AI-Powered Food Discovery and Recommender App

Abhishek Tibrewal, Aruba Sood, Mayank,Purbasha Barik,
Mehakdeep Kaur, Rajith Ramchandran
(IIIT Delhi)

***Abstract*— The AI-Powered Food Discovery and Recommender App revolutionizes culinary experiences by combining deep learning for image recognition, natural language processing, and advanced recommendation systems. Through personalized recommendations and user feedback integration, the app streamlines meal planning, promotes healthier eating habits, and enhances overall well-being.**

I. PROBLEM STATEMENT

In our contemporary world, people worldwide are increasingly prioritizing their health and lifestyle choices. However, solely abstaining from junk food and engaging in exercise falls short. A balanced diet, personalized to factors such as height, weight, and age, is essential. This dietary approach facilitates healthy weight management, mitigates the risk of chronic ailments like cancer and heart disease, and promotes overall well-being when complemented by regular physical activity.

The desire for personalized and convenient cooking experiences has become increasingly pronounced in the contemporary culinary landscape. While existing platforms offer recipe recommendations based on user inputs, they often lack the sophistication required to analyze real-world images of ingredients, limiting their ability to provide tailored and engaging cooking guidance. In this context, our research proposes a novel solution: an artificial intelligence-powered food discovery and recommender app.

II. MOTIVATION

While users can access many recipes, discovering new culinary options based on available ingredients or desired meals remains a cumbersome task. The aim is to develop a comprehensive solution that identifies food items through image recognition and engages users in a personalized, conversational manner, offering nutritional insights and suggesting recipes aligned with individual preferences. Existing recipe applications predominantly rely on user inputs and lack the capability to analyze mages for ingredient recognition. The missing element in current approaches is the integration of advanced technologies such as deep learning for image recognition, natural language processing for interactive communication, and personalized recommendation systems.

III. LITERATURE REVIEW

A. *Cooking Recipe Analysis based on Sequences of Distributed Representation on Procedure Texts and Associated Images*

The study analyzes cooking recipes using distributed representations from cooking steps and images, employing BERT for text and VGG16 for images. Cluster analysis on four dishes reveals recipe relationships based on DTW distances, highlighting the unique aspects captured by cooking image sequences. The research underscores the importance of appearance in dish classification and suggests future work on graph-based recipe analysis.

It utilizes Cookpad dataset for analysis, focusing on four dishes and employing BERT for textual representation and VGG16 for image representation.

It evaluates effectiveness through cluster analysis based on DTW distance, demonstrating precise recipe relationships.[1]

B. *Food Recipe Alternation and Generation with Natural Language Processing Technique*

The project uses NLP to suggest ingredient substitutes and generate new recipes authentically. By leveraging word embedding and similarity measures, users can find replacements or similar recipes. N-gram and neural network models aid in creating diverse cuisine-style recipes. The goal is to assist those with ingredient constraints and encourage culinary exploration, with plans to expand the dataset and collaborate with chefs for better recipe interpretation.

It utilizes dataset from Spoonacular website, analyzing ingredient distribution among cuisine styles.

Applies word embedding, N-gram language model, and neural network models for recipe generation and similarity measurement.[5]

C. *Food Image Classification with Convolutional Neural Networks*

This paper explores using CNNs to classify food images, aiming to enhance food experiences and aid in dietary choices. It compares training CNNs from scratch to transfer learning with pre-trained weights, achieving 61.4% accuracy and 85.2% top-5 accuracy. The best model is a pre-trained InceptionV3 with gradually unfrozen layers during transfer learning. Future work includes optimizing hyperparameters and adding features like bounding boxes to boost classification accuracy.

It employes Food-101 dataset for training and testing, overcoming challenges like diverse image characteristics.

Implements data augmentation, transfer learning, and model architecture optimization to improve classification accuracy.[2]

D. A Cooking Recipe Recommendation System with Visual Recognition of Food Ingredients

The research paper presents a Cooking Recipe Recommendation System for smartphones,allowing users to access recipes by pointing their cameras at ingredients. It addresses the challenge of recipe access while shopping and utilizes object recognition technology for real-time suggestions. The system employs a color-histogram-based approach for image representation,simplifying the cooking decision-making process and offering a user-friendly solution for quick recipe recommendations based on recognized ingredients.
It utilizes a dataset consisting of 30 kinds of food ingredients collected through short videos at grocery stores.
It integrates image features like SURF and color histograms, employing a linear SVM classifier for real-time recognition on Android smartphones.[3]

