

Image-Based Deep Learning System for Food Identification and Nutritional Profiling

Dr. C. Sridevi¹, M Vimalraj², R K Salini³, and U Kishore Kumar⁴

Department of Electronics Engineering, MIT Campus, Anna University, Chromepet, Chennai – 600044, India

Dr. C. Sridevi is an Associate Professor with the Department of Electronics Engineering, MIT Campus, Anna University, Chromepet, Chennai – 600044, India

ABSTRACT: Accurate dietary monitoring is essential for maintaining healthy eating habits and preventing lifestyle-related diseases. However, most existing calorie-tracking systems depend on manual food entry, which is time-consuming and often inaccurate. Although recent advances in deep learning have enabled image-based food recognition, reliable portion size and calorie estimation remain challenging, especially for Indian cuisine due to its wide variety of dishes and complex visual characteristics. This paper presents an image-based deep learning framework for automated food identification and nutritional profiling, with a specific focus on Indian food items. The proposed system integrates food detection, classification, segmentation-based portion size estimation, and calorie computation into a unified pipeline. Food classification is performed using VGG16 and YOLOv8n models, with YOLOv8n achieving superior performance and a mean Average Precision (mAP@50) of 89.6%. For food segmentation, K-means clustering, SVM-based segmentation, and YOLOv8s are evaluated, where YOLOv8s demonstrates the lowest area estimation error. The complete framework is deployed as a web-based application that enables users to upload food images, verify detected items, and obtain detailed nutritional information. Experimental results show that the proposed system provides an effective and practical solution for automated dietary assessment of Indian cuisine.

INDEX TERMS Food recognition, deep learning, YOLOv8, food segmentation, calorie estimation, nutritional profiling, web-based application.

I. INTRODUCTION

Maintaining a balanced diet is a fundamental requirement for leading a healthy lifestyle and preventing nutrition-related disorders such as obesity, diabetes, and cardiovascular diseases. With increasing awareness of health and fitness, there is a growing demand for systems that can accurately monitor daily food intake and nutritional consumption. However, most existing dietary tracking solutions rely on manual food logging, where users must enter food names and portion sizes. This approach is often inconvenient, time-consuming, and highly dependent on user memory and estimation, leading to inaccurate nutritional records.

In recent years, advancements in deep learning and computer vision have enabled automated analysis of food images, offering a promising alternative to manual dietary tracking. Image-based food recognition systems aim to identify food items directly from images captured using smartphones and estimate their nutritional content. Such systems reduce user effort and improve consistency in dietary monitoring. Despite these advantages, accurate food recognition and calorie estimation remain challenging due to variations in food appearance, lighting conditions, portion size, and food presentation.

These challenges are further amplified in the case of Indian cuisine. Indian food exhibits significant diversity in terms of ingredients, preparation styles, textures, and color patterns. Many dishes consist of mixed components, gravies, and side dishes served together on a single plate, making food detection and segmentation more complex. Additionally, the same food item can vary significantly in calorie content depending on cooking methods and ingredient proportions. As a result,

models trained primarily on Western food datasets often fail to generalize effectively to Indian food items.

Several existing studies focus mainly on food classification without considering portion size estimation, which is a critical factor in accurate calorie computation. Other approaches estimate calorie values using predefined portion assumptions, which limits their reliability in real-world scenarios. Furthermore, only a limited number of works integrate food recognition, segmentation, calorie estimation, and user interaction into a complete deployable system.

To address these limitations, this paper proposes an image-based deep learning framework for automated food identification and nutritional profiling, specifically tailored for Indian cuisine. The proposed system integrates food detection and classification using YOLOv8, segmentation-based portion size estimation, and calorie computation into a unified pipeline. In addition, the framework is deployed as a web-based application that enables real-time user interaction, food verification, and nutritional analysis.

