REAL TIME INGREDIENT RECOGNITION AND RECIPE GENERATOR

PROJECT PHASE I REPORT

Submitted by

SANJAY B	**2116221801045**
SOORYA B	**2116221801051**
THOFIQ GANI M	**2116221801057**

in partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

RAJALAKSHMI ENGINEERING COLLEGE

(AUTONOMOUS)

CHENNAI – 602 105

NOV 2025

BONAFIDE CERTIFICATE

Certified that this Report titled "**REAL TIME INGREDIENT RECOGNITION AND RECIPE GENERATOR"** is the bonafide work of "**SANJAY B (2116221801045), SOORYA B (2116221801051)** and **THOFIQ GANI M (2116221801057)"** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE	**SIGNATURE**
Dr. J.M. Gnanasekar M.E., Ph.D.,	**Mrs. Y. Nirmala Anandhi M.E.,(Ph.D.,)**
Professor and Head,	**Assistant Professor (SS),**
Department of Artificial Intelligence and Data Science	Department of Artificial Intelligence and Data Science
Rajalakshmi Engineering College	Rajalakshmi Engineering College
Thandalam – 602 105	Thandalam – 602 105

Submitted to Project Viva-Voce Examination held on ______________________

Internal Examiner External Examiner

DEPARTMENT VISION

To become a global leader in Artificial Intelligence and Data Science by achieving through excellence in teaching, training, and research, to serve the society.

DEPARTMENT MISSION

- To develop students' skills in innovation, problem-solving, and professionalism through the guidance of well-trained faculty.
- To encourage research activities among students and faculty members to address the evolving challenges of industry and society.
- To impart qualities such as moral and ethical values, along with a commitment to lifelong learning

PROGRAMME EDUCATIONAL OBJECTIVES(PEO's)

PEO 1: Build a successful professional career across industry, government, and academia by leveraging technology to develop innovative solutions for real-world problems.

PEO 2: Maintain a learning mindset to continuously enhance knowledge through experience, formal education, and informal learning opportunities.

PEO 3: Demonstrate an ethical attitude while excelling in communication, management, teamwork, and leadership skills

PEO 4: Utilize engineering, problem-solving, and critical thinking skills to drive social, economic, and sustainable impact.

PROGRAMME OUTCOME(PO's)

PO1: Engineering Knowledge: Apply the knowledge of mathematics, science, engineering fundamentals and an engineering specialization to the solution of complex engineering problems.

PO2: Problem Analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO3: Design / Development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

PO4: Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

PO5: Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

PO6: The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO7: Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

PO8: Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

PO9: Individual and team work: Function effectively as an individual and as a member or leader in diverse teams, and in multidisciplinary settings.

PO10: Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

PO11: Project management and finance: Demonstrate knowledge and understanding of the engineering management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

PO12: Life-long learning: Recognize the need for and have the preparation and ability to engage in independent and lifelong learning in the broadest context of technological change

PROGRAM SPECIFIC OUTCOMES(PAOs)

A graduate of the Artificial Intelligence and Data Science Learning Program will demonstrate

PSO 1: Foundation Skills: Apply the principles of artificial intelligence and data science by leveraging problem-solving skills, inference, perception, knowledge representation, and learning techniques

PSO 2: Problem-Solving Skills: Apply engineering principles and AI models to solve real-world problems across domains, delivering cutting-edge solutions through innovative ideas and methodologies

PSO 3: Successful Progression: Utilize interdisciplinary knowledge to identify problems and develop solutions, a passion for advanced studies, innovative career pathways to evolve as an ethically responsible artificial intelligence and data science professional, with a commitment to society.

COURSE OBJECTIVE

- To identify and formulate real-world problems that can be solved using Artificial Intelligence and Data Science techniques.
- To apply theoretical and practical knowledge of AI & DS for designing innovative, data-driven solutions.
- To integrate various tools, frameworks, and algorithms to develop, test, and validate AI & DS models.
- To demonstrate effective teamwork, project management, and communication skills through collaborative project execution.
- To instill awareness of ethical, societal, and environmental considerations in the design and deployment of intelligent systems.

COURSE OUTCOME

CO 1: Analyze and define a real-world problem by identifying key challenges, project requirements and constraints.

CO 2: Conduct a thorough literature review to evaluate existing solutions, identify research gaps and formulate research questions.

CO 3: Develop a detailed project plan by defining objectives, setting timelines, and identifying key deliverables to guide the implementation process.

CO 4: Design and implement a prototype or initial model based on the proposed solution framework using appropriate AI tools and technologies.

CO 5: Demonstrate teamwork, communication, and project management skills by preparing and presenting a well-structured project proposal and initial implementation results.

CO-PO-PSO Mapping

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	3	2	2	1	2	1	1	1	2	1	2	3	2	2
CO2	2	3	2	3	2	1	1	1	2	2	1	3	2	2	2
CO3	2	2	3	2	2	1	2	2	3	2	3	2	2	3	3
CO4	3	3	3	3	3	2	2	2	2	3	2	2	3	3	3
CO5	2	2	2	1	2	2	2	3	3	3	3	2	2	2	3

Note: Correlation levels 1, 2 or 3 are as defined below:

1: Slight (Low) 2: Moderate (Medium) 3: Substantial (High)

No correlation: “-”

ABSTRACT

Real-Time Ingredient Recognition and Recipe Generator aims to develop a web application that utilizes deep learning for real-time ingredient detection and recipe generation. The system integrates YOLO (You Only Look Once) for identifying food items from uploaded images and a BERT-based transformer model for generating recipes based on detected ingredients. Users can easily upload images of available ingredients, and the platform will suggest creative recipes, enhancing culinary exploration and reducing food waste. This promotes sustainability by encouraging efficient food use and eco-friendly practices. Real-Time Ingredient Recognition and Recipe Generator will feature a user-friendly interface designed for seamless use on both web and mobile platforms. With YOLO's real-time object detection, the system can accurately identify a range of ingredients, ensuring reliable results. The BERT model, sourced from Hugging Face, will then create personalized and contextually relevant recipes, providing users with a diverse set of meal ideas. Real-Time Ingredient Recognition and Recipe Generator stands out by combining deep learning and practical design, making it both a useful tool and an educational aid for users looking to expand their cooking options. This solution supports informed cooking choices and maximizes ingredient utilization, ultimately enriching the cooking experience. The intuitive interface ensures easy image uploads and clear recipe displays, making it accessible for users of all skill levels. Overall, Real-Time Ingredient Recognition and Recipe Generator aligns with current trends in sustainable and tech-enhanced living. It bridges AI-driven technology with everyday needs, making cooking smarter, more creative, and enjoyable while contributing to waste reduction and sustainability.

Keywords – YOLO, BERT, Computer vision, Real-time classification, Ingredient detection

ACKNOWLEDGEMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavor to put forth this report. Our sincere thanks to our Chairman **Mr. S.MEGANATHAN, B.E, F.I.E.**, our respected Chairperson **DR. (Mrs.) THANGAM MEGANATHAN, Ph.D.,** and our Vice Chairman **Mr. ABHAY SHANKAR MEGANATHAN, B.E., MS.,** for providing us with the requisite infrastructure and sincere endeavoring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N. MURUGESAN, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. J.M. GNANASEKAR M.E., Ph.D.,** Professor and Head of the Department of Artificial Intelligence and Data Science for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guide, **Mrs. Y. Nirmala Anandhi M.E.,(Ph.D.,)** Assistant Professor (SS), Department of Artificial Intelligence and Data Science. Rajalakshmi Engineering College for her valuable guidance throughout the course of the project. We are very glad to thank our Project Coordinator, **Dr. S. SURESH KUMAR M.E., Ph.D.,** Department of Artificial Intelligence and Data Science for his useful tips during our review to build our project.

