DeepFood: Food Image Analysis and Dietary Assessment Using VGG16-Based Faster R-CNN

KOLLIPARA KYATHILI PRIYA

PG Scholar. Department of MCA, DNR College, Bhimavaram, Andhra Pradesh

K. Rambabu

(Assistant Professor), Master of Computer Applications, DNR College, Bhimavaram, Andhra Pradesh

ABSTRACT

With the increasing awareness of health and nutrition, dietary monitoring has become an essential aspect of modern lifestyles. However, manual tracking of food intake is often tedious and prone to inaccuracies. Recent advancements in computer vision and deep learning have enabled automated food recognition systems that can assist users in analyzing their dietary habits. This project presents a deep learning-based system, "DeepFood," which performs food image classification and dietary assessment using a VGG16-based Faster R-CNN architecture.The proposed system utilizes the UECFOOD100 dataset, which contains a wide variety of food images along with bounding box annotations. The system is designed to detect food items in images and videos, classify them into predefined categories, and estimate basic nutritional information such as calories, fat, carbohydrates, and protein.The model integrates the VGG16 convolutional neural network as a backbone for feature extraction. The extracted features are then passed through two parallel heads: one for classification and another for bounding box regression. This dual-output structure enables the system to both identify the type of food and locate it within the image. The model is trained using a multi-task learning approach, optimizing both classification and localization simultaneously.During preprocessing, images are resized and normalized, and bounding box coordinates are scaled appropriately. The dataset is divided into training and testing sets to evaluate model performance. The system achieves effective classification accuracy and demonstrates the ability to detect food regions in real time.For user interaction, a graphical user interface (GUI) is implemented using Tkinter. It allows users to upload datasets, train the model, visualize performance graphs, and test the system on images or videos. The system also provides estimated nutritional information based on the detected food category.The proposed solution offers a practical and efficient approach for automated dietary assessment. It can be used in healthcare applications, fitness tracking systems, and smart kitchens. Future improvements may include integrating real nutritional databases, expanding food categories, and using advanced architectures such as EfficientNet or YOLO for improved accuracy and speed.

Keywords: Food Image Classification, Dietary Assessment, Faster R-CNN, VGG16, Computer Vision, Deep Learning, Nutrition Analysis, Object Detection, Image Processing, Health Monitoring

I. INTRODUCTION

The increasing prevalence of lifestyle-related diseases such as obesity, diabetes, and cardiovascular disorders has highlighted the importance of maintaining a balanced diet. Monitoring daily food intake is a critical step toward achieving better health outcomes. However, traditional methods of dietary tracking, such as manual logging or calorie counting, are often time-consuming and inaccurate.With the advancement of artificial intelligence, particularly in the field of computer vision, automated food recognition systems have emerged as a promising solution. These systems aim to identify food items from images and provide nutritional information without requiring manual input from users.Deep learning techniques, especially Convolutional Neural Networks (CNNs), have shown remarkable success in image classification tasks. Object detection models such as Faster R-CNN further enhance these capabilities by identifying both the location and category of objects within an image. This makes them highly suitable for food recognition applications, where multiple food items may be present in a single image.This project introduces “DeepFood,” a system that combines VGG16 and Faster R-CNN for food detection and classification. The system processes images and videos to identify food items and provide dietary insights. The use of a pre-trained VGG16 model allows efficient feature extraction, while the Faster R-CNN framework enables accurate object localization.The system also includes a graphical user interface (GUI) developed using Tkinter, making it accessible to non-technical users. Users can upload datasets, train the model, and test it on real-world images and videos.The integration of deep learning with dietary assessment has the potential to revolutionize healthcare and fitness applications. It can assist individuals in tracking their food intake, help dieticians provide better recommendations, and enable smart systems to monitor nutrition automatically.

II. LITERATURE SURVEY (WITH EXISTING METHODS)

Food recognition and dietary assessment have gained significant attention in recent years due to their applications in health monitoring and lifestyle management. Various approaches have been proposed, ranging from traditional image processing techniques to advanced deep learning models.Early methods relied on handcrafted features such as color histograms, texture analysis, and shape descriptors to identify food items. While these methods were computationally efficient, they lacked robustness and accuracy, especially when dealing with complex food images.The introduction of deep learning revolutionized this field. Convolutional Neural Networks (CNNs) became the standard approach for food image classification due to their ability to automatically learn hierarchical features. Models such as AlexNet, VGGNet, and ResNet have been widely used for this purpose.Object detection models such as Faster R-CNN, YOLO, and SSD further improved performance by enabling localization of food items within images. Faster R-CNN, in particular, is known for its high accuracy, as it uses region proposal networks to identify potential object locations.Recent research has also explored multi-task learning approaches, where models simultaneously perform classification and localization. This approach improves efficiency and reduces computational overhead.

Datasets such as UECFOOD100 and Food-101 have played a crucial role in advancing research in this domain. These datasets provide annotated images that help train and evaluate deep learning models.Despite these advancements, challenges remain. Variations in lighting, occlusion, and food presentation can affect model performance. Additionally, estimating accurate nutritional information from images is still a complex problem.The proposed system addresses these challenges by using a robust deep learning architecture and a well-annotated dataset. It combines classification and localization to improve accuracy and provides a practical solution for real-world applications.

III. EXISTING SYSTEM

Existing food recognition systems primarily rely on image classification techniques that identify food items without considering their location in the image. While these systems are effective for simple scenarios, they struggle when multiple food items are present.Some systems use traditional machine learning algorithms with handcrafted features, which often result in lower accuracy and poor generalization. These methods are sensitive to variations in lighting, orientation, and background.More advanced systems utilize deep learning models such as CNNs for food classification. However, many of these systems focus only on classification and do not provide localization of food items. This limits their applicability in real-world scenarios where multiple objects need to be detected.Object detection models like YOLO and SSD offer faster processing but may compromise accuracy compared to region-based methods like Faster R-CNN. Additionally, many existing systems do not provide dietary analysis, focusing solely on recognition.Another limitation is the lack of user-friendly interfaces. Many systems are developed for research purposes and are not easily accessible to end users.The proposed system overcomes these limitations by integrating Faster R-CNN with VGG16 for accurate food detection and classification. It also includes a GUI for ease of use and provides basic dietary information, making it a comprehensive solution for food analysis and nutrition assessment.

