The Bioscan
AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES
20(2): S2: 284-289, 2025 www.thebioscan.com

FOOD CLASSIFICATION AND CALORIE ESTIMATION USING DEEP LEARNING WITH INCEPTIONV3

Ms. Priya A
Assistant Professor
Department of CSE
priyaa@skcet.ac.in

Aswin S
Student - CSE, SKCET
Tamil Nadu, India
aswinsee796@gmail.com

Rohit Kumar S
Student - CSE, SKCET
Tamil Nadu, India
rohith.shiva786@gmail.com

Suman Adithya U
Student - CSE, SKCET
Tamil Nadu, India
sumanadithya1763@gmail.com

Varunadhithya M
Student – CSE, SKCET
Tamil Nadu, India
varunadhithyam@gmail.com

DOI: 10.63001/tbs.2025.v20.i02.S2.pp284-289

KEYWORDS
Food Classification, calorie estimation, Deep Learning, Inceptionv3 Architecture

Received on:

01-03-2025

Accepted on:

07-04-2025

Published on

11-05-2025

ABSTRACT

This paper provides a deep learning-based model for food classification and calorie estimation using the InceptionV3 architecture. The model uses transfer learning, with InceptionV3 serving as a feature extractor and specialized layers for classification, such as global average pooling, dropout, and thick layers. After it has been trained on the Food dataset, which consists of food classifications, it is optimized using the Stochastic Gradient Descent (SGD) optimizer. To enhance model generalization, data augmentation methods such shear range, zoom, and horizontal flips are employed. The system determines the caloric content of the observed food item in addition to classifying foods using a different dataset that associates each food class with its corresponding calorie value. Using the classification of the food item, a regression. The model's performance is monitored using validation loss and accuracy, ensuring that the best-performing model is kept for future projections. The model's decision-making process can be better understood using a range of visualization tools, including class activation maps, heatmaps, and activation layers. This technology, which combines food recognition and calorie estimation, has practical uses in health, fitness, and nutrition.

INTRODUCTION

The need for tools that promote nutrition management and nutritional literacy has been brought to light by the growing number of health issues associated with lifestyle choices. Deep learning has emerged as a powerful solution for several real-world issues, including food recognition and calorie estimation. This study uses the InceptionV3 architecture to develop a trustworthy system for classifying foods and determining their caloric content. Through the use of transfer learning techniques, the model's broad applicability is ensured by training it on the Food-101 dataset, which comprises a range of food classes. The system integrates feature extraction, data preprocessing, and customizable layers to achieve high accuracy and reliability. By combining classification and regression tasks to detect food items and provide calorie information, it meets crucial health monitoring demands. This technique has practical implications for anyone wishing to effectively track their eating habits and illustrates the intersection of artificial intelligence with nutrition and health.

1.1 FOOD CLASSIFICATION

The process of recognizing and classifying food items from pictures or other data inputs is known as food classification. Numerous applications, such as meal monitoring systems, nutritional assessment, and dietary control, depend on this function. With the use of deep learning methods, specifically Convolutional Neural Networks (CNNs), food classification has improved in precision and effectiveness. Systems can accurately identify food products by training models on extensive datasets like Food, which has a variety of food categories. Preprocessing photos, identifying relevant traits, and classifying them into predetermined groups are all steps in the classification process. Applications that focus on health and nutrition greatly benefit from accurate food classification, which serves as the basis for additional features like calorie estimate.

1.2 CALORIE ESTIMATION

The method of calculating a food's calorie content based on its visual characteristics, typically with the use of deep learning models, is known as "calorie estimate." This process must be included in applications intended to regulate food intake and promote healthy eating habits. By combining calorie data with food classification, the computer can accurately estimate the number of calories in a food item after identifying its kind. Many systems utilize a regression model to predict the number of calories, mapping the meal class to a certain range of calories. This work may be difficult because to differences in ingredient types, preparation methods, and portion sizes. But with the help of large, annotated datasets and deep learning architectures like InceptionV3, accurate calorie prediction is achievable, giving consumers useful information to track and control their food intake.

