLO over other algorithms can enhance the detection of real time food detection and nutrition analysis applications in the restaurants and similar systems for personalized recommendations as well. The techniques like Intersection Over Union (IoU) and Non Maximum Suppression (NMS) improves the ability to remove the redundant detections [33]. NutrifyAI uses YOLOv8 to detect the food items from image and then retrieves nutritional data from the

Edamam API, offering calorie counts, macronutrients breakdowns and personalized meal recommendations [34].

### Pros of using YOLO

The YOLO models have several advantages, including high performance, application-based suitability, and integration flexibility. These advantages are explained below.

**Performance metrics:** YOLO models consistently rank high in terms of mean Average Precision (mAP) and processing speed. For example, YOLOv8 achieved an impressive 0.963 mAP at 0.5 IoU on the Food Recognition 2022 dataset, a widely used benchmark for food detection. [32], [34]

**Application-based suitability:** YOLO's single-stage, grid-based architecture is ideal for real-time detection tasks. It's accuracy and processing speed aligns well with requirements in mobile health and dietary tracking applications, where users need rapid feedback. [32], [34]

**Integration and flexibility:** YOLO models can be easily integrated with other tools and APIs, such as Edamam for nutritional data retrieval. This makes them suitable for systems like NutrifyAI which provides calorie counts, macro nutrients breakdowns and personalized meal recommendations. [34]

### Cons of using YOLO

YOLO models have disadvantages like localization issues and the models also need of high computational resources which are explained below:

**Localization issues:** Single stage design of YOLO has some localization issues and high computational resources may require to train the model. The model struggles with highly overlapping or visually similar food items. It tends to misplace bounding boxes, especially when multiple objects fall into the same grid cell, struggle with varied aspect ratios and rely on coarse features due to down sampling [34], [35].

**High computational resources:** YOLO (especially modern variants like YOLOv4 and beyond) requires significant computational power and extensive training time. For instance the YOLOv4 improves performance through architectural enhancements at the cost of higher resource demands and longer training periods. [36], [37]

#### 2.1.1.3 MobileNetV2 Based Models

MobileNetV2 as described in [38] is another popular deep learning model used for food detection. It is designed for efficiency on mobile and embedded devices while maintaining strong accuracy in food classification and calorie estimation tasks. MobileNetV2 is typically paired with pre-trained models on larger food image datasets and these are fine-tuned for specific tasks such as nutritional analysis. [38], [39]

**Pros and Cons**

Being lightweight architecture this is ideal for mobile devices and for high efficiency where the resource environment is low. This model is comparatively less accurate than heavier models like YOLO or Faster R-CNN for the complex images and may require fine-tuning to handle the wide variety of foods. [38], [39]

#### 2.1.1.4 Faster R-CNN (ResNet-50) Based Models

As described in [40] faster R-CNN particularly with ResNet-50 as the backbone, is frequently used for more detailed food detection tasks. Faster R-CNN has a new layer ROI Pooling to extract the equal length feature vector and it has only single staged network making it lighter and faster compared to R-CNN [41]. The main contribution of R-CNN is this extract the features based on CNN consisting 3 modules to make more accurate compared to YOLO. The Extract region proposals extracts the feature vector from each of the region proposal [42], [43]. Pre trained SVM algorithm is used by 3rd region to classify the region proposals.´There is great improvement from R-CNN to Faster R-CNN in inference speed, where Fast R-CNN is 146 times faster than R-CNN and Faster R-CNN is 1460 times faster than R-CNN [44]. Although slower than YOLO, Faster R-CNN is renowned for its precision in identifying smaller and more intricate objects as described in [43] which can be important in food detection where items like nuts or garnishes might otherwise be missed. Faster R-CNN integrates both region proposal generation and object detection within a single convolutional network as described in [37], making it both efficient and powerful.

**Pros and Cons**

Excellent for detecting small and overlapping objects. High precision in challenging environments. It has proven to be a highly accurate object detection framework, particularly effective in scenarios requiring precise localization, as demonstrated by its successful application in detecting packaging defects and agricultural diseases [37], [43], [45]. However, this model has slower inference time so it is not suitable for real-time applications and is more computationally demanding and cannot be trained end to end. [37], [42], [43], [45]

#### 2.1.1.5 Transfer Learning Approaches

Transfer Learning approach as described in [46] involves using pre-trained models on large datasets and fine-tuning them for specific tasks in this literature review for food recognition. The transfer learning also has different stages starting from pre-training the base model, data pre-processing to the final output. In pre training the base model is trained on large data sets for specific tasks where feature extraction is happened. In data pre-processing stage data augmentation and data standardization process occurs. The features are extracted from pre trained models. Using a pre-trained CNN for extraction of features and training a new classifier on top of these features is important. Some layers of pre trained model are retrained and some layers remains fixed [46], [47]. Transfer learning has been particularly effective to improve the performance of food recognition systems. This approach extracts high-level semantic features from images which are crucial for accurate food recognition. [46], [47], [48]

#### Pros of Transfer Learning

Transfer learning approaches has advantages like increasing the accuracy, reducing the training time and simialr which are explained below.

**Improved accuracy:** These pre-trained models have already learned to extract high-level features from images, such as edges, textures, shapes and patterns. When applied to food detection, these features can be very effective in distinguishing between different types of food. This significantly improves the accuracy and are better in prediction. [46], [47], [49]

**Reduced training time:** Models are already pre trained so time to train the model is considerably short compared to train this from scratch [50].

#### Cons of Transfer Learning

Disadvantages of transfer learning can be described below.

**Dependency on pre trained models:** If pre-trained model to be used in transfer learning are not betteror accurate then it directly affects the performance from transfer learning. Because this approach is heavily relied on quality and relevance of pre-trained models [51].

**Limited Customization and complexity in integration:** Fine tuning may not always adapt the pre-trained models and there is always risk of overfitting if the specific dataset is too small or not diverse enough. Integration of the model as required is also challenging and need advance skills. [52], [53]

Recent advances in food recognition algorithms have seen the integration of various AI based algorithms like CNN, attention mechanisms and transfer learning. Models like DenseNet, Mask R-CNN, deep learning , YOLO and similar approaches are driving innovation in real-time food recognition, while multi-modal approaches and generative models are expanding the capabilities of these systems. These models are being applied in diverse detection systems, including mobile applications, healthcare, and dietary assessment, offering more accurate and efficient solutions for food recognition tasks. [29], [54]

The summary of popular models are explained with relevant articles in this section 2.1 are summarized in below Table 2.1

| Model | Architecture | Key Features | Pros | Cons |
|---|---|---|---|---|
| CNN | Multilayered (3 layered) | More accuracy in extraction and detection | High accuracy for complex images | High computational cost, slower inference |
| YOLOv8 | Single-stage, grid-based | High-speed, real-time and FPN (Feature Pyramid Network) | Fast, high mAP (mean Average Precision), suitable for real-time use | Struggles with overlapping small objects |
| MobileNet V2 based CNN | Pretrained and fine-tuned on food datasets | Lightweight, effective in low-resource environments | Ideal for mobile devices and higher efficiency | Less accurate for complex images, may require fine-tuning |
| Faster R-CNN (ResNet-50) | ROI (Region of Interest pooling), feature sharing across proposals | Precision in identifying smaller, intricate objects | Great for detecting small and overlapping objects | Not suitable for real-time applications |
| Transfer Learning Approaches | Pre-trained models with fine-tuning | Reusability and extendability | Faster training, better accuracy | Depends on base model; fine-tuning is computationally intensive |

Table 2.1: Architecture, features, pros and cons of presented Algorithms

### 2.1.2 Challenges From Previous Models and Approaches

The summary of popular models are summarized in above Table 2.1. The application of food recognition and nutrition analysis algorithms has been a rapidly growing field, particularly in versatile dietary monitoring applications [25], [48], [55]. However, despite significant progress in research, the studies evaluating these algorithms reveal several limitations in practical applications. Below is a critical description of recent studies, highlighting both the achievements and the gaps that still need to be addressed. AI models struggle with generalizability and reproducibility across different environments due to distributional shifts, limited data, hardware variations, differences in data pre-processing pipelines, hyperparameter sensitivity, software dependencies and lack of standardized evaluation protocols which can significantly impact the performance. [25], [56]

#### Data consistency and standardization

Different sources have different naming conventions, incorrectly detected names, terms and spellings. Many existing food recognition algorithms show high accuracy when evaluated on specific datasets like UEC-256 and Food-101, with models such as Deep Food and DenseNet but their generalization abilities to real-world scenarios remains a challenge. Studies have noted that models such as DeepFood report high accuracy on benchmarks, but performance drops by 10–20 percentage in real-world environments limiting practical usability. [23], [25]

#### Variability in food presentation

Food items vary significantly due to cooking methods, presentation styles (e.g., bowls vs. plates), lighting conditions, mixed or normally distributed, camera quality and camera angles. These variations pose a major challenge to the generalization of food recognition models. Occlusion between food items or dispersal on the plate makes accurate identification difficult.

#### Explainability, biases and transparency

Deep learning models often operate as black boxes, making it difficult to interpret their decisions. Key questions such as how the models were trained, how reliable the data is and how representative the datasets are unanswered. Lack of model transparency and potential data bias limits trust especially in the sensitive fields like healthcare [57].

#### Portion size and weight estimation

Recent studies found that current food image recognition techniques vary widely in accuracy and struggle with estimating portion sizes. A blended approach combining AI, contextual data and the user input could significantly improve dietary assessment [58]. Accurately estimating food quantity remains largely unsolved according to [59], the article has also discussed the validity of image-based dietary assessment methods. Foods differ in density, shape and layout which affects volume and weight estimation. While detection has improved still, estimation remains difficult and challenging [60].

#### Computational complexity

While deep learning improves accuracy, it increases the demand for computational resources too. Real-time applications may suffer from slower response times, API bottlenecks, data integration overhead and battery drain on mobile or embedded devices [61].

#### Challenges in developing multi-modal systems

Multimodal systems aim to combine image data with ingredient lists, nutrition databases and preparation methods. However, conflicts in naming, inconsistent categorization and partial coverage across sources make integration complex. A comprehensive survey is given by [18]. Merging these pipelines into one cohesive system is technically demanding and prone to errors [62].

These and other limitations above has emphasized the importance of robust, transparent and generalizable approach for food recognition. Addressing them in future research will greatly enhance practical deployment in real-world environments.

## 2.2 Existing AI Based Tools and Models in Food Analysis

In recent years several AI-based tools have emerged to address the requirements of accuracy in food recognition and nutritional estimation. Each tool has unique strengths and limitations, making them suitable for different applications in dietary monitoring, restaurant management and personalized nutrition. This section will review AI based popular latest tools and the existing challenges, limitations and errors from those tools in the context of this study.

### 2.2.1 Review of Popular AI Based Tools in Food Detection

AI based tools leverage deep learning, computer vision, transfer learning approaches, multimodal approaches and natural language processing (NLP) to enhance food identification accuracy, enable precise segmentation and generate detailed nutritional insights. AI based tools like LogMeal, FoodSAM, OpenAI vision and similar other models are getting more popularity in food detection and nutrition analysis these days.

#### LogMeal

LogMeal is a food recognition and nutritional analysis platform that offers both its own software solution and APIs for seamless integration into third-party applications. LogMeal API uses the deep learning approach for identifying the food in an image. LogMeal API is based on the deep learning algorithm and is very easy to integrate in software systems. The API can be used in restaurants, food tech startups, healthcare, consumers and food manufacturing business. First the image is uploaded or captured by the LogMeal based platform. The AI model then analyzes the image and food items in the image are identified by advanced image processing techniques to recognize different types of food based on their visual characteristics such as shape, color and texture. The identified food items are then compared against a comprehensive database of foods. The identified food items' detailed macro and micro nutritional information is provided by the LogMeal [26], [63]. The segmentation provided by LogMeal helps to calculate the area and weight of the food items. LogMeal APIs can be used in various applications like dietary tracking applications, in restaurants and similar other places where strict dietary requirement is needed. LogMeal detects wide range of food items, gives the segmentation of the food dish and also provides the detailed nutritional information. The algorithm is trained on large dataset of food images to provide better accuracy. The accuracy rate of LogMeal is about 85-90 percent for common food items. It recognizes over 1300 food dishes, it estimates the approximate portion of all the detected food items and also estimates the calories, macronutrients and micronutrients. It also helps to customize the recipes and dietary preferences. The LogMeal based applications can help to track the nutrients intake and can be used in self-checkout kiosks in the restaurants also. However, LogMeal's reliance on image based data alone may affect its accuracy in mixed dishes or foods with complex appearances, where exact portion sizes are harder to estimate and it can also affect the nutritional estimation. [26], [63]

**Segment Anything Model (SAM) Based Tools**

SAM is an advanced image segmentation model developed by Meta AI. It utilizes a Vision Transformer (ViT) architecture to process images and generate segmentation masks based on various prompts, including points, boxes and text descriptions. Trained on over 11 million images with more than 1.1 billion segmentation masks, SAM demonstrates impressive zero-shot performance across diverse tasks and domains [64]. Researchers are exploring SAM's capabilities across domains such as food image segmentation (FoodSAM, MealSAM, IngredSAM), medical imaging, and remote sensing, often enhancing it with fine-tuning, prompt-learning strategies, or feature-matching to adapt SAM to specific tasks and improve performance [55], [65], [66]. Prompt-based segmentation tools like MealSAM have demonstrated that such models can significantly reduce manual annotation time while encouraging involvement from non-AI specialists making high-quality food data annotation more accessible and efficient [65].

FoodSAM, known as a novel framework enhanced the Segment Anything Model (SAM) for food image segmentation by integrating coarse semantic masks with SAM-generated masks, enabling instance, panoptic, and promptable segmentation [55]. FoodSAM mainly integrates three models SAM (Segment Anything Model) that generates class-agnostic binary masks for objects present in image, semantic segmenter which match these masks with specific food categories and object detector that identifies non-food objects to provide a comprehensive segmentation. Extensive experiments have validated the FoodSAM's impressive performance, establishing it as a pioneering tool in food image segmentation [55]. In Segmentation process SAM creates multiple binary masks of the uploaded analyzed image. The semantic segmenter assigns the food category labels to the masks and the object detector identifies the nonfood items and background are distinguished from food items. Instance Segmentation recognizes the individual food items treating the individual food item's ingredients as separate entities. Panoptic Segmentation then combines the instance and semantic segmentation to provide a holistic view of the image, including both food and non-food items. Promotable segmentation supports the various prompt types to refine the segmentation based on the inputs from the users [55], [64].

**FoodMem**

FoodMem is a near Real-time and Precise Food Video Segmentation by Al Mughrabi Ahmad et al. [67]. FoodMem has described segmentation including videos for real world issues in health, technology and agriculture. FoodMem employs a two-phase solution: a transformer segmentation phase for initial masks and a memory-based tracking phase for monitoring in complex scenes. This framework as described in [67], [68] has been designed for segmentation of food items from video sequences of 360-degree unbound scenes. By generating the masks of food portions in a video sequence of 360-degree the limitations over existing semantic segmentation like flickering and prohibitive inference speeds in video processing has been resolved. It outperforms current models, enhancing mean average precision by 2.5 percentage and achieving 58 times faster processing. FoodMem demonstrates extremely high mAP and recall across both datasets compared to other state of the art food segmentation methods, indicating strong precision and instance retrieval performance [68].

**LLaVA**

LLaVA (Large Language and Vision Assistant) is a novel approach in detection of food as it integrates the visual image understanding with the language processing. LLaVA understands and generate the responses based on visual inputs and prompts. In LLaVA users can upload the image and ask about what are there in the image. The vision component of LLaVA processes the image to extract relevant features [63], [69]. The visual features extracted by the vision model are then integrated with a language model. The language model uses these features to understand the context of the image. This allows it to generate responses that are informed by both the visual and textual data. User can interact with food images through natural language queries providing the estimates based on visual recognition strengths lie in their ability to understand context and user needs. The integrated model generates responses in natural language, providing detailed descriptions, answering questions about the image. Users can interact with the system asking follow up questions [63], [70].

### 2.2.2 Generic Model OpenAI Vision API and ChatGPT

OpenAI's Vision API is not limited to general images only, it also supports a wide range of image-based use cases. The Vision API can be easily integrated into software applications for various image processing tasks. For example, when food images are uploaded, the API processes the images, extracts relevant features and detects objects within them. A recent research as described in [71] shows promising result with ChatGPT, the system is able to identify food items and provide nutritional information based on the uploaded images. [71], [72]

First, the uploaded image's contents such as shapes, colors, and textures are analyzed by a computer vision model to identify the food items present. Information about the names, portion sizes and nutritional values of the food can then be extracted through natural language processing (NLP) techniques. This includes detecting the name, weight and both macro- and micronutrient content of individual food items, as well as the portion size and the aggregated nutritional information of the entire dish, as specified by user prompts. ChatGPT (Chat Generative Pre-trained Transformer) is a conversational AI model developed by OpenAI, launched on November 30, 2022, and built on the GPT architecture to interact in natural language through text or voice. The can help to identify food dish, identify the components and weights, detailed nutritional information of the dish can be obtained which includes details like calorie count, macronutrients (proteins, fats, carbohydrates), and micronutrients (vitamins and minerals). The AI model continuously learns and improves over time. As more images are processed and more data is collected, the accuracy of food recognition and nutritional analysis is improved. OpenAI's Vision API can also understand the context of the food items. For example, it can differentiate between a whole apple and a sliced apple or recognize a dish with multiple ingredients for more accurate identification. [54], [71], [73]

By fine tuning these models on food image data sets, the vision API gives competitive accuracy. With additional details provided, follow up questions or corrections for initial identification the accuracy of detection and nutritional information is further improved. By connecting the vast database of nutritional data and user prompts the accuracy of some food items ranges around 80-90 percentage depending on image quality. [72], [73]

| **Tools** | **Key Features** | **Architecture** | **Pros** | **Cons** |
| --- | --- | --- | --- | --- |
| LogMeal | Deep learning-based API for food recognition, nutritional analysis and dietary recommendations | Uses computer vision to identify food items and a database for nutritional info | High accuracy, comprehensive nutritional data, personalized recommendations, and easy integration | Limited food database, subscription cost, learning curve for new users |
| LLaVA | Multimodal AI (language and vision), visual and language comprehension, chat capabilities | Combines visual data with language models for comprehensive analysis | Open-source, impressive chat abilities, multimodal GPT-4 level capabilities, Science QA accuracy | Limited dataset for initial training, potential biases, requires integration with other tools |
| FoodSAM | Food image segmentation, instance and panoptic segmentation, promptable segmentation | Uses SAM for segmentation, semantic segmenter for food categories, object detector for non-food items | Advanced segmentation capabilities, continuous learning, contextual understanding | May require high-quality images, limited to food-related applications |

| Tools | Key Features | Architecture | Pros | Cons |
|---|---|---|---|---|
| FoodMem | Food recognition, memory-based learning, nutritional analysis | Uses memory-based learning to recognize food items and provide nutritional info | Memory-based learning enhances accuracy, detailed nutritional information | Limited to food-related applications, potential data privacy concerns |
| ChatGPT | Image-based food detection via Natural Language Processing (NLP), conversational AI, broad knowledge base | Integrates with image recognition tools, uses NLP for nutritional info | Versatile, human-like conversations, continuous improvements, wide range of applications | Not specialized for food, inaccuracies, ethical and legal implications, limits and biases |
| OpenAI Vision API | Image analysis, object detection, visual search, medical image analysis | Uses advanced computer vision algorithms to analyze images and detect objects | High accuracy, customizable, versatile applications, continuous learning | Complexity, cost, dependence on data, limited interpretability, not food-specific |

Table 2.2: Architecture, features, pros and cons of the AI-based tools

### 2.2.3 AI Driven Systems for Food Analysis

Popular dietary related platforms in the form of mobile and system applications where versatile AI based Algorithms, APIs and databases are utilized are in existence. This section describes some of the latest platforms that are capable of providing highly accurate nutritional information. In some platforms hardware tools such as cameras, sensors, weighing scale, lighting systems and kiosks are integrated to achieve the most precise and reliable results. The summary of these systems is described in Table 2.3.

### Smart Nutrition Monitoring System

Smart Nutrition Monitoring System using Serverless Edge Computing was a research done by Javadi et al. [26]. The system has shown high level of accuracy as the system has focused in solving the errors in controlled environment and measurement of the intake of food. They integrated a camera, weighing scale, different deep learning models for food image segmentation, classification and food volume estimations. AWS Sage Maker has been used for training the both models for food segmentation and volume estimation [26]. Aqua Calc has converted the volume to mass and Fat Secret API has provided the nutritional values of food. The comparison of this system's food detection and nutritional analysis with the existing popular food recognition AI algorithms like FoodAI, Clarifai, LogMeal was done and the results were displayed after the analysis. The system has accurately estimated nutritional values with an average error of 6.1 percentage when evaluated using a real food database. This system with serverless edge computing architecture, is able to scan the food in just 10 seconds and the results has been reported to the users within approximately 110 seconds via mobile app in smartphones [26].

### NutriVision

NutriVision combines computer vision and machine learning for diet management by identifying food items, estimating the portion sizes of food and providing the nutritional contents. This system uses image-based detection through faster R-CNN [41], [42], [43], [45]. Faster region based deep convolutional neural networks (R-CNNs) for ingredients recognition and food classification has enhanced the accuracy with transfer learning approaches [74]. First the uploaded image is preprocessed with data augmentation to get the better prediction results. After extraction of the features then object detection process is continued with the faster R-CNN [42], [75] . Custom nutritional data base is used to calculate the detailed nutritional information [74]. NutriVision's features include multi-object detection, integration with smart healthcare devices and the ability to track dietary intake without manual input. It is highly effective for tracking general food categories but struggles with complex or custom dishes due to limited fine-grained recognition. The system is capable to integrate the health data of the specific users to give the dietary recommendations. The edge-cloud design improves processing speed, but network dependency can affect real-time performance, making it more suitable for users in stable internet environments [74], [75]. The NutriVision CNN model as described in [76] achieves exceptional performance in detecting child malnutrition with 97 percentage accuracy, 95 and 96 percent precision and recall result with a strong F1-score, making it suitable for clinical use. However, as described in [76] the effectiveness is constrained by a small, geographically and demographically limited dataset, potential inaccuracies in portion estimation and dependency on high-quality images for reliable diagnosis.

**NutrifyAI**

According to Han and Chen et al. [34], NutrifyAI combines advanced computer vision and nutritional analysis to solve the existed challenges in the food and nutrition. It utilizes YOLOv8 for food detection integrated with the Edamam Nutrition API and Google sheets to analyze the nutritional contents. The user interface is developed in the form of web and mobile application and server side in Flask. This application is tailored for personalized meal recommendations, suggestions can be adjusted based on the dietary goals and user's history data. Nutrify AI's strengths lie in its high detection accuracy (up to 96.3 percentage mean Average Precision) and integration with personalized recommendations, which enhances the user's engagement. However, it relies on accurate label mapping to ensure consistency, which can occasionally limit specificity when categorizing more granular food items. Overall, NutrifyAI is well-suited for users aiming to track nutrition while receiving tailored dietary recommendations in real time. This model lacks the diverse training dataset for more accurate recognition. It also lacks the user feedback mechanism that can help the system for individual meal recommendations.