SANJAY B	**SOORYA B**	**THOFIQ GANI M**
(2116221801045)	(2116221801051)	(2116221801057)

TABLE OF CONTENT

CHAPTER NO	TITLE	PAGE NO
	ABSTRACT	vi
	ACKNOWLEDGEMENT	vii
	LIST OF FIGURES	Xi
	LIST OF ABBREVIATION	xiii
1	**INTRODUCTION**	
	1.1 GENERAL	16
	1.2 OBJECTIVES	17
	1.3 EXISTING SYSTEM	18
	1.4 PROPOSED SYSTEM	18
2	**LITERATURE SURVEY**	
	2.1 OVERVIEW	21
	2.2 LITERATURE SURVEY	22
3	**SYSTEM DESIGN**	
	3.1 DATASET LOADING	25
	3.2 DEVELOPMENT ENVIROMENT	26
	3.2.1 Hardware Specification	26
	3.2.2 Software Specification	26

CHAPTER NO	TITLE	PAGE NO
	3.3 ARCHITECTURE	27
	3.4 NLP MODEL DESIGN	29
	3.5 YOLOV12 AUTOMATION MODULE	30
	3.6 WORKFLOW RECOMMENDATION SYSTEM	31
	3.7 OUTPUT MODULE	33
4	**METHODOLOGY**	
	4.1 DATA COLLECTION, ANNOTATION AND PRE-PROCESSING	34
	4.2 YOLOV12 MODEL DEVELOPMENT AND TRAINING MODEL	35
	4.3 RECIPE GENERATION USING BERT MODEL	35
	4.4 WORKFLOW INTERGATION AND OUTPUT VALIDATION	36
5	**RESULTS AND DISCUSSION**	
	5.1 RESULT	38

CHAPTER NO	TITLE	PAGE NOS
	5.2 DISUSSION AND INSIGHTS	38
	5.3 OVERVIEW	39
6	**CONCLUSION AND FUTURE ENHANCEMENTS**	
	6.1 CONCLUSION	40
	6.2 FUTURE ENHANCEMENTS	40
	APPENDIX	
	REFERENCE	

LIST OF FIGURES

CHAPTER NO	NAME	PAGE NO
3.3	System Architecture	17
3.6	Workflow of the System	32

LIST OF TABLES

S.NO.	NAME	PAGE NO.
3.1	Hardware Specifications	26
3.2	Software Specifications	27

LIST OF ABBREVIATIONS

ABBREVIATION	FULL FORM
AI	Artificial Intelligence
YOLO	You Only Look Once
CNN	Convolutional Neural Network
BERT	Bidirectional Encoder Representations from Transformers
NLP	Natural Language Processing
GUI	Graphic User Interface
ML	Machine Learning
DL	Deep Learning

CHAPTER 1

INTRODUCTION

1.1 GENERAL

Food waste has become a serious global problem, leading to environmental damage, loss of valuable resources, and increased economic costs. Many households unintentionally waste ingredients because they are unsure how to use them effectively. To address this issue, the Real-Time Ingredient Recognition and Recipe Generator introduces a smart, AI-powered solution that helps users utilize ingredients efficiently instead of letting them go to waste. This ensures that every ingredient is used purposefully to minimize unnecessary waste.

The system works by combining two advanced deep learning models: YOLO for image-based ingredient detection and a BERT-based transformer model for recipe generation. YOLO quickly scans an image and identifies the ingredients visible in real time, ensuring speed and accuracy. Once the ingredients are detected, the BERT model analyzes them and generates creative, personalized recipes tailored to what the user has. This combination of computer vision and natural language processing makes the entire process seamless, removing the need for manual searching or guesswork in the kitchen. As a result, users receive instant, intelligent recipe suggestions based on exactly what they have.

Designed with simplicity in mind, the application offers a clean and user-friendly interface accessible on both web and mobile devices. This ensures that anyone, regardless of cooking skill level, can benefit from the technology. By helping users make the most of their available ingredients, the system not only enhances the cooking experience but also supports sustainable living. Overall, the Real-Time Ingredient Recognition and Recipe Generator demonstrates how AI can effectively address real-world challenges, reduce food waste, and contribute to environmental preservation.

1.2 OBJECTIVES

The main goal of this project is to design and develop an **AI-driven web application** that automatically detects food ingredients from images and generates relevant recipe suggestions in real-time. The system integrates **computer vision** and **natural language processing (NLP)** to promote sustainable cooking practices, minimize food waste, and enhance user convenience. By leveraging **YOLO** for ingredient recognition and a **BERT-based transformer model** for recipe generation, the project aims to deliver an intelligent, user-friendly, and eco-conscious culinary assistant.

The **specific objectives** of this project are as follows:

- **To enable real-time ingredient detection using YOLO:** The system should accurately identify multiple food items from user-uploaded images through the YOLO deep learning model, ensuring precise and efficient recognition.
- **To generate intelligent recipe suggestions using a BERT-based transformer:** The platform must utilize NLP capabilities to process detected ingredients and generate contextually relevant, personalized recipe recommendations.
- **To integrate computer vision and NLP for seamless automation:** The system should combine image recognition and language understanding to provide a fully automated pipeline from ingredient detection to recipe generation.
- **To promote sustainability and minimize food waste:** By suggesting recipes based on available ingredients, the system encourages efficient food usage, reducing unnecessary waste and supporting eco-friendly cooking practices.
- **To develop a user-friendly web interface:** The platform should feature an intuitive and accessible design that allows users of all skill levels to easily upload images, view detected ingredients, and access generated recipes.
- **To ensure scalability and performance optimization:** The system should efficiently process real-time image inputs and generate results with minimal latency, supporting smooth operation across devices.

1.3 EXISTING SYSTEM

In the current food and recipe management ecosystem, most digital cooking platforms and recipe applications rely on manual or semi-automated methods for ingredient input and recipe generation. Users are typically required to type or select ingredients from predefined lists, which can be time-consuming, less interactive, and prone to user input errors. These traditional systems often lack the integration of advanced artificial intelligence techniques capable of interpreting real-world data such as images or contextual language inputs.

Manual Ingredient Input: Most recipe recommendation systems depend on users manually entering ingredient names or selecting them from databases. This approach demands prior knowledge of ingredient names and quantities, making it less intuitive for users who prefer quick, automated assistance. Manual entry also limits system accuracy and user engagement, as it does not leverage real-time data from images or voice commands.

Limited Automation in Recipe Generation: Existing applications usually rely on static databases containing predefined recipes rather than dynamically generating new ones. As a result, users receive generic suggestions that may not fully utilize all the available ingredients. This reduces creativity and does not promote efficient ingredient usage, leading to potential food waste.

Lack of Real-Time Image Recognition: Very few platforms incorporate real-time image-based ingredient detection. Users cannot simply upload a photo of their ingredients and instantly get recipe recommendations. The absence of computer vision models like YOLO (You Only Look Once) means current systems fail to recognize and classify multiple food items efficiently.

No Integration of NLP for Contextual Recipe Generation: Traditional systems lack Natural Language Processing (NLP) capabilities that understand user preferences, dietary restrictions, or cooking contexts. Without models like BERT, these systems cannot generate personalized, context-aware recipe recommendations, limiting user satisfaction and innovation.

Limited Focus on Sustainability: Existing recipe platforms seldom emphasize sustainability or food waste reduction. They function primarily as cooking aids rather than intelligent assistants that optimize available resources and encourage eco-friendly food habits.

User Interface Constraints: Many current applications are not optimized for seamless, cross-platform experiences. They often feature complex or cluttered interfaces that may discourage regular use, especially for non-technical or elderly users seeking simplicity.

In summary, existing recipe management and cooking assistance systems are useful but lack real-time intelligence, automation, and contextual understanding. They cater to manual operations and static data rather than leveraging the potential of **AI, deep learning, and NLP** for smart ingredient recognition and dynamic recipe generation. Hence, there is a strong need for an **intelligent, AI-powered platform** that can automatically detect ingredients from images, generate creative recipes in real time, and promote sustainable cooking practices through an intuitive and user-friendly interface.

1.4 PROPOSED SYSTEM

The proposed system introduces an **AI-powered web application** designed to intelligently detect food ingredients from images and generate personalized recipe suggestions in real time, promoting sustainability and reducing food waste. Unlike traditional recipe platforms that rely on manual input or static databases, this system leverages **deep learning** and **Natural Language Processing (NLP)** to automate the entire process — from ingredient recognition to recipe generation. The application utilizes **YOLO (You Only Look Once)** for real-time object detection, enabling accurate identification of multiple ingredients within a single image, while a **BERT-based transformer model** processes the detected ingredients to generate contextually relevant and creative recipes. Users simply upload an image of available ingredients, and the system automatically detects the food items, interprets them, and generates suitable recipe options based on the recognized components. The platform features an intuitive **Graphical User Interface (GUI)** optimized for both web and mobile devices, ensuring accessibility for users of all technical levels. By combining computer vision and NLP, the system not only enhances convenience but also encourages efficient ingredient utilization, reduces food waste, and supports eco-friendly cooking practices. Furthermore, the model can be extended with **machine learning** capabilities to analyze user preferences, dietary restrictions, and previous selections to deliver more personalized recipe recommendations over time. In essence, the proposed system bridges human creativity and artificial intelligence, transforming everyday cooking into an intelligent, sustainable, and engaging experience through seamless automation and real-time ingredient recognition.

