

# Deep Plate: A Deep Learning Approach to Recipe Generation from Food Images

Sakshi Londhe<sup>1</sup>, Smita Palnitkar<sup>2\*</sup>, Vishnupriya Kannadkar<sup>1</sup>, Shubhangi Magar<sup>1</sup>

## Abstract

*In the deep learning era, image understanding is advancing in sophistication, encompassing both semantic interpretation and the generation of meaningful image descriptions. To achieve this, deep neural networks must undergo specific cross-model training; these networks must be both simple enough to handle a wide range of inputs and complex enough to encode the fine contextual information associated with the image. An appropriate example of the previously described picture comprehension problem is the conversion of a food image to its preparation instructions. This study introduces a novel method that employs cross-model training of CNN, LSTM, and Bi-Directional LSTM to obtain a compressed representation of cooking instructions from a recipe image. Variable instruction length and the number of instructions per recipe is the biggest problem in this. The system addresses challenges such as varying instructions, differing numbers of instructions per recipe, and the presence of multiple food items in the image. The model successfully shrinks instructions with high similarity to original instructions, especially on Indian cuisine data. This model can be useful for information retrieval and automatic recipe recommendations. These results highlight the potential of deep learning in recipe generation from food images, offering a promising solution for culinary enthusiasts seeking convenient access to cooking instruction.*

**Keywords:** Training, image coding, deep learning, context modeling neural network, information retrieval, CNN, LSTM

## INTRODUCTION

In recent years, interest has surged in developing technology that connects visual information with culinary creativity. Food is crucial for human survival, providing energy, identity, and culture. Food-related activities like cooking, eating, and conversation is integral to our daily lives. It gives us energy and shapes our identity and culture.

As the old saying "we are what we eat" emphasizes the significant role food-related activities like cooking, eating, and discussing it play in our daily lives.

The research addresses the challenge of generating recipes from food images through the application of deep learning techniques. With the increasing popularity of food sharing on social media platforms automated recipe generation can greatly enhance user engagement and culinary creativity. This is precisely the vision behind the "Recipe-Generation-from-Food-Image using Deep Learning" project. The art of cooking has long been

### \*Author for Correspondence

Smita Palnitkar  
E-mail: [smita.palnitkar\\_skncoe@sinhgad.edu](mailto:smita.palnitkar_skncoe@sinhgad.edu)

<sup>1</sup>Student, Department of Electronics & Telecommunication Engineering, Smt. Kashibai Navale College of Engineering (SKNCOE), Savitribai Phule Pune University, Pune, Maharashtra, India

<sup>2</sup>Assistant Professor, Department of Electronics & Telecommunication Engineering, Smt. Kashibai Navale College of Engineering (SKNCOE), Savitribai Phule Pune University, Pune, Maharashtra, India

Received Date: July 03, 2024  
Accepted Date: July 31, 2024  
Published Date: August 08, 2024

**Citation:** Sakshi Londhe, Smita Palnitkar, Vishnupriya Kannadkar, Shubhangi Magar. Deep Plate: A Deep Learning Approach to Recipe Generation from Food Images. Journal of Operating Systems Development & Trends. 2024; 11(2): 15–22p.

---

a cherished human tradition, and with the proliferation of social media and food-centric platforms, food photography has become an integral part of sharing culinary experiences. Yet, this visually captivating medium remains largely untapped in extracting its full potential. The challenge lies in bridging the gap between the rich visual information contained in food images and the structured knowledge required to recreate a dish. Our project's significance extends beyond its ability to conjure recipes from images; it embraces the joy of culinary discovery, empowering home cooks and food enthusiasts to explore new flavors, experiment with diverse cuisines, and unlock their culinary creativity.

Discussing recipe generation from food images using deep learning showcases the cutting-edge technology and its application in a practical, everyday context. In the context of apps, websites, or platforms related to cooking and food, the ability to generate recipes from images enhances user engagement. It can make the user experience more interactive and user-friendly, which is important in consumer-oriented applications. Recipe generation from images can have applications in culinary education. In reports related to culinary schools or training programs, discussing how technology can assist in learning and skill development is important.

