Smart Fridge: Recipe Generator Using Object Detection Model, OpenCV, LLMs

BY

HARI HARAN R.B

UNDER THE GUIDANCE OF

Dr. ANURADHA J

ABSTRACT

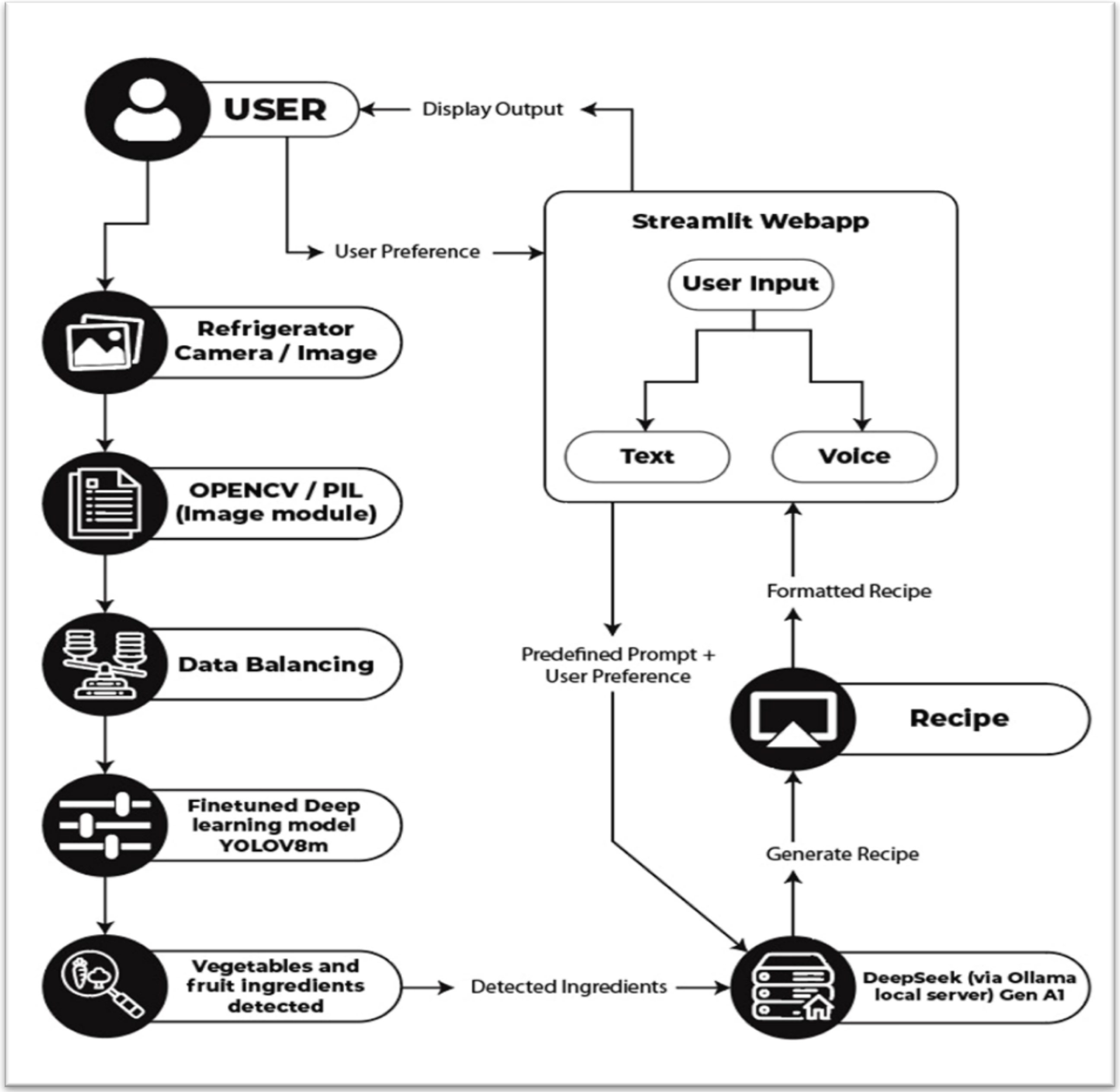
Recipe generation using multimodal language models has advanced as a result of the incorporation of artificial intelligence into culinary innovation. However, fine-tuning-based generative model's techniques are frequently not flexible enough to accommodate regionalized datasets and ingredient availability, especially when it comes to refrigeration inventory that are Indianized. This paper introduces a new AI-powered system that uses prompt engineering and object detection models to generate recipes in real time. This approach, in contrast to conventional fine-tuning techniques, improves inference efficiency while preserving accuracy and variety of recipes. The model was expanded for complicated ingredient recognition, tackling issues like non-transparent food containers and non-standardized ingredient storage, after being trained on a carefully selected dataset comprising 30 food classes. A comparative analysis of different language models reveals trade-offs between usability, latency, and performance. As a result, more context-aware and personalized culinary AI applications are being developed.