The main contributions of this work are summarized as follows:

- Development of a deep learning-based food classification system optimized for Indian food datasets.
- Comparative evaluation of multiple segmentation techniques to identify the most accurate method for portion size estimation.
- Integration of classification, segmentation, and nutritional profiling into an end-to-end automated framework.
- Deployment of the proposed system as a user-friendly web-based application for real-world dietary monitoring.

The remainder of this paper is organized as follows. Section II reviews related work in food image analysis and dietary assessment. Section III describes the proposed methodology in detail. Section IV presents the experimental results and performance analysis. Section V discusses the results and identifies common challenges and sources of error. Finally, Section VI concludes the paper and outlines future research directions.

II. RELATED WORK

Automated food recognition and nutritional assessment have received increasing attention in recent years due to their potential applications in healthcare, fitness monitoring, and lifestyle management [1]. Early research in this domain primarily relied on handcrafted visual features such as color histograms, texture descriptors, and shape-based representations combined with traditional machine learning classifiers. Although these approaches showed promising results in controlled environments, their performance was often affected by changes in lighting conditions, background clutter, and food presentation, which limited their reliability in real-world applications [2].

The rapid advancement of deep learning has significantly transformed food image analysis. Convolutional neural networks (CNNs) have emerged as the dominant approach for food classification tasks due to their ability to automatically learn discriminative features from raw image data. Gupta *et al.* demonstrated the effectiveness of transfer learning for nutrition monitoring, showing that pretrained CNN models can accurately recognize food items while reducing training complexity [1]. Similarly, Jiang *et al.* introduced the DeepFood framework, which leverages deep neural networks for food image analysis and calorie estimation, highlighting the capability of deep models to capture complex visual patterns present in food images [2].

To further improve recognition performance, several studies have explored fine-grained food classification techniques. Arslan *et al.* investigated fine-grained classification methods using the UEC FOOD-100 dataset and emphasized the importance of learning subtle visual differences between visually similar food categories [6]. However, most classification-oriented approaches assume fixed or predefined portion sizes, which restricts their effectiveness for accurate calorie estimation in real-world scenarios.

Recognizing the importance of portion size estimation, researchers have incorporated segmentation-based techniques to isolate food regions from the background. Traditional segmentation methods such as K-means clustering and support vector machine (SVM)-based segmentation have been widely used due to their simplicity and low computational cost [7]. Nevertheless, these methods often struggle with complex backgrounds, overlapping food items, and mixed dishes, which are common in practical food images.

More recent studies have adopted deep learning-based object detection and segmentation frameworks to address these challenges. Models such as YOLO and Mask R-CNN [3], [8] enable simultaneous localization and classification of multiple food items within a single image, making them

suitable for real-time dietary assessment applications [3], [8]. Additionally, hybrid approaches that combine vision-based analysis with nutritional databases have been proposed to enhance calorie estimation accuracy [7].

Despite these advancements, relatively limited research has focused on end-to-end food recognition systems specifically designed for Indian cuisine. Indian food presents unique challenges due to its diversity, mixed ingredients, and wide variations in preparation styles. Moreover, only a few existing works integrate food classification, segmentation, calorie estimation, and user interaction into a deployable web-based platform. In contrast, the proposed system addresses these gaps by focusing on Indian food items and providing a complete pipeline that combines deep learning-based analysis with a practical web-based application for nutritional profiling.

III. PROPOSED METHODOLOGY

The proposed system is designed as an end-to-end deep learning framework for automated food identification and nutritional profiling. The methodology integrates multiple stages, including food image acquisition, food detection and classification, segmentation-based portion size estimation, calorie computation, and result visualization through a web-based application. The overall workflow ensures minimal user intervention while maintaining reliable accuracy in real-world scenarios.

A. Overall System Architecture

The overall architecture of the proposed system follows a modular pipeline, where each module performs a specific function and contributes to the final nutritional analysis. Initially, a food image is captured using a smartphone or uploaded through the web interface. The image is then processed by the food detection and classification module to identify the food items present. Following this, a segmentation model is applied to estimate the food region and portion size. Based on the estimated portion size or user-provided weight, the system computes the nutritional values and displays the results to the user.

