Journal of
Artificial Intelligence
Research & Advances

ISSN: 2395-6720
Volume 11, Issue 3, 2024
September–December
DOI Journal: 10.37591/JoAIRA

https://journals.stmjournals.com/joaira

Review JoAIRA

Recipe-fusion: Multimodal Food Recipe Recommendation System

Karthik Anilkumar Nair[1,*], Sanjog Gajanan Chavhan[2], Tejas Nilkanth Pawar[3], Anmol Jagdish Reddy[4], K.S. Charumathi[5]

Abstract

The food recipe recommendation system using data science is a software solution designed to help users discover new and delicious food options based on their food history and other relevant data. This system recommends various recipes based on the input given by the user and it helps to filter out the recipes on course type, diet type, and nature of the food (including non-veg, and veg) using a recommendation technique. The system also analyzes food trends and popular dishes that the users might not have considered before. The system can recognize the input ingredients provided by the user and give the recipe based on the ingredients mentioned. Various machine learning and deep learning approaches like convolutional neural networks (CNN) and recurrent neural networks (RNN) will be used for the implementation of this system. The CNN model is trained and implemented to recommend the recipes based on the image similarity score. The RNN model is trained and implemented for the prediction of the recipes from the input (recipe name) given by the user. Both models (CNN and RNN) are used simultaneously for the implementation of the system, which is why it is named as "Multimodal Food Recipe Recommendation System". This system will help the person learn and make recipes based on the input given to it. This system will simplify finding relevant recipes for the user.

Keywords: Recommendation, content-based, machine learning, deep learning approaches, convolutional neural network

INTRODUCTION

A multimodal food recipe recommendation system requires a large dataset of recipes along with their metadata such as ingredients, cooking time, nutritional information, and ratings. The system must pre-process the recipe data and extract relevant information, such as ingredient names, quantities, and cooking instructions. A multimodal food recipe recommendation system must represent recipe data using multiple modalities, such as text, images, and videos. This enables the system to provide a rich and engaging user experience and better understand user preferences. For example, food images can be processed using convolutional neural networks (CNN). Natural language processing (NLP) techniques can be used to analyze user reviews and ratings and extract sentiments and opinions about specific recipes. The existing system of food recipe recommendations is designed to help users discover new and delicious meal ideas, based on their preferences and dietary restrictions. This system uses various algorithms and machine learning techniques to analyze user data, such as past searches, ratings, and ingredient lists, to provide personalized recipe recommendations. By incorporating user feedback and learning from their

***Author for Correspondence**
Karthik Anilkumar Nair
E-mail: nairkart20comp@student.mes.ac.in

[1-4]Student, Department of Computer Engineering, Pillai College of Engineering, New Panvel, Maharashtra, India
[5]Professor, Department of Computer Engineering, Pillai College of Engineering, New Panvel, Maharashtra, India

Received Date: June 28, 2024
Accepted Date: August 21, 2024
Published Date: September 11, 2024

Citation: Karthik Anilkumar Nair, Sanjog Gajanan Chavhan, Tejas Nilkanth Pawar, Anmol Jagdish Reddy, K.S. Charumathi. Recipe-fusion: Multimodal Food Recipe Recommendation System. Journal of Artificial Intelligence Research & Advances. 2024; 11(3): 82–91p.

choices, the system continuously improves its recommendations, making them more accurate and relevant over time. Moreover, the system can offer nutritional information for each recipe, enabling users to easily make informed dietary choices. Overall, the food recipe recommendation system offers a convenient and effective way for users to explore new and exciting meal options and provides a valuable resource for those seeking to maintain a healthy and balanced diet [1].

The system begins by leveraging a CNN to analyze food images. These CNN models extract intricate features from food images and capture visual cues such as ingredients, presentation, and style. These features were then used to create a high-dimensional image representation, enriching the understanding of the dishes in the dataset. Simultaneously, the recurrent neural networks (RNN) component processed the textual data associated with each recipe, including its name, ingredients, and instructions. RNN models are well suited for handling sequential and textual data.

The RNN analyzes textual information to comprehend the semantics and cooking process of each recipe, allowing it to capture the nuances and complexity of food preparation. The representations, images, and text of the two modalities are fused in a multimodal feature integration step. This fusion helps align the visual and textual contexts and creates a holistic understanding of each recipe. This will help in recommendation purposes and understanding the personalization or diet of an individual. The user will then obtain recommendations of the recipes to which they are familiar or have an interest. The combined architecture enriches the recommendation system by considering both the visual appeal and textual content of recipes, making it well suited for a wide range of users with diverse preferences and dietary requirements. The multimodal approach enhances the recommendation quality, making it a powerful tool for culinary enthusiasts and individuals seeking personalized and visually appealing recipe suggestions.

