FoodieNet: Transforming Food Images into Nutritional Insights with a Focus on Allergenic Substances

Dongin Kim
dkim949@gatech.edu

Bixi Liu
bliu426@gatech.edu

Sangjun Lee
slee3426@gatech.edu

Tetsuya Nakajima
tnakajima7@gatech.edu

Abstract

There is growing interest among individuals to make informed nutritional decisions, especially for those with food allergies. The primary objective of this paper is to tackle the difficulties associated with automatically guessing ingredients from food images and establishing connections with respective allergen information. Leveraging inspiration from the InverseCooking model by Facebook Research [7], we introduce FoodieNet, a modified architecture designed to enhance the performance of ingredient inference and allergen relation. Our approach involves 1) significant modifications to the image encoder of the original networks by replacing with CLIP as well as 2) the incorporation of a Large Language Model (LLM) at the end of the architecture. Despite facing limitations in local computing power and storage, FoodieNet surpasses the original Inverse Cooking model's image encoder, exhibiting remarkable improvements in 1) detecting 41 additional classes of ingredients that InverseCooking missed, specifically allergens such as lobster, crab, octopus, squid, and red snapper 2) improved F1 score on 56% of ingredient classes. These achievements demonstrate FoodieNet's enhanced ability to automatically generate valuable insights about ingredients and their relation to major allergens. The implications of this research extend to applications in dietary analysis, food labeling, and allergen-aware meal planning.

1. Introduction

1.1. Motivations

In the current era, there is a growing inclination among individuals to make informed nutritional choices, particularly driven by an increasing awareness of the significance of dietary decisions. This trend is particularly noteworthy for those managing food allergies, considering that recent surveys indicate nearly 6.2% of adults in the U.S. grapple with food allergies, with half of them developing new allergies not present in childhood [5]. The motivation here is to supply supplementary information about dietary choices, aiming to not only promote informed nutritional decisions but also prevent potential health risks associated with consuming allergens.

1.2. Objectives

The primary objective of this project is to leverage the capabilities of deep neural networks to predict ingredients and allergenic information from food images. The overarching goal is to empower consumers with enhanced knowledge about the nutritional content of their food. This predictive model envisions practical application through a web or mobile platform, allowing users to input a food image and automatically receive comprehensive information about ingredients and major allergens. The anticipated outcome is a user-friendly tool that contributes to promoting dietary awareness and facilitates smarter food choices.

2. Related Work

There have been several attempts to recognize ingredients from food images utilizing convolutional neural networks. The Food-101 dataset was introduced by Bossard et al. [1] in 2014 to perform food categorization. It consists of 101 food categories with 101,000 food images. In 2019, Marin et al. [4] introduced Recipe 1M+, a new large-scale, structured corpus of over 1M cooking recipes and 13M food images. Recipe 1M+ is the largest publicly available collection of food image and recipe data. Since the release of Recipe 1M+, several studies have been done to improve upon the initial research on food image recognition and recipe retrieval. The current state-of-the-art model for image to recipe retrieval is VLPCook [8], which achieves an R@1 of 74.9% on image-to-recipe tasks based on Recipe 1M+ dataset and an accuracy of 89.14% on food recognition tasks on Food-101 dataset. InverseCooking [7] remains the only approach using Visual Set Transformer to predict ingredients from images.

3. Approach

There are two approaches in image to ingredient prediction, namely Generation vs Retrieval.

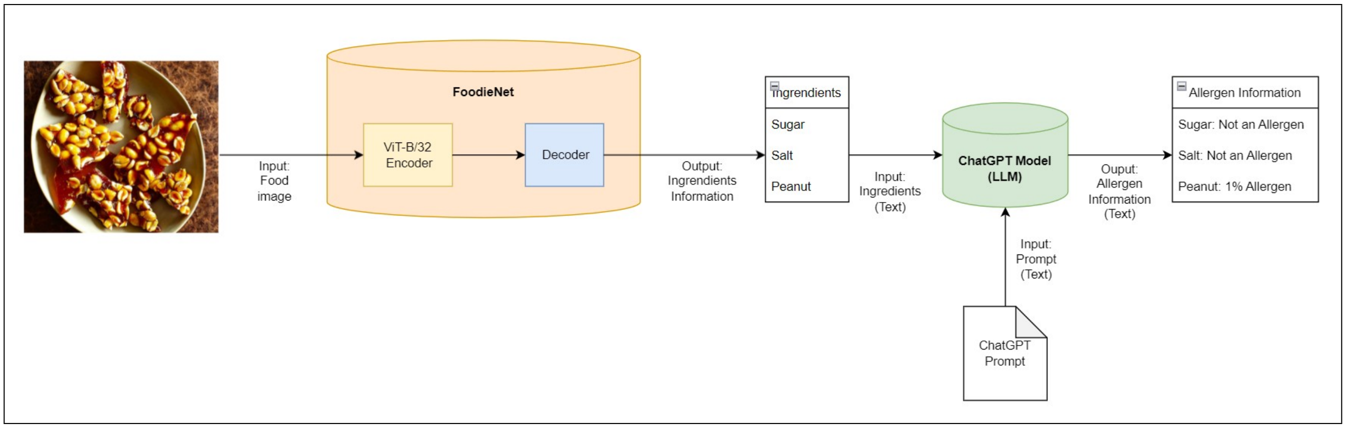


Figure 1: FoodieNet: Image to Allergen Information generation model. We extract image features with CLIP Visual (Vit-B/32) Encoder. Ingredients are predicted by the ingredients decoder. Then output of ingredients information was prompted to the LLM model to generate Allergen Information.