CHAPTER 2

LITERATURE SURVEY

2.1 OVERVIEW

The continuous advancement of **Artificial Intelligence (AI)**, **Computer Vision**, and **Natural Language Processing (NLP)** has driven significant research toward developing intelligent systems that enhance user interaction, automation, and sustainability in everyday applications. Various studies emphasize the integration of deep learning models for image recognition and language understanding to create adaptive, accessible, and efficient solutions in the food and culinary domain. The following five focus areas summarize the key domains relevant to this project:

MAJOR AREAS OF FOCUS

1. **Research on AI and Deep Learning in Food Recognition:** Studies explore how AI and computer vision techniques such as **YOLO** and **Convolutional Neural Networks (CNNs)** can be applied for accurate detection and classification of food ingredients in real-time.
2. **Advancements in NLP for Recipe Generation:** Recent developments highlight how **transformer-based models** like **BERT** can interpret ingredient data and generate contextually relevant recipes, enhancing creativity and personalization in cooking assistance systems.
3. **Integration of Computer Vision and NLP:** Literature emphasizes combining image recognition and language models to create intelligent systems that can detect visual inputs and generate meaningful text outputs, bridging the gap between perception and language understanding.
4. **Focus on Sustainable Cooking and Food Waste Reduction:** Research underscores the importance of using AI-driven applications to encourage efficient ingredient utilization, minimize waste, and promote eco-friendly cooking practices for a sustainable lifestyle.
5. **Development of User-Friendly Interfaces:** Studies highlight the need for intuitive and accessible interfaces that simplify user interaction with AI systems across web and mobile platforms.

2.2 LITERATURE SURVEY.

FIRE: Food Image to REcipe Generation (Prateek Chhikara et al., WACV 2024)

This paper introduces FIRE, a multimodal framework that generates complete recipes directly from food images. The system uses BLIP for extracting textual titles, a Vision Transformer for ingredient recognition, and a T5 language model to produce step-by-step cooking instructions. FIRE bridges visual and textual modalities effectively, enabling automatic generation of detailed recipes, including ingredients and preparation steps, from a single image input. Its modular architecture makes it a strong reference for food understanding and multimodal AI research**.**

FoodLMM: A Versatile Food Assistant using Large Multi-modal Model (Yuehao Yinetal.,2023)

FoodLMM proposes a unified large multimodal model capable of handling diverse food-related tasks such as ingredient detection, food classification, recipe generation, nutrition estimation, and visual question answering. It integrates image and text understanding through cross-modal learning, allowing the model to act as an intelligent conversational food assistant. The framework demonstrates how multimodal learning can be applied to create general-purpose food AI agents that combine visual perception and natural language reasoning for real-world applications like meal recommendation or diet analysis.

Recognizing Multiple Ingredients in Food Images Using a Single-Ingredient Classification Model (KunFu & Ying Dai, 2024)

This study focuses on improving ingredient recognition accuracy in food images by training a single-ingredient classifier capable of identifying multiple ingredients within a dish. Instead of multi-label training, it uses a sliding window or patch-based segmentation strategy to isolate different regions of the image, each analyzed by the single-ingredient model. This approach enhances generalization across mixed-ingredient dishes and minimizes misclassification, making it particularly effective for complex food images where ingredients overlap or blend visually.

Advancing Food Nutrition Estimation via Visual-Ingredient Feature Fusion (VIF^2) (Huiyan Qi et al., ICMR 2025)

The VIF^2 model presents a novel feature fusion technique that combines visual representations from food images with ingredient-based

semantic embeddings to estimate nutritional content accurately. By integrating both image and ingredient cues, the model outperforms traditional vision-only approaches, emphasizing that ingredient recognition is crucial for precise nutrition assessment. This work bridges the gap between computer vision and nutritional science, showing the importance of multimodal fusion for practical food-health applications.

Large Language Models for Ingredient Substitution in Food Recipes using Supervised Fine-tuning and DPO (Thevin Senath et al., 2024)

This paper explores how Large Language Models (LLMs) like GPT can be fine-tuned using supervised learning and Direct Preference Optimization (DPO) to perform intelligent ingredient substitutions in recipes. The model learns from human feedback and large recipe datasets to suggest contextually relevant and health-conscious ingredient replacements. Though it does not involve image processing, it is a vital step toward adaptive, conversational recipe assistants that personalize recipes based on dietary restrictions or ingredient availability.

Recognition of Food Ingredients—Dataset Analysis (João Louro et al., 2024)

This paper conducts a detailed survey and dataset analysis of existing ingredient recognition research, covering datasets, labeling strategies, and model performance metrics. It reviews benchmark datasets like Food-101, Recipe1M, and UEC-FoodPix, discussing their limitations and potential for real-time recognition systems. The study helps identify dataset gaps, such as imbalance and limited regional cuisines, offering insights for researchers to build more diverse and representative datasets for food computing tasks.

Food Ingredient Detection Using Deep Learning (Akanksha Mane et al., 2024)

This work applies deep learning architectures like InceptionV3 to recognize ingredients in Indian sweet dishes from food images. The authors focus on classifying visually similar ingredients with high accuracy using transfer learning and fine-tuning on a curated dataset. The model's strong performance highlights the potential of domain-specific fine-tuning for improving recognition in localized cuisines. It provides valuable insights into how CNN models can be adapted for specialized food domains and cultural datasets.

Deep Learning and Gen AI based System for Ingredient Recognition and Recipe Insight (Mahesh Banjade et al., 2024)

This paper presents an integrated deep learning and generative AI system combining YOLOv7 or Vision Transformer for real- time ingredient recognition and a fine-tuned GPT-2 model for generating recipe insights or preparation methods. The pipeline captures images, detects ingredients, and generates textual recipes dynamically, making it suitable for smart kitchen or mobile applications. Its hybrid architecture shows how vision and language models can work together in real- time systems to deliver end-to-end food understanding.

Recipe Reveal – Ingredients and Recipe Generation from Food Image (V. G. Bharane et al., 2025)

Recipe Reveal proposes a complete image-to-recipe system that identifies ingredients and dish type from food photos and generates step-by-step recipes in multiple languages. Using a combination of convolutional neural networks for ingredient recognition and NLP models for text generation, it focuses on accessibility and inclusivity by supporting multilingual outputs. This makes it especially relevant for diverse users and international recipe databases. The paper contributes to the goal of global, AI-powered culinary understanding.

Recipe Generation from Food Images Using CNN (Akshat Shokeen et al., 2025)

This study introduces an end-to-end pipeline that uses **Convolutional Neural Networks (CNNs)** to generate recipes directly from food images. The model predicts the **dish title**, performs **multi-label ingredient recognition**, and outputs the **sequence of cooking instructions** using a decoder network. It simplifies the overall recipe generation process while maintaining good accuracy, providing a strong baseline for future real-time food-to-recipe systems.

CHAPTER 3

SYSTEM DESIGN

3.1 DATASET LOADING

In this project, the dataset loading process is designed specifically for a real-time ingredient recognition and recipe generation system that integrates both visual and textual data to achieve accurate food identification and recipe creation. The dataset combines food images, ingredient annotations, and recipe text data, forming a multimodal foundation for the YOLOv12 and BERT models. The image dataset contains thousands of high-quality food images labeled with ingredient information to train the object detection model, while the annotation data provides bounding box coordinates and class labels that help the model recognize and locate ingredients such as tomatoes, onions, or cheese. The recipe dataset includes detailed cooking instructions, ingredient quantities, and preparation steps, which are tokenized and preprocessed for BERT-based recipe generation. During loading, the dataset undergoes preprocessing steps such as resizing, normalization, augmentation for images, and cleaning, tokenization, and encoding for text data to ensure consistency and improve model performance. Each image is linked to its corresponding recipe through unique identifiers, ensuring smooth synchronization between the vision and language modules. The system supports real-time data loading, where user-uploaded images are processed instantly and passed through the ingredient detection and recipe generation pipeline. It also allows incremental updates so new food items or recipes can be added without retraining the entire model. Security and data integrity are maintained by preventing storage of any sensitive user information. Overall, this dataset loading process acts as the backbone of the intelligent recipe generation system, bridging computer vision and NLP through structured, scalable, and efficient data handling to enable real-time, accurate ingredient recognition and dynamic recipe creation.