LITERATURE REVIEW

In [2], the authors proposed a machine learning technique using a CNN in the system. This presents the importance of food and the difficulty of selecting the correct recipe based on an ingredient, which can be challenging for both beginner and expert cooks. CNN models cannot classify multiple objects. CNN models can only recognize a single food ingredient, and not various food ingredients, from a given image at a given time. Many existing systems rely solely on past user interactions to make recommendations, without considering the user's personal preferences or dietary restrictions. There is a need to incorporate explicit user feedback or explicitly model user preferences to improve the quality of the recommendations [3].

In [4], the authors proposed collaborative filtering for a recipe system. The technique is used whenever the system recommends recipes to the user based on the ratings of similar users. The system could only rate and recommend recipes to users. The system does not display lists of top-rated recipes in the system. Additionally, the system cannot send notifications. Some existing systems tend to recommend popular or commonly searched recipes, resulting in a lack of diversity in recommended recipes.

In [5], the authors solved the problem of the image processing of ingredients using machine learning and deep learning approaches. In this study, a model was implemented to identify food ingredients, and an algorithm was designed to recommend recipes based on the identified ingredients. It focuses on analyzing user interactions with recipes to make personalized food recommendations. Incorporating diversity-aware recommendation techniques could lead to more varied and personalized recipe recommendations.

In [6], content-based filtering was applied to a recipe system. It provides recommendations based on user preferences. Products cannot be recommended until they do not contain information about the user. More appropriate recommendations were provided in the user based approach. Most of the existing systems do not consider the availability of ingredients or budget constraints when making recommendations. Integrating information regarding ingredient availability and budget constraints could lead to more practical and useful recommendations for users.

In [7], the authors proposed a recipe recommendation system that operates as an interactive web application on a user's smartphone. This system employs a recommendation method in which the user's ingredients are input, an analysis process is performed using a dataset, and suitable dishes or recipes are then suggested to the user. Many existing systems utilize only a subset of the available multimodal data, such as images or text descriptions of the recipes. There is a need to investigate ways to effectively integrate all available data sources, including user-generated content and user feedback, to improve the quality of recommendations.

In [8], the authors proposed a machine learning technique using k-nearest neighbors (kNN). It introduces an AI-driven food recommendation system that suggests recipes based on the ingredients that the user currently has. While using a large dataset that has many variances, it is better to opt for a solution within a deep learning domain, and kNN would have slower and less accurate results on large datasets with high intraclass variance features. Although many existing systems are evaluated using user studies or offline evaluation metrics, there is a need to evaluate these systems in real-world scenarios to better understand their practical effectiveness and user acceptance.

METHODOLOGY

Techniques

Natural Language Processing

Natural language processing techniques can be used to analyze recipe text and extract key information such as ingredients, cooking instructions, and nutritional information. This will enable personalized recipe recommendations tailored to user preferences and dietary restrictions [9].

Image Recognition

Image recognition techniques can analyze food images to identify ingredients and cooking methods. This information can be used to make recipe recommendations based on the visual characteristics of a dish.

Reinforcement Learning

Reinforcement learning is a method of training an agent to take action within an environment to maximize a reward. In the realm of food recipe recommendation systems, reinforcement learning can be used to enhance recipe suggestions based on user feedback.

Recommendation Approaches

Recommendation approaches involve hybrid, collaborative, and content-based approaches. Hybrid approaches integrate various recommendation techniques to deliver more accurate and personalized recommendations. For example, a hybrid approach includes collaborative and content-based filtering to suggest recipes that are both popular among similar users and similar to recipes that a user has enjoyed in the past [10].

Machine Learning Algorithms

Recipe recommendation systems utilize machine learning algorithms to analyze extensive datasets, including user preferences, ingredient combinations, and cooking techniques. By examining these data, these algorithms can offer personalized recommendations tailored to each user's unique taste and previous interactions with the system.

Implementation Details

- *Data collection:* A large dataset of food recipes, along with their ingredients, cooking steps, and accompanying images. We can scrape data from recipe websites or use publicly available data sets.

