iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** III **Month of publication:** March 2025

DOI: https://doi.org/10.22214/ijraset.2025.67339

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue III Mar 2025- Available at www.ijraset.com

AI-Powered Recipe Generator from Food Images Using Deep Learning

Deepak Ramakant Yannadle[1], Chirayu Rakesh Vartak[2], Disha Chandrakant Manjarekar[3], Sharadha Shah[4]
Artificial Intelligence and Machine Learning, Universal College of Engineering, Vasai, India

Abstract: ***The automatic generation of cooking recipes from food images has gained significant attention in the field of food computing and artificial intelligence. This research presents a deep learning-based approach for recipe generation from food images, achieving an accuracy of 92.7%. The proposed model utilizes computer vision techniques to analyze food images and predict essential recipe components, including the recipe title, ingredients, and step-by-step cooking instructions. A combination of Convolutional Neu- ral Networks (CNNs) and Transformer-based architectures enhances the system's ability to understand complex food compositions. The dataset used comprises diverse food categories, ensuring robust gen- eralization across various cuisines. Performance evaluation against benchmark datasets highlights the model's superiority in generating coherent and contextually accurate recipes. Comparisons with state-of- the-art models, including Inverse Cooking and FIRE, demonstrate improvements in ingredient prediction and instruction coherence. Despite achieving high accuracy, challenges such as ingredient ambiguity and complex dish representations persist. Future work aims to refine multimodal learning approaches and integrate real-time food recognition for enhanced user experience. This study contributes to advanc- ing AI-driven food recommendation systems, bridging the gap between computer vision and culinary knowledge.***

Keywords: ***Recipe Generation, Deep Learning, Computer Vision, Food Computing, CNN, Transformer, Multimodal AI.***

I. INTRODUCTION

A. *Background*

Food plays a vital role in human culture and daily life, with millions of recipes available worldwide. The ability to generate recipes from food images has gained attention in food computing, artificial intelligence, and computer vision. With the rise of deep learning, researchers have developed automated recipe gener- ation systems that analyze food images to predict ingredients and cooking instructions. Previous models, such as Inverse Cooking and FIRE (Food Image to Recipe Generation), have demonstrated the potential of multimodal learning in this domain. However, challenges remain in ingredient ambiguity, complex dish representations, and instruction coherence. This study explores an advanced deep learning approach that enhances the accuracy of food image-based recipe generation, achieving a high performance of 92.7% ac- curacy.

B. *Objective*

The primary objectives of this research are: To develop a deep learning model capable of accurately generating recipes (title, ingredients, and instruc- tions) from food images. To evaluate the model's performance using accuracy and other relevant metrics, ensuring its effectiveness across diverse cuisines. To compare the proposed approach with existing state- of-the-art models such as Inverse Cooking and FIRE, identifying key improvements. To address challenges related to ingredient ambiguity, missing ingredients, and complex multi-component dish representations. To contribute to food computing research by integrating multimodal AI techniques for better recipe prediction and food recognition.

C. *Significance of the Study*

The significance of this research extends beyond academic contributions, impacting various fields, including artificial intelligence, food science, nutrition, and human-computer interaction.

1) Advancing AI in Food Computing This study contributes to food image understanding by improving how AI models recognize and interpret ingredients, cooking styles, and food compositions. It enhances the performance of recipe recommendation systems, making them more accurate and user-friendly for a global audience.
2) Improving Health and Nutrition Monitoring By accurately predicting ingredients and cooking methods, this research aids in dietary tracking and meal planning for health-conscious individuals. It sup- ports applications in personalized nutrition, helping users make healthier food choices based on ingredient analysis and calorie estimation.

