NutriParse AI

Owen Pannucci, Abhishek Sharma, Krish Arulalan
Department of Computer Science, Rice University
op21@rice.edu, as537@rice.edu, ka72@rice.edu

Abstract

Understanding and interpreting nutrition labels can be challenging for consumers. In this project, we develop a system that combines image Vision-Language Models (VLMs) and Large Language Models (LLMs) to automatically extract text from nutrition labels and answer users queries using the variety of information provided. In addition to real images of nutrition labels, we also created a dataset of synthetically generated images with labels to evaluate the OCR (Optical Character Recognition) model. Our pipeline makes use of the "Phi-3-Vision-128K-Instruct Model" [11] created by Microsoft to do OCR and is then followed by the "Llama-3.2-1B LLM" [10] and "Llama-2-7B-chat-hf" [9] created by Meta to generate natural language explanations and answers to user queries about the label contents.

1. Introduction

Recent advances in Vision-Language Models for OCR tasks have significantly expanded their practical applications, driven by improvements in accuracy and reliability. At the same time, the rapid development of LLMs has evolved the field of information retrieval, offering intelligent, dependable, and human-like capabilities - so much so that traditional search engines are becoming obsolete for many information-based questions. The convergence of these two technologies, OCR-focused VLMs and LLMs, opens the door to new possibilities of bridging visual text data with Natural Language Processing (NLP). In this work, we propose a practical approach to integrate these models into a unified pipeline, demonstrating their utility on a focused use case: the extraction and evaluation of nutritional information from food labels. Although our implementation focuses on this specific application, the framework can be easily generalized to other tasks involving the retrieval and analysis of textual information from images.

2. Project Pipeline

Our model processes a user-provided nutrition label image and prompt to generate a response to the user's query. This is done by first passing the image through an OCR model to extract the text. Then the extracted text is combined with the prompt and sent to the LLM, which produces the final response. A visual representation of the pipeline can be seen in Figure 1.

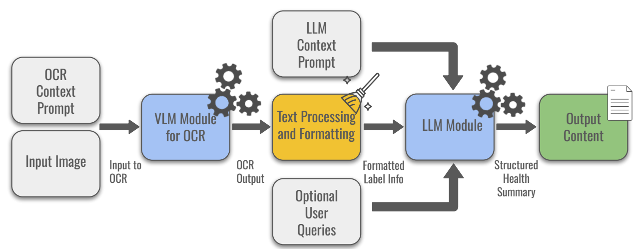


Figure 1: Pipeline to Extract and Analyze a Nutrition Label.

3. Related Work

Early OCR methods like Tesseract [3, 18] provided a foundation for digitizing printed text but often struggled with images captured in real-world conditions, such as those taken with smartphones. These approaches typically relied on bounding boxes to extract structured information from food packaging. With the rise of CNNs and their strong performance in object recognition, OCR systems began incorporating CNN-based architectures—as seen in [16, 5], where nutrition-related terms and values were extracted with improved accuracy. More recently, broader vision-language models like GPT-4o [14] and Claude Sonnet 3.5 [2] have been used for OCR tasks, outperforming traditional methods such as EasyOCR [6, 15], PyTesseract [4, 17], and TrOCR [7, 8].

[19] explores how LLMs such as GPT-3.5 [12] and GPT-4 [13] can assist in estimating nutritional content from food images and natural language descriptions by comparing their outputs to ground truth data from the USDA and Open Food Facts. While the models produced plausible estimates,

they often lacked precision and consistency. Building on this, NutriBench [1] introduces a benchmark dataset of over 11,000 human-verified meal descriptions paired with nutrient values to evaluate LLM performance. The study compares models like GPT-4 [13] and Claude, showing that although they perform reasonably well, they remain prone to hallucinations and inconsistencies, especially on unfamiliar inputs.