- *Data pre-processing:* Clean and pre-process the data by removing duplicates and correcting or eliminating any incorrect or irrelevant information. Text cleaning, normalization, and tokenization may be required to prepare textual data for modeling.
- *Feature Extraction:* Extract features from textual data using methods such as word embeddings or bag-of-words and from images using computer vision techniques such as CNNs.
- *Multimodal Fusion:* Combine textual and visual features using a multimodal fusion technique such as concatenation or late fusion. This creates a joint representation of the recipe that captures both the textual and visual aspects.
- *Recommendation Model:* Train a recommendation model for the joint representation of recipes. Model building can be achieved using recommendation techniques such as collaborative filtering, matrix factorization, or neural networks. We made recommendations using a CNN model.
- *Evaluation:* Assess the performance of recommendation models using metrics such as precision, recall, and F1-score while also conducting user studies to evaluate the overall user experience of the system [11–15].
- *Deployment:* The system is deployed on a web or mobile platform where users can interact with it. We can use APIs to enable the system to receive input from users and provide recipe recommendations in real time.

Proposed System

The proposed system architecture for a multimodal food recipe recommendation system, integrating CNN and RNN, offers a comprehensive approach to enhance the food recommendation experience [16].

The architecture combines the strengths of both CNN and RNN for processing multimodal data. The system begins by leveraging a CNN to analyze food images. These CNN models extract intricate features from food images. Simultaneously, the RNN model was used to generate recipes based on the textual input provided by the user. The system has two main functions: searching for and fetching recipes. When a user searches for recipes, the system searches the database for relevant recipes. When a recipe is selected, the system fetches the recipe details and displays them to the user [17–20].

The RNN analyzes textual information to comprehend the semantics and cooking process of each recipe, allowing it to capture the nuances and complexity of food preparation. With the help of the RNN and CNN models, recipes are generated and accordingly recommended to the user. The app was deployed using the Flask software. The user must enter the recipe's name as an input, and the recipe will be generated with the help of our trained RNN model, and the recommendation based on image similarity will be performed with the help of the trained CNN model.

The multimodal approach enhances recommendation quality, making it a powerful tool for culinary enthusiasts and individuals seeking personalized and visually appealing recipe suggestions, as shown in Figure 1.

RESULTS

We prepared a dataset to train the CNN and RNN models. Various cooking sites were used to collect information to prepare the dataset. Figure 2 shows a snapshot of this dataset. In Figure 3, a line of code assigns the value of the expression cnn_model.eval() to the variable image_features. The expression cnn_model.eval() evaluates the CNN model and returns the features extracted from the image. Variable image_features contain a list of features suitable for subsequent processing. This model is used to predict recipes with similarities in images. Figure 4 shows the output of the recommendation based on image similarity using the CNN model. The metrics used to evaluate the CNN model for recommendation purposes, along with the scores, are listed in Table 1.

Table 1 shows the metrics used to evaluate the recommendation system using CNN in terms of ranking the relevant items suggested in the recommendations.