**Food Tracker**

Food Tracker developed by researchers at McGill University, recognizes the multi object meals in single image to provides the nutritional facts utilizing the YOLOv2 and deep CNN [25], [37]. It merges the deep learning models with YOLO [34] architecture, offering multi-object recognition for real-time food and nutrient analysis in the form of mobile applications. Without the need of user effort [77]. This tool provides fast and real time nutritional information upon capturing the meal image and helps the users to monitor their dietary habits. Food Tracker's mobile-friendly implementation has made it highly useful for the daily use. However, as with other image-based models, it has limitations as it relies on pre-trained Deep Convolutional Neural Network(DCNN) that may not generalize well in identifying the complex food compositions [78]. The accuracy for the dishes with multiple and versatile food components is not much effective. Additionally, the model's efficiency also varies depending on the device specifications, as high-resolution image processing demands more resources [77].

The summary of these popular system is described in Table 2.3 below.

| System | Key Features | Architecture | Pros | Cons |
|---|---|---|---|---|
| Smart Nutrition Monitoring System | Automated nutrition monitoring, IoT integration, real-time analysis | Uses IoT-enabled sensors and deep learning algorithms to monitor and analyze food | Accurate real-time monitoring, useful for infants and daycare settings | May require specific hardware, initial setup can be complex to start and process the data. |
| Nutri Vision | Food recognition, nutritional analysis, meal recommendations | Combines computer vision with nutritional databases to identify food and analyze nutrients | High accuracy in food recognition, personalized meal recommendations | Dependent on quality of food images, may require manual corrections |
| NutrifyAI | Real-time food detection, nutritional analysis, personalized meal plans | Utilizes YOLOv8 model for food detection, Edamam API for nutrient analysis | Immediate dietary insights, high food recognition accuracy | Requires internet connection, potential privacy concerns with data handling |
| FoodTracker | Food logging, calorie tracking, nutritional insights | Uses image recognition and user input to log food and track nutritional intake | Easy to use, integrates with fitness apps, provides detailed nutritional insights | Manual input can be tedious, accuracy depends on the user compliance |

Table 2.3: Summary of the AI Driven Applications

As reviewed in a Forbes Health article [79], several calorie-counting applications have been evaluated for their usability and accuracy like Lose It! Calorie Counter, MyFitnessPal, Calorie Mama, Nutritionix Track, YAZIO Calorie Counter and Diet, Calory, Calorie

Counter by MyNetDiary, Calorie Counter by FatSecret, Carb Manager Keto Diet Tracker, Lifesum: Diet and Food Tracker. There are many other applications like NutriTrackPro, Nutritrack, m2Calories, Bite.ai, Food visor, Snap Calorie, Diet Camera AI etc. Many applications and researches have claimed to solve the issues existed in previous similar products and has better accuracy comparing to previous studies.

### 2.2.4 Comparative Analysis of AI Based Models and Systems

Previous sections 2.2.1 and 2.2.3 has described the architecture, features, pros and cons of popular AI based models and applications in food analysis with summary tables 2.2 and 2.3. This section analyzes various models and tools based on multiple aspects, including the nature of the training data, detection accuracy, component-based analysis, semantic segmentation capability and the user experience efficiency. It also examines how these tools integrate with real-world applications and how the specific characteristics of each model influence the overall effectiveness in detecting food items and estimating nutritional contents. Furthermore, the performance of these tools is often influenced by the quality of integrated APIs, underlying software frameworks, hardware components and the diversity of testing environments. Additional factors such as model architecture, computational efficiency, generalization ability and scalability are also considered to provide a comprehensive comparison.

**Detection of Food Dishes**

The results obtained from different AI based tools like LogMeal, ChatGPT, OpenAI vision, LLaVA and FoodSAM all are highly affected by the trained data set. These AI modals has shown the high accuracy with the trained data but their limited food database may not cover all the regional cuisines and the versatile food items [54]. OpenAI vision API depends on integrating the image recognition tools but it can give the different results depending upon the prompts. OpenAI vision API is not designed primarily for the food detection but can show great results based on the prompts [54]. LLaVA [63] is very effective in detecting the food dishes but initial training dataset limitations can affect the performance of LLaVA as well. It requires extensive data for improvement and integration with other tools for optimal results. FoodSAM [55] is also effective in detecting food dishes but it requires high-quality images for accurate segmentation and the segments are also fragmented. It excels in detailed segmentation only and can't provide other information. FoodMem [67] is very accurate in detecting the known food items but very limited to its memory-based learning model and struggles with new or uncommon dishes.

### Detection of Individual Food Components

There can be different food items in the image of the food dish. Different algorithms have different abilities in identifying the different food items and giving the detection names. LogMeal has high accuracy in semantic segmentation of individual food items in the image but it can be affected by poor image quality, trained data and other factors also. It provides very detailed nutritional information and personalized recommendations for each food items but may not account for variations in preparation methods and dishes. The OpenAI Vision API is very effective in identifying the individual food items and displaying weights and the nutritional information for whole dish and total but it requires high-quality images and extensive training data [54]. LLaVA [63] has the high accuracy in identifying individual food items but needs integration with other tools for optimal performance. Excellent segmentation capabilities can be obtained with FoodSAM but it is dependent on the image clarity. The detailed segmentation and contextual understanding can be obtained from FoodSAM. High accuracy is obtained in identifying the individual food item via FoodMem due to its memory-based learning. It requires frequent updates to maintain accuracy and relevance.

**User Experience Challenges**

LogMeal has user-friendly interface and has its own device for food analysis and nutrition detection and also provides APIs for easy integration into the health and fitness applications. Conversational experience and easy user interaction is provided by the ChatGPT and it produces the responses based on the user uploaded image and communication and is even capable for generation of dietary plans. OpenAI's Vision API can be customized for the specific applications easily but this is not specialized for food industry. Open-sourced and accessible LLaVA comes with the impressive chat abilities. It requires integration with other tools for detail nutritional information. Advanced segmentation capabilities is provided by FoodSAM but it may requires high-quality images and is limited to only segmentation.

**Scalability**

LogMeal is scalable for use in health and fitness apps but limited by its food database for regional cuisines. ChatGPT and OpenAI vision are highly scalable due to versatility and broad knowledge but these are not meant only for food. It can be integrated into various applications. The OpenAI Vision API is scalable for applications requiring high precision but is also resource-intensive and is not only for food. LLaVA combines the visual and textual data for comprehensive analysis and is scalable with extensive data and integration with other tools. FoodSAM is scalable for food-related applications' segmentation but it requires high-quality images. FoodMem is also scalable with frequent updates to its memory-based learning model.

**Integration and Customization**

LogMeal is easy to integrate with the health and fitness apps and can also be customized for personalized dietary recommendations. ChatGPT is versatile and this can be integrated into various applications. OpenAI Vision API is also customizable for specific applications but requires technical expertise and is very suitable for high-precision tasks. LLaVA [63] also require the integration with other tools for full functionality. Customizable for combining visual and textual data. FoodSAM can be used for detailed segmentation but require computational resource and skills. FoodMem can also be customizable for food-related applications.

In conclusion, LogMeal is great for standardized, accurate food detection and nutrients estimation due to its structured dataset and versatile features. In contrast, ChatGPT , OpenAI vision API, FoodSAM and LLaVA offer higher flexibility and adaptability in conversational queries but with more variability in accuracy, especially in the complex or mixed food images. Standardizing the data sources across these tools could improve accuracy, as demonstrated by studies emphasizing the benefits of multimodal data integration and consistent labeling protocols for more accurate results.

Each AI tool has its unique strengths and limitations depending upon the needs of the application, making them suitable for different purposes and user needs. Single AI tool only is not enough to face the challenges present in the current food industry. Multimodal approaches and integration of the AI tools with other databases can help solving the existing problems. Even for proper analysis of tools there is lack of platform for real time analysis from versatile pipelines, data integration, correction and analysis. This thesis will address the challenges in previous research developing a platform.

## 2.3 Previous Studies in AI Based Food Analysis and Challenges

Commonly, previous studies apply methods like controlled dataset testing, real-world testing and standardized metrics to analyze the model performance. The Image-based dietary assessment (IBDA) methods comparing to traditional approaches such as 24-hour dietary recalls (24-HDR) and weighed food records (WFR) show the significant measurement errors. Further research is needed to better understand the sources of these inaccuracies [59]. Meta-analyses which is used for the comparison of traditional evaluation methods (such as food records, 24-hour recalls, and food frequency questionnaires) and image-based evaluations (IBDAs) underestimates the energy intake when compared to the gold standard, doubly labeled water (DLW) with errors averaging 179–263 kcal/day [80].

### Prior Research Methodologies

The methodologies employed in food recognition research are essential for evaluating the reliability and applicability of food detection algorithms and tools. Commonly previous studies apply methods like controlled dataset testing, single approach from AI tools and no correction methods were applied for improving the model's performance. Dietary assessment relies on both correct food identification and accurate portion size estimation. Without knowing how much was eaten, nutrient intake can't be reliably calculated [58]. These limitations mean image recognition tools cannot yet replace traditional methods like food diaries or recalls.

### Benchmark dataset testing

Many studies utilize benchmark datasets like Food-101 or UEC-256, as these provide a structured and standardized approach to measuring model performance under controlled conditions. As Mansouri et al. [23] has highlighted, these datasets allow for a comparative assessment of algorithms using consistent data inputs, reducing the potential for variability due to environmental factors. However, these datasets typically lack the of real world applications, often focusing on clear, fine images that do not reflect everyday challenges like low lighting or occlusions [23].

### Real-world T testing and environmental variability

Some studies adopt real world testing environments, where food images are captured in settings with varying lighting, camera angles and food presentations. Javadi et al. [26] has emphasized that real-world testing provides insight into how models perform under conditions users are likely to encounter, such as at home or in a restaurant. However, such methodologies can be challenging to standardize and without consistent protocols, results may lack comparability across studies.

### Evaluation metrics

Precision, recall, F1 score, mean absolute error(MAE), root mean square error (RMSE) are commonly used metrics for evaluating food recognition models. These metrics quantify accuracy and reliability by analyzing the proportion of true positives, false positives, and errors. Liu et al. [25] note that while these metrics are effective for model comparison, they may not fully capture the nutritional accuracy needed in real world applications where small deviations can lead to significant nutritional miscalculations.

## 2.3.1 Research Gaps in Previous Studies

To effectively utilize technology-based applications for dietary assessment, it is essential to accurately recognize and estimate food portions in order to calculate energy and nutrient intake using food composition databases. In theory, food recognition, supported by appropriate classification techniques and deep learning models capable of identifying food ingredients can facilitate the estimation of nutrient intake for dietary evaluation.

#### Standardization across studies

A recurring issue is the lack of standardized protocols for model evaluation, which makes it difficult to compare tools directly. Studies like those by Alif Rabbani et al. [32]

have called for a unified framework, particularly for dietary applications, where accuracy and consistency are critical for user health outcomes. Inconsistent dataset use and non-standardized performance made difficult to predict the model's true efficiency across different context [32]. This thesis will try to address this by integrating multiple AI based model's data pipelines and re-correct the AI based data based on weight and standard recipe menu.

**Dataset diversity and bias**

Different models are trained in different datasets so this can already cause bias and evaluation of this models may can't give better analysis results. Many food recognition models are trained on datasets that lack cultural or dietary diversity and are trained in controlled environment images , which can bias the models against certain types of foods [59]. For instance, Food-101 is heavily weighted towards the Western dishes, limiting its applicability for more culturally diverse diets. Zhang et al. [57] has argued that expanding datasets to include a broader range of foods can improve model performance and reduce biases that affect accuracy for non-Western cuisines. Moving food measurement technologies from labs to realistic settings like virtual reality, naturalistic labs and real food environments is more important for more valid results. It also highlights the need for collaboration and pooling resources to overcome limitations in sample size and study duration [81]. This thesis will test the developed AIoT system in real environment and compare the data with standardized data.

**Portion estimation and multimodal data integration**

Van Asbroeck and Matthys [58] conducted a comparison study evaluating the performance of seven image recognition platforms using 185 food and beverage images. While some platforms demonstrated moderate accuracy in identifying food types (e.g., Calorie Mama API with a Top-1 accuracy of 63 percent), none were able to accurately estimate food quantities which is essential component for the nutritional analysis. As a promising solution to overcome the limitations in current methodologies. The findings concluded that current food image recognition platforms vary in accuracy and are unreliable for portion size estimation [58]. A blended approach using AI, contextual data, and user interaction is recommended to improve dietary assessment [23]. Even Weight prediction can be inaccurate mainly due to image occlusion and visual similarity between food items, which made precise weight estimation from 2D images challenging [82]. While CNN based models are accurate in detecting food types, they are less effective in estimating portion sizes, especially in complex dishes. Integrating weight or volume data alongside image analysis could improve portion estimation, providing more accurate nutritional information for mixed dishes. Additionally, Javadi et al. [26] has explored the use of multi modal data sources, using the scale data with image recognition, as a promising solution to overcome the limitations in current methodologies. Portion size based analysis after integration of multiple pipelines and recipe menu can help to reduce this error in this study.

#### Explainability and transparency in AI predictions

Deep learning based food recognition models lack interpretability, making it difficult to understand how they make predictions. Many models function as black boxes, raising concerns about biases, data authenticity, and accountability. Implementing explainable AI (XAI) frameworks that provide visual or textual explanations for each classification decision. In this thesis we try to overcome this error by using more open source based and most popular models and modular system to integrate other models later.

#### Integration with external APIs and databases

Few studies explore the integration of AI food recognition models with nutritional databases (e.g., USDA FoodData Central, MyFitnessPal, Edamam API). NutrifyAI as described by [57] demonstrated improved dietary recommendations by combining YOLO-based food detection with API-driven nutritional data. Developing standardized API frameworks to enable real-time integration with nutritional databases and health monitoring applications. It is important to combine the multiple data sources to improve the accuracy and completeness of the application, and to solve the practical challenges related to data quality, participant involvement, privacy, and resource demands collaboration across different disciplines is important [81].

This thesis will address this challenge by using multi modal approach like sensor data(weight), implementing multiple AI based tools and re-correction of names and nutritional information with standardized menu.

### 2.3.2 Research Direction for This Study

The key issues identified in previous studies and proposed solution based on this study has been summarized in table 2.4. Most previous studies have concluded that food recognition models and applications perform well on benchmark datasets (e.g., Food-101, UEC-256), but their performance drops significantly on real-world images. This limitation is directly addressed in RQ 1.5, as this study will use a real-time dataset in a testing environment with diverse food items to evaluate model robustness and generalizability.

| Research Gap | Key Issues Identified | Proposed Solution |
|---|---|---|
| Lack of Standardization | Inconsistent datasets and evaluation methodologies | Unified benchmarking framework |
| Dataset Diversity | Bias in food recognition models due to Western-centric training datasets | Expanding datasets with versatile food items. |
| Portion Size Estimation | 2D image-based models fail to estimate weight and volume accurately | Hybrid AI and sensor based models (integrated weight scales and possibly other sensors) |
| Lack of Explainability | AI models function as black boxes, limiting user to trust and use the system | Implementing explainable AI (XAI) in the system to make it easy to understand and use for different use cases |
| Multimodal Integration | Limited combination of AI models with nutritional databases and external APIs | Developing API-driven multimodal system with different AI based tools for detection and re-correction |

Table 2.4: Summary of Key Research Gaps and Future Directions

Many existing applications lack automated data processing, extraction and analysis but this study will propose a platform for real time data analysis and extraction from multiple pipelines with minimal manual effort. Furthermore, although most models can recognize food names, they often fail to provide accurate weight or volume estimates. To overcome this, our approach combines image-based food recognition with weight-based estimation using an integrated scale to refine nutritional information. This aligns with the research questions in Section 2 and specifically address RQ 2.1 and RQ 2.2 through the integration of a standardized menu for name correction and nutritional recalibration.

Additionally, most AI-based systems focus solely on visual information and overlook ingredient lists, cooking methods and nutrition labels. This study will integrate camera input, scaled weight measurements, LogMeal API, OpenAI vision and other detection algorithms and standardized menu data from authentic food databases to deliver more accurate real time food detection and nutrient estimation. Finally, this thesis aims to propose a more standardized and modular framework for food detection, incorporating diverse AI-based APIs and exploring multimodal approaches. This framework directly responds to RQ 3.1 by improving system accuracy and user-friendliness, and to RQ 3.2 by providing a foundation for more reliable dietary monitoring and nutrition research.

This thesis proposes a novel multi modal approach for food detection and nutritional analysis by integrating computer vision, weight measurement, and menu-based nutritional information into a unified, real-time system. Unlike previous works that rely solely on benchmark image datasets, controlled testing and mostly vision-only models, this approach incorporates multi modal inputs and automated pipelines to enhance accuracy, adaptability and scalability in real-world scenarios. The modular system will be able to integrate many AI based pipelines into a single integrated platform.

# 3 Study Design

This section has presented a plan for a series of experimental studies designed to evaluate the effectiveness of AI-based food recognition models in identifying food items and estimating their nutritional values. The studies focus on specific challenges such as multi-item dish recognition, portion size estimation, comparison with standardized menu data, and cross-model evaluation.

## 3.1 Experimental Studies Plan and Objectives

This section focuses in general methodologies and the overall approach to execute the studies. The technical details regarding data collection process, system, AI based APIs, re-correction and testing the system has been described in the methodology Section 4 in detail. Summary of experimental studies with objectives, procedure and data analysis is described in Table 3.1.

A series of experimental studies were conducted to assess the ability of different AI systems in recognizing food items and estimating their nutritional contents as summarized in Table 3.1. The main focus has been on assessing the systems' accuracy and reliability in various scenarios when comparing with the ground truth data involving different food types, portion sizes, and complexities. The goal of the Experimental studies is to evaluate the performance of AI-based food recognition systems in a real-world context. We aim to address the specific challenges as below.

- Total food dish recognition and weight estimation directly.
- Multi-item food components' recognition and displaying the weight information.
- Displaying total dish and food components' weight and nutritional information.
- Comparison of the data obtained from the correction of device with ground truth.
- Multi modal AI based and re-correction pipelines integration and evaluation of performance comparing with the ground truth.

## 3.2 Overview for Data Collection and Pre-processing

Data collection and pre-processing was really important for the empirical part of the thesis. This section has provided an outline of how the Experimental studies which has ben structured and implemented. Each Experimental study will follow a clear methodological plan, and the implementation of the studies (including hardware setup, data collection, and testing) which has been described in the methodology Section 4 in detail. The overview for data collection, extraction and analysis is shown in Figure 3.1.

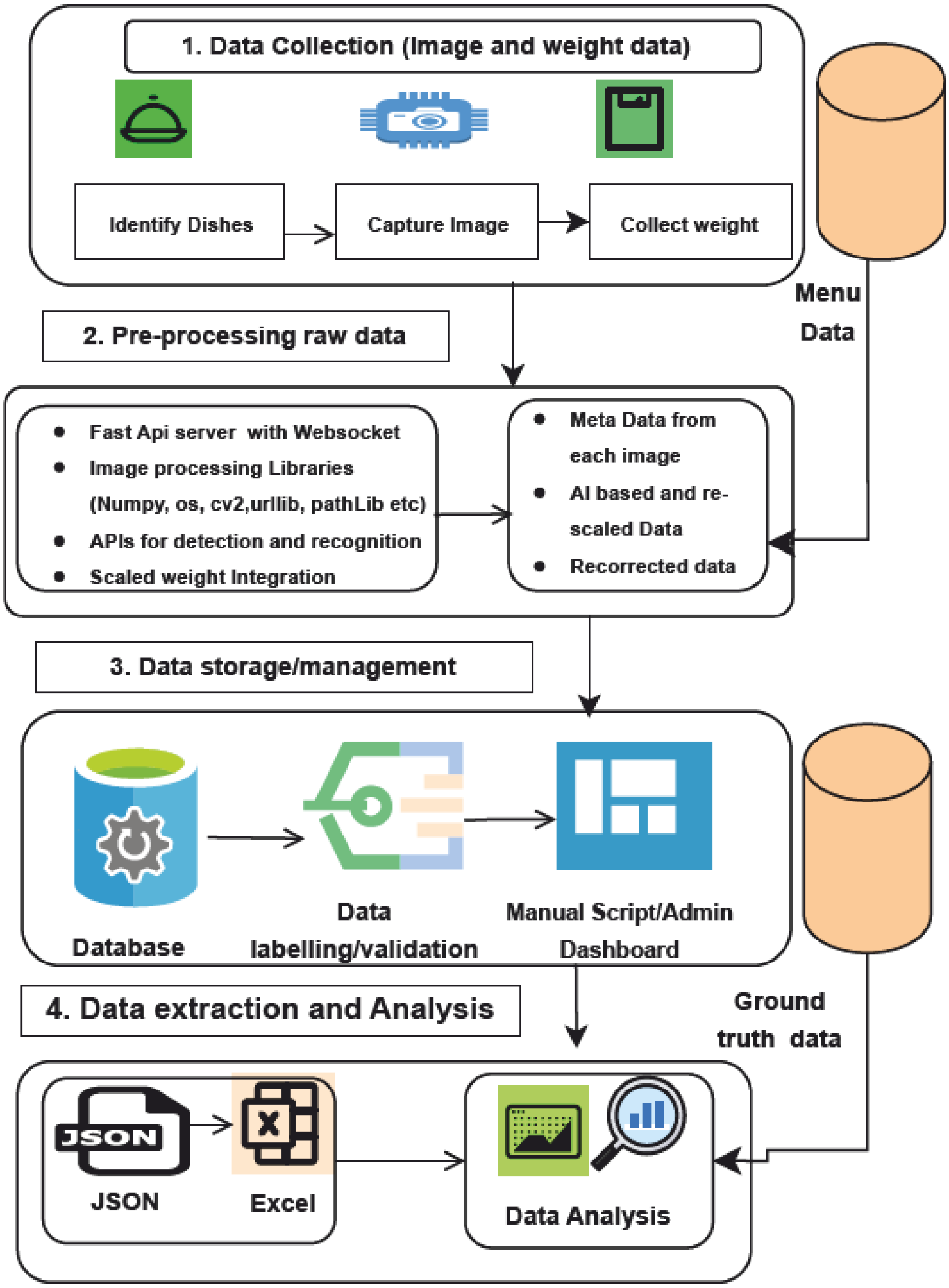
### 3.2.1 Identify the Food Dishes for Study

The selection of diverse set of dishes that vary in complexity was available in Flavoria restaurant [8] as taken by the volunteers which is defined in detail in 4. The selection of dishes represents the common meals that might pose different challenges for food recognition systems.

### 3.2.2 Taking the Images and Collecting the Data from Scale

A high-resolution camera was used to capture the images of the food dishes. The Telpo C50 kiosk has been used in later stages for prototype studies and it has excellent camera.