3.1. Retrieval

Retrieval method involves comparing features from the input food image with the features extracted from pre-existing images in the database. The system then retrieves the ingredients and recipes corresponding to the most similar images in the dataset. Image features are extracted from Deep Learning models.

The main challenge is that the accuracy of model output is highly reliant on the completeness of existing data in the dataset, and the accuracy of feature extraction and comparison. The Retrieval method is limited to existing database entries and cannot generate new recipes.

3.2. Generation

Contrastingly, Generation models employ a transformer decoder architecture, trained to predict ingredients based on features derived from images. This approach enables the identification of ingredients from previously unseen images. In our research, we have implemented the Generation model by developing a Visual Set Transformer. This model processes inputs from an image encoder and predicts ingredients using an ingredient decoder, thus offering a more dynamic and expansive method for ingredient prediction.

3.3. Implementation Details

Our model architecture (Figure 1) is built based of InverseCooking code base[1]. We followed the same data augmentation approach as [7] by resizing the images to 256 pixels and taking random crops of 224 x 224 for training.

Multimodal models trained on both images and text have become the key to obtain SoTA results in recent research and development. Inspired by VLPCook [8] where the author achieved significant uplift in model results on a retrieval model by adopting Vision-Language Pre-Training model, we undertook a similar approach for our ingredient generation task.

For image embeddings, we replaced the last convolution layer of the ResNet-50 model with Visual Transformer (ViT-B/32) from the pretrained Constructive Language-Image Pre-training (CLIP) [6] model. We maintained the frozen state of ViT-B/32 model parameters throughout training and introduced an auxiliary linear projection layer atop ViT-B/32 for fine-tuning purposes. Implementation of CLIP was inspired by LLaMA-Adapter[2] [3]. The models were trained with Adam Optimizer until early-stopping criteria was met.

CLIP [6] was trained on a large scale of image and textual description pairs. This enables the model to understand and relate visual concepts in the context of natural languages. ViT-B/32 is based on the Transformer architecture which has shown remarkable success in understanding contextual relationships in data. Unlike ResNet-50, ViT processes the entire image in patches and captures global context, allowing it to understand more complex and abstract visual patterns. Food images often contain complex context, CLIP model has shown significant advantage in capturing the complex context presented in food images by comparing Zero-Shot CLIP classifier model on Food-101 dataset with a supervised linear classifier fitted on ResNet-50 features. CLIP model score outperforms ResNet 50 by 22.5%.

Vision transformers [2] are also more efficient than CNNs in processing large models, this is reflected in model training time in Table 4. While CNN trains faster than Vit-B/32 on smaller sample (4,002 image recipe pairs),

[1] https://github.com/facebookresearch/inversecooking

[2] https://github.com/OpenGVLab/LLaMA-Adapter

when training on larger sample (25,052 image recipe pairs), model training with ViT-B/32 visual encoder is 30% faster than CNN model with ResNet-50 encoder.

The updated model using ViT-B/32 visual encoder also showed significant improvement in the quality of ingredient prediction.

3.4. Ingredients Generation

We have kept the same ingredients decoder architecture as [7], which is a Set Transformer where Softmax probabilities are pooled across time to avoid penalizing for order [7]. Ingredients are selected without repetition. The outputs are then aggregated across different time steps by max pooling operation. The model was trained by minimizing binary cross-entropy loss between predicted ingredients and the ground truth. To learn when ingredient generation should stop, an additional eos loss is added to the model. A cardinality penalty was also introduced as a regularization term that penalizes the model's predictions if the number of predicted ingredients deviates significantly from what is expected. This forces the model to predict sets of ingredients with a more realistic size.

For our team project, we implemented an image to text generation model that was trained on the Recipe 1M dataset and used LLM to map ingredients output to allergens (Figure 1). To be more specific, our aim is to replace the original encoder structure detailed in InverseCooking [7] with CLIP [6] vision encoder to improve its performance.

	Recipe1M+	InverseCooking	Our model
Image Encoder	VGG-16 ResNet-50	ResNet-50	ViT-B/32
Ingredients Encoder	Bidirectional LSTM	N/A	N/A
Ingredients Prediction	Through Retrieval	Decoder	Decoder

Table 1: Model architecture comparison

3.5. LLM Prompt

The outputs from the image-to-ingredient model were then used to prompt Large Language Models (LLMs) to identify potential allergenic substances through few-shot learning. We provided the LLMs with example inputs and expected outputs to generate information on allergens, symptoms, and the percentage of the population affected.

Example Input: pepper, oil, peanut

Example Output: Pepper: Not an allergen.

Oil: Not an allergen.

Peanut: Allergen information: Approximately % of the population is allergic to peanuts. Symptoms include ...

4. Experiments

In this section, we will discuss different experiment set ups, challenges, evaluation results and compare it with the baseline InverseCooking model. All experiments were trained with early stopping criteria, to prevent overfitting. We compared and evaluated different models and experiment results using F1 score and Intersection Over Union (IoU) score on the test set.

