International Journal of Applied Food Science and Nutrition

PRIME OPEN ACCESS

Volume 1, Issue 1

Research Article

Date of Submission: 20 February, 2025
Date of Acceptance: 04 June, 2025
Date of Publication: 30 June, 2025

An Automatic Recognition of Indian Dish and Allergen Prediction Using Deep Learning

Viren Kadam*, Anita Thengade, Keval Pambhar, Mihir Agarwal and Tarush Bachal

School of Computer Science and Engineering, MIT World Peace University, India

***Corresponding Author:**
Viren Kadam, School of Computer Science and Engineering, MIT World Peace University, India.

Citation: Kadam, V., Thengade, A., Pambhar, K., Agarwal, M., Bachal, T. (2025). An Automatic Recognition of Indian Dish and Allergen Prediction Using Deep Learning. *Int J Appl Food Sci Nutr, 1*(1), 01-09.

Abstract

The research addresses the issue of image variability in practical settings by introducing a deep learning-based system for allergy prediction and Indian dish recognition. Food plays a big part in supporting a healthy lifestyle, which is becoming more and more popular globally. Recognizing the different types of food and the allergies that are in it is crucial. With the help of convolutional neural networks and visual transformers, the model can predict possible allergies and identify foods with accuracy, which is important for managing diet and raising awareness of allergens. Its uses are extensive in the food service, medical, and nutrition domains, providing workable answers for safer and better-informed nutritional selections in Indian food. Extensive analysis validates the model's effectiveness and potential influence on many sectors. In this study, various CNN architectures, including ResNet50, VGG16, VGG19, and a custom CNN, were employed for Indian Dish Recognition and Allergy Prediction. To enhance model accuracy, techniques such as image augmentation were integrated. While certain models showed promising results, reaching a maximum accuracy of 81%, others did not meet the desired expectations. To improve performance, Vision Transformers (ViTs) were employed, resulting in a remarkable accuracy of 92%.

Keywords: Deep Learning, Allergens, Food Detection, Image Detection, Vision Transformers

Introduction

In India, there is an increased risk of developing food allergies especially due to the complex nature of Indian cuisine which comprises of a wide variety of ingredients and preparation methods. One particular difficulty, though, is the absence of systems designed especially for Indian food. Because of the variety and complexity of Indian cuisine, automated food recognition and allergy prediction are still in their infancy. This makes it extremely difficult to ensure food safety, particularly for people who are allergic to common allergens like dairy, gluten, and nuts.

Despite remarkable progress in the use of Machine Learning (ML) and Deep Learning (DL) algorithms for food recognition and allergen prediction, the majority of existing systems are unable to handle the complexities of Indian cuisine. Current research uses conventional machine learning algorithms like Random Forest (RF), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Decision Trees (DT) and focuses mostly on simpler food categories or more general international cuisines. The accuracy levels of these techniques range from 85 to 95%, however they have trouble with the diverse meal compositions and component variations found in Indian cooking. Furthermore, the datasets utilized in these studies frequently don't have the specificity and diversity required to correctly identify Indian foods.

Recent studies have shown that Deep Learning architectures including Vision Transformers (ViTs), ResNet, VGG, and Convolutional Neural Networks (CNNs) perform better than other models, with accuracy levels ranging from 90 to 99%. Areas like as medical image analysis, object detection, and food image recognition for different cuisines have seen the successful application of these specialized approaches [1-4]. Still, not much research has been done on applying these models to Indian food, which has a variety of preparation techniques, regional variations, and possible allergies. There is a crucial gap in the development of automated systems for this purpose in the lack of research on Indian food recognition and allergy prediction.

This study aims to fill this gap by putting forth a novel deep learning-based framework created especially for allergy prediction and Indian dish recognition. This system, which makes use of CNNs, ViTs, and advanced image processing algorithms, can correctly identify Indian foods and anticipate possible allergens including dairy, gluten, and nuts. In addition to taking into consideration the natural variety in Indian cuisine, the suggested system tackles issues brought on by actual circumstances, like changes in image size and quality. By thoroughly evaluating the system's performance the study demonstrates how the system can offer safer food options and improved allergen control in context of Indian cuisine. The research represents a significant advancement over current food recognition systems that have mostly disregarded the complexities of Indian cuisine because of its customized approach to Indian food recognition and allergen prediction.