Figure 3.1: Overview of the Data Collection and Analysis



The Telpo C50 device is highly effective. Simultaneously, recording the weight of each food item and dish using a calibrated digital scale.

### 3.2.3 Image Pre-processing

Image pre-processing involved normalizing the images to a standard resolution and format and applying the image augmentation techniques such as cropping, rotation, and scaling to simulate real-world variations.

### 3.2.4 Integrating the Food Recognition APIs

Processing of the real time images' data using the LogMeal API, FoodSAM API, OpenAI vision API and other possible relevant AI Algorithms and extraction of the data can be done. Collection of the identified food item's name, calculated weight of the individual food items based on the segmentation names, re-scaled weight of the individual food items and nutritional values (e.g., calories, macro nutrients and micro nutrients of individual and whole food dish). The data was exported from these versatile tools for the analysis.

### 3.2.5 Ground Truth Data Collection

Flavoria restaurant's lunchline was used for the ground true data collection [82] for comparison.

### 3.2.6 System Development

An IoT based application has been developed to display, extract, re-correct and analyze the real time data. Developing a user-friendly user interface to display the analysis results of the food and re-corrected information with different pipelines. Integration of APIs, WebSocket server, multimodal approaches to facilitate the communication between different modules (image processing, menu, database and detail analysis).

## 3.3 Data Storage and Management

This section has outlined the process for handling the data collected during the experiments, including storage, extraction, re-correction and analysis. The workflow begins with server-side data processing and continues through structured database management and performance evaluation.

### 3.3.1 Database Setup

The process started with running the test image data on the server, where the AI models and integrated system has performed food detection and nutritional estimation. All output data, including detected food names, estimated weights and nutritional values were automatically stored on the server-side PostgreSQL database. The re-corrected data values are stored as well.

### 3.3.2 Data Extraction

After the data collection process, the stored information was extracted from the server using Python script for analysis. The structured database contains tables for raw images, detection results, and metadata such as re-scaled weights and AI-based nutritional predictions.

### 3.3.3 Data Labeling

Each dataset entry was labeled with identifiers representing the dish, its components and the experimental conditions under which the data was collected (e.g., test day, lighting, and menu context). Proper mapping between the ground truth records and AI-generated outputs ensured consistency and reliability during validation.

### 3.3.4 Name and Weight re-Correction

To improve accuracy, the food names and weights estimated by the AI models were normalized and re-corrected using a combination of the weighing scale, test day's menu data and manually validated ground truth information. The re-corrected data was re-checked and manually validated as needed. For components weights and nutritional analysis this was done manually.

### 3.3.5 Nutrient Calculation and Adjustments

The nutrient estimations obtained from AI algorithms were refined using the corrected food names and weights. These adjusted nutrient values were compared with the ground truth nutrient data to evaluate each model's accuracy and calibration capability.

### 3.3.6 Data Analysis

After extraction and correction, the final dataset was analyzed to assess the overall performance of the AI-based models and APIs. Both quantitative metrics (e.g., precision, recall, and F1-score) and qualitative analyses were performed to identify systematic errors and evaluate model generalization.

### 3.3.7 Python script and Admin Dashboard for Analysis

Python Jupyter notebook has been used to extract the required data from data base in the form of Excel for further analysis. Later a web-based admin tool, Flex Admin, has been implemented to facilitate the reprocessing of images, correction of data entries and re-running of specific test cases when required. For data handling, the extracted results were first stored in `JSON` format and later converted to Excel files for further analysis. The ground truth data including verified food names, actual weights and nutrients was obtained from the lunch line records [82].

## 3.4 Experimental Studies

This section provides the Objectives, structure and procedure for each Experimental study. This is also displayed as as summary Table 3.1.

### Experiment 1: Comparison of Accuracy of Names

#### Objectives

This study was aimed to address the research question RQ 1.1, to evaluate how well different food recognition systems like LogMeal and FoodSAM recognizes the names of food items in comparison with the general vision model OpenAI vision and ground truth. How these algorithms detect the simple, managed and even complex dishes with the multiple food components is important here.

#### Procedure

The procedure involved presenting a mixed dish under the camera of the device and collecting the identified food items data. These names were normalized and the results are then compared with the ground truth names.

#### Data Analysis

First the names were extracted then normalized using the mapping dictionary. The same food items were detected with similar names based on trained data so normalizing was important. Data analysis includes using metrics like precision, recall and F1 score to quantify the performance of each system and performing a qualitative analysis of common errors and challenges faced by the recognition systems.

### Experiment 2: Comparison of Accuracy Of Weights

#### Objectives

This experimental study aimed to address the research question RQ 1.2, by evaluating how well different food recognition systems like LogMeal, OpenAI vision and system's scale can detect the total weight of food items on a plate and compare this estimation with the Lunch line based ground truth data. This study has also tried to address the food component based weight analysis part to answer RQ 1.5.

#### Procedure

In this study, food items were analyzed by the system which is integrated to the different food recognition systems (OpenAI vision, LogMeal and FoodSAM). The weights data from the AI models was saved and extracted for analysis. The ground truth weight data and AI based results were compared for the accuracy. The re-scaled total weight from LogMeal and FoodSAM is almost same as the both weights were calculated based on pixel segmentation using the scaled weight.

#### Data Analysis

The accuracy of each system in estimating total weight were evaluated using error metrics, such as Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE). A comparison of the performance across different approaches will provide idea into the strengths and weaknesses of each algorithm. The analysis was able to identify the common factors that may influence the accuracy of weight estimation, such as dish type, image quality or the complexity of the food arrangement.

### Experiment 3: Comparison of Accuracy Of Nutrients

#### Objectives

This Experimental study has aimed to address the research question RQ 1.3, to evaluate how accurately different food recognition systems such as LogMeal, OpenAI vision can estimate the total energy and total macro nutrients contents (Proteins, Fats, Carbohydrates) of the food dish. The primary focus was on comparing the results obtained from these AI-based systems with known ground truth data from Lunchline.

#### Procedure

In this study, whole dish's food items' nutrients were analyzed by different food recognition systems (LogMeal, OpenAI vision, LogMeal re-scaled, OpenAI re-scaled, Final corrected). These systems attempted to estimate the total nutrients and macro nutrients of the food on a plate, which were compared with the Lunch Line based nutritional information (ground truth data).

#### Data Analysis

The accuracy of the energy and macro nutrients predictions were evaluated by calculating error metrics such as Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE). A comparative analysis were conducted to assess the performance of the different recognition systems, even the correction mechanism from the system and identifying the strengths and weaknesses of each in terms of estimating the total and macro nutrient contents of the food dish.

### Experiment 4: Impact of Varying Dish Weights on Accuracy

#### Objectives

This experimental study was aimed to answer RQ 1.4, by assessing how different sizes of food items can influence the accuracy of food recognition and estimation of the weights and nutrients. By examining the detection based on the plates size, the accuracy of the recognition algorithms and their ability to estimate nutritional values accurately has been analyzed.

#### Procedure

To conduct this study, selection of 20 least weights ground truth based data and 20 maximum ground truth based images were done. The results from the approaches were compared against the ground truth data to assess the accuracy of portion size recognition and nutrients calculation. The images were analyzed separately and the errors were checked.

#### Data Analysis

The analysis has focused on how the varying portion sizes impact the accuracy of the food recognition systems. By comparing the AI results with the ground truth information, any trends or biases in the algorithms regarding the portion size can be identified. This analysis has helped in understanding the strengths and limitations of the current system and provide the insights for further improvements.

### Experiment 5: Comparison of Component Based Analysis

#### Objectives

This experimental study was aimed to answer RQ 1.5, by assessing how accurately various food recognition detects weights and nutrients of individual food components. The performance of the AI systems and corrections from AI systems were qualitatively analyzed to see the significance and possibility of further analysis.

#### Procedure

In this study, weights and nutrients of the different food components present in the plate were analyzed by different food recognition systems. These systems has attempted to estimate the weights and total nutrients and macro nutrients of the food components on the plate, which was compared with the ground truth data.

#### Data Analysis

The accuracy of the individual food items' macros and total nutrients predictions were extracted and observed. Possibility of calculating the qualitative error metrics such as Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE) was observed.

### Experiment 6: Comparison Of Corrected Weights and Nutritional Data

The final corrected names, weights and nutritional information from the system with the Lunch line based values were compared in this study.

#### Objectives

This study has aimed to answer RQ 2.1 and RQ 2.2, to evaluate the performance of the food recognition systems by comparing the re-scaled outputs and final corrected outputs against the ground truth, first correction only using the scaled weight and second correction by integrating the menu based nutritional information.

#### Procedure

The procedure has involved the correction of AI predictions of the food dishes using the scaled weight and standardized menu. The corrected food items, weights and their nutritional values were then compared with the standardized menu data based information(lunch line based data).

### Data Analysis

This process has included conducting the statistical examination to measure the deviation of the results from the standardized ground truth data. Additionally, the study has also explored the consistency and standardization.

Below is a summary table that shows summary of the case studies explained above.

| Experiments | Objectives | Procedure | Data Analysis |
|---|---|---|---|
| Comparison of Individual Food Items' Names | Evaluate the accuracy of food item identification | Present mixed dishes, collect identified items and compare with ground truth data. | Measure precision, recall, F1 score and analyze the errors. |
| Comparison of Total Weight and Individual Weights of Food Items | Compare AI-based weight detection with ground truth data | Compare estimated vs. actual weights of food items | MAE and RMSE to compare system performance. |
| Comparison of Total nutrients and total Macro nutrients of the Dish | Compare the nutrient estimation (macros) from algorithms with ground truth | Process food image, compare estimated nutrients with the ground truth | Calculate MAE, RMSE, and compare the accuracy across systems. |
| Comparison of Name and Nutrition of the Food Item with Varying weights | Assess recognition accuracy across different dish weights | Selecting the maximum ground truth weights 20 data and least weights 20 data for analysis | Compare the portion sizes' impact on accuracy, identify biases. |
| Possibility of comparison of the Individual Food Item's weights and nutrients | Macro nutrients and total nutrients detection for individual food components | Compare the weights and nutrients for each food items with the ground truth | MAE and RMSE to assess accuracy, compare systems. |
| Comparison of Name, weights and nutrition of the food dish with varying weights | Assess recognition accuracy across different dish weights | Selecting the maximum ground truth weights 20 dish's data and least ground weights 20 dish's data for the analysis | Compare the portion sizes' impact on accuracy, identify biases. |
| Comparison of the Corrected Nutrition Estimation by the System with the rescaled weights and standardized Menu | Compare final system's corrected results with the predefined menu data | Process food, compare corrected results with ground truth data | Measure deviation from ground truth data, and analyze consistency. |

Table 3.1: Comparison of Experimental Studies

# 4 Study Implementation

This section has described how each case study has been implemented, including the hardware and software setup, data collection methods and analysis techniques. How the Flavoria Flex system has integrated the AI models and how it has facilitated the data collection and testing is described in this section.

## 4.1 Integration of Hardware and Software

The system setup using laptop, separate scale and camera and integrated Telpo C50 kiosk is as shown in Figure 4.1. The Telpo C50 runs only Android and system has a Java/Kotlin Frontend app showing a webview. The app streams camera and weight/qr/rfid data to a Python backend with React frontend on a separate device as our system. The server side and user interface components can be run on a standard laptop or desktop as well. The hardware components like USB camera, consumer-grade digital scale, and QR code reader can be easily integrated in a Laptop supporting boarder adoption.

### 4.1.1 Hardware Components

**Camera**

A camera has been integrated in the system as shown in Figure 4.1b. The Telpo C50, device is equipped with a high-resolution camera Figure 4.1b to capture clear, high- quality images of the food. The camera has been calibrated for optimal lighting and angles to

Figure 4.1: Telpo C50 kiosk and laptop setup

(a) Telpo C50 kiosk

(b) Setup using laptop, webcam, and scale

ensure that the food images are consistent and of high quality.

### Scale

The system has also included a scale to weigh the food Figure 4.1b , which will be important for comparing portion sizes and estimating nutritional content based on the known weights. The Scale Calibration has been done by weighing the standard items with known weights to verify the accuracy of the scale. The Telpo C50 itself has integrated scale.

### QR code scanner

QR code scanner has helped to scan the QR code of the user to identify the user and the food taken by the user in real use case setting.

### Tray support and other components:

The device has a good place to support the tray of the food items. Even Figure 4.1 b it is easy to have a tray support as shown. The Telpo C50 system also has light, NFC, printer, QR code reader, memory, power plug, communications port for Wi-Fi, Bluetooth and ethernet.These all components can be integrated using laptop as a server also.

### 4.1.2 Software Components

As shown in the **Architecture diagram** the whole system as shown in Figure 4.2 has integration of software and hardware components like camera, scale, food detection APIS, React,Redux, Material UI libraries, WebSocket, Fast APIs, Python libraries, correction Logic, data extraction and analysis logics as Flavoria Flex admin.

#### Food Recognition AI based tools

The system is integrated with multiple AI based APIs including LogMeal, OpenAI Vision and FoodSAM to enable detection of food items names, weights and estimation of nutritional content from different pipelines.

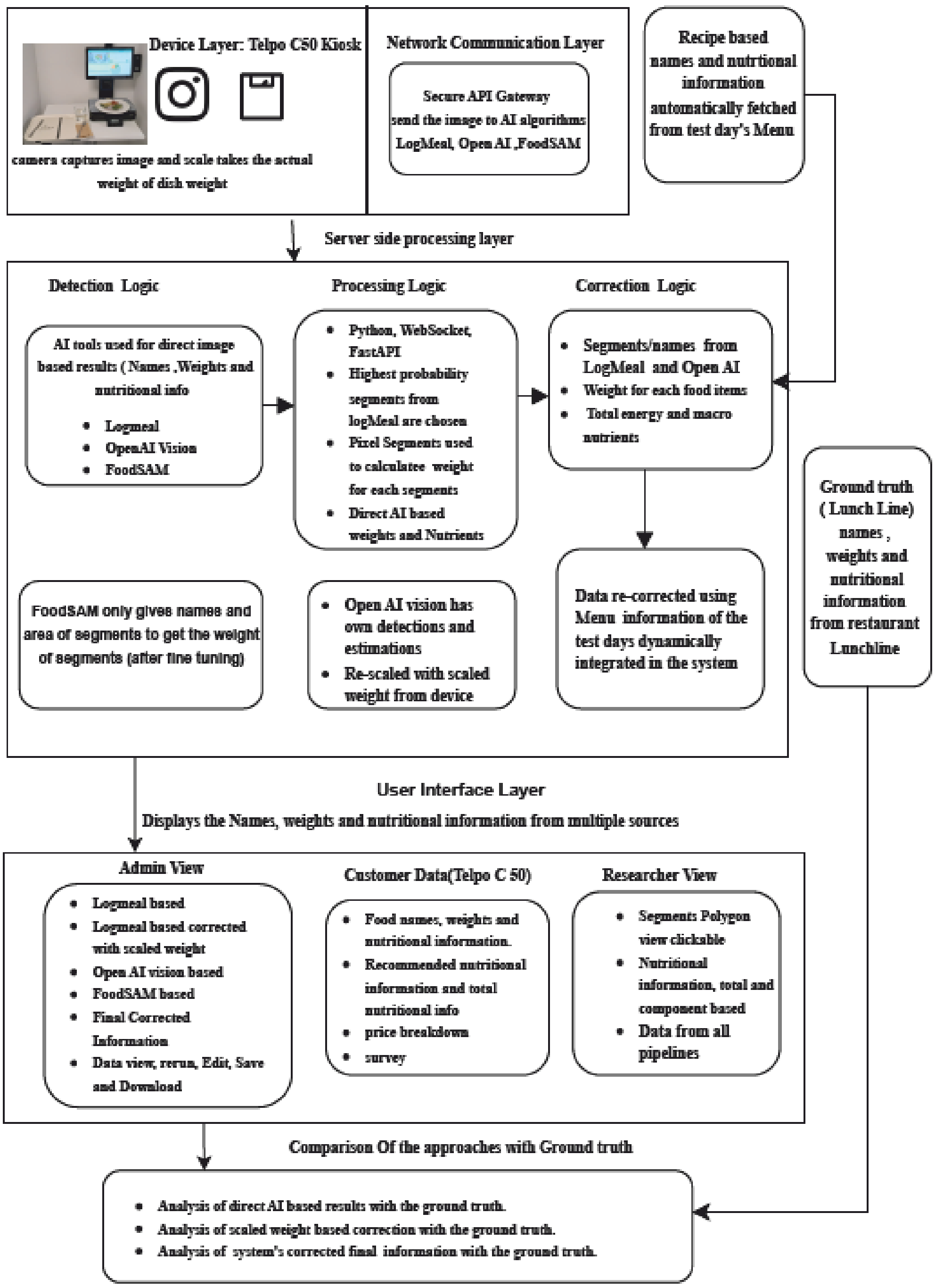
#### System Integration tools

The platform has integrated AI based models and the hardware into a machine vision system. The components like camera, scale, food recognition APIs are integrated into a single platform using WebSocket, FastAPI server(using Python as a programming language) for server side,various image processing and the segmentation libraries, React, Redux and material UI libraries for front end.

#### Dynamic Menu for ground truth comparison

The direct AI based results after re-scaling has been integrated with the restaurant's menu of the test day, So the system has provided correction of names and nutritional information improving the the accuracy.

Figure 4.2: Architecture Diagram



**User Interface tools**

Modular UI for researchers, users and admins has been developed using React, Redux and Material UI libraries to displays real-time food names, weights and the nutritional details obtained from the AI based pipelines.

## 4.2 Integration of AI-based APIs

For developing the platform for data collection Web-Socket, LogMeal based API, Fast API, OpenAI vision API, React, Redux and Java/Kotlin has been used. Design and development of the UI was crucial for easy displaying and comparison of the food recognition and nutritional data. The UI has rovided a seamless experience, displaying food names, weights and nutritional data in real-time from various algorithms and pipelines. The application has been developed in modular way considering the needs of both the client/customers in different use cases and researchers too. Configuration flags has been used for different approaches and views based on the needs of users.

### 4.2.1 Identification and Segmentation

The hardware tools has been integrated with Web socket Fast API as Python's web application server for taking and analyzing the real time images. The real time image as shown in Figure 4.3, above has been taken by the camera and the scale of the device is taking the weight. The user Interface has been developed in React, Redux, material UI libraries and different other similar libraries. The Python based libraries for image processing are used in the server side for the image encoding and data analysis purposes.

**Detection of Names**

LogMeal API has been primarily used for detection and segmentation of names from the food dish subsection 2.2.1. OpenAI vision has been integrated with the application to get

Figure 4.3: Sample test dish image

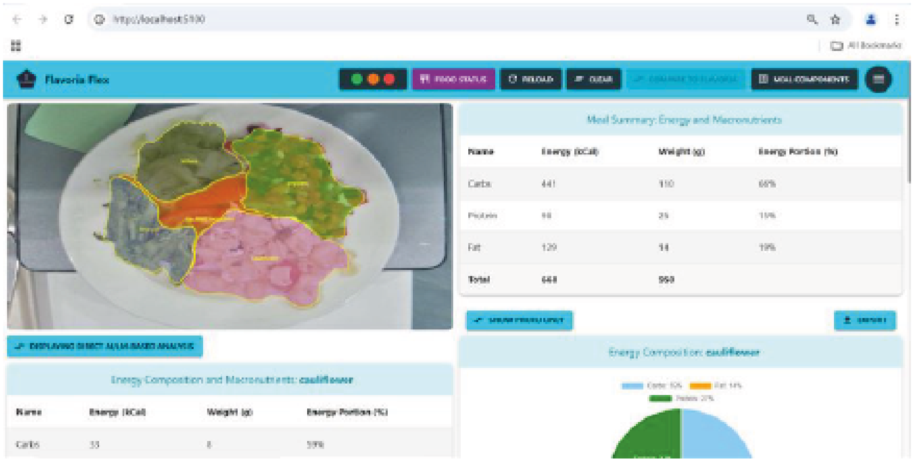
the names of the dish of the food items and this has given more accurate names. FoodSAM based names of the food dishes are also taken by running the images fine tuning the model and the names are then compared with the ground truth names.

**Segmentation of Dish**

LogMeal API has divided the whole food into distinct segments as described in subsection 2.2.1. The term segmentation means division of the whole dish into the distinct parts or components based on the specific characteristics they have. Pixel segmentation divides the image into distinct more manageable regions based on the certain features of the pixels as shown in Figure 4.4. The segmentation has helped for simplifying the tasks here for the object recognition and image analysis. In Semantic segmentation each pixel of the image has been labeled and it comes as a part of the specific food item.

Deep Learning based LogMeal API has shown high accuracy in pixel segmentation and even provide detailed nutritional analysis. There is also the option to see the probabilities of the guesses based on LogMeal's guesses and re-correct the names of the detected food

Figure 4.4: Food segmentation, total and components nutritional info display




items based on the probabilities. The nutritional information that we obtained from the direct LogMeal guesses has been converted to the scaled weight later.

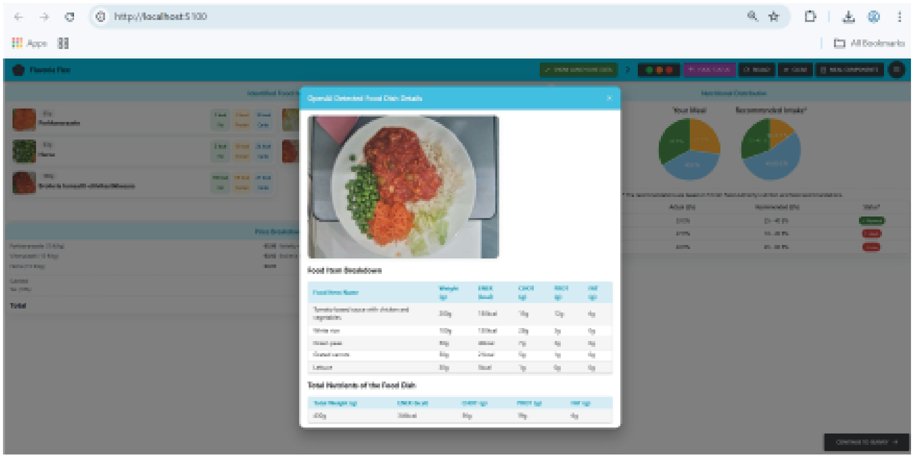
### 4.2.2 Estimation Of Weight

Method for estimation of total weight of the dish and individual food components' weight is described in this section.

#### Total Weight of Dish

The weight of the food dish and the individual food component has been obtained from the LogMeal API, FoodSAM and the OpenAI Vision API's direct estimation. The total weight of the dish is obtained from the Flavoria-Flex device's scale. The ground truth weight has been obtained from the Flavoria research restaurant's lunchline [82]. The restaurant's lunch line for each menu item is equipped with integrated scales to measure the weight of each food component. Finally, the data for whole dishes and individual items, including names, weights and the nutritional information is collected.