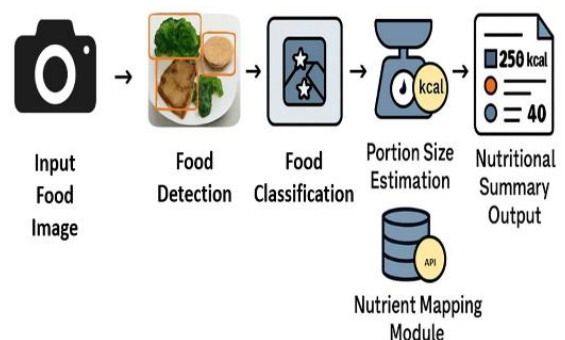


Figure 1: Overall system architecture

This modular design improves system scalability and allows individual components to be enhanced independently. It also enables seamless integration of additional food categories and nutritional parameters in future extensions.

B. Food Detection and Classification

Food detection and classification represent the first and most important stage of the proposed framework. In this work, two deep learning models are explored for this purpose: a transfer learning-based VGG16 model and a YOLOv8n model. The VGG16 model is used as a baseline approach and is fine-tuned using an Indian food dataset so that it can better recognize region-specific dishes. Although this model provides acceptable classification results, it follows a separate preprocessing and classification pipeline, which increases computation time and makes it less efficient for real-time use. On the other hand, the YOLOv8n model offers a more practical and faster solution by performing food detection and classification together in a single network. This combined approach allows the system to identify and locate multiple food items in one image without additional processing steps. The model is trained on annotated Indian food images and optimized for lightweight and real-time performance. Experimental results show that YOLOv8n achieves a mean Average Precision (mAP@50) of 89.6%, clearly outperforming the VGG16-based model in terms of speed, accuracy, and overall efficiency. These results indicate that YOLOv8n is more suitable for real-world food recognition applications.

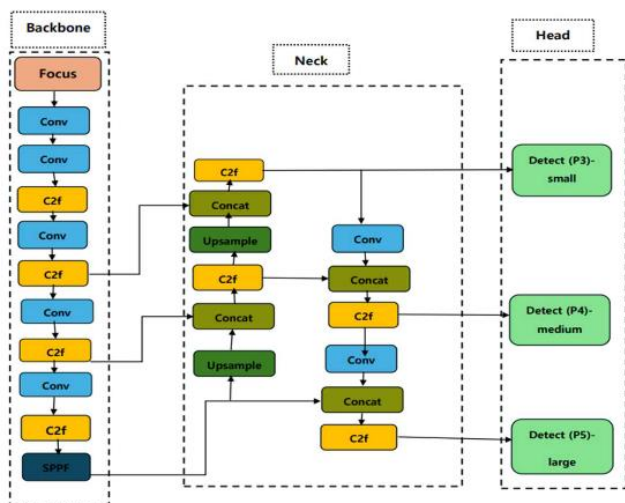


Figure 2: YOLOv8 Architecture

C. Food Segmentation and Portion Size Estimation

Accurate calorie estimation depends heavily on precise portion size measurement. To achieve this, the proposed system incorporates food segmentation techniques to isolate food regions from the background and estimate their area.

Three segmentation approaches are implemented and compared: K-means clustering, SVM-based segmentation, and YOLOv8s segmentation. K-means clustering groups pixels based on color similarity, but its performance degrades under varying lighting conditions. SVM-based segmentation improves classification of food and non-food regions but requires handcrafted feature extraction.

YOLOv8s, a deep learning-based segmentation model, provides instance-level segmentation with higher robustness

and accuracy. It effectively handles complex backgrounds and overlapping food items. Experimental results indicate that YOLOv8s achieves the lowest area estimation error among the evaluated methods, making it the most suitable choice for portion size estimation.