Based on the above findings, we arrive at two primary hypotheses:

1. Modern OCR methods that incorporate vision models are highly accurate and can generally be trusted.

2. LLMs, on the other hand, can be unreliable and may require extensive human evaluation and large, domain-specific datasets to ensure consistent and accurate performance.

4. Methodology

4.1. Vision Model for OCR

In this work, we used the pretrained "Phi-3-Vision-128K-Instruct" [11] Vision-Language Model developed by Microsoft as our OCR system to extract text from nutrition labels. Both an input image and a context prompt are passed to the model. The context prompt provides additional guidance to the VLM, explaining how to interpret the image and format the extracted output. After processing by the OCR model, the output undergoes a post-processing step to correct formatting issues. Specifically, the post-processing ensures that percentages align correctly with their corresponding rows in the nutrition tables and that ingredients spanning multiple lines are merged into single-line entries. An example of the OCR output after post-processing can be found in Table 3 in Appendix 5.5.

4.2. Parsing Output and Prompt Construction

Once OCR extraction and post-processing is complete, the structured data is passed to a prompt construction function, which formulates a query for the LLM. The function incorporates the nutritional information alongside a set of predefined queries, which include "*Provide a brief summary of the nutritional profile*", "*Any health risks or red flags?*", and "*Dietary suitability*". Additionally, the function allows for user-specific queries, enabling personalized evaluations - for example, "*Can I eat this if I am allergic to peanuts?*". This mechanism of prompt construction forms the basis for evaluating the LLM's capability to correctly analyze and summarize nutritional information in a meaningful and user-specific manner. An example of this can be seen in Listing 1.

4.3. Language Model

The formatted prompt is passed to the LLM model, "Llama-3.2-1B" [10] or "Llama-2-7B-chat-hf" [9], in the form of a structured message list, following a role prompting dialogue pattern. The system message establishes the context for the model by fist explaining that it is a food, nutrition, and allergy expert whose task is to analyze text extracted from food labels and answer the user's questions. Other context is also included in this prompt, such as how to process the given information and structure the output. Additionally, the user message includes the parsed nutritional information from the OCR module along with the user queries. The LLM processes this input and generates a natural language response that addresses the questions and summarizes the nutritional profile accordingly. See Listing 1 for an example case.

5. Evaluation and Discussion

To evaluate the OCR model, a synthetic dataset with annotations was created and used for assessment. The evaluation of the entire pipeline involved manual inspection of the OCR and LLM outputs for 30 images: 20 synthetically generated images and 10 real-world nutrition labels sourced from an everyday kitchen.

5.1. Synthetic Dataset Creation

To get started with the evaluation of the OCR model we first used GROK 3 [20] to help create a script that generates a synthetic data set of nutrition labels along with the annotations. The synthetic dataset of nutrition labels was created by randomly generating nutritional information, including headers, nutrients, micronutrients, ingredients, and product details, using predefined ingredient lists and randomized values for nutritional content. The data was rendered into images with realistic layouts, featuring one or two-column formats, tables, and varied fonts and colors to mimic real-world nutrition labels. The annotations were formatted to have all relevant parts on the nutrition label on the same line to match with the behavior of the OCR model we used. This was done to make sure that the dataset is suitable to evaluate the text extraction from the nutrition labels. An example of some of the images generated can be seen in Figures 2, and 3. An example of an annotation can be seen in Table 3.

5.2. OCR Validation Using Synthetic Dataset

The OCR model was evaluated using a synthetic dataset of nutrition label images. The OCR predictions were compared to the ground truth annotations to determine correctness. Predictions and annotations were treated as sets of lines to allow order-independent matching. A prediction was considered correct if every line had an exact match in the annotation. Matched lines were removed iteratively until all were paired or a mismatch occurred. If any unmatched lines remained in either the predictions or annotations, the image was marked as incorrect. Accuracy was measured as the percentage of images with fully matching lines. We used Levenshtein distance to identify mismatches and provide detailed reports on missing and extra lines. An example of a detailed report generated for an incorrect OCR result can be seen in Listing 2 for the image in Figure 4.

