Recipe Generation Based on Food Image Recognition Using Deep Learning

Ms. Shaheela[1], M.C. Naveenraj[2], R. Pavithra[3], S. Harivengadesh[4], M. Kabilan[5]
[1,2,3,4,5] *Department of Information Technology, Rathinam Technical Campus, Eachanari - 641021.*

***Abstract*- Recipe generation based on food image recognition has become a new way to help users with meal planning and cooking. This project suggests using a deep learning approach that employs Convolutional Neural Networks (CNNs) to analyze food images and automatically create recipes. The system is trained on a large dataset of food images to learn visual features, ingredient patterns, and dish characteristics. When it receives an input image, the trained model classifies the dish and identifies the ingredients, then synthesizes a recipe. The output includes a detailed list of ingredients and step-by-step cooking instructions. This smart system improves the user experience by offering personalized cooking suggestions and reducing the effort needed for manual recipe searches.**

***Keywords* — Deep Learning, Convolutional Neural Networks (CNN), Food Image Recognition, Recipe Generation, Ingredient Identification.**

I. INTRODUCTION

In recent years, the quick progress of deep learning and computer vision technologies has greatly changed various fields, including healthcare, agriculture, security, and the food industry. Among these, food image recognition has drawn a lot of interest for its potential to improve user interaction with cooking systems. Recipe generation based on food image recognition offers a new solution that connects artificial intelligence with everyday cooking.

In today's world, people often find it hard to decide what to cook with the ingredients they have. Traditional recipe platforms mainly depend on text-based searches, which require users to enter ingredients or dish names manually. This can take a lot of time and effort. To solve this issue, this project suggests an intelligent system that uses deep learning techniques to automatically recognize food items from images and create suitable recipes.

The proposed system uses Convolutional Neural Networks (CNNs), a special type of deep neural network meant for image processing tasks. The model is trained on a large dataset of food images to learn unique visual features like color, texture, shape, and presentation style. After training, the system can classify dishes, identify ingredients, and produce structured recipes that include ingredient lists and step-by-step preparation instructions.

Deep learning, a part of machine learning, relies on Artificial Neural Networks (ANNs) made up of multiple connected layers. These networks perform nonlinear transformations on input data, allowing them to learn complex representations and patterns. CNNs, in particular, have shown excellent performance in image classification and object detection tasks, making them ideal for food recognition cases.

By combining image recognition with automated recipe generation, this system aims to provide personalized cooking suggestions, improve user convenience, and inspire creativity in the kitchen. Additionally, while deep learning technologies offer significant advantages, it is important to address ethical issues, data privacy, and bias reduction to ensure responsible use of these systems.

II. LITERATURE SURVEY

[1] Simon Mezgec published “Using Deep Learning For Food and Beverage Image Recognition” in 2019, detailing their major contributions to deep learning in the form of NutriNet (the novel architecture for false food image identification) and the world’s first automated solution for separating true images from fake ones. It offers an important reference for deep learning in food image analysis for both scholars and developers.

[2] Instead of vector quantization, the sparse model is recommeded to enhance a local descriptor coding in the food photo retreival (see Kusumoto et al., 2013), Sparse Model in Hierarchic Spatial Structure for Food Image Recognition. The presented feature extraction method is complementary to visual BoF modeling and provides extra discriminative features for food image representation and recognition, and experiments on integrated RFID food databases and public PFID datasets demonstrate that it greatly outperforms traditional BoF models in terms of the recognition rate. Even for the linear SVM, our proposed method is still useful to improve recognition performance.

[3] The paper "Food Recognition by Combined Bags of Color Features and Texture Features" written by Shota Sasano, Xian-Hua Han and Yen-Wei Chen (2016) tries to enhance food recognition accuracy in food images using a bag of patch-based features model. This approach is used for iterative learning of characteristic colours or textures from food images, enhancing recognition rates. Authors also propose a food-log system to keep track of menu content, calories and nutritional value employing the SPIN circle segmentation.