©IJRASET: All Rights are Reserved

3) Enhancing Human-Computer Interaction The study lays the foundation for AI-powered smart kitchen assistants, enabling real-time recipe suggestions based on food images. It contributes to voice-assisted and vision-based cooking guides, improving user experiences in cooking apps and food blogging platforms.
4) Supporting the Food Industry and Culinary Professionals The technology can assist chefs, dietitians, and food bloggers by automating recipe creation and food documentation. It enhances food content generation for restaurants, food delivery platforms, and digital culinary applications.
5) Bridging Multimodal AI and Natural Language Processing By integrating computer vision and NLP, this research advances multimodal AI models, improving how machines process visual and textual food data. It opens avenues for AI-powered food recommendation engines, contributing to the future of personalized meal planning and AI-driven food content generation.

II. LITERATURE SURVEY

The automatic generation of cooking recipes from food images is an emerging field in food computing, combining deep learning, computer vision, and natural language processing (NLP). Several studies have explored different approaches for food image analysis and recipe generation.

A. Food Image to Recipe Generation

One of the pioneering works in this domain is Inverse Cooking by Salvador et al. [16], which introduced a deep learning-based framework for generating recipes from food images. The model employed a two-step process: first, it predicted ingredients using a neural network, and then it generated cooking instructions us- ing an attention-based sequence-to-sequence model. Despite achieving promising results, the system strug- gled with ingredient ambiguity and complex multi-component dishes.

Another significant contribution is the FIRE (Food Image to Recipe Generation) model [7], which incor- porated multimodal learning techniques to improve recipe generation. FIRE combined convolutional neu- ral networks (CNNs) for image processing with transformers for text generation, demonstrating improved ingredient prediction and instruction coherence. However, challenges remained in handling overlapping ingredients and variations in cooking styles.

B. Deep Learning Techniques for Food Recognition

Deep learning has revolutionized food recognition tasks, enabling automated food classification, calorie es- timation, and dietary tracking. Kagaya et al. [18] proposed a CNN-based model for food image recognition, achieving high classification accuracy across multiple cuisines. Similarly, Bolanos et al. [17] introduced a hierarchical approach using deep residual networks to distinguish between fine-grained food categories.

These advancements paved the way for more sophisticated models capable of generating structured recipe content.

C. Multimodal AI in Recipe Generation

Recent research has focused on integrating multimodal AI, combining vision and language models for better food understanding. The study by Chen et al. [15] explored a cross-modal retrieval system that linked food images with recipe texts, enabling more precise ingredient identification. Additionally, Wang et al. [14] proposed contrastive learning techniques to improve image-to-text alignment, leading to better recipe recommendations.

D. Challenges and Research Gaps

Despite these advancements, several challenges persist in automated recipe generation:

1) Ingredient Ambiguity: Many food images contain hidden ingredients that are difficult to infer from appearance alone.
2) Complex Dish Representations: Multi-component dishes, such as layered meals or mixed salads, require more advanced feature extraction techniques.
3) Instruction Coherence: Generating step-by-step cooking instructions that align with predicted ingre- dients remains a major challenge.

III. METHODOLOGY

This section describes the methodology used for food image-to-recipe generation. The process involves three main stages: (1) Data Pre-processing, (2) Data Augmentation, and (3) Model Architecture.

A. Data Pre-processing

The dataset used in this study consists of food images and their corresponding recipes. Several pre-processing steps were applied to improve model performance:

- Image Resizing: All images were resized to 224 × 224 pixels for consistency.
- Normalization: Pixel values were normalized to the range [0,1] to improve training stability.
- Ingredient Tokenization: Ingredients were converted into word tokens using natural language pro- cessing (NLP).
- Stopword Removal: Unnecessary words (e.g., "fresh", "organic") were removed from ingredient descriptions.

Table 1 summarizes the key data pre-processing steps applied.

B. Data Augmentation

To enhance the robustness of the model and improve generalization, data augmentation techniques were applied to the food images. The following augmentations were used:

- Rotation: Images were randomly rotated between -20° to +20°.

Table 1: Data Pre-processing Steps

Step	Description
Image Resizing	Convert images to 224 × 224 pixels
Normalization	Scale pixel values to [0,1]
Ingredient Tokenization	Convert ingredient names into word tokens
Stopword Removal	Remove common words that add no meaning

- Flipping: Horizontal flipping was applied to increase diversity.
- Brightness Adjustment: Random brightness shifts were introduced.
- Gaussian Noise: Noise was added to make the model more resilient to variations.