In this project for generating a recipe from food image we need to use a deep learning architecture. The text generation process combines Convolutional Neural Networks (CNNs) for image processing with Recurrent Neural Networks (RNNs) or Transformer-based models. Some popular image processing architectures include VGG or Inception. Text generation models like LSTMs, GRUs, or Transformer-based models like GPT-3 can be used for generating recipe text. Pass the food image through the image processing CNN to extract meaningful features and by using RNN the recipe is generated.

## LITERATURE SURVEY

It is common for literature surveys to involve collaboration among multiple researchers, experts, or authors who contribute their knowledge, expertise, and insights to compile a comprehensive review of the literature in a specific field or topic. The collaboration could involve anywhere from one researcher to a team of experts, depending on the depth and breadth of the survey and the resources available for conducting the research.

It is based on the principle of reinforcement learning applied to generation of logical recipe particularly in context of bridging the gap between image and planning [1]. To create a system that can generate structured and logical recipe based on image is a fascinating area for research at intersection of machine learning, Computer vision, and NLP.

"Learning cross-modal embeddings for cooking recipes and food images" [2]: The idea is to structure information or data into multiple level or layers; each layer represents a different level of abstraction, high level structure, Ingredients selection, and Cooking Steps. It will generate not only creative but also well organized and easy to follow recipes.

"Hierarchical attention network for visually-aware food recommendation" [3]: The idea is to break down information or data into several layers, each reflecting a different level of detail, like overall structure, ingredient choices, and cooking procedures. This method will help produce recipes that are both inventive and systematically arranged, making them straightforward to follow.

"Cross-modal recipe retrieval with rich food attributes" [4]: It is based on statistical language model which consists of NLP processing statistical language model. To generate unique recipes, we use an extensive data-driven approach that combines the creativity of a human chef with computational power to develop novel culinary ideas.

"Deep reinforcement learning for dialogue generation" [5]: This paper introduced Inverse cooking: Recipe generation from food images. The system predicts ingredients using a novel architecture,

modeling their dependencies without order, and generates cooking instruction by examining both images and inferred ingredients.

The study explores the use of semantic text-image embeddings in “Cross-modal retrieval in the cooking context: Learning semantic text-image embeddings” [6]. This paper introduces a novel unsupervised learning approach to generate sentence-level tree structures for cooking recipes, which can improve recipe generation and food cross-modal retrieval tasks. In this recipe generation model, we propose using an RNN to generate tree structures from food images and employ the inferred tree to enhance the generation process.

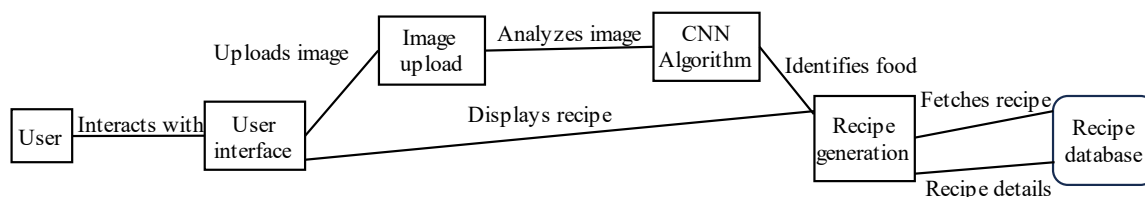
The study employs deep learning techniques to generate cooking recipes from videos, using natural language processing and neural network models to create authentic flavors. The paper, titled “Never fry carrots without chopping” Generating Cooking Recipes from Cooking Videos Using Deep Learning Considering Previous Process” is discussed in the International Journal of Networked and Distributed Computing, (2019) [7]. In this paper, the state of art natural language processing techniques, including word embeddings, are used to help people choose alternative recipe and build language models, N-gram and Neural 6 network model to generate authentic flavor of certain Cuisine style. To improve the performance of model training part of speech of tagging and only kept verbs and nouns cooking instruction steps the exact amount of each ingredient in the recipe was also ignored. Created dataset of 3433 recipes with 9 features.

The study explores “A culinary computational creativity system” [8] that uses deep learning to generate recipe sentences from YouTube-generated cooking videos. A method that adds vector of sentence that already generated in same recipe; to the input of captioning model, data processing is used; to improve accuracy, simplest encoder and decoder is used. Metrics indicate that the proposed method improves accuracy.