Figure 4.5: Detections from OpenAI Vision (names, weights and nutritional info)



**Pixel Segmentation For the Weight of Segments**

The area of pixel segments based on the total scaled weight has been integrated with the Telpo C-50 kiosk. Then re-calculation of the weight of each individual food items has been done. This integration has ensured that the system delivers highly reliable outputs with reduced errors for weights which is later used for re-correction of nutrients. The total area and the area percentages of the polygons/segments were calculated. After this the polygons weight was used to get the weight of each food components. Direct AI based weight guessed by LogMeal and other AI based sources has been taken into account while evaluating.

### 4.2.3 Estimation of Nutrients

The total energy information, including macro nutrients and micro nutrients of the dish, has been initially obtained from the LogMeal API and OpenAI vision API. Figure 4.5, shows the names, weights and nutritional information of the food components and whole dish based on the OpenAI vision's prediction. The information was obtained from these AI tools based on the trained data of the tools. The re-scaled weight of the whole food item and the individual components has been used to refine and re-correct the nutritional information derived from these AI pipelines.

### 4.2.4 Correction of Direct AI Based Predictions

Direct AI model based names, weights and nutritional information are re-corrected to enhance the accuracy of the system.

#### Correction of Names

The detected names from different sources have been mapped to get the corrected names using the restaurant menu of each test day. Different hybrid approaches has been tried and the names are re-corrected to the most matching menu based names. Manual dictionary has also been used to map the names before analysis.

#### Versatile Hybrid Approaches

Segmentation is first performed by LogMeal, which provides pixel-level segments, detected names and labels for segments. For effortless and dynamic correction, the OpenAI Vision API is then used to provide corrected names based on Menu data. The system dynamically fetches the recipe menu names from the Flavoria restaurant's menu for the test days for correction. OpenAI vision has then re-corrected the names dynamically and returns the final corrected names. The result is explained in the results and discussion Section 5.1. In the hybrid approach versatile popular libraries has been tried such as **Fuzzy library, Rapid Fuzzy library, String Similarity** library, **custom mapping dictionary** and other approaches. The logic is to re-map the AI detected names (from LogMeal and OpenAI vision) to the ground-truth menu names of the test day.

Different hybrid approaches has re-corrected direct AI based (LogMeal, FoodSAM and OpenAI names are re-corrected to the correct ground-truth menu names which is described in Section 6.1.

#### Correction of Weight

The direct AI based weight of the whole dish and food components are re-scaled using the weight taken from the device's scale. The pixel segmentation method as described in Section 4.2.2 has given us more accurate weights of each food segments. The re-scaled weight data, when integrated with image segmentation has provided more accurate estimates of food segments' weights and enabled refined nutritional information. Specific areas of food items (e.g., a slice of pizza, an apple) has be linked to the weights. This approach also improved efficiency by reducing discrepancies between direct AI-based predictions and ground-truth weight as described in Section 6.2.

#### Correction of Nutritional Information

The direct AI based Nutritional information for whole food dish and the individual food components has been re-corrected using the re-scaled weight and corrected names of the segments. The re-scaled nutritional information is calculated from re-scaled weights of the segments. The nutritional information (total nutrients, macro nutrients and micro nutrients) for the menu items based on the recipe, standardized per 100 grams of each food component are based on Fineli database. After re-correcting the names as described in Section 4.2.4 and re-scaling the weights as in Section 4.2.4 of the segments, the nutritional information has been re-calculated. The re-scaled nutritional information is derived from the re-scaled weights of each segment and whole dish, providing a more accurate and reliable assessment of the dish's nutritional content as discussed in Section 6.3.

### 4.2.5 User Interface

Real time processing of the images and the smooth user interaction has enhanced the usability. The user interface has been used to display the food item names, weights and nutritional information. The AI based names, weights and nutritional information against the ground truth data. The final user interface is as shown in Figure 4.6.

Figure 4.6: Final User Interface of System with re-Corrected Total and Individual Food Components


### Real-Time Processing

The system is able to processes the images and the weights data instantly for the quick analysis and display in the user interface. The real time image, scaled weight and the nutritional information with recommendation has been displayed in the user interface. The recommended intake for the user is also shown.

### User Interaction

The user interface has enabled the dynamic exploration of food details via hover actions or click-based on the segments to get the weight and detailed nutritional information of each segment as shown in Figure 4.4. If researcher click on cauliflower as shown in image then detail nutritional information from that food item is displayed.

### Displaying Final Corrected Information

Shows individual and aggregate summarized nutritional data, allowing the users to analyze the detailed meals information comprehensively. Users can also see the recommended nutritional information as shown in Figure 4.6.

**Validation**

The system has ensured the displayed data aligns with verified nutritional standards ground truth data which has been collected from Flavoria Lunchline. Which can be visible in the same UI screen for instant comparison as well. Menu names and nutritional information is also visible.

**Storing and Extraction of the data from the server**

Separate Admin dashboard has been created integrating with the server side layer and postgreSQL database. The meta data regarding names, weights and nutritional information obtained from all of the AI sources are extracted as JSON and re-converted to the Excel using Python script.The correction pipelines has been integrated and the data is visible in real time.

**Modular User Interface**

Different user interface based on the configuration flags for clients, developer and researchers is available.Other hardware, sensor tools, AI based pipelines and correction logic can be easily integrated in the system for further research.

## 4.3 Experiments

Each experimental study has represented the specific dimension of food recognition performance using the system. The methodology across experiment studies has followed a common pipeline, starting from dish preparation, image and weight capturing, AI-based recognition, weight and nutrients re-correction, data extraction followed by the analysis. Variations in test data, procedure, objectives, and evaluation metrics has distinguished each experiment. The experiments are summarized in Table 4.1. Details regarding how experiment was done is described below:

### 4.3.1 Comparison of Predicted and Actual Food Components of Dish

This experiment was conducted to evaluate the ability of AI models to detect multiple food items within complex dishes by answering RQ 1.1. AI based algorithms like LogMeal, OpenAI and FoodSAM were used. The total of 350 images were collected across diverse meals, with each dish containing 2–5 distinct items, including salads, side dish and main dish. After data cleaning 167 images was used for the analysis.

**Preparation of the dish**

The dish of the food items according to the menu for that particular were taken by the volunteers from the restaurant. There are versatile food items in the dish taken from the menu for the different test days. Different images have similar or very different meals based on the user's intake. The dishes that contain 2-5 different food items were selected including salads, main dishes and bowls.

**The camera captures the image**

The test dish of the food can be placed on the device's scale, same was done by the volunteers. The camera was positioned at the optimal angle to get the image of food items clearly. The image was captured by the camera.

**The food recognition API is used for detection**

The food recognition LogMeal API was used to process the image to provide the list of predicted food items. Additionally, OpenAI Vision, FoodSAM was also used to gather the alternate name predictions for research.

**Corrections**

The names obtained from different sources were re-corrected using menu, manual methods, dictionary approach, OpenAI vision, Fuzzy library, string similarity, Rapid Fuzz library, dictionary and many other techniques to get the corrected names that matches with the menu of the day.

**Extraction of data**

The output data as names from different sources and also the most corrected names from different AI based system were extracted. The ground truth names were extracted and collected too.

**Further Analysis**

The output data as names from different sources and also the most corrected names from different AI based system were analyzed to compare the accuracy of models in detecting the food components of dish. Precision, recall and F1 score was used for the name accuracy comparison.

### 4.3.2 Comparison of Predicted and Actual Weight of Dish

This experiment study was used to evaluate the accuracy of AI algorithms in comparing the total weight of the dish by answering RQ 1.2. Then re-scaled weight of the dish was calculated by the device scale so that helps to re-correct the nutritional information based on the actual scaled weight of the food.

**Preparation of the dish**

Dishes with multiple food items were taken as described in 4.3.1, the volunteers have the ground truth weights of each food items measured from the restaurant Lunch line.

**Taking the Image and Weight data**

The images were captured by the camera and the total weight was measured using the integrated scale in the device.

**Weight Estimation**

The food recognition algorithm like LogMeal, OpenAI vision was used to estimate the weights of the entire food dish and individual food components. Scale was used for collecting the scaled weight from the system too. The total weight of food components from the scale and the image segmentation based weight helped later to re-calculate the weights of the food components.

**Data Extraction**

The estimated direct weight from the algorithms and the re-scaled weight based on the pixel segmentation were extracted from the system.

**Data Analysis**

Comparison of the estimated weight with the actual weights was conducted to evaluate the error metrics like MAE, MAPE and RMSE to evaluate the AI model's accuracy.

### 4.3.3 Comparison of Predicted Nutrients and Actual Nutrients

This experiment study was used to evaluate the accuracy of AI algorithms in estimating the macro nutrients and total energy of the whole dish to address RQ 1.3.

**Preparation of the dish**

Dishes with multiple food items were available in the restaurant menu. Ground truth macro nutrients and micro nutrients data was collected from lunch line.

**Image Capturing**

High-quality images of the dish were captured as explained in previous sections.

#### Food Recognition and Nutrient Estimation

AI based models (LogMeal,OpenAI vision) estimated the macro nutrients and total energy of the food dishes.

#### Data Extraction

Nutrients data from the direct AI systems was extracted for each food component and total dish.

#### Data Analysis

The AI-estimated data of nutrients were compared with the ground truth data using error metrics like MAE and RMSE to evaluate the AI model's accuracy.

### 4.3.4 Comparison of Predicted Nutritional Values with Corrected Results

The re-scaled weights and predefined menu of the dishes with known nutritional values for each test days was calculated by system as well. These menus are from the Flavoria restaurant's lunch line menu. The data in the menu is from the standardized database Fineli for the test days. This experimental study will answer RQ 2.1 and RQ 2.2 to evaluate the accuracy of corrected names and nutritional information with the ground truth menu based names and nutritional information. For RQ 2.1 direct AI based data is re-scaled based on the scaled weight. Then again for RQ 2.2 the nutritional data is re-corrected based on the menu information for the test days.

#### Preparation of The Dish

The predefined menu of the dishes with known nutritional values for each test days are available in Flavoria Menu. The menu is based on more authentic databases like Fineli or nutritional guidelines for the test days.

### Image and Weight Integrated

High-quality image from camera, weight of the dish from device's scale were recorded and further analyzed with the automated system.

### Food Recognition and Nutrient Estimation

AI systems (LogMeal, re-scaled LogMeal, OpenAI Vision and re-scaled OpenAI vision data) estimated the total macro nutrients and energy of the food dishes based on direct AI guess. The data were re-scaled using system's scaled weight. Then extraction of the re-scaled based weights data of the food dish was done.

### Data correction

Then the identified names were re-corrected with the nearest available menu names for the particular detected food items which was mapped with the menu and menu based final correction of names and nutritional information. The menu has the nutritional information for 100 gram of each food item. First re-correction of the name to match with menu name was done, then we re-correction of the weight based on the pixel segmentation weight was done. The corrected name and weight was used to obtain the nutritional information based on menu. For example if direct AI has incorrectly detected potato as cauliflower then the system will re-correct that to potato and the incorrectly detected weight is re-corrected based on menu.

### Data Extraction

The re-corrected data based on re-scaled weight and based on the re-scaled weight and the nutritional information was extracted for each food component and total dish.

### Data Analysis

The final corrected names and nutritional information was compared with the ground truth names and nutritional data. The re-corrected nutrients data were then compared with the ground truth data using error metrics like MAE and RMSE to evaluate the developed system's accuracy.MAE and RMSE were used as performance metrics.

### 4.3.5 Dish Weights Impact on Food Detection Accuracy

This study was done to examine the impact of dish size variation on food recognition accuracy and the correction pipeline as in RQ 1.4. To this end, 20 dishes with the highest and 20 with the lowest ground truth weights were selected from the full dataset for separate analysis.

**Preparation of the dish**

The maximum weighed 20 dishes and the least weighed 20 dishes were selected from the cleaned data set for the analysis.

**Image Capture**

High-quality images and data of the dish was captured, data is cleared and selected as described in above all experiments

**Food Recognition**

The LogMeal API, OpenAI Vision and other tools were used for detection of names, weights and nutritional information as described in above sections

**Weights and Nutritional Estimation**

Similarly as described in the above sections the system collected the weights and nutritional data.

**Data Extraction**

Selected images data were extracted for further analysis as explained in previous experiments.

#### Further Analysis

Weights and nutrition accuracies has been compared separately using different statistical analysis methods for the largest and smallest dishes based on the ground truth weights. Precision, recall, F1 score, MAE and RMSE was used as error metrics for analysis.

### 4.3.6 Component Based Analysis

This experimental study will try to focus on possibility of component based accuracy analysis on weights, total energy and macro nutrients of the components RQ 1.5. This will be later discussed in Section 6.4

#### Preparation of the dish

The dishes has been prepared similar as other case studies as explained above.

#### Image and weight Capture

High-quality images and weights of the dish was captured as in experimental study 1.

#### Food Recognition API

The LogMeal API, OpenAI Vision and FoodSAM was used for detection of names, weights and nutritional information for the components.

#### Nutritional Estimation

Direct AI based estimations was refined via the re-scaled weight from the device and names has been re-corrected. Finally corrected names, weights and nutritional information for the components were obtained.

#### Data Extraction

Recognition results and corrected names, weights and nutritional information were compiled and extracted for analysis.

### Further Analysis

Due to non uniform numbers of detection of components from AI sources, the manual mapping has been tried to solve the problem. The analysis of the components names, weights and nutrients from different sources with examples has been displayed in the results section. The idea was to explore and see the possibility of this kind of component based analysis in food which is explained in results and discussion sections.

| **Experimen-ts** | **Objective** | **Procedure** | **Data Analysis** |
|---|---|---|---|
| Study 1: Research Question RQ 1.1 | Evaluate AI models in recognition of multiple food components in complex dishes. | LogMeal API, OpenAI Vision, FoodSAM to predict the food names | Compare AI-generated names with the ground truth data using precision, Recall . |
| Study 2: Research Question RQ 1.2 | Evaluate AI algorithms for estimating total weight of dish. | AI(LogMeal, OpenAI vision's) directly estimated total food weights, are compared with the ground weight | Comparison of AI-estimated weights with the actual ground truth weight using error metrics like MAE, MAPE and RMSE. |
| Study 3: Research Question RQ 1.3 | Evaluate the accuracy of AI models in estimating the macro nutrients and total nutrients of the dish. | LogMeal based and OpenAI Vision based data is compared with the ground truth | Compare the AI-estimated nutrients data with the ground truth data using error metrics (MAE, RMSE). |

| Experiments | Objective | Procedure | Data Analysis |
|---|---|---|---|
| Study 4: Research Questions RQ 2.1 and RQ 2.2 | Evaluate the accuracy of final corrected names and nutritional information with the ground truth | Re-scaled LogMeal and OpenAI Vision based data corrected with Menu of the test days is compared with the ground truth | Compare the final corrected nutrients data with the ground truth data using the error metrics (MAE, RMSE). |
| Study 5: Research Question RQ 1.4 | Evaluate the impact of the weights of the dish (maximum and minimum) for food detection accuracy. | LogMeal API, OpenAI Vision and correction pipelines are used for detection of names, weights, and nutritional info. | Compare results of analysis from 20 maximum weighed dishes and 20 minimum weighed dishes using the statistical analysis like (precision, recall, F1 score , MAE, MAPE and RMSE) with ground truth. |
| Study 6: Research Question RQ 1.5 | Explore the possibility of food components based analysis | AI systems estimated weights and nutrients (macros and total nutrients) of each component. | Analysis of direct AI based data and re-corrected data with the ground truth data. |

Table 4.1: Summary of experimental studies

## 4.4 Data Collection and Analysis

The Flavoria research restaurant's [8] lunch line was used for testing the system. In this phase, data has been extracted from the server for each experiments as summarized in Table 4.1. The overview for data collection, extraction and analysis is shown in Figure 3.1. The collected dataset typically included the names of food items recorded during the case studies, along with their corresponding weights across various dishes. This stage has also covered the nutritional information from the ground truth data collection of the Flavoria restaurant [8].

### 4.4.1 Ground Truth and System Test in Flavoria

The Flavoria research restaurant's [8] lunch line was used for ground truth data collection. It has scales to collect the weight of each food components. The weight is collected only for the chosen food components , the volunteers select the food items according to their choice. The screen above each food menu guides the users in the weighing process. Volunteers can check the weights of food taken immediately while taking food from Lunch line to the plate. The overview for data collection, extraction and analysis is shown in Figure 3.1.

#### Timeline for data collection

The total timetable of the span was 1 week. Week 1 was for collecting and preparing the dataset of the mixed dish images from restaurant using the device. These images were processed using multiple AI-based food recognition APIs, including LogMeal, OpenAI Vision API, FoodSAM and menu based corrected nutritional information. Simultaneously, the ground truth data for each food item has been integrated in the system including weight measurements using a calibrated digital scale and nutritional values sourced from authorized food database like Fineli [83](Menu nutritional data is from Fineli data base).

**Number of images and volunteers**

The Flavoria research restaurant serves meals to customers according to its regular menu. Volunteers were recruited to participate in the study for a free meal, and their meals were selected by them based on personal choice. The volunteers were observed and carefully instructed to ensure accurate ground truth data collection. If any mistakes were identified during data recording, the corresponding dishes were flagged and excluded from later analysis.

A total of 350 images and their metadata were collected, including mixed dishes with corresponding ground truth data. Each image has been analyzed using multiple food recognition APIs, and the identified food items' weights and estimated nutritional values has been stored in a database. Re-corrected names and weights based and menu-based re-corrected nutrients results has been recorded/saved for further analysis. Finally, the outputs were compared to evaluate the performance of the recognition systems.

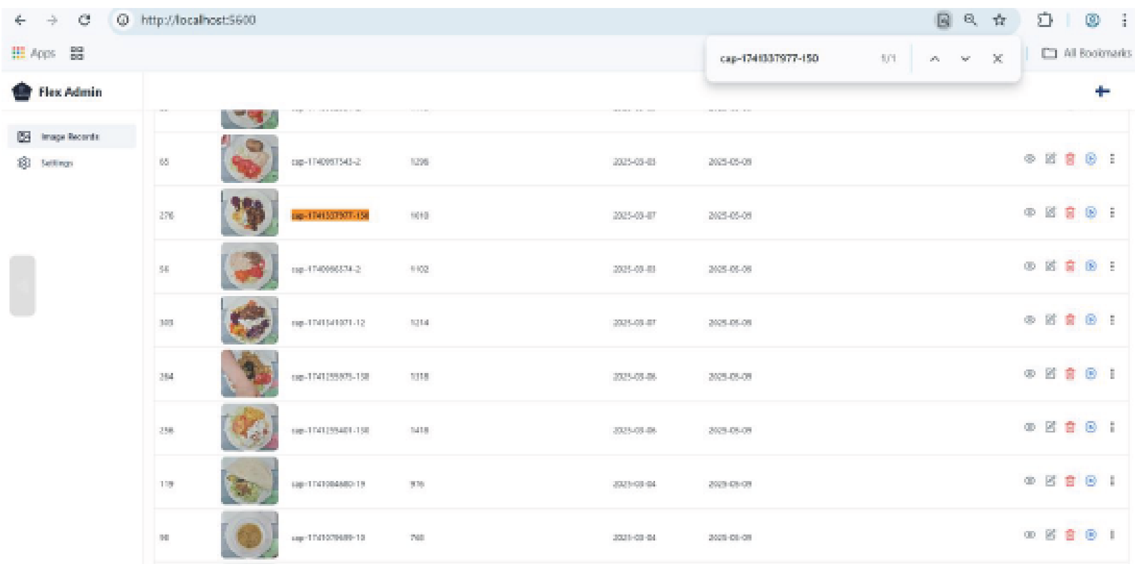
### 4.4.2 Data Extraction

Data saved in the database was extracted using the Python script in the form of Excel. The results generated by AI-based systems—such as recognized food items, weights and their estimated nutritional values—are gathered. These outputs were then compared against the ground truth values obtained from labeled sources to validate the performance and accuracy of the AI models.

### Data cleaning and pre-processing

This stage has helped to ensure that the data is prepared for further analysis by resolving issues like missing or inconsistent values.

Figure 4.7: Admin user Interface of the system




### Data cleaning

The first step in data cleaning involved removing missing or duplicate values, removing wrongly taken data by the volunteers. The images with the null entries, comments, wrongly taken data, data with extreme outliners, redundant and irrelevant images' were removed and finally 167 images' samples were used for analysis.

### Automation of data processing and analysis

Once new data was loaded, the cleaning and normalization functions run automatically in the sequence and the results were saved in database so it was then extracted in the Excel form for further analysis. The data was automatically updated in the separate Admin application as shown in 4.7 for previewing, adding, editing, deleting and re-running the pipelines with the system. Saving and re-correction of data is possible using this admin dashboard.

#### Standardizing the column names

Next, column names were standardized for consistency. For example, the column labeled "Food" was renamed to "Food Item Name" using for example the rename () function in pandas. This helped maintain uniformity across all datasets. The names of columns has been standardized uniformly as we have many columns and data from different approaches. For example for total energy alone the columns are like Lunch line based total energy, LogMeal based total energy, OpenAI vision API based total energy, LogMeal re-scaled total energy, OpenAI vision re-scaled total energy and final corrected total energy.

#### Conversion to data types

Data type conversion was then applied to ensure the appropriate formats for easy comparisons. The weight column was converted to numeric using coercion, and food item names were formatted as strings for the consistency and compatibility in the analysis. The data columns were updated to uniform columns.

#### Analysis of the Names

The names from different sources were different based on the trained data types of the models. LogMeal based API has given us the Spanish dishes and OpenAI has given more western dishes. The FoodSAM based names has also contained the Chinese names. For example, the food item name “Pasta” has been detected as “Spaghetti” from one modal and "pesto pasta" from another modal but the name in Menu is Pesto pasta so the dictionary has helped us to normalize the names to a common name. For comparison of detection of names from different sources the names are normalized using mapping dictionary.The dictionary converts the names from different AI models to a common name from the recipe information of the menu for further analysis.

**Performance Metrics for Name Accuracy Analysis**

Precision, Recall and F1 scores were used as qualitative measures for the normalized names to compare the accuracy from different sources against ground truth.

**TP (True Positives)** shows the predicted food items that actually appear in the ground truth for a sample data.

**FP (False Positives)** means the predicted food items that are not in the ground truth sample list.

**FN (False Negatives)** means the ground truth food items that the AI based prediction has missed. These are counted per sample (per row) and then aggregated across the entire dataset as summary data.

**Precision** says ,how often are the positive predictions from different approaches were correct when comparing with the ground truth (lunch line -based names). High precision compares to a low rate of false positive predictions. Precision = (TP / (TP + FP))

**Recall** shows the rate of the food items which are actually present in the dish detected by the method. High recall compares to a low rate of false negatives. Recall= (TP / (TP + FN))

**F1 score** is the harmonic mean derived from Precision and Recall this shows the balance between the two is f1 score. This represents both the false positives (precision) and false negatives (recall) so this is especially useful when there is an imbalance between the two. F1 Score: 2 × (Precision × Recall) / (Precision + Recall)

**Example: Dish-Level Evaluation Metrics**

Consider the dish **"Grilled Chicken Salad"**.