Motivation

This research is influenced by the growing number of food allergies and dietary restrictions as well as the complexity of Indian food. Food allergies and intolerances impact a large percentage of the world's population, hence automated systems that can correctly identify dishes and identify any potential allergens within them are desperately needed.

Given the variety of ingredients and intense flavours of Indian food, automated dish recognition and allergy prediction face particular difficulties. Automatically recognizing Indian food is a difficult task due to its complex composition, which is further complicated by regional differences and cooking styles. Even more aggravating these difficulties are the lack of established recipes and the variation in ingredient amounts.

This project aims to overcome these challenges and further automate food recognition systems by using a deep learning-based framework specifically designed for Indian dish recognition and allergen prediction. In addition to providing useful solutions for people with dietary restrictions or intolerances, the suggested framework seeks to improve nutritional management and allergy knowledge in Indian cuisine. The ultimate goal of the study is to enable better allergy control and safer food choices in a variety of ethnic cooking environments.

The key contributions of this study are as follows:

- This study introduces a novel deep learning-based framework specifically designed for Indian dish recognition and allergen prediction using Visual Transformers (ViTs).
- Through rigorous testing and assessment, the framework is thoroughly examined, showcasing its adaptability and practical utility in real-world scenarios with varying picture quality and size. This results in dependable options for safer food choices and improved allergy control in Indian cuisine.

Literature Survey

Related Work

Deepanshu Pandey et al., have demonstrated their work titled "Object Detection in Indian Food Platters using Transfer Learning with YOLOv4" [5]. This study proposes a transfer learning approach with YOLOv4 for object detection in Indian cuisine platters. Making use of the YOLOv4 model's capabilities, the study suggests a transfer learning strategy suited to the complexities of Indian cooking. Improved object detection accuracy is obtained by optimizing pre-trained models on datasets related to Indian food. The work advances automatic food recognition systems in the field of computer vision applied to cultural contexts. The research illustrates the efficacy of the suggested strategy in precisely recognizing different food items within Indian culinary settings through trial and analysis.

E. D. Cherpanath et al., have demonstrated their work titled "Food Image Recognition and Calorie Prediction Using Faster R-CNN and Mask R-CNN" [6]. The goal of the research is to utilize deep learning techniques for calorie prediction and food image recognition. The study discusses health issues like obesity, overeating in food, and related conditions like diabetes and hypertension. The authors provide a strategy for precisely recognizing food items in photos and estimating their calorie content by utilizing Faster R-CNN and Mask R-CNN models. These models allow for more accurate analysis of food images by providing improvements in object detection and classification. The goal of the research is to aid in the creation of technologies that help people control their caloric intake, potentially reducing health concerns listed by agencies such as the World Health Organization.

Seon-Joo Park et al., have demonstrated their work titled "The development of food image detection and recognition model of Korean food for mobile dietary management" [7]. The research aims to enable mobile nutritional control by using a food image detection and identification technology specifically tailored for Korean cuisine. By utilizing machine learning and computer vision techniques, the scientists developed a system that can recognize and classify different Korean foods from photos taken with mobile devices. The image recognition technology in this model allows users to precisely record the amount of food they eat, which improves dietary tracking. In addition to helping people better track their eating habits, the study advances the area of nutrition by providing a useful tool. By focusing on the distinctive qualities of Korean cuisine, the study broadens the application of food identification technology to a variety of cultural settings, which may enhance dietary control and encourage better eating practices.

Research Gap

Despite its potential, the research on object detection in Indian cuisine platters has some limitations. Firstly, the wide range of ingredients and presentation techniques in Indian cuisines makes it difficult to generalize the method to other varieties and regional variants. Second, the suggested model's performance and scalability can be constrained by the quantity and caliber of annotated datasets for Indian cuisine. Furthermore, the article might not include a comprehensive comparative analysis with other approaches or standards, which would make it more difficult to evaluate the paper's efficacy in comparison to current solutions. Finally, it is possible to forget about the computational resources needed for inference and training [5].