After validating on a synthetic dataset of 500 images, the OCR model produced correct results for 473 of the 500 nutrition labels, which corresponds to a validation accuracy of **94.60%**. The complete validation results can be found in the OCR validation results document linked in Appendix C 5.5. Detailed reports for incorrect OCR results reveal that errors arise mainly from missing lines, leading to the omission of critical information such as product codes, names, or nutritional details such as calories. Additionally, content that should appear on separate lines in the true label is often concatenated into a single line in the prediction. Other issues include formatting errors, such as incorrect nutrient percentages, misaligned lines, and stray percentage values that distort the structure of the nutritional data. There are also typographical errors in ingredient lists, like misspelled words such as "Cashews" being misspelled as "Cashens", which cause minor but noticeable inaccuracies.

5.3. OCR Manual Inspection Analysis

To get a more in-depth evaluation of the OCR model, the outputs were manually inspected for 30 images, where 20 images were synthetically generated, and 10 images are images from the real world. The results of the manual inspection can be found in the full pipeline manual inspection results linked in Appendix C 5.5. After analyzing the manual inspection results, it was found that the OCR model was able to correctly perform OCR for 28 of the 30 images. In image 1 the model did not include "calories" and in image 22 the daily value percentage of calcium was incorrectly labeled as 2% when it is truly 0%, incorrectly stated the daily value percentage of iron as 8% when it is truly 2%, did not include the daily value percentage for potassium, and in the ingredients, it labeled "sodium sulfate" instead of "sodium acid sulfate." Everything else in the OCR for these images was present and correct. Another observation was that in the real images, some related content appeared on new lines. While this isn't a major issue since the information is still contextually linked by being in the same vicinity, it is still worth noting. Based on these results, it can be seen that the OCR model performs relatively well for both synthetic and real nutrition labels, however, it is important to note that it can still make mistakes such as reporting incorrect values or ingredients, and having missing items.

5.4. LLM Evaluation Using Retrieved Information

To empirically evaluate the performance of our language models, we tested them on the basis of precision and recall. The primary evaluation task involved determining whether a food product was gluten-free and vegan, based on the presence or absence of gluten-containing or animal-derived ingredients. Table 1 presents the results for the Llama-2-7B-chat-hf model, and Table 2 presents the results for the Llama-3.2-1B model. The evaluation was done by comparing the "Dietary Suitability" report from the LLMs to the ingredients found in the OCR output in Appendix C 5.5.

For Llama-2-7B-chat-hf, both the precision and recall were **0.6176**. For Llama-3.2-1B, the precision was **0.7895** and the recall was **0.4167** . While the empirical results suggest only modest differences between the models, they do not fully capture the overall model behavior. In addition to the gluten-free and vegan classification task, we manually evaluated the quality of responses to unique user queries, nutrition profile summaries, and health risk assessments. In these areas, both models demonstrated generally stronger performance.

		Predicted	
		Positive	**Negative**
Actual	**Positive**	21	13
	Negative	13	13

Table 1: Confusion Matrix for Llama-2-7B-chat-hf

		Predicted	
		Positive	**Negative**
Actual	**Positive**	15	21
	Negative	4	20

Table 2: Confusion Matrix for Llama-3.2-1B

5.5. LLM Manual Inspection Analysis

The results of the of the manual inspection can be found in the full pipeline manual inspection results linked in Appendix C 5.5.

Qualitatively, Llama-3.2-1B appeared to produce slightly better outputs overall. Its responses were more neatly formatted and provided more accurate and detailed nutritional summaries. Performance on user queries was comparable between models; both were generally accurate.