1.3 DEEP LEARNING

"Deep learning" is a subfield of machine learning that focuses on modeling and solving complex problems using multi-layered neural networks, or "deep neural networks." These models

automatically create hierarchical representations of the data by processing input in layers, each of which learns a unique feature or pattern. In deep learning, the model is trained on large amounts of labeled data, and if properly trained, it can generate incredibly accurate classifications or predictions. In fields including speech recognition, image identification, and natural language processing, it has shown great use due to its capacity to process and learn from vast volumes of data as well as to find intricate correlations within that data. Convolutional Neural Networks (CNNs) for image processing are one example of a deep learning model that has revolutionized several industries by offering state-of-the-art performance in tasks that were previously challenging for traditional methods.

1.4 INCEPTIONV3 ARCHITECTURE

InceptionV3 belongs to Google's Inception family of deep convolutional neural network architectures. It makes use of several state-of-the-art techniques to improve computing performance and model correctness. One of the key innovations in InceptionV3 is the use of Inception modules, which apply multiple convolutional filters with different kernel sizes in parallel and then concatenate the results. This allows the model to capture data at different scales while reducing the computational cost. Additionally, InceptionV3 uses techniques like factorized convolutions, global average pooling, and dimensionality reduction to help reduce overfitting and parameter count while maintaining high performance. It is widely used for tasks like object recognition and picture classification because of its ability to extract rich properties from photos and its relatively low computational cost. The architecture's exceptional accuracy on large image datasets like ImageNet has made it attractive for transfer learning applications.

II. LITERATURE REVIEW

According to Joachim Dehais, [1] et al., This system Due to the increasing prevalence of diet-related chronic diseases and the shortcomings of traditional diet management strategies, new tools are required to accurately and automatically assess meals. Proposals for computer vision-based systems that assess the content of food photographs have surfaced recently. For people assessing their meals, estimating food portions is the most difficult activity and the least studied topic. Based on two pictures of a meal taken using a mobile device, the current study proposes a three-step process for calculating portion sizes. The first step is to comprehend how the different views are arranged. The two pictures are then used to create a dense 3D model, and finally, this 3D model is used to extract the volume of the different objects. The system was extensively tested using 77 real dishes with known volumes, and it was able to achieve an average inaccuracy of less than 10% in 5.5 seconds each dish. Due to its computational tractability and lack of human input, the proposed pipeline is a viable option for fully automated nutritional assessment. Patients are usually asked to remember and assess their meals as part of food frequency surveys and self-maintained dietary diaries. Although these methods are widely used, they are not always effective, particularly when applied to children and adolescents who often lack the requisite skills and drive. The main known source of errors is the measurement of food portion sizes, which has an average error of over 20%. Even well-trained diabetic people have difficulties when undergoing intensive insulin therapy. Mobile applications that let consumers measure their food intake have been made possible by high-speed cellular networks and smart mobile devices. These consist of electronic food diaries, barcode scanners for nutritional data, picture logs, and remote dieticians. Meanwhile, automated systems for assessing photos of meals have been made practical by advancements in computer vision.