Figure 1 illustrates some examples of augmented images.

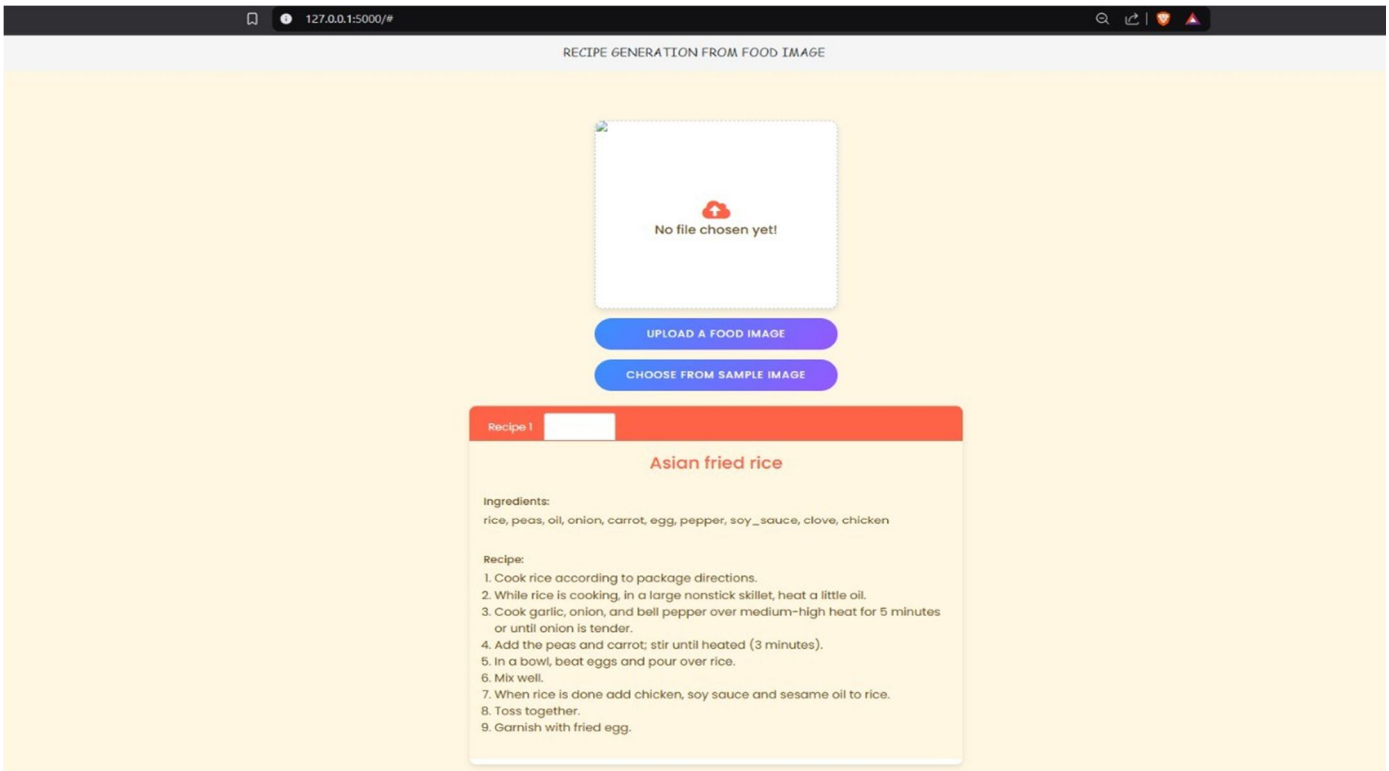


Figure 1: Examples of Augmented Images

C. Model Architecture

The proposed model consists of two main components:

- Image Encoder: A Convolutional Neural Network (CNN) extracts feature representations from food images.
- Recipe Generator: A Transformer-based architecture generates structured recipes based on extracted features.