**Ground Truth:** Chicken, Lettuce, Tomato, Cucumber, Olive

**AI Prediction:** Chicken, Lettuce, Tomato, Cheese

- **TP (True Positives):** Chicken, Lettuce, Tomato → 3

- **FP (False Positives):** Cheese $\rightarrow 1$
- **FN (False Negatives):** Cucumber, Olive $\rightarrow 2$
- **Precision:** $\frac{TP}{TP+FP} = \frac{3}{3+1} = 0.75$
- **Recall:** $\frac{TP}{TP+FN} = \frac{3}{3+2} = 0.6$
- **F1 Score:** $2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \approx 0.667$

#### Normality Tests and Non-parametric Tests

To determine the appropriate statistical tests for comparing nutritional and weight data, normality of the data was first assessed using the Shapiro–Wilk test.

The **Shapiro–Wilk test** evaluated whether a dataset is normally distributed or not. In our analysis, most p-values from the Shapiro–Wilk test were less than 0.05, indicating that we reject the null hypothesis and conclude that the data are not normally distributed.

This test was important because many parametric tests, such as the paired t-test or independent t-test, assumes the normality. When this assumption is violated, parametric tests can give misleading results. Therefore, non-parametric tests that do not rely on normality were used in our analysis.

**Mann–Whitney U Test** was applied to compare two independent groups (e.g., weight_ll vs. weight_logMeal) where each value came from different observations. It was a non-parametric alternative to the independent t-test and tests whether one group tends to have higher or lower values than the other when the data are not normally distributed.

**Wilcoxon Signed-rank Test** is a non-parametric test which was used to compare the median differences between the two paired datasets (e.g., the same dish measured before and after the correction). It is the alternative to the paired t-test when data are not normally distributed in this study.

In summary, if the Shapiro–Wilk test showed $p > 0.05$, it is not possible to reject the null hypothesis and consider the data approximately normal, potentially allowing parametric tests. Otherwise, non-parametric tests like Mann–Whitney U (for independent samples) and Wilcoxon signed-rank (for paired samples) were preferred and done in this study as the data was not following the normal distribution.

### 4.4.3 Statistical Analysis

After pre-processing statistical analysis was performed on the dataset to identify patterns, correlations and trends in the data as below:

**Descriptive statistics**

Mean, Median, Standard deviation, Q1, Q3 and maximum were considered for descriptive statistics summary for all data. Which are explained below:

**Mean** is the average of all values, calculated by summing the values and dividing by the number of observations.In this study, example can be mean weight of the dish, mean total energy of the dish and similar others in whole data set.

**Median** is the middle value when the data are sorted in ascending order. It is less sensitive to outliers than the mean. In this study, the example can be mid value between largest and smallest value of weight. If one dish had an unusually large portion (say 4000 g), the median would remain stable while the mean would increase sharply , it is not affected by outliners.

**Standard Deviation (SD)** is a measure of how spread out the data are around the mean. A higher SD indicates more variability in the data. In this study if the variation in the weight of dishes is large then it has high standard deviation. Lower Standard deviation means the dishes are more consistent across models.

**25th Percentile (Q1)** is the value below which 25% of the data fall. Also known as the first quartile. In this study it indicates the lower range of dish weights or nutrients. If Q1 for dish weight is 250 g, it says 25% dishes has weights less than 250 g.

**75th Percentile (Q3)** is the value below which 75% of the data fall. Also called the third quartile. In this study it indicates the upper range of dish weights or nutrient contents. If Q3 is 420 g then 75% dishes weights are less than 420 g.

**Maximum** is the largest observed value in the dataset. In this study It shows the heaviest dish or the dish with the highest nutrient content.

### Correlation Analysis

Checking relationships between different columns, e.g., weight and calories. The differences between pairs were not normally distributed. Correlation analysis (e.g., Pearson or Spearman correlation) checked if two variables tend to move together (strength and direction of relationship). More correlated variables were ignored in this study for further analysis as those don't have much impact on data analysis.

### Performance Evaluation Metrics

The goal of these tests was to figure out how close are the predicted weights and nutrients values to the true/reference values, how reliable are the methods in the estimation of nutrients and which method is better and gives less errors. Comparison between different models and approaches (LogMeal, OpenAI Vision, FoodSAM and different correction pipelines) in estimating the names, weights, nutrients and total energy was done by these tests. These estimation can never be perfect and they always can have errors compared to the actual measured or reference values. Advanced statistical methods used in this study were Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) , which are summarized in Table 4.2 and explained below:

**Mean Absolute Error (MAE)** measures the average size of the errors (differences) between predicted and actual values,without considering their direction (positive or negative). MAE was used in this study because it has given how close are the predicted data (weights and nutrients) values to the actual ones. This has accessed the accuracy and reliability of different prediction methods.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

where:

- $y_i$ = actual nutrient (or weight) value
- $\hat{y}_i$ = predicted nutrient (or weight) value
- $n$ = number of observations

A smaller MAE indicates that the model's predictions are closer to the actual values, on average.

**Root Mean Square Error (RMSE)** measures the square root of the average squared errors. This penalizes the larger errors as the errors are squared before averaging. This method was used because in this type of study the large errors are critical. If RMSE is much larger than MAE it means there can be very big mistakes in AI predictions and correction mechanisms.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

where:

- $y_i$ = actual nutrient (or weight) value
- $\hat{y}_i$ = predicted nutrient (or weight) value
- $n$ = number of observations

A lower RMSE indicated that the prediction errors are smaller and more consistent. It is sensitive to large deviations in predictions.

**Mean Absolute Percentage Error (MAPE)** can measures the average magnitude of the errors predicted as a percentage of the actual ground true values of the data. It explains us, on average, how distant are the predicted values (nutrients or weights) from the true values, relative to their actual size.

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

where:

- $y_i$ = actual nutrient (or weight) value
- $\hat{y}_i$ = predicted nutrient (or weight) value
- $n$ = number of observations

MAPE allowed to express the error as a percentage, making it easier to compare prediction accuracy across nutrients of different magnitudes. For example, a MAPE of 8% means that, on average, the model's nutrient predictions are off by 8% from the actual values.

The summary of these error evaluation metrics are in Table 4.2.

| Metrics | Definition | Purpose in This Study |
|---|---|---|
| **MAE (Mean Absolute Error)** | Average absolute difference between actual and predicted values. | Indicates the average prediction error in natural units (e.g., grams or kcal). Lower MAE value means better performance. |
| **RMSE (Root Mean Square Error)** | Square root of the mean squared differences between predicted and actual values. | Penalizes large errors more strongly. Lower RMSE value means better performance. |
| **MAPE (Mean Absolute Percentage Error)** | Prediction error displayed as a percentage of the actual value. | Relative measure of prediction accuracy for comparison across nutrients and weights with different scales. Lower MAPE value means better performance. |

Table 4.2: Evaluation Metrics for Model Performance in Predicting Dish Weights and Nutrients

# 5 Results and Evaluation

This section presents the results obtained from the experimental studies to evaluate the comparative accuracy of AI-based food recognition models. To evaluate the accuracy of AI-based nutritional estimation systems, approximately 350 images' data were collected. After filtering out incorrect, redundant, and unnecessary data, data from 167 samples were retained for the analysis. The dataset includes direct AI-based outputs, re-scaled outputs and menu-based corrected samples, enabling a comprehensive comparison with Lunch Line (LL) estimates.

## 5.1 Name Recognition Accuracy Comparison

First, after data extraction and cleaning, the unique names from various sources were identified, and a canonical mapping dictionary was created to normalize these names for comparison to answer the research question, RQ 1.1.

| Column Name | Unique Food Items Count |
|---|---|
| Food Component Names Lunch Line | 58 |
| Food Component Names by LogMeal | 98 |
| Food Component Names by FoodSAM | 54 |
| Food Component Names by OpenAI Vision | 264 |

Table 5.1: Unique Food Item Counts Across Different Systems

The above Table 5.1, shows the unique food items from different AI based sources. Highest number of food items are shown by OpenAI vision as 264 and minimum from FoodSAM which is 54.

Analysis of the names from different sources are displayed in Table 5.2. Analysis was done after normalizing the names from all the sources to nearest menu based names. Precision from all four sources LogMeal, FoodSAM and OpenAI vision shows almost a similar results 0.7. However, OpenAI vision shows more precision after normalizing the names to 0.77.

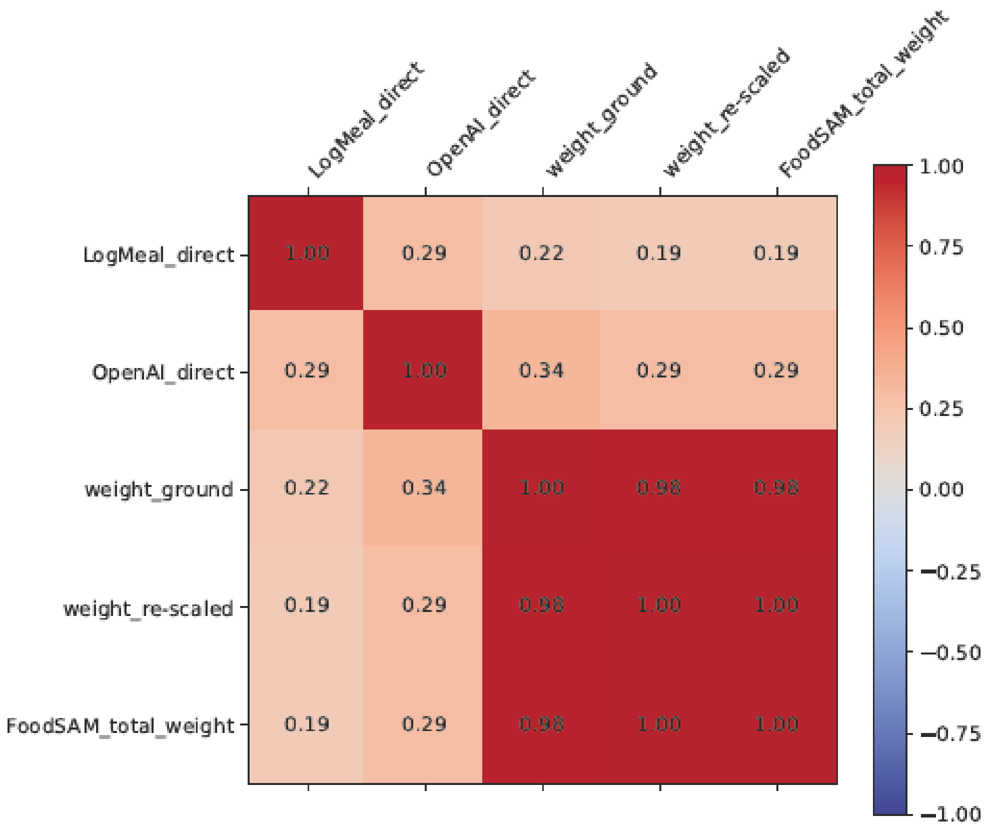
| **Methods** | **Precision** | **Recall** | **F1 Score** |
|---|---|---|---|
| LogMeal Normalized | 0.74 | 0.55 | 0.62 |
| FoodSAM Normalized | 0.72 | 0.70 | 0.70 |
| OpenAI Vision Normalized | 0.77 | 0.74 | 0.75 |
| Corrected Normalized Names | 0.74 | 0.65 | 0.70 |

Table 5.2: Comparison of accuracy of names from analysis methods

The recalls from FoodSAM 0.74 and the OpenAI 0.75 respectively.The recall from FoodSAM is 0.75 and OpenAI Vision is 0.75. Direct LogMeal based names shows the lower recall as 0.62. OpenAI vision corrected names also shows the lower recall 0.61.Detected Food Names (FoodSAM Norm) and OpenAI norm achieved the F1 score of 0.81.

Direct AI based approaches like (LogMeal's direct, OpenAI vision's direct) does not align well with the ground truth, displaying the inaccuracies in the estimation. The Heatmap's Figure 5.1 shows ground truth weight (weight ground) and re-scaled weight are highly correlated with each other (0.975–1.0), meaning the re-scaled weight from the system has given almost correct weight estimates. This re-scaled weight has been used for further correction of nutrients information later. FoodSAM based total weights and system's rescaled total weights are same because the total weight is used to get the weight of components, so in study FoodSAM is not included for further analysis.

Figure 5.1: Heatmap displaying corelation



## 5.2 Weight Accuracy Comparison

The accuracy of weight from LogMeal, OpenAI and FoodSAM is as shown in Table 5.3. The heatmap shows the corelation between approaches.

As shown in Table 5.3, the mean portion weight from lunch line is 490.81 g, which is almost similar to the re-scaled weight from the system, 491.4 g. The mean of LogMeal's direct weight is 760 g, which has vast difference from the ground truth weight by 271 g. However, a noticeable mean difference in total weight of approximately 0.59g was observed after system scaled is used.

| Summary | Lunch-Line/Ground truth (g) | LogMeal (g) | LogMeal vs Ground Truth | Open-AI Vision (g) | Open-AI vs Ground Truth | Device Scaled (g) | System re-scaled vs Ground Truth |
|---|---|---|---|---|---|---|---|
| Count | 167 | 167 | – | 167 | – | 167 | – |
| Mean | 490.81 | 760.18 | +54.88% | 363.85 | -25.87% | 491.40 | +0.12% |
| Std | 143.01 | 279.10 | +95.17% | 66.02 | -53.83% | 142.11 | -0.63% |
| Min | 239.00 | 140.00 | -41.42% | 170.00 | -28.87% | 240.00 | +0.42% |
| 25% | 393.00 | 559.50 | +42.31% | 330.00 | -16.03% | 392.00 | -0.25% |
| Median (50%) | 468.00 | 732.20 | +56.44% | 360.00 | -23.08% | 464.00 | -0.85% |
| 75% | 554.00 | 946.65 | +70.89% | 400.00 | -27.80% | 561.00 | +1.26% |
| Max | 997.00 | 2068.97 | +107.50% | 600.00 | -39.82% | 942.00 | -5.52% |

Table 5.3: Comparison of total weights and relative differences from the ground truth(Lunch Line)

The minimum to maximum variation from the ground truth with the LogMeal based direct weight is wide range spanning from 270 g to 1071 g, which is suggesting variability across individual samples. The error percentages between LogMeal vs ground truth, OpenAI vs ground truth is displayed in Table 5.3. LogMeal has over estimated the weights in average mean weights by +54.88% but OpenAI has underestimated the weights -25.87%. System re-corrected weight is almost same to real weight only 0.12% error.

| Source | Mean Absolute Error (g) | Root Mean Square Error (g) | Mean Absolute Percentage Error (%) |
|---|---|---|---|
| LogMeal Direct Weight | 267.37 | 312.50 | 70.48% |
| OpenAI Based Weight | 144.84 | 192.01 | 26.43% |
| Device re-scaled Weight | 23.64 | 31.75 | 5.21% |

Table 5.4: Comparison of total weights from different estimation methods

As shown in Table 5.4 which has addressed the research question RQ 1.2, The MAE for the Device re-scaled weight with ground truth is just 23.64, comparing to 267.37g for non-scaled LogMeal weight and OpenAI vision API based weight is 144.84 g.

## 5.3 Energy and Nutrients Accuracy Comparison

As shown in Table 5.5, the corrected total mean energy using the correction method is 533 kcal while the ground truth mean energy is 618.5 kcal. Total mean energy in kilocalories (kcal) from the LogMeal re-scaled weight is 513.16 kcal.

| **Summary** | **Lunch Line** | **LogMeal Direct** | **OpenAI Direct** | **OpenAI re-scaled** | **LogMeal re-scaled** | **Final Cor-rected** |
|---|---|---|---|---|---|---|
| Count | 167 | 167 | 167 | 167 | 167 | 167 |
| Mean | 618.5 | 790.86 | 442.00 | 601.92 | 513.99 | 533.16 |
| Std | 446.5 | 381.70 | 111.00 | 220.50 | 225.68 | 245.42 |
| Min | 155 | 102 | 180 | 232.75 | 132.73 | 125.22 |
| 25% | 387 | 546.50 | 375.00 | 445.86 | 352.75 | 376.80 |
| Median (50%) | 492 | 728.00 | 434.00 | 584.33 | 460.25 | 477.52 |
| 75% | 641.5 | 987.00 | 521.00 | 717.38 | 629.36 | 619.30 |
| Max | 2698 | 2539 | 739 | 1612.00 | 1382.36 | 1444.94 |

Table 5.5: Comparison of Energy estimations across the different methods

The OpenAI vision API based results 442 is bit closer than direct LogMeal based result which is 618.5 kcal according to the Table 5.5. The re-scaled values from LogMeal 514 kcal is better than the OpenAI vision based direct estimate 442 kcal. The LogMeal API's directly -guessed total energy is about 791 kcal and OpenAI vision API guessed value is 442 kcal and OpenAI vision rescaled is 601 kcal.

| **Approaches** | **Total Carbs (g)** | **Total Protein (g)** | **Total Fat (g)** | **Total Energy (kcal)** |
|---|---|---|---|---|
| MAE (Direct LogMeal vs Ground Truth) | 36.78 | 21.79 | 21.36 | 442.47 |
| MAE (OpenAI Vision Direct) | 18.70 | 7.54 | 7.00 | 252.24 |
| MAE (re-scaled LogMeal vs Ground truth) | 18.46 | 11.00 | 10.07 | 241.30 |
| MAE (OpenAI Vision Rescaled) | 15.30 | 10.51 | 10.85 | 245.13 |
| MAE Final (Corrected) | 16.90 | 6.03 | 7.61 | 234.65 |

Table 5.6: Mean Absolute Error (MAE) for macro nutrients and energy estimations

MAE comparisons is as shown in Table 5.6, for the total energy in kilocalories between the device final corrected data and LogMeal's direct data is 234.65 and 442.47 kcal respectively for research questions, RQ 2.1 and RQ 2.2. The OpenAI vision API based data in comparison with the lunch line-based ground truth data 252.4 kcal and OpenAI vision re-scaled is 245.13 kcal for research question RQ 1.3. For Carbs the final corrected result is 16.9 but direct LogMeal and OpenAI vision based errors are 36.78 and 18.70 respectively. Re-scaled LogMeal for Carbs from both OpenAI vision and LogMeal is 18.47 and 15.3 kcal respectively. The MAE for Protein and Fat is also in the table.

| Approaches | Total Carbs (g) | Total Protein (g) | Total Fat (g) | Total Energy (kcal) |
|---|---|---|---|---|
| RMSE (Direct LogMeal) | 45.70 | 30.64 | 29.8 | 601.32 |
| RMSE (LogMeal re-scaled) | 24.37 | 16.18 | 14.62 | 451.64 |
| RMSE (OpenAI vision direct) | 23.90 | 9.51 | 9.00 | 483.8 |
| RMSE (OpenAI vision re-scaled) | 20.11 | 14.7 | 14.13 | 446.57 |
| RMSE (Final Corrected) | 25.18 | 6.05 | 12.29 | 480 |

Table 5.7: Root Mean Square Error (RMSE) for macro nutrients and energy estimations

RMSE for the Total nutrients and total macro nutrients is displayed in Table 5.7. Direct AI based RMSE for total Energy from LogMeal and OpenAI vision are 601.32 and 483.8 kcal respectively, whereas the re-scaled values for LogMeal and OpenAI vision are 451.64 and 446.57 kcal respectively. The RMSE for other macro nutrients are displayed in table for research question RQ 1.3.

## 5.4 Component Based Analysis

The qualitative analysis of food components from AI based models versus ground truth components is described in Section 5.1. Weight and nutritional accuracy comparison between the dish components remains challenging due to inconsistent detection counts from the AI tools, fragmented and/or unequal number of segments and also incorrectly detected food items. To illustrate, we take the example data from 2 images out of our total datasets as shown in Table 5.8 to explain the research question RQ 1.5.

| **Ground Truth Components and Weights (Count)** | **LogMeal Components and Weights (Count)** | **FoodSAM Components and Weights (Count)** | **OpenAI Components and Weights (Count)** |
|---|---|---|---|
| ['*Broileria tomaatti-oliivikastikkeessa*' (Broiler in tomato olive sauce)*: 187g, 'Cooked rice': 72g, 'Green salad': 33g, 'Grated carrot': 19g] (4) | ['beet': 100g, 'rice': 100g, 'carrot': 70g, 'lettuce': 150g, 'cuban style rice': 279g] (5) | ['carrot': 2.75g, 'chicken duck': 207.72g, 'lettuce': 34g, 'other ingredients': 1g, 'red beans': 61.6g, 'rice': 30.34g, 'salad': 6.59g, 'sauce': 14.7g, 'tomato': 3.58g] (9) | ['Carrots': 30g, 'Beets': 50g, 'Rice': 50g, 'Lettuce': 20g, 'Chicken stew': 100g] (5) |

| **Ground Truth Components and Weights (Count)** | **LogMeal Components and Weights (Count)** | **FoodSAM Components and Weights (Count)** | **OpenAI Components and Weights (Count)** |
|---|---|---|---|
| ['Mashed potatoes': 262g, '*Lihapyörykät kermakastikkeessa* (meatballs in cream sauce)*': 206g, 'Green salad': 90g, 'Tomato': 79g, 'Grated carrot': 47g, 'Pickled cucumber': 36g] (6) | ['mashed potato': 223g, 'tomato': 123g, 'carrot': 70g, 'lettuce': 150g, 'meatballs': 263g, 'tomato': 123g, 'tomato': 123g, 'meatballs': 263g, 'rice with vegetables': 211g] (9) | ['carrot': 39g, 'lettuce': 29g, 'other ingredients': 16g, 'pepper': 41g, 'pork': 89g, 'potato': 185g, 'shiitake': 16g, 'steak': 47g, 'tomato': 245g] (9) | ['meatballs': 100g, 'mashed potatoes': 150g, 'tomatoes': 50g, 'shredded carrots': 30g, 'cabbage': 30g, 'pickles': 50g, 'sunflower seeds': 20g] (7) |

Table 5.8: Comparison of Food Component Detection and Weights: Ground Truth, LogMeal, FoodSAM, and OpenAI

The comparison between ground truth food components and LogMeal's detected outputs reveals significant differences in both the identification of food items and their estimated weights as shown in Table A.1. Overall, the system has often misclassified the food items, over or underestimated the number of components, and exhibits inconsistency in weight predictions.

For instance, in the first image example, the ground truth included four food items: Chicken in tomato-olive sauce(187 g), Cooked rice (72 g), Green salad (33 g), and Grated carrot (19 g). However, LogMeal has detected five components, including beet (100 g), rice (100 g), carrot (70 g), lettuce (150 g), and Cuban style rice (279 g). While some items like rice and carrot may loosely align with the ground truth, other components like beet and Cuban style rice was detected for the main protein *Broileria tomaatti-oliivikastikkeessa* (Broiler in tomato olive sauce)*. As shown in Table 5.8 for the same first image FoodSAM has detected too fragmented as 9 components,whereas OpenAI vision has detected 4 and OpenAI vision based names were almost aligned with the ground truth names.