Food image segmentation is a crucial and dispersible task for creating health-related applications like calculating food calories and nutrients, according to Xiong Wei Wu [2] et al. Food's complex appearance makes it difficult to locate and identify ingredients in food images, as the same ingredient may appear clearly in different food images or overlap in the same image. These two factors contribute to the poor performance of current food image segmentation models: (1) there are few high-quality food image datasets with pixel-wise location masks and fine-grained ingredient labels; the ones that do exist are either small or carry coarse ingredient labels. FoodSeg103 (and its extension FoodSeg154) is a new dataset of 9,490 food pictures that we created for this study. We annotate these images with 154 ingredient classes, and each image contains an average of six ingredient names and pixel-wise masks. Additionally, we present ReLeM, a multi-modality pre-training technique that, in particular, offers a segmentation model with extensive and semantic food knowledge. We test ReLeM on our new datasets and compare it against three popular semantic segmentation methods (based on Vision Transformer, Feature Pyramid, and Dilated Convolution) as baselines. We believe that the FoodSeg103 (and its expansion FoodSeg154) and the pre-trained models using ReLeM can serve as a baseline to support future work on fine-grained food image interpretation. For the purpose of studying food image segmentation, we create the extensive image dataset FoodSeg103 (and its extension FoodSeg154). We use roughly 10,000 photographs and annotate 60k segmentation masks in total, which include 154 components with drastically varied appearances. Additionally, we validate the effectiveness of our multi-modality-based pre-training method ReLeM by integrating three baseline semantic segmentation methodologies and conducting extensive tests on the FoodSeg103, or the normal environment, and the FoodSeg154, or the challenging cross-domain scenario.

In this system, Xiang-Yong Kong [3] et al. have suggested The purpose of the food calorie estimation system (FCES) is to track dietary data for diabetes patients in order to estimate the number of calories they consume. FCESs have recently made use of deep learning technologies. In order to determine the calorie content of food photos, we employ a neural network for pattern recognition. To segment the food photos and extract the area feature, we create a semantic segmentation network model based on Seg Net + Mobile Net, as opposed to the conventional convolutional neural network. It is possible to estimate and determine the number of calories in food by figuring out the corresponding relationship between the meal's calorie value and its area feature. Diabetes has progressively become a well-known and common occurrence in many contemporary nations in recent years due to changes in lifestyle and population ageing. Specifically, the most prevalent kind of diabetes in the twenty-first century is type 2.(1) High blood sugar levels are a hallmark of diabetes, a chronic metabolic condition. Polyuria, polydipsia, polyphagia, and weight loss are common symptoms.(2) Most individuals tend to underestimate the risks of diabetes after learning of raised blood sugar and receiving a diagnosis since the symptoms are not immediately apparent and they neglect diabetes treatment. In this study, we suggested a new method based on calorie conversion and food image recognition. It mostly explains how to create a food image dataset and the precise techniques for identifying images and calculating caloric content from food photos. We developed a deep learning-based algorithm to forecast food image classification and segmentation. As a result, we used picture segmentation to determine the food area. Additionally, we used a variety of dishes to ascertain the corresponding relationship between the food calorie content and the unit area of the food image.

In this system, Sri Kalyan Yarlagadda [4] et al. Nutritionists and researchers have been able to increase the accuracy of dietary evaluation through the development of image-based technologies, which use wearable technology or smartphones to take pictures of the food they eat. The energy and nutritional value of the foods are then estimated by analyzing these photos using computer vision techniques. A key component of this technique is food picture segmentation, which identifies the areas of an image that contain food. Because current methods rely on data, they are not well suited to generalizing to other forms of food. We provide a class-agnostic food image segmentation technique to solve this issue. We employ two images of dining scenes: one before we begin eating and one after we finish. Without any prior knowledge about the food class, we may segment food photos by identifying the salient missing objects using information from both the before and after eating photographs. Using a challenge to identify the salient missing objects in a pair of images, we simulate a top-down saliency paradigm that directs the human visual system's (HVS) attention. The food photos from a dietary study that we

used to validate our strategy yielded encouraging findings. Dietary practices are proven to have a significant influence on an individual's overall health and well-being. Although eating a healthy, nutritious diet is crucial for overall well-being, numerous studies have shown that unhealthy eating patterns can cause a number of illnesses and health issues. For instance, research from the World Health Organization (WHO) has demonstrated that a poor diet is a major modifiable risk factor for the emergence of several noncommunicable diseases, including diabetes, cancer, and heart disease, which are the world's top causes of death.