The study has some shortcomings even if it offers encouraging developments in the use of deep learning for food image analysis. First of all, environmental elements like lighting and background clutter in images may have an effect on how effective the suggested method is, as well as the variety and variability of food items. Furthermore, the system's ability to adapt to new datasets or dietary preferences may be limited by its dependency on pre-trained models such as Mask R-CNN and Faster R-CNN. Furthermore, variables like portion size estimation and cooking process differences may have an impact on the accuracy of the calorie projection. Improving the suggested approach's adaptability and effectiveness in real-world circumstances would require addressing these constraints [6].

Although the research offers improvements in the identification and detection of food images for Korean cuisine, there are a few drawbacks to take into account. The model's accuracy in real-world applications may be impacted by variables including lighting, camera quality, and food presentation methods, which can differ in the model's performance. Additionally, the wide variety of Korean cuisine may not be adequately captured in the training and evaluation dataset, which could introduce biases or inaccuracies into the recognition results. To improve the model's practical application and dependability, future study might incorporate a more diverse dataset, optimize the model's robustness to environmental influences, and investigate its usefulness in other cultural cooking situations [7].

Material and Methods

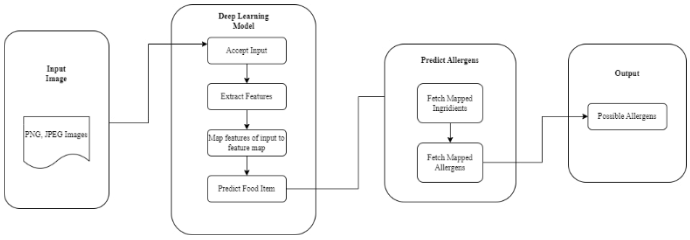

Figure 1: Process Flow of the Proposed System

In the first step food photos are fed into the allergy detection system (Figure 1). To make sure that the deep learning models used for allergy detection are compatible with these photos, pre-processing is applied, which includes scaling and normalization.

Dataset

In this paper, we used the publicly available "Indian food classification" dataset [8]. It consists of different images of food in 20 different classes with about 300 images per class. Some of the classes are of Indian food. All the images are extracted from google. The dataset contains: burger, butter_naan, chai, chapati, chole_bhature, dal_makhani, dhokla, fried_rice, idli, jalebi, kaathi_rolls, kadai_paneer, kulfi, masala_dosa, momos, paani_puri, pakode, pav_bhaji, pizza, samosa.

The "Indian food classification" dataset is a number of images intended for use in image classification applications, specifically related to food identification. For researchers and practitioners looking to create and evaluate algorithms in the fields of computer vision and food recognition, this dataset is an invaluable resource.

Food Class	Number of images	Description
burger	309	Images of different types of burgers.
butter_naan	329	Traditional Indian bread with butter.
chai	381	Images of Indian tea in various cups.
chapati	413	Flatbread commonly eaten in India.
chole_bhature	411	Chickpea curry with fried bread.
dal_makhani	321	Lentil-based curry popular in India.

dhokla	289	Fermented steamed cakes made from rice and chickpeas.
fried_rice	355	Various preparations of fried rice.
idli	310	Steamed rice cakes, a staple in South India.
jalebi	297	Sweet, deep-fried dessert soaked in syrup.
kaathi_rolls	293	Indian flatbread rolls with various fillings.
kadai_paneer	412	Spicy curry made with paneer (Indian cottage cheese).
kulfi	237	Traditional Indian ice cream.
masala_dosa	311	Rice crepes filled with spicy potato filling.
momos	319	Tibetan-style dumplings.
paani_puri	130	Hollow, crispy puris filled with spicy water.
pakode	278	Various types of deep-fried snacks.
pav_bhaji	353	Spicy vegetable mash served with bread rolls.
pizza	261	Various types of pizza.
samosa	262	Deep-fried pastries with savory filling.

