



Ingredient Detection and Recipe Recommendation Using Deep Learning

Divya Bharathi P¹, Gobburu Harish², Guna R³, Hemanth R⁴, Madhusudan B⁵

¹Assistant Professor/CSE, Sri Sairam College of Engineering, Bangalore – 560126

^{2,3,4,5} Final Year/CSE, Sri Sairam College of Engineering, Bangalore – 560126

ABSTRACT

In response to the hectic pace of modern life, there's a growing need for a smartphone web app that streamlines meal preparation. Our project aims to address this need by developing a sophisticated recipe recommendation system powered by technologies such as computer vision and machine learning. The primary objective is to simplify the culinary experience for users who often find themselves uncertain about what to cook with the ingredients they have on hand. By leveraging computer vision techniques, our system can accurately identify the ingredients available to the user. This information is then processed using machine learning algorithms to generate tailored recipe suggestions. This approach eliminates the need for extensive meal planning or manual recipe searches, saving users valuable time and effort. To tackle this, we prepared an ingredient dataset containing 12,558 images across 15 food ingredient classes. The YOLOv8 object detection model was used to detect and classify food ingredients. Additionally, the recommendation system was built using machine learning. In the end, we achieved an accuracy of 96%, which is quite impressive.

1 Introduction

In a world where convenience frequently takes precedence over creativity in culinary trends, our project seeks to revolutionize home cooking. Our commitment lies in creating an innovative mobile application that tackles a basic kitchen problem: what to make using the ingredients at hand. Our app signifies a change in the way people approach cooking; it goes beyond conventional recipe apps by utilizing object detection technology to evaluate photos of ingredients and provide recipe recommendations based on the detected ingredients. At its core lies the YOLOv8 model, used in real-time object detection, which serves as the foundation for identifying ingredients within culinary images. Complementing this, the system utilizes word embeddings, a technique rooted in natural language processing, to discern semantic similarities among ingredients and curate tailored recipe recommendations. Powered by a fast API backend, the system presents a streamlined interface for seamless interaction, facilitating efficient data transmission and processing. Through meticulously crafted endpoints, it offers a gateway to the underlying functionalities, enabling users to effortlessly harness the system's capabilities. There is a trend in the modern world toward more individualized and health-conscious culinary experiences. People are looking for creative ways to find and try new recipes as a result of the wide variety of products and recipes that are becoming more and more accessible. Still, it can be difficult to find ingredients and select recipes that make use of what



is available, especially for people with dietary limitations or no culinary experience. Intelligent systems that can help users identify components and suggest meals based on their preferences and available resources are therefore becoming more and more necessary. By creating a deep learning-based system for component detection and recipe recommendation, this research seeks to address this issue. This system uses picture identification and sophisticated computer vision and natural language processing algorithms to recognize substances in photographs.

2 Recent Works

Utilizing state-of-the-art methods in computer vision and natural language processing, recent developments in artificial intelligence have propelled breakthroughs in ingredient detection and recipe recommendation systems. For training models to recognize ingredients and produce recipes, datasets such as Recipe1M, which has more than a million recipes associated with ingredients and photographs, have established themselves as industry standards. Recurrent neural networks (RNNs) and convolutional neural networks (CNNs) are used in systems like Pic2Recipe to match photos with recipes and identify ingredients. Transformer-based models, such as BERT4Rec, have improved recipe generation by capturing contextual interactions among components, while attention mechanisms have further improved ingredient recognition by allowing models to focus on pertinent regions of an image. Hybrid systems that combine collaborative and content-based filtering techniques have shown promise in recommendation, enabling suggestions that take user preferences and ingredient properties into account. While customized meal planners provide weekly suggestions in line with customers' health objectives, other systems, such as Health-aware Recipe Generation models, employ nutritional data to Customize recipes to certain dietary requirements, such as low-sodium or gluten-free diets. End-to-end systems like INSTAFOOD combine text, image, and recipe data with multi-modal networks to establish a direct connection between recipe suggestions and ingredient recognition in a smooth user experience. A major step toward smart kitchen assistants, these intelligent, health-conscious, and user-specific suggestion systems encourage effective ingredient use and minimize food waste while making cooking more approachable and customized for users.

