RV College of Engineering®

Go, change the world

Department of AIML

Recipe Generation from Food Images

Presented by Students

KM Amogha-1RV22AI070

Pranshu Bhatt- 1RV22AI071

Sujay Arun Kudtarkar- 1RV22AI059

Faculty Mentors: Prof Somesh Nandi

CONTENTS

1. Agenda
2. Introduction
3. Literature Survey
4. Summary of LS
5. Requirement analysis – hardware and software specification
6. System architecture (-- ANN-DL architecture) Eg .. architecture of CNN
7. Methodology
8. Module specification –

a.Module 1 : data collection and pre processing

i.Input ii. Process iii output

b.Module 2 : Implementation of ANN / DL algorithm

i.Input ii. Process iii output

c.Module 3 : testing and validation

i.Input ii. Process iii output

AGENDA

The agenda of this project is to build an AI-powered system that takes an image of food as input and generates a recipe based on the detected ingredients. The system offers personalization options (dietary restrictions) and dynamic scaling of ingredient quantities. The main objectives include:

1. Accurate ingredient prediction using image analysis.
2. Dynamic recipe generation and scaling.
3. Providing a user-friendly interface for interaction.

INTRODUCTION

This project focuses on developing a system that can identify food items from images and generate detailed, step-by-step recipes. It combines the power of computer vision for image analysis and natural language processing for recipe creation.

Objective:

- To create an intelligent tool that recognizes food images, extracts relevant features, and produces human-readable recipes. The aim is to enhance user experience in the kitchen by providing automated, accurate recipe suggestions.

Applications:

- Smart Kitchen Assistants: Helps users in real-time by suggesting recipes based on available food.
- Dietary Planning: Assists in creating personalized meal plans based on visual input.
- Convenience: Simplifies cooking for individuals by automating recipe generation for various cuisines.

Literature Survey

Sl No	Author and Paper title	Details of Publication	Summary of the Paper
1.	B. Jamalpur, S. P. G, S. Venkat, A. R, S. Kavitha and G. C. Babu, "Optimizing Food Image Classification Using Black Widow Algorithm and Deep Learning Techniques,"	2024 Second International Conference on Advances in Information Technology (ICAIT), Chikkamagaluru, Karnataka, India, 2024, pp. 1-6	This research introduces the FIC-BWODL model for food image classification, integrating CLAHE for preprocessing, CapsNet for feature extraction, and CAE for classification. The Black Widow Optimization Algorithm is used for hyperparameter tuning, enhancing performance and leveraging unique dataset features.
2.	H. V. Hasti, S. A. Jatan, T. Nahar, V. Dubey and C. Parmar, "Image Processing and Machine Learning Methods for Assessing Food Quality,"	2024 International Conference on Data Science and Network Security (ICDSNS), Tiptur, India, 2024, pp. 01-06,	This review explores advanced techniques for agricultural quality assessment, including CNN-based wheat contamination detection, 92.57% accuracy in apple freshness estimation, and 97.33% accuracy in apple bruise classification using hyperspectral imaging. It also highlights automated meat quality scanning and cost-effective rice classification, enhancing efficiency and industry competitiveness.

Literature Survey

Sl No	Author and Paper title	Details of Publication	Summary of the Paper
3.	S. P. Singh, D. Siddharth, S. Prabakeran, R. Das and A. Balaji, "Food-Lens: Improving Culinary Experiences with AI-Driven Meal Analysis and Recipe Generation,"	2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 2024, pp. 1385-1391	This research presents an AI-based system using CNN and MobileNetV2 to identify food from images and generate recipes, including ingredients and instructions. promotes healthier eating through features like recipe suggestions and text-to-speech capabilities
4.	R. Krutik, C. Thacker and R. Adhvaryu, "Advancements in Food Recognition: A Comprehensive Review of Deep Learning-Based Automated Food Item Identification,"	2024 2nd International Conference on Electrical Engineering and Automatic Control (ICEEAC), Setif, Algeria, 2024, pp. 1-6	This review highlights advancements in food recognition using deep learning, emphasizing its role in dietary monitoring and nutritional analysis. It examines motivations, state-of-the-art architectures, and publicly available datasets, while identifying research gaps and proposing future directions in this evolving field.

