

Investigating the Efficacy of the Gemini 2.5 Pro Model in the Estimation of Macronutrient Content

Sonali Guttikonda
University High School
Carmel, IN, USA
sonali.gutti@gmail.com

He Tu
New York University Shanghai
Shanghai, China
ht2604@nyu.edu

Junyao Chen
NYU Stern, New York University
New York, NY, USA
junyao.chen@nyu.edu

Zhengyuan Zhou
NYU Stern, New York University
New York, NY, USA
zz26@stern.nyu.edu

Abstract

Health guidance is a topic that has been gaining increasing traction in the past few years. Specifically, apps utilizing AI to provide health guidance and allow users to track macronutrient content have been gaining prevalence in recent research. However, current apps require users to manually input all information into the app. Therefore, this work seeks to evaluate the effectiveness of the Gemini 2.5 Pro model, a multimodal large language model (LLM), in analyzing the macronutrient content (i.e., proteins, fats, and carbohydrates) in food images. We tested 100 dishes randomly sampled from the Nutrition5k dataset. Each image was queried three times with three different prompt phrasings and the estimated values were compared with the ground truth values in the dataset. For the baseline prompt, the mean absolute error (MAE) was 4.1 g for protein, 3.6 g for fat, and 6.1 g for carbohydrates, and the median absolute percentage errors were 29.9%, 29.2%, and 32.1% respectively. The mean absolute percentage error (MAPE) was 38.5% for protein but 211.2% for fats and 442.7% for carbohydrates, which were inflated due to a small number of dishes containing less than 0.5 g of fat or carbohydrate, where errors of a few grams produced errors in the thousands. Without dishes with less than 2 g of the nutrient, the MAPE was 38.5%, 33.6%, and 33.5%. On metrics not sensitive to near-zero values, error was of similar size across the three macronutrients, and protein was underestimated on average, with the largest errors concentrated in a subset of dishes. MAE differed by at most 0.6 g across prompt wordings, although fat MAPE was higher under both alternative prompts. Because no confidence intervals or significance tests were computed, these are observations from a single 100-dish sample rather than established differences. Additionally, the 100 dishes selected for the study came from a 400-dish subset whose selection cannot be reproduced.

Keywords

Multimodal Large Language Models, Gemini 2.5 Pro, Macronutrient Estimation, Food Image Analysis, Nutrition5k, Prompt Engineering

ACM Reference Format:

Sonali Guttikonda, Junyao Chen, He Tu, and Zhengyuan Zhou. 2026. Investigating the Efficacy of the Gemini 2.5 Pro Model in the Estimation of Macronutrient Content. In *Proceedings of International Journal of Secondary Computing and Applications Research (IJSCAR VOL. 3, ISSUE 4)*. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.67149/yhjs2024.5/c6ey9odw>

1 Introduction

Artificial Intelligence has gained increasing prevalence in the past few years in multiple aspects such as academic research and everyday tasks. Particularly, the use of AI models for estimating nutritional and mass information about a meal is a topic that has been increasingly explored in the past few years. This research is significant for combating widespread misinformation about nutrition online as well as providing accurate guidance for users seeking a healthier lifestyle. Recently, researchers have explored utilizing AI to estimate nutritional information based on a single image as opposed to current applications that require users to manually input food data. Previous research in this area has utilized the YOLO computer vision models. YOLO models belong to a family of object detection models that use single-stage detection. YOLO models simply localize and detect images without any language reasoning; therefore our research primarily focuses on utilizing a Gemini 2.5 Pro model, which is a multimodal model that is designed for language reasoning and image analysis. Additionally, we utilized the Nutrition5k dataset, which is a dataset containing images of about 5,000 dishes with different types of food along with their nutritional information, along with the Gemini model. We evaluate how accurately Gemini estimates per-dish macronutrient content from a single image, using the Nutrition5k dataset as the ground truth.

2 Related Work

2.1 Large Language Models (LLMs) for Macronutrient Estimation

Recent studies have tested LLMs for use in nutrition estimation with both text and image inputs. For text input, Carrillo-Larco utilized a fine-tuned LLM in order to estimate energy and macronutrient content from 24-hour dietary recalls with chain-of-thought prompting [1]. Additionally, Lawabni et al. found a high similarity in reference values between four LLMs (ChatGPT 5.2, Claude Opus 4.5, Gemini 3 Pro Preview, and Llama 4 Maverick) for carbohydrates,

This paper is published under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International (CC-BY-NC-ND 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

