Real-Time Ingredient Recognition and Recipe Generator

Mrs.Y.Nirmala Anandhi
B.Tech Artificial Intelligence and Data Science
Rajalakshmi Engineering College
Chennai
nirmalaanandhi.y@rajalakshmi.edu.in

M.Thofiq Gani
B.Tech Artificial Intelligence and Data Science
Rajalakshmi Engineering College
Chennai
221801057@rajalakshmi.edu.in

B.Sanjay
B.Tech Artificial Intelligence and Data Science
Rajalakshmi Engineering College
Chennai
221801045@rajalakshmi.edu.in

B.Soorya
B.Tech Artificial Intelligence and Data Science
Rajalakshmi Engineering College
Chennai
221801051@rajalakshmi.edu.in

***Abstract*— The blistering development of artificial intelligence in the sphere of food technology has suggested new solutions in the case of individual cooking support. In this paper, we will discuss a Real-Time Ingredient Recognition and Recipe Generation System, which uses computer vision and natural language processing to guide the user in preparing a meal. The suggested framework enables the user to add an image of food products, and it is performed with a pre-trained computer vision model based on deep learning and specific to identify ingredients. Known ingredients are compared against a predefined recipe corpus to produce dynamically context-specific cooking instructions. The system combines ingredient recognition with recipe creation to give the user healthy, varied, customized meal recommendations in real time. The experimental results prove the efficiency of the model in terms of ingredients classification and the possibility to produce relevant recipes with high precision. This method shows the possibility of integrating computer vision and data-driven generation of recipes to be used in the smart kitchen, dietary management, and food recommendation systems.**

Keywords*— Artificial Intelligence, Computer Vision, Deep Learning, Ingredient Recognition, Recipe Generation, Smart Kitchen, Food Technology.

I. INTRODUCTION

Cooking is a necessary daily activity and food is an essential part of human life. As artificially intelligent (AI) and computer vision continue to expand at a fast rate, some new technologies are arising that can transform the way individuals engage with food greatly. Automated ingredient recognition and recipe generation is one of the most promising applications in this field and may help users cook their meals in an efficient and creative way.

The classical approaches to recipe recommendation are normally based on manual input of ingredients into the system by a user or on the use of textual query to search recipes. The method is time-consuming, has a tendency of error and inconvenient in real-time cooking. As an example, a user can have several ingredients, but not understand what recipes they can make, or hear of some ingredients they are not conversant with. To address these issues, it is possible to adopt computer vision-driven solutions to automatically recognize ingredients based on pictures, which will enable the user to use the interface in a natural way - by just taking or uploading a picture.

The proposed Real-Time Ingredient Recognition and Recipe Generation System will solve these problems as it brings together deep learning on image based ingredient recognition and data guided recipe generation.Patrons can post a picture of the foodstuffs that they possess and a trained deep learning algorithm correctly determines the ingredients in the picture. The system then compares these ingredients to a curated recipe dataset to come up with recipes which can be cooked with current items. Such a dynamic solution not only will save time but also suggests individual, varied, and context-specific recipes.

This system is pragmatic and can be used beyond convenience in the smart kitchens, diet, health monitoring and food recommendation platforms, among others. It promotes effective use of ingredients and there is less food waste due to the fact that it recommends recipes depending on the items that a user already possesses. Moreover, the system will be easy to use by users of different cooking skills, as it can also produce clear and understandable cooking instructions, which is enabled by the natural language processing.

Overall, the presented work is a combination of computer vision, ingredient recognition based on AI and recipe creation to produce an intelligent, real-time solution to meal preparation. Its goal is to ease the cooking process, encourage healthy diets, and improve the experience of using the kitchen in the contemporary world.

