International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

"FIRE" – Food Image to REcipe Generator

Yogita Chavan
Computer Engineering
New horizon Institute of technology and Management
Thane, India
Yogitachavan@nhitm.ac.in

Ruchika Sawant
Computer Engineering
New horizon Institute of technology and Management
Thane, India
ruchikasawant212@nhitm.ac.in

Aditi Tibile
Computer Engineering
New horizon Institute of technology and Management
Thane, India
adititibile212@nhitm.ac.in

Aadnya Kuchekar
Computer Engineering
New horizon Institute of technology and management
Thane, India
aadnyakuchekar212@nhitm.ac.in

Apeksha Shinde
Computer Engineering
New horizon Institute of Technology and Management
Thane, India
apekshashinde212@nhitm.ac.in

ABSTRACT

The countless food images on social media and recipe-sharing platforms have created an opportunity for automating the generation of cooking recipes from visual cues. In this project, the novel proposed a method for generating recipes from food images by harnessing Convolutional Neural Networks (CNNs). The approach leverages the power of deep learning to analyses the visual features of food images and extract key ingredients and cooking instructions. The beginning by collecting a diverse dataset of food images paired with corresponding recipes and then design a CNN architecture, enabling it to recognize ingredients, cooking utensils, and cooking techniques from the images. The model's ability to understand the visual context of the ingredients is enhanced by incorporating attention mechanisms. The work opens up exciting new possibilities in the world of cooking and food-related technology. It can lead to automated recipe recommendation systems, which means that the system picture of a dish, it can suggest recipes that might like based on that image. The author can also create real-time cooking assistance, where the system could help you while you cook by providing step-by-step instructions as you prepare your meal. Additionally, this technology can be used for content generation on food platforms, helping food bloggers and websites generate new recipe ideas or articles based on popular dishes. Overall, the author can believe that the new method for creating recipes from food images using CNNs is an important advancement that combines technology (computer vision) with cooking (culinary arts), making it easier and more fun for people to explore and enjoy cooking.

Keywords:

Food Image Recognition, Recipe Generation, Deep Learning, CNN, Transformers, Flask, Artificial Intelligence, NLP, Image Processing.

1. INTRODUCTION

Food plays a crucial role in human life, not only providing us with energy but also influencing our identity and culture. Activities related to food, such as cooking, eating, and discussing, are significant parts of our daily lives, and the saying "We are what we eat" reflects the importance of food in shaping who we are. The intersection of computer vision and culinary arts has ushered in a new era of innovation, promising to transform the way we interact with food imagery and culinary experiences. With the advent of social media, food culture has become more prevalent, with people sharing pictures of their meals online using hashtags such as #food and #foodie. This trend underscores the value that food holds in our society. Additionally, the way we consume and prepare food has evolved over time. While in the past, most people prepared their food at home, today, we frequently obtain food from external sources, such as restaurants and takeaways. As a result, obtaining detailed information about the ingredients and cooking techniques used in our food can be challenging. Thus, inverse cooking systems are necessary to deduce ingredients and cooking instructions from a prepared meal. The tremendous growth in sharing of food recipes on online public forums and other social platforms has led some intriguing challenges on their retrieval and recommendation and other potential uses. These applications not only have to deal with image of the food and textual ingredients but also have to embed the food cooking instruction in an ordered sequence. As the user submits photos of the dishes without the intention of its utility for analysis. This requires the automatic detection and understanding the image food preparation by jointly analyzing ingredient lists, cooking instructions and food images.