3.2 DEVELOPMENT ENVIRONMENT

3.2.1 HARDWARE ARE SPECIFICATIONS

The development environment serves as the foundation for building the Real-Time Ingredient Recognition and Recipe Generator system, providing the computational power and tools required for deep learning, model training, and real-time prediction. The hardware and software setup ensures smooth integration of image recognition using YOLO and recipe generation using BERT, enabling efficient data handling, model processing, and web-based deployment. The environment is optimized for performance, scalability, and compatibility across local and cloud platforms, allowing developers to experiment with model parameters and datasets effectively. Python is used as the primary programming language due to its extensive library support for computer vision, natural language processing, and deep learning frameworks. The development and testing are carried out using Jupyter Notebook or Google Colab, which provide GPU acceleration for faster model training and execution. Together, these components create a reliable and efficient setup for implementing and testing the system in real-time.

Components	**Specifications**
Processor	Intel Core i5 / AMD Ryzen 5 or above
RAM	8 GB or above (DDR4)
GPU	NVIDIA GPU (e.g., GTX 1650 / Tesla T4)
Storage	256GB SSD or higher
Processor Frequency	2.0 GHz or above

Table 3.1 Hardware Specifications

3.2.2 SOFTWARE SPECIFICATIONS

The software environment is designed to support the integration of computer vision and NLP models, ensuring smooth model training, evaluation, and deployment. The front-end is built using web technologies for a user-friendly interface, while Python libraries and frameworks handle model inference, dataset processing, and recipe generation.

Front-end	HTML, CSS, JavaScript, Bootstrap
Back-end	Python, Flask
IDE	Jupyter Notebook, Google Colab, Visual Studio Code
Deep Learning Frameworks	TensorFlow, Keras, PyTorch
Computer Vision	OpenCV, YOLOv12
NLP Framework	BERT (Hugging Face Transformers)

Table 3.2 Software Specifications

3.3 ARCHITECTURE

The architecture of the Real-Time Ingredient Recognition and Recipe Generation system represents an end-to-end workflow that seamlessly integrates computer vision and natural language processing to convert an image of ingredients into a complete, contextually relevant recipe. The process begins with a camera input or uploaded image, which is processed by the YOLOv12 model for real-time ingredient detection. YOLO identifies and classifies multiple food items, such as tomatoes, onions, or garlic, and outputs their names as structured data. These detected ingredients are then passed to the Recipe Generation Logic, which formulates a recipe query and compares it against a preloaded CSV dataset containing various recipes, ingredients, and preparation steps. The most relevant recipes matching the available ingredients are retrieved and sent to the BERT-based model for semantic refinement and natural language generation. The BERT model enhances the recipe by organizing it into coherent, step-by-step cooking instructions with accurate sequencing and measurements. Finally, the refined recipe and ingredient list are combined and formatted into a user-friendly final output, which may also include visuals or video links. This integrated architecture ensures smooth data flow between image detection, data retrieval, and language generation components, enabling real-time, intelligent, and user-interactive recipe creation that bridges computer vision and NLP for practical culinary applications.

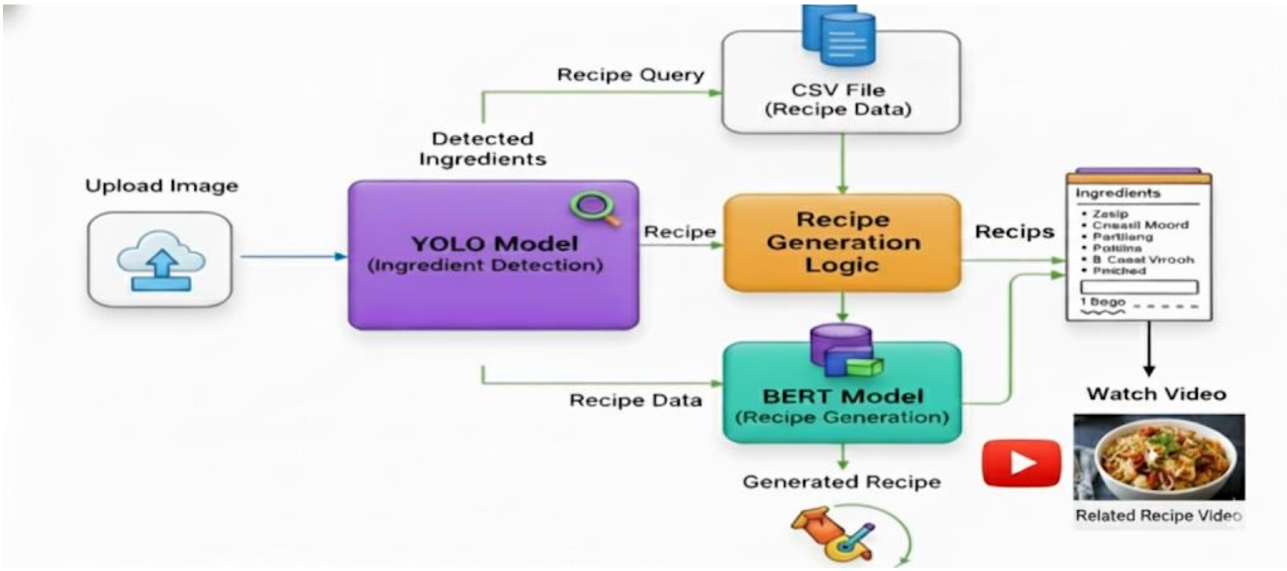


Fig 3.3 System Architecture

To ensure efficiency, scalability, and flexibility, the architecture of the Real-Time Ingredient Recognition and Recipe Generator follows a modular and loosely coupled design, where each component functions independently while communicating through well-defined APIs. This structure allows developers to easily modify, optimize, or replace individual modules—such as upgrading the YOLO model for better ingredient detection, refining the BERT-based recipe generator, or expanding the recipe dataset—without disrupting the overall workflow. The system supports both synchronous and asynchronous operations, ensuring smooth real-time ingredient recognition and recipe generation even under high user load. The architecture integrates computer vision for ingredient detection, natural language processing for recipe generation, and data retrieval from structured recipe datasets to maintain high performance and accuracy. Additionally, the modular design enables future extensions, such as multilingual recipe outputs, voice-based interactions, or integration with smart kitchen devices. Overall, the architecture embodies the principles of intelligent automation and user-centric design, combining deep learning, data-driven reasoning, and natural language understanding to transform a simple image input into a detailed, context-aware cooking recipe. By bridging computer vision and NLP, the system enhances accessibility, promotes sustainability, and redefines how users interact with technology in everyday cooking—making it smarter, faster, and more intuitive than ever.