4.1. Data

Our model training and evaluation were conducted on the Recipe1M dataset [4]. Original dataset has 720,639 training, 155,036 validation and 154,045 test recipes. To ensure data quality, we initially filtered out recipes with either too few (less than 2) or too many (more than 20) ingredients or instructions, following a similar approach as in [7]. This process resulted in 645,114 training, 138,743 validation, and 138,070 test recipes. Due to computational constraints, we further randomly downsized the training data into 10,000 and 60,000 samples to experiment with different model architecture and hyper-parameter set ups before training on the full sample.

Many of the original Recipe1M recipes do not contain matching images, approximately 40% of the recipes contain matching images. For training, only recipes with corresponding images were considered.

Experiment name	Sample size	Training size (Image recipe pairs)
10K	10,000	4,002
60K	60,000	25,052
FULL	645,114	252,547

Table 2: Experiment training size comparison

In addition, we applied the same data augmentation techniques as [7]. The original Recipe1M dataset contained 16,823 ingredients. We adopted the referenced approach, which involved: Eliminating plurals and words appearing less than 10 times. Merging ingredients that either began or finished with the same last two words. Clustering ingredients with shared starting or ending words. These preprocessing steps condensed the ingredient list to a more manageable 1,488.

4.2. Challenges

One of the foremost challenges we faced was the extended duration of training time. Our experimentation revealed various constraints associated with the Windows Operating System.

LMDB files: Converting images into LMDB (Lightning Memory-Mapped Database) files is a common practice in

OS	GPU	Experiment name	Image encoder	Batch Size	F1	IoU	Training time (40 epochs)
Windows	3070	CNN_10K	ResNet-50	150	36.03%	22.47%	1.15 hrs
Windows	3070	CLIP_10K	ViT-B/32	150	40.23%	25.60%	1.88 hrs

Table 3: Experiments ran on 10K sample (4,002 image recipe pairs used in the image-ingredient model)

OS	GPU	Experiment name	Image encoder	Batch Size	F1	IoU	Training time (40 epochs)
Windows	3070	CNN_60K	ResNet-50	150	37.17%	23.25%	19.24 hrs
Windows	3070	CLIP_60K	ViT-B/32	150	44.74%	29.32%	12.71 hrs
Linux	4070Ti	CNN_60K	ResNet-50	64	38.05%	23.82%	46.94 mins
Linux	4070Ti	CLIP_60K	ViT-B/32	64	44.70%	29.26%	42.58 mins
Linux	4070Ti	CNN_60K	ResNet-50	128	37.56%	23.49%	38.19 mins
Linux	4070Ti	CLIP_60K	ViT-B/32	128	45.00%	29.53%	34.59 mins
Linux	4070Ti	CNN_60K	ResNet-50	256	37.46%	23.41%	37.35 mins

Table 4: Experiments ran on 60K (25,052 image recipe pairs used in the image-ingredient model)

training image-based deep learning models involving images. However, we noted a significant disparity in file handling between operating systems. Windows tends to preemptively allocate disk space equivalent to the size of the original input map, which can result in the unnecessary use of a large amount of disk space. On the other hand, Unix systems incrementally expand the LMDB file size, scaling up to match the input map size.

num_workers in DataLoader: Multiprocessing limitations on Windows systems necessitates setting num_workers to 0. This restriction results in a marked increase in data loading times when compared to the Unix system. Midway through training we gained access to Linux systems with superior GPU configuration, which yielded a notable enhancement in performance. It is important to note that training times are only directly comparable when conducted within the same operating system environment. Given the constraints of our computing resources, we decided to experiment the new architecture on a smaller training dataset as a starting point and gradually increase the sample size.

4.3. Training Process

For all of the experiments, we used a learning rate of 1e-4 and learning decay rate of 0.99.

10K sample: Table 3 outlines the comparative performance of models trained using ResNet-50 and ViT-B/32 as image encoders. The model with the ViT-B/32 encoder demonstrated notable gains, with an 11.66% increase in the F1 score and a 13.93% rise in the IoU score when contrasted with its ResNet-50 encoder counterpart. It is important to note, however, that the training duration for the ViT-B/32 encoder model was 63.48% longer than that for the ResNet-50 encoder.

In light of these findings, we decided to increase the sample size to 60K and experiment with different hyper parameter settings.

60K sample: Table 4 provides a side-by-side performance evaluation of models that were trained using ResNet-50 and ViT-B/32 as image encoders under various hyper parameter settings. Experiment with 128 batch size achieved the highest F1 score and IoU on the test sets for both models. Specifically, the model using ViT-B/32 encoder achieved a 19.81% higher F1 score and 25.71% higher IoU. Despite the extended training time observed in the previous experiment, increasing the sample size by sixfold resulted in training time reduction of up to 33.94% on Windows OS, and 9.43% on a Linux OS. This indicates that Vision Transformers are more efficient when applied to larger datasets.

Full sample: Memory limitations prevented us from experimenting with larger batch sizes, so we proceeded with the optimal hyperparameter settings identified during the 60K sample trials. Table 5 presents a comparative analysis of evaluation metrics against the baseline InverseCooking model. While our model exhibited a modest decrease in F1 score of 1.38% when measured across the entire test set, it achieved a marginally higher IoU of 0.75%, and demonstrated a substantial enhancement in prediction quality.