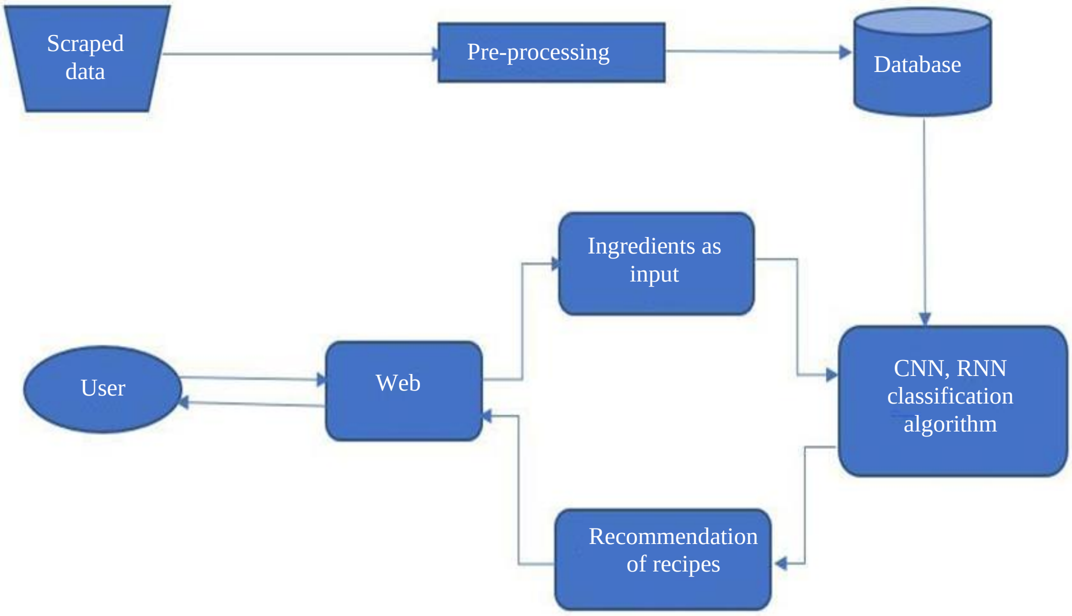

Figure 1. Proposed system.

	A	B	C	D	E	F	G	H	I	J	K	L
1	name	image_url	descriptio	cuisine	course	calories &	diet	prep_time	cook_time	ingredien	instructio	image_id
2	Indian Fry	https://m	Fried brea	Indian	Appetizer	682 & 4	Veg	45 minute	20 minute	4 cups flo	Mix flour,	0
3	Baked Ind	https://s3	Delicious	Indian	Dessert	581 & 4	Veg	25 minute	60 minute	AÂ½ cup c	Preheat O	1
4	Cabbage F	https://s3	Cabbage F	Indian	Side Dish	134 & 8	Veg	10 minute	0 minutes	1/1 a med	Place the	2
5	Shepuchi	http://gre	Shepuchi	Indian	Main Cour	31 & 4	Veg	10 minute	20 minute	1 bunch D	1.Soak mo	3
6	Beetroot (	https://s3	To make b	Indian	Appetizer	190 & 4	Veg	20 minute	15 minute	Beetroot -	Pressure c	4
7	Mattar Sa	https://s3	Mattar sar	Indian	Appetizer	46 & 4	Veg	20 minute	25 minute	For crust,	For crust,	5
8	GAJAR KA	https://s3	Gajar ka H	Indian	Dessert	239 & 10	Veg	15 minute	30 minute	1/4 c. corn	Saute grat	6
9	Tawa Fry S	http://gre	Tawa fry s	Indian	Main Cour	213 & 6	Veg	20 minute	30 minute	Mix veget	To make t	7
10	Capsicum	https://s3	Capsicum	Indian	Main Cour	278 & 2	Veg	15 minute	15 minute	Ingredien	Wash and	8
11	Pizza on T	https://s3	Tawa pizz	Indian	Main Cour	478 & 2	Veg	25 minute	20 minute	for the Piz	For the Pi:	9
12	Doodhi Th	http://gre	Doodhi Th	Indian	Main Cour	197 & 4	Veg	15 minute	10 minute	1 cup grat	1.Mix whe	10
13	Pav Bhaji	https://s3	Pav Bhaji i	Indian	Main Cour	758 & 4	Veg	20 minute	20 minute	15 buns (c	Boil all the	11
14	Palak Pan	https://he	Palak Pan	Indian	Main Cour	117 & 4	Veg	20 minute	40 minute	500grams	Rinse the	12
15	Kosambar	https://s3	Kosambar	Indian	Salad	92 & 1	Veg	30 minute	0 minutes	1 Cup grat	Wash carr	13
16	Punjabi St	http://gre	"Punjabi S	Indian	Main Cour	404 & 2	Veg	10 minute	30 minute	Medium-s	Medium-s	14
17	Veggie De	https://im	A deliciou	Indian	Main Cour	400 & 4	Veg	20 minute	15 minute	Pizza dou	Preheat o	15
18	Spinach &	https://be	Savory stu	Indian	Appetizer	150 & 6	Veg	15 minute	20 minute	Mushroon	Clean mus	16
19	Lentil & V	https://i0.	Hearty an	Indian	Appetizer	250 & 6	Veg	15 minute	30 minute	Lentils,Ca	Saute onic	17
20	Quinoa Sa	https://m	Refreshin	Indian	Salad	300 & 4	Veg	20 minute	15 minute	Quinoa,Cu	Cook quin	18
21	Chickpea :	https://w	Quick and	Indian	Main Cour	350 & 4	Veg	15 minute	20 minute	Chickpeas	Stir-fry gir	19
22	Eggplant F	https://cd	Classic Ita	Indian	Main Cour	450 & 4	Veg	30 minute	40 minute	Eggplant,I	Bread and	20
23	Black Bear	https://im	Spicy and	Indian	Main Cour	320 & 4	Veg	25 minute	15 minute	Black bear	Cook blacl	21