Literature Survey

Sl No	Author and Paper title	Details of Publication	Summary of the Paper
5.	P. Chhikara, D. Chaurasia, Y. Jiang, O. Masur and F. Ilievski, "FIRE: Food Image to REcipe generation,"	2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2024, pp. 8169-8179	FIRE is a multimodal methodology for generating recipes from food images, using models like BLIP for titles, Vision Transformers for ingredients, and T5 for instructions.
6.	E. D. Cherpanath, P. R. Fathima Nasreen, K. Pradeep, M. Menon and V. S. Jayanthi, "Food Image Recognition and Calorie Prediction Using Faster R-CNN and Mask R-CNN,"	2023 9th International Conference on Smart Computing and Communications (ICSCC), Kochi, Kerala, India, 2023, pp. 83-89	This project proposes using deep learning to address health issues related to obesity by calculating food calories from user images. The approach involves image capture, food classification, and calorie prediction, aiding in dietary awareness and calorie control to prevent obesity-linked conditions.

Literature Survey

Sl No	Author and Paper title	Details of Publication	Summary of the Paper
7.	F. S. Konstantakopoulos, E. I. Georga and D. I. Fotiadis, "A Review of Image-Based Food Recognition and Volume Estimation Artificial Intelligence Systems,"	in IEEE Reviews in Biomedical Engineering, vol. 17, pp. 136-152, 2024,	This review examines food recognition and volume estimation using computer vision on smartphone images, analyzing methods for segmentation, classification, and volume computation. It highlights strengths, limitations, and future directions for dietary assessment systems.
8.	T. M. L. Rosaline, C. A. Suciawan, W. Ng, S. Achmad and J. V. Moniaga, "AI-Powered Mobile Application for Image-Based Food Ingredient Detection and Recipe Generation,"	2024 International Conference on Information Management and Technology (ICIMTech), Bali, Indonesia, 2024, pp. 123-128, doi: 10.1109/ICIMTech63123.2024.10780810.	This study presents a mobile app leveraging YOLOv8 and CNN to detect food ingredients and assess nutrition. Trained on 24,583 images, the model improves accuracy and provides calorie counts with tailored recipes. Future work includes expanding datasets for enhanced precision.

RV College of Engineering®

Literature Survey

Go, change the world

Sl No	Author and Paper title	Details of Publication	Summary of the Paper
9.	M. S. M. Rabby, M. Uddin, E. Islam, M. Khaliluzzaman, M. N. I. Shanto and A. Islam, "A Modified Transfer Learning-Based Framework for Efficient Food Image Classification,"	2024 IEEE International Conference on Computing, Applications and Systems (COMPAS), Cox's Bazar, Bangladesh	This study explores food image classification using EfficientNetB7, ResNet50, and VGG19 with the Food-11 dataset. EfficientNetB7 achieved 87.38% validation accuracy, and VGG19 variants, enhanced with augmentation and regularization, reached up to 83%.
10.	P. K. Singh and S. Susan, "Transfer Learning using Very Deep Pre-Trained Models for Food Image Classification,"	2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India,	This study evaluates food image classification using deep pre-trained CNN architectures on the Food-101 dataset, which includes 101 food categories. Among Inception-v3, EfficientNet-B0, Xception, DenseNet-121, and MobileNet, Xception achieved the highest accuracy of 84.54%, outperforming the others significantly.

Literature Survey

Sl No	Author and Paper title	Details of Publication	Summary of the Paper
11.	S. S, P. Kokil, A. R and N. V. Sai Manoj, "Enhanced Food Classification System Using YOLO Models For Object Detection Algorithm,"	2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-6	This study evaluates YOLOv5 and YOLOv7 for food classification, highlighting YOLOv5's superior performance in accuracy (0.851), recall (0.836), and mAP (0.892 at 0.5 IoU). The findings underscore YOLOv5's potential for enhancing food categorization, nutrition, and safety in the food industry.
12.	J. Sultana, B. M. Ahmed, M. M. Masud, A. K. O. Huq, M. E. Ali and M. Naznin, "A Study on Food Value Estimation From Images: Taxonomies, Datasets, and Techniques,"	in IEEE Access, vol. 11, pp. 45910-45935, 2023	This review analyzes methods and datasets for automating nutritional value estimation from food images using deep learning. It categorizes existing research, evaluates performance metrics like accuracy and precision, and highlights current trends, challenges, and future directions in the field, offering insights for researchers, health practitioners, and nutritionists.

Literature Survey

Sl No	Author and Paper title	Details of Publication	Summary of the Paper
13.	S. Chaudhary, B. Soni, A. Sindhavad, A. Mamaniya, A. Dalvi and I. Siddavatam, "ChefAI.IN: Generating Indian Recipes with AI Algorithm,"	2022 International Conference on Trends in Quantum Computing and Emerging Business Technologies (TQCEBT), Pune, India, 2022, pp. 1-6	Focuses on generating unique Indian recipes using the Autochef algorithm supported by mutation and similarity techniques. It employs models like NLP, LSTM, to develop and refine recipes, enhancing accuracy and efficiency for Indian cuisine.
14.	J. N. V. R. S. Kumar, C. Jyothsna, K. Srinivas, A. E. Sravanth, A. T. Kumar and D. Indira, "Self-Attention Architecture for Ingredients Generation from Food Images,"	2022 Second International Conference on Next Generation Intelligent Systems (ICNGIS), Kottayam, India, 2022, pp. 1-7	This research presents an image-to-recipe generation system that predicts a dish's title, ingredients, and cooking instructions from its image, with a focus on Indian cuisines.