The work introduces a recipe generation system that uses an image to predict ingredients from food images, generating a recipe with a title and ingredients. The collection of research articles represents a diverse range of studies addressing critical aspects of recipe generation from food images using deep learning techniques [9]. Major works include the recipe generation system using Convolutional Neural Networks (CNN) and Recurrent Neural networks (RNN) models, which highlights the progress of automatically generating culinary recipes from food images, Additionally, research on feature recognition by Bayesian networks and Hidden Markov Models (HMMs) has significantly improved feature recognition from dish images. Several studies emphasize the importance of recipe generation system in facilitating cooking and meal preparation for individuals, and they discuss challenges and methodologies. Further explorations delve into cuisine-specific recipe generation, such as Italian cuisine, and its real time applications, offering innovative techniques for different culinary traditions. The literature also investigates developments in RNNs, including LSTM and IndRNN, to address training efficiency and enhance recipe quality. The collection encompasses advancements in dish recognition using both visual and text data, reducing ambiguity and enabling real-time recipe suggestions and cooking guidance. This diverse collection of research serves as a valuable resource for informing and advancing the development of automated recipe generation system from food images using deep learning [10].

## **BLOCK DIAGRAM AND DESCRIPTION**

The system begins with the Input Module, where users enter or upload data. This data is then processed by the Processing Unit, which utilizes NLP and image recognition to interpret the input and fetch or generate appropriate recipes from the Recipe Database. The Display Module presents the recipes to the user, and the Output Module ensures the recipes are delivered in a suitable format for the user's device. User feedback collected through the Feedback Loop helps refine and enhance the system's performance over time as shown in Figure 1.



**Figure 1.** Block diagram of displays recipe.

### User

The user provides input to the system by uploading an image of food through the user interface. This input could be photo taken with a camera or selected from the images available in the image database. The user interface is the means by which users interact with the system. It provides options for the user to upload an image of food.

### Image Upload

This block handles the image uploaded by user. It passes the uploaded image to next block that is CNN algorithm.

### CNN Algorithm (Convolution Neural Network)

This is the heart of system. The CNN takes the food image as input and extracts relevant features from it through multiple layers of convolution and pooling. These features capture important aspects of the food image, such as ingredients, textures, and shapes.

### Recipe Generation

Once the CNN has extracted features from the food image, this block uses those features to generate a recipe. This could involve mapping the features to known recipes in database or using a generative model to create a new recipe based on extracted features which displays on user interface.

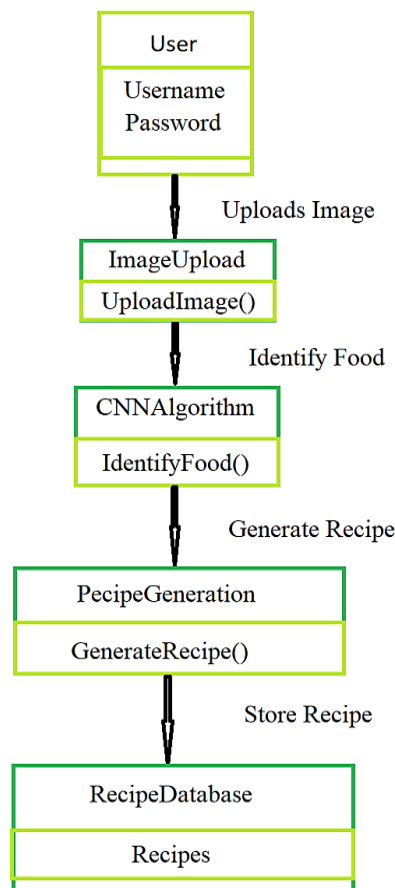
### Recipe Database

This block stores a collection of recipes. In some systems, the generated recipe might be matched to existing recipes in database, while in other, it might be added as a new entry.

## DATA FLOW DIAGRAM

Data flow diagrams (DFD) outline the processes involved in transferring data from input to file storage and report generation within a system. Recipe generator system operates through a series of interconnected steps of AI/ML. Initially, users authenticate themselves to access the system, ensuring security and personalized features. Upon uploading an image of food, the system leverages a Convolutional Neural Network (CNN) algorithm to accurately identify the food item depicted in the image. CNNs are very effective for image recognition because they identify patterns and features within images. Once the food is identified, the system proceeds to generate a recipe tailored to the recognized food item. This recipe generation process may involve accessing a database of existing recipes and utilizing information related to the identified food, such as ingredients and cooking methods, to create a suitable recipe. The generated recipe is then stored in a dedicated recipe database for future reference and potential recommendation to other users. By utilizing AI and machine learning techniques, the system can continuously improve its recipe generation capabilities, adapt to user preferences, and enhance the overall user experience by providing personalized and relevant recipe suggestions (Figure 2).