Figure 2. Dataset.

```
In [70]: image_features = []
         cnn_model.eval()

Out[70]: RecipeCNN(
           (conv1): Conv2d(3, 16, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
           (pool): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
           (fc1): Linear(in_features=200704, out_features=128, bias=True)
           (fc2): Linear(in_features=128, out_features=64, bias=True)
           (fc3): Linear(in_features=64, out_features=59, bias=True)
         )
```

Figure 3. CNN model.


```
# Example: Recommend recipes similar to recipe at index 0
recommended_recipes = recommend_recipes_by_image_similarity(0, num_recommendations=5)
print("Recommended Recipes based on Image Similarity:")
print(recommended_recipes)
```

```
Recommended Recipes based on Image Similarity:
               name calories & servings  \
572     Mango Curry            250 & 4
155    Carrot Halwa            512 & 3
606    Mutton Curry            271 & 3
293      Bread Upma            200 & 4
645  Chicken Lazeez            375 & 3

                                           ingredients  \
572  2 ripe mangoes, peeled and diced , 1 onion, fi...
155  4 C Carrots, peeled - thickly grated, 1 Tbsp G...
606  Mutton - 500 gms, boneless, cleaned, washed,On...
293  6 slices of bread, preferably whole wheat , 2 ...
645  500g chicken pieces (boneless or with bones),1...

                                           instruction
572  Heat the vegetable oil in a large skillet over...
155  In a non stick wide bottomed pot, add ghee, as...
606  Heat oil or ghee in a large, heavy-bottomed po...
293  Trim the edges of the bread slices and cut the...
645  Heat oil or ghee in a pan and add cumin seeds....
```

Figure 4. Recommendation using the CNN model.

Table 1. Metrics used for evaluating the recommendation system in terms of ranking relevant items (considering the top 5 recommendations).

Evaluation metrics	Scores
Average Recall@5	0.004340277777777777
Average MRR	1.0
Average MAP@5	1.0
Average NDCG@5	1.0

Figure 5 shows the output of the rnn_model.summary() function in Keras. This function displays a summary of the model's architecture, detailing the types and quantities of layers, output shapes for each layer, and the total number of trainable parameters. The model in the image is a sequential model with three layers.

- Embedding layer with 415 inputs and 200 outputs. This layer converted each input word into a dense vector with 200 dimensions.
- LSTM layer with 200 units. This layer learns long-range dependencies in the sequence data.
- A dense layer with 3,043 outputs is designed to predict the next word in the sequence.
- The model had 1541043 trainable parameters. This model is employed for NLP tasks, such as text classification or machine translation.

Table 2 lists the metrics used to evaluate the trained RNN model, which is used for the prediction of the recipe by accuracy, f1-score, and loss.

These metrics can be used to enhance the efficiency of the model in the future. Figure 6 shows the home page of our system where the user searches for recipes, and Figure 8 shows the output of the recipe that the user had searched for. The user must first log in to land on our home page, as shown in Figure 6, where he/she will search for recipes that will land on the recipe generation page [21–25].


```
rnn_model.summary()

Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding_1 (Embedding)     (None, 415, 200)          608600

 lstm_1 (LSTM)               (None, 200)               320800

 dense_1 (Dense)             (None, 3043)              611643

=================================================================
Total params: 1541043 (5.88 MB)
Trainable params: 1541043 (5.88 MB)
Non-trainable params: 0 (0.00 Byte)
```

Figure 5. RNN model.

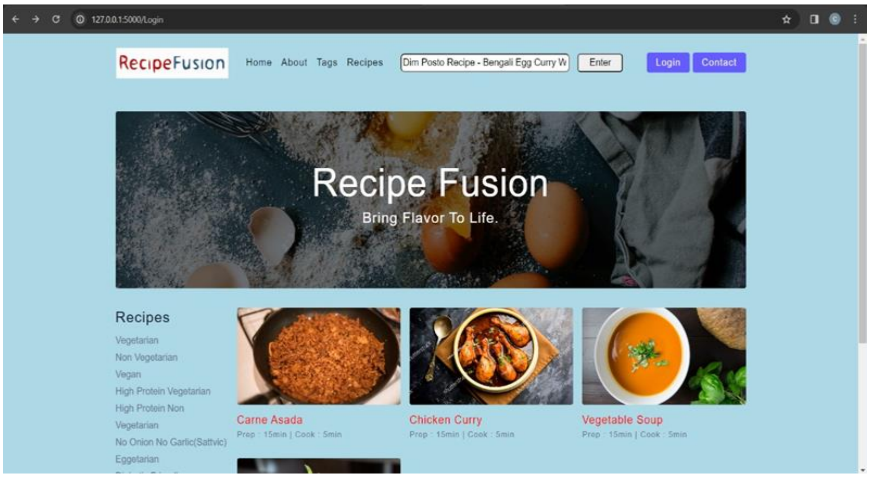

Figure 6. Home page of the website where the user will search for the recipe.