3 Proposed Work

In order to provide a more efficient cooking experience, the proposed study focuses on creating a deep learning-powered system for component detection and recipe selection. Users will be able to take a picture of the ingredients they have on hand, have the system automatically identify those ingredients, and get recipe suggestions that are specific to their ingredients. Recipe recommendation and ingredient detection are the two primary parts of the system. To accurately identify common kitchen goods, a deep learning-based object identification model—like YOLO or Faster R-CNN—will be refined using a bespoke dataset that contains photos tagged with ingredients. In order to recognize many ingredients in a single image and standardize ingredient names for recommendation accuracy, attention methods will improve



the model's focus on pertinent image regions. A hybrid recommendation engine will integrate collaborative filtering and content-based filtering in the recipe suggestion component. Cosine similarity and multi-hot encoding will be used to match detected items with recipes, and collaborative filtering will take advantage of user preferences and dietary limitations to provide results that are tailored to each individual. A knowledge graph will handle ingredient substitutions, enabling recipe modifications depending on dietary requirements, further increasing the adaptability of recipe recommendations. The performance of each component will be evaluated using conventional measures, such as precision and recall for ingredient detection and Precision@K and NDCG for the recommendation engine. The overall goal of this system is to reduce food waste, streamline meal planning, and encourage culinary exploration by providing an end-to-end solution for component recognition and customized recipe recommendation.

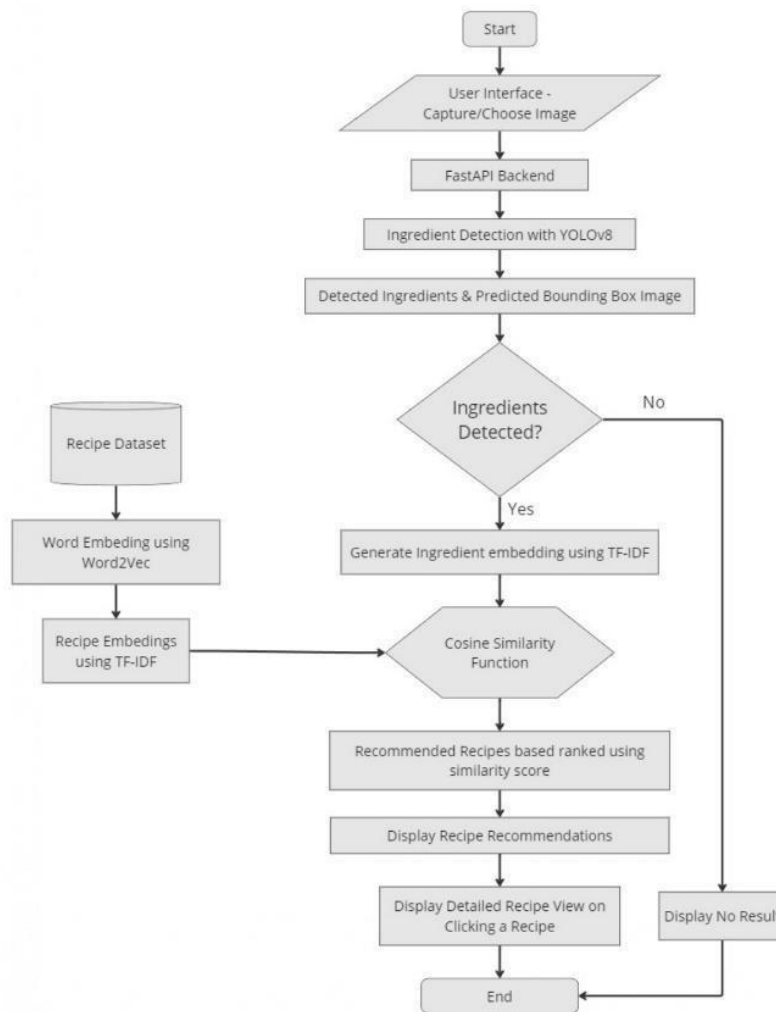


Figure 1: *proposed block diagram*



4 Mathematical Expression and Symbols

Both the ingredient detection and recipe recommendation components can function more clearly when mathematical expressions and symbols are used. Below is a summary of some important terms and symbols that may be helpful:

4.1 Identification of Ingredients

Assume that the objective of ingredient detection is to locate and categorize the various ingredients included in an image I . This procedure can be stated as follows:

4.2 Model Output for Object Detection

Let the model produce a set of bounding boxes, $B = \{b_1, b_2, \dots, b_n\}$, where each $b_i = (x_i, y_i, w_i, h_i)$ represents the coordinates and dimensions of a recognized object (ingredient).

A class name c_i and a confidence score s_i are assigned to each bounding box b_i .

The formula $b_i = \text{model}(I) = \{(x_i, y_i, w_i, h_i), c_i, s_i\}$ If $b_i = \text{model}(I) = \{(x_i, y_i, w_i, h_i), c_i, s_i\}$

4.3 Attention Mechanism

- The attention mechanism focuses on relevant image regions. If we denote attention weights by α , the final feature vector f for each region i can be computed as:
 $f_i = \alpha_i \cdot \text{Feature}(b_i)$
- Here, α_i represents the weight assigned to each region by the attention model.