Several interesting behaviors were observed during evaluation. Both models occasionally misclassified ingredients such as "rice flour" and "oat flour," mistakenly associating them with gluten due to the presence of the term "flour" (e.g., examples 5 and 8). Additionally, instances of internal contradiction were observed. For example, in example 16, Llama-3.2 correctly identified "whole wheat flour" as containing gluten but still claimed the product was suitable for gluten-intolerant individuals. Similarly, in example 13, Llama-2 simultaneously stated that a product was not gluten-free but contained no gluten-containing ingredients.

Both models tend to summarize the nutritional profile of a food product by labeling each nutrient as either “high” or “low” without considering the size of a single serving. For instance, in example 24, Llama-2 labeled the food item as “high in calories” simply because it contains 140 calories, but ignores the fact that this is for a serving size of 240 mL, a relatively large volume that makes the calorie content more moderate or even low when considered in context.

Both models are more prone to making mistakes when answering general questions like “Are there any health risks?” compared to more specific queries. For instance, in example 29, both models fail to mention the presence of coconut and soy allergens when asked about general health risks. However, when specifically prompted about allergens, they correctly identify them.

Despite these issues, both models performed well in many specific cases. Most user queries were answered correctly. One notable success was in example 24, where Llama-2 correctly identified that a product containing the lactose enzyme would still be suitable for someone who is lactose intolerant. In example 18, Llama-2 also correctly identified that the presence of honey made a product non-vegan, a subtle detail often overlooked.

Overall, both LLMs performed well and accurately answered most queries, though they occasionally misclassified ingredients and allergens, made contradictory statements, or oversimplified nutrition details.

Conclusion

In this project, we explored the application of deep learning models for optical character recognition and natural language text generation for nutrition and health information. We combined the frameworks of vision language models and large language models into a unified pipeline that provides an overview of a food item’s nutritional profile, health risks, dietary suitability, and simultaneously answers unique user queries.

Through an experimental evaluation using a synthetically created dataset, we demonstrated that our OCR model was able to capture the majority of the information presented in the nutrition label images. Evaluation of the language model outputs revealed areas for improvement in contextual awareness and logical reasoning. These findings are consistent with observations from related work, where modern OCR models tend to be highly reliable, while LLMs often require domain-specific fine-tuning to achieve high levels of accuracy.

Manual inspection of outputs suggested a generally satisfactory LLM performance. Most user queries were answered correctly and supported by information drawn from the nutrition labels. Notably, while the two benchmarked LLM models produced similar content, the Llama-3.2 model delivered outputs with detailed formatting and structure, resulting in a more user-friendly experience.

Based on these results, the area with the most potential for improvement is in the LLM component. Due to computing limitations, we were restricted to models that could fit within approximately 12-15GB of memory, leading us to utilize simpler Llama models with 1B and 7B parameters. If this constraint were removed, it would be possible to implement more advanced models such as Llama-4-17B, which could substantially improve both contextual understanding and response quality.

Overall, this project successfully demonstrated the feasibility of combining OCR and LLMs to extract and interpret nutrition label information for end users.

Acknowledgments

We used GROK 3 [20] to help create the script for the synthetic dataset and to help correct errors that occurred in the code and ChatGPT-4o [14] to assist in editing and style.

References

[1] E. Akyürek, X. Liu, X. Li, et al. Nutribench: A benchmark for evaluating large language models on nutrition estimation. *arXiv preprint arXiv:2407.12843*, 2024.

[2] Anthropic. Claude sonnet 3.5. `https://www.anthropic.com/news/claude-3-5-sonnet`, 2025.

[3] S. Bajracharya and N. Shrestha. Implementation of tesseract ocr and bounding box for text extraction on food nutrition labels. *ResearchGate*, 2024.

[4] S. Hoffstaetter. Pytesseract. `https://pypi.org/project/pytesseract/`, 2024.

[5] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman. Reading text in the wild with convolutional neural networks. *International Journal of Computer Vision*, 116(1):1–20, 2016. Originally released as arXiv:1412.1842 (2014).