PROJECT DESCRIPTION

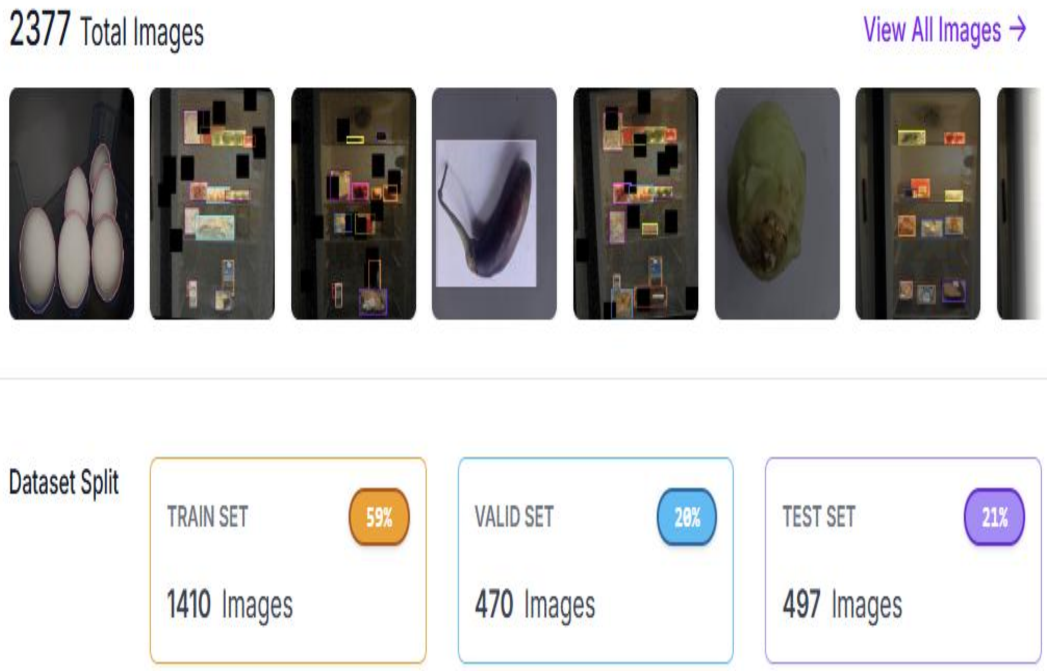
This paper presents an AI-powered Smart Fridge Assistant that uses computer vision (YOLOv8 for real-time ingredient recognition) and Generative AI (Deepseek R1 1.5b for recipe development) to improve kitchen management. The technology scans uploaded fridge photos for accessible components, allowing users to create personalized recipes depending on their dietary choices, cuisine type, and serving size. It also has an Ingredient Availability Checker, which compares detected goods to recipe requirements, identifies missing ingredients, and suggests replacements. The software offers voice input for hands-free operation and contains a Shopping List Analyzer that compares grocery lists to available items, reducing food waste. Built using Streamlit for an interactive UI, the system incorporates multimodal AI (vision + NLP) to deliver real-time culinary recommendations, making meal planning quick and flexible. Future improvements include real-time fridge monitoring via IoT sensors and nutritional analysis to recommend better meals. This initiative fills the gap between AI-powered automation and practical kitchen demands, providing a user-friendly solution for modern households.

SYSTEM FLOW DIAGRAM

Category	Technology/Library	Purpose
Computer Vision	YOLOv8 (via Ultralytics)	Real-time ingredient detection from fridge images using a fine-tuned object detection model.
Generative AI (Local LLM)	DeepSeek-V3 (via Ollama)	Locally-run LLM for recipe generation, eliminating cloud API dependencies.
Speech Recognition	SpeechRecognition (Google STT)	Voice-to-text conversion for hands-free input.
Web Framework	Streamlit	Interactive UI for user-friendly interaction.
Image Processing	Pillow (PIL)	Image handling and preprocessing.
Environment Management	python-dotenv	Secure API key/configuration management.



Dataset Details



The dataset used in this study is a combination of three different sources. The first source is a vegetable dataset featuring publicly available annotated photos, which were used to train models for reliable identification of individual vegetables. The second source consists of photos of food products inside refrigerators gathered online to portray crowded situations and different lighting conditions. The third source is a proprietary dataset of Indianized food products from local refrigerators, which reflects distinctive food storage patterns specific to Indian homes.