D. Calorie and Nutritional Estimation

Once the food region is segmented, the pixel area is converted into an approximate portion size using reference scaling and empirical density values. The estimated portion size is then mapped to calorie and nutritional information using standard nutritional databases. To improve reliability, the system allows users to manually input food weight when available, which overrides the automated estimate. This hybrid approach balances automation with user input, resulting in improved accuracy in real-world scenarios.

E. Web-Based Application Implementation

To demonstrate the practical applicability of the proposed deep learning-based framework, a web-based application was designed and implemented as the primary user interface. The application enables users to upload food images captured using standard cameras or mobile devices, which are then processed by the trained model for automated food identification and verification.

The platform also allows users to input portion size information through predefined serving options or manual quantity selection to improve nutritional estimation accuracy. Based on the recognized food item and portion details, the system performs nutritional analysis by mapping the results to a structured nutritional database and computes key parameters such as calorie content and macronutrient composition.

The application is designed to be lightweight, responsive, and user-friendly, ensuring efficient performance with minimal computational overhead. A modular architecture separates the frontend from the backend inference engine, enhancing scalability and maintainability, while cross-device compatibility ensures seamless access across both desktop and mobile platforms.

F. Food Image Upload and Preview

The initial stage of user interaction is the food image upload and preview interface. As shown in Fig. 8, users can upload a food image through the web page, which is then displayed for verification. This step ensures that the correct image is selected before initiating food detection and classification. Once confirmed, the uploaded image is forwarded to the backend deep learning models for processing.



Figure 3: File upload and food image preview interface.

I. Food Weight Entry and Verification

Accurate calorie estimation depends on reliable portion size information. To enhance estimation accuracy, the application provides an interface for entering or verifying food weights, as illustrated in Fig. 9. The detected food items are listed with corresponding weight fields, allowing users to either accept the automatically estimated values or manually input actual weights when available. This hybrid approach improves flexibility and reduces errors associated with visual size estimation alone.

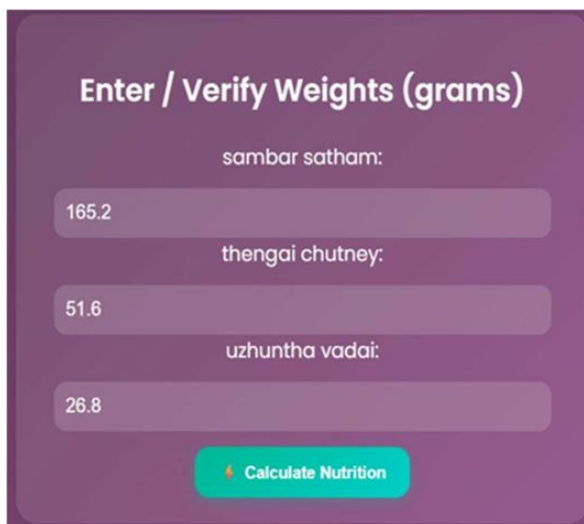


Figure 4: Food weight estimation and user verification interface.

II. Nutritional Estimation and Result Display

After food identification and weight confirmation, the system computes nutritional values using standard nutritional databases. The resulting calorie and macronutrient breakdown is displayed to the user, as shown in Fig. 10. The interface

presents individual calorie values for each food item along with the total calorie count for the meal, enabling users to easily interpret their dietary intake.



Figure 4: Nutrition estimation results using estimated portion size.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

This section presents the experimental setup, evaluation metrics, and performance analysis of the proposed food identification and nutritional profiling system. The experiments were conducted to evaluate the effectiveness of the food classification and segmentation models, as well as to compare their performance with traditional approaches.