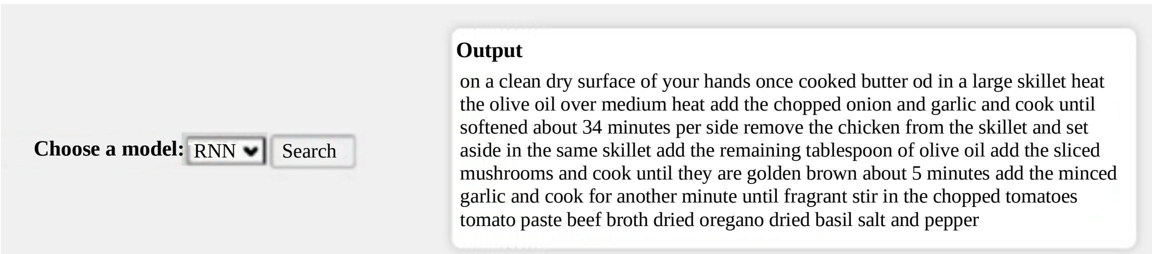

Figure 7. Prediction of instructions of the recipe using RNN.

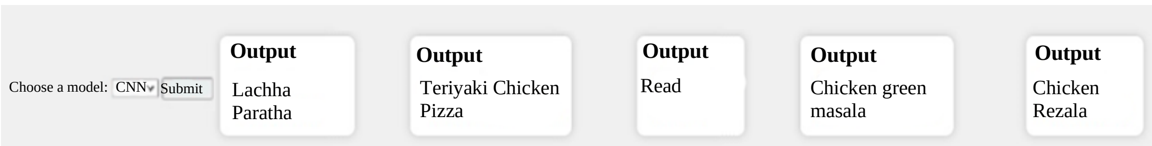

Figure 8. Recommendation of recipe name using CNN based on the input given by the user.

In Figure 7, there is an option to select the CNN or RNN model for recommendation or prediction of the mentioned recipe.

Figure 8 shows the recommended recipe names that were generated once the user selected the CNN model, and Figure 9 shows the instructions for the recommended recipes using the CNN model. Figures 8 and 9 contain five recommended recipes based on the input given by the user as a recipe name.