Dataset Overview:

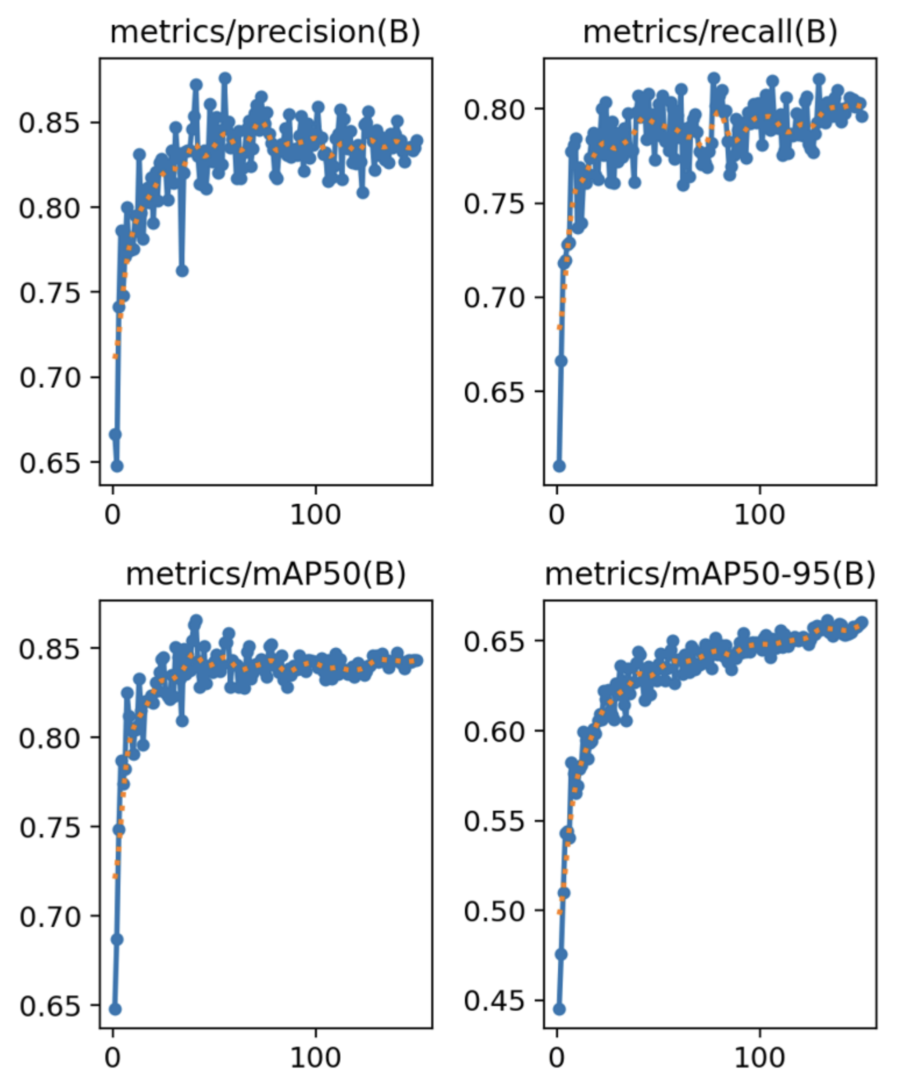
•**Total Images**: 2377
•**Classes (Food Items)**: 30
•**Dataset Split**:

- **Train Set**: 1410 images (59%)
- **Validation Set**: 470 images (20%)
- **Test Set**: 497 images (21%)