[6] JaidedAI. Easyocr. `https://github.com/JaidedAI/EasyOCR`, 2020.

[7] S. Kim, J. Baudru, W. Ryckbosch, et al. Early evidence of how llms outperform traditional systems on ocr/htr tasks for historical records. *arXiv preprint arXiv:2501.11623*, 2024.

[8] M. Li, T. Lv, J. Chen, L. Cui, Y. Lu, D. Florencio, C. Zhang, Z. Li, and F. Wei. Trocr: Transformer-based optical character recognition with pre-trained models. In *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7464–7474, 2021.

[9] MetaAI. Llama 2 7b chat-hf. `https://huggingface.co/meta-llama/Llama-2-7b/tree/main`, 2023.

[10] MetaAI. Llama 3.2 1b. `https://huggingface.co/meta-llama/Llama-3.2-1B`, 2024.

[11] Microsoft. Phi-3 vision 128k instruct. https://huggingface.co/microsoft/Phi-3-vision-128k-instruct, 2024.

[12] OpenAI. Gpt-3.5 turbo. https://platform.openai.com/docs/models/gpt-3.5-turbo, 2022.

[13] OpenAI. Gpt-4. https://openai.com/research/gpt-4, 2023.

[14] OpenAI. Chatgpt-4o. https://openai.com/index/hello-gpt-4o/, 2024.

[15] B. K. Pattanayak, A. K. Biswal, S. R. Laha, S. Pattnaik, B. B. Dash, and S. S. Patra. A novel technique for handwritten text recognition using easyocr. In *Proc. International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, pages 1115–1119. IEEE, 2023.

[16] R. Shrestha. Word recognition in nutrition labels using convolutional neural networks. Master's thesis, Utah State University, 2020.

[17] S. Sisodia, S. Mukherjee, H. Tyagi, N. Singh, and S. Bhardwaj. Ocr using python and its applications. *Journal of Advanced Zoology*, 44(S-3):1083–1092, 2023.

[18] R. Smith. An overview of the tesseract ocr engine. In *Proc. 9th International Conference on Document Analysis and Recognition (ICDAR)*, pages 629–633. IEEE, 2007.

[19] B. Truong. Investigating the potential of large language models in nutritional analysis. Master's thesis, Kennesaw State University, 2023.

[20] xAI. Grok 3. https://x.ai/grok, 2025.

Appendix A

Nutrition Facts

Serving Size 1 piece (34g)
Servings Per Container about 13

Amount Per Serving	%Daily Value*
Calories 95	
Total Fat 7g	10%
Saturated Fat 2g	10%
Trans Fat 1g	
Cholesterol 10mg	3%
Sodium 101mg	0%
Total Carbohydrate 9g	3%
Dietary Fiber 2g	8%
Sugars 4g	8%
Protein 2g	

Vitamin D	1%
Vitamin C	1%
Calcium	3%
Iron	4%
Potassium	2%

Ingredients: Rice Flour, Molasses, Palm Kernel Oil, Sugar, Salt, BHT (Preservative), Citric Acid, Cinnamon, Yellow 5, Blue 1, Sodium Benzoate (Preservative).

RB24963 CRUNCHY SNACKS 373

Figure 2: Synthetic Nutrition Label Example 1.

Nutrition Facts

Serving Size 3 pieces (36g)

Servings Per Container about 10

Amount Per Serving	%Daily Value*
Calories 81	
Total Fat 8g	12%
Saturated Fat 2g	10%
Trans Fat 1g	
Cholesterol 2mg	0%
Sodium 114mg	0%
Total Carbohydrate 9g	3%
Dietary Fiber 0g	0%
Sugars 4g	8%
Protein 2g	