IV-A. Experimental Setup

The proposed system was evaluated using Indian food datasets for both classification and segmentation tasks. All experiments were conducted on annotated food images containing multiple Indian dishes captured under varying lighting conditions and backgrounds. The dataset was divided into training, validation, and testing subsets to ensure unbiased performance evaluation. For food classification, two deep learning models—VGG16 and YOLOv8n—were trained and evaluated. VGG16 was implemented using transfer learning, where pretrained ImageNet weights were fine-tuned on the Indian food classification dataset. YOLOv8n was trained for end-to-end food detection and classification using labeled bounding boxes. For food segmentation, three approaches were implemented: K-means clustering, SVM-based segmentation, and YOLOv8s segmentation. These methods were compared to assess their ability to accurately segment food regions and estimate portion size.

IV-B. Evaluation Metrics

The performance of the food classification models was evaluated using standard metrics such as precision, recall, F1 and mean Average Precision (mAP@50). These metrics provide a comprehensive assessment of the model's ability to correctly identify and localize food items.

For food segmentation, performance was evaluated based on area estimation accuracy. The predicted food area obtained

from segmentation was compared with the ground truth area, and the percentage error was calculated. Lower area estimation error indicates better segmentation performance and more reliable portion size estimation.

Fig. 5 represents the Precision–Confidence curve of the trained model. Precision increases steadily with higher confidence thresholds, indicating a reduction in false positive detections. This curve assists in selecting an optimal confidence threshold to ensure reliable food classification performance.

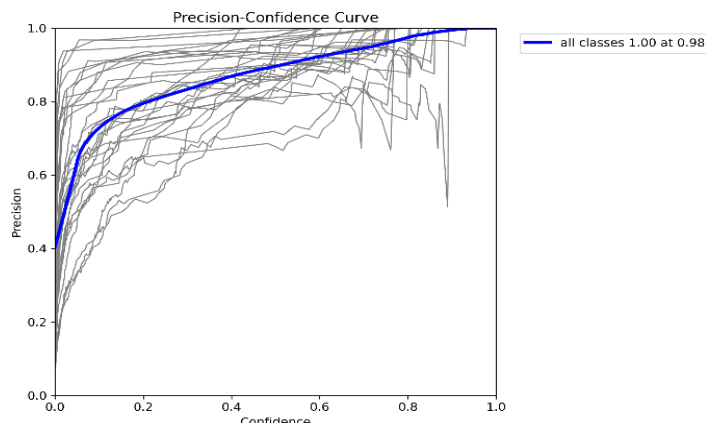


Figure 5: Precision–Confidence Curve

Fig.6 represents the Recall–Confidence curve of the YOLOv8n-based food detection model, showing high recall at lower confidence thresholds and a gradual decrease at higher thresholds due to stricter prediction filtering, highlighting the trade-off between detection sensitivity and confidence reliability.

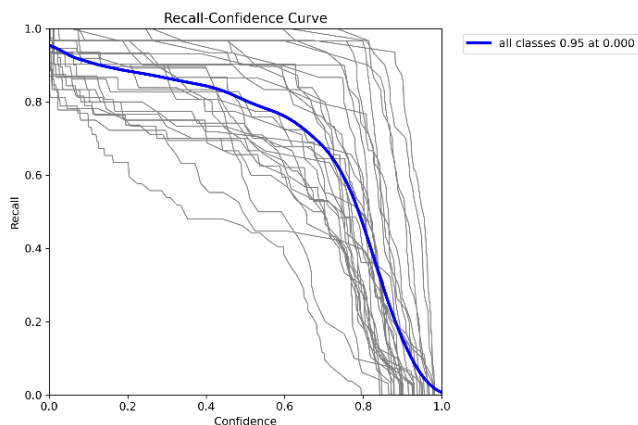


Figure 6: Recall–Confidence Curve

Fig.7 represents the F1–Confidence curve, illustrating the balance between precision and recall. The peak F1 score indicates the optimal confidence threshold at which the model achieves balanced detection accuracy. This operating point is critical for deploying the system in real-world food recognition applications