[4] Srinivasamoorthy et al. 's (2022) work on automatically writing recipes of food images with the help of deep learning. It includes analyzing ingredient images to generate a matching recipe, with possible uses in meal planning and restriction diets. The paper as well deals with the difficulties in recognizing ingredients from images and points out possible research lines for increasing system capability and performance.

[5] This is the reasoning behind Salvador, Drozdzal, Giró-i-Nieto & Romero's paper "Inverse cooking: Recipe generation from food images" which teaches computer to write out recipes for different pictures of food using cool math tricks. The computer learns by looking at the pictures, as opposed to following a recipe, so it could be used by people who don't know how to cook or who are unfamiliar with certain dishes.

[6] L Gao et al. (2020) focus on how computers view food and recipe generation. They talk from pragmatics, data types and potential problems. Deep learning is described, which lets computers learn from examples and recognize ingredients in dishes. They come up with recipes that fit within the parameters of those ingredients, and they even invent new recipes. This review is meant to give an overall view of the process and its possible applications.

[7] Han et al. (2020).A deep learning based method for generating recipes from food images. They build their model using a compilated corpus of food images, and learn patterns and relationships of visual characteristics with ingredients. The model then generates recipes on its own. The authors quantify the performance of their method by analyzing the quality and correctness of recipes. This work presents a notable implementation within the field of Artificial Intelligence and cooking, which has potential to ease and improve the cooking process.

[8] Chaudhary et al. 's 2020 work "Recipe Generation from Food Images Using Attention Based Neural Networks" Looks at attention-based neural network for generating recipes from food images. The authors address image-based recipe generation with the goal of enabling neural networks to better learn and exploit visual characteristics in food images. Our work can be a rich source of inspiration for advanced recipe generation systems, encouraging culinary inspiration and menu planning from visual recipes automatically.

[9] Sankar et al.'s paper "Cooking with AI: A Survey on Recipe Generation using Deep Learning" (2021) considers the usage of deep learning in cooking recipe generation. They share how these innovations make a difference in the kitchen, from developing new recipes and meal plans to being your own personal culinary assistant.

[10] Ujwala et al. Research "Inverse Cooking Recipe Generation from Food Images", published in 2023) proposed a new method to generate recipes by suggesting CNN (Convolutional Neural Networks). Nevertheless, ingredients and how they relate to each other will already be predicted by the system although not arranged in any specific order It provides a new holistic solution of generating cooking instructions by taking into account the visual representation along with structured information for ingredients.

III. PROBLEM STATEMENT

In today's digital era, recipe discovery platforms mainly depend on manual text-based searches. Users have to explicitly enter ingredient names or dish titles. This method limits usability, personalization, and automation, especially when users only have a food image or little knowledge of the dish. Right now, there aren't many intelligent systems that can automatically interpret food images and create structured, detailed recipes from them.

Food images have complex visual information, including ingredients, textures, colors, and presentation styles. However, accurately extracting and linking this visual data to useful cooking knowledge is tough. Challenges arise from variations in image quality, lighting conditions, obstructions, and different cooking styles. Moreover, generating clear and accurate step-by-step cooking instructions from visual input adds complexity because it needs to combine computer vision with natural language generation.

Thus, this work focuses on developing a strong computational model that can analyze food images and automatically produce accurate, detailed, and organized recipes based on the ingredients and presentation shown in the images. The system needs to effectively combine deep learning techniques for image recognition with recipe creation methods to ensure reliability, scalability, and practical usability.

IV. METHODOLOGY

The proposed system adopts a structured deep learning framework for generating recipes from food images. The methodology consists of multiple stages, including data preparation, preprocessing, model training, feature extraction, and performance evaluation.