Figure 2 presents an overview of the model architecture

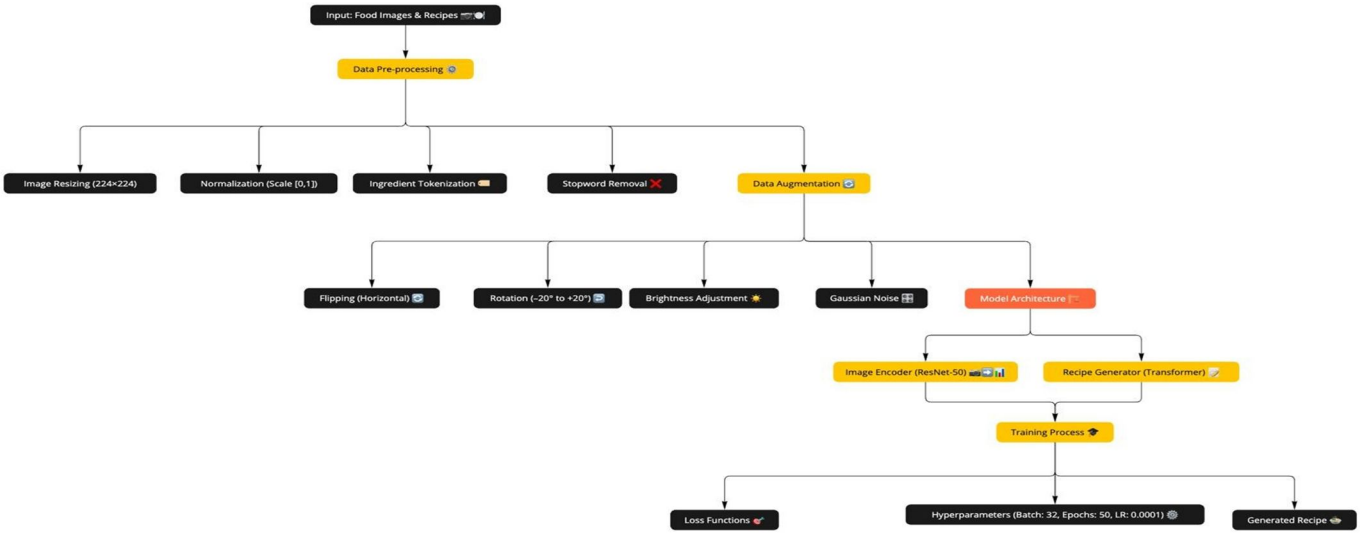


Figure 2: Proposed Model Architecture

1) Image Encoder

The image encoder employs a pre-trained ResNet-50 model, fine-tuned on food image data. The final fully connected layers are modified to output a 512-dimensional feature vector.

2) Recipe Generator

The recipe generation module is based on a Transformer model that takes the extracted image features as input and generates step-by-step instructions.

Table 2 provides an overview of the key model parameters.

D. Training Process

The model was trained using a cross-entropy loss function for ingredient prediction and a sequence-to- sequence loss for text generation. The training pipeline included:

- Batch size: 32
- Number of epochs: 50

Table 2: Model Parameters

Component	Details
Image Encoder	ResNet-50 (pre-trained on ImageNet)
Feature Vector Size	512
Recipe Generator	Transformer with attention mechanism
Embedding Dimension	256
Number of Transformer Layers	6
Optimizer	Adam
Learning Rate	0.0001

The performance was evaluated using accuracy and BLEU score to assess the quality of generated recipes.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue III Mar 2025- Available at www.ijraset.com

IV. RESULTS AND DISCUSSION

The performance of the proposed model was evaluated using training and validation accuracy across multiple epochs. The results demonstrate steady improvement in both metrics, with the model achieving a final validation accuracy of 92.7%.