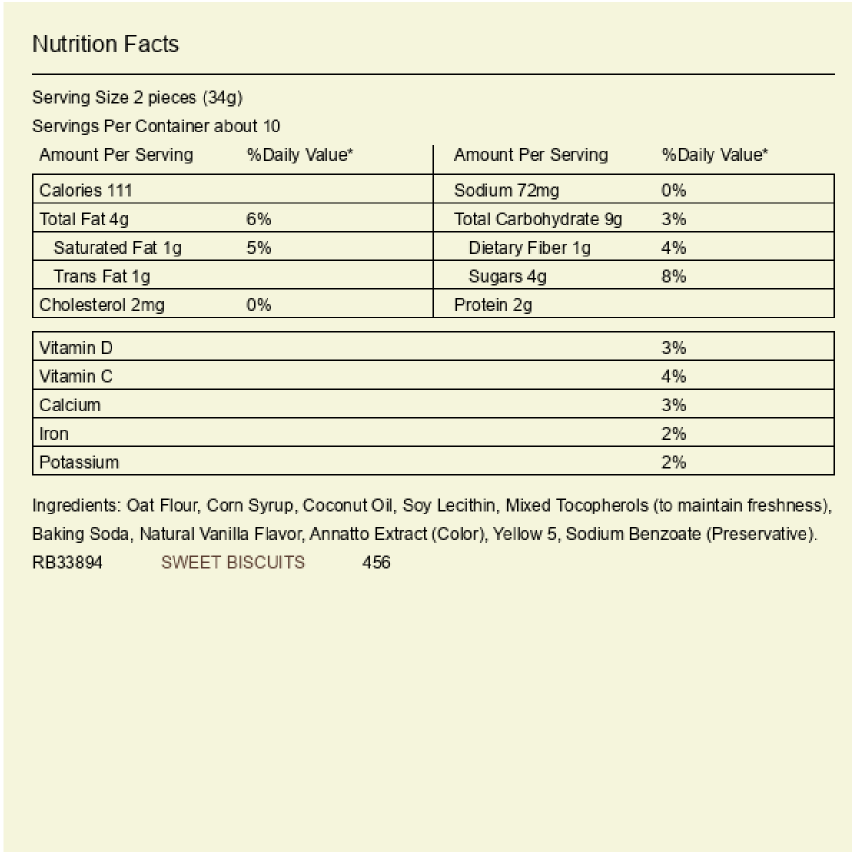
Vitamin D	2%
Vitamin A	4%
Vitamin C	1%
Calcium	1%
Iron	3%
Potassium	2%

Ingredients: Enriched Flour (Wheat Flour, Niacin, Reduced Iron, Thiamine Mononitrate, Riboflavin, Folic Acid), High Fructose Corn Syrup, Coconut Oil, Corn Syrup Solids, Mono- and Diglycerides, Mixed Tocopherols (to maintain freshness), BHT (Preservative), Cinnamon, Red 40, Blue 2, Sodium Benzoate (Preservative).

RB87580 VANILLA WAFERS 535

Figure 3: Synthetic Nutrition Label Example 2.

Image



Nutrition Facts

Serving Size 2 pieces (34g)
Servings Per Container about 10

Amount Per Serving	%Daily Value*	Amount Per Serving	%Daily Value*
Calories 111		Sodium 72mg	0%
Total Fat 4g	6%	Total Carbohydrate 9g	3%
Saturated Fat 1g	5%	Dietary Fiber 1g	4%
Trans Fat 1g		Sugars 4g	8%
Cholesterol 2mg	0%	Protein 2g	

Vitamin D	3%
Vitamin C	4%
Calcium	3%
Iron	2%
Potassium	2%

Ingredients: Oat Flour, Corn Syrup, Coconut Oil, Soy Lecithin, Mixed Tocopherols (to maintain freshness), Baking Soda, Natural Vanilla Flavor, Annatto Extract (Color), Yellow 5, Sodium Benzoate (Preservative).