4.4. Results

Our model demonstrated an enhanced ability to detect a wider range of ingredients compared to the InverseCooking model. Overall F1 score is weighted by how frequently an ingredient occurs across all test data. The InverseCooking

OS	GPU	Experiment name	Image encoder	Batch Size	F1	IoU	Training time (40 epochs)
Linux	**4070Ti**	**CLIP_FULL**	**ViT-B/32**	**128**	**48.00%**	**32.35%**	**5.32 hrs**
Linux	Unknown	InverseCooking	ResNet-50	300	48.61%	32.11%	Unknown

Table 5: Experiments ran on the full training sample (252,547 image recipe pairs used in the image-ingredient model)

model has a higher F1 score on some of the most frequently used ingredients in cooking i.e. salt, pepper, cheese. Despite not surpassing the SoTA F1 score evaluated on the full set of ingredients, when assessing F1 score on individual ingredient classes, our model excelled in recognizing 10% more ingredient classes. Our model predicted 312 classes of ingredients with positive F1 score, outperforming the InverseCooking model's prediction of 284 classes.

Our model successfully identified numerous ingredients that are known to trigger fish and shellfish allergies which InverseCooking model had overlooked. Among the 41 newly detected classes of ingredients, several are recognized allergens, such as lobster, crab, octopus, squid, red snapper. Appendix C includes a full list of 41 newly detected classes of ingredients by FoodieNet.

For ingredients that were predicted by both models, ours show a superior F1 score for a greater number of ingredients. Upon removing ingredients where both models scored an F1 of zero, our model outperformed InverseCooking by achieving a higher F1 score for 56% of the remaining ingredients.

4.5. Prompt Engineering

After obtaining the best image-to-ingredient model, we experimented with several LLMs offered by OpenAI, including gpt-4-1106-preview, gpt-4, gpt-4-0613, gpt-4-0314, and gpt-3.5-turbo, and compared their outputs qualitatively. Thus far, gpt-4-1106-preview has produced the most satisfying results. Appendix D provides a comparison of different LLM outputs on list of sample input images.

Following extensive testing with various prompts, it was established that employing precise and straightforward language is the most effective strategy to guide ChatGPT toward generating desired results. As a result, the prompt has been meticulously formulated to incorporate a clear and direct desired output format, thereby ensuring more accurate and relevant responses from the Language Model.

In the context of prompt engineering within our research, our primary goal was to provide explicit instructions to the machine, outlining specific guidelines for its designated tasks. Our approach involved defining particular conditions and elucidating the information that would be supplied to the model. Stressing the significance of clarity and adherence to a predefined format, we tasked the model with generating responses that center on the most critical allergens with substantial impacts on the human body. To refine our expectations, we concluded the prompt engineering process by furnishing a concrete example input and output, thereby further specifying the desired outcome to enhance precision in the model's responses.

5. Sample Outputs

5.1. Ingredients Output

Input image	Our Model	InverseCooking
	pepper, oil, parsley, **lobster**, clove, **butter**, salt	oil, pepper, salt, parsley, **butter**, clove

Table 6: Ingredient output comparison

Both models predicted common ingredients such as pepper, oil, parsley, salt, clove, and butter. However, the InverseCooking model failed to identify the key ingredient, lobster, which is a source of shellfish allergy that could lead to severe and even life-threatening reactions like anaphylaxis.

5.2. LLM Generated Allergen Alerts

1) Pepper: Not an allergen

2) Oil: Not an allergen (with exceptions)

3) Parsley: Not an allergen

4) Lobster: Allergic information: Lobster is a type of shellfish, which is a known allergen. Approximately 2% of the population is allergic to shellfish, but it is important to note that shellfish allergies are generally more common in adults than children. Shellfish allergy symptoms can range from mild reactions, such as hives or a stuffy nose, to severe and even life-threatening reactions like anaphylaxis. It is advised for individuals with shellfish allergies to avoid all types of shellfish.

5) Clove: Not an allergen

6) Butter: Allergic information: Butter is a dairy product and contains milk proteins, which can be allergenic. Around 0.2-0.5% of adults and 2.5% of children are allergic to cow's milk. The majority of children outgrow their

milk allergy by age five. Milk allergy symptoms can include hives, vomiting, digestive problems, and in severe cases, anaphylaxis. Individuals with a milk allergy should avoid butter and other dairy products to prevent allergic reactions.

7) Salt: Not an allergen

6. Conclusion

In conclusion, this paper presents FoodieNet, a novel approach to automatically deduce ingredients from food images and establish connections with allergen information, addressing the growing need for informed nutritional decisions, especially among individuals with food allergies. By enhancing the original Inverse Cooking model with a modified architecture featuring CLIP-based image encoding and the incorporation of a Large Language Model (LLM), FoodieNet surpasses previous versions in several critical metrics, including the F1 Score for the majority of ingredient categories. The experimentation with different sample sizes overcomes local computing constraints and showcases FoodieNet's scalability and potential for widespread application. It is important to note that while the immediate objective may not be to offer precise information for patients with pathological allergies, the research lays the foundation for future enhancements. Specifically, expanding the training dataset beyond Western-style diets to encompass a variety of cultural food images is expected to extend the model's applicability and relevance in diverse nutritional contexts. This research contributes not only to the advancement of automated ingredient inference but also sets the stage for a more inclusive and versatile application of such models in the realm of dietary awareness and allergen-conscious meal planning. Additionally, it encourages users to verify health-related information, crucial in preventing potentially life-threatening allergic reactions.