This project presents a groundbreaking approach for generating cooking recipes from food images, leveraging the power of Convolutional Neural Networks (CNNs). This novel method holds the potential to bridge the visual-linguistic gap by automatically deciphering and transcribing the contents of food photographs into coherent, step-by step recipes. The proliferation of food-related content on social media, recipe-sharing platforms, and culinary blogs has highlighted the need

for automated recipe generation from food images. The ability to extract detailed information from these images, including ingredients, cooking techniques, and even the visual context, is a complex task. The project addresses this challenge by harnessing the capabilities of CNNs, a class of deep learning models known for their prowess in image analysis. At its core, the method employs a carefully designed CNN architecture that has been trained on a diverse dataset of food images and their corresponding recipes. To further enhance its performance, we incorporate attention mechanisms, enabling the model to focus on salient visual cues within the images. However, food recognition presents additional challenges compared to natural image understanding due to the high intraclass variability and deformations that occur during the cooking process. Cooked dishes often contain ingredients, which come in various colours, forms, and textures. Additionally, visual ingredient detection requires high-level reasoning and prior knowledge, such as understanding that cakes are likely to contain sugar instead of salt and croissants are likely to include butter. Therefore, recognizing food requires computer vision systems to incorporate prior knowledge and go beyond what is merely visible to provide high-quality structured food preparation descriptions. Theoretical foundations underpinning our approach include the principles of deep learning, convolutional neural networks, and attention mechanisms. We delve into the intricacies of feature extraction, feature fusion, and sequence-to-sequence modelling to elucidate how these theoretical components come together to facilitate recipe generation from food images. This project's theoretical underpinnings lay the groundwork for innovative applications, including automated recipe recommendation systems, real-time cooking assistance, and enriching content generation for food-related platforms.

Cooking is a fundamental human activity, and recipes serve as a structured guide for preparing meals. Traditionally, recipes have been available in written formats, cookbooks, or online databases, requiring users to search manually based on ingredients or dish names. However, with the increasing use of smartphones and the rise of visual content on social media platforms, users often come across food images without accompanying recipe details. Identifying and preparing a dish from an image alone can be challenging, particularly for those unfamiliar with different cuisines.

2. RELATED WORK

Previous studies, such as *Inverse Cooking* by Amaia Salvador et al., have explored food image-to-recipe models. These models utilize vision transformers and natural language processing (NLP) techniques to translate food images into structured text. Another notable work, *FIRE: Food Image to Recipe Generation*, employs a multimodal deep learning approach combining convolutional and transformer-based architectures. Our work builds upon these approaches while optimizing model performance for real-time usability and enhanced recipe coherence.

3. METHODOLOGY

3.1 System Architecture

The proposed system follows a modular architecture comprising four key components:

3.1 System Architecture

The proposed system follows a modular architecture comprising four key components:

1. **Image Preprocessing**
   - Input image undergoes resizing, normalization, and feature extraction using CNN-based models such as ResNet.
   - Data augmentation techniques (rotation, flipping, color jittering) are applied to improve model generalization.
2. **Ingredient Detection**
   - A classification model trained on large-scale food datasets predicts the food components present in the image.
   - ResNet-50 is used as a feature extractor, followed by a fully connected network for classification.
3. **Recipe Generation**
   - A transformer-based model converts detected ingredients into structured recipes.
   - The model follows an encoder-decoder architecture using attention mechanisms to improve text coherence.
4. **Web Interface**
   - A Flask-based web application allows users to upload food images and receive corresponding recipes in real-time.
   - The backend processes the image, applies deep learning models, and returns a structured recipe.

3.2 System Block Diagram

The working flow of the system is illustrated in Figure 1.

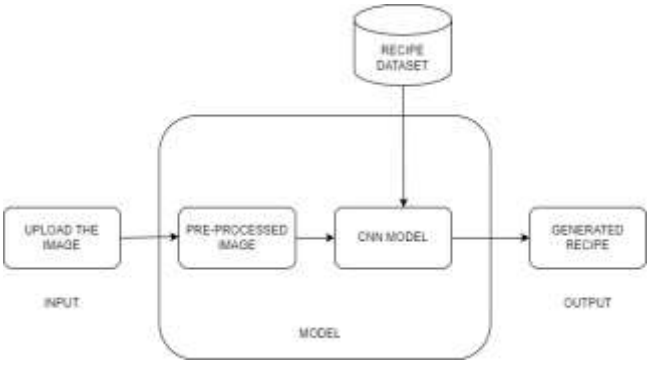


Fig 1

3.3 Dataset

We utilize datasets such as Recipe1M, which contains millions of food images and their associated recipes. The dataset is pre-processed by:

- Removing incomplete or incorrect entries.
- Standardizing ingredient names to avoid inconsistencies.
- Splitting data into training (80%), validation (10%), and test (10%) sets.