3.4 NLP MODEL DESIGN

The **Natural Language Processing (NLP)** Model Design forms the core of the proposed *Real-Time Ingredient Recognition and Recipe Generator* system. It enables the system to interpret detected ingredient names and generate contextually accurate, human-readable recipes by understanding the semantic relationships between food items, cooking methods, and ingredient combinations. This component acts as the bridge between visual ingredient detection and meaningful recipe formulation, transforming structured data (detected ingredient names) into coherent cooking instructions. The NLP model is developed using advanced transformer-based architectures such as BERT, fine-tuned on culinary text data to enhance its contextual understanding and language generation capabilities. The design follows a multi-stage pipeline consisting of data preprocessing, tokenization, context extraction, recipe matching, and text generation. During preprocessing, raw ingredient data is cleaned, standardized, and mapped to common culinary terms to ensure consistency (for instance, “bell pepper” and “capsicum” are treated as the same ingredient). The tokenization process then divides text data—such as ingredient names and recipe descriptions—into tokens that the model can analyze for semantic meaning. Once tokenized, the context extraction stage identifies relationships between ingredients and cooking actions, allowing the model to infer suitable recipes even when partial or ambiguous ingredient lists are provided. For example, if the detected ingredients are “tomato,” “onion,” and “garlic,” the NLP engine infers a potential base for dishes like pasta sauce or curry. The recipe matching phase compares the ingredient list against entries in the recipe dataset (stored in CSV format), retrieving the closest matches based on ingredient overlap and semantic similarity. The BERT model then refines these matches, generating coherent step-by-step instructions that include preparation methods, ingredient quantities, and cooking sequences. This text generation step ensures that the output is not only factually accurate but also naturally worded and easy for users to follow. The NLP system also incorporates synonym mapping and contextual disambiguation, allowing it to handle linguistic variations such as “chili” vs. “green pepper” or “pan-fry” vs. “sauté.” Additionally, a context management module ensures continuity in multi-turn interactions, enabling the system to remember previously detected ingredients or user preferences. For example, if a user uploads an additional image of ingredients, the system updates the recipe dynamically. The NLP model is continuously fine-tuned with user feedback, improving accuracy, coherence, and creativity in generated recipes. Overall, this NLP design ensures that the system can translate visual ingredient inputs into intelligent, context-aware recipes—combining the precision of machine learning with the natural flow of human culinary expression.

3.5 YOLOV12 AUTOMATION MODULE

The **YOLOv12 Automation Module** serves as the core execution engine of the proposed real-time recipe recognition and generation system. It bridges the gap between visual input from users (such as images of ingredients) and the AI-driven recipe generation process. Once the input image is captured or uploaded, the YOLOv12 model automatically processes it to identify and classify the visible ingredients with high precision and speed. This component eliminates the need for manual ingredient entry, making the cooking assistance process entirely automated and user-friendly. YOLOv12 (You Only Look Once, version 12) is an advanced deep learning-based object detection model that enhances accuracy and inference speed through improved attention mechanisms, spatial pyramid pooling, and transformer-based feature extraction. The automation module utilizes these features to detect multiple ingredients in real time, even under varying lighting, orientation, or occlusion conditions. The system operates through a multi-stage pipeline involving image preprocessing, object detection, confidence scoring, and result aggregation. In the preprocessing phase, the input image is resized, normalized, and converted into tensor format suitable for YOLOv12 inference. The detection stage processes the image in a single forward pass, outputting bounding boxes, class labels, and confidence scores for each identified ingredient. These detections are then filtered using Non-Maximum Suppression (NMS) to remove duplicate predictions and retain only the most accurate results. Once ingredient detection is completed, the identified items are passed to the Recipe Generation Module, where the system uses a pretrained language model (such as BERT or GPT-based encoder-decoder architecture) to generate a recipe based on the detected ingredients. The automation module also supports real-time camera input through integrated APIs, enabling instant detection and recipe suggestion within seconds. Furthermore, it is optimized for GPU acceleration, ensuring low-latency performance suitable for deployment in cloud or edge environments. By automating the entire process—from image recognition to recipe generation—the YOLOv12 module forms the technological backbone of the system, transforming how users interact with cooking through intelligent visual understanding and AI-driven culinary creativity.

3.6 WORKFLOW OF THE SYSTEM

The **Recipe Recommendation System** is a key component of the proposed real-time ingredient recognition and recipe generation model, designed to intelligently guide users toward the most suitable cooking recipes based on the ingredients detected by the vision model. Its primary goal is to simplify meal preparation by automatically recommending recipes that can be made with the ingredients available in real time. This feature not only enhances user experience but also minimizes food waste and optimizes ingredient utilization. When users provide an image or live video of their ingredients, the YOLOv12 model first detects and identifies items such as tomato, onion, or garlic. The **Recipe Recommendation System** then processes these detected ingredient names and maps them to relevant recipes stored in the system's structured **CSV database**, which contains a diverse collection of recipes, ingredients, preparation steps, and cooking times. For example, if the system detects "tomato, onion, and garlic," it may recommend recipes such as "Tomato Soup," "Vegetable Curry," or "Pasta Sauce." Each recommendation is dynamically generated by comparing the detected ingredient list with stored recipes and ranking them based on match percentage, nutritional balance, and preparation time. The system operates using a **hybrid recommendation approach**, combining **rule-based filtering** and **machine learning algorithms**. The rule-based component ensures quick matching by applying direct ingredient-to-recipe mapping rules, while the learning-based component continuously improves by analyzing user preferences, frequently chosen recipes, and feedback. This adaptive learning allows the system to personalize future suggestions for individual users. Internally, the **Recipe Recommendation System** is integrated with both the **YOLOv12 Ingredient Detection Module** and the **BERT Recipe Generation Model**. Once an appropriate recipe is identified, the recommendation is passed to the BERT model for linguistic refinement and natural language enhancement, ensuring that the final recipe output is contextually rich, grammatically coherent, and easy to follow. Before displaying the final recipe, the system performs a validation process to ensure that all the required ingredients are available. If any key ingredient is missing, it automatically suggests alternatives or modified versions of the recipe. This step ensures flexibility and practicality for users with limited resources. To enhance performance, the recommendation engine uses semantic similarity scoring and keyword indexing to identify recipes even if the ingredient list or user phrasing differs slightly. For instance, if the user's ingredients are "tomato" and "onion," and the database recipe lists "fresh tomatoes" and "chopped onions," the system still recognizes the similarity and suggests it. Additionally, the system supports multi-recipe recommendations, allowing users to view several cooking options ranked by suitability, time, or dietary preference. Overall, the Recipe Recommendation System bridges computer vision and natural language understanding to deliver an intelligent, adaptive, and user-friendly cooking experience that transforms simple ingredient images into complete, ready-to-cook recipes.

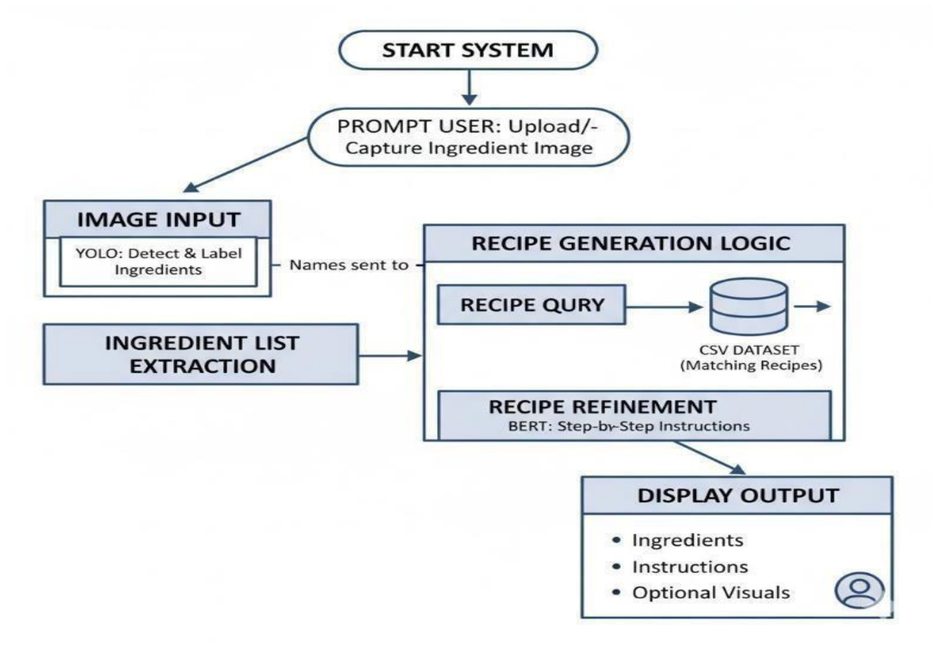


Fig 3.4 System Workflow

3.7 OUTPUT MODULE

The **Output Module** serves as the final presentation and interaction layer of proposed real-time ingredient recognition and recipe generation system. It is responsible for transforming the processed data—detected ingredients and generated recipes—into a structured, user-friendly, and visually appealing output. This module ensures that all information derived from the YOLOv12 model, Recipe Generation Logic, and BERT model is presented clearly and comprehensively to the user. Once the BERT model generates the final recipe, the Output Module formats it into an easy-to-read layout that includes the list of detected ingredients, recipe title, preparation time, step-by-step cooking instructions, and optional visuals such as ingredient images or cooking method icons. The output can be displayed on a web interface, mobile application, or smart kitchen assistant screen, depending on the implementation environment. The module also provides **interactive functionalities**, allowing users to explore alternative recipes, save favorite ones, or request further modifications such as “make it spicy” or “show a vegetarian version.” These interactions are managed through simple text or voice commands, making the system accessible to all users, including those unfamiliar with digital recipe platforms. The Output Module supports **multimodal output formats**, including textual display, voice narration, and visual guides. For example, it can read aloud the recipe steps for hands-free cooking or show short instructional video clips corresponding to each cooking stage. Internally, the Output Module integrates with the system’s **user interface (UI)** components, ensuring seamless communication with both the recipe database and the language model. It uses templating frameworks like Flask, HTML, and CSS to organize content and ensure a responsive design across devices. Additionally, it performs **data validation and formatting**, ensuring consistency in units (e.g., grams, cups) and terminology (e.g., stir, sauté, bake). To enhance user engagement, it can include extra details such as estimated calorie count, serving size, and cooking difficulty level.