Literature Survey

Sl No	Author and Paper title	Details of Publication	Summary of the Paper
15.	D. P. Papadopoulos, E. Mora, N. Chepurko, K. W. Huang, F. Ofli and A. Torralba, "Learning Program Representations for Food Images and Cooking Recipes,"	2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022, pp. 16538-16548	This work generates cooking programs from images or recipes using a Vision Transformer-based encoder for images, a Transformer-based encoder for recipes, and a program decoder. The model aligns image and recipe embeddings in a shared space via self-supervision for program generation.
16.	N. -I. Galanis and G. A. Papakostas, "An update on cooking recipe generation with Machine Learning and Natural Language Processing,"	2022 IEEE World Conference on Applied Intelligence and Computing (AIC), Sonbhadra, India, 2022, pp. 739-744	Generating recipes based on user-provided ingredient lists or by suggesting ingredients. Leveraging advancements in natural language processing and deep learning, these methods enable the creation of innovative, personalized, and healthier recipes while providing inspiration and context for culinary endeavors.

Literature Survey

Sl No	Author and Paper title	Details of Publication	Summary of the Paper
17.	H. Wang, G. Lin, S. C. H. Hoi and C. Miao, "Learning Structural Representations for Recipe Generation and Food Retrieval,"	in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 3, pp. 3363-3377 , 2022	Unsupervised method to generate level tree structures for cooking recipes, enhancing recipe generation and food cross-modal retrieval tasks. By leveraging ON-LSTM, the approach extracts paragraph structures, generates recipe trees from images, and integrates them into generation and retrieval models.
18.	M. Goel et al., "Ratatouille: A tool for Novel Recipe Generation,"	2022 IEEE 38th International Conference on Data Engineering Workshops (ICDEW), Kuala Lumpur, Malaysia, 2022, pp. 107-110	This research introduces Ratatouille, a web application for generating novel recipes using neural network-based LSTMs and the transformer-based GPT-2 model. By incorporating ingredient quantities, GPT-2 outperformed LSTMs with a BLEU score of 0.806.

Literature Survey

Sl No	Author and Paper title	Details of Publication	Summary of the Paper
19.	F. P. W. Lo, Y. Sun, J. Qiu and B. Lo, "Image-Based Food Classification and Volume Estimation for Dietary Assessment: A Review,"	in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 7, pp. 1926-1939 , 2020	This study reviews image-based dietary assessment methods, comparing algorithms and models for food recognition and volume estimation. It highlights challenges and opportunities in improving accuracy, speed, and efficiency in dietary analysis, emphasizing the potential of integrated systems combining deep learning and other approaches for better dietary intake assessment.
20.	A. -S. Metwalli, W. Shen and C. Q. Wu, "Food Image Recognition Based on Densely Connected Convolutional Neural Networks,"	2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Fukuoka, Japan	This study introduces DenseFood, a densely connected convolutional neural network designed for food image recognition. The model employs softmax and center loss functions to enhance intra-category consistency and inter-category distinction. Using the VIREO-172 dataset, DenseFood achieved 81.23% accuracy, surpassing fine-tuned DenseNet121 and ResNet50 models in performance.

Requirement Analysis

Hardware Requirements(Developer Side)

- Storage :
  - Minimum: ~5 GB (including project files and dependencies).
- Memory (RAM):
  - Minimum: 4 GB RAM.
- Processor (CPU):
  - Minimum: Dual-core (2.0 GHz or higher) CPU (Intel Core i3 or equivalent).
- Graphics Processing Unit (GPU):
  - Minimum: Integrated graphics (e.g., Intel UHD or AMD Radeon Vega).

Software Requirements(Developer Side)

- Operating System:
  - Minimum: Windows 10 or Ubuntu 18.04.
- Programming Languages:
  - Minimum: Python 3.7+.
  - Java or Kotlin (for Android app development).
- Machine Learning Libraries:
  - Minimum: TensorFlow 2.0+ or PyTorch (for model development and training).
- Mobile Development Frameworks:
  - Minimum: Android SDK.