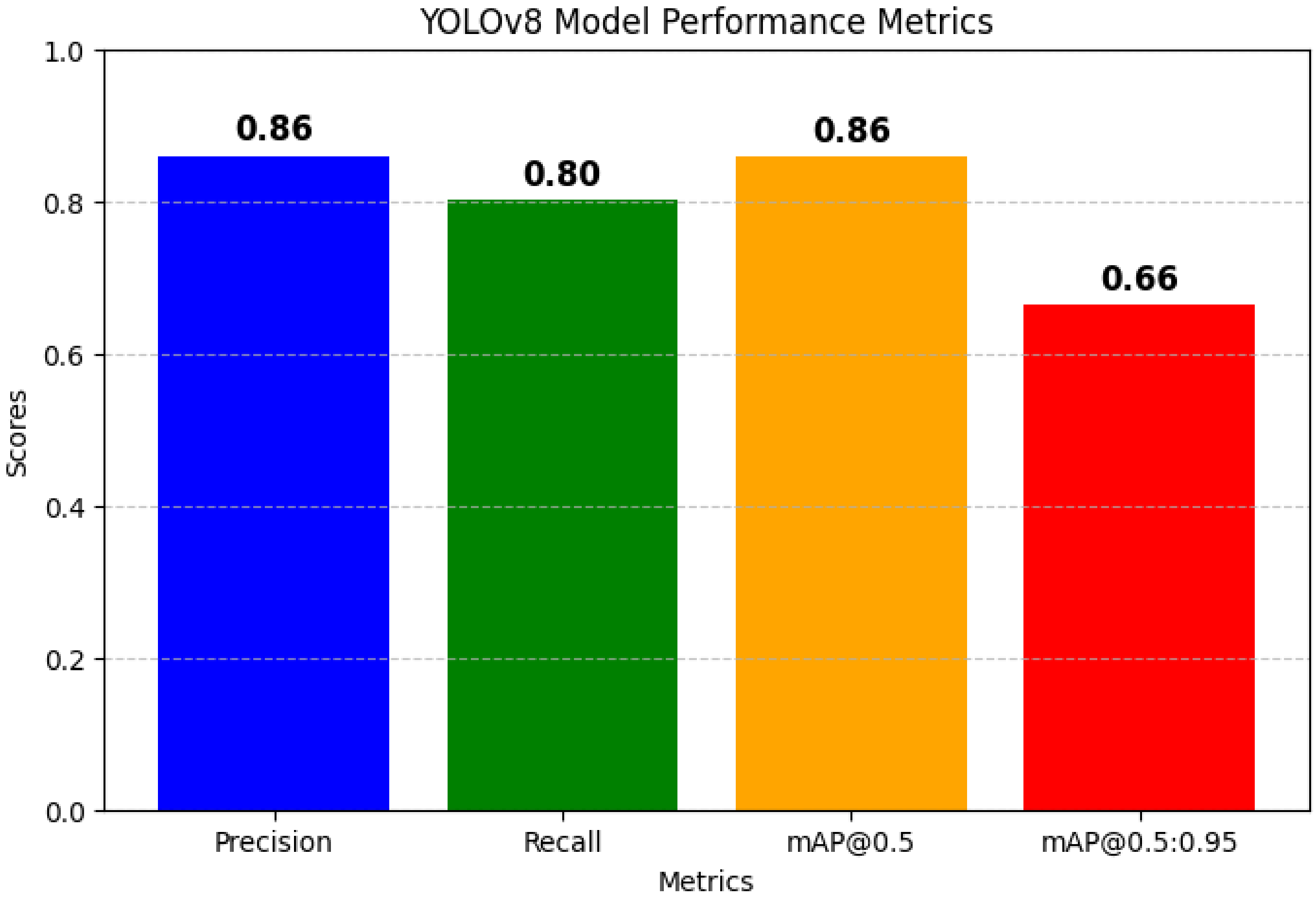
EVALUATION METRICS FOR FINE-TUNED YOLOV8M, YOLOV11,YOLO-NASS AND ROBOFLOW 3.0 MODEL

	YOLOv8m	YOLOv11	YOLO-NASS	ROBOFLOW 3.0
Accuracy	**84%**	81%	79%	81%
Precision	**0.8481**	0.8211	0.8192	0.8364
F1 Score	**0.8435**	0.8063	0.7841	0.8027
mAP	**0.8697**	0.8472	0.8097	0.8455
Recall	**0.8032**	0.7961	0.7867	0.7782

YOLO V8 Training Graphs



...
- Precision: 0.8603
- Recall: 0.8009
- mAP@0.5: 0.8594
- mAP@0.5:0.95: 0.6631



Representation of performance of the finetuned YOLOv8m object detection model in identifying ingredients in a Indianized refrigerator.