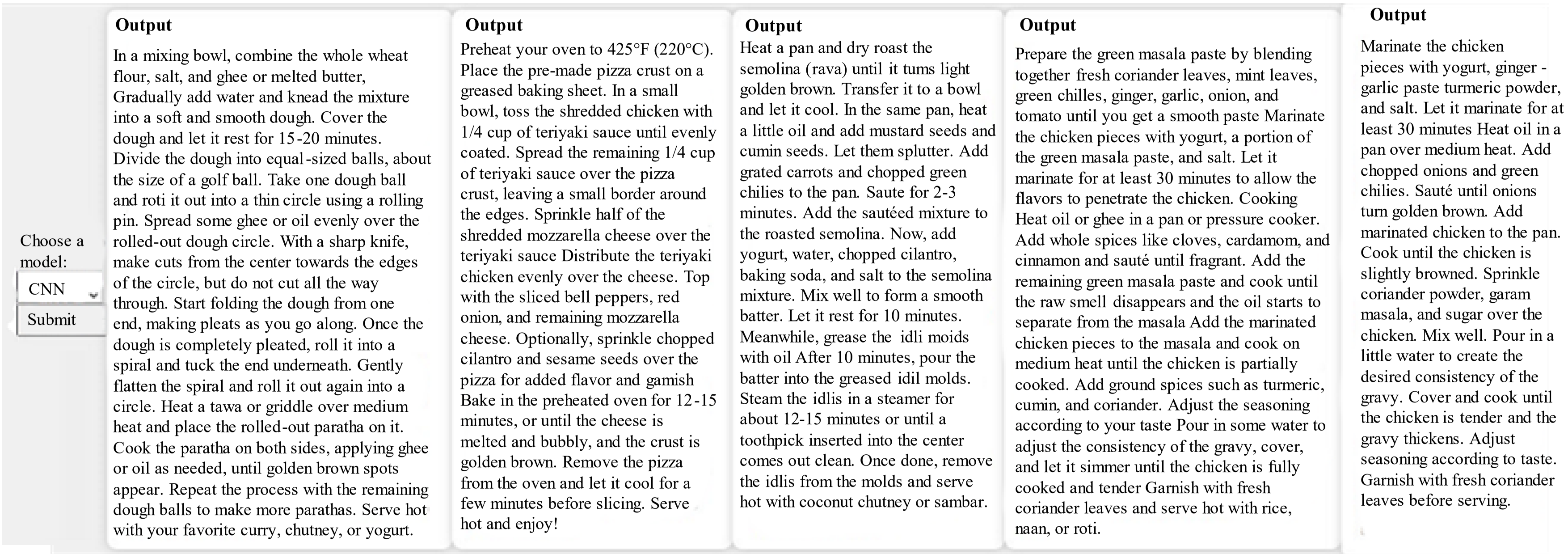

Figure 9. Recommended recipe instructions using CNN.

Table 2. RNN model evaluation.

Training	Testing
Accuracy= 82.04%	Accuracy= 53.6%
Loss= 0.95295	F1 Score= 52.4%

CONCLUSION

We implemented this system to simplify the users' need to cook food. This will help people find various food recipes without spending a lot of time. Our system enhances the user's cooking experience and simplifies the recipe discovery process. The system recommends recipes that match the user's requirements. The system provides users with a personalized and seamless recipe discovery experience by leveraging multiple modalities, such as text, images, and user preferences. The system will help the user to try new and innovative recipes. This will help users to learn and cook recipes efficiently. This system will help every person, whether it is a bachelor' or a chef. While building this system, we learned various techniques that we would not have discovered.

Future Work

The following features can be improved in our system:

- Improve the accuracy of the model.
- Introduce more features like the calorie count of foods with recipes.
- Make recipes available in various languages to help people from various regions understand them.
- Allow users to capture images of the ingredients to search for the recipe instead of typing the recipe.
- Apply other machine learning approaches and compare the accuracy.
- Introduce a feature where a user can provide input in the form of multiple modalities- text, image, and audio.

REFERENCES

1. Patel R, Thakkar P, Ukani V. CNNRec: Convolutional neural network-based recommender systems: A survey. Eng Appl Artif Intell. 2024;133:108062. doi:10.1016/j.engappai.2024.108062.
2. Forouzandeh S, Rostami M, Sheikhpour R. Hfrs-Han: Health-aware food recommendation system based on the heterogeneous attention network. SSRN. 2023. doi:10.2139/ssrn.4378052.
3. Khilji AFUR, Sinha U, Singh P, Ali A, Dadure P, Manna R, Pakray P. Multimodal recipe recommendation system using deep learning and rule-based approach. SN Comput Sci. 2023;4:421. doi:10.1007/s42979-023-01870-6.
4. Castiglia G, El Majjodi A, Starke AD, Narducci F, Deldjoo Y, Calò F. Nudging towards health in a conversational food recommender system using multi-modal interactions and nutrition labels. 4th edn of Knowledge-aware and Conversational Recommender Systems (KaRS) Workshop @ RecSys, Seattle, USA, September 2023.
5. Zhang J, Li M, Liu W, Lauria S, Liu X. Many-objective optimization meets recommendation systems: A food recommendation scenario. Neurocomputing. 2022;503:109-117. doi:10.1016/j. neucom.2022.06.081.
6. Jia N, Chen J, Wang R. An attention-based convolutional neural network for recipe recommendation. Expert Syst Appl. 2022;201:116979. doi:10.1016/j.eswa.2022.116979.
7. Lei Z, Ul Haq AU, Zeb A, Suzauddola M, Zhang D. Is the suggested food your desired?: Multi-modal recipe recommendation with demand-based knowledge graph. Expert Syst Appl. 2021;186:115708. doi:10.1016/j.eswa.2021.115708.
8. Chavan P, Thoms B, Isaacs J. A recommender system for healthy food choices: Building a hybrid model for recipe recommendations using big data sets. In: Proc 54th Hawaii Int Conf Syst Sci, HICCS-54, CA, USA, Jan 2021. doi:10.24251/HICSS.2021.458.
9. Manna R, Das D, Gelbukh A. Question-answering and recommendation system on cooking recipes. Computacion y Sistemas. 2021;25:223-235. doi:10.13053/cys-25-1-3899.