IV. PROPOSED METHOD

The proposed system, **DeepFood**, is an intelligent food recognition and dietary assessment framework that leverages deep learning and computer vision techniques to automate food analysis. The system is designed to detect, classify, and localize food items from images and videos while simultaneously providing estimated nutritional information.The architecture is based on a **VGG16-powered Faster R-CNN model**, which integrates feature extraction, object detection, and classification into a unified framework. Faster R-CNN is chosen due to its high detection accuracy and ability to handle complex visual scenes with multiple objects. It utilizes a Region Proposal Network (RPN) to identify potential food regions, followed by classification and bounding box regression.The system processes input data from the **UECFOOD100 dataset**, which includes annotated food images. During preprocessing, images are resized, normalized, and augmented to improve generalization. Bounding box annotations are scaled appropriately to match model requirements.

The model employs **multi-task learning**, where classification and localization tasks are optimized simultaneously. This approach improves efficiency and reduces computational redundancy. The trained model predicts both the class label and the spatial coordinates of detected food items.Additionally, the system includes a **Graphical User Interface (GUI)** developed using Tkinter. The GUI allows users to upload datasets, train the model, visualize accuracy graphs, and test the system on images and videos.The dietary assessment module provides approximate nutritional values such as calories, fat, carbohydrates, and protein. While current values are randomly generated, the system is designed to integrate real nutritional databases in future enhancements.Recent studies show that deep learning-based food recognition systems significantly outperform traditional approaches due to their ability to learn complex visual patterns . The proposed system aims to provide an efficient, scalable, and user-friendly solution for automated dietary monitoring.

V. IMPLEMENTATION

The implementation of the DeepFood system involves multiple stages, including data preprocessing, model training, evaluation, and deployment through a graphical interface.

1. Data Collection and Preprocessing

The system uses the **UECFOOD100 dataset**, which contains labeled food images and corresponding bounding box annotations. The preprocessing phase includes:

- Reading image data using OpenCV
- Resizing images to a fixed size (80×80 pixels)
- Normalizing pixel values between 0 and 1
- Converting labels into categorical format
- Extracting bounding box coordinates and scaling them relative to image dimensions

The dataset is shuffled and split into training and testing sets using an 80:20 ratio.

2. Model Development

The core model is built using **VGG16 as a backbone network** for feature extraction. VGG16 is pre-trained on ImageNet and used with frozen layers to leverage transfer learning.

Two output branches are added:

- **Classification Head**: Uses fully connected layers with softmax activation
- **Bounding Box Head**: Uses dense layers with sigmoid activation to predict coordinates

The model is compiled using:

- Loss functions: categorical cross-entropy (classification) and mean squared error (localization)
- Optimizer: Adam with a learning rate of 0.0001

3. Model Training

The model is trained for 30 epochs with a batch size of 32. Multi-task learning ensures simultaneous optimization of classification and localization tasks. Training history (accuracy and loss) is saved for visualization.Faster R-CNN-based architectures are widely used due to their high precision in object detection tasks, despite higher computational cost .

4. Model Evaluation

Performance is evaluated using:

- Classification accuracy
- Loss curves (visualized using Matplotlib)
- Bounding box prediction quality

5. Deployment with GUI

A Tkinter-based GUI is implemented with the following features:

- Dataset upload
- Data preprocessing
- Model training
- Accuracy graph visualization
- Image and video-based classification

Users can upload images or videos, and the system displays predicted food labels along with bounding boxes.

6. Dietary Assessment Module

The system generates approximate nutritional values (calories, fats, carbohydrates, protein). Future improvements will integrate real-world nutrition databases.

VI. ALGORITHMS

1. Faster R-CNN Algorithm

Faster R-CNN is a two-stage object detection algorithm:

1. **Feature Extraction**: Input image is passed through VGG16 CNN
2. **Region Proposal Network (RPN)**: Generates candidate object regions
3. **ROI Pooling**: Extracts fixed-size feature maps
4. **Classification and Regression**: Predicts class labels and bounding boxes

This architecture ensures high detection accuracy and robustness.

2. VGG16 Feature Extraction Algorithm

- Input image is passed through convolutional layers
- Feature maps are extracted using ReLU activation
- Max-pooling reduces spatial dimensions
- Final feature maps are flattened for classification

3. Multi-task Learning Algorithm

The model simultaneously optimizes:

- Classification loss (categorical cross-entropy)
- Localization loss (mean squared error)

Total Loss = Classification Loss + Bounding Box Loss

This improves performance and reduces training time.

4. Prediction Algorithm

1. Load trained model
2. Preprocess input image
3. Perform forward pass
4. Extract predicted bounding box and class
5. Display results with bounding box

Recent advancements in object detection show a shift toward hybrid and transformer-based models, but CNN-based methods like Faster R-CNN remain highly effective .

VII. SYSTEM DESIGN

The system design of DeepFood follows a modular architecture consisting of four main components: input module, processing module, model module, and output module.

1. Input Module

This module handles:

- Dataset upload
- Image/video input
- User interaction through GUI

Users can upload datasets or test images/videos via file dialog options.

2. Preprocessing Module

This module performs:

- Image resizing and normalization
- Label encoding
- Bounding box scaling
- Dataset splitting

Efficient preprocessing ensures improved model performance and generalization.