References

[1] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. "Food-101 – Mining Discriminative Components with Random Forests". In: *European Conference on Computer Vision*. 2014.

[2] Alexey Dosovitskiy et al. *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. 2021. arXiv: 2010.11929 [cs.CV].

[3] Peng Gao et al. "LLaMA-Adapter V2: Parameter-Efficient Visual Instruction Model". In: *arXiv preprint arXiv:2304.15010* (2023).

[4] Javier Marin et al. *Recipe1M+: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images*. 2019. arXiv: 1810.06553 [cs.CV].

[5] *More than a quarter of U.S. adults and children have at least one allergy.* Accessed: 2023-12-10. URL: https://www.cdc.gov/nchs/pressroom/nchs_press_releases/2022/20220126.htm.

[6] Alec Radford et al. *Learning Transferable Visual Models From Natural Language Supervision*. 2021. arXiv: 2103.00020 [cs.CV].

[7] Amaia Salvador et al. "Inverse Cooking: Recipe Generation From Food Images". In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2019.

[8] Mustafa Shukor, Nicolas Thome, and Matthieu Cord. *Vision and Structured-Language Pretraining for Cross-Modal Food Retrieval*. 2023. arXiv: 2212.04267 [cs.CV].

Appendix

A. Team Member Contribution

Student Name	Contributed Aspects	Details
Dongin, Kim	1) Research 2) Model Training 3) Prompt Engineering 4) Paper Write-up	Research on the project topics and possible algorithm options for the implementation. Research on the image encoder - InverseCooking by the Facebook Research. Training the image encoder models and conducting experiments with different hyperparameters. Prompt Engineering on the recipe generator. Writing up the overall research paper.
Sangjun, Lee	1) Research 2) Prompt Engineering 3) Coding 4) Paper Write-up	Researched on project topics and implementation methods. Worked on prompt engineering on the recipe generator, and tested various LLM models to compare results. Developed inferencer.py and openai_text_generator.py for the inferencer module and text generation module. Combined all modules into a single working code. Writing up paper on general architecture and inferencer and LLM part of the project.
Bixi, Liu	1) Research 2) Identify training data 3) Training data preparation 4) Update model architecture to use ViT-B/32 5) Model Training & evaluation 6) Experiment result summarization 7) Overall paper write-up	Researched and compared different approaches to model food images to ingredients generation. Identified Recipe1M training data, and preprocessed data for training. Modified InverseCooking model architecture to replace ResNet-50 with Vit-B/32, trained multiple experiments and compared results. Summarized approaches and findings in the final research paper. Writing up the overall research paper and formatting in LaTeX.
Tetsuya, Nakajima	1) Research 2) Model Training 3) Paper Write-up	Research on the project topics with comparison of related papers. Training the image encoder. Writing up research paper.

Table 7: Contributions of team members.