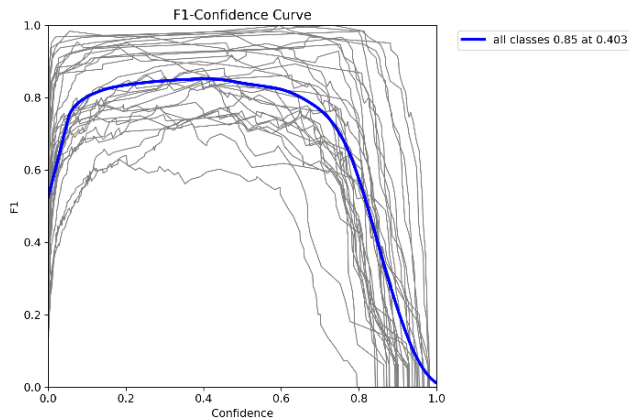


Figure 7: F1–Confidence Curve

Fig.8 represents the Precision–Recall curve for all food classes. The high area under the curve demonstrates that the proposed YOLOv8n model maintains strong precision even at higher recall levels. The achieved mAP@0.5 value confirms the model’s robustness in accurately detecting and classifying food items

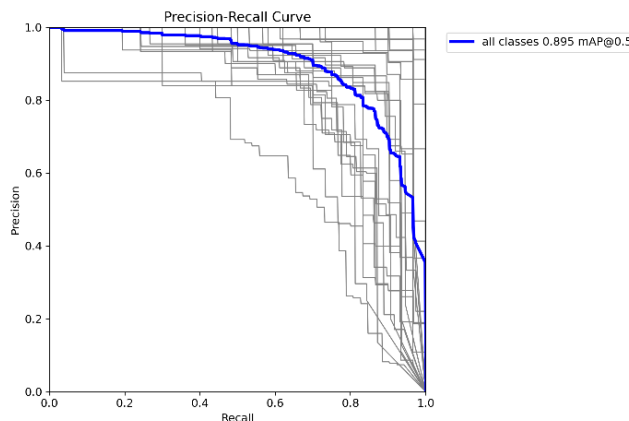


Figure 8: Precision–Recall Curve

IV-C. Food Classification Results

The classification performance of the VGG16 and YOLOv8n models was evaluated on the test dataset to assess their effectiveness in recognizing Indian food items. The VGG16-based model achieved reasonable classification accuracy when applied to single food images; however, its performance degraded in scenarios involving visually similar food items and mixed dishes. Additionally, VGG16 requires independent classification of each image and lacks inherent object localization capability, which limits its applicability in multi-food scenarios. In contrast, the YOLOv8n model demonstrated superior performance by simultaneously detecting and classifying multiple food items within a single image. The YOLOv8n model achieved a mean Average Precision (mAP@50) of 89.6%, significantly outperforming the VGG16-based approach. This improvement can be attributed to YOLOv8’s end-to-end detection framework, which effectively captures both spatial and contextual features. The ability of

YOLOv8n to learn object-level representations and contextual relationships enables more robust classification, particularly in complex scenes containing multiple food items and background variations.

analyzed. The segmentation output generated by the YOLOv8s model demonstrates its ability to accurately isolate food regions while preserving clear object boundaries, which is essential for reliable portion size estimation.



Figure 9: Food classification results

Fig. 9 presents sample food classification results obtained using the proposed YOLOv8n model. The model accurately detects and classifies food items under varying lighting conditions and background complexities. The results indicate strong robustness in distinguishing visually similar food categories, which is a common challenge in Indian cuisine. Overall, the classification outputs demonstrate the effectiveness of YOLOv8n for real-world food recognition applications.