RB33894 SWEET BISCUITS 456

Annotation

Nutrition Facts
Serving Size 2 pieces (34g)
Servings Per Container about 10
Amount Per Serving
%Daily Value*
Calories 111
Total Fat 4g 6%
Saturated Fat 1g 5%
Trans Fat 1g
Cholesterol 2mg 0%
Sodium 72mg 0%
Total Carbohydrate 9g 3%
Dietary Fiber 1g 4%
Sugars 4g 8%
Protein 2g
Vitamin D 3%
Vitamin C 4%
Calcium 3%
Iron 2%
Potassium 2%
Ingredients: Oat Flour, Corn Syrup, Coconut Oil, Soy Lecithin, Mixed Tocopherols (to maintain freshness), Baking Soda, Natural Vanilla Flavor, Annatto Extract (Color), Yellow 5, Sodium Benzoate (Preservative).
RB33894
SWEET BISCUITS
456

Table 3: Generated Annotation for a Nutrition Label.

Appendix B

Listing 1: Example Prompt Construction and Response

```
-> Nutriton Label OCR:

Nutrition Facts
Serving Size 3 pieces (39g)
Servings Per Container about 8
Amount Per Serving
%Daily Value*
Calories 146
Total Fat 9g 13%
Saturated Fat 1g 5%
Trans Fat 0g
Cholesterol 3mg 1%
Sodium 106mg 0%
Total Carbohydrate 9g 3%
Dietary Fiber 3g 12%
Sugars 4g 8%
Protein 1g
Vitamin D 3%
Calcium 3%
Iron 3%
Potassium 2%
Ingredients: Rice Flour, Honey, Canola Oil, Almonds, Corn Syrup, Soy Lecithin, Baking Soda,
↪ Dried Egg Whites, Natural Flavor, Yellow 5 Lake, Red 3, Peanut Butter (Peanuts, Salt).
RB39246
HONEY BITES
826

-> Predefined Prompts:

Provide a brief summary of the nutritional profile.
Any health risks or red flags?
Dietary suitability.

-> User Prompts:

What's the sodium content per serving?
Are there any vitamins or minerals listed (like Vitamin D, calcium, iron)?
How much Vitamin C is there?

-> User Message After Prompt Contrsuction:

Here is the nutrition information:
Nutrition Facts
Serving Size 3 pieces (39g)
Servings Per Container about 8
Amount Per Serving
%Daily Value*
Calories 146
Total Fat 9g 13%
Saturated Fat 1g 5%
Trans Fat 0g
Cholesterol 3mg 1%
Sodium 106mg 0%
Total Carbohydrate 9g 3%
```

```
Dietary Fiber 3g 12%
Sugars 4g 8%
Protein 1g
Vitamin D 3%
Calcium 3%
Iron 3%
Potassium 2%
Ingredients: Rice Flour, Honey, Canola Oil, Almonds, Corn Syrup, Soy Lecithin, Baking Soda,
↪ Dried Egg Whites, Natural Flavor, Yellow 5 Lake, Red 3, Peanut Butter (Peanuts, Salt).
RB39246
HONEY BITES
826

Provide:
1. Provide a brief summary of the nutritional profile.
2. Any health risks or red flags?
3. Dietary suitability.
4. What's the sodium content per serving?
5. Are there any vitamins or minerals listed (like Vitamin D, calcium, iron)?
6. How much Vitamin C is there?

-> LLM Response:

1. **Nutritional Profile Summary:**
The nutritional profile of Honey Bites, a sweetened rice flour-based snack, is as follows:
- Calories: 146
- Total Fat: 9g (13% of the Daily Value (DV))
- Sodium: 106mg (0% of the DV)
- Total Carbohydrate: 9g (3% of the DV)
- Dietary Fiber: 3g (12% of the DV)
- Sugars: 4g (8% of the DV)
- Protein: 1g

2. **Health Risks or Red Flags:**
- Presence of peanuts: The product contains peanuts, which is a significant allergen.
↪ Individuals with peanut allergies should avoid consuming Honey Bites.

3. **Dietary Suitability:**
- The product is gluten-free: The ingredient list does not contain gluten-containing
↪ ingredients (rice flour, wheat, barley, rye). However, there is a risk of
↪ cross-contamination due to shared equipment or facilities.

4. **Sodium Content per Serving:**
- Sodium per serving: 106mg

5. **Vitamins and Minerals:**
- Vitamin D: Present in the product (3% of the DV)
- Calcium: Present in the product (3% of the DV)
- Iron: Present in the product (3% of the DV)

6. **Vitamin C:**
- Present in the product (0% of the DV)
```