B. Evaluation Metrics Comparison Between Runs

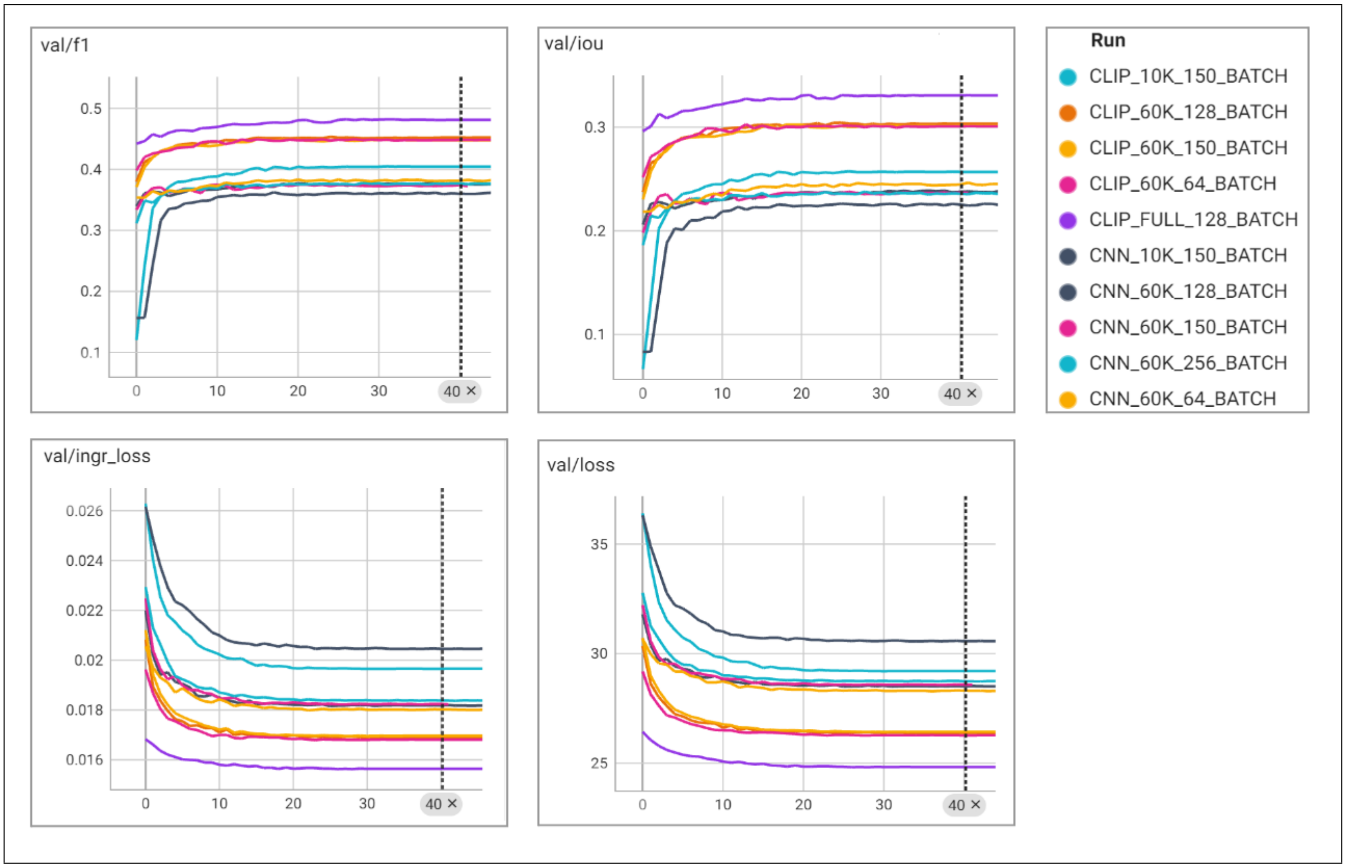


Figure 2: Evaluation metrics comparison between experiments on the validation dataset

C. Predicted Ingredients Comparison

41 classes of ingredients detected by FoodieNet model but missed by InverseCooking

Frisee , miniature_phyllo_cup , radicchio , bagel , tostadas , champagne , tater_tots , lobster , octopus , endive , plantain , red_snapper , maraschino_cherry , spring_roll_wrappers , croissant , crab , grits , kahlua , plum , lavender , stuffing_mix , bulgur , frosting , pesto_sauce , wrap , squid , melon , gluten , wheat , parsnip , marinara_sauce , wasabi , espresso , bran , cashews , fruit , cardamom , beer , meat , hot_sauce , cooking_spray.

13 classes of ingredients detected by InverseCooking model but missed by FoodieNet

Turnip , catalina_dressing , pancake , apricot , cacao , miracle_whip , cornbread , lemongrass , chile , horseradish , bay_leaf , allspice, chives.

D. LLM Model Output Comparison

LLM Model	Output
gpt-3.5-turbo	1) Sugar: Not an allergen Sugar, specifically refined white sugar, is not a commonly known allergen. Allergic reactions to sugar are extremely rare and usually attributed to other components within the product that contains sugar, such as preservatives or additives. 2) Butter: Not an allergen Butter is not commonly known to cause allergies. However, some individuals who are allergic to dairy may experience allergic reactions to butter due to its dairy content. It is important for individuals with dairy allergies to avoid butter and opt for dairy-free alternatives. 3) Salt: Not an allergen Salt, which is primarily composed of sodium chloride, does not typically cause allergic reactions. However, in rare cases, individuals may be allergic to certain impurities or additives found in specific brands of salt. It is always advisable to check the ingredients label for any potential allergens present in salt substitutes or flavored salts. 4) Syrup: Not an allergen Generic syrup, such as corn syrup or maple syrup, is not an allergen itself. However, flavored syrups, especially those with additives or natural flavors, may contain allergens such as nuts or dairy. It is important to carefully read the ingredient labels of syrups to determine if any allergenic ingredients are present. 5) Peanuts: Allergic information: Approximately 0.6-1.4% of the population is allergic to peanuts. Symptoms include itching or tingling in the mouth, hives, swelling, difficulty breathing, abdominal pain, vomiting, and anaphylaxis in severe cases. Peanut allergies can be life-threatening and individuals with peanut allergies must strictly avoid peanuts and products containing peanuts. It is essential to carefully read food labels and be cautious of cross-contamination in food preparation environments. Epinephrine auto-injectors are often carried by individuals with peanut allergies in case of accidental exposure.
gpt-4	1) Sugar - Not an allergen: Sugar is generally not an allergen and reactions are often a result of intolerance rather than an allergy. However, some people may have hypersensitivity or adverse reactions to specific types of sugar such as fructose or glucose. Such cases are not common. 2) Butter - Allergen: Approximately 2-3% of the population is allergic to milk and dairy products including butter. This is more common in infants and children, although most outgrow it by age 16. This allergic reaction is caused by the proteins in milk such as casein and whey. Symptoms commonly include hives, wheezing, vomiting, diarrhea and anaphylaxis in severe cases. 3) Salt - Not an allergen: Salt, whether sea salt, table salt, or kosher salt, isn't an allergen. Some people may have a sensitivity or an intolerance toward certain types of salt, but it does not cause an immune response. 4) Syrup - Conditional allergen: Most syrups, such as corn syrup or cane syrup, are not allergens. However, if the individual is allergic to a specific ingredient in the syrup, such as corn, then an allergic reaction may occur. Additionally, some individuals may have an intolerance or sensitivity to high levels of sugar found in syrups. 5) Peanuts - Allergen: Approximately 1-2%of the population is allergic to peanuts. This type of food allergy is one of the most common and it often lasts into adulthood. The immune system in someone with a peanut allergy has been sensitized to react to the proteins found in peanuts. Symptoms may range from mild, such as hives, to severe, such as anaphylaxis—a potentially life-threatening reaction.