Table 1: Summary of Indian Food Classification Dataset

Table 1 provides a clear overview of the dataset used in the research enhancing the understanding and diversity of the scope of the data.

Model Structure of Visual Transformers (ViTs)

ViTs known as Vision Transformers, are a class of deep learning models created especially for computer vision tasks; they are an important shift from the conventional convolutional neural networks (CNNs), which have long been the standard architecture for image classification tasks [9].

ViTs indicate a paradigm shift in computer vision by modifying the Transformer architecture, which was first developed for natural language processing, to visual tasks.

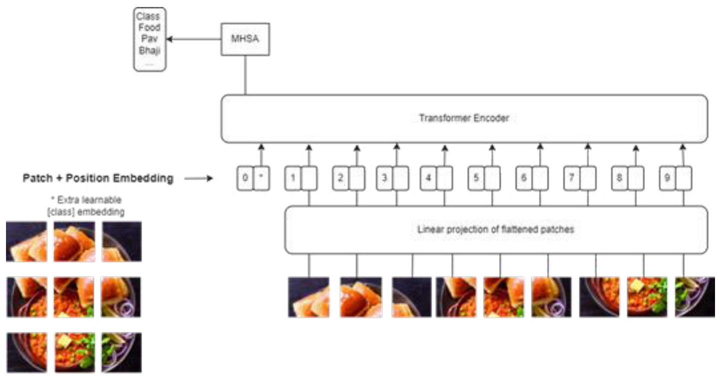

Figure 2: Architecture of ViT

Conventional CNNs work in a hierarchical fashion, beginning with local features and progressively combining them to create a global understanding. As opposed to this, ViTs process images in a comprehensive manner by first splitting them into fixed-size patches, which are subsequently flattened and fed into a Transformer encoder. This enables ViTs to better understand physical connections inside images by using self-attention mechanisms to collect both local and global dependencies simultaneously. ViTs are exceptionally skilled at tasks like image classification because they are able to capture long-range dependencies and context by utilizing self-attention. ViTs also have the benefit of being scalable to higher image resolutions without appreciably raising computing complexity. ViTs are trained to predict image labels using supervised learning on labeled data, showcasing their adaptability and effectiveness in a range of visual identification tasks.

The primary transformer architecture consists of a decoder and an encoder, which have similar structures. Since we are focused on the classification of a food dish, only the encoder portion of the transformer is required. The basic structure of a vision Transformer comprises four blocks; Patch-generation, Patch-Embedding, Multi-Head Self-Attention (MHSA) and Classification block, shown in Figure2.

Patch Generation: A H×H sized input image is divided up into smaller square patches. A matrix with dimension N×h×h, where N is the number of patches and h×h is the resolution of each patch, is created from the image. The formula for the number of patches N is H2/h2.

Patch Embedding: N vectors are created by flattening the obtained patches. Then a learnable D-dimensional projection operator is applied to project these vectors into D dimensions. By adding a Positional Embedding to every vector, positional information is kept intact. After that, an output tensor is acquired and sent to a transformer block.

Multi-Head Self-Attention (MHSA): The MHSA layer receives the output tensor from the Patch-Embedding layer after it has been normalized. After the projected patches are normalized, they are run through a number of self-attention brains. Each head rates the patches according to how important they are in relation to one another. A representation of patches with a weighted average is the end product.

Classification Block: The attention-enhanced patches pass through a Transformer layer and a Feed Forward Neural Network as needed, ultimately arriving at a classification block that uses the learnt characteristics to produce the final prediction.

Allergen Mapping System
The system uses an allergen mapping component after the deep learning models have predicted the food item (Figure 1). To make identification easier for users with food allergies, this component links anticipated allergens to particular ingredients that are present in the predicted food item in step 2.

Output Generation
Ultimately, the system produces output with predictions for allergens. Enabling users to utilize their allergy profiles to make educated food consumption decisions.

Technologies Used
Deep Learning Frameworks:
TensorFlow provides both lower-level APIs for complex customisation and higher-level APIs like Keras for customizing and optimizing deep neural networks, including CNNs and ViTs. Another well-liked open-source deep learning framework is PyTorch, which is perfect for CNN and ViT prototype and experimentation because of its flexible design and dynamic computational graphs.