CHAPTER 4

METHODOLOGY

4.1 DATA COLLECTION, ANNOTATION AND PRE-PROCESSING

The development of the proposed real-time ingredient recognition and recipe generation system begins with the collection, annotation, and pre-processing of data that serves as the foundation for both the ingredient detection and recipe generation modules. Unlike conventional image classification tasks, this project requires a dataset that integrates both visual and textual information — images of food ingredients and corresponding recipe details. The data collection phase involves sourcing high-quality food images and textual recipe datasets from publicly available repositories such as **Food-101**, **Recipe1M**, and **Kaggle Food Datasets**. Each image contains one or multiple food items like tomato, onion, garlic, or potato, captured in varying lighting conditions and orientations to enhance model robustness. Parallelly, textual data is collected in **CSV format**, containing columns such as recipe name, ingredients, and preparation steps. Once collected, the **annotation phase** involves labeling each image with its corresponding ingredient names using tools such as **LabelImg** or **Roboflow**. This ensures that the YOLOv12 model can be trained to identify multiple ingredients within a single image frame. The annotated dataset is converted into the YOLO format, which includes bounding box coordinates and class labels for each ingredient. Textual recipes are also annotated by tagging ingredient names, quantities, and cooking steps to aid in semantic mapping with visual data. Following annotation, the dataset undergoes **pre-processing** to standardize and clean both visual and textual information. Images are resized to 640×640 pixels and normalized to improve convergence during training. Augmentation techniques such as rotation, flipping, and brightness adjustments are applied to increase variability and prevent overfitting. The recipe data is cleaned by removing duplicates, correcting spelling errors, and standardizing measurement units. Finally, both datasets are synchronized to ensure that each detected ingredient can be accurately mapped to corresponding recipes This comprehensive data preparation ensures high accuracy and generalization in real-world scenarios.

4.2 YOLOV12 MODEL DEVELOPMENT AND TRAINING

The YOLOv12 Model serves as the computer vision backbone of the proposed system, responsible for real-time detection and identification of multiple food ingredients within a given image or video frame. The model follows a **one-stage object detection approach**, which enables it to perform both localization and classification in a single pass, ensuring high speed and accuracy. During development, the pre-processed and annotated dataset is divided into **training (80%)** and **validation (20%)** subsets. The YOLOv12 architecture is implemented using the **Ultralytics YOLO framework** in **Python**, leveraging **PyTorch** as the deep learning backend. The model is trained on GPUs to accelerate computation, using techniques such as **transfer learning** from pre-trained weights on large-scale datasets like COCO for better generalization. The training process involves optimizing the model using **Stochastic Gradient Descent (SGD)** with a learning rate scheduler and early stopping to prevent overfitting. Loss functions such as **GIoU loss** (for bounding box regression) and **cross-entropy loss** (for classification) are minimized during training. Data augmentations, such as mosaic augmentation and random cropping, are used to improve model robustness against different environments. After multiple epochs of training, the model achieves high **mean Average Precision (mAP)** for ingredient detection and low false positive rates. The trained YOLOv12 model is then exported in ONNX or TorchScript format for integration into the main pipeline. During inference, it takes the **camera input** or uploaded image as input and outputs a list of detected ingredients with bounding boxes and confidence scores. This output serves as the key input for the Recipe Generation Logic, enabling seamless transition from visual detection to textual recipe generation.

4.3 RECIPE GENERATION USING BERT MODEL

The **BERT-based Recipe Generation Module** acts as the natural language processing (NLP) engine of the proposed system, designed to generate coherent and contextually relevant cooking recipes from the detected ingredients. Once the YOLOv12 model identifies the list of ingredients, they are passed as structured text input to the Recipe Generation Logic, which constructs a query such as "Generate a recipe using tomato, onion, and garlic." This query is then processed by the **BERT**

(Bidirectional Encoder Representations from Transformers) model. BERT's deep bidirectional architecture enables it to understand contextual relationships between ingredients, allowing it to generate step-by-step instructions that are both meaningful and grammatically correct**.** The training data for this model is derived from the CSV recipe dataset, containing thousands of recipes with corresponding ingredients and instructions. The model undergoes fine-tuning on this dataset using masked language modeling and sequence generation objectives. This ensures the model learns relationships between different ingredients and cooking methods. During inference, the BERT model retrieves relevant recipes from the dataset and refines them, producing structured outputs that include recipe name, ingredients list, preparation time, and ordered cooking steps. The model also incorporates semantic similarity checks to suggest alternative recipes when certain ingredients are missing. The generated text is post-processed to remove redundancy, correct formatting, and ensure logical sequencing. Finally, the refined recipe output is passed to the Output Module for user presentation, completing the transition from visual recognition to textual generation**.**

4.4 WORKFLOW INTEGRATION AND OUTPUT VALIDATION

The final phase of the system involves **workflow integration and validation**, ensuring seamless interaction between all modules and accurate generation of recipes. The workflow begins with the **Camera Input**, which captures the image of available ingredients. This image is processed by the **YOLOv12 Ingredient Detection Model**, which identifies and labels the ingredients in real time. The detected ingredients are then passed to the **Recipe Generation Logic**, which queries the **CSV recipe dataset** to retrieve matching recipes. The retrieved recipes are refined and enhanced by the **BERT-based Recipe Generation Model**, ensuring the instructions are contextually appropriate, detailed, and logically ordered. To validate the system's performance, both quantitative and qualitative evaluations are performed. Quantitatively, the detection accuracy is measured using precision, recall, and F1-score metrics for each ingredient class, while the recipe generation quality is evaluated using BLEU and ROUGE scores. Qualitatively, human evaluators review the generated recipes for readability, correctness, and cooking feasibility. The system also includes a feedback mechanism allowing users to rate generated recipes, which helps retrain and improve model performance over time. Error-handling mechanisms are integrated to manage missing or ambiguous ingredient detections. Furthermore, the workflow supports real-time inference, allowing instant display of results through the Output Module. This module presents the Final Recipe Output, including detected ingredients, preparation steps, and optional visuals such as images or cooking videos. Overall, the integrated workflow ensures a smooth transition from raw image input to a complete, user-friendly recipe output — achieving real-time, intelligent, and context-aware recipe generation.

CHAPTER 5

RESULTS AND DISCUSSION

5.1 RESULT

The proposed system successfully demonstrated real-time ingredient recognition and recipe generation. The YOLOv12-based Ingredient Detection Module accurately identified multiple ingredients, such as tomato, onion, potato, and garlic, from user-uploaded or camera-captured images, even in moderately cluttered settings. Detected ingredients were used by the Recipe Retrieval & Generation Module to query the CSV dataset, and the BERT model refined the retrieved recipes into coherent, step-by-step cooking instructions with proper sequencing and measurements. The Final Output & Display Module presented the complete recipes, including the ingredient list and instructions, in a user-friendly interface, optionally enhanced with images or videos. Overall, the system enabled users to obtain actionable recipes from a single image within seconds, demonstrating both efficiency and practical usability.