The AI/ML-based recipe generator system works by allowing users to upload food images, which are then analyzed using a Convolutional Neural Network (CNN) to identify the food item. Once identified, the system generates a recipe based on the recognized food item and stores it in a recipe database for future access. This process enables personalized recipe recommendations and enhances user experience through automated recipe generation.



**Figure 2.** Data flow diagram of recipe generator system.

## EXPECTED RESULTS AND DISCUSSION

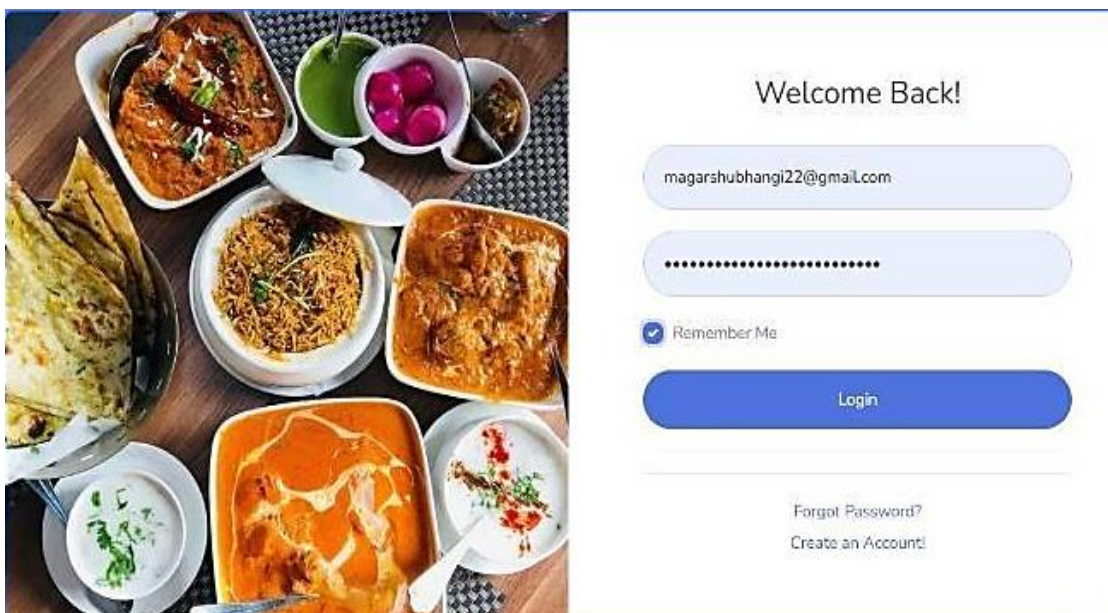
The image displays a user login interface for a website or application, with a food-themed background suggesting that the platform is likely related to culinary content, such as a recipe sharing service. The design is clean and straightforward, featuring a central white login box with a welcoming message "Welcome Back!" at the top. The login box contains fields for the user's email address and password, both of which have been filled in, though the password is obscured for security. Below the input fields, there is a "Remember Me" checkbox that likely allows users to stay logged in for future visits, enhancing user convenience. The primary action button is the "Login" button, which is designed to stand out with a solid blue color. At the bottom of the login box, there are additional links for users who have forgotten their password ("Forgot Password?") or new users who need to create an account ("Create an Account"), providing clear navigation for different user needs. The background image on the left side of the interface showcases a variety of dishes, which adds a visual appeal and context to the platform's focus on food. The overall design uses a simple color scheme with blue accents that match the login button, creating a cohesive and user-friendly interface (Figure 3).