Other Classification Algorithms
For comparison we are using transfer learning by loading ImageNet weights to every model, freezing the internal layers, and swapping out the top layers for specially created layers intended for food dish classification. VGG19, EfficientNet, and ResNet are the basic models that were employed in this comparison.

VGG (Visual Geometry Group): It is the basic CNN architecture for its efficiency and simplicity. It is composed of several convolutional layers with modest convolutional filters, followed by fully linked layers.

ResNet (Residual Network): ResNet is a deep CNN design that popularized the idea of residual connections. By reducing the vanishing gradient issue, it makes training extremely deep networks possible.

EfficientNet: The EfficientNet family of CNN designs combines cutting-edge functionality with computational resource savings. In order to balance network depth, width, and resolution, it employs the compound scaling technique.

The mentioned networks underwent 50 epochs of training with an early stopping condition. The learning rate was set at 1 e-4, which will be reduced by a factor of 0.1. Softmax is utilized as an activation function in multi-label classification tasks, as it produces outputs that are mutually exclusive.

Result
We assess models in this research using most common evaluation metrics such as accuracy and Confusion Matrix. The proposed Vision transformer gives a test accuracy of 92%.

Models	Accuracy
ResNet50	34%
ResNet50 with regularization	40%
VGG16	56%
VGG19	65%
Custom CNN	81%
ViT	92%

Table 2: Accuracy of Different Models

Table 2 presents the accuracy scores of the various models used in the study. Different CNN models showed promising results reaching a maximum accuracy of 81% while others did not meet desired expectations. Utilization of Vision Transformers (ViTs) resulted in a remarkable accuracy of 92%.

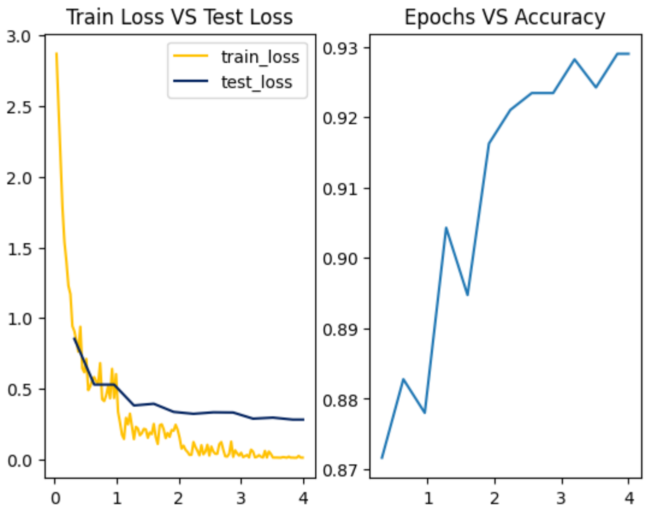

Figure 3: ViT (Final Selected Model) Training Results

The figure 3 demonstrates the graph of Train Loss VS Test Loss and Epochs Vs Accuracy.
Train Vs Test Loss: This graph plots both the training loss and test loss on the same y-axis against epochs on the x-axis. In the graph, A significant decrease in both training loss (from 2.9 to 0.2) and test loss (from 0.9 to 0.4) within just 4 epochs indicates the model is efficiently learning to fit the data.

Epochs Vs Accuracy: This is a separate graph with epochs on the x-axis and accuracy (usually training accuracy) on the y-axis. In the graph, an increase in training accuracy from 87% to 93% in 4 epochs shows the model is getting better at classifying the Indian food images in the training data.