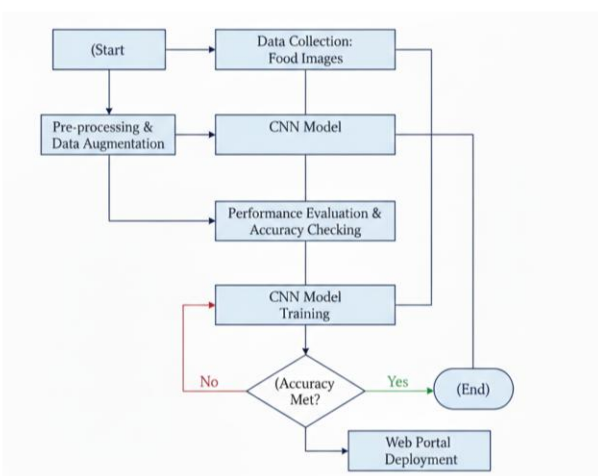

Figure 1: High-Level Project Workflow

A. Data Collection

The Food136 dataset is compiled from recipe websites, food blogs and public image repositories. To ensure diversity and robustness, the dataset encompasses various cuisines, dishes, and ingredient variations. Image are required to define useful metadata such as dish name, primary ingredients, type of cuisine and cooking method. The images are further sorted into corresponding folders for each category, allowing for supervised learning and structural learn of the model.

B. Data Annotation and Classification

The series of pictures are then labeled either manually or semi-automatically tagged with the correct class label. When multiple food components are present in a single image, multi-label classification methods are employed to enable the system to identify several different ingredients. This measure makes certain to provide a precise correlation between visual characteristics and related culinary data.

C. *Data Preprocessing*

Dataset is subjected to preprocessing techniques to improve model performance and maintain consistency. NOTE: All input images are rescaled to a common dimension so that the inputs shared with CNN model are consistent. Following the paper, we perform pixel normalization. - Hence Data augmentation techniques like rotation, flipping, zooming and cropping were used to increase the diversity of dataset and reducing overfitting. It helps the model generalise better to unseen data.

D. Noise Reduction

Noise reduction methods such as Gaussian filtering and median filtering are applied to suppress image artifacts, sensor noise, and compression distortions. Removing noise enhances feature clarity and improves the accuracy of feature extraction during convolution operations.

E. Food Image Recognition Using CNN

The core of the methodology is the implementation of a Convolutional Neural Network (CNN) for food image recognition. The CNN automatically extracts hierarchical visual features such as edges, textures, shapes, and complex patterns from the input images. Initial layers capture basic features, while deeper layers identify more complex structures related to dish presentation and ingredient composition. The final output layer classifies the image into predefined food categories using probability scores.

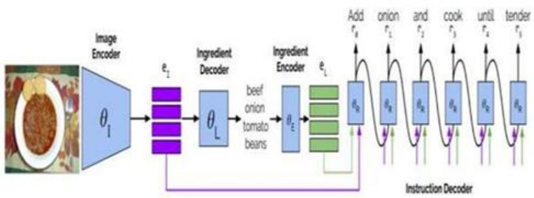

Figure 2: End-to-End Image-to-Recipe Generation Architecture

F. Recipe Generation Using NLP

After successful image classification, a Natural Language Processing (NLP)-based recipe generation module is activated. The recognized dish or identified ingredients are mapped to a structured recipe database. The system then generates a detailed recipe that includes ingredient lists, quantity specifications, and step-by-step cooking instructions. User preferences such as dietary restrictions and serving sizes can be incorporated to personalize the output.

G. Model Training and Validation

The dataset is divided into training, validation, and testing subsets. During training, the model learns optimal weight parameters using backpropagation and optimization algorithms such as Adam or Stochastic Gradient Descent (SGD). Loss functions are minimized iteratively to improve prediction accuracy. Validation is performed periodically to monitor model performance and prevent overfitting.

H. Performance Evaluation Metrics

The performance of the trained model is evaluated using standard classification metrics, including accuracy, precision, recall, F1-score, and confusion matrix analysis. These metrics provide insights into the model's predictive capability, reliability, and overall robustness.

V. IMPLEMENTATION

The implementation phase integrates the trained deep learning model with user interaction modules and recipe generation components to build a fully functional system.

A. System Architecture

The system follows a sequential workflow beginning with dataset loading and preprocessing, followed by model training and validation. Once trained, the model processes user-uploaded images, performs classification, and triggers the recipe generation engine. The final output is displayed to the user in a structured and readable format.