In this system, Sekhwan Park [5] et al. have suggested Data gathering and labelling for network training are crucial but time-consuming steps in the process of employing deep neural networks to intelligently categories foods in photos for diet control. To address the challenges of data gathering and annotation, this research suggests a real-world food segmentation technique using synthetic data. By arranging several objects on a meal plate and utilizing the open-source 3D graphics program Blender to create synthetic data, we can do food segmentation on healthcare robot systems, like meal assistance robot arms, and train Mask R-CNN for instance segmentation. Additionally, we construct a mechanism for gathering data and validate our segmentation model using actual food data. Consequently, the model trained just on synthetic data can segment food instances that were not trained using 52.2% mask AP all on our real-world dataset, and after fine-tuning, it performs +6.4% better than the model trained from scratch. Furthermore, we validate the potential for the enhancement of performance on the public dataset for impartial analysis. According to some analysts, if the current trend continues, 20 percent of adults worldwide will be fat and 38 percent will be overweight by 2030 [1]. The importance of maintaining a healthy diet and consuming a balanced diet has lately grown due to the rising obesity epidemic. Specifically, systems that use photographs of food to be consumed to automatically compute and record meal types and calories are being developed progressively. The most crucial technology in this service is food recognition, which is applicable to many different types of service robots, such as those that assist with meals, serve food, and cook. Many researchers are putting a lot of effort into creating food-aware datasets and improving food recognition since food-aware activities are becoming more and more important. Food classification, food detection, and food segmentation are the three ways to identify food. Since food classification is comparatively simpler than other tasks during the data collecting and labelling phase, many public datasets are also made available. Food classification is the task of matching the type of food in an image through a single image.

III.EXISTING SYSTEM

In order to address modern health issues like obesity, diabetes, and cardiovascular problems, it has become more crucial than ever to maintain a nutritious diet and a balanced intake of essential nutrients. A precise assessment of meal nutrients is necessary for the proper management of these illnesses. As smartphone applications gain pace, there is a growing interest in developing machine learning-based systems that can accurately and automatically recognize and evaluate the nutritional value of food in real-time using user-captured images. The most recent machine learning methods for food identification, segmentation, and volume estimation—all crucial components of dietary assessment systems—are examined in this work. In order to provide a comprehensive analysis of food detection models, classification techniques, and volume estimation algorithms while connecting their efficacy to the characteristics of various datasets, the study systematically organizes the data that was extracted from the reviewed works. Furthermore, by summarizing the benefits and drawbacks of various strategies, this analysis offers helpful recommendations and insights that could guide future research and development in machine learning-based food detection systems.

IV.PROPOSED SYSTEM

The proposed system aims to provide an intelligent food classification and calorie prediction application through deep learning. Using the InceptionV3 architecture, the system employs transfer learning to recognize food photographs from the Food-dataset, which includes multiple food categories. The design uses special layers including dropout, global average pooling, and thick layers to increase model durability and accuracy. The algorithm goes beyond simple classification by estimating the calorie content of the identified food items using a regression layer. In order to predict the calorie value based on the categorized food class, this layer makes use of an extra dataset that links food categories to the pertinent calorie information. Examples of data augmentation techniques used during training to improve the model's generalization include shear range, zoom, and horizontal flips. The model's performance is evaluated using validation loss and accuracy, ensuring that the best model is selected for deployment. To further illuminate the decision-making process, the system makes use of advanced visualization techniques including class activation maps and activation layers. This integrated system aims to help users understand the type of food they are eating and its calorie content, with practical applications in nutrition management and health monitoring.

A. LOAD DATA

The focus of this course is the Food dataset, which consists of labeled images that depict various food classes. By organizing the dataset, it ensures a clear separation of training and testing data. Additional datasets with calorie information for every meal class are combined in calorie estimation challenges. This module ensures data availability for additional preparation in addition to controlling file organization.

B. DATA PREPROCESSING

In this step, the raw image data is prepared by scaling, normalizing, and transforming it into a format that is compatible with the model input. Examples of data augmentation methods that are used to enhance dataset diversity and model generalization include random shear, zoom, and horizontal flips. This module ensures that the data is clean, consistent, and ready for training.