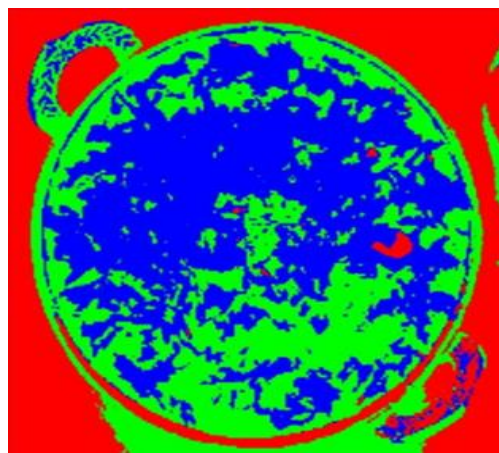


Figure 11: (a) K-means clustering

shown in Fig. 11(a), K-means clustering segments the image based on pixel intensity and color similarity. Although this approach roughly separates the food region from the background, it is highly sensitive to lighting variations and often includes non-food regions, resulting in inaccurate boundaries. This limitation makes K-means unsuitable for precise portion size estimation.

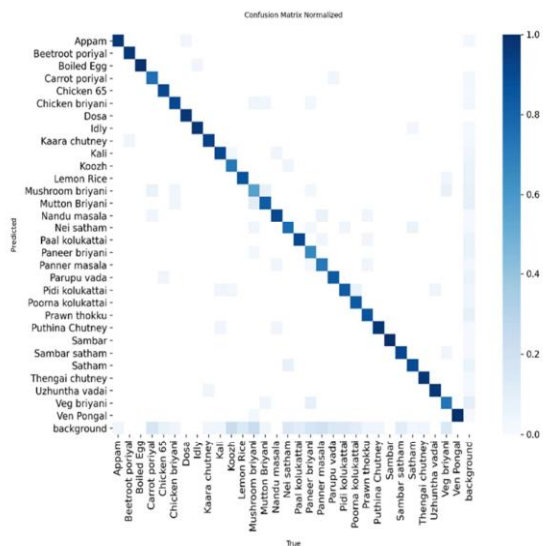


Figure 10: Confusion matrix for food classification

Fig. 10 shows the confusion matrix analysis shows that YOLOv8n accurately classifies most food categories, with minor misclassifications occurring primarily among visually similar dishes.

IV-D. Food Segmentation Results

To further evaluate the effectiveness of the proposed segmentation approach, qualitative segmentation results were



Figure 11(b): SVM-based segmentation

Fig.11(b) illustrates the result of SVM-based segmentation. Compared to K-means clustering, SVM provides improved separation between food and background regions. However, segmentation accuracy degrades in the presence of complex backgrounds and adjacent objects, leading to over-segmentation and inaccurate food area estimation.



Figure 11(c): YOLOv8s segmentation

In contrast, the YOLOv8s-based segmentation result shown in Fig. 11(c) demonstrates accurate localization with clear boundary delineation of the food item. By effectively excluding background regions, the model enables precise area estimation, highlighting the suitability of deep learning-based segmentation for real-world portion size estimation.



Figure 12: Multi-class segmentation outputs using YOLOv8s

IV-E. Comparison of Segmentation Techniques

A quantitative comparison of segmentation methods was conducted by evaluating the predicted food area against the ground truth area. The results are summarized in Table I. Techniques

Table 1: comparison of segmentation techniques

Method	Predicted Area (pixels)	Area Error (%)
YOLOv8s	164,622	6.796
SVM-based Segmentation	198,280	28.632
K-means Clustering	102,098	33.940

YOLOv8s achieved the lowest area estimation error of 6.796%, significantly outperforming the traditional segmentation approaches. This demonstrates the effectiveness of deep learning-based segmentation for accurate portion size estimation.

IV-F. Analysis of Nutritional Estimation Results

The accuracy of nutritional estimation depends on both food recognition and portion size estimation. By combining YOLOv8-based classification and segmentation with optional user-provided weight input, the proposed system achieves reliable calorie estimation. Experimental results from multiple test cases show that allowing users to verify or manually enter food weight reduces estimation errors and improves overall nutritional profiling accuracy.