Nutrition Facts

Serving Size 1 piece (21g)

Servings Per Container about 10

Amount Per Serving	%Daily Value*	Amount Per Serving	%Daily Value*
Calories 134		Sodium 116mg	0%
Total Fat 5g	7%	Total Carbohydrate 13g	4%
Saturated Fat 1g	5%	Dietary Fiber 1g	4%
Trans Fat 1g		Sugars 4g	8%
Cholesterol 5mg	1%	Protein 0g	

Vitamin D	0%
Vitamin A	5%
Vitamin C	4%
Calcium	1%
Iron	4%
Potassium	2%

Ingredients: Whole Wheat Flour, High Fructose Corn Syrup, Palm Oil, Walnuts, Citric Acid, Sodium Benzoate (Preservative), Mixed Tocopherols (to maintain freshness), Natural Caramel Flavor, Blue 1, Red 3, Whey Powder (Milk).

RB62555 OAT CRACKERS 711

Figure 4: Nutrition Label that has an Incorrect OCR Result.

Listing 2: Example of Detailed Report For Incorrect OCR Output

```
Image: nutrition_label_093.png

True Label:
Nutrition Facts
Serving Size 1 piece (21g)
Servings Per Container about 10
Amount Per Serving
%Daily Value*
Amount Per Serving
%Daily Value*
Calories 134
Total Fat 5g 7%
Saturated Fat 1g 5%
Trans Fat 1g
Cholesterol 5mg 1%
Sodium 116mg 0%
Total Carbohydrate 13g 4%
Dietary Fiber 1g 4%
Sugars 4g 8%
```

```
Protein 0g
Vitamin D 0%
Vitamin A 5%
Vitamin C 4%
Calcium 1%
Iron 4%
Potassium 2%
Ingredients: Whole Wheat Flour, High Fructose Corn Syrup, Palm Oil, Walnuts, Citric Acid,
↪ Sodium Benzoate (Preservative), Mixed Tocopherols (to maintain freshness), Natural
↪ Caramel Flavor, Blue 1, Red 3, Whey Powder (Milk).
RB62555
OAT CRACKERS
711

Prediction:
Nutrition Facts
Serving Size 1 piece (21g)
Servings Per Container about 10
Amount Per Serving
%Daily Value*
Amount Per Serving
%Daily Value*
Calories 134
Sodium 116mg 0%
Total Fat 5g 7%
Total Carbohydrate 13g 4%
Saturated Fat 1g 5%
Dietary Fiber 1g 4%
Trans Fat 1g
Sugars 4g 8%
Cholesterol 5mg 1%
Protein 0g
Vitamin D 0%
Vitamin A 5%
Vitamin C 4%
Calcium 1%
Iron 4%
Potassium 2%
Ingredients: Whole Wheat Flour, High Fructose Corn Syrup, Palm Oil, Walnuts, Citric Acid,
↪ Sodium Benzoate (Preservative), Mixed Tocopherols (to maintain freshness), Natural
↪ Caramel Flavor, Blue 1, Red 3, Whey Powder (Milk).
RB6255
OAT CRACKERS
711

Levenshtein Distance: 104

Details:
Missing lines (in true but not in pred):
['rb62555']
Extra lines (in pred but not in true):
['rb6255']
```

Appendix C

Link to OCR Validation Results

Link to Full Pipeline Manual Inspection Results