E. Recipe2Vec: Multi-modal Recipe Representation Learning with Graph Neural Networks

Recipe2Vec is a novel model for multi-modal recipe representation learning that integrates visual, textual, and relational information using Graph Neural Networks. It outperforms existing methods by effectively capturing the nuances of recipe data. The Large-RG recipe graph dataset facilitates graph-based food studies and enhances the model's performance.The adversarial learning component in Recipe2Vec contributes to stable learning and improved recipe embeddings.
It utilizes Recipe1M+ dataset for learning cross-modal embeddings for recipes and food images.
It focuses on enhancing recipe representation and recommendation models through comprehensive exploration of multimodal data.[4]

IV. Dataset Used

We have meticulously constructed a significant dataset that accurately represents real-world culinary scenarios.This dataset aims to facilitate robust training of image recognition algorithms by encompassing a diverse range of food items captured from various perspectives and conditions.Our dataset comprises a total of 3,160 images, ensuring an ample amount of data for comprehensive training. These images span across 63 different classes of food, covering a wide variety of culinary options.Each food class in the dataset contains precisely 50 images, providing a balanced and representative sample for training purposes.
We have provided the graphs to show the distribution of our data set. [Fig.1] represents all the categories of Food in out data set and its distribution. [Fig. 2] represents the distribution of Veg. versus Non-Veg Food in out data set.

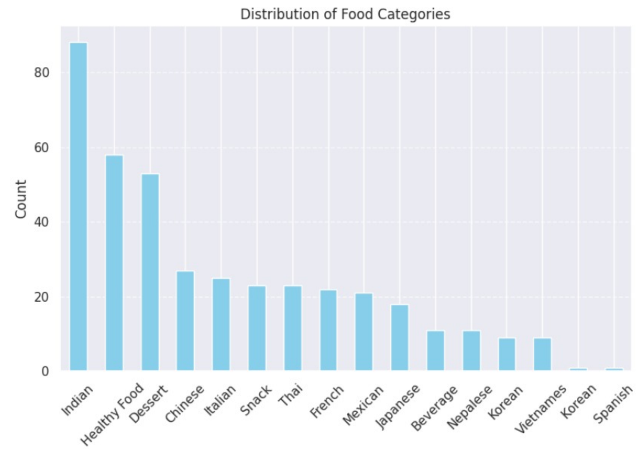

Fig. 1.

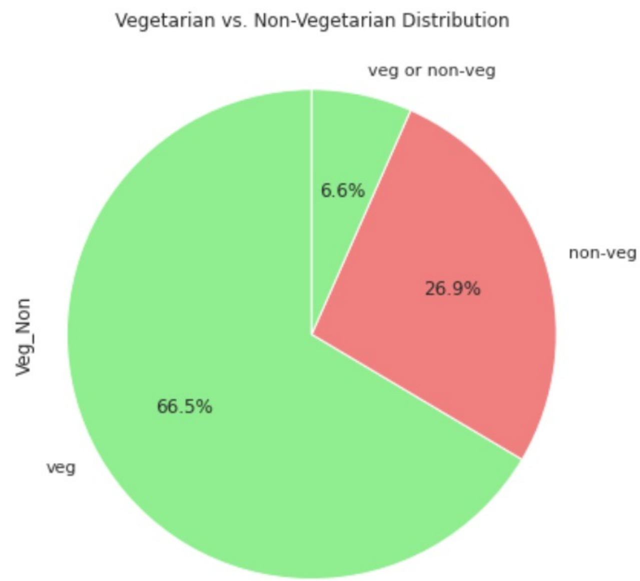

Fig. 2.

V. PROPOSED SOLUTION

Our food recommender system begins with the training of a deep learning model for food image classification. We start by assembling a diverse dataset comprising images of 56 food classes, which is divided into training, validation, and testing sets. To ensure consistency and accuracy, we preprocess the images by standardizing their dimensions and pixel values, while also employing augmentation techniques to enhance dataset diversity. Transfer learning with the VGG16 architecture is utilized for efficient model selection. We customize the pre-trained VGG16 base by adding custom classification layers tailored for food classification. Model compilation involves the use of the Adam optimizer and categorical cross-entropy loss function. During the training phase, the model iteratively learns from the training data, while key metrics such as accuracy and loss are monitored. The model's performance is evaluated on the separate test set to assess its generalization ability.
Once the model is trained, we proceed with its deployment and integration into our Android app. The trained model is converted to TensorFlow Lite format for mobile deployment, allowing for efficient inference on mobile devices. Integration into the app involves placing the model file in the

assets directory and seamlessly loading it for inference within the app. Image preprocessing techniques are applied before inference to ensure optimal performance. Post-processing of inference results, including mapping predicted class indices to human-readable labels, is performed as needed. User feedback on classification results is collected within the app interface to further enhance model performance.