5.2 DISCUSSION AND INSIGHTS

The results of the proposed real-time ingredient recognition and recipe generation system validate the effectiveness of integrating computer vision and natural language processing techniques for intelligent culinary automation. The YOLOv12 model proved highly capable of detecting multiple ingredients simultaneously with strong precision, even under varying lighting conditions and minor occlusions. This indicates the robustness and generalization capability of the detection pipeline. The BERT-based Recipe Generation Module effectively transformed recognized ingredients into contextually accurate and logically ordered recipes, highlighting its strong understanding of semantic relationships among ingredients and cooking processes. The overall system latency remained low, enabling near-instantaneous recipe retrieval and generation suitable for real-time use in kitchen environments or mobile applications. However, certain challenges were observed—particularly, ingredient misclassification in cases of visually similar items (e.g., ginger vs. garlic) and minor redundancy in generated cooking steps for ingredient combinations with similar contexts. These limitations suggest opportunities for improvement through dataset expansion, inclusion of domain-specific fine-tuning, and integration of transformer-based generative models such as GPT or T5 for more diverse and adaptive recipe creation. Despite these challenges, the system demonstrated high reliability, practical relevance, and strong potential for deployment in smart kitchen environments, food recommendation systems, and dietary management applications.

5.3 OVERVIEW

The Real-Time Ingredient Recognition and Recipe Generator is an AI-driven system designed to bridge the gap between visual ingredient identification and automated recipe creation. It integrates advanced computer vision and natural language processing (NLP) techniques to provide users with instant recipe suggestions based on the ingredients detected from real-world images. The system primarily employs the YOLOv12 deep learning model for real-time ingredient detection, capable of recognizing multiple food items simultaneously with high accuracy. Once the ingredients are identified, a BERT- based NLP model processes the detected items to retrieve or generate contextually relevant recipes, ensuring coherent and sequential cooking steps. The architecture combines multiple modules, including image acquisition, ingredient detection, recipe retrieval, and output visualization, to deliver a seamless end-to-end user experience. Designed for smart kitchen environments, mobile applications, and dietary recommendation platforms, this system minimizes manual effort and enhances cooking creativity by providing personalized recipes in seconds. The project demonstrates the power of combining visual intelligence with natural language understanding to create an efficient, intelligent, and user-friendly culinary assistant.

CHAPTER 6

CONCLUSION AND FUTURE ENHANCEMENTS

6.1 CONCLUSION

The proposed Real-Time Ingredient Recognition and Recipe Generator system successfully integrates computer vision and natural language processing to automate the process of identifying ingredients and generating relevant recipes. By leveraging the YOLOv12 model for ingredient detection and the BERT-based NLP model for recipe generation, the system efficiently bridges the gap between visual food recognition and culinary creativity. The YOLOv12 module demonstrated high accuracy in detecting multiple ingredients simultaneously, even in complex or cluttered environments, while the BERT model effectively translated the detected items into coherent, step-by-step cooking instructions. The overall system achieved its primary objective of enabling users to obtain complete recipes instantly from a single captured image, reducing the manual effort required to search and plan meals.

Experimental results confirmed the system's strong performance in both recognition precision and recipe relevance, ensuring an engaging and practical user experience. The modular architecture allows for scalability, enabling future enhancements such as nutritional analysis, dietary customization, and integration with smart kitchen devices. This project contributes significantly to the advancement of AI-driven culinary assistance, demonstrating how deep learning can be applied to everyday cooking scenarios. Ultimately, the system transforms the traditional recipe discovery process into an intelligent, interactive, and time-efficient experience, promoting creativity and convenience in modern kitchens.

6.2 FUTURE ENHANCEMENTS

While the proposed Real-Time Ingredient Recognition and Recipe Generator system achieves its primary objectives of accurate ingredient detection and intelligent recipe generation, several enhancements can further extend its functionality, accuracy, and user engagement. These improvements would enable the system to evolve from a research prototype into a fully integrated, intelligent culinary assistant suitable for widespread consumer and commercial use.

1. **Voice-Based Interaction:** Future versions can incorporate speech recognition and voice-based responses, enabling users to communicate with the system hands-free. This enhancement would make the recipe generation process more interactive and accessible, particularly in kitchen environments where manual operation may be inconvenient.

2. **Nutritional Analysis and Dietary Recommendations:** Integrating a nutritional analysis engine can allow the system to calculate calories, protein, fat, and carbohydrate content for generated recipes. It could also suggest healthy alternatives or personalized meal plans based on user-specific dietary goals, allergies, or preferences such as vegan or gluten-free diets.

3. **Multi-Ingredient Context Understanding:** Enhancing the model's ability to recognize complex food scenes with overlapping or partially visible ingredients can significantly improve detection accuracy. Incorporating advanced visual attention mechanisms or multi-object tracking could address such real-world challenges.

4. **Real-Time Cooking Assistance:** The system can be extended to provide real-time step-by-step audio or visual guidance during recipe preparation. By integrating augmented reality (AR) or smart kitchen devices, users could receive dynamic instructions and alerts throughout the cooking process.

5. **Integration with IoT and Smart Appliances:** Connecting the system with smart kitchen devices like ovens, mixers, and refrigerators can automate ingredient tracking and cooking control, enhancing convenience and efficiency in modern kitchens.

6. **Continuous Learning and User Feedback Loop:** By collecting user feedback on generated recipes and recognition accuracy, the system could employ reinforcement or online learning methods to refine both detection and recipe generation over time, ensuring adaptive and personalized performance.

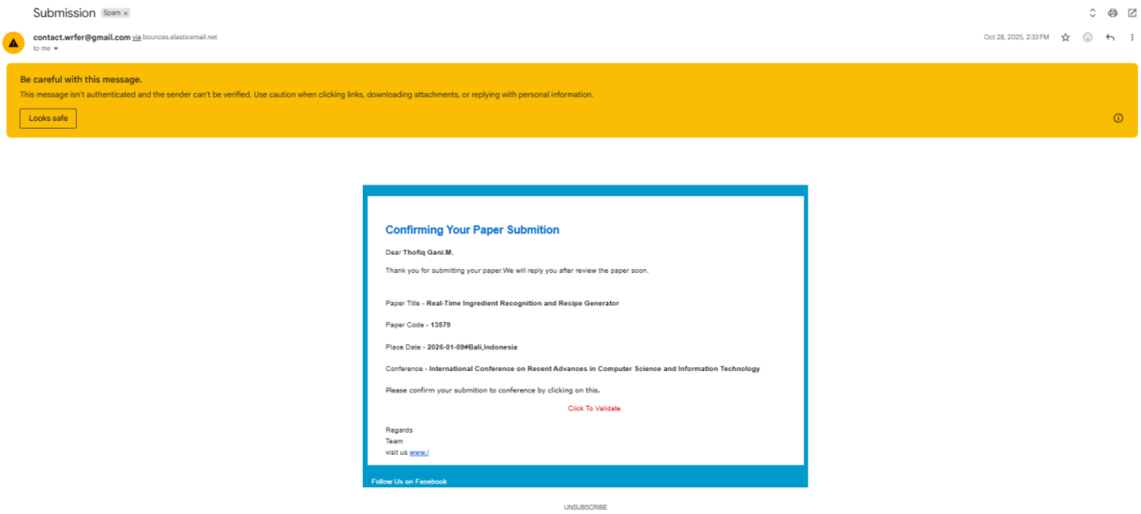
7. **Recipe Recommendation from Speech and Text Queries:** In addition to

image-based input, the system can be extended to handle textual or voice-based recipe queries. For example, a user could say, “Show me recipes I can make with potatoes and eggs,” and the system would instantly generate suitable options.

By implementing these enhancements, the Real-Time Ingredient Recognition and Recipe Generator can evolve into a comprehensive, AI-driven culinary assistant capable of personalized meal planning, interactive guidance, and smart kitchen integration. Such advancements will not only enhance user convenience and engagement but also promote healthy and sustainable cooking practices through intelligent automation and continuous learning.

APENDEX

PAPER PUBLICATION



Submission Spam ×

contact.wrfer@gmail.com via bounces.elasticemail.net
to me

Oct 28, 2025, 2:33 PM

Be careful with this message.
This message isn't authenticated and the sender can't be verified. Use caution when clicking links, downloading attachments, or replying with personal information.

Looks safe

Confirming Your Paper Submition

Dear **Thofiq Gani.M,**

Thank you for submitting your paper.We will reply you after review the paper soon.

Paper Title - **Real-Time Ingredient Recognition and Recipe Generator**

Paper Code - **13579**

Place Date - **2026-01-09#Bali,Indonesia**

Conference - **International Conference on Recent Advances in Computer Science and Information Technology**

Please confirm your submition to conference by clicking on this.