C.FEATURE EXTRACTION

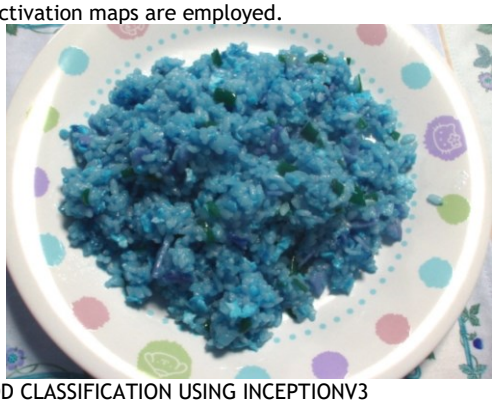
The InceptionV3 model is the pre-trained feature extractor. In this lesson, high-level information is extracted from food photographs using transfer learning. The model's convolutional layers extract rich, hierarchical characteristics to optimize for the specific tasks of food classification and calorie estimation. These features are then processed by specialized layers such as global average pooling, dropout, and dense layers.

D.TRAINING AND TESTING

This module trains the model on the processed dataset using the Stochastic Gradient Descent (SGD) optimizer. In order to prevent overfitting, it separates the data into training and validation sets, employs early stopping techniques, and monitors performance using metrics such as accuracy and validation loss. To determine the trained model's ability to generalize and ensure dependable performance, testing entails evaluating it using unseen data.

E. MODEL EVALUATION

The model undergoes a thorough evaluation to determine its efficacy after training. We calculate metrics like F1-score, recall, accuracy, and precision to assess classification performance. Regression metrics such as Mean Squared Error (MSE) and Mean Absolute Error (MAE) are computed for calorie estimation. To shed light on the model's choices, visualization tools like heatmaps and class activation maps are employed.



F. FOOD CLASSIFICATION USING INCEPTIONV3

This module combines all the parts to classify food images in real time. It determines the food category from an input image using

the optimized InceptionV3 architecture. To provide a comprehensive solution for both categorization and nutrition assessment, the calorie estimation layer additionally forecasts the caloric content linked to the recognized food class.

Data Normalization

X norm = σ/X−μ

Where:

X is the original pixel value,

μ is the mean of the pixel values of the entire dataset,

σ is the standard deviation of the pixel values.

Training (Using SGD Optimizer)

θt+1=θt−η∇θJ(θt)

Where:

- θt\ theta_tθt represents the model parameters at iteration t,
- η \ etaη is the learning rate,
- ∇Θj (θt) \nabla_\theta J(\ theta_tθt)∇θJ(θt) is the gradient of the loss function J(θ)J(\theta) J(θ) with respect to the parameters θ\ thetaθ.

Evaluation Metrics

F1-score

F1=2×Precision+Recall/ Precision × Recall

Precision

Precision=TP+FP/TP

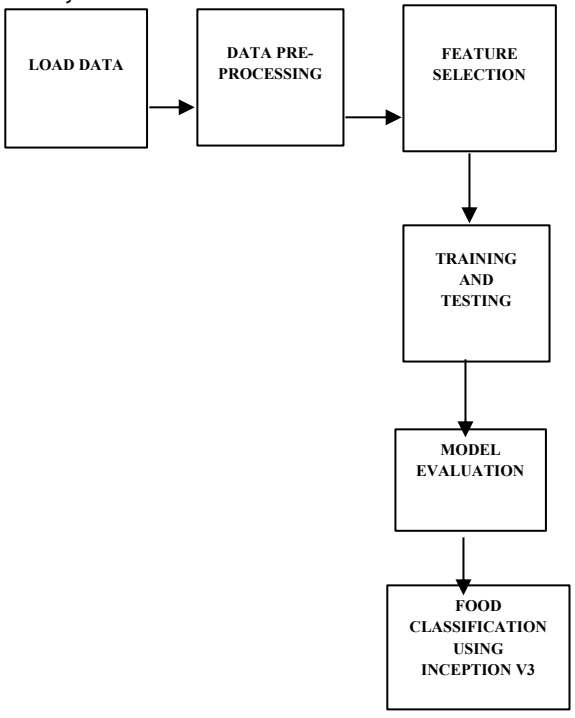
Recall

Recall=TP+FN/TP

Where:

- TPTPTP is the number of true positives,
- FPFPFP is the number of false positives,
- FNFNFN is the number of false negatives.

Accuracy=TP+TN+FP+FN/TP+TN



SYSTEM FLOW DIAGRAM

ALGORITHM DETAILS

This food classification and calorie calculation system uses a deep learning algorithm that makes use of the InceptionV3 architecture. Convolutional neural networks (CNNs) like InceptionV3 are made to process and categories images quickly and accurately. To improve the model's capacity for generalization, the first step in the process is importing and pre-processing the food photos, which entails scaling them to a uniform shape and using data augmentation methods like rotation, flipping, and zooming. The InceptionV3 model is then employed as a feature extractor after having been pre-trained on a sizable dataset such as ImageNet. In order to forecast the food class, a custom classification head that incorporates global average pooling, dropout layers, and thick layers sharpens the model's pre-trained weights.

LOAD dataset from 'Food-dataset' folder

SEPARATE dataset into training and testing sets

LOAD additional calorie information dataset

MERGE calorie dataset with food class data

FOR each image in dataset:

SCALE image to a consistent size

NORMALIZE pixel values between 0 and 1

APPLY data augmentation (flip, zoom, shear) to increase diversity

LOAD pre-trained InceptionV3 model

EXTRACT features from image using InceptionV3's convolutional layers

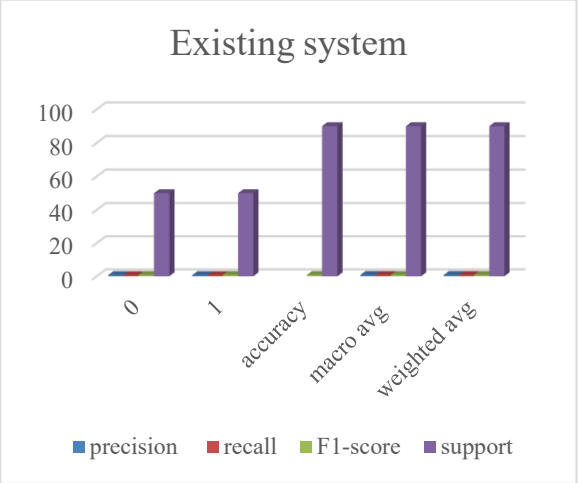
PROCESS extracted features through global average pooling and dropout layers