3. Model Module

The model module is the core component and includes:

- VGG16 backbone for feature extraction
- Region Proposal Network (RPN)
- Classification head
- Bounding box regression head

The model processes input images and outputs:

- Food category
- Bounding box coordinates

4. Training Module

Responsible for:

- Model training
- Loss optimization
- Saving trained model
- Recording training history

5. Prediction Module

This module:

- Loads trained model
- Processes new images/videos
- Predicts food class and location
- Displays results visually

6. GUI Module

The GUI provides:

- Buttons for each functionality
- Output display panel
- Graph visualization

7. Output Module

Outputs include:

- Classified food label
- Bounding box visualization
- Nutritional information

System Architecture Flow

Input → Preprocessing → Feature Extraction → Region Proposal → Classification & Localization → Output

Design Advantages

- Modular and scalable
- User-friendly interface
- Real-time processing capability
- Extendable for advanced models

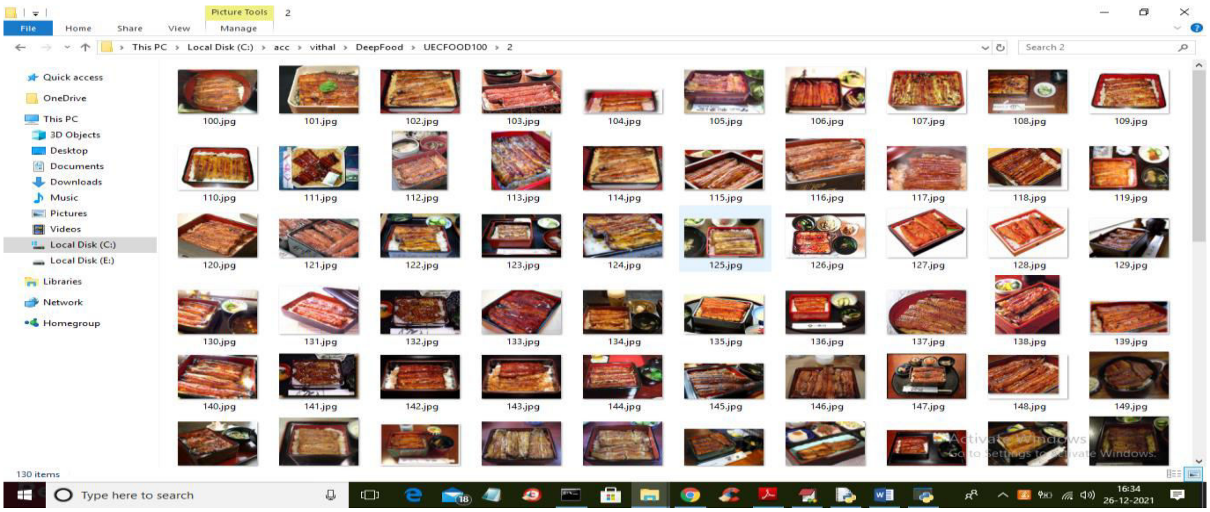
SYSTEM DESIGN IMAGES

Food is an important for human life and its mandatory for them to know dietary details before consuming them to lead a healthy life and to automatically identify dietary details author of this paper has introduce REGION base convolution neural network algorithms which get trained by using regions from the image and this regions will be in the form of food and this algorithm not only detect region of food but also classify food and based on that food classification dietary details will be displayed.To implement this project author has used VGG16 based Faster RCNN (Region Convolution Neural Network) algorithm and to trained this algorithm author has used UECFOOD 100 and 250 dataset. This dataset contains bounding boxes only for one food in the plat so this algorithm can efficiently identified on food from plate. We searched a lot to find dataset with multiple bounding boxes in plat but we don't find any such dataset so we also trained this FRCNN algorithm using UECFOOD 100 dataset. Author is saying he has built new dataset but he has not publish that dataset so our algorithm can detect one food from plate and this same algorithm can be trained to detect multiple foods in a plate if we found such dataset.

To implement this project we have designed following modules

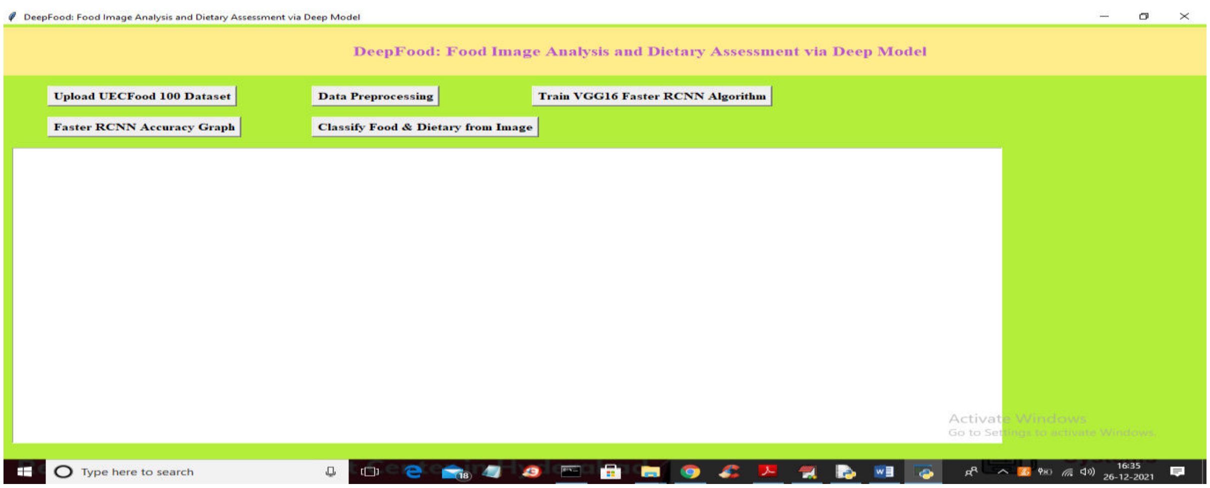
1) Upload UECFood 100 Dataset: using this module we will upload food dataset to application
2) Data Preprocessing: using this module we will read all food images and bounding boxes from dataset and this bounding boxes helps in extracting REGIONS from images. In this module training data will be generated using images and bounding boxes regions
3) Train VGG16 Faster RCNN Algorithm: using this module we will trained VGG16 FRCNN algorithm by using preprocessed images and bounding boxes region and then calculate accuracy and loss of the training model and our able to achieve 92% food classification accuracy.
4) Faster RCNN Accuracy Graph: using this module we will plot FRCNN training accuracy and loss graph
5) Classify Food & Dietary from Image: using this module we will upload test image and then application will apply FRCNN model on test image to classify food from the plate and display dietary details