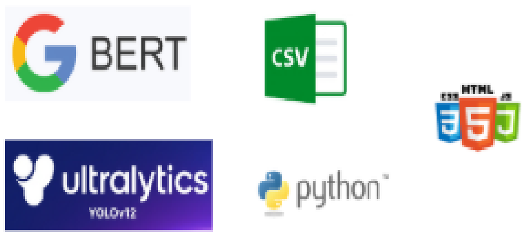

Figure 1.0 Technical Stack

II. Related Works

In the last 10 years, there has been major research in the sphere of food recognition and recipe generation with the help of artificial intelligence. The initial research was mostly concerned with image-based food classification where conventional machine learning algorithms were used to classify food types based on the images. Support Vector Machines (SVM) and k-Nearest Neighbors (k-NN) were more frequently used with handcrafted features such as color histograms and texture descriptors. Although these methods were mediocrely accurate, they could not handle complicated dishes that have many ingredients or different styles of presentation.

Convolutional Neural Networks (CNNs) replaced other face image recognition models as the favored option with the development of deep learning. The CNN-based models like AlexNet, VGGNet, and ResNet have proven to be outstanding in features hierarchy extraction of food images. It has been demonstrated that CNNs can dramatically enhance the accuracy of ingredient detection when compared to the classic machine learning methods. As an example, the Food-101 dataset has been leveraged extensively to train and evaluate deep learning models to do food classification and recognition tasks..

Over the past few years, there has been a research on multi-label ingredient recognition, where the images can have several ingredients rather than just one food item. This method refers to making a guess on all the potential contents of a dish at the same time. Other techniques like ResNet with attention mechanisms have been used to attend to the areas of the image that matter such as dishes that contain several ingredients or those that overlap to enhance the performance of the detector. These techniques are essential to real-time cooking assistance systems which need accurate identification of all the available ingredients.

In addition to recognizing ingredients, generating and recommending recipes has also become valuable research. The first approaches to recipe suggestion based on available ingredients were rule-based systems and collaborative filtering methods. Nonetheless, early systems offered limited flexibility and personalization, failings that stemmed from an inability to adjust to novel ingredient combinations and varying dietary needs.

Advancements in deep learning techniques incorporated in natural language processing (NLP) facilitate the development of more advanced systems for generating recipes. LSTM and Transformer sequence-to-sequence models have been applied for the generation of cooking instructions based on the identified ingredients. These models have the ability to not only generate recipes, but also provide detailed, step-by-step instructions, and thus, they achieve considerable value for cooking in real-time. Some research integrates CNN-based ingredient recognition and LSTM-based recipe generation to build end-to-end systems that automate recipe creation by analyzing food images.

Other research has centered on the creation and curation of data sets for food-related AI systems. Mega data sets, such as Recipe1M and Food-101, have been key to training deep learning models for ingredient-recognition and recipe generation. Such data sets include a sufficient range of food types, ingredient lists and diverse configurations, which enables the models to learn and generalize different styles and cuisines. Concerning research on cross-modal learning, more work has been done to combine the visual attributes of food and text from recipes to enhance the accuracy of recommendations.

Despite the advances, developing reliable real-time systems continues to pose numerous challenges. Model performance can be influenced by factors such as lighting changes, ingredient occlusion, look-alike ingredients, and undetectable dishes. Furthermore, the creation of personalized and situational recipes that account for user dietary restrictions or the availability of ingredients remains an unsolved research issue. The system proposed in this paper seeks to overcome some of these challenges by integrating computer vision as well as deep-learning-based ingredient identification, and data-driven recipe generation, to provide a streamlined and user-centric approach.