CNN Architecture

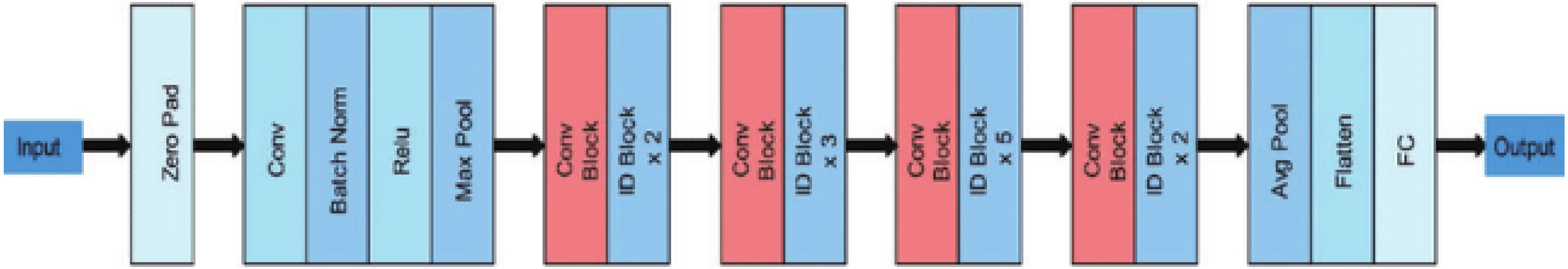


Fig 4.3 CNN Resnet Architecture

CNN Architecture

1. **Convolutional Layers**

The first layers of ResNet-101 consist of multiple convolutional layers that extract low-level visual features from the input food images.

- What it does: Detects basic patterns such as edges, textures, and color variations that help differentiate various food items.
- Importance in recipe generation: Helps in identifying key visual characteristics of dishes, such as whether a dish has pasta, bread, or leafy greens.

Example:

If the input image is a pizza, the convolutional layers detect circular edges, cheese texture, and tomato toppings, which later contribute to accurate classification.

2. **Residual Blocks**

One of the most powerful aspects of ResNet-101 is the use of residual blocks, which are designed to solve the vanishing gradient problem in deep networks.

- What it does: Allows information to flow directly across multiple layers via skip connections, making the training process more stable and efficient.
- Why it's important: Traditional deep CNNs suffer from gradient vanishing, where deeper layers struggle to learn. Residual connections help maintain meaningful gradients, ensuring better feature extraction for food images.

Example:

If two different images of a burger have different lighting conditions, residual blocks ensure that the model focuses on relevant features (such as the presence of a bun and patty) rather than getting confused by lighting differences.

CNN Architecture

3. **Global Average Pooling (GAP) Layer**

After extracting deep features from food images, ResNet-101 uses a Global Average Pooling (GAP) layer to convert feature maps into a compact representation.

- What it does: Instead of using fully connected layers, which have a large number of parameters, GAP reduces each feature map to a single value by taking the average of all activations.
- Why it's important: Reduces overfitting and helps in making the model more generalized across different food images.

Example:

For a spaghetti dish, GAP ensures that important high-level features (such as noodle patterns) are retained, rather than memorizing unnecessary details like the plate or background.

4. **Fully Connected (FC) Layer**

The extracted features from the convolutional layers and GAP layer are then passed through a Fully Connected (FC) layer.

- What it does: Converts extracted features into meaningful numerical representations that correspond to dish types or ingredients.
- Why it's important: The FC layer is responsible for mapping the visual representation of the dish to a text-based output, enabling the model to predict dish names and ingredients.

Example:

If an image of pasta is passed, the FC layer outputs a probability distribution that maps the features to categories like:

- Pasta - 95%
- Curry - 3%
- Burger - 2%

CNN Architecture

5. **Softmax Layer (Final Classification)**

The final layer of ResNet-101 is a softmax layer, which takes the outputs from the FC layer and converts them into a probability distribution over possible dish categories.

- What it does: Assigns probabilities to different dish types based on extracted features.
- Why it's important: Helps in identifying the most probable food category, which is then used to retrieve the corresponding ingredients and recipe steps from the RecipeNLG dataset.