4.4 Collaborative Filtering (Matrix Factorization)

- Let $R \in \mathbb{R}^{m \times n}$ be a user-recipe interaction matrix, where R_{ij} is the rating of recipe j by user i
- Factorize R as $R \approx U \cdot V^T$, where $U \in \mathbb{R}^{m \times k}$ and $V \in \mathbb{R}^{n \times k}$ are latent feature matrices $R_{ij} \approx U_i \cdot V_j^T$

Hybrid Recommendation Score

The final recommendation score s_{rsr} for a recipe r combines content-based similarity and collaborative filtering, weighted by parameters α and β :

$$S_r = \alpha \cdot \text{sim}(d, r) + \beta \cdot (U_i \cdot V_j^T)$$

4.5 Knowledge Graph for Substitutions

- Represent ingredient substitutions as edges in a knowledge graph $G = (V, E)$, where each node $v \in V$ is an ingredient, and each edge $e = (v_1, v_2) \in E$ indicates a possible substitution.
- For an ingredient v_d not in d , find substitutions $S(v_d)$ that minimize a distance metric δ in the graph: $S(v_d) = \arg v \in V \min \delta(v_d, v)$

5 Results and Discussion

Two distinct models, YOLOv8 and Word2Vec, were trained for the tasks of ingredient detection and recipe recommendation for photos in this deep learning experiment. A dataset of



ingredient photos and recipes was used for training, and each model's architecture was put into practice. A dataset including around 12,558 photos and 15 ingredient classes with 36,159 annotations overall was used to train the YOLOv8 algorithm. 50 epochs were used to train the model.

Through rigorous testing and assessment, the Web App's implemented YOLOv8 object detection model's accuracy is confirmed to be high in recognizing ingredients from photos. On the test set, the model's remarkable 96% accuracy rate is achieved. The model also did well on a number of other evaluation metrics, such as recall, precision, and F1 score. Ingredient similarities between the uploaded photographs and the recipes in the dataset are efficiently calculated by semantic analysis using Word2Vec and TF-IDF with cosine similarity. The accuracy of the recommendation system can be assessed by gauging how well the suggested recipes fit the ingredients and how pertinent they are to the user's preferences, even though the Word2Vec model can offer recommendations based on the detected ingredients.

The ingredient substitution knowledge graph's effectiveness highlights the system's adaptability, making it a helpful resource for users with restricted pantry ingredients or dietary restrictions. In order to further improve the model's applicability across a variety of cuisines and dietary preferences, future study could expand the ingredient and recipe datasets and incorporate user-driven feedback for ongoing model refining. In conclusion, the suggested system successfully blends user-centered design with deep learning approaches, demonstrating encouraging outcomes in both recipe recommendation and ingredient recognition. These findings open the door for intelligent kitchen aides that provide useful, individualized assistance with meal preparation and cooking.



Figure 2:YOLOv8 Predictions on



Figure 3: Application Interface the Test Data test



Figure 4: Results after Detection of



Figure 5: Detailed Recipe View



6 Conclusion

In conclusion, the implemented system successfully addresses the need for personalized recipe recommendations based on culinary images. Through the integration of technologies such as YOLOv8 for object detection and Word2Vec with TF-IDF and cosine similarity for semantic analysis, the system achieves accurate and relevant recipe suggestions. The comprehensive dataset collection ensures the quality and diversity of recipe recommendations, enhancing the user experience. Thorough testing and evaluation validate the system's effectiveness, the object detection model achieved an accuracy score of 96% along with high-quality recommendations consistently delivered. Moving forward, continuous refinement and optimization can further improve the system's performance and expand its capabilities. Overall, this project showcases the potential of advanced technologies in revolutionizing culinary exploration and enhancing user satisfaction in discovering new and exciting recipes.

References

1. Maheshwari; Suyash; Manas Chourey, Year: 2019, "Recipe recommendation system using machine learning models," International Research Journal of Engineering and Technology (IRJET), Vol: 6, no: 9, pp. 366-369.
2. Thi Phuong Chung; Dang; Dinh Van Tai, Year: 2019, "A fruits recognition system based on a modern deep learning technique," In Journal of physics: conference series, Vol 1327, no: 1, pp. 012050.
3. Morol; Md Kishor; Md Shafaat Jamil Rokon; Ishra Binte Hasan; Saif; Rafid Hussain Khan; Shuvra Smaran Das, Year: 2022, "Food recipe recommendation based on ingredients detection using deep learning," In Proceedings of the 2nd International Conference on Computing Advancements, pp. 191-198.
4. Chhipa; Shubham; Vishal Berwal; Tushar Hirapure; Soumi Banerjee, Year: 2022, "Recipe recommendation system using TF-IDF," In ITM web of conferences, Vol: 44, pp. 02006.
5. Rodrigues; Miguel Simoes; Filipe Fidalgo; Angela Oliveira, Year: 2023 "RecipeIS—Recipe Recommendation System Based on Recognition of Food Ingredients," Vol: 13, no. 13, pp. 7880.