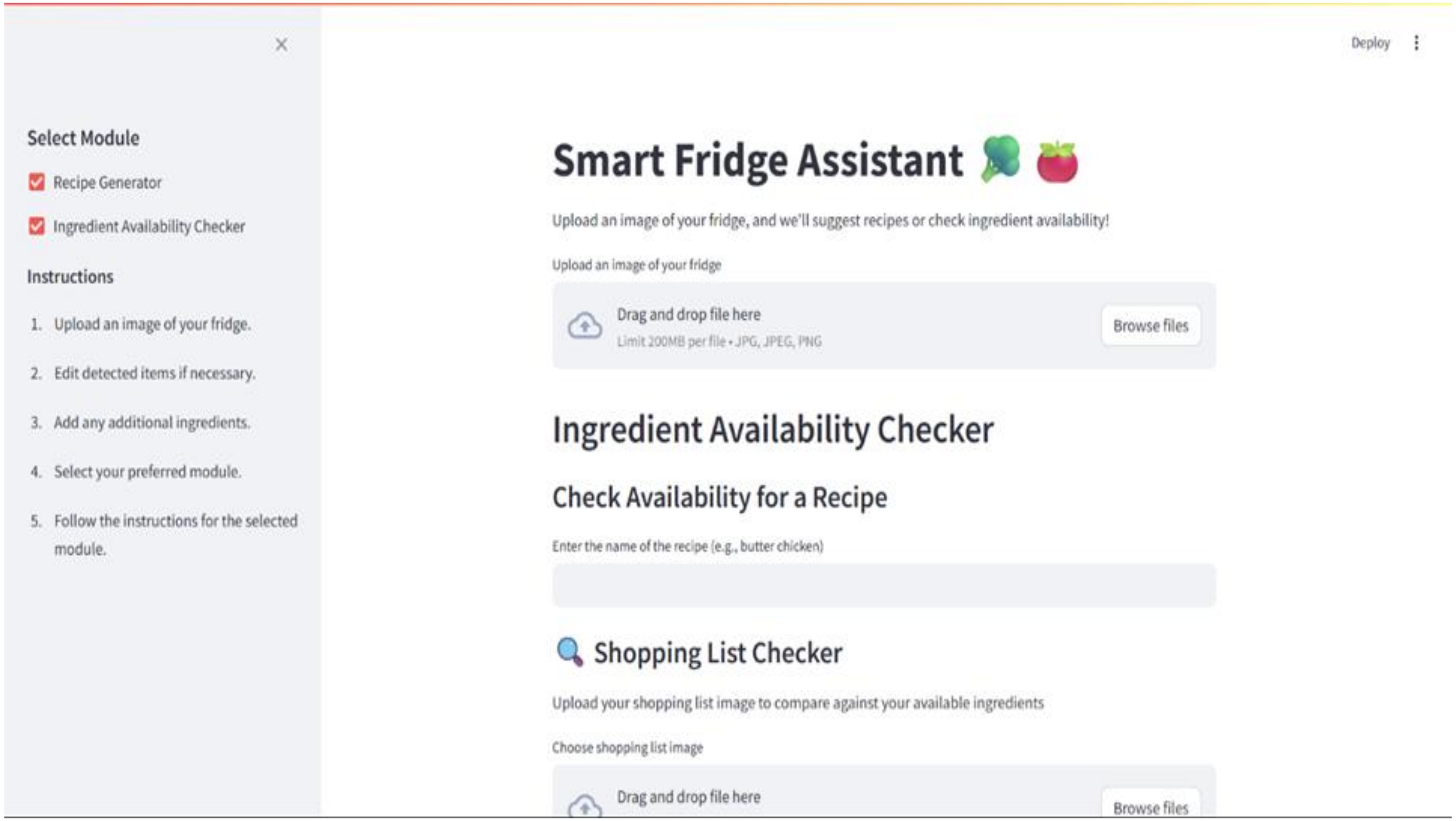
Comparative Analysis of Generative Model Performance

Feature	ChatGPT-3.5	Gemini 2.0 Pro Experimental	DeepSeek R1 1.5B
Model Size	~175B parameters (estimated)	Not officially disclosed	1.5B parameters
Deployment	API-based (OpenAI)	API-based (Google AI)	Runs locally via Ollama
Latency (the time delay between making a request to an LLM API and receiving a response)	0.30/ms	0.59/ms	60.48/ms
Context window (the amount of text, in tokens, that the model can remember at any one time)	Input - 16,385 tokens Output - 4,096 tokens	Input - 2M tokens Output – 8k tokens	Input - 128K tokens Output - 32K tokens
Inference Time (the process of generating output based on an input prompt)	Fast (Cloud-Optimized)	Fast (Cloud-Optimized)	Slower (Local Inference)
Availability	Proprietary	Proprietary	Opensource