Below is the dataset images used to train FRCNN algorithm

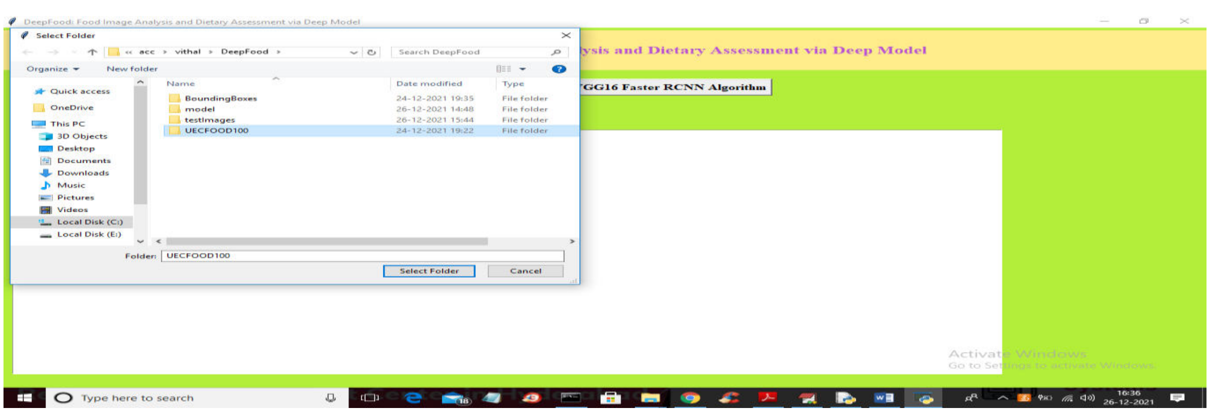


SCREEN SHOTS

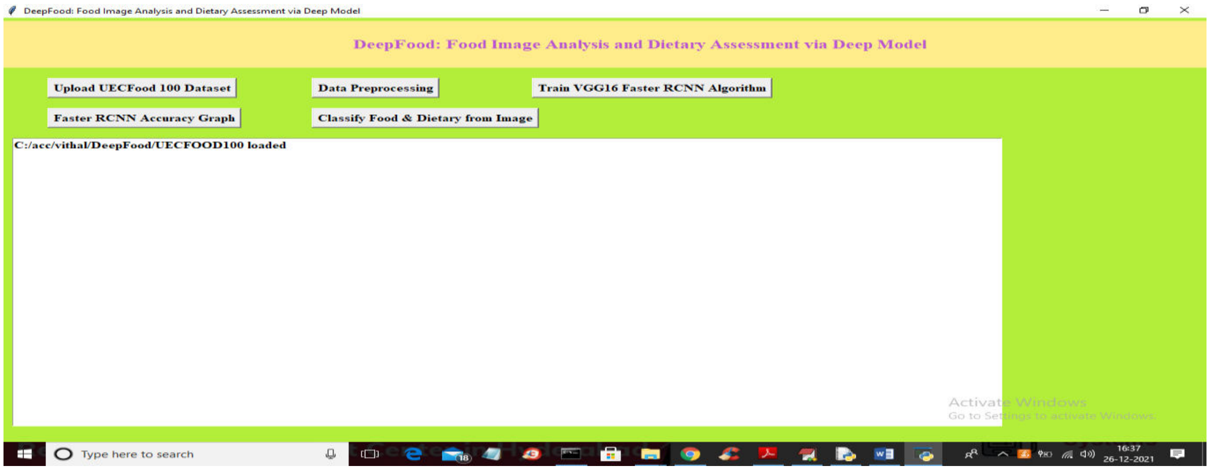
To run project double click on ‘run.bat’ file to get below screen



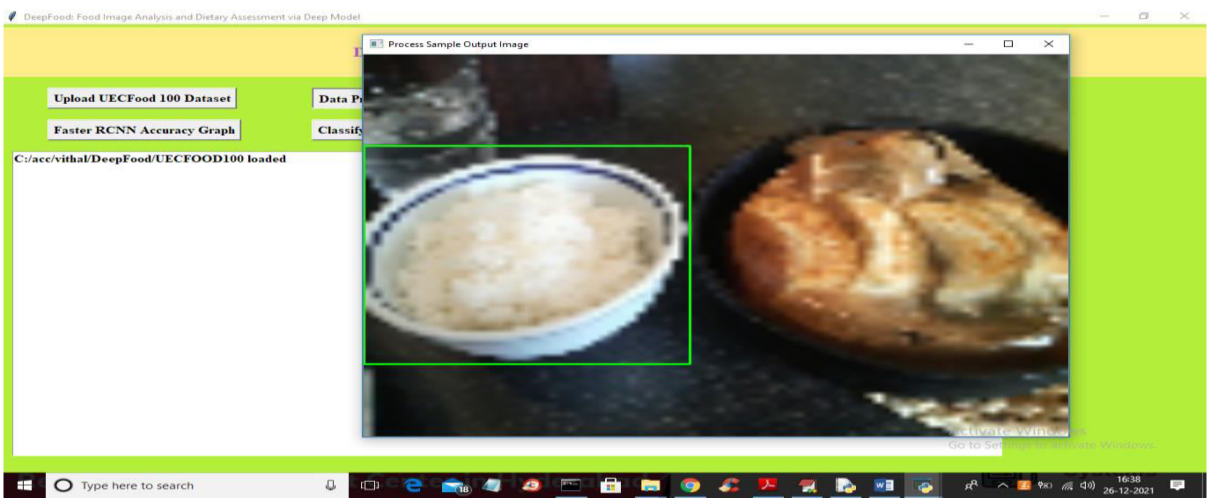
In above screen click on ‘Upload UECFood 100 Dataset’ button to upload dataset and to get below screen