Table 8: LLM Model output comparison when given ingredients of Peanut Brittle - part 1

LLM Model	Output
gpt-4-0314	1) sugar: Not an allergen 2) butter: Allergic information: Butter is primarily made from cow's milk, and contains milk proteins that can cause an allergic reaction. About 2-3% of children and 0.2-0.4% of adults are allergic to cow's milk. Symptoms can range from mild to severe, including hives, wheezing, digestive problems, and anaphylaxis. 3) salt: Not an allergen 4) syrup: Not an allergen. However, some people might have an intolerance or sensitivity to certain types of syrup containing artificial additives, which could cause mild side effects. 5) peanuts: Allergic information: Approximately 1.4-1.6% of the population is allergic to peanuts. Symptoms include hives, swelling, shortness of breath, digestive issues, and anaphylaxis. Peanut allergies are often severe and can be life-threatening.
gpt-4-0613	1) Sugar - Not an allergen 2) Butter - Allergen: Butter contains milk, which is a common allergen. Approximately 2-3% of infants worldwide are allergic to milk, though many outgrow the allergy by their school-age years. Symptoms can range from mild, such as hives, to severe, like anaphylaxis. This can also cause lactose intolerance though it's not an allergy. 3) Salt - Not an allergen 4) Syrup - Potential Allergen: It depends on what the syrup is made from. If it's made from corn (high fructose corn syrup), it is generally not an allergen. However, maple syrup might cause allergic reaction to those people who are allergic to maples. 5) Peanuts - Allergen: Approximately 1% of the population worldwide are allergic to peanuts. Peanut allergy symptoms can range from a minor irritation to a life-threatening reaction (anaphylaxis). For some people with peanut allergy, even tiny amounts of peanuts can cause a serious reaction.
Gpt-4-1106-preview (latest model)	1) Sugar: Not an allergen - Sugar itself is not considered a common allergen. However, some people may have sensitivities or dietary conditions that require them to avoid sugar. 2) Butter: Allergen - Allergic information: Butter is a dairy product and contains milk proteins, which can cause an allergic reaction in some individuals. Milk is one of the eight major allergens recognized by the U.S. Food and Drug Administration. Approximately 2-3% of children under the age of three are allergic to milk, but many outgrow it by the age of five. Symptoms of a milk allergy can range from mild (such as hives, itching or eczema) to severe (such as difficulty breathing or anaphylaxis). 3) Salt: Not an allergen - Salt is not known to trigger an allergic reaction. It is an essential mineral for the human body, though some individuals may have to limit their salt intake due to certain health conditions, like hypertension. 4) Syrup: Not a specific allergen - The term "syrup" can refer to many different types, such as corn syrup, maple syrup, and so on. Some syrups, like those that contain flavorings or additives, may cause reactions in sensitive individuals, but syrup itself is not a common allergen. However, if syrup contains ingredients derived from common allergens (e.g., almond syrup), then it may be allergenic. 5) Peanuts: Allergen - Allergic information: Approximately 1-2% of the U.S. population is allergic to peanuts. Peanut allergy is one of the most common food allergies in children, and it tends to be lifelong. Symptoms can range from mild (redness, hives, and digestive issues) to severe and potentially life-threatening (anaphylaxis). Due to the risk of severe allergic reactions, strict avoidance of peanuts and peanut-containing products is recommended for individuals with a peanut allergy.

Table 9: LLM Model output comparison when given ingredients of Peanut Brittle - part 2

Comparing different ChatGPT LLM models revealed consistency in identifying sugar and salt as non-allergens. The earlier gpt-3-5-turbo model differed on its view on butter, not labeling it as an allergen unlike later gpt models such as gpt-4 series, which acknowledged it as dairy allergen. All models recognized peanut as allergies consistently, emphasizing its prevalence and severity. Comparing LLM models, it was evident that the later models, particularly gpt-4-1106-preview went further by offering detailed statistics and symptoms associated with each allergen. Thus, after the comparison experiment, our team decided to use the gpt-4-1106, which was the latest model.

E. Example Outputs

	Our Model	Inverse Cooking	LLM Allergen Alert using our model	Comments
	sugar, egg, butter, milk, salt, cream, extract	sugar, egg, milk, cream, butter, salt	1) Sugar: Not an allergen Sugar is not a commonly known allergen and allergies to sugar itself are extremely rare. Most people can ingest various types of sugars (like glucose, fructose, and sucrose) without any problems. People may have intolerances or reactions to sugar additives, but these are not the same as allergies. 2) Egg: Allergen information Approximately 1-2% of children are allergic to eggs, making it one of the most common food allergies in children. However, many children outgrow egg allergy by the age of 16. Symptoms can range from mild reactions like hives to severe anaphylactic responses. It is important to check product labels for hidden sources of egg proteins in various food products. 3) Butter: Allergen information (due to milk content) Butter is made from milk and can cause allergic reactions in individuals with a milk allergy. It is estimated that about 0.2-0.5% of the general population is allergic to cow's milk, which is a more common allergy in infancy and early childhood, with many outgrowing it by their school-age years. Reactions can include gastrointestinal upset, rash, hives, vomiting, and in severe cases, anaphylaxis. 4) Milk: Allergen information Milk is one of the eight major food allergens recognized by the FDA. Approximately 2-3% of children under the age of three are allergic to milk, but many children outgrow this allergy by the time they are school-aged. Symptoms of a milk allergy can include hives, vomiting, respiratory issues, and in severe cases, anaphylaxis. Milk allergy should not be confused with lactose intolerance, which is an inability to digest the sugar lactose, found in milk. 5) Salt: Not an allergen Salt, in its pure form (sodium chloride), is not considered an allergen. There are no known allergic reactions to common table salt, and it is a mineral that is essential to human physiology. 6) Cream: Allergen information (due to milk content) Cream, being a dairy product, contains milk proteins, which can cause allergic reactions in people with milk allergies. It carries the same allergen potential as milk, affecting approximately 2-3% of infants and young children, though as with milk, many children outgrow it. The symptoms of a cream allergy are similar to those described for butter and milk allergies, such as skin reactions, gastrointestinal symptoms, or respiratory problems, and can include anaphylactic shock in severe cases.	Our model predicted extract (vanilla extract) being used in Crème brûlée which was missing in the InverseCooking model. Both models predicted the allergic ingredients egg, butter, milk and cream in this example.