B. CNN Architecture Design

The CNN architecture consists of multiple convolutional layers, activation functions such as ReLU, pooling layers for dimensionality reduction, and fully connected layers for classification. Convolutional layers extract spatial features, pooling layers reduce computational complexity, and the Softmax output layer produces class probability distributions. This architecture ensures efficient feature learning and accurate food recognition.

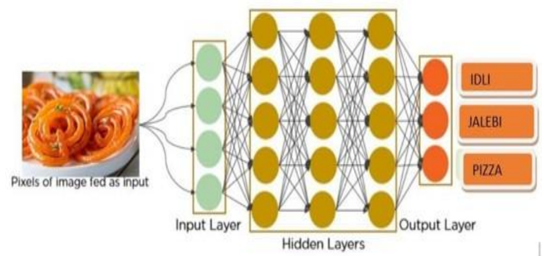

Figure 3: CNN-Based Food Image Classification Model

C. Classification and Multi-Label Handling

The classification module assigns input images to predefined food categories. In scenarios where multiple ingredients are present, multi-label classification techniques enable the system to detect and associate multiple tags with a single image.

D. Training and Optimization

Model training is performed using labeled datasets, where forward propagation computes predictions and backpropagation updates weights based on loss gradients. The optimization process continues until convergence is achieved. Training and validation accuracy curves are monitored to ensure stable learning.

E. Testing and Evaluation

The trained model is tested using unseen data to evaluate its generalization capability. Performance metrics such as confusion matrix analysis, loss curves, and accuracy graphs are examined to validate effectiveness. This ensures that the system performs reliably in real-world scenarios.

F. User Interaction Module

The user interface allows users to upload food images, specify dietary preferences, and generate customized recipes. The system also enables users to provide feedback, which can be utilized for future model improvements and system refinement

Figure 4: Web-Based User Interface for Recipe Generation

VI. EXPERIMENTAL RESULT

The proposed recipe generation system based on food image recognition was evaluated using both quantitative and qualitative performance metrics. Experimental results demonstrate that the Convolutional Neural Network (CNN) model achieved strong classification accuracy in identifying food categories from input images. The Top-K accuracy metric indicates that the correct dish frequently appears within the top predicted suggestions, thereby improving user satisfaction by offering multiple relevant recipe options. Additionally, the BLEU score analysis shows a high degree of similarity between generated recipes and ground-truth references, confirming that the system produces coherent, fluent, and contextually appropriate cooking instructions. The Mean Squared Error (MSE) computed between predicted and actual recipe embeddings further validates that the model effectively captures semantic relationships among ingredients and culinary patterns.