PASS processed features through dense layers for classification

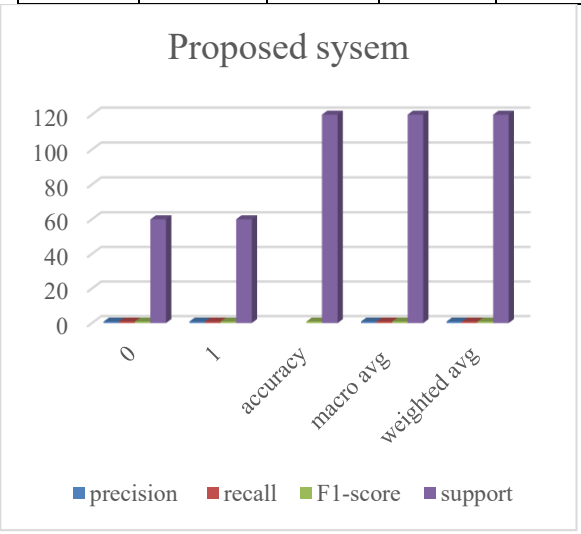
V. RESULT ANALYSIS

The proposed system for food classification and calorie calculation is quite effective in detecting food items and determining their caloric values. The model uses the InceptionV3 architecture with transfer learning to categorize 101 distinct food classes in the Food-101 dataset. The model's generalization abilities are significantly enhanced by including data augmentation techniques like shear, zoom, and horizontal flips, which enable it to perform well on previously unseen data. Throughout the evaluation phase, the model continuously exhibits excellent accuracy and low validation loss, proving its dependability. The calorie estimation module accurately predicts the caloric content of the identified food items by utilizing a regression layer to connect the food class to its appropriate calorie value. The model's accuracy in nutritional analysis is demonstrated by the error metrics, such as Mean Absolute Error (MAE), which show minimal variations in calorie estimates. Examples of visualizations that enhance interpretability include heatmaps and class activation maps, which highlight the regions of the image that influenced the model's decision-making and confirm that it paid attention to relevant details.

Existing	precision	recall	F1-score	support
0	0.7	0.7	0.8	50
1	0.7	0.7	0.8	50
accuracy			0.77	90
macro avg	0.78	0.78	0.78	90
weighted avg	0.78	0.78	0.78	90



Proposed	precision	recall	F1-score	support
0	0.87	0.9	0.89	60
1	0.9	0.87	0.88	60
accuracy			0.88	120
macro avg	0.88	0.88	0.88	120
weighted avg	0.88	0.88	0.88	120



CONCLUSION

Using the InceptionV3 architecture, this work presents a sophisticated system for food classification and calorie estimate that makes use of deep learning techniques to produce precise predictions. The system efficiently detects food items and calculates their caloric content by combining transfer learning, data augmentation, and a regression-based calorie prediction mechanism. The incorporation of sophisticated visualizations improves the model's interpretability and guarantees that its predictions are accurate and comprehensible. The suggested approach offers a workable solution for dietary monitoring and advising, and it shows great promise for use in health and nutrition applications. In order to provide more comprehensive nutritional analysis, future improvements might entail improving the calorie estimation model and adding new food categories to the dataset.

FUTURE WORK

Future iterations of this system will concentrate on improving its accuracy, scalability, and practicality. The model will be more applicable to a wider range of user bases if the dataset is more diverse and includes more food products and regional cuisines. A more comprehensive dietary analysis will be produced by calculating calories as well as macronutrients and micronutrients. Using complex deep learning structures and techniques, including ensemble approaches or attention mechanisms, may help further optimize model performance. Adding real-time food image recognition capabilities to wearable or mobile devices can increase end users' accessibility and convenience. Additionally, working with nutrition experts and integrating user feedback could enhance the system's functionality and ensure that it conforms with actual dietary and health needs.