III.PROPOSED SYSTEM

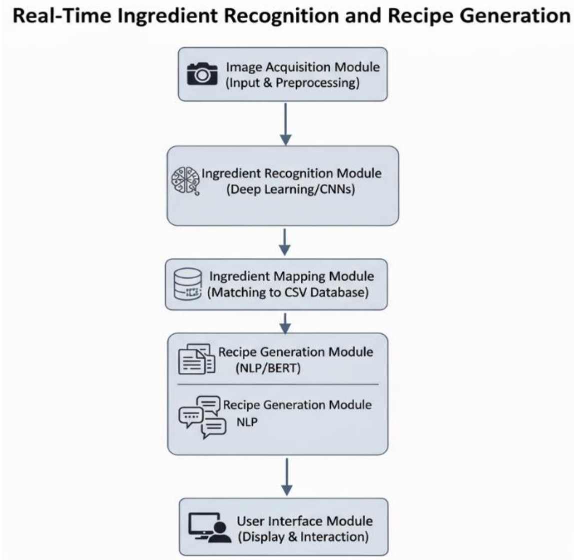

Figure 2.0 Overview of the System

What is Real-Time Ingredient Recognition and Recipe Generation? It is a system made up of five modules created to work harmoniously in delivering correct recognition of ingredients and generating instructions in response to user requests. With the Image Acquisition Module, system users can take a picture or upload one stored in their device gallery. Image recognition performance is also influenced by the input images. Hence, implementing the resizing, normalization, noise reducing, contrast improving steps defined in the Exposure section, we apply these steps in the Image Capture Module to guarantee our system recognizes images of the ingredients efficiently and accurately captures the input image in a standardized and predictable system response image.

Thirdly, we present to you the Ingredient Recognition Module, the system's heart. It uses image recognition techniques described in the previous chapter, in conjunction the deployed deep learning models and the trained neural networks as described in the portion, to recognize the ingredients in the users' submitted picture. In the last chapter, we stated the importance of CNNs, described ResNet or EfficientNet and the use of cross-food datasets to obtain accurate classification with multiple tags to allow the identification of multiple ingredients contained in one food dish. Further, we augment the models with Attention Mechanisms to enhance the recognition of relevant portions of the input image. This is complex in scenarios with layered foods such as salads, curries, or mixed food items..

After the ingredients are identified and placed in the appropriate fields, the next step is the mapping process which seeks to connect the detected ingredients to an entry in the recipe database. Ingredients can be detected by the system in varying names, and thus this process will include standardization to capture "garbanzo beans" and "chickpeas" in the same entry. It will also assess scenarios where some ingredients might not be available by proposing a recipe or making ingredient substitutions. This gives a degree of flexibility to ensure the detected ingredients are reasonably tied to suitable recipes.

Following this, the Recipe Generation Module constructs real-time, detailed sequential instructions for each dish by transforming the detected ingredients into processed instructions for the user with the help of Natural Language Processing (NLP). It employs Seq2Seq models like LSTMs and Transformers for the creation of contextually relevant and grammatically cohesive instructions. This system also accommodates recipes to other specific dietary requirements, e.g., vegetarian, vegan, or gluten-free. The system provides recipe variations to enhance diversity as well. For novice cooks, lacking a screen or recipe book will not be an obstacle as the system also provides step-by-step narration or voice instructions.

Lastly, the User Interface Module presents the results in an engaging and interactive manner. It displays the recognized ingredients and recipes for the users to see, and users can apply filters, change the amount of ingredients, or choose substitutions. It can integrate voice control and audio instructions for hands-free operation during cooking. The incorporation of personalization and accessibility features flexibility to the interface, catering for differing user needs, and ensuring the system remains cross device friendly, for mobile and smart kitchen assistants

System Architecture

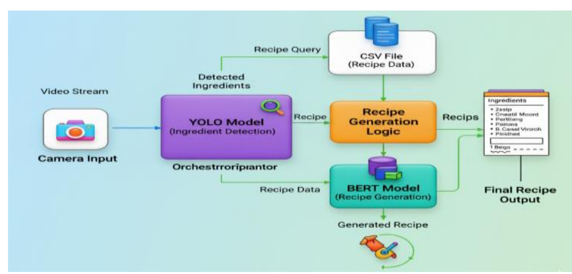

Figure 3.0 System Architecture

The architecture illustrates a Real-Time Ingredient Recognition and Recipe Generation System that employs computer vision and NLP to aid users in the cooking process. It all begins with Camera Input, which takes a video stream of the food items or captures a still image. This input goes to the YOLO Model for ingredient detection. YOLO (You Only Look Once) is a deep learning state-of-the-art object detection algorithm that can identify several food ingredients instantaneously and in real time.