Example:

For a fruit salad image, the softmax layer outputs:

- Fruit Salad – 97%
- Pasta – 2%
- Soup – 1%

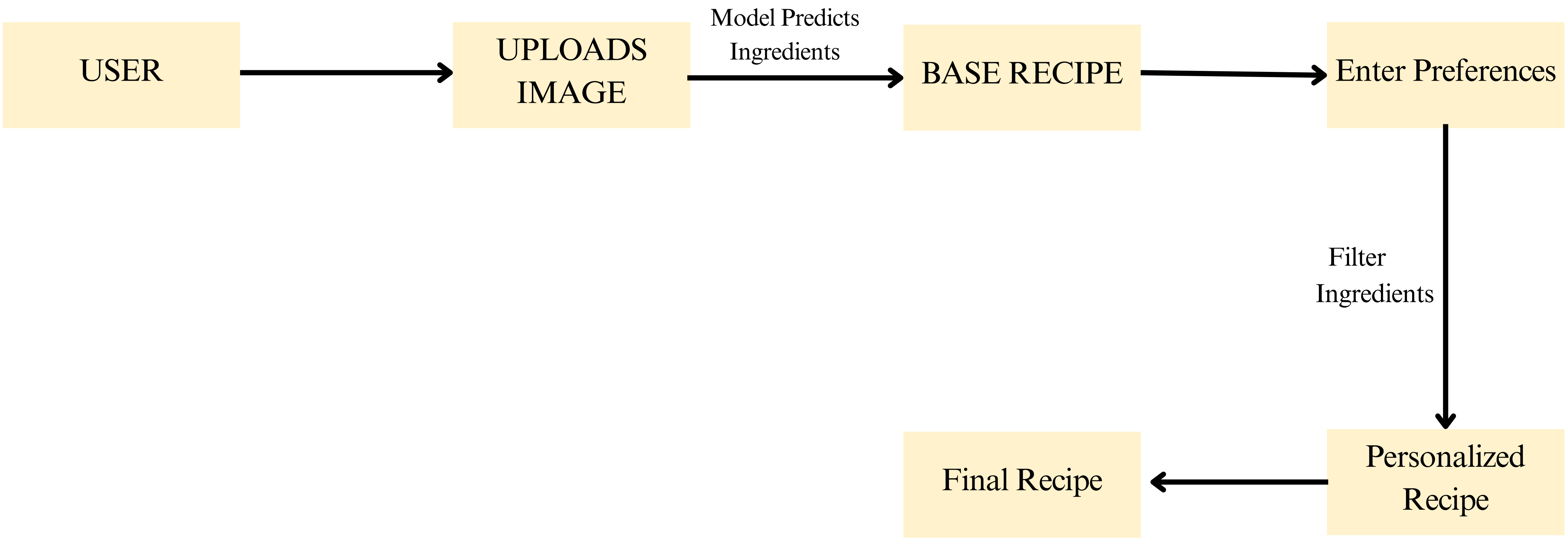
This classified dish name is then used to fetch its ingredients and recipe steps.

Final Role of ResNet-101 in Recipe Prediction

1. Processes the food image to extract meaningful features.
2. Classifies the image into the most probable dish type.
3. Maps the dish type to a list of ingredients from the RecipeNLG dataset.
4. Generates a recipe based on the dish and ingredients.

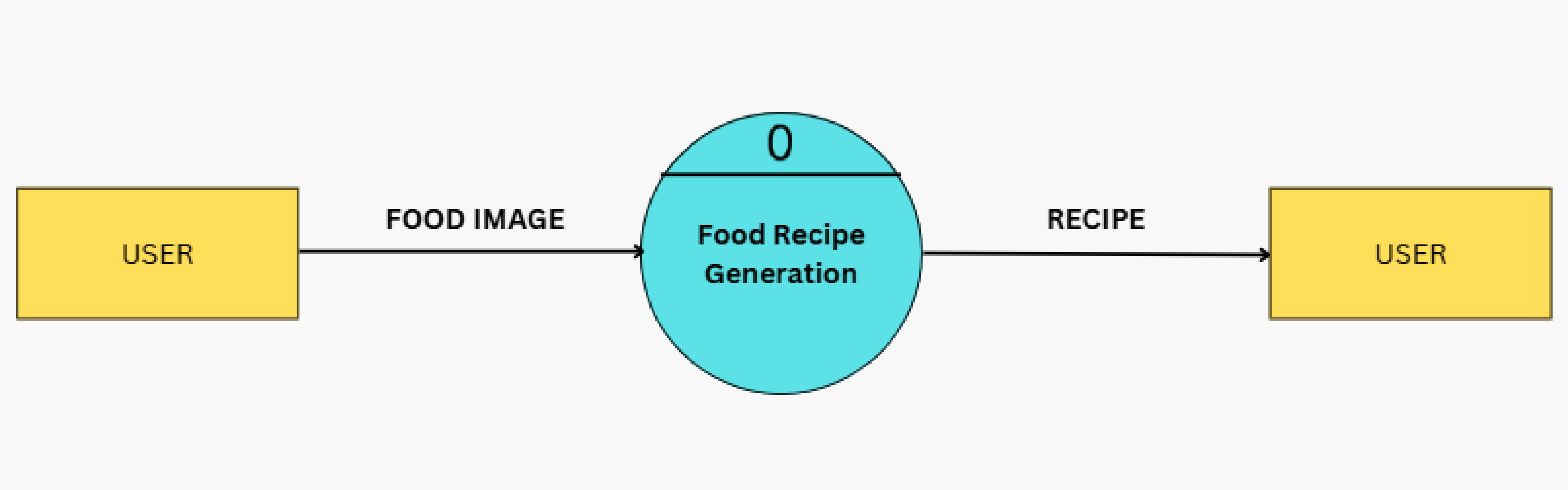
Thus, ResNet-101 acts as the backbone of the entire image-to-recipe prediction system, ensuring that food images are correctly identified and linked to appropriate recipes.