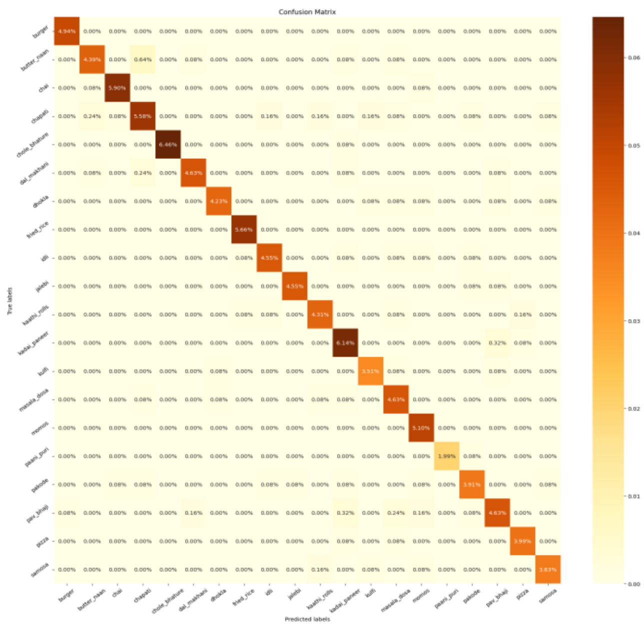

Figure 4: Confusion Matrix

The confusion matrix helps understand the performance of a classification model.It provides a breakdown of how many predictions were correct and incorrect for each class in the dataset. Misclassifications are obvious between similar dishes such as butter naan and chapati. Color intensity indicates the proportion of prediction. Darker colors represent higher values helping to quickly identify which classes are being correctly or incorrectly predicted.

Indian Food Classes	Precision	Recall	F1-score
burger	0.98	1.0	0.99
butter_naan	0.92	0.83	0.87
chai	0.97	0.97	0.97
chole_bhature	1.0	0.99	0.99
dal_makhani	0.95	0.91	0.93
dhokla	0.96	0.91	0.94
fried_rice	0.97	1.0	0.99
idli	0.94	0.92	0.93
jalebi	0.98	0.96	0.97
kadhai_paneer	0.87	0.94	0.91

Table 3: Performance of Indian Food Classes

Table 3: represents the precision, recall and f1-score of different classes in the dataset.

System	Accuracy
D.Pandey[1]	91%
S.Joo[3]	88%
K.Srigurulekha[12]	86.85%
S.Mezgec[4]	86.72%
Proposed Model	**92%**

Table 4: Comparison with State of Arts

Table 4 represents the comparison of the proposed model with other state of arts. Our research investigates food image recognition using a comprehensive dataset of approximately 20 food items, achieving an accuracy of around 92%. In contrast, our broader dataset and focus on a wider variety of food items strengthens the generalizability and robustness of our findings.

Figure 5: Food-Item and Allergen Prediction Using Image

In Figure 5 the proposed model is correctly predicting the food item as kadai paneer and its allergens are listed.

Conclusion

In conclusion, the creation and application of the allergy detection system mark a substantial development in computer vision and medical technology. By utilizing both Convolutional Neural Networks (CNNs) and Visual Transformers (ViTs)—with the proposed vision transformer being used for the first time in a food recognition system—the system shows impressive potential for properly recognizing allergens in food photos. Utilizing cutting-edge models and technology, the system offers consumers insightful information to help them make educated dietary choices and reduce the health risks related to food allergies [10-17].

Furthermore, the study presented in this paper emphasizes how crucial it is to use machine learning and image recognition techniques to address practical issues with allergen management and food safety. With its four steps—input processing, allergen mapping, deep learning model selection, and output generation—the suggested methodology provides a solid foundation for creating allergy detection systems that have real-world uses in the food business, consumer wellness, and healthcare.

As advancements in deep learning continue to accelerate, future research efforts should focus on enhancing the scalability, accuracy, and interpretability of allergy detection systems. Additionally, there is a need for collaborative initiatives between researchers, healthcare professionals, and industry stakeholders to ensure the widespread adoption and impact of such systems in improving public health outcomes and enhancing quality of life for individuals with food allergies.

In conclusion, the allergy detection system presented in this research paper holds great promise for revolutionizing allergen management practices and empowering individuals to make informed dietary choices, thereby contributing to a safer and healthier future for all.

Conflict of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data used in this study are from kaggle repository.
https://www.kaggle.com/datasets/l33tc0d3r/indian-food-classification?resource=download

Funding Statement

No funding was received for conducting this study.