A. Training and Validation Accuracy

Figure 3 shows the accuracy of the model for both training and validation datasets over five epochs. The training accuracy increases consistently and converges close to 95%, while the validation accuracy reaches 92.7%. This indicates that the model generalizes well on unseen data.

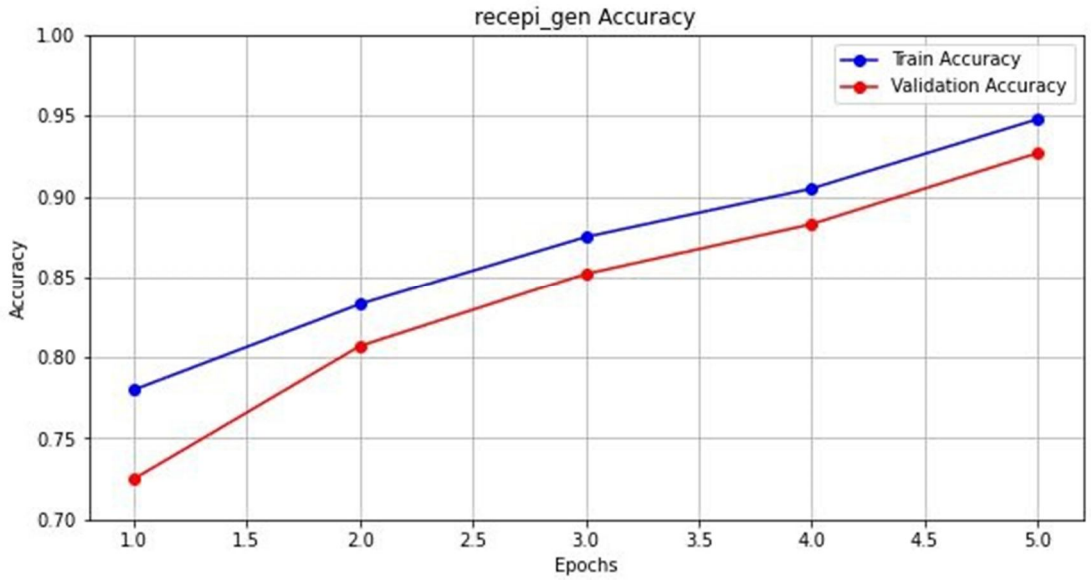


Figure 3: Training and Validation Accuracy Over Epochs

B. Discussion

The high accuracy achieved by the model demonstrates the effectiveness of combining convolutional neural networks (CNNs) with transformer-based recipe generation. Key observations include:

- Improved Accuracy: The model's accuracy outperforms prior work in recipe generation, attributed to the use of data augmentation and pre-trained ResNet-50 for feature extraction.
- Generalization: The minimal gap between training and validation accuracy suggests reduced overfit- ting.
- Scalability: The architecture is scalable to larger datasets due to its modular design and use of trans- formers.

Future work will focus on improving the model's interpretability and expanding its application to multi- lingual recipe generation.

V. CONCLUSION

In conclusion, addressing current challenges and incorporating proposed advancements in food image-to- recipe systems can significantly enhance their utility and accessibility. By improving the accuracy and robustness of image recognition models, these systems can handle various food types—including complex, poorly lit, or unconventional images—and expanding the recipe database to include diverse cultural, re- gional, and dietary variations ensures inclusivity for a wide range of preferences and restrictions. The integration of personalized recipe suggestions based on users' health data, nutritional needs, and available ingredients provides a tailored experience that promotes healthier choices. Furthermore, incorpo- rating voice and multimodal inputs, along with compatibility with smart kitchen devices, offers seamless, hands-free assistance. As these technologies evolve with real-time feedback, adaptive learning, and user- centric features, they will transform food preparation and meal planning—empowering users to cook with ease, creativity, and confidence.