Conversely, under-detection is also evident in more complex and culturally specific dishes as shown in 4th image data of the Table 5.8. For example, a dish containing five components—Tikka masala tofu (195 g), Basmati rice (166 g), Pickled cucumber (30 g), Green salad (28 g), and Grated carrot (21 g) was reduced by LogMeal to only two items: brown rice (70 g) and carrot (70 g). In terms of weight estimation also there were problems. For example, the LogMeal has estimated lettuce at 150 g when the ground truth salad component weighed only 28 g, while rice was underestimated at 70 g compared to the actual 166 g.There is same discrepancies from LogMeal and OpenAI vision as well.

### 5.4.1 Mapping the Components and Weights

In below analysis I have manually tried to map the food components for the image 1 and align the detected components from multiple sources with the Lunchline based menu for comparison. The total plate weight from the device is 362 g. Number of food components of the dish is: 4

#### Mapping LogMeal Components to menu

Now the food item is based on the above lunch line based menu as shown in Table 5.9 above here 'cuban style rice' is mapped as *Broileria tomaatti-oliivikastikkeessa* (Broiler in tomato olive sauce)*. Union of Beet and lettuce is the Green salad of the menu. Carrot is Grated carrot of the Menu. The weights are also now for the mapped names.

Now we have below table where I have mapped FoodSAM based names and weights

| **Ground Truth Components** | **Ground Truth Weights (g)** | **LogMeal Components** | **LogMeal Weights (g)** | **Mapped LogMeal Components** | **Mapped Weights (g)** |
|---|---|---|---|---|---|
| *Broileria tomaatti-oliivikastikkeessa* (Broiler in tomato olive sauce)*, Cooked rice, Green salad, Grated carrot | 187, 72, 33, 19 | beet, rice, carrot, lettuce, cuban style rice | 100.0, 100.0, 70.0, 150.0, 279.7 | *Broileria tomaatti-oliivikastikkeessa* (Broiler in tomato olive sauce)*, Grated carrot, Green salad (beet + lettuce), Cooked rice | 279, 70, 250, 100 |

Table 5.9: Comparison of Ground Truth and LogMeal Component Detection with Manually Mapped Labels and Weights)

with the menu based names and weights as shown in Table 5.10.

## Mapping FoodSAM components to menu

| **FoodSAM Components** | **FoodSAM Weights (g)** | **Mapped Components (Menu)** | **Mapped Weights (g)** |
|---|---|---|---|
| carrot, chicken duck, lettuce, other ingredients, red beans, rice, salad, sauce, tomato | 2.75, 207.72, 34.03, 0.72, 61.58, 30.34, 6.59, 14.70, 3.58 | Green salad, Chicken in tomato-olive sauce(chicken duck + tomato + sauce), Cooked rice, Grated carrot | 103, 226.00, 30.34, 2.75 |

Table 5.10: Comparison of Ground Truth and FoodSAM Component Detection (with Mapped Labels and Weights)

As shown in Table 5.10, manual approach was tried to fix the number of food components to match with the ground truth based result and the modified components based weights, which are displayed in the table. ['lettuce', 'other ingredients', 'red beans', salad] is the 'Green salad' of the Menu.['chicken duck','sauce', 'tomato'] is the '*Broileria tomaatti-oliivikastikkeessa* (Broiler in tomato olive sauce)*' of the Menu. So weight was also calculated by summing the fragments of major menu based components.

### Mapping OpenAI vision components to menu

The OpenAI vision based components are mapped with Menu as shown in Table 5.11.

| **OpenAI Components** | **OpenAI Weights (g)** | **Mapped Components (Menu)** | **Mapped Weights (g)** |
|---|---|---|---|
| carrots, Beets, Rice, Lettuce, Chicken stew | 30.0, 50.0, 50.0, 20.0, 100 | Green salad, *Broileria tomaatti-oliivikastikkeessa* (Broiler in tomato olive sauce)*, Cooked rice, Grated carrot | 70, 100, 50, 30 |

Table 5.11: Comparison of Ground Truth and OpenAI vision based components Detection (with Mapped Labels and Weights)

## 5.4.2 Comparative Analysis of Mapped Components for Weights

The component based analysis can be shown in Table 5.12.

| **Components Manually (Mapped)** | **Ground Truth (g)** | **LogMeal (g)** | **FoodSAM (g)** | **OpenAI (g)** | **System Corrected Components** | **System Corrected (g)** |
|---|---|---|---|---|---|---|
| Chicken in tomato-olive sauce | 187 | 279 | 226.00 | 100 | Chicken tomato olive sauce | 145 |
| Cooked rice | 72 | 100 | 30.34 | 50 | Cooked rice | 43 |
| Green salad | 33 | 250 (beet + lettuce) | 103 (lettuce + salad + others) | 70 (beets + lettuce) | Green Salad | 84 |
| Grated carrot | 19 | 70 | 2.75 | 30 | Grated carrot | 55 |
| | | | | | Rosolli | 48 |

Table 5.12: Comparison of Ground truth and Mapped Component Weights

Ground truth weights are realistic and consistent as shown in Table 5.12 e.g., chicken (187 g), rice (72 g), salad (33 g), carrot (19 g). LogMeal often overestimated the portion sizes, e.g., rice (279 g vs actual 72 g), lettuce (150 g vs actual 33 g). OpenAI Vision shows mixed estimates chicken stew (120 g vs actual 187 g), rice (100 g vs 72 g), carrot (70 g vs 19 g). Automatic correction by the system has displayed beet as Rosolli and 48 gram of weight as it was in the menu for the test day.

### 5.4.3 Comparative Analysis for Total Energy and Nutrients

As shown in Table B.1, the results of corrected energy and macro nutrients values for components are described for 1 selected image sample's data and explained the research question, RQ 1.5.

**Energy (Kcal)**

**Ground truth** has Chicken dish 181 kcal, rice 79 kcal, salad minimal (4 kcal), carrot 6 kcal and direct **LogMeal API's** values are very inflated due to oversized or undersized weights (e.g., rice: 207 kcal vs. 79 kcal ground truth). Again **OpenAI Vision** Underestimated the energy for chicken (120 kcal vs. 181 kcal) but overestimated for salad and carrot in some images. Final **Corrected System,** shows the chicken as 140 kcal vs. 181 kcal, rice 47 kcal vs. 79 kcal, carrot 19 kcal vs. 6 kcal.

**Carbohydrates(g)**

**Ground truth** rice provides (17 g), small contributions from salad (0.3 g) and carrot (1 g), whereas LogMeal has predicted very high rice carbs (24 g vs 17 g), carrot also inflated (4.5 g vs 1 g). **OpenAI vision** showed rice 14 g (quite close to ground truth), carrot 3 g (triple than the ground truth), salad as 0.5 g (slightly Over than ground truth). The **Corrected by System** shows rice as 10.3 g (under count 17 g), carrot 3.2 g (slightly Over), salad 0.8 g (over double).

**Protein(g)**

**Ground Truth** from Chicken (14.7 g), rice and vegetables contribute very little (1–2 g total). **LogMeal API** prediction for chicken is underestimated (5.4 g vs 14.7), rice is also underestimated (1.15 vs 1.77). The **OpenAI Vision** has underestimated the chicken (10 vs 14.7) and rice too (1.4 vs 1.77). **Corrected System** has provided the chicken as 11.3 g (closer, still 25 under truth), salad 1.05 (good), rice 1.05 (accurate).

**Fats(g)**

**Ground Truth** Chicken fats is 10.2 g, rice and vegetables have negligible (<0.5 g each).**LogMeal** based fat's value from the chicken is (10.3 vs 10.2, fine) but underestimated from rice (0.11 vs 0.35). **OpenAI** based fat from chicken is underestimated (6 vs 10.2), vegetables slightly off. The **Corrected System's** fat from Chicken is 8 g (under by 20%), rice 0.2 (slightly under) and for vegetables closer.

In this section, Finnish food names are presented in italics with their English translations in parentheses (e.g., *Lihapyörykät kermakastikkeessa* (meatballs in cream sauce)* (meatballs in cream sauce)*). * translated from Finnish

## 5.5 Dish Weights Impact in Accuracy

In this section the portion size based data is taken from 20 food items with the maximum weights and 20 food items with the minimum weights. Those data has been analyzed separately. The idea here is we are evaluating the accuracy based on the weight of meal for the research question, RQ 1.4.

### 5.5.1 Name Recognition Accuracy Compared to Ground Truth

The Table 5.13, shows that for the 20 least weight dishes, food item names from OpenAI and Final corrected methods achieved the highest F1 scores (0.77 and 0.76, respectively),

| Methods | Precision | Recall | F1 Score |
|---|---|---|---|
| Food item names LogMeal | 0.836 | 0.654 | 0.726 |
| Food item names FoodSAM | 0.758 | 0.770 | 0.754 |
| Food item names OpenAI | 0.763 | 0.799 | 0.770 |
| Final corrected names | 0.836 | 0.708 | 0.757 |

Table 5.13: Performance metrics for the 20 minimum weighted dishes

while detected food names from FoodSAM and food item names from LogMeal also performed reasonably well in precision and recall.

| Methods | Precision | Recall | F1 Score |
|---|---|---|---|
| Food item names LogMeal | 0.72 | 0.46 | 0.55 |
| Food item names FoodSAM | 0.64 | 0.59 | 0.61 |
| Food item names OpenAI | 0.71 | 0.68 | 0.68 |
| Final corrected names | 0.86 | 0.60 | 0.70 |

Table 5.14: Performance metrics for the 20 maximum weighted dishes

The Table 5.14 indicates that Final corrected names achieved the highest precision (0.86) and F1 score (0.70) among all methods for maximum weighed 20 images' data, showing the most accurate alignment with the ground truth. OpenAI and LogMeal names also performed well, with balanced precision and recall, while FoodSAM names showed moderate performance.

### 5.5.2 Weight Estimation Errors Compared to Ground Truth

This table 5.15 combines the errors in estimating the total energy (calories) and macro nutrients (Fat, Carbs, and Protein) for the minimum and maximum sized dishes.

| Methods | MAE (smallest dishes, g) | RMSE (smallest dishes, g) | MAE (largest dishes, g) | RMSE (largest dishes, g) |
|---|---|---|---|---|
| LogMeal Direct | 337.33 | 385.48 | 346.20 | 453.87 |
| Device Scaled | 22.00 | 30.94 | 16.85 | 22.13 |
| OpenAI Vision | 55.65 | 66.79 | 381.55 | 389.47 |

Table 5.15: Comparison of dish size based weight Estimation Errors

MAE smallest portions in Table 5.15 for comparison with the ground weight is 337 and 55 g for LogMeal direct and OpenAI vision whereas the device's re-scaled MAE is only 22 g.Re-scaling has reduced the MAE for weights for portion size as well, as described in Section 5.2. MAE for largest dish for direct LogMeal is 346 and 381.55 for OpenAI vision .RMSE info is also displayed in Table 5.15.

### 5.5.3 Energy and Nutrients Estimation Error

**Minimum sized dishes (images with least 20 ground truth weight**

The MAE for Total Energy for minimum sized 20 dishes are as shown in Table 5.16. The MAE is much higher 309.05 kcal for direct LogMeal and 143.25 kcal for OpenAI Vision but final correction has reduced the error to 113.94 kcal. MAE for Carbs, proteins and Fats are described in Table 5.16.

| Comparison | MAE (Total Energy, kcal) | MAE (Fat, g) | MAE (Carbs, g) | MAE (Protein, g) |
|---|---|---|---|---|
| LL vs LogMeal Direct | 309.05 | 16.10 | 33.60 | 11.65 |
| LL vs LogMeal re-scaled | 134.97 | 5.40 | 13.61 | 6.35 |
| LL vs OpenAI | 143.25 | 4.84 | 13.36 | 5.97 |
| LL vs OpenAI re-scaled | 138.73 | 5.20 | 12.99 | 5.18 |
| LL vs Corrected | 113.94 | 3.70 | 8.40 | 3.44 |

Table 5.16: Comparison of Energy and Macro nutrients (Minimum Ground Weighted dishes)

LogMeal re-scaled and OpenAI vision re-scaled show significantly lower MAE, suggesting that scaling these methods helps in more accurate energy estimation for smaller portions as shown in Section 5.3.

**Maximum sized dishes(images with maximum ground truth weight)**

As shown in Table 5.17, LogMeal direct continues to have the high MAE 395.6 comparing to OpenAI vision 294 kcal, pointing to its difficulty in accurately estimating energy for larger portions as described in Section 5.3.

| Comparison | MAE (Total Energy, kcal) | MAE (Fat, g) | MAE (Carbs, g) | MAE (Protein, g) |
|---|---|---|---|---|
| LL vs LogMeal Direct | 395.60 | 26.25 | 39.85 | 28.70 |
| LL vs LogMeal re-scaled | 220.69 | 18.77 | 27.20 | 21.71 |
| LL vs OpenAI | 294.05 | 11.15 | 39.08 | 9.38 |
| LL vs OpenAI re-scaled | 240.69 | 18.29 | 39.08 | 23.24 |
| LL vs Corrected | 249.41 | 14.61 | 36.25 | 14.94 |

Table 5.17: Comparison of Energy and Macro nutrients (Maximum Ground Weighted Dishes)

LogMeal re-scaled and OpenAI re-scaled still show lower MAE values 220.69 and 240.69 respectively , indicating that re-scaling these methods reduces the error even for larger portions. Table 5.17 shows errors for macro nutrients as well.

# 6 Discussion

This discussion chapter has interpreted the results of our study, highlighting the performance of different food recognition systems, explaining observed discrepancies and outlining implications for future research and improvements.

## 6.1 Performance Analysis of Name Accuracy

This section has analyzed the name accuracy comparison answering the research question, RQ 1.1. As shown in section 5.1, the normalized LogMeal-detected food names has achieved higher precision but lower recall and F1 score, indicating weaker performance (see Table 5.2). In contrast, the normalized OpenAI and OpenAI-corrected methods performed best overall, achieving the highest precision, recall, and F1 scores among all methods. This suggests that the OpenAI vision-based approach is the most effective for food component recognition in our dataset.

The segmentation and detected names from FoodSAM were more detailed and fragmented compared to the recipe menu. This has indicated that integrating and re-correcting the FoodSAM based data with the application's recipe menu presents additional challenges, as described in section 5.4. Components examples are shown in Table 5.8.

In contrast to the study by Javadi et al. [26], which reports accuracy based on only five images (insufficient for robust conclusions), our analysis uses a significantly larger dataset of 167 food images. Our approach integrates multiple pipelines, correction methods, and a modular design.

Furthermore, compared to the benchmark study by Van Asbroeck et al. [58], Calorie Mama achieved the highest Top-1 accuracy of approximately 63% among seven food recognition platforms, while LogMeal achieved only 48.6%. However, system's re-corrected names achieved a very competitive F1 score of 0.70. These explained results, together with the prior researches have highlighted that **name recognition alone remains an ongoing challenge** in food image analysis, especially for complex multi-item dishes. A detailed systematic review by Zhang et al. [57], has reported lower accuracy (where top-1 ranged from 63–9% and top 5 ranged from 88–24%).

## 6.2 Performance Analysis of Weight Accuracy

This section has analyzed the weight accuracy comparison answering the research question RQ 1.2. As shown in Table 5.3, the mean portion weight from the lunch line (ground truth) was 490.81 g, which is very close to the LogMeal API's re-scaled weight of 491.4 g. However, the LogMeal API's direct weight estimate was 760 g, which deviates significantly from the ground truth. In comparison, the OpenAI Vision API predicted 363 g, which is more aligned with the actual portion weight.

The standard deviations across metrics as shown in Table 5.3 were relatively high, reflecting the variability in the collected dataset. Minimum and maximum weights based analysis has further illustrated this as the weights deviation of dishes ranged from 239 g to 997 g.

As shown in Table 5.4, the MAE error in estimation of weights from OpenAI vision API (141.31 g) is lower compared to LogMeal API's direct guess (311.03 g). However, the LogMeal's approach using the re-scaled weight has significantly reduced the MAE result to 23.64 g. As described in a systematic review by Javadi et al. [26] the weight estimation error is from 6-23.69 percent for the sample test images which is more than our system's re-corrected results. These results display the limitations of AI models, where the direct

AI based results have high error in estimation of weights. As shown, the Flavoria Flex (system's) scale has re-scaled the weight and reduced the error significantly. As reviewed by Tahir and Loo (2021), ingredient recognition is a multi-label task that often suffers from misclassification causing error for portion level estimations [18].

## 6.3 Performance Analysis of Energy and Nutrients Accuracy

This section has analyzed the weight accuracy comparison answering the research question, RQ 1.3. As shown in section 5.3, the MAE from LogMeal and OpenAI vision based direct approach are more compared to re-scaled and re-corrected LogMeal and OpenAI vision based data in Table 5.6. This suggests potential over estimation for LogMeal and OpenAI vision API's directly guessed data compared to the re-scaled and re-corrected data. For example for energy, Flavoria Flex system's (Final re-corrected data) shows the lowest MAE (234.65 kcal).

Table 5.7 shows that the RMSE values for the total energy and macro nutrients of the dish. RMSE from direct Log Meal for the total energy is 601.32 kcal. For Carbs the OpenAI vision based methods are competitive and better. RMSE for re-scaled Log Meal based result is 451 kcal showing the impact of integrated re-scaled data for research question RQ 2.1. RMSE for OpenAI vision based direct result is 483 kcal and re-scaled OpenAI based value is 446.57 kcal. Final re-corrected total energy values using correction approach by the system is 480 kcal showing impact of menu based correction for research question RQ 2.2. For Fats and Carbs the OpenAI vision methods are competitive or better as shown in Table 5.7.

Large gaps between RMSE and MAE (like in Direct LogMeal) warn about the presence of big outlier errors.For example, Direct LogMeal has a big gap (Carbs: 36.78 vs 45.7), indicating some large outliers.Some methods like(e.g., LogMeal protein) show MAE =

RMSE, which implies errors are quite consistent with few large spikes.

For Protein, Final (Corrected) has the lowest MAE and RMSE (6.05), indicating very good accuracy and consistency. For Carbs the OpenAI vision based re-scaled has the lowest RMSE (20.11) and MAE (15.3), closely followed by system's Corrected value.For Fat, OpenAI vision direct prediction shows the lowest MAE (7.00) and RMSE (9.00), which is slightly better than final corrected from the system.

For Total Energy, Final corrected data shows the lowest MAE (234.65) explaining the research question RQ 2.2. The re-scaling and menu based correction consistently reduces the MAE and RMSE, significantly improving the accuracy. OpenAI vision methods performed well, mostly better than the raw LogMeal based results. In this studies data, the LogMeal (re-corrected) and OpenAI vision re-corrected method shows very strong performance.

A review titled "Evaluating the Quality and Comparative Validity of Manual Food Logging and Artificial Intelligence-Enabled Food Image Recognition in Apps for Nutrition Care" [13] evaluated 18 popular nutrition apps. It found that manual food-logging apps tend to overestimate the calories by up to 363 kcal, indicating that **this thesis'** approach has achieved better accuracy than the review. The study reported that MyFitnessPal underestimated energy intake by approximately 3%, while Foodvisor underestimated intake by approximately 47% on average. HealthifyMe and Fastic showed smaller but mixed discrepancies (8% overestimation and 44% overestimation, respectively). Notably, Foodvisor and Fastic exhibited the largest errors in energy estimation for some items. For simple foods such as boiled eggs and potato chips, estimation errors were smaller, but for mixed dishes, the discrepancies increased significantly, highlighting the challenges of estimating portion sizes and calories accurately from complex meals.

As systematically reviewed by Shonkoff et al. [54] titled "AI-based Digital Image Dietary Assessment Methods Compared to Humans and Ground Truth," the average overall relative errors (AI vs. ground truth) ranged up to 38.3% for calories, which is far higher than the system's re-corrected MAE errors [54]. Another systematic review by Mansouri et al. [23] reported MAE errors for calories and macronutrient estimations of up to 33%, again substantially higher than the re-corrected results presented in Table 5.7. As concluded in Shonrekoff et al. [54] and Zhang et al. [57], image-based dietary assessment methods hold potential for food recognition and nutrient estimation, but current systems still face low accuracy and are not yet reliable for practical use. Nonetheless, AI methods can match or exceed human accuracy in estimating nutrients from food images, provided that standardized testing and proper error reporting are implemented.

## 6.4 Performance Analysis of Component Based Accuracy Evaluation

This section has analyzed the component based accuracy comparison possibility answering the research question, RQ 1.5. As shown in the example Table 5.8 in section 5.4 of 4 selected images' data for the detection vs ground truth. LogMeal has over-detected or under-detected the food items, often duplicating the food components if they are spread in different place of dish(e.g., "tomato" listed multiple times in one of the image), and sometimes includes unrelated items (e.g., "Cuban style rice" instead of "cooked rice"), this happens because of biased trained data of LogMeal. FoodSAM shows the highest granularity, often identifying very small quantities (e.g., 2.75 g of carrot, 1 g of "other ingredients"), but also introduces ambiguous labels such as "other ingredients" or "salad. FoodSAM based results are better comparing to LogMeal and OpenAI vision but are more fragmented. OpenAI vision, on the other hand, returns more aggregated results,like "Chicken in Tomato Sauce" or "Mixed Vegetables," while not displaying the minor components. This leads to simplified outputs with generally fewer items.

Regarding the components based weight estimations as shown in Table 5.12, all the systems show varying degrees of error. For instance, in one dish, the actual rice weight was 166 g, while LogMeal estimated as 70 g, FoodSAM as 30.34 g, and OpenAI vision as 100 g. These discrepancies suggest significant challenges in visual portion size estimation, particularly in mixed or layered dishes. FoodSAM based components are based on the re-scaled weight and pixel area percentages of the segments in comparison to total area.The detection names from all the sources are mostly semantically similar comparing to the ground truth.

As shown in Table B.1, the re-corrected energy and macro nutrient values shows a significant deviation from the direct AI based estimations for components also for the example data. This shows the component based analysis is far inaccurate than the total weights and nutritional information based analysis as explained in previous sections and further analysis is required for analyzing this kind of data properly.

| **Criteria** | **LogMeal** | **FoodSAM** | **OpenAI** |
|---|---|---|---|
| **Avg. Components Detected** | High (often > GT) | Matches or slightly above | Low (abstracted items only) |
| **Redundant Detections** | Frequent (e.g., tomato x3) | Rare | None |
| **Granularity** | Moderate | High | Low (generalized items) |
| **Missed Key Components** | Yes (e.g., sauces) | Sometimes | Frequent |
| **Weight Accuracy** | Inconsistent | Slight underestimation | Tends to generalize |
| **Best For** | Comprehensive logs | Detailed nutrition analysis | Calorie estimation, simplicity |

Table 6.1: Qualitative Comparison of Food Recognition Systems

As reviewed by Tahir and Loo [18] ingredients recognition is a multi-label task that often suffers from fragmentation and misclassification in complex dishes as the result described above. Furthermore, the survey [18] has highlighted that single-image portion and volume estimation is inherently limited due to scale and depth ambiguities, which directly explains the weight errors observed in our component-based analysis. As described in [58] current commercial food image recognition platforms can identify food items but fail to accurately estimate portion sizes, severely limiting their utility for precise dietary assessment. Furthermore, the performance of analysis [58] on mixed dishes remains inconsistent, highlighting the need for improved methods to quantify individual food components as the result has discussed above.