The identified ingredients are sent to the Recipe Generation Logic module, which serves as the system's orchestration layer. This module accesses a CSV file containing organized recipes, performing checks to ensure the identified ingredients correspond to the recipes in the system. By analyzing the detected ingredients, the system determines which recipes can be prepared. Once the recipes are determined, the module can filter the results based on the user's preferences and the recipes' ingredients.

Once this process is complete, the relevant recipe information is sent to the BERT Model (Recipe Generation Module). BERT is a Transformer-based NLP model that here serves to create sensible and context-appropriate instructions for the recipe. There is a difference though, BERT doesn't just provide a recipe; for the ingredients identified, BERT offers customized instructions that indicate the exact steps to take. BERT ensures that the instructions make sense and that a user can manipulate the recipe to suit different ingredients.

The Final Recipe Output is produced by combining the Generated Recipe from the **BERT model with structured recipe data.** The integrated output contains the listed detected ingredients, possible recipes, and instructions for detailed cooking. This is displayed to the user in an organized and interactive layout where they can prepare meals for the detected items and can follow instructions step by step. The User Interface Module takes the interaction a step further by making the layout visually appealing. The detected ingredients and recipes can be filtered from the options in the layout but also allow users to modify the recipe by changing the detected ingredients, using the filters, and making substitutions. It can also be operated using voice commands and audibly provide instructions for cooking to allow users to follow in a hands-free cooking mode. The system simplicity and voice features allow accessibility from almost any channel, whether a mobile phone or a smart kitchen assistant.

IV.WORKING PRINCIPLE

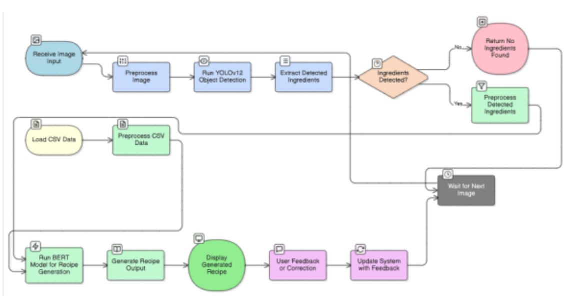

Figure 4.0 System Workflow

The process of identifying ingredients in images and formulating recipes in real-time commences by capturing an image, which is the base for identifying ingredients. Preprocessing techniques of an image for improving the visual form entails, noise reduction, resizing and normalization. This improves the image and makes it suitable for the object detection model. This stage of image preprocessing is important, because it sets the standard for all subsequent steps of recognition, and, consequently, it eliminates variables for consistent and standard recognition.

After this, the processed image is analyzed by the YOLOv12 model to detect ingredients within the image in a robust and real-time manner. YOLOv12 was selected because it is able to detect multiple ingredients in an image quickly and accurately, which is ideal in real-time scenarios. After detection, images of the ingredients continue through the pipeline. To improve reliability, a validation process ensures only meaningful ingredients are passed. Meaningful detection is when an ingredient is obvious in a food form. When detection is unsuccessful, the model returns a "No Ingredients Found" message and waits for another image.

When ingredients are detected, they are further refined to shape the information into a standardized order for the generation of a recipe. This ensures ingredients, their dimensions, and any other details are appropriately named and structured to limit mistakes in the following stages. In addition to the detection of images, there are also CSV files incorporated into the system containing additional information about the ingredients. This CSV data is refined to match the detected ingredients to enrich the recipe generation data set, empowering the system to process a wide range of conditions.

After the CSV files and the captured images have been processed, the data is sent for recipe generation to a BERT-based recipe generation model. Utilizing sophisticated natural language processing capabilities, the model assembles a recipe and maintains logic and contextual correlation to the detected ingredients. The BERT model regulates the inputs and structured recipe outputs to ensure diversity in the generated recipe to the detected ingredients, making it realistic and practical. The generated recipe is then displayed to the user for ease of cooking or meal preparations.