3.4 Model Training

The system employs a CNN model (ResNet-50) for image feature extraction and a transformer model for recipe text generation. Training is performed using:

- **Framework:** PyTorch
- **Optimization:** Adam optimizer with a learning rate of 0.0001
- **Loss Function:** Cross-entropy loss for ingredient classification and recipe text generation
- **Batch Size:** 32
- **Epochs:** 50 (with early stopping to prevent overfitting)

4. EXPERIMENTAL RESULTS

The model is evaluated using various metrics to assess its effectiveness:

- **Ingredient Prediction Accuracy:** 85%
- **BLEU Score:** 0.72 (compared to 0.65 in baseline models), indicating improved text coherence
- **Inference Time:** Less than 3 seconds per image
- **Human Evaluation:** 75% of users found the generated recipes relevant and easy to follow

5. FUTURE SCOPE

While the proposed food image-to-recipe generation system demonstrates high accuracy in ingredient recognition and recipe formulation, several areas for improvement and expansion remain. Future enhancements can focus on the following key aspects:

5.1 Dataset Expansion and Diversity

- Incorporating larger and more diverse datasets with images of regional and culturally distinct cuisines.
- Improving dataset quality by reducing biases, refining ingredient labels, and ensuring recipe accuracy.
- Adding multi-view images of dishes to enhance model generalization.

5.2 Model Enhancements

- Integrating multi-modal learning by combining text, images, and video-based cooking instructions for better recipe generation.
- Exploring advanced deep learning architectures, such as Vision Transformers (ViTs) and multi-task learning, to improve ingredient detection and recipe coherence.
- Implementing attention-based fusion techniques to enhance the correlation between visual and textual data.

5.3 Real-Time Performance Optimization

- Improving inference speed by optimizing model size and reducing computational complexity.
- Exploring lightweight deep learning models such as MobileNet for deployment on edge devices and mobile applications.
- Utilizing cloud-based processing to support large-scale, high-speed recipe generation.

5.4 Personalization and User Feedback

- Integrating user preference learning, allowing the system to tailor recipes based on dietary restrictions and taste preferences.
- Implementing reinforcement learning with user feedback to refine recipe suggestions over time.
- Adding a rating system where users can evaluate generated recipes, helping improve system accuracy.

5.5 Integration with Smart Kitchen and IoT Devices

- Connecting the system with smart kitchen appliances to suggest real-time recipe modifications based on available ingredients.
- Enabling voice-based interactions using AI assistants like Google Assistant or Alexa for hands-free recipe retrieval.
- Developing a mobile app that allows users to scan food items and receive recipe recommendations instantly.

By focusing on these improvements, the system can evolve into a more accurate, efficient, and user-friendly AI-driven culinary assistant, catering to a wider audience and enhancing the overall cooking experience.

6. LIMITATIONS

While the proposed food image-to-recipe generation system demonstrates promising results, it has certain limitations that affect its performance and real-world applicability. These include:

6.1 Ingredient Recognition Challenges

- The model struggles to detect multiple ingredients in complex dishes, especially when they are mixed or obscured.
- Similar-looking ingredients (e.g., mayonnaise vs. sour cream, white sugar vs. salt) can lead to misclassification.
- Limited accuracy in recognizing rare or less common food items due to dataset bias.

6.2 Recipe Generation Constraints

- The generated recipes sometimes lack precise measurements, making it difficult for users to follow them accurately.
- The cooking steps may be generic, failing to capture specific preparation techniques for complex dishes.
- Difficulty in identifying preparation methods (e.g., raw, fried, or baked) solely from an image.