Based on the manual checking of other 167 images data samples, Table 6.1 above has summarized the approaches and summaries. For the use cases requiring high granularity FoodSAM may be more appropriate. For applications prioritizing the simplicity and major ingredients tracking (e.g., calorie tracking apps), the OpenAI Vision model is better. LogMeal based approach is better in cases where we have to re-correct the segments name based on recipe-menu as it gives us the segments and Labels. For generalized items OpenAI Vision based approach is better and OpenAI Vision does not have redundant detections as well and the weights accuracy is better than other approaches.

The above conclusions for the research question RQ 1.5 to analyze the possibility for components based analysis indicate that while component-based analysis is theoretically feasible for dietary assessment, it faces substantial practical challenges that limit its reliability. Key issues include inconsistent component detection, large errors in weight estimation, misclassification of food items, and nutrient mismatches arising from database and recipe variations. Even after manual correction, discrepancies between estimated and true values remained significant as described in above sections.Significant methodological improvements is required for component based analysis.

## 6.5 Performance Analysis of Impact of Dish Weight in Accuracy

This section has analyzed the research question, RQ 1.4. Dish size also has impact in the detection of food components. The comparison of the two tables Table 5.13 and Table 5.14 shows that across all of the evaluated methods, OpenAI based detections and the final re-corrected names consistently achieved the higher precision and F1 scores than LogMeal and FoodSAM. The final re-corrected names combine high precision with improved recall, making them the most balanced and reliable method for identifying food items, especially for the least weighted dishes. Comparing the results with section 5.1 especially the least weighted dishes are higher in precision, recall and F1 score across all domains Table 5.2.

#### 20 images with least ground truth weight

MAE values are relatively higher for "LogMeal Direct" and "OpenAI Vision", indicating that these methods struggle more when estimating weights for smaller size.

LogMeal re-scaled showed much lower MAE and RMSE, suggesting that these methods are better at estimating weight for smaller portions, likely due to their ability to scale or adjust based on size.

#### 20 images with maximum ground truth weight

LogMeal Direct continues to show higher errors, especially in RMSE, indicating this method might perform poorly on large portions.

The system's re-scaled and FoodSAM methods still perform well with lower MAE and RMSE, suggesting their robustness across both smaller and larger portions.

OpenAI vision has high MAE for larger portions, indicating that while it is good for smaller portions, it struggles with larger ones, as seen in the higher error values.

The error increases when the weights increased, when dealing with larger portions (maximum weight) — particularly for LogMeal Direct and OpenAI Vision direct data.

#### Energy comparison on the basis of dish size

The tables Table 5.16 and Table 5.17 in estimating total energy (calories) and macro nutrients (Fat, Carbs, and Protein) MAE for the minimum 20 and maximum 20 dish size.

Re-scaling methods (like LogMeal re-scaled and OpenAI vision re-scaled) result in lower energy estimation errors, for smaller portions, showing that recalibrating the methods for the dish size leads to better accuracy.

Direct methods like LogMeal's Direct and OpenAI vision direct showed the higher errors for both portion sizes, especially for large portions similar as previous sections. Re-scaling and correction methods (like OpenAI re-scaled and Final Corrected) improved the macro nutrients estimation accuracy for both small and large portions, indicating the importance of adjusting the model based on portion size.

This analysis for the research question RQ 1.4 has highlighted how portion size (least vs. maximum ground truth weight) has significantly affected the estimation errors for both weight and macro nutrients. This suggests that re-scaling or re-correcting based on menu is essential for the accurate estimations across different dish sizes and larger the weight of food then larger is the errors in weights and nutrients estimations.

## 6.6 Limitations and Sources of Error

While the research methodology explained in this study addresses existing challenges in food detection, there are still several limitations, some of which stem from prior studies and some that are specific to this research. The system is not yet fully accurate, as the project is ongoing, and further research, as explained below, is required. Key sources of error include:

**Limitation of time and resources**

This is the most important factor in this advanced and broad research topic. The development of a more accurate system requires more time, research, and developers but due to limited resources, we were unable to implement many additional research ideas into the existing system for example creating our own model with our own data sets and investigating about versatile cameras, density volume related research, detail component based analysis and similar other are explained below. Mainly this study has used three models for food detection, but with more time and resources, more detection methods could have been tested and integrated in the system to improve the overall accuracy.

**Limitation in the studies**

While the study and the developed system has been expanded to increase the accuracy using re-scaled weight, re-corrected names and nutritional information based on menu however, the system still struggles with occlusion, portion variability, lighting issues and similar others. It has also highlighted that there is still much work to be done to overcome the challenges in food detection systems and achieve higher accuracy, particularly in real-world, uncontrolled environments.

**Inaccurately trained data**

The recognition systems exhibit biases reflecting their training data. LogMeal frequently produced Spanish dish names, FoodSAM generated Chinese or other Asian names, and OpenAI Vision tended to aggregate items into broad Western categories. In contrast, the ground truth and menu data is rooted in Finnish cuisine. This mismatch between the models' learned representations and the dataset is a major source of recognition and detection error.

**Lighting and image quality**

Variations in lighting,camera quality, reflections, or shadows and model's sensitivity has also affected the recognition accuracy.

**Food texture and appearance variability**

Different cooking styles, ingredient variations, or garnishes has also caused mis-classification in or case (for example food that contains soup, Chicken hidden by soup so incorrectly detected as soup).

### Inaccurate weight estimation

Volunteers sometimes added sauces and desserts on top of dishes, which were not captured by the Flavoria Lunchline weighing scales. This caused inaccuracies in the total ground-truth weight. Additionally, the pixel-level segmentation method used for component weight estimation lacks sufficient precision, as noted in the Results section.

### Occlusion and portion variability

Some dishes are partially hidden by other food or desserts and served in non-standard portions, leading to errors in both recognition, segmentation and weight estimation.

### Incomplete correction logic

The current re-correction method for AI-generated predictions is still under development. Alternative correction strategies, improved camera technology, and a density-volume database for weight estimation could enhance accuracy.

### Mixed and messy dishes

All models struggle with mixed or non-uniform dishes. Food components obscured by other items (e.g., sausages, sauces, desserts) or stacked under other items are often mis-detected, leading to recognition and weight estimation errors.

# 7 Conclusion

This thesis includes background and literature review sections, where the evaluation of existing food detection algorithms, the need for assessing the accuracy of food detection systems, and previous research approaches are discussed and tailored how this approach can be better the previous studies. It also covers the factors influencing the accuracy of food detection systems, potential approaches for improving current methods and how to measure results using the platform.

The machine vision-based automated food detection and recognition system, Flavoria Flex, used in this study has integrated AI based models in single application and simplified the diet assessment and analysis process. It has significantly reduced the time required for researchers by automating dietary intake data collection and evaluation. Furthermore, it eliminates memory-based bias, as data collection and analysis were conducted in real time. This section will answer the main research questions which are addressed in this thesis. In addition we have also answered the main future research direction and potential sources of errors as well.

## 7.1 Key Findings and Their Interpretation

Based on our analysis. The normalized names, re-scaled weight and menu based corrected result is far more accurate than direct AI based estimations as described in previous sections.

**Comparative Accuracy of Food Recognition Algorithms**

**RQ 1.1** *How do the accuracies of image recognition algorithms, including food-specific models such as LogMeal API and FoodSAM, and general-purpose image recognition models such as the OpenAI Vision API — compare in identifying various dishes against the ground truth names?*

Food recognition accuracy varies across the algorithms Section 5.1. OpenAI Vision showed the highest overall performance in food identification, followed by FoodSAM and LogMeal Table 5.2. In Qualitative analysis the OpenAI Vision model has performed competitively with food-specific models, even for complex or mixed dishes followed by corrected approach and FoodSAM.

**RQ 1.2** *How accurate are the total weights of the dishes estimated by each algorithm and estimated by the system compared to the ground truth weight?*

Direct AI-based estimations showed variable accuracy across models. Among them, the OpenAI vision method consistently achieved the lowest MAE and RMSE for weight estimation as shown in Table 5.4 but the weight from the device scale has lowest MAE and RMSE showing the importance of weight integration in the system.

**RQ 1.3** *How accurate are the total energy and macro nutrients of the dish compared to the ground truth?*

The total nutrients and macro nutrients estimation vary across the algorithms as shown in Table 5.6. The OpenAI vision has demonstrated better performance than LogMeal when compared with ground truth.

**RQ 1.4** *How consistent are the measurements across different weights of the food dishes?*

Consistency in performance varied by the weights of the food dishes, with the algorithms showing higher accuracy for lighter (lower-weight), uniform portions compared to heavier(maximum weighed) more heterogeneous meals.

**RQ 1.5** *What is the possibility of individual component based analysis based on the available test data?*

Component level analysis has significant challenges. The algorithms' estimation of individual food components' weights and nutrient values are generally inaccurate, often differing substantially from the ground truth as described **in the result and conclusion section in above Table 5.12**. Algorithms are able to identify which components are present in the dish as described in Section 5.1 and Section 5.4 but unable to provide reliable weights and dietary information. Further research is required to improve for components' level weights and nutrients estimation.

**Data Integration and Enhanced Accuracy**

**RQ 2.1** *How effective is the integration of a machine vision system with a weighing scale in improving nutrient estimation accuracy compared to direct image-based methods (e.g., LogMeal, OpenAI Vision)?*

The re-scaled LogMeal and re-scaled OpenAI vision based nutritional information has demonstrated the improved performance with the reduced MAE error when comparing with the ground truth (Table 5.6). These findings indicate that re-scaling machine vision system's data can significantly reduce the errors in estimating the total energy and macro nutrients relative to the ground truth.

**RQ 2.2** *How accurate are the nutrient estimates from the machine vision system integrated with a weighing scale when re-calibrated against the ground truth menu-based information?*

Accuracy improved significantly for all algorithms when re-scaled outputs were further re-calibrated using menu-based information, confirming that reference data is essential for more accurate estimation (Table 5.6). Re-calibration using menu-based ground truth further enhances the accuracy, producing the lowest MAE and RMSE.

**Applications and Implications**

**RQ 3.1** *How can the findings of this research be applied to develop more accurate and user-friendly food tracking applications?*

Findings suggest dietary applications can achieve higher accuracy by incorporating re-scaled image data and menu-based corrections. This modular system can allow integration of other AI models and re-correction logic. The system is able to display , extract and analyze the real time data with minimal effort which is really valuable even for researchers in nutrition science.

**RQ 3.2** *What implications do these findings have for nutrition research and AI-assisted dietary assessment?*

These findings has suggested that the dietary applications need to offer more accurate nutrient estimation. This modular approach has been tested in real restaurant environments and demonstrated potential for nutrition tracking system. The application design allows integration with other open-source models, supporting further research and development, and has implications for automated, real-time nutrient tracking for large-scale nutrition studies, and clinical applications requiring dietary intake data.

## 7.2 Recommendations for Future Research

Future research should focus on improving accuracy through multiple avenues. One of the key areas is the **integration of open-source models**. By incorporating additional accurate open-source models into the processing pipelines, we can enhance detection, segmentation, and nutrient estimation capabilities, providing a more robust system for food analysis.

Another important area for development is the **implementation of a density-volume database**. Such a database can map food densities to volumes, allowing for more accurate estimation of energy and macro nutrients from the segmented images. This will help to improve the reliability of nutrient estimation from visual data.

In addition, there is a need for **dedicated model training**. The collection of sufficient, domain-specific data tailored to specific use cases, such as Finnish cuisine for restaurants and kitchens in Finland, can enable the development of specialized models that can better handle regional variations in food and cooking styles.

Advancements in **camera technology** could also play a crucial role. Integrating depth cameras, 3D imaging, and multimodal sensing technologies have potential to improve portion size estimation and enhance component detection accuracy, especially in complex food items.

Improving the **controlled imaging conditions** is another avenue to explore. Optimizing lighting, camera angles and positions along with uniform backgrounds can reduce the variability in image capture, thereby improving the consistency and reliability of the analysis.

Additionally, more research on **component-based analysis** is necessary. Further studies should address the current limitations in identifying individual components of food, as discussed in the conclusion section, and help improve the reliability of such analyses.

Another important focus is **food waste tracking**. Monitoring and analyzing food waste from each dish can provide valuable insights into consumption patterns and help reduce discrepancies in dietary assessment.

Lastly, it is also easy to **calculate actual consumption**. By capturing images of leftovers post-consumption and comparing them with original portions, it is be possible to more accurately estimate actual energy and nutrient intake by the consumers, providing more precise dietary assessments.

# References

[1] A. L. Eldridge et al., “Evaluation of new technology-based tools for dietary intake assessment—an ilsi europe dietary intake and exposure task force evaluation”, *Nutrients*, vol. 11, no. 1, 2023. DOI: 10.3390/nu15040847. [Online]. Available: https://www.mdpi.com/2072-6643/11/1/55.

[2] A. Izbassar and P. Shamoi, “Image-based dietary assessment: A healthy eating plate estimation system”, *School of Information Technology and Engineering, Kazakh-British Technical University*, 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2403.01310.

[3] A.-J. Tessier et al., “Optimal dietary patterns for healthy aging”, *Nature Medicine*, vol. 31, pp. 1644–1652, 2025. DOI: 10.1038/s41591-025-03570-5. [Online]. Available: https://www.nature.com/articles/s41591-025-03570-5.

[4] C. T. McEvoy, N. Temple, and J. V. Woodside, “Vegetarian diets, low-meat diets and health: A review”, *Public Health Nutrition*, vol. 15, no. 12, pp. 2287–2294, 2012. DOI: 10.1017/S1368980012000936. [Online]. Available: https://www.cambridge.org/core/journals/public-health-nutrition/article/vegetarian-diets-lowmeat-diets-and-health-a-review/CFE7D0A7ADA80651A3DC03892287BABA.

[5] G. P. Danezis, A. S. Tsagkaris, F. Camin, V. Brusic, and C. A. Georgiou, “Food authentication: Techniques, trends & emerging approaches”, *TrAC Trends in Analyt-*

*ical Chemistry*, vol. 85, pp. 123–132, 2016, On-site and In-vivo Instrumentation and Applications. DOI: https://doi.org/10.1016/j.trac.2016.02.026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165993615301291.

[6] R. Fallaize et al., "Association between diet-quality scores, adiposity, total cholesterol and markers of nutritional status in european adults: Findings from the food4me study", *Nutrition and Health*, vol. 10, no. 1, p. 49, 2018. DOI: 10.1177/0260106017741056. [Online]. Available: https://doi.org/10.3390/nu10010049.

[7] M. Bryant et al., "Understanding and applying food systems approaches to improve dietary health", *Proceedings of the Nutrition Society*, pp. 1–9, Apr. 2025, Part of: Nutrition Society Congress 2024. DOI: 10.1017/S0029665125001661. [Online]. Available: https://www.cambridge.org/core/journals/proceedings-of-the-nutrition-society.

[8] L. Koivunen et al., "Increasing customer awareness on food waste at university cafeteria with a sensor-based intelligent self-serve lunch line", in *2020 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, Sep. 2020. DOI: 10.1109/ICE/ITMC49519.2020.9198571.

[9] J. Poore and T. Nemecek, "Reducing food's environmental impacts through producers and consumers", *Science*, vol. 360, no. 6392, pp. 987–992, 2018. DOI: 10.1126/science.aaq0216. [Online]. Available: https://www.science.org/doi/10.1126/science.aaq0216.

[10] C. Carvalho, D. Correia, S. A. Costa, C. Lopes, and D. Torres, "Assessing the environmental impact of diet – influence of using different databases of foods' environmental footprints", *Journal of Cleaner Production*, vol. 410, p. 137 973, 2023.

DOI: 10.1016/j.jclepro.2023.137973. [Online]. Available: https://doi.org/10.1016/j.jclepro.2023.137973.

[11] M. A. Clark, M. Springmann, J. Hill, and D. Tilman, "Multiple health and environmental impacts of foods", *Proceedings of the National Academy of Sciences*, vol. 116, no. 46, pp. 23 357–23 362, 2019. DOI: 10.1073/pnas.1906908116. [Online]. Available: https://www.pnas.org/doi/abs/10.1073/pnas.1906908116.

[12] Z. Liu, X. Yu, N. Liu, C. Liu, A. Jiang, and L. Chen, "Integrating ai with detection methods, iot, and blockchain to achieve food authenticity and traceability from farm-to-table", *Trends in Food Science & Technology*, vol. 158, p. 104 925, 2025. DOI: https://doi.org/10.1016/j.tifs.2025.104925. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924224425000615.

[13] X. Li, A. Yin, H. Y. Choi, V. Chan, M. Allman-Farinelli, and J. Chen, "Evaluating the quality and comparative validity of manual food logging and artificial intelligence-enabled food image recognition in apps for nutrition care", *Nutrients*, vol. 16, no. 15, p. 2573, 2024. DOI: 10.3390/nu16152573. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11314244/.

[14] H. Kassem, A. A. Beevi, S. Basheer, G. Lutfi, L. C. Ismail, and D. Papandreou, "Investigation and assessment of ai's role in nutrition—an updated narrative review of the evidence", *Nutrients*, vol. 17, no. 1, p. 190, 2025. DOI: 10.3390/nu17010190. [Online]. Available: https://doi.org/10.3390/nu17010190.

[15] B. Amoutzopoulos et al., "A systematic review of portion size estimation elements used in dietary assessment: Are image-based tools more accurate than household measures?", *Nutrition Reviews*, vol. 78, pp. 885–900, 2020. DOI: 10.1093/

nutrit/nuz107. [Online]. Available: https://doi.org/10.1093/nutrit/nuz107.

[16] S. Bhetuwal et al., "Automated image recognition system for determining energy composition of meals by ai-powered detection and identification of food items – a study utilizing flavoria flex", in *2025 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT)*, To appear, Morocco: IEEE, Nov. 2025.

[17] P. Zhou, W. Min, J. Song, Y. Zhang, and S. Jiang, "Synthesizing knowledge-enhanced features for real-world zero-shot food detection", vol. 33, pp. 1285–1298, 2024. DOI: 10.1109/TIP.2024.3360899. [Online]. Available: https://doi.org/10.1109/TIP.2024.3360899.

[18] G. A. Tahir and C. K. Loo, "A comprehensive survey of image-based food recognition and volume estimation methods for dietary assessment", vol. 9, no. 12, p. 1676, 2021. DOI: 10.3390/healthcare9121676. [Online]. Available: https://api.semanticscholar.org/CorpusID:237420412.

[19] A. Fakhrou, J. Kunhoth, and S. Al Maadeed, "Smartphone-based food recognition system using multiple deep cnn models", *Multimedia Tools and Applications*, vol. 80, no. 21, pp. 33 011–33 032, 2021. DOI: 10.1007/s11042-021-11329-6. [Online]. Available: https://doi.org/10.1007/s11042-021-11329-6.

[20] P. Bohlol, S. Hosseinpour, and M. S. Firouz, "A deep learning model with machine vision system for recognizing type of the food during the food consumption", *Scientific Reports*, vol. 15, no. 1, p. 29 734, 2025. DOI: 10.1038/s41598-025-29734-1. [Online]. Available: https://www.nature.com/articles/s41598-025-15755-6.

[21] A. T, P. T, K. Nfor, K. JI, and K. HC, “Applications of artificial intelligence, machine learning, and deep learning in nutrition: A systematic review”, *Nutrients*, vol. 16, no. 7, p. 1073, 2024. DOI: 10.3390/nu16071073. [Online]. Available: https://doi.org/10.3390/nu16071073.

[22] Y. Liu, H. Pu, and D.-W. Sun, “Efficient extraction of deep image features using convolutional neural network (cnn) for applications in detecting and analysing complex food matrices”, *Trends in Food Science & Technology*, vol. 113, pp. 193–204, 2021, ISSN: 0924-2244. DOI: https://doi.org/10.1016/j.tifs.2021.04.042. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924224421003022.

[23] M. Mansouri, S. B. Chaouni, and e. a. S. J. Andaloussi, “Deep learning for food image recognition and nutrition analysis towards chronic diseases monitoring: A systematic review”, *SN Computer Science*, vol. 4, p. 513, 2023. DOI: 10.1007/s42979-023-01972-1. [Online]. Available: https://link.springer.com/article/10.1007/s42979-023-01972-1.

[24] S. Kaushal, D. K. Tammineni, P. Rana, M. Sharma, K. Sridhar, and H.-H. Chen, “Computer vision and deep learning-based approaches for detection of food nutrients/nutrition: New insights and advances”, *Trends in Food Science & Technology*, vol. 146, p. 104 408, 2024, ISSN: 0924-2244. DOI: 10.1016/j.tifs.2024.104408. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924224424000840.

[25] C. Liu, Y. Cao, Y. Luo, G. Chen, V. Vokkarane, and Y. Ma, “Deepfood: Deep learning-based food image recognition for computer-aided dietary assessment”, Lecture Notes in Computer Science, vol. 9677, pp. 37–48, 2016. DOI: 10.48550/arXiv.1606.05675. [Online]. Available: https://doi.org/10.1007/978-3-319-39601-9_4.

[26] B. Javadi, A. Khan, N. Jayasuriya, and Y. Guo, "Smart nutrition monitoring system using serverless edge computing", *IEEE Transactions on Consumer Electronics*, vol. 70, no. 3, pp. 6363–6375, 2024. DOI: 10.1109/TCE.2024.3411552. [Online]. Available: https://ieeexplore.ieee.org/document/10552155.

[27] K. A. Nfor, T. A. T. Poupi, K. P. Ismaylovna, M. Joo, and H.-C. Kim, "An explainable cnn and vision transformer-based approach for real-time food recognition", *Nutrients*, vol. 17, no. 2, p. 362, 2025, ISSN: 2072-6643. DOI: 10.3390/nu17020362. [Online]. Available: https://www.mdpi.com/2072-6643/17/2/362.

[28] C. M. Badgujar, A. Poulose, and H. Gan, "Agricultural object detection with you only look once (yolo) algorithm: A bibliometric and systematic literature review", *Computers and Electronics in Agriculture*, vol. 223, p. 109 090, 2024. DOI: 10.1016/j.compag.2024.109090. [Online]. Available: https://doi.org/10.1016/j.compag.2024.109090.

[29] S. Sudharson, P. Kokil, R. Annamalai, and N. V. S. Manoj, "Enhanced food classification system using yolo models for object detection algorithm", in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Jul. 2023, pp. 1–6. DOI: 10.1109/ICCCNT56998.2023.10307943. [Online]. Available: https://doi.org/10.1109/ICCCNT56998.2023.10307943.