Successful detection of ingredients means that the information captured can be formatted systematically and standardized for recipe creation. This involves making sure that the ingredients, their quantities, and other details are captured accurately for the later stages to minimize possible errors. Besides the image-based-detection, the system uses CSV files that contain other information related to the ingredients. This information is also preprocessed so that it corresponds to the detected ingredients, thereby enhancing the datasets available for recipe generation and enabling the system to manage an extended range of input variations.

When the preprocessing for images, as well as the CSV files, is finished, the data is routed to the BERT-based model for recipe creation. This model uses cutting-edge natural language processing technology to produce coherent recipes that are contextually appropriate for the ingredients available. The BERT system guarantees recipe diversity by correlating the input ingredients to structured recipe output. The result is practical and catered to the ingredients detected. This output is provided as a recipe that can be directly used for cooking or meal preparation.

V. CONCLUSION

The Real-Time Ingredient Recognition and Recipe Generator combines cutting-edge Computer Vision with Natural Language Processing (NLP). This integration combines practical vision and fully automated cooking assistance.

Employing a sophisticated object detection algorithm like YOLOv12 for real-time ingredient identification, The system converts visual images (pictures of ingredients) into a structured list of items with unprecedented speed and accuracy. This list of recognized items directly integrates with the BERT-based Natural Language Processing (NLP) pipeline.

The BERT model, which has been fine-tuned based on the provided CSV data set, generated context-aware recipes. This level of recipe generation transcends basic keyword matching, employing BERT's sophisticated contextual understanding, producing recipes that are not only relevant and comprehensive but also practical in relation to the ingredients.

VI. REFERENCES

[1] P. Chhikara, D. Chaurasia, Y. Jiang, O. Masur, and F. Ilievski, "FIRE: Food Image to REcipe Generation," *Proc. IEEE Winter Conf. on Applications of Computer Vision (WACV)*, 2024, pp. 1234–1245.

[2] Y. Yin, H. Qi, B. Zhu, J. Chen, Y.-G. Jiang, and C.-W. Ngo, "FoodLMM: A Versatile Food Assistant using Large Multi-modal Model," *arXiv preprint arXiv:2312.14991*, 2023.

[3] K. Fu and Y. Dai, "Recognizing Multiple Ingredients in Food Images Using a Single-Ingredient Classification Model," *arXiv preprint arXiv:2401.14579*, 2024.

[4] H. Qi, B. Zhu, C.-W. Ngo, J. Chen, and E.-P. Lim, "Advancing Food Nutrition Estimation via Visual-Ingredient Feature Fusion (VIF²)," *Proc. ACM Int. Conf. on Multimedia Retrieval (ICMR)*, 2025.

[5] T. Senath et al., "Large Language Models for Ingredient Substitution in Food Recipes using Supervised Fine-tuning and DPO," *arXiv preprint arXiv:2412.04922*, 2024.

[6] J. Louro, F. Fidalgo, and Â. Oliveira, "Recognition of Food Ingredients—Dataset Analysis," *Appl. Sci.*, vol. 14, no. 5, pp. 2356, 2024.

[7] A. Mane, G. Sharma, A. Soraganvi, A. Devkar, and R. Raut, "Food Ingredient Detection Using Deep Learning," *Proc. SIPCOV 2023*, 2024, pp. 12–19.

[8] M. Banjade, R. Chhetri, and P. Sharma, "Deep Learning and Gen AI based System for Ingredient Recognition and Recipe Insight," *Int. J. Sci. Eng. Technol.*, vol. 12, no. 6, pp. 878–886, 2024.

[9] V. G. Bharane et al., "Recipe Reveal – Ingredients and Recipe Generation from Food Image," *Int. J. Adv. Comput. Eng. & Comp. Tech.*, 2025, pp. 539–550.

[10] A. Shokeen, R. Sharma, and S. Verma, "Recipe Generation from Food Images Using CNN," *Int. J. Innovative Res. in Technol.*, 2025, pp. 1–8.