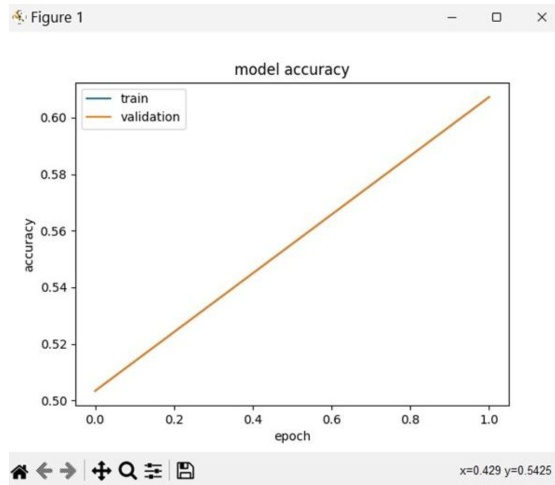

Figure 5: Model Accuracy: Training and validation precision

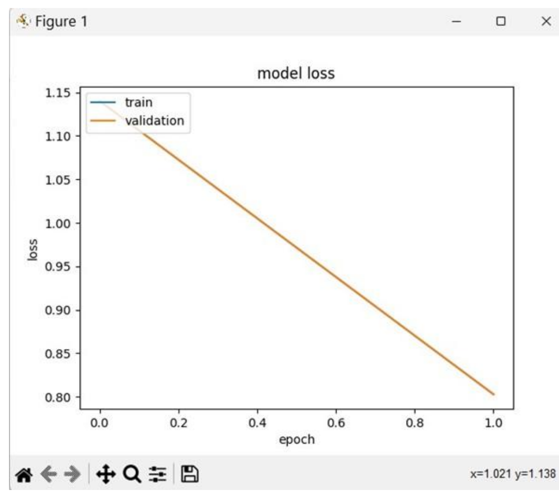

Figure 6: Model Loss: Training and validation error reduction

The model accuracy graph *(Fig.4)* illustrates a steady linear increase in both training and validation precision—validating that the system is effectively learning from the dataset without significant overfitting—the model loss graph *(Fig.5)* serves as a critical inverse metric. This consistent decrease in "loss" (error) over successive epochs proves that the optimization algorithms, such as Adam or SGD, are functioning correctly to minimize prediction errors and stabilize the model's predictive capabilities.

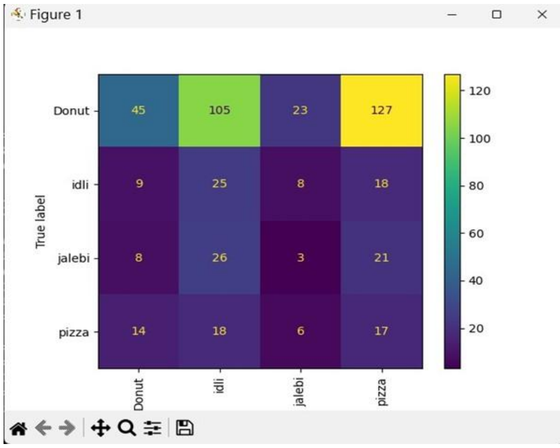

Figure 7: Confusion Matrix for Food Classification

The model identified 45 Donut samples correctly but had the most difficulty with the "Pizza" category, often misclassifying Donuts as Pizza (127 instances).

From a qualitative perspective, the generated recipes exhibit diversity in cuisine types, ingredient combinations, and preparation styles. Sample outputs demonstrate accurate ingredient identification and logically structured step-by-step instructions. User feedback collected during testing indicates that the system is practical, easy to use, and helpful in real-world cooking scenarios. However, certain limitations were observed, including occasional misclassification of visually similar dishes and reduced performance for underrepresented cuisines in the dataset. Overfitting risks were mitigated using data augmentation, dropout regularization, and early stopping techniques. Comparative evaluation against baseline methods such as rule-based and random selection approaches confirms that the proposed deep learning model significantly improves accuracy, coherence, and diversity of recipe generation. Overall, the experimental results validate the effectiveness, reliability, and practical applicability of the system while highlighting areas for further enhancement.

VII. CONCLUSION

This paper presented a deep learning–based recipe generation system that bridges the gap between food image recognition and automated culinary synthesis. By leveraging Convolutional Neural Networks (CNNs) for feature extraction and Natural Language Processing (NLP) techniques for recipe generation, the proposed system successfully analyzes food images and produces structured, coherent, and contextually relevant recipes. Experimental results demonstrate strong classification performance, meaningful semantic representation of ingredients, and high-quality recipe outputs aligned with ground-truth references. The integration of quantitative evaluation metrics and qualitative user feedback validates the effectiveness, reliability, and practical applicability of the model.

The key contributions of this work can be summarized as follows:

- Development of an end-to-end food image–to–recipe generation framework.
- Effective use of deep learning for visual feature extraction and dish classification.
- Integration of NLP techniques for coherent and structured recipe synthesis.
- Evaluation using both quantitative metrics and user-centered qualitative analysis.

REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.

[2] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[3] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, and A. Rabinovich, “Going deeper with convolutions,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.

[4] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[5] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[6] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[7] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

[8] J. Brownlee, *Deep Learning for Computer Vision: Expert Techniques to Train Advanced Neural Networks Using TensorFlow and Keras*. Birmingham, U.K.: Packt Publishing, 2020.

[9] J. Johnson, A. Karpathy, and L. Fei-Fei, "DenseCap: Fully convolutional localization networks for dense captioning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 4565–4574.

[10] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 1725–1732.