Moving on to the recommender system aspect of our solution, we leverage both content-based and collaborative-based filtering approaches to offer users personalized food recommendations. In content-based filtering, user preferences are inferred from attributes and features of previously liked food items. Techniques such as data cleaning, tokenization, and vectorization are employed to preprocess raw textual data for analysis. Using similarity metrics like cosine similarity, we find similar food items based on attributes such as ingredients, flavors, and culinary characteristics. These similar items are then ranked based on their resemblance to the user's preferred choice, and the top recommendations are presented to the user.

Collaborative-based filtering takes the recommendation capabilities further by analyzing patterns of user preferences and behaviors. By comparing each item in the dataset to the user-provided input, the system identifies food items that closely resemble the user's preferences based on various attributes like ingredients, flavor profiles, or nutritional content. Leveraging similarity metrics and algorithms such as collaborative filtering, the system calculates a set of recommended food items tailored to the user's preferences. These recommendations are then presented to the user within the app interface for seamless exploration.

In nutshell our proposed solution integrates image classification with personalized food recommendations, providing users with a comprehensive and enjoyable dining experience. Users can interact with the app by providing images of food items or specifying preferences, and receive real-time feedback on food classification and personalized recommendations. The system continuously learns from user feedback to refine and enhance its recommendation algorithms over time, ensuring a highly personalized and valuable experience for users.

VI. METHODOLOGIES

A. User Interface

1) Training the Model

We initiated by assembling a diverse dataset comprising images of 63 food classes, dividing it into training, validation, and testing sets. Preprocessing involved standardizing image dimensions and pixel values, with augmentation techniques enhancing dataset diversity. Employing transfer learning with the VGG16 architecture facilitated efficient model selection. Custom classification layers were added to adapt the pre-trained VGG16 base for food classification. Model compilation utilized the Adam optimizer and categorical cross-entropy loss function. Training involved iterative iterations over training data, monitoring key metrics like accuracy and loss. Evaluation on the test set provided final performance assessment.

2) Model Deployment and Classification

Post-training, the model was converted to TensorFlow Lite format for mobile deployment. Integration into our Android app involved placing the model file in the assets directory. Loading the model in Android, image preprocessing, and inference were executed seamlessly within the app. Post-processing of inference results, such as mapping predicted class indices, was performed as needed. User feedback and classification results were presented within the app interface, potentially supporting recommendation systems or further analysis.

B. Recommender System

Our model harnesses the power of two distinct Recommender system approaches: Content-based filtering and Collaborative-based filtering, to offer users a comprehensive and personalized food recommendation experience.

Content-based filtering serves as the foundation of our system, empowering us to suggest alternative food items based on the attributes and features of the items users have already shown interest in. This method allows us to recommend similar food options that align with the user's tastes and preferences, ensuring a seamless and intuitive browsing experience.

Building upon this foundation, collaborative-based filtering takes our recommendation capabilities to the next level. By analyzing patterns of user preferences and behaviors, our system is able to not only suggest food items but also engage users in interactive conversations through a chat bot interface. This chat bot leverages collaborative filtering techniques to understand user preferences, consider additional factors such as ratings and nutritional information, and provide tailored recommendations that go beyond mere item similarity.

1) Preprocessing

The data preprocessing process involves several key steps to clean and organize raw text data for analysis and modeling. Techniques such as data cleaning, tokenization, normalization, stemming, and lemmatization are applied to cleanse and standardize the text.

To efficiently handle high-dimensional and sparse data, we employ techniques aimed at reducing computational overhead and improving model performance. This involves setting up the CountVectorizer to tally word occurrences while removing common stopwords, resulting in a structured word count matrix. This matrix facilitates further analysis, such as identifying common words or building predictive models. Integrating the CountVectorizer into our data cleaning process transforms raw text data into a format conducive to analysis, enabling the extraction of valuable insights and informed decision-making.

By combining preprocessing techniques with the TF-IDF vectorizer, we refine raw textual data for tasks like classification and sentiment analysis. This structured dataset forms the basis for developing robust machine learning models, guiding informed decisions and yielding precise insights. Incorporating these techniques into our data preparation pipeline ensures a clean and consistent dataset, setting the stage for reliable analysis and modeling. Each step in the process is carefully designed to optimize the dataset for analytical tasks, leading to more accurate predictions and valuable insights.