10. Wang W, Duan LY, Jiang H, Jing P, Song X, Nie L. Market2Dish: Health-aware food recommendation. ACM Trans Multimedia Comput Commun Appl. 2021;17:1-19. doi:10.1145/3418211.
11. Pham HX, Guerrero R, Pavlovic V, Li J. CHEF: Cross-modal hierarchical embeddings for food domain retrieval. In: Proc AAAI Conf Artif Intell. 2021 May 18;35:2423-2430. doi:10.1609/aaai.v35i3.16343.
12. Deldjoo Y, Schedl M, Hidasi B, Wei Y, He X. Multimedia recommender systems: Algorithms and challenges. In: Recommender Systems Handbook. Springer: New York, USA; 2021. p. 973-1014.
13. Marin J, Biswas A, Ofli F, Hynes N, Salvador A, Aytar Y, Weber I, Torralba A. Recipe1M+: A dataset for learning cross-modal embeddings for cooking recipes and food images. arXiv Preprint arXiv:1810.06553. 2018 Oct 14.
14. Xie Z, Liu L, Wu Y, Li L, Zhong L. Learning TFIDF enhanced joint embedding for recipe-image cross-modal retrieval service. IEEE Trans Serv Comput. 2021;15:3304-3316. doi:10.1109/TSC.2021.3098834.
15. Tian Y, Zhang C, Metoyer R, Chawla NV. Recipe representation learning with networks. In: Proc 30th ACM Int Conf Inf Knowl Manag. 2021 Oct 26;1824-1833. doi:10.1145/3459637.3482468.
16. Chen Y, Zhou D, Li L, Han JM. Multimodal encoders for food-oriented cross-modal retrieval. In: Web and Big Data. Proc 5th Int Joint Conf, APWeb-WAIM 2021, Guangzhou, China, 23–25 Aug 2021, Vol 5. Springer International Publishing; 2021. p. 253-266.
17. Pan L, Chen J, Liu S, Ngo CW, Kan MY, Chua TS. A hybrid approach for detecting prerequisite relations in multi-modal food recipes. IEEE Trans Multimedia. 2020;23:4491-4501. doi:10.1109/TMM.2020.3042706.
18. Pecune F, Callebert L, Marsella S. A recommender system for healthy and personalized recipes recommendations. In: HealthRecSys@ RecSys. 2020 Sep 26;15-20.
19. Rostami A, Pandey V, Nag N, Wang V, Jain R. Personal food model. In: RT-PA of the 28th ACM Int Conf Multimedia. 2020 Oct 12;4416-4424.
20. Theodoridis T, Solachidis V, Dimitropoulos K, Daras P. A cross-modal variational framework for food image analysis. In: 2020 IEEE Int Conf Image Process (ICIP). IEEE Publications; 2020. p. 3244-3248. doi:10.1109/ICIP40778.2020.9190758.
21. Twomey N, Fain M, Ponikar A, Sarraf N. Towards multi-language recipe personalisation and recommendation. In: Proc 14th ACM Conf Recommender Syst. 2020 Sep 22;708-713. doi:10.1145/3383313.3418478.
22. Lin AS, et al. A recipe for creating multimodal aligned datasets for sequential tasks. arXiv Preprint arXiv:2005.09606. 2020.
23. Liu A, Yuan S, Zhang C, Luo C, Liao Y, Bai K, Xu Z. Multi-level multimodal transformer network for multimodal recipe comprehension. In: Proc 43rd Int ACM SIGIR Conf Res Dev Inf Retrieval. 2020 Jul 25;1781-1784. doi:10.1145/3397271.3401247.
24. Meng L, Feng F, He X, Gao X, Chua TS. Heterogeneous fusion of semantic and collaborative information for visually aware food recommendation. In: Proc 28th ACM Int Conf Multimedia. 2020 Oct 12;3460-3468. doi:10.1145/3394171.3413598.
25. Fu H, Wu R, Liu C, Sun J. Mcen: Bridging cross-modal gap between cooking recipes and dish images with latent variable model. In: Proc IEEE/CVF Conf Comput Vis Pattern Recognit. 2020;14558-14568. doi:10.1109/CVPR42600.2020.01458.