V. DISCUSSION AND COMMON CHALLENGES

Although the proposed system demonstrates reliable performance in food classification, segmentation, and nutritional estimation, several challenges were observed during experimentation. These limitations are inherent to image-based dietary assessment systems and are discussed to ensure transparency.

challenge, as dishes with comparable color and texture—such as gravies and mixed rice varieties—can occasionally lead to misclassification, particularly under poor lighting conditions. This issue arises due to high intra-class similarity and inter-class variation.

Segmentation errors may occur when food items overlap or when complex backgrounds are present. While YOLOv8s significantly improves segmentation accuracy, minor boundary inaccuracies can still affect portion estimation in such scenarios.

Portion size estimation introduces additional uncertainty when reference scaling or depth information is unavailable, which may lead to deviations in calorie estimation. Allowing users to manually input food weight helps mitigate this limitation and improves accuracy.

Variations in food preparation styles and user-related factors, such as improper image capture angles or partial food visibility, also contribute to estimation errors. Despite these challenges, the proposed system provides a practical and effective solution for automated nutritional profiling. Future improvements will focus on dataset expansion, depth-based estimation, and adaptive nutritional databases.

VI. CONCLUSION AND FUTURE SCOPE

This paper presented an image-based deep learning system for automated food identification and nutritional profiling. By leveraging YOLOv8-based models and a web-based deployment strategy, the system provides an efficient and practical solution for dietary monitoring. Future work will focus on incorporating depth estimation, expanding the dataset to include more regional foods, and enhancing calorie estimation accuracy using 3D volume reconstruction techniques.

REFERENCES

Basic format for books:

- [1] Base paper A. Gupta, A. Das, M. P. Karnik, D. S. Wankhede, U. Mahajan, and T. Patel, "Nutrition Monitoring based on Food Image Classification Using Transfer Learning," IEEE, 2023, doi: 10.1109/ACCESS.2023.10837768.
- [2] L. Jiang, B. Qiu, X. Liu, C. Huang, and K. Lin, "DeepFood: Food image analysis and dietary assessment via deep model," IEEE Access, vol. 8, pp. 47477–47489, 2020, doi: 10.1109/ACCESS.2020.2973625.
- [3] H. Hu, Q. Zhang, and Y. Chen, "NIRSCam: A Mobile Near-Infrared Sensing System for Food Calorie Estimation," IEEE Internet Things J., vol. 9, no.19, pp.18934–18946, Oct.2022, <https://ieeexplore.ieee.org/document/9745595> -
- [4] L. Jiang et al., "Deep Food: Food Image Analysis and Dietary Assessment via Deep Model," IEEE Access, vol. 8, pp. 47477–47491, Mar. 2020, <https://ieeexplore.ieee.org/document/8998172> -
- [5] N. Theera-Ampornpant, P. Treepong, N. Phokabut, and S. Phlaichana, "KediNet: A Hybrid Deep Learning Architecture for Thai Dessert Recognition," IEEE Access, vol. 13, pp. 86935–86948, May 2025, <https://ieeexplore.ieee.org/document/8998172> -
- [6] B. Arslan, S. Memiş, E. B. Sönmez, and O. Z. Batur, "Fine-Grained Food Classification Methods on the UEC FOOD-100 Database," IEEE Trans. Artif. Intell., vol.3, no.2, pp.238–246, Apr.2022, <https://ieeexplore.ieee.org/document/8998172> -
- [7] S. Turmchokkasam and K. Chamnongthai, "The Design and Implementation of an Ingredient-Based Food Calorie Estimation System Using Nutrition Knowledge and Fusion of Brightness and Heat Information," IEEE Access, vol.6, pp.46863–46876, Sept.2018, <https://ieeexplore.ieee.org/document/8359388> -
- [8] S. Espinoza, C. Aguilera, L. Rojas, and P. G. Campos, "Analysis of Fruit Images With Deep Learning: A Systematic Literature Review and Future Directions," IEEE Access, vol. 12, pp. 3837–3852, Jan. 2024. <https://ieeexplore.ieee.org/document/10368014> -