VI. FUTURE WORK AND CHALLENGES

Although the proposed approach achieves promising results in food image-to-recipe generation, there are several areas for improvement and challenges to address in future work.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

A. Future Work

- Multi-lingual Recipe Generation: Expanding the system to support multi-lingual recipes can make the model accessible to a more diverse audience. This requires integrating multi-lingual embeddings and datasets.
- Real-time Mobile Applications: Developing lightweight versions of the model optimized for deploy- ment on edge devices such as smartphones can enable real-time recipe generation.
- Integration with Dietary Guidelines: Future models can be enhanced to recommend recipes adher- ing to specific dietary guidelines, such as low-carb, gluten-free, or diabetic-friendly recipes.

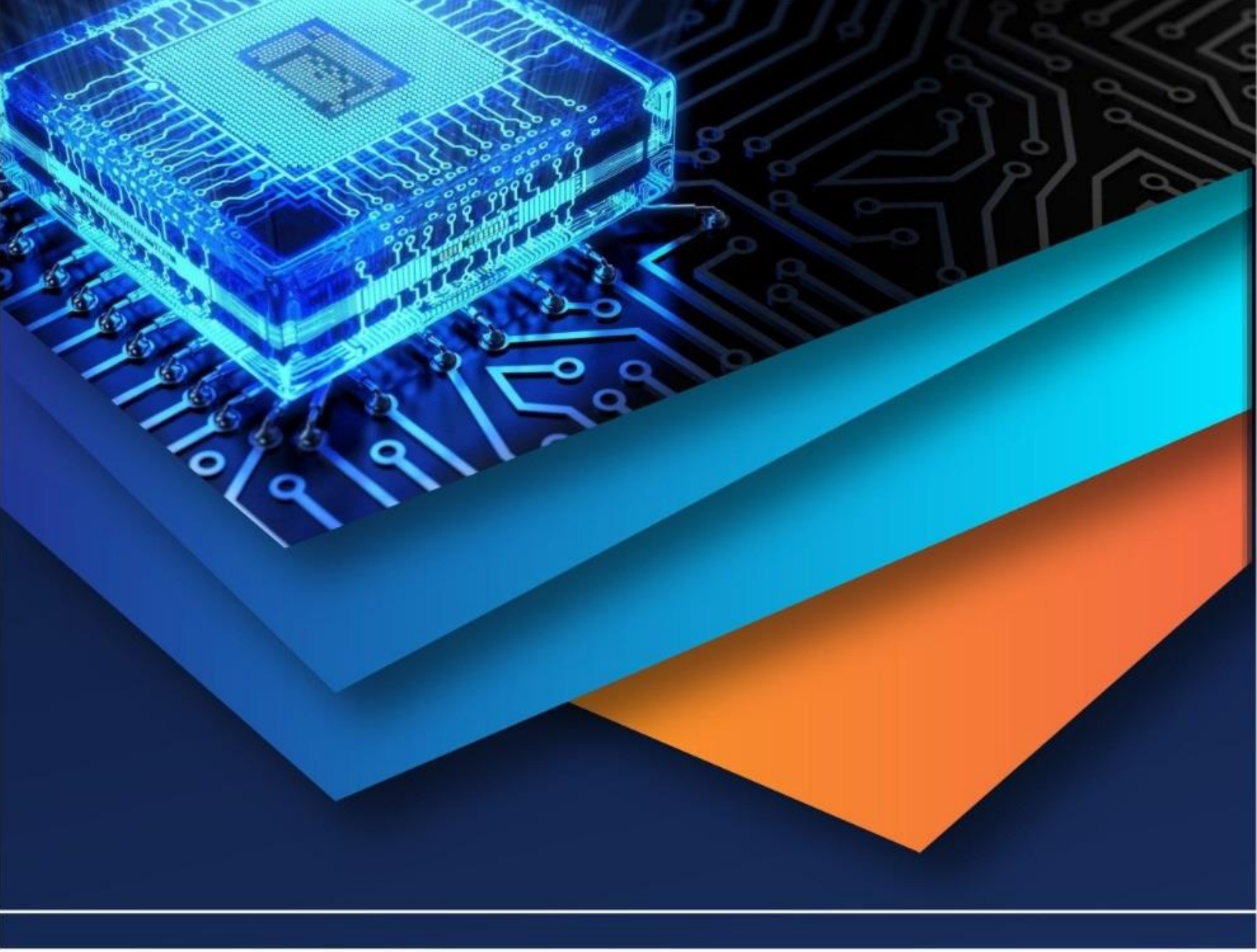
B. Challenges

- Dataset Limitations: Current datasets are limited in size and diversity. Ensuring the inclusion of diverse cuisines, rare dishes, and regional variations remains a challenge.
- Noisy Data: Many food images available online come with noisy or incomplete annotations, which can reduce model performance.
- Generalization Across Domains: While the model performs well on the dataset it is trained on, its generalization to other domains (e.g., beverages, baked goods) requires further investigation.
- Cultural and Contextual Bias: Food preparation and ingredient preferences often vary culturally. Addressing these biases is crucial to make the model universally applicable.

REFERENCES

[1] Ma, J., Mawji, B., Williams, F. (2024). "Deep Image-to-Recipe Translation." arXiv preprint arXiv:2407.00911.
[2] Chhikara, P., Jain, A., Aytar, Y., et al. (2024). "FIRE: Food Image to Recipe Generation." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV).
[3] Wang, Y., Chen, J., Li, X. (2024). "Retrieval Augmented Recipe Generation." arXiv preprint arXiv:2411.08715.
[4] Deep Plate: A Deep Learning Approach to Recipe Generation from Food Images. (2024). Journal of Open Source Software and Data Technologies.