WORKFLOW

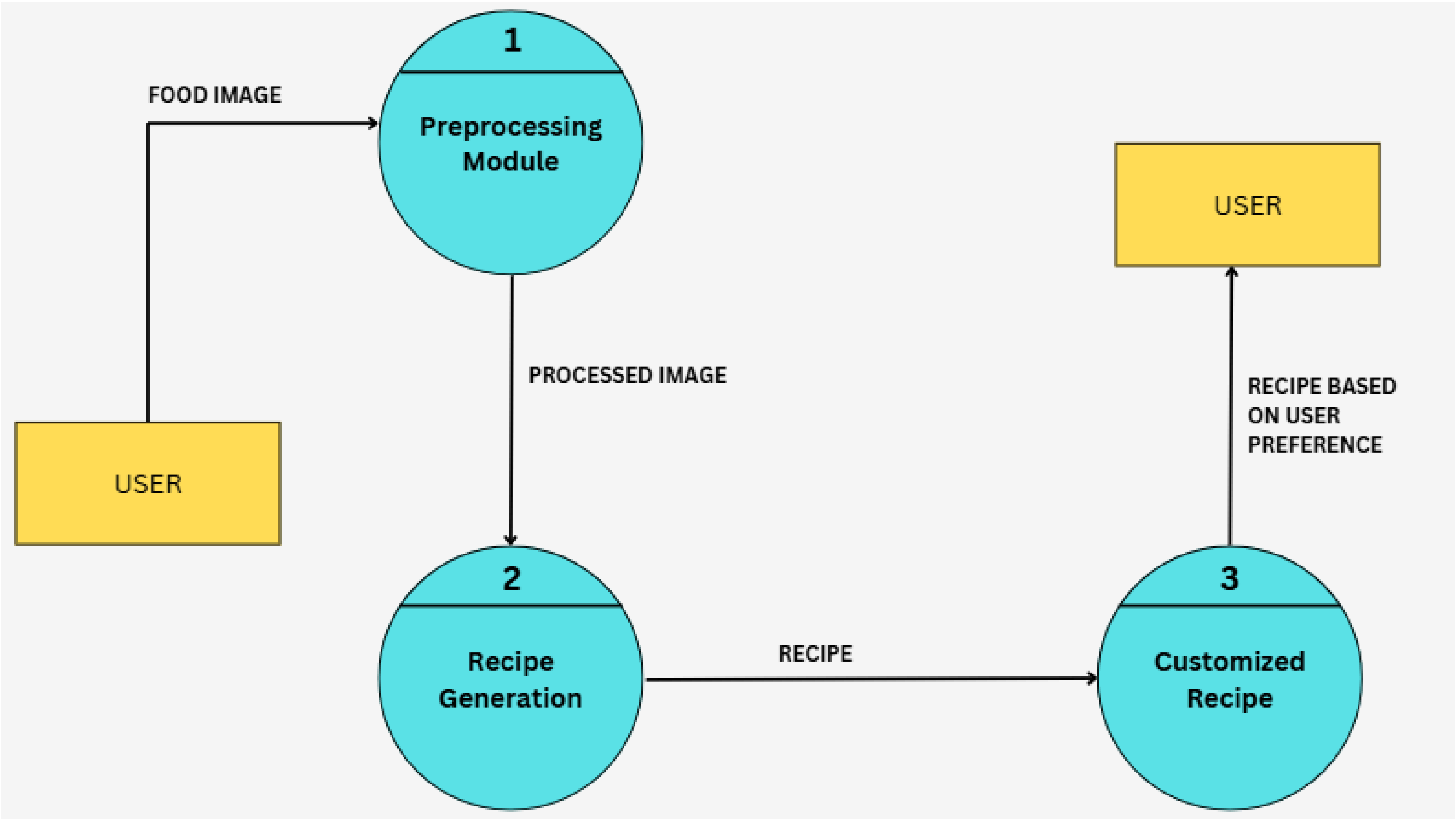

USER
UPLOADS IMAGE
Model Predicts Ingredients
BASE RECIPE
Enter Preferences
Filter Ingredients
Personalized Recipe
Final Recipe

DATA FLOW DIAGRAM

LEVEL 0



LEVEL 1


1
Preprocessing Module
FOOD IMAGE
USER
PROCESSED IMAGE
2
Recipe Generation
RECIPE
3
Customized Recipe
RECIPE BASED ON USER PREFERENCE
USER

METHODOLOGY

Module 1: Data Preprocessing

Input: Food images from RecipeNLG and other datasets, with metadata (dish name, ingredients, instructions, cuisine).