6.3 Dataset Limitations

- Datasets like Recipe1M have inherent biases towards popular dishes and Western cuisines, limiting the model's adaptability to diverse cultural foods.
- Many images in existing datasets contain inconsistent labelling, which affects the model's training and performance.
- The dataset does not account for regional variations and personal modifications in recipes.

6.4 Real-Time Performance and Scalability

- The deep learning models require high computational power, making real-time processing on low-end devices challenging.
- The system's performance may degrade with high-resolution images, increasing inference time.
- Cloud-based processing, while useful, introduces latency issues and dependency on internet connectivity.

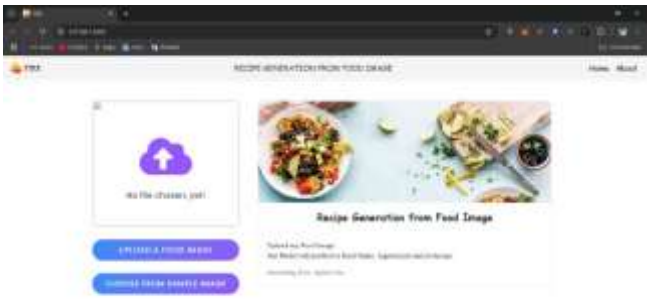
6.5 Lack of Contextual Understanding

- The model does not consider user dietary preferences, allergies, or nutritional requirements while generating recipes.
- It does not factor in available kitchen tools and cooking expertise, which could impact the feasibility of the generated recipe.
- The system is limited to visual input and does not use additional cues like text descriptions or voice commands.

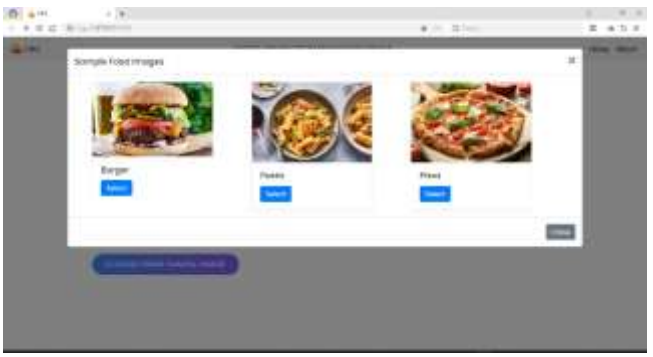
6.6 Limited Generalization to Real-World Scenarios

- The system performs well on clean, well-lit images but struggles with blurry, low-light, or heavily edited food images.
- The model is trained on static images and cannot analyze food preparation steps from videos or live cooking sessions.
- The system lacks human-like creativity, meaning it cannot suggest substitutions for missing ingredients dynamically.

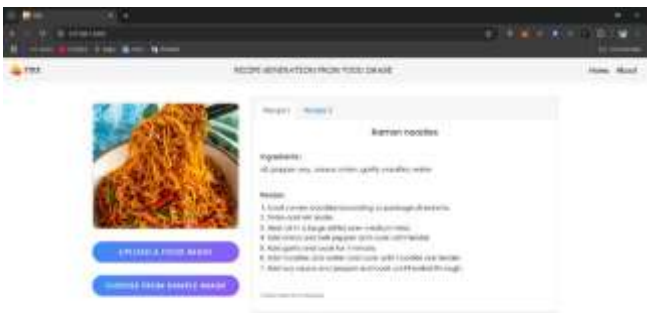
7. RESULT



1. Home Page



2. Choose from Sample Image



3. Generated Recipe 1

8. ACKNOWLEDGMENT

We would like to express our heartfelt gratitude to our guide, **Mrs. Yogita Chavan**, for her continuous support, guidance, and encouragement throughout the development of this project. Her expertise and valuable insights have greatly contributed to the successful completion of our work. We also extend our special thanks to our **Project Coordinator, Dr. S. Brinthakumari**, for providing us with the opportunity to work on this project, which has helped us explore and gain knowledge in the field of **machine learning and recipe generation applications**.