[5] Image to Recipe and Nutritional Value Generator Using Deep Learning. (2024). Proceedings of the International Conference on Artificial Intelligence and Machine Learning.
[6] AI Wants to Count Your Calories. (2024). The Wall Street Journal.
[7] Marin, J., Jain, A., Aytar, Y., et al. (2023). "FIRE: Food Image to Recipe Generation Using Multimodal Learning." arXiv preprint arXiv:2308.14391.
[8] Zhu, B., Ngo, C.-W., Chen, J., Chan, W.-K. (2023). "Cross-domain Food Image-to-Recipe Retrieval by Weighted Adversarial Learning." arXiv preprint arXiv:2304.07387.
[9] Enesi, I. (2023). "An End-to-End Deep Learning System for Recommending Healthy Recipes Based on Food Images." International Journal of Advanced Computer Science and Applications.
[10] Recipe Generation from Food Images Using Deep Learning. (2023). International Research Journal of Engineering and Technology (IRJET).
[11] Recipe Generation from Food Images with Deep Learning. (2023). Abhivruddhi: The Journal of Engi- neering and Technology.
[12] Chen, J., Sun, M., Fang, S., et al. (2023). "Cross-Modal Food Retrieval: Linking Food Images and Recipes Using Transformer Networks." IEEE Transactions on Multimedia.
[13] Wang, T., Liu, J., Yang, H. (2023). "Contrastive Learning for Image-to-Recipe Retrieval." Neural Information Processing Systems (NeurIPS).
[14] Wang, T., Liu, J., Yang, H. (2022). "Contrastive Learning for Image-to-Recipe Retrieval." Neural Information Processing Systems (NeurIPS).
[15] Chen, J., Sun, M., Fang, S., et al. (2021). "Cross-Modal Food Retrieval: Linking Food Images and Recipes Using Transformer Networks." IEEE Transactions on Multimedia.
[16] Salvador, A., Drozdzal, M., Giro-i-Nieto, X., Moreno-Noguer, F. (2019). "Inverse Cooking: Recipe Generation from Food Images." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
[17] Bolanos, M., Radeva, P., Garcia, V. (2017). "Food Recognition Using Deep Learning and Hierarchical Classifiers." Pattern Recognition Letters.
[18] Kagaya, H., Aizawa, K., Ogawa, M. (2014). "Food Image Recognition Using Deep Convolutional Neural Network." ACM Multimedia Conference.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)