References

1. Thengade, A., & Rajurkar, A. M. (2020). Segmentation of Knee Bone Using MRI. In Applied Computer Vision and Image Processing: Proceedings of ICCET 2020, Volume 1 (pp. 237-246). Springer Singapore.
2. Desai, V., & Thengade, A. (2022, July). Chest Abnormality Detection from X-Rays Using Deep Learning. In 2022 Second International Conference on Next Generation Intelligent Systems (ICNGIS) (pp. 1-6). IEEE.
3. Bhalekar, M., & Bedekar, M. (2022). D-CNN: a new model for generating image captions with text extraction using deep learning for visually challenged individuals. Engineering, Technology & Applied Science Research, 12(2), 8366-8373.
4. Sharma, J., Aher, J., Kalsariya, T., Nyaykare, K., Darwade, A., Makwana, F., (2024). 'SafeBite Fruit Scanner: A New Era in Fruit Quality and Safety Analysis', ISSN 2349-6002, UGC Journal International Journal of Innovative Research in Technology (IJIRT), 10(12), 1206-1214.
5. Pandey, D., Parmar, P., Toshniwal, G., Goel, M., Agrawal, V., Dhiman, S., ... & Bagler, G. (2022, May). Object detection in indian food platters using transfer learning with yolov4. In 2022 IEEE 38th International conference on data engineering workshops (ICDEW) (pp. 101-106). IEEE.
6. Cherpanath, E. D., Nasreen, P. F., Pradeep, K., Menon, M., & Jayanthi, V. S. (2023, August). Food image recognition and calorie prediction using Faster R-CNN and Mask R-CNN. In 2023 9th International Conference on Smart Computing and Communications (ICSCC) (pp. 83-89). IEEE.
7. Park, S. J., Palvanov, A., Lee, C. H., Jeong, N., Cho, Y. I., & Lee, H. J. (2019). The development of food image detection and recognition model of Korean food for mobile dietary management. Nutrition research and practice, 13(6), 521-528.
8. Sheng, G., Min, W., Zhu, X., Xu, L., Sun, Q., Yang, Y., ... & Jiang, S. (2024). A lightweight hybrid model with location-preserving vit for efficient food recognition. Nutrients, 16(2), 200.
9. Srigurulekha, K., & Ramachandran, V. (2020, January). Food image recognition using CNN. In 2020 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-7). IEEE.
10. Mezgec, S., & Seljak, B. K. (2019, December). Using deep learning for food and beverage image recognition. In 2019 IEEE International Conference on Big Data (Big Data) (pp. 5149-5151). IEEE.
11. Tian, Y. (2020). Artificial intelligence image recognition method based on convolutional neural network algorithm. Ieee Access, 8, 125731-125744.
12. Sultana, F., Sufian, A., & Dutta, P. (2018, November). Advancements in image classification using convolutional neural network. In 2018 Fourth international conference on research in computational intelligence and communication networks (ICRCICN) (pp. 122-129). IEEE.
13. Fakhrou, A., Kunhoth, J., & Al Maadeed, S. (2021). Smartphone-based food recognition system using multiple deep CNN models. Multimedia Tools and Applications, 80(21), 33011-33032.
14. Xiao, L., Lan, T., Xu, D., Gao, W., & Li, C. (2021). A simplified CNNs visual perception learning network algorithm for foods recognition. Computers & Electrical Engineering, 92, 107152.

15. Tarannum, S., Jalal, M. S., & Huda, M. N. (2024). HALALCheck: a multi-faceted approach for intelligent halal packaged food recognition and analysis. IEEE Access.
16. https://www.kaggle.com/datasets/l33tc0d3r/indian-food-classification
17. Rohini, B., Pavuluri, D. M., Kumar, L. N., Soorya, V., & Aravinth, J. (2021, March). A framework to identify allergen and nutrient content in fruits and packaged food using deep learning and ocr. In 2021 7th International conference on advanced computing and communication systems (ICACCS) (Vol. 1, pp. 72-77). IEEE.