The image shows a web-based interface for a recipe generation tool that uses food images to predict recipes. The design is minimalist and functional, with a clean and modern aesthetic. The main feature of the interface is a central panel where users can upload an image of food. In this case, an image of a dish has already been uploaded, as indicated by the thumbnail preview within the panel. Below the image preview, there is a prominent blue "Predict" button, which suggests that clicking this button will initiate the recipe prediction process based on the uploaded food image. The color of the button is a deep blue, which stands out against the light background, drawing the user's attention and indicating that it is the next step in the process. On the left side of the interface, there is a vertical navigation bar

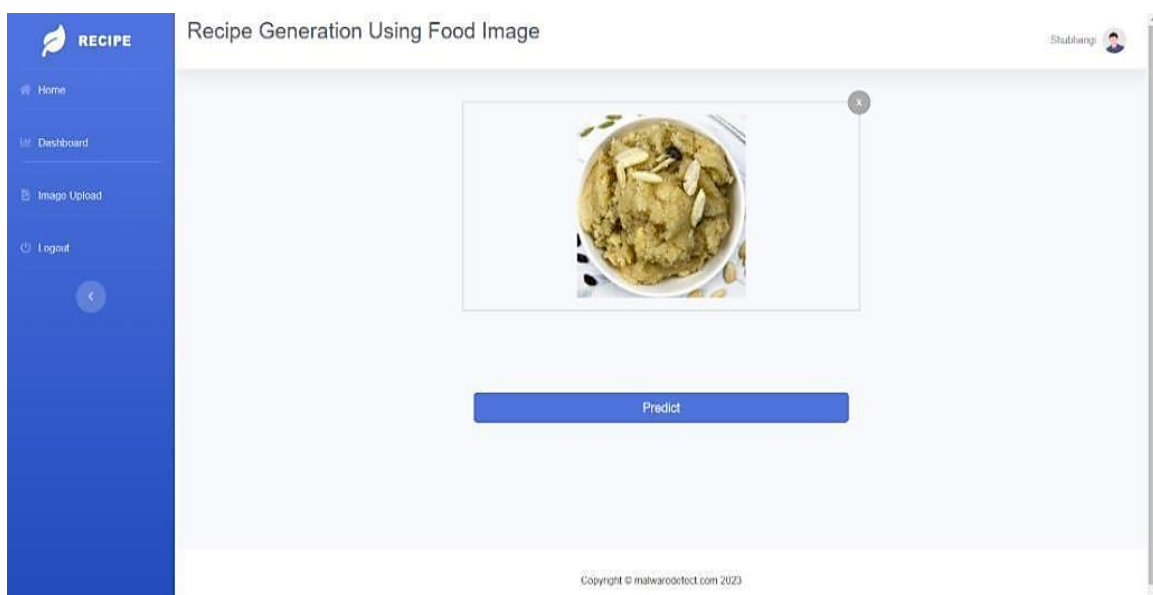
with options for "Home", "Dashboard", "Image Upload", and "Logout", indicating different sections of the application. The navigation bar uses icons and text, making it intuitive for users to move between different parts of the application.

The design is straightforward, focusing on usability and functionality without any unnecessary distractions. The layout is well-organized, with ample white space that contributes to an uncluttered look, making it easy for users to understand and interact with the tool (Figure 4).

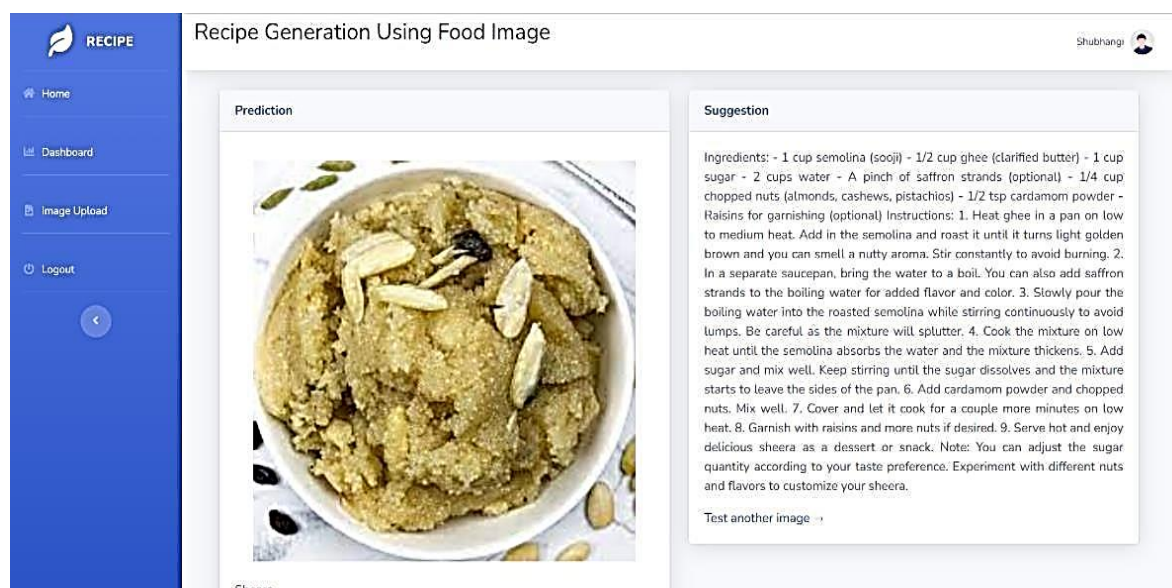
Subsequently, a recipe generation module processes this information to create a corresponding recipe based on the identified food. The generated recipe, including ingredients and cooking instructions, is stored in a Recipe Database for future reference. This seamless flow of data from image upload to recipe generation and storage enables users to easily access and recreate recipes corresponding to the food images they upload (Figure 5).