Table 10: Model output for Crème brûlée

	Our Model	InverseCooking	LLM Allergen Alert using our model	Comments
	pepper, oil, parsley, lobster, clove, butter, salt	oil, pepper, salt, parsley, butter, clove	1) Pepper: Not an allergen Pepper, including black, white, and bell peppers, is not commonly recognized as a major allergen and does not affect a significant portion of the population with allergies. While it's not impossible for individuals to have an allergy to pepper, these cases are rare and not typically included in standard allergy information. 2) Oil: Not an allergen (with exceptions) The term "oil" encompasses a vast range of products. Most cooking oils by themselves, like canola oil, sunflower oil, or olive oil, are not major allergens. However, oils derived from allergenic sources, such as peanut, soybean, and sesame, can cause allergic reactions in people who are allergic to the source food. These specific oils would carry the same allergenic potential as the source. 3) Parsley: Not an allergen Parsley is not recognized as a common allergen and it does not affect a significant portion of the population. Allergies to spices, including parsley, are possible but occur infrequently. 4) Lobster: Allergic information: Lobster is a type of shellfish, which is a known allergen. Approximately 2% of the population is allergic to shellfish, but it is important to note that shellfish allergies are generally more common in adults than children. Shellfish allergy symptoms can range from mild reactions, such as hives or a stuffy nose, to severe and even life-threatening reactions like anaphylaxis. It is advised for individuals with shellfish allergies to avoid all types of shellfish. 5) Clove: Not an allergen Clove, like most spices, is not considered a common allergen. While it's possible for some people to have a sensitivity or allergy to cloves, this is relatively rare and lacks the prevalence to be classified alongside major allergens. 6) Butter: Allergic information: Butter is a dairy product and contains milk proteins, which can be allergenic. Around 0.2-0.5% of adults and 2.5% of children are allergic to cow's milk. The majority of children outgrow their milk allergy by age five. Milk allergy symptoms can include hives, vomiting, digestive problems, and in severe cases, anaphylaxis. Individuals with a milk allergy should avoid butter and other dairy products to prevent allergic reactions. 7) Salt: Not an allergen Salt is not recognized as an allergen. It is a mineral composed primarily of sodium chloride and, while excessive consumption can lead to health issues, it does not provoke an immune response that causes allergic reactions.	Inverse Cooking model missed the key ingredient lobster which is a common allergen.

Table 11: Model output for Grilled Lobster

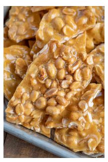
	Our Model	InverseCooking	LLM Allergen Alert using our model	Comments
	sugar, butter, salt, syrup, peanuts	sugar, butter, syrup, peanuts, salt, baking soda	1) Sugar: Not an allergen - Sugar itself is not considered a common allergen. However, some people may have sensitivities or dietary conditions that require them to avoid sugar. 2) Butter: Allergen - Allergic information: Butter is a dairy product and contains milk proteins, which can cause an allergic reaction in some individuals. Milk is one of the eight major allergens recognized by the U.S. Food and Drug Administration. Approximately 2-3% of children under the age of three are allergic to milk, but many outgrow it by the age of five. Symptoms of a milk allergy can range from mild (such as hives, itching or eczema) to severe (such as difficulty breathing or anaphylaxis). 3) Salt: Not an allergen - Salt is not known to trigger an allergic reaction. It is an essential mineral for the human body, though some individuals may have to limit their salt intake due to certain health conditions, like hypertension. 4) Syrup: Not a specific allergen - The term "syrup" can refer to many different types, such as corn syrup, maple syrup, and so on. Some syrups, like those that contain flavorings or additives, may cause reactions in sensitive individuals, but syrup itself is not a common allergen. However, if syrup contains ingredients derived from common allergens (e.g., almond syrup), then it may be allergenic. 5) Peanuts: Allergen - Allergic information: Approximately 1-2% of the U.S. population is allergic to peanuts. Peanut allergy is one of the most common food allergies in children, and it tends to be lifelong. Symptoms can range from mild (redness, hives, and digestive issues) to severe and potentially life-threatening (anaphylaxis). Due to the risk of severe allergic reactions, strict avoidance of peanuts and peanut-containing products is recommended for individuals with a peanut allergy.	Inverse cooking generated baking soda in the ingredient list which was missed in our prediction. Both models predicted the allergic ingredients of peanuts and butter in this scenario.

Table 12: Model output for Peanut Brittle