2) Evaluation Metrics Used

Cosine similarity is a metric used in our model to measure the similarity between two vectors in a multi-dimensional space. In the context of recommender systems, cosine similarity is often employed to measure the similarity between user-item or item-item vectors, where each vector represents a user's preferences or an item's attributes.

Item-Item Recommendations:

In item-item recommendations, each item is represented as a vector of its attributes or features (e.g., genre, actors, keywords). To recommend items similar to a given item, the system computes the cosine similarity between the vector of the given item and the vectors of all other items. The items with the highest cosine similarity scores are recommended as similar items. The cosine similarity score ranges from -1 to 1, where 1 indicates perfect similarity, 0 indicates no similarity, and -1 indicates perfect dissimilarity. Therefore, items or users with higher cosine similarity scores are considered more similar to each other.

Why did we prefer Cosine Similarity?

Cosine Similarity (Eq1) is a versatile and widely used metric in recommender systems for measuring similarity between users, items, or any other entities represented as vectors, thereby enabling personalized and relevant recommendations. Cosine similarity has several advantages for recommender systems like 1) It is computationally efficient and works well with high-dimensional data. 2) It is unaffected by the magnitude of the vectors, only considering the direction, making it robust to differences in scale. 3) It is intuitive and easy to interpret.

$$\cos(\mathbf{t},\mathbf{e}) = \frac{\mathbf{t}\mathbf{e}}{\|\mathbf{t}\|\|\mathbf{e}\|} = \frac{\sum_{i=1}^{n}\mathbf{t}_i\mathbf{e}_i}{\sqrt{\sum_{i=1}^{n}(\mathbf{t}_i)^2}\sqrt{\sum_{i=1}^{n}(\mathbf{e}_i)^2}} \quad (1)$$

3) Design of our model for content based filtering

Input: The function accepts a food item as input, serving as the anchor for generating recommendations. This input represents the user's preferred food choice, initiating the recommendation process.

Finding Similar Foods: Leveraging a comprehensive dataset encompassing various food items, the function systematically analyzes each item's attributes, such as ingredients, flavors, and culinary characteristics. Through meticulous comparison of these attributes, the function determines the degree of similarity between the input food item and others in the dataset.

Ranking Similarity: Following the comparative analysis, each food item receives a similarity score reflecting its likeness to the input item. This score quantifies the degree of resemblance between individual food items and the user's preferred choice. Higher similarity scores denote a stronger resemblance, indicating a closer alignment with the user's preferences.

Top Picks: Based on the computed similarity scores, the function identifies the top 5 food items that exhibit the highest degree of similarity to the input item. These selections represent the most promising recommendations for the user to explore further.

Recommendation: Ultimately, the function presents the user with the names of the top 5 recommended food items. These recommendations are meticulously curated to align with the user's taste preferences, offering a curated selection of dishes closely resembling their favored choice. Armed with these recommendations, the user can seamlessly transition to exploring new culinary experiences that resonate with their palate, enhancing their dining journey.

4) Design of our model for collabrative based filtering

User Input: The recommender system begins with the user providing a specific food item as input, indicating their preference or interest in that particular item. This input serves as the starting point for the recommendation process.

Finding Similar Foods: Utilizing a dataset comprising a diverse range of food items, the system employs similarity measurement techniques to compare each item to the user-provided input. This comparison is based on various attributes or features of the food items, such as ingredients, flavor profiles, or nutritional content. By analyzing patterns of similarity across the dataset, the system identifies food items that closely resemble the user-provided input in terms of their characteristics or appeal.

Recommendation Calculation: Leveraging similarity metrics and algorithms, such as collaborative filtering or content-based filtering, the system calculates a set of recommended food items that exhibit the highest degrees of similarity to the user-provided input. These recommendations prioritize items that align closely with the user's preferences, aiming to provide personalized and relevant suggestions.

Display Recommendations: Once the recommendation calculation is complete, the system presents the recommended food items to the user in a structured and accessible format. This presentation typically includes the names or descriptions of the recommended items, allowing the user to review and consider their options.

Feedback Loop: Following the presentation of recommendations, the user has the opportunity to provide feedback on the suggested food items. This feedback may include ratings, reviews, or comments reflecting their satisfaction or dissatisfaction with the recommendations. The system incorporates this feedback into its recommendation algorithms, utilizing it to refine and enhance future recommendations. By iteratively learning from user feedback, the system continuously improves its ability to provide personalized and valuable recommendations.