**Figure 3.** Login page.



**Figure 4.** Input Image.



**Figure 5.** Generated Recipe.

## CONCLUSION

The application of deep learning in generating recipes from food images holds great promise in enhancing culinary experiences and assisting in meal planning. This technology leverages neural networks to recognize food items, analyze their composition, and propose relevant recipes. However, challenges remain in accurately capturing complex culinary nuances, ensuring diversity in recipe suggestions, and accounting for cultural variations. As the field advances, refining model accuracy and incorporating user preferences will be key for realizing the full potential of recipe generation from food images.

## Acknowledgment

We would like to express our sincere gratitude to all those who have supported and contributed to the successful completion of the first Stage of Project. Their guidance, encouragement, and assistance have been invaluable throughout this journey. First and foremost, we are deeply grateful to our project supervisor, Ms. Smita Palnitkar, for their constant support, valuable insights, and expert guidance. Their vast knowledge and experience in the field have been instrumental in shaping the direction of this project. We would like to express our gratitude towards our parents and our classmates for their kind cooperation and encouragement which help us in completion of this project. My thanks and appreciations also go to my colleagues in developing the project and people who have willingly helped me out with their abilities.

## REFERENCES

1. Salvador A, Drozdal M, Giró-i-Nieto X, Romero A. Inverse cooking: Recipe generation from food images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019; 10453–10462.
2. Salvador A, Hynes N, Aytar Y, Marin J, Ofli F, Weber I, Torralba A. Learning cross-modal embeddings for cooking recipes and food images. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2017; 3020–3028.
3. Gao X, Feng F, He X, Huang H, Guan X, Feng C, Ming Z, Chua TS. Hierarchical attention network for visually-aware food recommendation. IEEE Trans Multimedia. 2019 Oct 4; 22(6): 1647–59.
4. Chen JJ, Ngo CW, Chua TS. Cross-modal recipe retrieval with rich food attributes. In Proceedings of the 25th ACM international conference on Multimedia. 2017 Oct 23; 1771–1779.
5. Li J, Monroe W, Ritter A, Galley M, Gao J, Jurafsky D. Deep reinforcement learning for dialogue generation. arXiv preprint arXiv:1606.01541. 2016 Jun 5.

- 
6. Carvalho M, Cadène R, Picard D, Soulier L, Thome N, Cord M. Cross-modal retrieval in the cooking context: Learning semantic text-image embeddings. In the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. 2018 Jun 27; 35–44.
  7. Fujii T, Sei Y, Tahara Y, Orihara R, Ohsuga A. “Never fry carrots without chopping” Generating Cooking Recipes from Cooking Videos Using Deep Learning Considering Previous Process. *Int J Networked Distrib Comput*. 2019 Jul; 7(3): 107–12.
  8. Pinel F, Varshney LR, Bhattacharjya D. A culinary computational creativity system. *Computational creativity research: Towards creative machines*. Paris: Atlantis Press; 2015; 327–46.
  9. Teng CY, Lin YR, Adamic LA. Recipe recommendation using ingredient networks. In *Proceedings of the 4th annual ACM web science conference*. 2012 Jun 22; 298–307.
  10. Srinivasa CK, Harshitha S, *et al*. Recipe Finder Web Application Using Convolutional Neural Network. *International Journal of Research Publication and Reviews (IJRPR)*. 2024; 5(5): 10071–10076.