Process: Images are resized to 224x224 pixels, normalized to a 0-1 range, and passed through ResNet-101 for feature extraction.

Output: Preprocessed dataset with extracted feature vectors.

Module 2: Food Image Classification

Input: Preprocessed images with extracted features.

Process: Features are input into a classification model (e.g., ResNet-101) to predict food categories.

Output: Predicted food category (e.g., "Pasta," "Burger").

Module 3: Ingredient Prediction

Input: Predicted food category and feature vector.

Process: Dish is matched with potential ingredients, and a model predicts likely ingredients, filtering irrelevant ones.

Output: List of ingredients for the dish.

Module 4: Recipe Generation

Input: Predicted dish and ingredients list.

Process: System retrieves similar recipes or generates a new recipe using an AI-based model (LSTM, Transformer).

Output: Complete recipe with dish name, ingredients, instructions, and cooking time.

METHODOLOGY

Module 5: Recipe Presentation & User Interaction

Input: Generated recipe.

Process: Recipe displayed in a UI with options to save, modify, or share. User feedback helps improve future predictions.

Output: Fully formatted recipe for viewing, saving, or sharing.

RESULTS

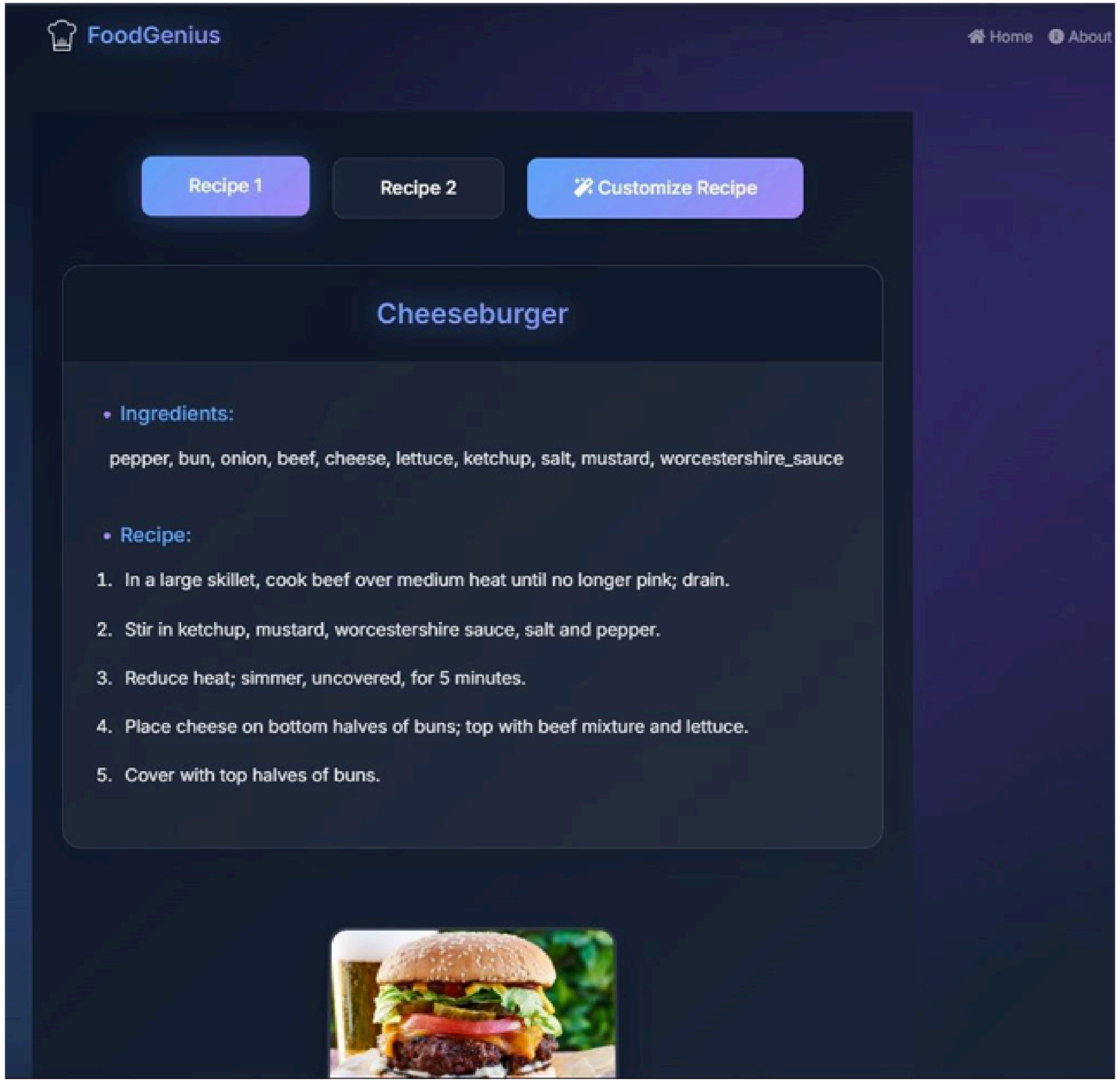


Fig 5.9 UI Implementation

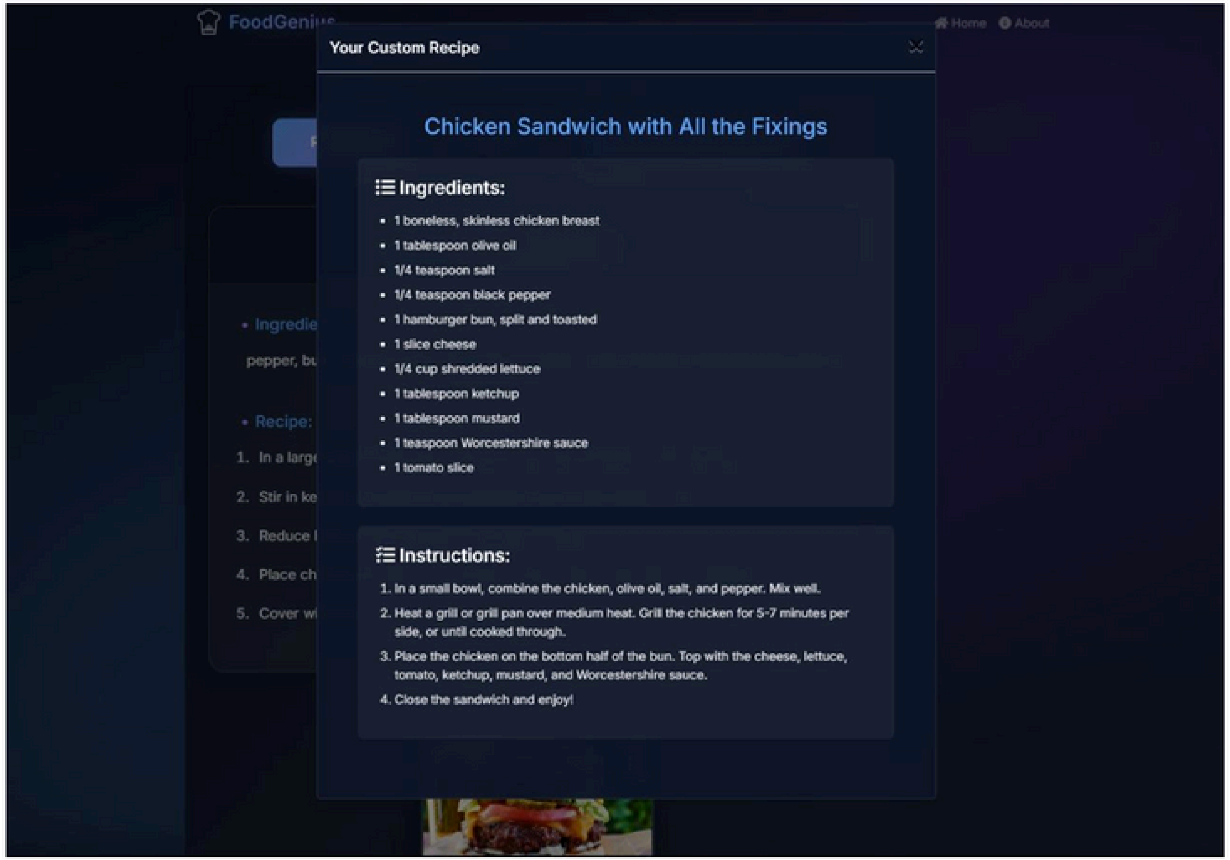


Fig 5.10 Customized recipe output

THANK YOU