Comparative Analysis of Generative Model Performance

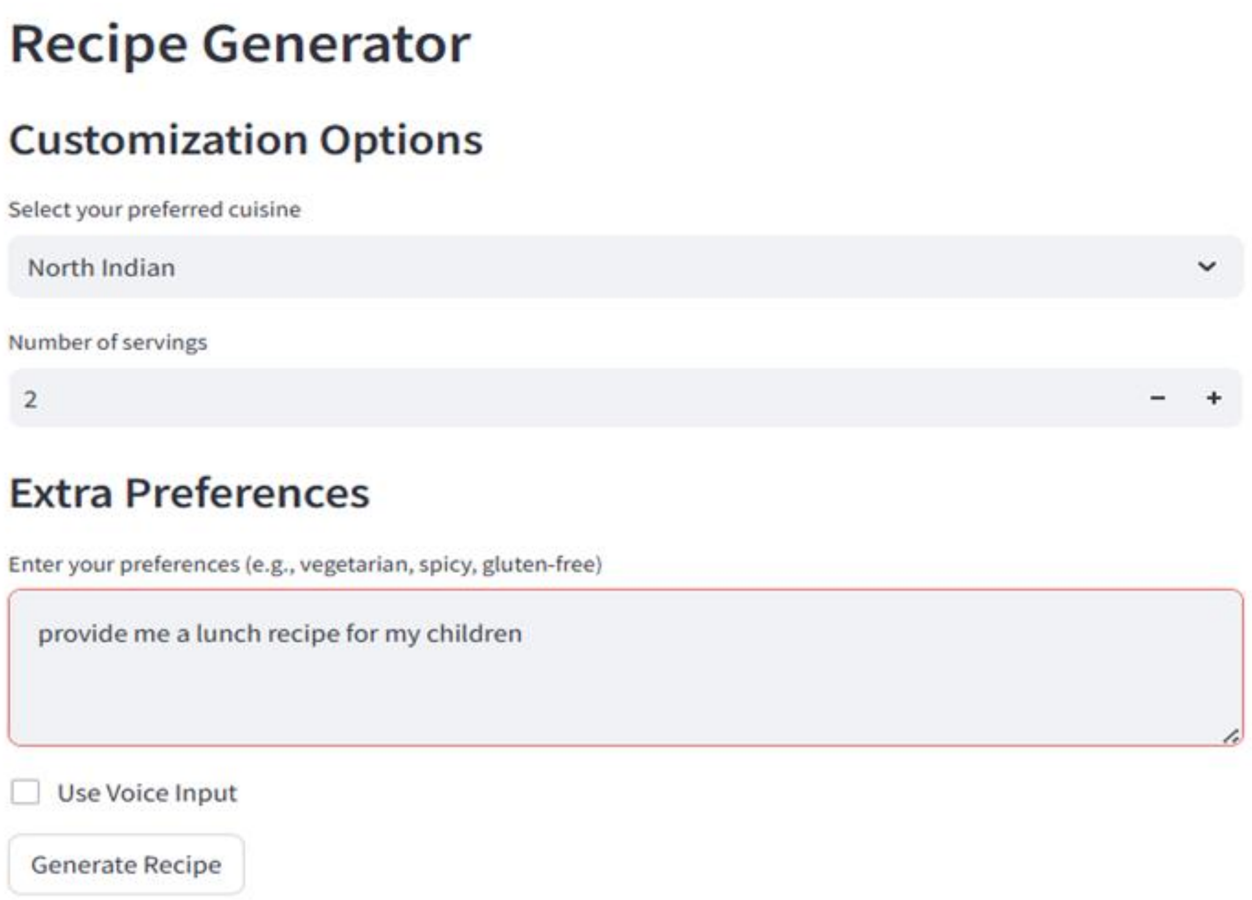
Throughput (The number of output tokens per second an inference server can generate across all users and requests)	90 tokens/s	120 tokens/s	0.9 tokens/s
GPQA (Graduate-Level Google-Proof Q&A Benchmark)	30.8%	64.7%	71.5%
DROP (Discrete Reasoning Over Paragraphs)	70.2%	-	92.2%
Quality index	44	49	89

The Smart Fridge Assistant's User Interface



The Streamlit based multimodal assistant Web app is user-friendly and allows users to upload images of their refrigerator interiors. The application allows for the manual modification and enhancement of the detected ingredients, ensuring flexibility and accuracy. It has two main modules: Recipe Generator and Ingredient Availability Checker.

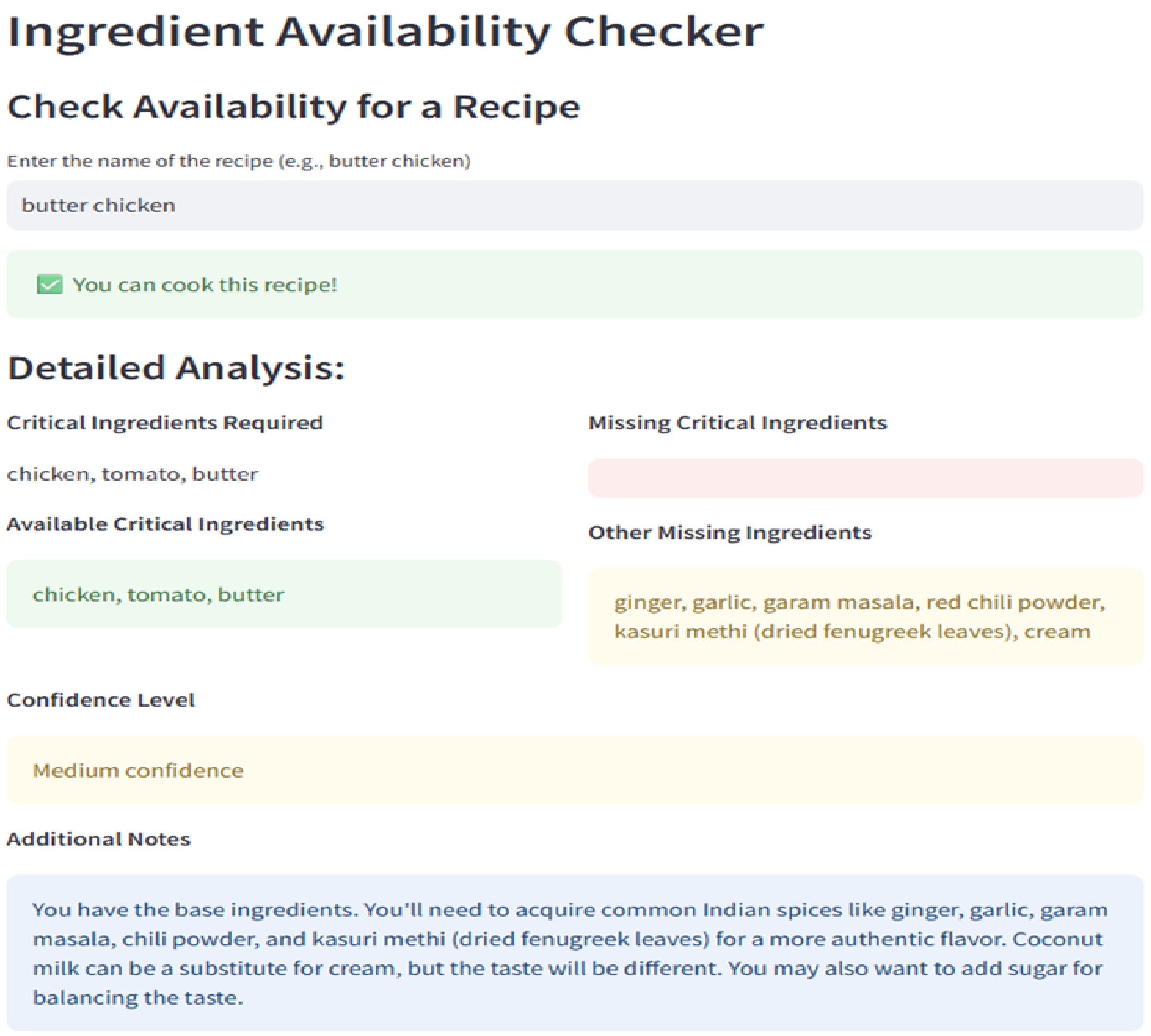
The Smart Fridge Assistant's User Interface