REFERENCES

- F[1] Konstanta Kopoulos and colleagues, "Using stereo vision approaches for 3D reconstruction and meal volume estimate," in Proceedings of the IEEE 21st International Conference on Bioinf. Bioeng., 2021, pp. 1–4.
- [2] X. Wu and colleagues, "A large-scale benchmark for food image segmentation," in Proceedings of the 29th ACM International Conference on Multimedia, 2021, pp. 506–515
- [3] "Refined image segmentation for calorie estimation of multiple-dish food items," by P. Poply and J. A. A. Jothi, in Proc. Int. Conf. Compute., Communed., Intell. Syst., 2021, pp. 682–687.
- [4] Saliency-aware class-agnostic food image segmentation [S. K. Yarlagadda et al., ACM Trans. Comput. Healthcare, vol. 2, no. 3, pp. 1–17, 2021.
- [5] "Food instance segmentation using deep learning based on synthetic data," in Proc. 18th Int. Conf. Ubiquitous Robots, 2021, pp. 499–505, D. Park et al.
- [6] Neil Houlsby, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Xi aohua Zhai, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, and Alexey Dosovitskiy. 2021. A Picture Is Worth 16x16 Words: Scalable Image Recognition Transformers. In ICLR
- [7] Baining Guo, Stephen Lin, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, and Ze Liu. 2021. Swin Transformer: Shifted Windows-Based Hierarchical Vision Transformer. The preprint arXiv is arXiv:2103.14030 (2103).
- [8] Antonio Torralba, Javier Marín, Yusuf Aytar, Ingmar Weber, Ferda Ofli, Nicholas Hynes, Amaia Salvador, and Aritro Biswas. 2021. Recipe1M+: A Dataset for Cross-Modal Embedding Learning for Food Photos and Cooking Recipes. 2021 TPAMI, 187–203
- [9] Keiji Yanai and Kaimu Okamoto. 2021. UEC-FoodPIX Complete: An extensive dataset for food image segmentation. In MadiMa
- [10] Quin Thames, Tobias Weyand, Shawn Norris, Arjun Karpur, Fangting Xia, Liviu Panait, and Jack Sim. 2021. The goal of Nutrition5k is to automatically understand the nutritional value of generic foods. In C
- [11] Deng-Ping Fan, Kaitao Song, Ding Liang, Tong Lu, Ping Luo, Ling Shao, Enze Xie, Xiang Li, and Wenhai Wang. 2021. A flexible backbone for dense prediction without convolutions is the pyramid vision transformer. 2021's arXiv preprint is arXiv:2102.2122.
- [12] Contributors to MM Segmentation. 2020. MM Segmentation: Benchmark and Open MM Lab Semantic Segmentation Toolbox. mm segmentation: https://github.com/open-mmlab/.
- [13] Philips Kokoh Prasetyo, Yue Liu, Ee-Peng Lim, Lav R Varshney, Helena H. Lee, Ke Shu, and Palakorn Achananuparp. 2020. RecipeGPT is a system that generates and evaluates cooking recipes using generative pre-training. 181–184 in WWW.
- [14] Keiji Yanai and Yuji Matsuda. 2012. Multiple-food recognition using manifold ranking while taking co-occurrence into account. In ICPR, 2017–2020

- [15] Weiqing Min, Xiaolin Wei, Xiaoming Wei, Zhiling Wang, Zhengdong Luo, and Linhu Liu. 2020. ISIA Food-500: A Stacked Global-Local Attention Network Dataset for Large-Scale Food Recognition. ACM International Conference on Multimedia Proceedings, 393–401.
- [16] Chunyan Miao, Steven CH Hoi, Hao Wang, and Guosheng Lin. 2020. Recipe Generation from Images Using a Structure-Aware Generation Network. 359–374 in ECCV.
- [17] Philip HS Torr, Xiatian Zhu, Zekun Luo, Yabiao Wang, Yanwei Fu, Jianfeng Feng, Tao Xiang, Hengshuang Zhao, Sixiao Zheng, Jiachen Lu, et al. 2020. Rethinking Semantic Segmentation with Transformers from a Sequence-to-Sequence Perspective. (2020) arXiv preprint arXiv:2012.15840.
- [18] Zhaoyan Ming, Wanli Zuo, Tat-Seng Chua, Chong-Wah Ngo, Yunan Wang, and Jing-jing Chen. 2019. Multi-label learning for the recognition of mixed dishes. 11th Workshop on Multimedia for Cooking and Eating Activities Proceedings, 1–8
- [19] Adriana Romero, Xavier Giro-i Nieto, Michal Drozdzal, and Amaia Salvador. 2019. Inverse Cooking: Creating Recipes From Pictures of Food. From 10453 to 10462 in CVPR
- [20] Palakorn Achananu parp, Doyen Sahoo, Wang Hao, Shu Ke, Xiongwei Wu, Hung Le, Ee-Peng Lim, and Steven C. H. Hoi. 2019. FoodAI: Deep Learning-Based Food Image Recognition for Intelligent Food Recording. KDD. 2260–2268