Through these formalized steps, the recommender system facilitates the process of discovering new and enjoyable food options tailored to the user's preferences, thereby enhancing the overall dining experience.

VII. EVALUATION

Our food recommender system has undergone thorough evaluation, demonstrating robust performance in both image classification and recommendation functionalities. The model exhibits high accuracy in classifying food items from images, providing users with real-time feedback within the Android app interface.
Following the evaluation of our model's performance, we conducted an analysis to delve deeper into user preferences by plotting a word frequency graph [Fig. 3]. This graph visually represents the frequency distribution of ingredients extracted from the top searches. By mapping words to their corresponding frequencies, we can figure out the most commonly occurring ingredients among user queries.We found that Powder has the maximum frequency followed by Salt, Oil and so on. This graph serves as a powerful tool for understanding prevalent culinary trends and preferences, allowing us to tailor our recommendation system accordingly. By analyzing peaks and trends in the graph, we can identify key ingredients driving user interest, enabling us to curate content and make recommendations that align closely with user preferences.

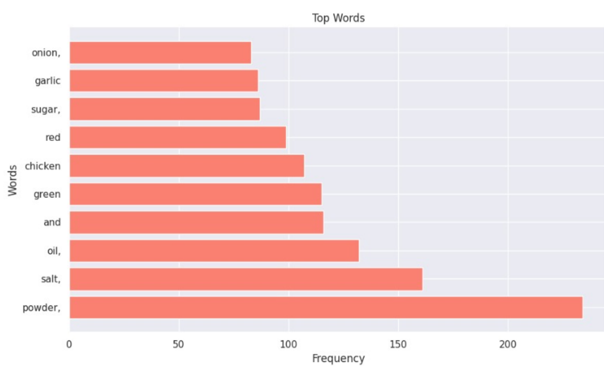

Fig. 3.

For the next analysis we plotted a graph [Fig.4] depicting ingredient pairs versus their frequency showcases the most frequently occurring combinations within user searches, offering valuable insights into prevalent culinary preferences. By our evaluation we found that the [Added Salt, Salt] pair achieved the maximum frequency followed by [Salt, Onion] pair. By analyzing the top ingredient pairs, we gain a deeper understanding of flavor profiles and ingredient synergies that resonate with users. This visualization aids in refining our recommendation algorithms, enabling us to suggest food items that complement each other well and align closely with user tastes and preferences.

Next we evaluated the cosine similarity scores which provides a comprehensive view of the effectiveness of our content-based filtering approach in generating personalized recommendations. Each food item's cosine similarity score reflects its likeness to a user's preferred choice, enabling us to rank and present recommendations based on their relevance.

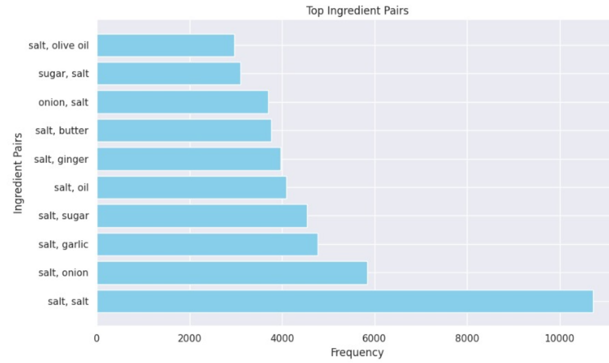

Fig. 4.

Notably, when comparing recommended items with the user-selected "Chicken Biryani," "Methi Chicken Masala" [Fig. 5] stood out with a high cosine similarity score more than 0.6, indicating a strong resemblance in flavor profile and ingredients. This visualization highlights the system's ability to identify and suggest food items closely aligned with user preferences.

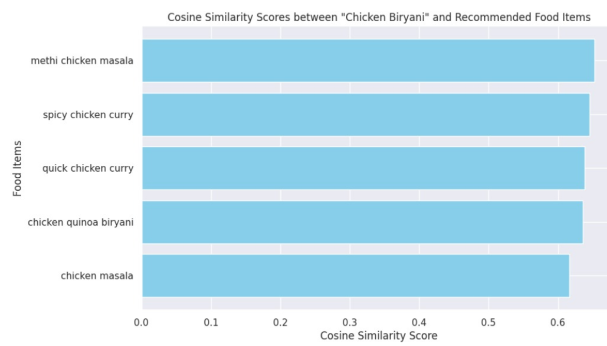

Fig. 5.