The Recipe Generator uses DeepSeek r1 1.5B to generate customized, cuisine-specific meals based on detected and available items, user preferences, and the number of servings. It also enables voice-based preference input via Google Speech Recognition, which improves the app's accessibility and user experience.

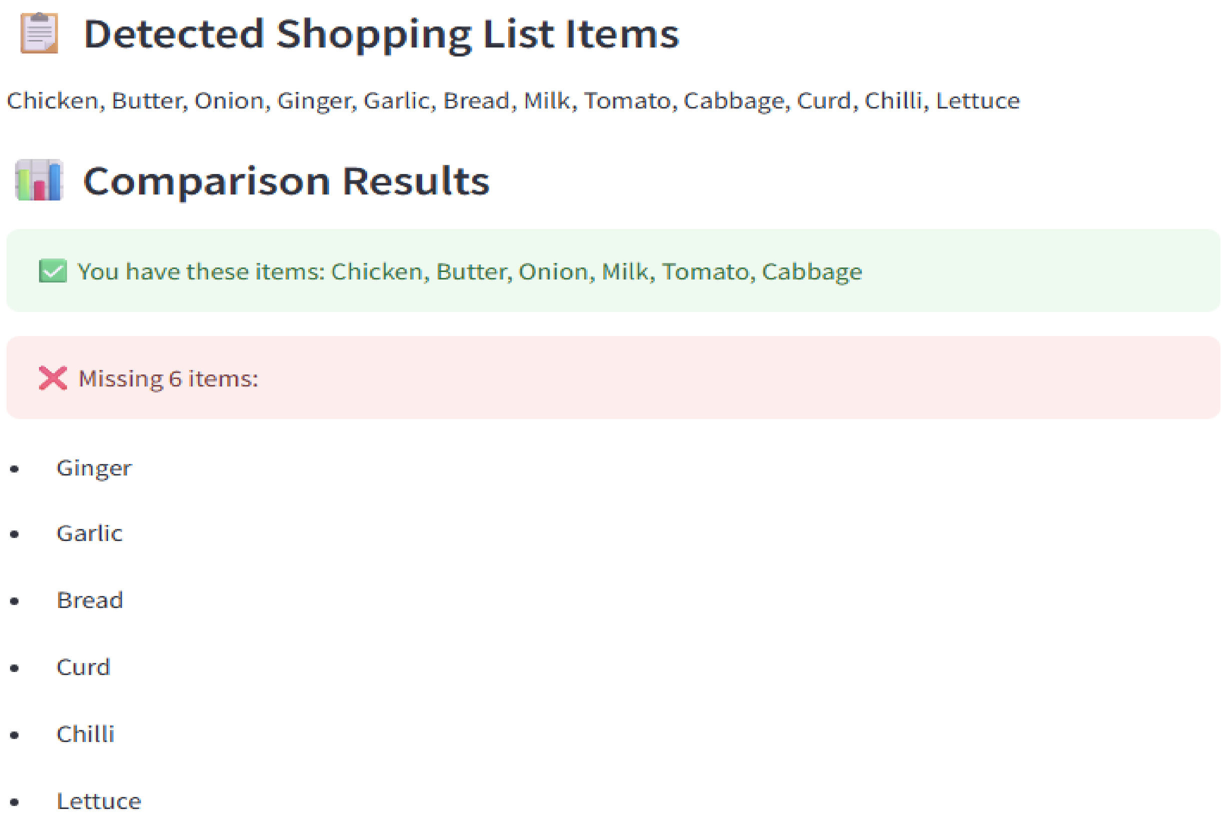
Ingredient detection with YOLOv8, displaying the recognised food items from the uploaded fridge image. Users can change or add missing ingredients to ensure correctness

The Smart Fridge Assistant's User Interface



The Ingredient Availability Checker allows users to enter a recipe name, and the DeepSeek model determines whether the ingredients in the refrigerator are sufficient to make it, providing a detailed breakdown of critical and missing ingredients, confidence levels, and substitution suggestions.