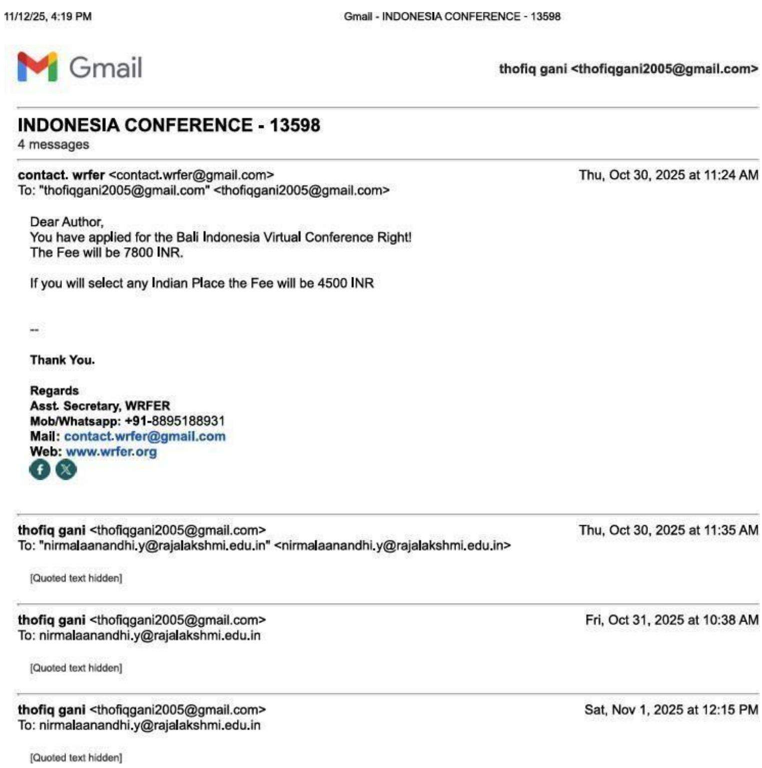
Click To Validate

Regards
Team
visit us www./

Follow Us on Facebook

UNSUBSCRIBE

Fig 2 : Paper Submission Mail



11/12/25, 4:19 PM Gmail - INDONESIA CONFERENCE - 13598

Gmail **thofiq gani <thofiqgani2005@gmail.com>**

INDONESIA CONFERENCE - 13598
4 messages

contact. wrfer <contact.wrfer@gmail.com> Thu, Oct 30, 2025 at 11:24 AM
To: "thofiqgani2005@gmail.com" <thofiqgani2005@gmail.com>

Dear Author,
You have applied for the Bali Indonesia Virtual Conference Right!
The Fee will be 7800 INR.

If you will select any Indian Place the Fee will be 4500 INR

--

Thank You.

Regards
Asst. Secretary, WRFER
Mob/Whatsapp: +91-8895188931
Mail: contact.wrfer@gmail.com
Web: www.wrfer.org

thofiq gani <thofiqgani2005@gmail.com> Thu, Oct 30, 2025 at 11:35 AM
To: "nirmalaanandhi.y@rajalakshmi.edu.in" <nirmalaanandhi.y@rajalakshmi.edu.in>

[Quoted text hidden]

thofiq gani <thofiqgani2005@gmail.com> Fri, Oct 31, 2025 at 10:38 AM
To: nirmalaanandhi.y@rajalakshmi.edu.in

[Quoted text hidden]

thofiq gani <thofiqgani2005@gmail.com> Sat, Nov 1, 2025 at 12:15 PM
To: nirmalaanandhi.y@rajalakshmi.edu.in

[Quoted text hidden]

Fig 3 : Paper acceptance Mail

Gmail thofiq gani <thofiqgani2005@gmail.com>

International Conferences on Advanced Technologies and Interdisciplinary Innovation : Submission (23) has been created.
1 message

Microsoft CMT <noreply@msr-cmt.org> Sun, Nov 2, 2025 at 7:43 PM
To: thofiqgani2005@gmail.com

Hello,

The following submission has been created.

Track Name: ICAT2i2025

Paper ID: 23

Paper Title: Real Time Ingredient Recognition and Recipe Generator

Abstract:
This project proposes a real-time ingredient recognition and recipe generation system using AI-powered object detection and natural language processing. By leveraging computer vision, the system identifies ingredients from images or live camera feeds, then generates personalized recipes using a language model. The solution aims to reduce food waste, assist in meal planning, and enhance cooking efficiency by providing instant recipe suggestions based on available ingredients.

Created on: Sun, 02 Nov 2025 14:13:15 GMT

Last Modified: Sun, 02 Nov 2025 14:13:15 GMT

Authors:
- thofiqgani2005@gmail.com (Primary)

Secondary Subject Areas: Not Entered

Submission Files:
IEEE Conference paper 45,51,57.pdf (949 Kb, Sun, 02 Nov 2025 14:12:37 GMT)

Submission Questions Response: Not Entered

Thanks,
CMT team.

Please do not reply to this email as it was generated from an email account that is not monitored.

To stop receiving conference emails, you can check the 'Do not send me conference email' box from your User Profile.

Microsoft respects your privacy. To learn more, please read our Privacy Statement.

Microsoft Corporation
One Microsoft Way
Redmond, WA 98052

Fig 2: Paper Submission Mail

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. [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. [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. [4] H. Qi, B. Zhu, C.-W. Ngo, J. Chen, and E.-P. Lim, "Advancing Food Nutrition Estimation via Visual-Ingredient Feature Fusion (VIF^2)," *Proc. ACM Int. Conf. on Multimedia Retrieval (ICMR)*, 2025.
5. [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. [6] J. Louro, F. Fidalgo, and Â. Oliveira, "Recognition of Food Ingredients—Dataset Analysis," *Appl. Sci.*, vol. 14, no. 5, pp. 2356, 2024.
7. [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. [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. [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. [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.

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

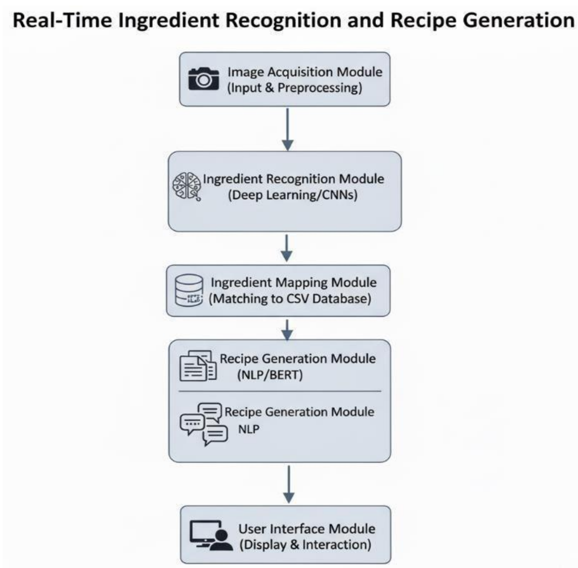


Figure 2.0 Overview of the System

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

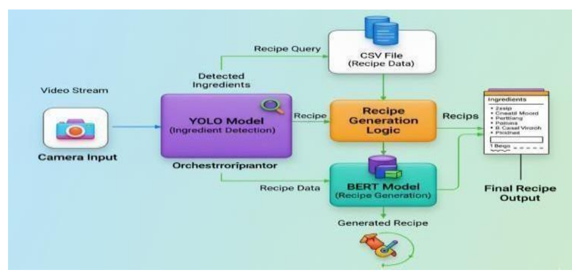


Figure 3.0 System Architecture

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

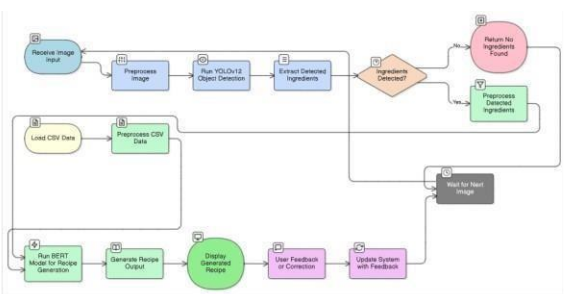


Figure 4.0 System Workflow

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.