IJSCAR VOL. 3, ISSUE 4,
© 2026 The Author(s). Published by IJSCAR.
<https://doi.org/10.67149/yhjs2024.5/c6ey9odw>

protein, and fat content [7]. Utilizing photographs for nutrient estimation is more difficult since models must recognize the foods and interpret portion sizes. Fridolfsson et al. compared three different LLMs on food images: GPT-4o, Claude 3.5 Sonnet, and Gemini 1.5 Pro. GPT-4o and Claude had MAPEs of 36% and 37.3%, while Gemini had a significantly higher MAPE [3]. Certain studies specifically used the Nutrition5k dataset. For example, Khlaisamniang et al. first described the dish and then estimated nutrient content, with MAEs of 7.38 g for protein, 5.62 g for fat, and 7.76 g for carbohydrates using the Gemini 1.5 Pro model [6]. Additionally, Yan et al. utilized their DietAI24 framework and reduced MAE by 63% [11]. Our work evaluates the more updated Gemini 2.5 Pro model with three different single-step prompts on the Nutrition5k dataset to test how prompt variation and variability affect macronutrient estimation accuracy.

2.2 YOLO Models for Macronutrient Estimation

Another type of work dedicated to macronutrient estimation utilizes object detection models, specifically in the YOLO family. These models localize foods within a single pass and then compare them to nutrition values found in a database. Kalaimani et al. used the YOLOv8 and YOLOv10 models and reported accuracies around 95–99% for proteins, fats, and carbohydrates [5]. Matan et al. used YOLOv8 detection alongside a nutrition database, resulting in portion-aware estimates with an MAE of 4.5 g [8]. Dhanush et al. used the YOLOv11 model with the Segment Anything Model in order to isolate macronutrients and micronutrients from dishes [2]. Selvaraju et al. used YOLOv5, YOLOv8, and Mask R-CNN models in conjunction, resulting in a precision range of 0.95–0.97 [9]. All of these studies utilized models specifically designed for detection as compared with a multimodal LLM, which can identify foods as well as reason about their portions without specific training. Additionally, they utilize different comparison metrics, so they cannot be compared directly to our study's results.

3 Methodology

3.1 Dataset and Sampling

All of the data was primarily collected from one dataset, the Nutrition5k dataset, which has both photographs of diverse dishes and information on their macronutrient content. The images and ground-truth metadata were selected by Junyao Chen as a subset of the Nutrition5k dataset. This subset contained 400 randomly sampled overhead dish images from cafeteria 1 with matching nutrition labels. The seed used to select the 400-dish subset was not recorded. Ground truth values for each dish (total mass, protein, fat, and carbohydrates, in grams) were taken from the Nutrition5k dish metadata for cafeteria 1 (dish_metadata_cafe1) [4, 10]. Only the subset was used due to the API cost and time constraints of the study. Of these 400 dish images we used a random sample of 100 images with a fixed random seed (42) in order for our sample to be unaffected by dish ordering in the dataset.

3.2 Model Pipeline

This research primarily utilized a Gemini 2.5 Pro model with images from the Nutrition5k dataset as the input. The images were first resized to 1024 px by 1024 px and input to the Gemini 2.5 Pro model

through the Google Gen AI Python SDK, with a prompt requesting a JSON object with total mass, protein, fat, and carbohydrates in grams alongside a list of visible ingredients. The temperature was set to 1.0, which is the API default. Since the LLM outputs varied between calls, each image was queried three times per prompt and these estimates were averaged. Any responses that were not valid were re-requested with up to three attempts. However, all 100 images returned valid predictions for every prompt.

3.3 Prompt Variants

In order to determine whether prompt wording affected the accuracy of the model, we compared three different prompts that requested the same output: baseline, stepwise, and diet app. The baseline asked directly for the estimates. The stepwise prompt asked the model to identify each ingredient and portion size prior to estimating nutrient content. Finally, the diet-app prompt framed the task as if the model was logging a meal in a diet-tracking app. The full prompts are shown in Figure 1.

3.4 Error Calculation

Across all dishes, we calculated the mean absolute error (MAE), the mean signed error, the mean absolute percentage error (MAPE), the median absolute percentage error (MdAPE), and the symmetric mean absolute percentage error (sMAPE) using the following formulas:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |E_i - T_i| \quad (1)$$

$$\text{Mean signed error} = \frac{1}{N} \sum_{i=1}^N (E_i - T_i) \quad (2)$$

$$\text{MAPE} = \frac{100}{N} \sum_{i=1}^N \frac{|E_i - T_i|}{T_i} \quad (3)$$

$$\text{MdAPE} = \text{median}_i \left(100 \cdot \frac{|E_i - T_i|}{T_i} \right) \quad (4)$$

$$\text{sMAPE} = \frac{100}{N} \sum_{i=1}^N \frac{2 |E_i - T_i|}{|E_i| + |T_i|} \quad (5)$$

Here i is the dish index, E_i is the estimate, T_i is the true value in grams, and N is the total number of dishes. Negative signed error indicates underestimation. Dishes with $T_i = 0$ are excluded from MAPE and MdAPE; the number excluded was 0 for total mass, 1 for protein, 1 for fat, and 2 for carbohydrates. Since MAPE becomes very large when T_i is close to zero, we also report MAPE computed over dishes with $T_i \geq 2$ g. To measure run-to-run variability, we computed the standard deviation of the three repeated estimates for each dish and report the mean of these values across dishes.