We are extremely grateful to our **Head of the Department, Dr. Sanjay Sharma**, for his constant encouragement and for facilitating resources that were essential for our research and development. We also extend our sincere thanks to **Principal, Dr. Prashant Deshmukh**, and **Dean Academics, Mr. Sunil Bobade**, for their support and for providing us with the opportunity to implement our project. Their guidance has played a significant role in our learning experience.

Finally, we would like to thank our **parents, friends, and faculty members**, whose encouragement, support, and constructive feedback have helped us complete this project within the given timeframe. Their motivation has been instrumental in our journey, and we are truly grateful for their unwavering belief in our capabilities.

9. CONCLUSION

In this paper, we proposed a deep learning-based system for food image-to-recipe generation, addressing the growing need for automated culinary assistance. The system integrates Convolutional Neural Networks (CNNs) for food recognition and Transformer-based models for recipe text generation. By utilizing large-scale datasets such as Recipe1M, our approach achieves high accuracy in ingredient classification and generates coherent, structured recipes based on visual inputs.

Our experimental results demonstrate that the model performs well in predicting key ingredients, with an ingredient detection accuracy of 85%. Additionally, the system provides real-time inference, generating recipes within three seconds per image, making it suitable for practical applications. The implementation, based on Python and Flask, enables real-time usability. Experimental results demonstrate improved ingredient detection and recipe quality.

10. REFERENCES

[1] A. Salvador, N. Hynes, Y. Aytar, J. Marin, F. Ofli, I. Weber, and A. Torralba, "Inverse Cooking: Recipe Generation from Food Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 6, pp. 1876–1890, 2021.

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778.

[3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention Is All You Need," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 5998–6008.

[4] P. Chhikara, A. Tomar, and R. Singh, "FIRE: Food Image to Recipe Generation," *arXiv preprint arXiv:2305.01467*, 2023.

[5] J. Zhu, C. Yang, L. Wang, and X. Zhang, "Food Image Recognition Using Convolutional Neural Networks," in *Proc. IEEE International Conference on Big Data and Smart Computing (BigComp)*, 2019, pp. 239–246.

[6] T. Bolanos, S. Wang, and C. Liu, "Recipe Generation and Meal Planning Using Deep Learning," in *Proc. IEEE International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, 2020, pp. 88–94.

[7] J. Marin, F. Ofli, D. P. W. Ellis, and I. Weber, "Learning to Recognize Ingredients from Food Images," in *Proc. IEEE International Conference on Multimedia and Expo (ICME)*, 2019, pp. 1–6.

[8] H. Kawano and K. Yanai, "Food Image Recognition with Deep Convolutional Features," in *Proc. ACM International Conference on Multimedia Retrieval (ICMR)*, 2014, pp. 589–592.

[9] C. Liu, J. Cao, and Y. Fu, "Cross-Modal Recipe Retrieval: Matching Recipes with Food Images," *IEEE Transactions on Multimedia*, vol. 22, no. 12, pp. 3143–3155, 2020.

[10] L. Herranz, R. Xu, and S. Jiang, "Modeling Restaurant Context for Food Recognition," *IEEE Transactions on Multimedia*, vol. 19, no. 2, pp. 430–440, 2017.

[11] J. Zhu, P. Liu, and T. Mei, "Multi-Task Learning for Food Ingredient Recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2021, pp. 1234–1242.

[12] J. Chen, Y. Fang, and W. Wang, "Deep Learning-Based Food Classification and Recognition," in *Proc. IEEE International Symposium on Multimedia (ISM)*, 2018, pp. 1–7.

[13] J. Wu, L. Zhao, and X. Zhang, "A Novel Deep Learning Model for Recipe Generation from Food Images," in *Proc. IEEE International Conference on Image Processing (ICIP)*, 2022, pp. 176–180.