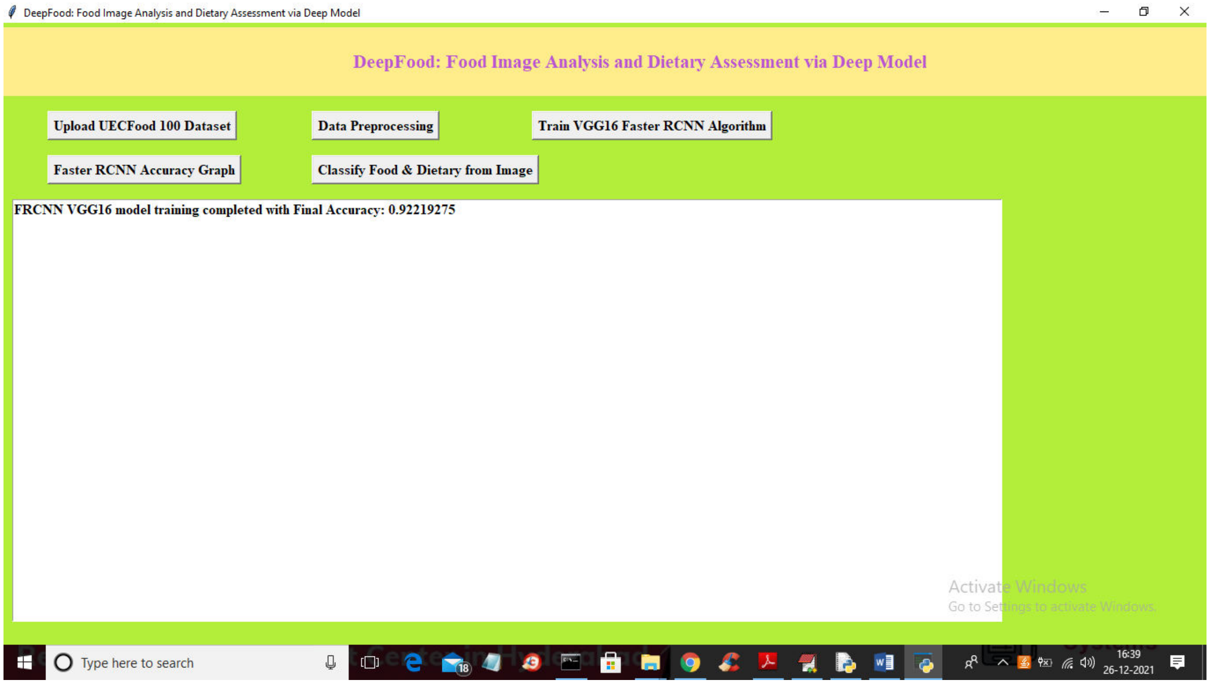
In above screen selecting and uploading 'UECFood100' dataset folder and then click on 'Select Folder' button to load dataset and to get below screen



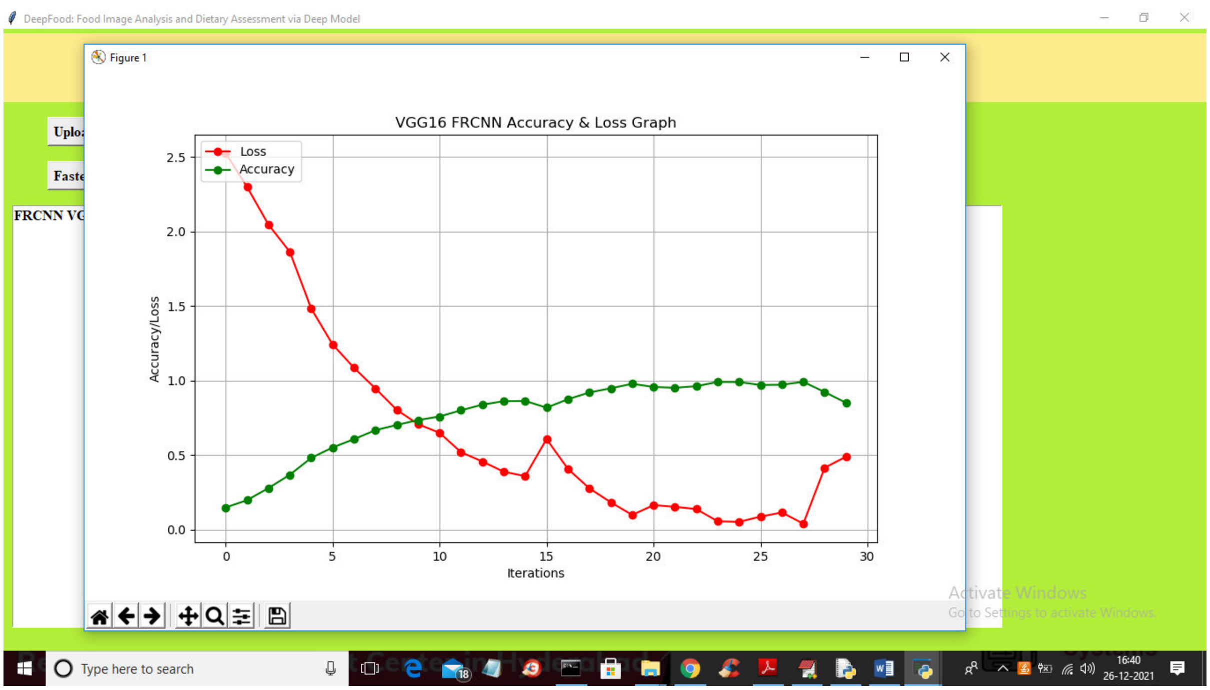
In above screen dataset loaded and now click on 'Data Preprocessing' button to read all images and then build a training data and to get below screen