[30] H. Li and G. Yang, "Dietary nutritional information autonomous perception method based on machine vision in smart homes", *Entropy*, vol. 24, no. 7, p. 868, 2022. DOI: 10.3390/e24070868. [Online]. Available: https://doi.org/10.3390/e24070868.

[31] D. Sharma, R. Baviskar, V. Chavan, V. Sharma, L. Prasad, and D. H. Deshmukh, "A yolo-based image and video recognition system with integrated recipe recommendation functionality", in *Intelligent Computing and Communication Techniques*, 1st, CRC Press, 2025, p. 8. DOI: 10.1201/9781003530190. [Online]. Available: https://doi.org/10.1201/9781003530190.

[32] M. A. R. Alif and M. Hussain, *Yolov1 to yolov10: A comprehensive review of yolo variants and their application in the agricultural domain*, 2024. DOI: 10.48550/arXiv.2406.10139. [Online]. Available: https://arxiv.org/abs/2406.10139.

[33] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 779–788. DOI: 10.1109/CVPR.2016.91. [Online]. Available: https://doi.org/10.1109/CVPR.2016.91.

[34] M. Han, J. Chen, and Z. Zhou, *Nutrifyai: An ai-powered system for real-time food detection, nutritional analysis, and personalized meal recommendations*, 2024. arXiv: 2408.10532 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2408.10532.

[35] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas", *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1680–1716, 2023. DOI: 10.3390/make5040083. [Online]. Available: https://doi.org/10.3390/make5040083.

[36] Y. Zhao, D. Yang, S. Cao, B. Cai, M. Maryamah, and M. I. Solihin, "Object detection in smart indoor shopping using an enhanced yolov8n algorithm", *IET*

*Image Processing*, vol. 18, no. 14, pp. 4745–4759, 2024. DOI: 10.1049/ipr2.13284. [Online]. Available: https://doi.org/10.1049/ipr2.13284.

[37] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using yolo: Challenges, architectural successors, datasets and applications", *Multimedia Tools and Applications*, vol. 82, no. 6, pp. 9243–9275, Aug. 2022. DOI: 10.1007/s11042-022-13644-y. [Online]. Available: https://doi.org/10.1007/s11042-022-13644-y.

[38] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, *Mobilenetv2: Inverted residuals and linear bottlenecks*, 2018. DOI: 10.1109/CVPR.2018.00474. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/CVPR.2018.00474.

[39] S. Rokhva, B. Teimourpour, and A. H. Soltani, *Computer vision in the food industry: Accurate, real-time, and automatic food recognition with pretrained mobilenetv2*, 2024. DOI: https://doi.org/10.1016/j.foohum.2024.100378. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2949824424001538.

[40] R. Gavrilescu, C. Zet, C. Foşalău, M. Skoczylas, and D. Cotovanu, "Faster r-cnn: An approach to real-time object detection", in *2018 International Conference and Exposition on Electrical And Power Engineering (EPE)*, 2018, pp. 0165–0168. DOI: 10.1109/ICEPE.2018.8559776. [Online]. Available: https://doi.org/10.1109/ICEPE.2018.8559776.

[41] W. Li, "Analysis of object detection performance based on faster r-cnn", in *Journal of Physics: Conference Series*, vol. 1827, IOP Publishing, 2021, p. 012 085. DOI: 10.1088/1742-6596/1827/1/012085. [Online]. Available: https://doi.org/10.1088/1742-6596/1827/1/012085.

[42] J. Xu, H. Ren, S. Cai, and X. Zhang, "An improved faster r-cnn algorithm for assisted detection of lung nodules", *Computers in Biology and Medicine*, vol. 153, p. 106 470, 2023. DOI: 10.1016/j.compbiomed.2022.106470. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482522011787.

[43] P. K. Sethy, N. K. Barpanda, A. K. Rath, and S. K. Behera, "Rice false smut detection based on faster r-cnn", *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, no. 3, pp. 1590–1595, Sep. 2020, ISSN: 2502-4752. DOI: 10.11591/ijeecs.v19.i3.pp\%25p. [Online]. Available: http://ijeecs.iaescore.com.

[44] H. Nguyen, "Fast object detection framework based on mobilenetv2 architecture and enhanced feature pyramid", *Journal of Theoretical and Applied Information Technology*, vol. 98, no. 5, pp. 812–819, Mar. 2020, Published on 15th March 2020, ISSN: 1992-8645. [Online]. Available: http://www.jatit.org/volumes/Vol98No5/4Vol98No5.pdf.

[45] J. Xia, K. Wang, and H. Zhou, "Food packaging defect detection by improved network model of faster r-cnn", *Journal of Food Machinery*, vol. 39, no. 11, pp. 131–137, Dec. 2023. [Online]. Available: https://doi.org/10.13652/j.spjx.1003.5788.2023.80475.

[46] V. G. VijayaKumari, V. Priyanka, and P. Vishwanath, "Food classification using transfer learning technique", *Global Transitions Proceedings*, vol. 3, no. 1, pp. 225–229, 2022, International Conference on Intelligent Engineering Approach(ICIEA-2022), ISSN: 2666-285X. DOI: https://doi.org/10.1016/j.gltp.2022.03.027. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666285X22000334.

[47] L. Bu, C. Hu, and X. Zhang, "Recognition of food images based on transfer learning and ensemble learning", *Public Library of Science*, vol. 19, no. 1, 2024. DOI: 10.1371/journal.pone.0296789. [Online]. Available: https://doi.org/10.1371/journal.pone.0296789.

[48] B. N. Lakshmi, G. S. Shanmukh, T. N. Suraj, M. R. Reddy, and P. V. Rao, "Food recognition and calorie estimation using deep learning", *International Journal of Creative Research Thoughts (IJCRT)*, vol. 12, no. 3, 2024, ISSN: 2320-2882. [Online]. Available: https://www.ijcrt.org/papers/IJCRT2403977.pdf.

[49] M. Tan and Q. V. Le, "Efficientnetv2: Smaller models and faster training", pp. 10 096–10 106, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:232478903.

[50] S. J. Pan and Q. Yang, "A survey on transfer learning", *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010. DOI: 10.1109/TKDE.2009.191. [Online]. Available: https://ieeexplore.ieee.org/document/5288526.

[51] F. Zhuang et al., "A comprehensive survey on transfer learning", *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2021. DOI: 10.1109/JPROC.2019.2950664. [Online]. Available: https://ieeexplore.ieee.org/document/9134370.

[52] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning", *Journal of Big Data*, vol. 3, no. 1, pp. 1–40, 2016. DOI: 10.1186/s40537-016-0043-6.

[53] S. Bozinovski and A. Bozinovski, "The influence of training dataset size on transfer learning performance", in *2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO)*, IEEE, 2020, pp. 555–560.

DOI: 10.23919/MIPRO48935.2020.9245070. [Online]. Available: https://doi.org/10.1007/978-981-15-5971-6_83.

[54] E. Shonkoff et al., "Ai-based digital image dietary assessment methods compared to humans and ground truth: A systematic review", *Nutrition Research*, vol. 128, pp. 105–114, 2024, ISSN: 0271-5317. DOI: https://doi.org/10.1016/j.nutres.2024.07.002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0271531724000915.

[55] X. Lan et al., "Foodsam: Any food segmentation", *IEEE Transactions on Multimedia*, vol. 27, pp. 2795–2808, 2023. DOI: 10.48550/arXiv.2308.05938. [Online]. Available: https://doi.org/10.48550/arXiv.2308.05938.

[56] J. K. Chaudhary, "Algorithmic foundations for generalizable artificial intelligence models: A multi-domain study", Doctoral Dissertation, University of Turku, 2025. [Online]. Available: https://urn.fi/URN:ISBN:978-952-02-0260-6.

[57] S. Zhang, V. Callaghan, and Y. Che, "Image-based methods for dietary assessment: A survey", *Food Measure*, vol. 18, pp. 727–743, 2024. DOI: 10.1007/s11694-023-02247-2. [Online]. Available: https://doi.org/10.1007/s11694-023-02247-2.

[58] S. V. Asbroeck and C. Matthys, "Use of different food image recognition platforms in dietary assessment: Comparison study", *JMIR Formative Research*, vol. 4, no. 12, e15602, 2020. DOI: 10.2196/15602. [Online]. Available: https://formative.jmir.org/2020/12/e15602.

[59] D. K. N. Ho et al., "Validity of image-based dietary assessment methods: A systematic review and meta-analysis", *Clinical Nutrition*, vol. 39, no. 10, pp. 2945–2959, Oct. 2020, ISSN: 0261-5614. DOI: https://doi.org/10.1016/j.clnu.

2020.08.002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0261561420304039.

[60] G. A. Tahir, C. K. Loo, F. M. Moy, and N. A. Kong, *A review of critical features and general issues of freely available mhealth apps for dietary assessment*, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:235795257.

[61] J. Chen and X. Ran, "Deep learning with edge computing: A review", *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1655–1674, 2019. DOI: 10.1109/JPROC.2019.2921977. [Online]. Available: https://ieeexplore.ieee.org/document/8763885.

[62] M. Wyawahare, M. Rane, V. Patil, Y. Telkhade, R. Tele, and A. Walsepatil, "Nutri scan on desk: A food sensitivity testing system", in *2024 5th International Conference on Image Processing and Capsule Networks (ICIPCN)*, IEEE, 2024, pp. 198–203. DOI: 10.1109/ICIPCN63822.2024.00040. [Online]. Available: https://ieeexplore.ieee.org/document/10660819.

[63] M. Bolaños, M. Valdivia, and P. Radeva, "Where and what am i eating? image-based food menu recognition", in *Computer Vision–ECCV 2018 Workshops*, Springer International Publishing, 2019, pp. 590–605, ISBN: 978-3-030-11024-6. [Online]. Available: https://doi.org/10.1007/978-3-030-11024-6_45.

[64] A. Kirillov et al., "Segment anything", 2023. arXiv: 2304.02643 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2304.02643.

[65] L. A. Rahman, I. Papathanail, L. Brigato, and S. Mougiakakou, "A sam based tool for semi-automatic food annotation", *Frontiers in Artificial Intelligence and Applications (ECAI 2024)*, vol. 392, pp. 4475–4478, 2024. DOI: 10.3233/FAIA241033. [Online]. Available: https://doi.org/10.48550/arXiv.2410.19756.

[66] L. Chen, B. Wang, and J. Zhang, "Ingredsam: Open-world food ingredient segmentation via a single image prompt", *Journal of Imaging*, vol. 10, no. 12, p. 305, 2024. DOI: 10.3390/jimaging10120305. [Online]. Available: https://www.mdpi.com/2313-433X/10/12/305.

[67] A. AlMughrabi, A. Galán, R. Marques, and P. Radeva, "Foodmem: Near real-time and precise food video segmentation", *Pattern Recognition Letters*, vol. 192, pp. 59–64, 2025, ISSN: 0167-8655. DOI: https://doi.org/10.1016/j.patrec.2025.03.014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865525001060.

[68] A. G. Pacheco, "Foodmem: A fast and precise food video segmentation", Bachelor's Thesis, Universitat de Barcelona, Faculty of Mathematics and Computer Science, Barcelona, Jun. 2024. [Online]. Available: https://hdl.handle.net/2445/215722.

[69] S. Liu et al., "Llava-plus: Learning to use tools for creating multimodal agents", in *Computer Vision – ECCV 2024*, Springer, Nov. 2024, pp. 126–142. DOI: 10.1007/978-3-031-72970-6_8. [Online]. Available: https://doi.org/10.1007/978-3-031-72970-6_8.

[70] L. Lu, Y. Deng, C. Tian, S. Yang, and D. Shah, *Purrfessor: A fine-tuned multimodal llava diet health chatbot*, arXiv preprint arXiv:2411.14925, University of Wisconsin–Madison, 2024. [Online]. Available: https://arxiv.org/abs/2411.14925.

[71] B. Hieronimus, S. Hammann, and M. C. Podszun, "Can the ai tools chatgpt and bard generate energy, macro-and micro-nutrient sufficient meal plans for different dietary patterns?", *Nutrition Research*, vol. 128, pp. 105–114, 2024, ISSN: 0271-5317. DOI: https://doi.org/10.1016/j.nutres.2024.07.002. [Online].

Available: https://www.sciencedirect.com/science/article/pii/S0271531724000915.

[72] M. Cascella, J. Montomoli, V. Bellini, and E. Bignami, “Evaluating the feasibility of chatgpt in healthcare: An analysis of multiple clinical and research scenarios”, *Journal of Medical Systems*, vol. 47, no. 1, p. 33, 2023. DOI: 10.1007/s10916-023-01925-4. [Online]. Available: https://doi.org/10.1007/s10916-023-01925-4.

[73] G. Ciocca and P. Napoletano, “Food recognition: A new dataset, experiments, and results”, *IEEE Journal of Biomedical and Health Informatics*, vol. 21, pp. 588–598, 2017. DOI: 10.1109/JBHI.2016.2636441. [Online]. Available: https://api.semanticscholar.org/CorpusID:2040113.

[74] M. Veeramreddy, A. K. Pradhan, S. Ghanta, L. Rachakonda, and S. P. Mohanty, *Nutrivision: A system for automatic diet management in smart healthcare*, 2024. [Online]. Available: https://arxiv.org/abs/2409.20508.

[75] A. Holappa, O. Myllymäki, and J. Sotaniemi, “Nutrivision: Harnessing ai for personalized healthy food recommendations”, M.S. thesis, University of Oulu, Jun. 2025. [Online]. Available: https://oulurepo.oulu.fi/bitstream/handle/10024/56684/nbnfioulu-202506064191.pdf.

[76] Ibraheem, A. Raza, M. Mubarik, Q. U. N. Kamal, and S. Ayaz, “Nutrivision: Early detection of malnutrition using advanced deep learning”, in *2025 International Conference on Emerging Technologies in Electronics, Computing, and Communication (ICETECC)*, 2025, pp. 1–7. DOI: 10.1109/ICETECC65365.2025.11070276. [Online]. Available: https://ieeexplore.ieee.org/document/11070276.

[77] J. Sun, K. Radecka, and Z. Zilic, “Foodtracker: A real-time food detection mobile application by deep convolutional neural networks”, 2019. arXiv: 1909.05994 [cs.CV]. [Online]. Available: https://arxiv.org/abs/1909.05994.

[78] Y. Luo, P. Liu, and E. K. Choe, “Co-designing food trackers with dietitians: Identifying design opportunities for food tracker customization”, in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*, Glasgow, Scotland, UK: ACM, May 2019, pp. 1–13, ISBN: 9781450359702. DOI: 10.1145/3290605.3300881. [Online]. Available: https://doi.org/10.1145/3290605.3300881.

[79] S. Davis and A. Hall. “Best calorie counting apps of 2025”. Audited & Verified: Jul 18, 2024, 8:41am. Accessed: Oct 22, 2025, Forbes Health. [Online]. Available: https://www.forbes.com/health/wellness/best-calorie-counting-apps/.

[80] X. Wan, Y. Song, H. Zhou, and M. Shao, “Layered double hydroxides for oxygen evolution reaction towards efficient hydrogen generation”, *Energy Material Advances*, vol. 2022, pp. 1–17, 2022. DOI: 10.34133/2022/9842610. [Online]. Available: https://www.researchgate.net/publication/363378040_Layered_Double_Hydroxides_for_Oxygen_Evolution_Reaction_towards_Efficient_Hydrogen_Generation.

[81] X. Wan, Y. Song, H. Zhou, and M. Shao, “Layered double hydroxides for oxygen evolution reaction towards efficient hydrogen generation”, *Energy Material Advances*, vol. 2022, pp. 1–17, Sep. 2022. DOI: 10.34133/2022/9842610. [Online]. Available: https://doi.org/10.34133/2022/9842610.

[82] T. Sarapisto, L. Koivunen, T. Mäkilä, A. Klami, and P. Ojansivu, “Camera-based meal type and weight estimation in self-service lunch line restaurants”, in *Proceedings of the 2022 12th International Conference on Pattern Recognition Systems*

*(ICPRS)*, 2022, pp. 1–7. DOI: 10.1109/ICPRS54038.2022.9854056. [Online]. Available: https://doi.org/10.1109/ICPRS54038.2022.9854056.

[83] Finnish Institute for Health and Welfare (THL), *Food diary (fineli)*, https://fineli.fi/fineli/en/ruokapaivakirja, Accessed: 2025-09-13, 2025.

# Appendix A Example Of the Components Names And Weights Data

| Ground Truth Components and Weights (Count) | LogMeal Components and Weights (Count) | FoodSAM Components and Weights (Count) | OpenAI Components and Weights (Count) |
|---|---|---|---|
| ['Broileria tomaatti-oliivikastikkeessa': 187g, 'Cooked rice': 72g, 'Green salad': 33g, 'Grated carrot': 19g] (4) | ['beet': 100g, 'rice': 100g, 'carrot': 70g, 'lettuce': 150g, 'cuban style rice': 279g] (5) | ['carrot': 2.75g, 'chicken duck': 207.72g, 'lettuce': 34g, 'other ingredients': 1g, 'red beans': 61.6g, 'rice': 30.34g, 'salad': 6.59g, 'sauce': 14.7g, 'tomato': 3.58g] (9) | ['Carrots': 30g, 'Beets': 50g, 'Rice': 50g, 'Lettuce': 20g, 'Chicken stew': 100g] (5) |

| **Ground Truth Components and Weights (Count)** | **LogMeal Components and Weights (Count)** | **FoodSAM Components and Weights (Count)** | **OpenAI Components and Weights (Count)** |
|---|---|---|---|
| ['Mashed potatoes': 262g, 'Lihapyörykät kermakastikkeessa': 206g, 'Green salad': 90g, 'Tomato': 79g, 'Grated carrot': 47g, 'Pickled cucumber': 36g] (6) | ['mashed potato': 223g, 'tomato': 123g, 'carrot': 70g, 'lettuce': 150g, 'meatballs': 263g, 'tomato': 123g, 'tomato': 123g, 'meatballs': 263g, 'rice with vegetables': 211g] (9) | ['carrot': 39g, 'lettuce': 29g, 'other ingredients': 16g, 'pepper': 41g, 'pork': 89g, 'potato': 185g, 'shiitake': 16g, 'steak': 47g, 'tomato': 245g] (9) | ['meatballs': 100g, 'mashed potatoes': 150g, 'tomatoes': 50g, 'shredded carrots': 30g, 'cabbage': 30g, 'pickles': 50g, 'sunflower seeds': 20g] (7) |
| ['Pesto pasta salad': 259g, 'Green salad': 79g, 'Grated carrot': 72g, 'Peas': 26g] (4) | ['carrot': 70g, 'green peas': 100g, 'pasta with pesto': 174g] (3) | ['cabbage': 21.5g, 'carrot': 128g, 'green beans': 41.4g, 'potato': 170g, 'tomato': 55g] (5) | ['Pesto Pasta': 150g, 'Grated Carrots': 50g, 'Peas': 30g] (3) |
| ['Tikka masala tofusta': 195g, 'Basmati rice': 166g, 'Pickled cucumber': 30g, 'Green salad': 28g, 'Grated carrot': 21g] (5) | ['brown rice': 70g, 'carrot': 70g] (2) | ['carrot': 6g, 'chicken duck': 38g, 'cucumber': 49g, 'fried meat': 34g, 'lettuce': 27g, 'other ingredients': 12g, 'rice': 304g, 'tomato': 3g] (8) | ['Rice': 150g, 'Paneer Curry': 100g, 'Mixed vegetables': 50g] (3) |

| Ground Truth Components and Weights (Count) | LogMeal Components and Weights (Count) | FoodSAM Components and Weights (Count) | OpenAI Components and Weights (Count) |
|---|---|---|---|

Table A.1: Comparison of Food Component Detection and Weights: Ground Truth, LogMeal, FoodSAM, and OpenAI

# Appendix B Detailed Nutritional Data Table

Table B.1: Comparison of Ground Truth, AI Systems (LogMeal and OpenAI), and Corrected System Nutritional Values for One Sample Case (per food item)

| Source | Food Item and Weight | Energy | Carbs (g) | Protein (g) | Fat (g) |
|---|---|---|---|---|---|
| Ground Truth | Chicken Tomato sauce: 187g | 181 | 6.7 | 14.7 | 10.23 |
| | Cooked rice: 72g | 79 | 17.24 | 1.77 | 0.35 |
| | Green salad: 33g | 4.3 | 0.33 | 0.37 | 0.066 |
| | Grated carrot: 19g | 6.4 | 1.0 | 0.11 | 0.04 |
| | Beet: — | — | — | — | — |
| LogMeal AI | Cuban style rice: 279g | 207 | 24.08 | 5.44 | 10.29 |
| | Rice: 100g | 56 | 12.06 | 1.15 | 0.11 |
| | Lettuce: 150g | 14.2 | 0.72 | 1.02 | 0.25 |
| | Carrot: 70g | 19.2 | 4.52 | 0.41 | 0.09 |
| OpenAI AI | Chicken Stew: 120g | 120 | 8 | 10 | 6 |
| | Rice: 100g | 65 | 14 | 1.4 | 0.2 |
| | Lettuce: 150g | 3 | 0.5 | 0.2 | 0 |
| | Carrot: 70g | 12 | 3 | 0.3 | 0.1 |
| Corrected by System | Chicken Tomato sauce: 140.43g | 140.42 | 5.18 | 11.30 | 8 |
| | Cooked Rice: 43g | 47.33 | 10.3 | 1.05 | 0.20 |
| | Green Salad: 84g | 11.08 | 0.83 | 1.05 | 0.34 |
| | Grated Carrot: 55g | 18.71 | 3.17 | 0.34 | 0.11 |
| | Rosolli: 48g | 19.2 | 3.26 | 0.43 | 0.00 |

# Appendix C Example Prompts Used with AI Tools

This section includes example prompts used with AI tools during the development of this thesis. These prompts were used to support writing, clarify technical concepts, and assist in image-based food detection tasks.

## ChatGPT Prompts

- **Prompt 1:** "Can you help me to improve the grammar and clarity of this paragraph on food image segmentation?"
- **Prompt 2:** "Act yourself as an expert in AI algorithms and explain the difference between CNN and YOLO in simple terms."
- **Prompt 3:** "Act yourself as an expert translator and suggest a more formal Finnish way to phrase the sentence: '."
- **Prompt 4:** "Suggest a more academic way to phrase the sentence: 'The system was good to detect food components in the images.'"
- **Prompt 5:** "Can you translate this sentence to Finnish the sentence is: The system can be developed using AI models, sensors and laptop as server."
- **Prompt 6:** "Act yourself as a food expert and answer me what is English trans-

lation for below food? Is this common food item in English dishes? : 'Broileria tomaattikastikkeessa'."

## OpenAI Vision (GPT-4 with Image Input) Prompts

- **Prompt 4:** [Uploaded image of a dish] "What food items can you identify in this image?"
- **Prompt 5:** [Uploaded image of a misclassified dish] "Can you point out what's incorrect in this food classification?"
- **Prompt 5:** [Uploaded image of a dish] "What food items can you identify in this image, can you please give me the total and individual food components weights and nutritional information using your vision abilities?"