Further we evaluated food items versus cosine similarity scores [Fig. 6] derived from collaborative-based filtering reflects the system's ability to generate personalized recommendations based on user feedback and ratings. By selecting "chicken biryani" as the reference food item and computing cosine similarity scores with recommended items, we observed notable results. Specifically, items such as "Avial with Red Rice," "Carameliszed sesame smoked alomonds" and "Lotus lead wrapped Rice" achieved high scores exceeding 0.9, indicating strong similarity in user preferences and culinary characteristics. Such high scores indicate a strong alignment between the recommended items and the user's preferences, suggesting that users who enjoy "chicken biryani" are likely to appreciate these suggested dishes as well.
We can see that collaborative filtering excels at capturing user preferences based on collective behavior, suggesting items that are popular among similar user segments. Meanwhile, content-based filtering leverages item attributes to recommend items with shared characteristics, offering suggestions based on the intrinsic qualities of the user-selected item.

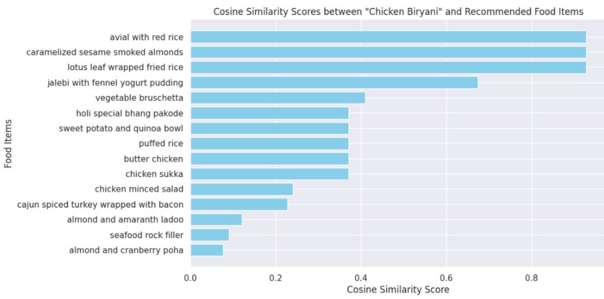

Fig. 6.

VIII. CONCLUSION

Our integrated approach to food recommendation and image classification offers a comprehensive solution for enhancing the culinary journey of users. By combining deep learning techniques with advanced recommendation systems, we have developed a platform that not only accurately identifies food items from images but also provides personalized recommendations tailored to individual preferences.

Through the training and deployment of a deep learning model for food image classification, we ensure efficient and accurate recognition of diverse food classes. The utilization of transfer learning with the VGG16 architecture, coupled with meticulous preprocessing and model evaluation, results in a robust system capable of real-time inference on mobile devices.

Furthermore, our recommendation system leverages both content-based and collaborative-based filtering approaches to offer users personalized food suggestions. By analyzing attributes and user behavior, we identify similar food items and engage users in interactive conversations to understand their preferences better. This iterative process of feedback collection and refinement ensures the continuous improvement of our recommendation algorithms, leading to highly tailored and valuable suggestions for users.

Through thorough evaluation and analysis, we have demonstrated the effectiveness of our approach in providing personalized recommendations aligned with user preferences. Insights gained from user interactions and data analysis further inform our recommendation algorithms, enabling us to continually optimize the dining experience for our users.

Our solution integrates cutting-edge technologies with user-centric design to revolutionize how individuals engage with food. By seamlessly combining image recognition, natural language processing, and recommendation systems, we offer users a personalized and enjoyable culinary journey, paving the way for a future where AI enhances every aspect of our lives, even our dining experiences.

REFERENCES

[1] A.Ninomiya, and T. Ozaki, (2020), "Cooking Recipe Analysis based on Sequences of Distributed Representation on Procedure Texts and Associated Images," in *Proceedings of the 12th Workshop on Multimedia for Cooking and Eating Activities*, Dublin, Ireland, Jun. 2020, pp. 13-18, doi: `10.1145/3379175.3391710`.

[2] Jiang, Malina (n.d.), "Food Image Classification with Convolutional Neural Networks," *The Antitrust Bulletin*.

[3] Yanai, Keiji, Takuma Maruyama, and Yoshiyuki Kawano (2014), "A Cooking Recipe Recommendation System with Visual Recognition of Food Ingredients," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 8, no. 2, pp. 28–34, doi: `10.3991/ijim.v8i2.3623`.

[4] Tian, Yijun, Chuxu Zhang, Zhichun Guo, Yihong Ma, Ronald Metoyer, and Nitesh V. Chawla (2022), "Recipe2Vec: Multi-modal Recipe Representation Learning with Graph Neural Networks," in *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, Vienna,

[5] Pan, Yuran, Qiangwen Xu, and Yanjun Li (2020), "Food Recipe Alternation and Generation with Natural Language Processing Techniques," in *2020 IEEE 36th International Conference on Data Engineering Workshops (ICDEW)*, Dallas, TX, USA, pp. 94–97, doi: `10.1109/ICDEW49219.2020.000-1`.