In above screen we can see application process all images and then region bounding boxes and we can see dataset has only one region or bounding box for each image and now close above image and then click on 'Train VGG16 Faster RCNN Algorithm' button to start training FRCNN algorithm and to get below screen

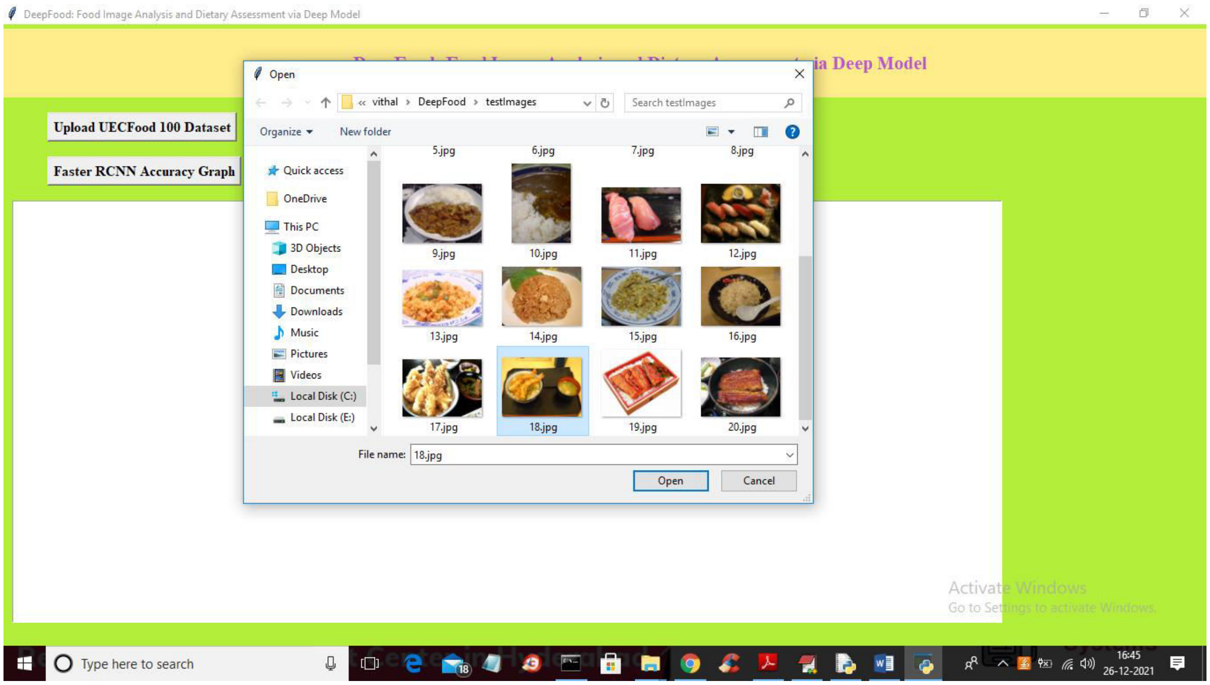


In above screen we can see FRCNN training completed and we got 92% accuracy and now click on 'Faster RCNN Accuracy Graph' button to get below graph



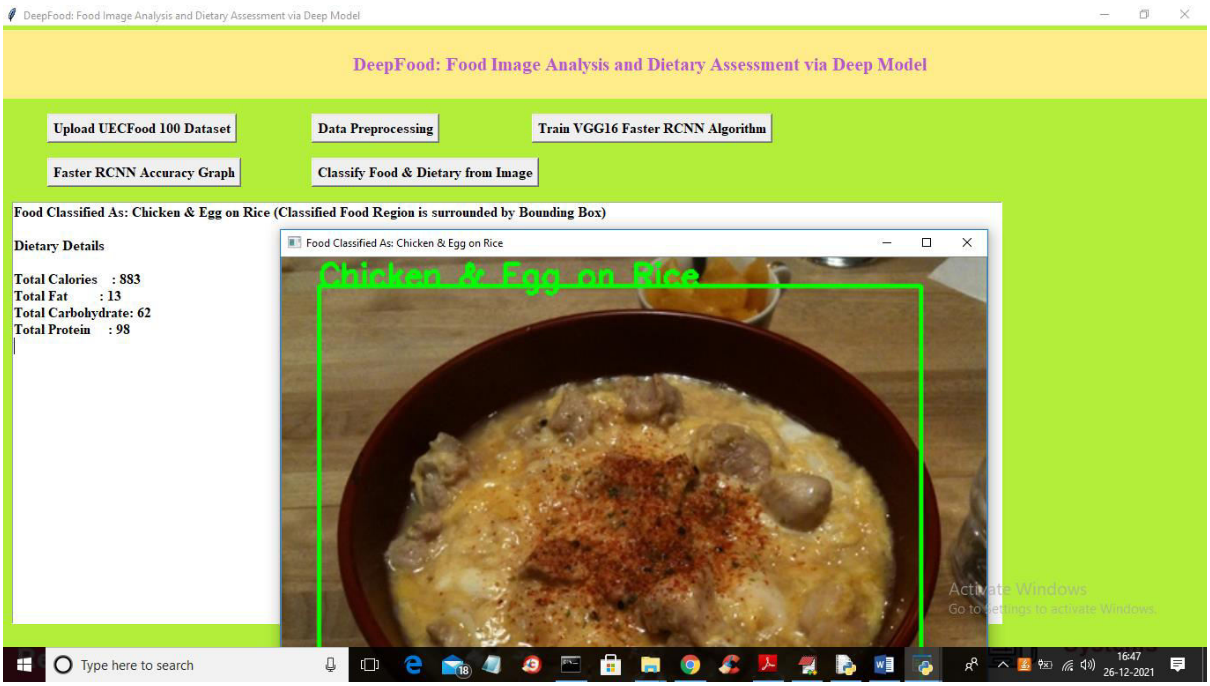
In above screen green line represents accuracy and red line represents loss and in x-axis we have NUMBER OF EPOCH and y-axis represents accuracy and loss values. In above graph we can see with each increasing epoch accuracy got increase and loss got decrease.