The Smart Fridge Assistant’s User Interface



Shopping List Analyzer module allows users to upload shopping list photos, which are then cross-referenced with accessible fridge items using either YOLO-based detection or fallback processes that mimic OCR-based extraction. The app then highlights which products on the list are already available and which must be purchased. The entire system is modular, responsive, and user-centric and provides simple step-by-step instructions for users.

CHALLENGES

The lack of an Indianized fridge dataset was one of the primary challenges encountered in this study, which made it challenging to successfully train the model for food ingredients odd to a specific place. I gathered and annotated a small dataset to address this, but I still plan to extend it in the future. Furthermore, only 30 classes were used to train the model; for better accuracy, a larger dataset featuring more classes is needed. The existence of vessels and opaque containers inside Indian refrigerators presented another significant obstacle, making it challenging for the model to accurately identify food items. Additionally, the model found it difficult to distinguish between components like curry leaves or coriander leaves kept in plastic bags in the absence of accurate annotations. The dataset has to be increased further with better-labelled photos and a variety of storage settings in order to perform better. I used prompt engineering over fine-tuning a Generative AI model for recipe production because it is more flexible in responding to different user inputs, lowers computational costs, and does not require large amounts of labelled training data. Furthermore, domain-specific datasets and substantial hardware resources are needed to fine-tune a big model, and these were not easily accessible. In order to improve real-world performance, future developments will concentrate on diversifying datasets, improving annotations, and reducing inference.

CONCLUSION

The multi-modal hybrid system handled ingredient detection and recipe development efficiently by merging computer vision (YOLOv8) with generative AI (DeepSeek). This combination improved item detection even in congested refrigerators, allowing the system to create contextually relevant meals based on human tastes and cultural contexts. By employing prompt engineering rather than fine-tuning, the system minimized its need on large labelled datasets and processing resources. This multi-modal smart fridge recipe generator overcame major challenges by creating and annotating a dataset of Indian specific ingredients, improving ingredient recognition with better annotations and diverse training data, and implementing a robust hybrid system that seamlessly combined computer vision and generative AI. The system achieves high accuracy and efficiency by combining a range of datasets and refining model architectures, hence improving the kitchen management user experience. This innovative technology presents itself as a valuable supplement to intelligent home automation systems, with significant promise for reducing food waste and supporting healthy eating habits.

FUTURE WORK

Future research will concentrate on adding more extensive datasets, such as fine-tuning object detection model with more ingredients or labelled dataset, barcode-recognized packaged foods, to strengthen the system's resilience. To meet the demands of a wide range of users, multilingual support and dietary preference customisation will be expanded, increasing user accessibility. Furthermore, improving real-time performance through model inference optimisation for edge devices would open the door for a broader integration of smart kitchen technology in contemporary homes.

REFERENCES

1. H. Lee, Helena, Ke Shu, Palakorn Achananuparp, Philips Kokoh Prasetyo, Yue Liu, Ee-Peng Lim, and Lav R. Varshney. "RecipeGPT: Generative pre-training based cooking recipe generation and evaluation system." In *Companion Proceedings of the Web Conference 2020*, pp. 181-184. 2020.

2. Marın, Javier, Aritro Biswas, Ferda Ofli, Nicholas Hynes, Amaia Salvador, Yusuf Aytar, Ingmar Weber, and Antonio Torralba. "Recipe1m+: A dataset for learning cross-modal embeddings for cooking recipes and food images." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, no. 1 (2021): 187-203.

3. R. S. Latha et al., "Fruits and Vegetables Recognition using YOLO," 2022 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2022, pp. 1-6, doi: 10.1109/ICCCI54379.2022.9740820.

4. Tian, Yijun, Chuxu Zhang, Zhichun Guo, Chao Huang, Ronald Metoyer, and Nitesh V. Chawla. "RecipeRec: A heterogeneous graph learning model for recipe recommendation." *arXiv preprint arXiv:2205.14005* (2022).

5. B. Chopra, G. Jain, N. Mahendra, S. Sinha and S. Natarajan, "Ingredient Detection, Title and Recipe Retrieval from Food Images," 2023 Fourteenth International Conference on Ubiquitous and Future Networks (ICUFN), Paris, France, 2023, pp. 289-293, doi: 10.1109/ICUFN57995.2023.10199735.

Thank You