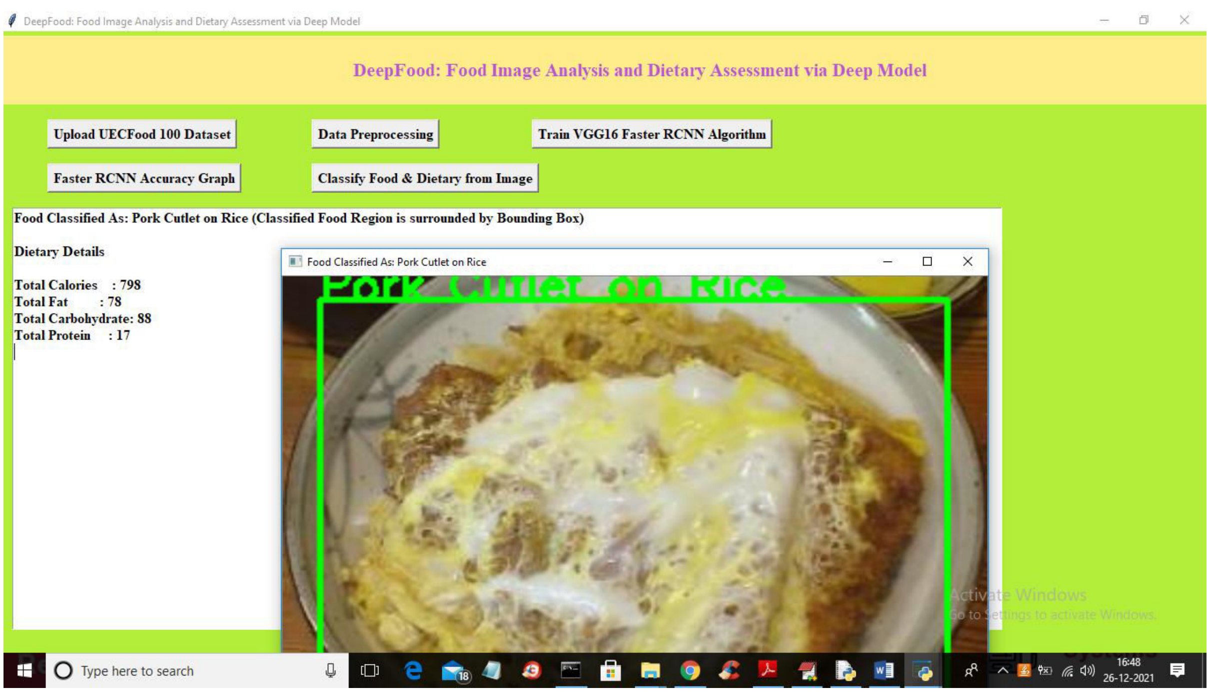
Now close above graph and then click on ‘Classify Food & Dietary from Image’ button to upload test image and classify food



In above screen I am selecting and uploading ’18.jpg’ file and then click on ‘Open’ button to get below output

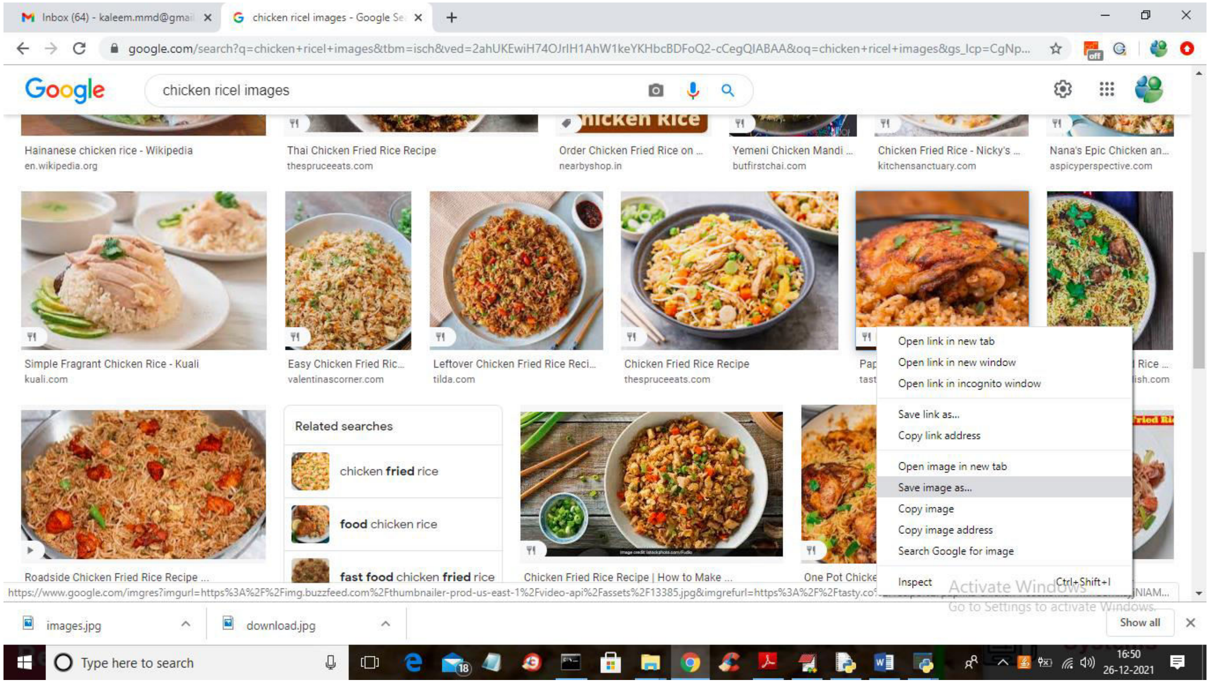
In above screen in text area we can see application classify food name as ‘Tempura Bowl’ and then displaying dietary details with REGION bounding box across identified food and now test other images

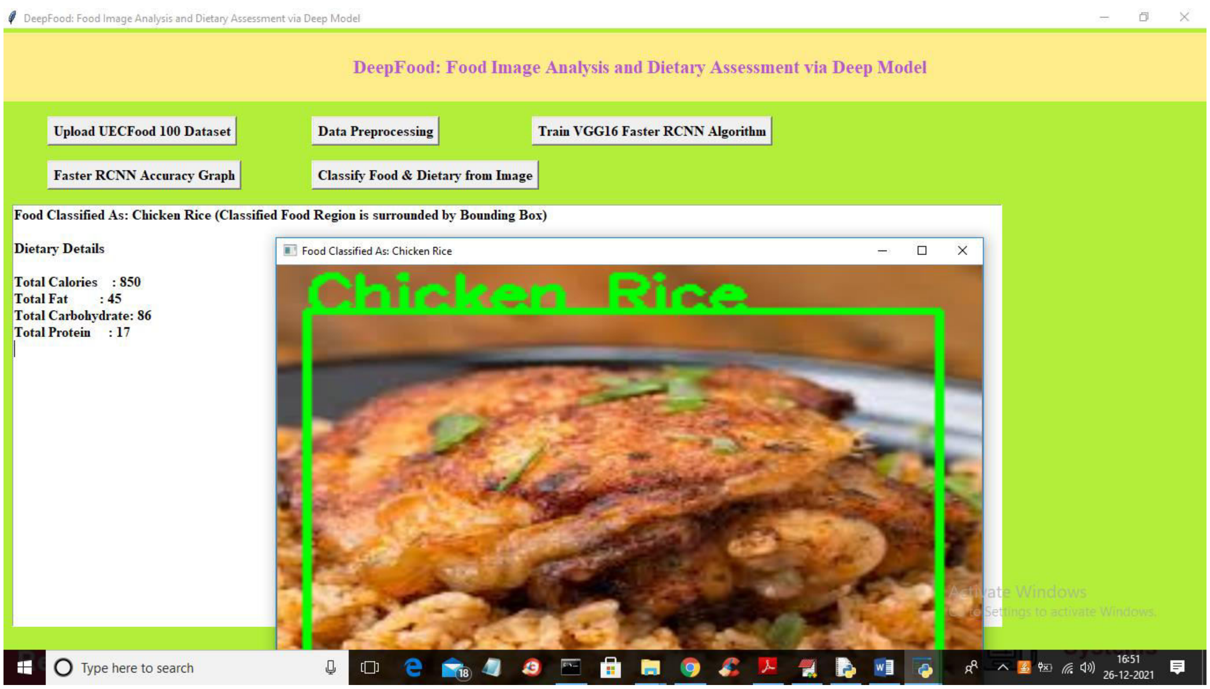




Similarly you can upload other images and test it

Below image we are downloading from GOOGLE and application is correctly classifying it





In above screen application correctly classifying GOOGLE image as Chicken Rice

VIII. CONCLUSION

The DeepFood system demonstrates an effective approach for automated food recognition and dietary assessment using deep learning techniques. By integrating VGG16 with Faster R-CNN, the system achieves accurate food classification and localization, making it suitable for real-world applications.The use of transfer learning enhances feature extraction while reducing training time and computational cost. Multi-task learning enables simultaneous optimization of classification and bounding box prediction, improving overall system efficiency.The implementation of a user-friendly GUI ensures accessibility for non-technical users, allowing easy interaction with the system. The ability to process both images and videos further enhances its practicality.Although the current system provides approximate dietary values, it lays the foundation for integrating real nutritional databases and advanced estimation techniques. Recent research highlights the growing importance of AI-driven dietary monitoring systems in healthcare and wellness applications .Overall, the proposed system offers a scalable, efficient, and practical solution for intelligent food analysis and dietary monitoring.

REFERENCES

Core Food Image Recognition (Recent & Highly Relevant)

1. Liu, D., et al., “Deep Learning in Food Image Recognition: A Comprehensive Review,” *Applied Sciences*, 2025.
2. Boyd, L., et al., “Fine-Grained Food Image Recognition Using Optimised CNNs,” *Journal of Imaging*, 2024.
3. Zuo, E., et al., “Recent Advances in Food Image Recognition Using Deep Learning,” *Applied Sciences*, 2025.

Advanced Architectures (Latest Research Trends)

1. Zhang, Y., et al., “CNN-Transformer Hybrid Model for Fine-Grained Food Recognition,” *Journal of Food Composition and Analysis*, 2025.
2. Zou, S., et al., “Fraesormer: Adaptive Sparse Transformer for Food Recognition,” *arXiv*, 2025.
3. Liu, L., & Xiao, Z., “Feature-Enhanced TResNet for Fine-Grained Food Classification,” *arXiv*, 2025.
4. Song, J., & Liu, X., “SalientFusion: Zero-Shot Food Recognition Using Context-Aware Models,” *arXiv*, 2025.

Object Detection & Faster R-CNN-Based Studies

1. Tan, S. W., et al., “Food Detection and Recognition with Deep Learning: A Comparative Study,” *IEEE ICoICT*, 2023.
2. Ren, S., et al., “Faster R-CNN: Towards Real-Time Object Detection,” *IEEE TPAMI* (foundational but still widely cited), 2015
3. Magdy, A., et al., “Lightweight Faster R-CNN for Efficient Object Detection,” *Scientific Reports*, 2025

CNN-Based Food Recognition Systems

1. Rao, D. K., et al., “Deep Learning-Based Food Image Classification Using Enhanced CNN,” *IJCERT*, 2024.
2. Kagaya, H., et al., “Food Detection and Recognition Using CNN,” *ACM Multimedia*, 2014

Dietary Assessment & Applications

1. Myers, A., et al., "Im2Calories: Towards Automated Mobile Vision Food Diary," *ICCV*, 2015
2. Min, W., et al., "Food Image Recognition for Dietary Assessment: A Survey," *IEEE Transactions*, 2023
3. Chen, C., et al., "Res-VMamba: Fine-Grained Food Classification Using State Space Models," *arXiv*, 2024