4 Results

We evaluated Gemini 2.5 Pro on 100 randomly sampled dishes (fixed seed 42) from the 400 available Nutrition5k images. Each image was queried three times per prompt at temperature 1.0, and the three estimates were averaged. All 100 images returned valid predictions for every prompt.

```

JSON_FORMAT = (
    '{"total_mass_g": 0, "protein_g": 0, "fat_g": 0, "carbohydrates_g": 0, '
    '"ingredients": [{"name1": "name2"}]}\n'
    "total_mass_g is total food mass in grams.\n"
    "protein_g is total protein content in grams.\n"
    "fat_g is total fat content in grams.\n"
    "carbohydrates_g is total carbohydrate content in grams.\n"
    "ingredients is a list of visible ingredient names."
)
)
PROMPTS = {
    "baseline": "Look at this food image. Reply with JSON only:\n" + JSON_FORMAT,
    "stepwise": (
        "Look at this food image. First identify each visible ingredient and estimate "
        "its portion size, then add them up to get the totals. Reply with JSON only:\n"
        + JSON_FORMAT
    ),
    "diet_app": (
        "Look at this food image and estimate its nutrition as accurately as possible, "
        "as if logging it in a diet-tracking app. Reply with JSON only:\n" + JSON_FORMAT
    ),
}

```

Figure 1: Three different prompts input into Gemini 2.5 Pro for macronutrient estimation.

Table 1: Estimation error metrics of Gemini 2.5 Pro on 100 randomly sampled Nutrition5k dishes (with baseline prompt and three queries per image, averaged). MAE = mean absolute error. Bias = mean signed error (estimate – true); negative values indicate underestimation. MdAPE = median absolute percentage error. sMAPE = symmetric mean absolute percentage error. MAPE = mean absolute percentage error, over dishes with a true value above 0 g; “MAPE, dishes ≥ 2 g only” excludes dishes with a true value below 2 g. Repeat SD = mean standard deviation of the three repeated estimates per dish. Dishes = 0 g (excluded) = number of dishes with 0 g of each nutrient, excluded from the MAPE calculation.

Metric	Mass	Protein	Fat	Carbs
MAE (g)	24.2	4.1	3.6	6.1
Bias (g)	-0.3	-3.2	-0.9	+1.2
MdAPE (%)	20.4	29.9	29.2	32.1
sMAPE (%)	24.2	43.5	46.6	45.7
MAPE (%)	23.8	38.5	211.2	442.7
MAPE, dishes ≥ 2 g only (%)	23.0	38.5	33.6	33.5
Dishes with true value < 2 g	1	42	55	18
Repeat SD (g)	8.93	0.71	1.11	1.60
Dishes = 0 g (excluded)	0	1	1	2

Overall accuracy. Table 1 shows the error for the baseline prompt. The mean absolute error (MAE) was 4.1 g for protein, 3.6 g for fat, 6.1 g for carbohydrates, and 24.2 g for total dish mass. Median absolute percentage errors (MdAPE) were similar for the three macronutrients (29.2–32.1%), as were the symmetric MAPE values (43.5–46.6%).

Sensitivity of MAPE. Standard MAPE was substantially higher for fat (211.2%) and carbohydrates (442.7%) than for protein (38.5%). This difference does not come from the number of low-value dishes alone. Protein had 42 dishes below 2 g (one of which had a true

Table 2: Effect of prompt wording on estimation error (100 dishes, three queries per image averaged). MAE = mean absolute error. MdAPE = median absolute percentage error. sMAPE = symmetric mean absolute percentage error.

Nutrient	Prompt	MAE (g)	MdAPE (%)	sMAPE (%)
Total mass	Baseline	24.2	20.4	24.2
	Stepwise	23.9	18.5	24.6
	Diet app	23.3	17.1	24.8
Protein	Baseline	4.1	29.9	43.5
	Stepwise	4.0	26.5	43.3
	Diet app	4.0	23.0	43.3
Fat	Baseline	3.6	29.2	46.6
	Stepwise	3.8	37.0	50.1
	Diet app	3.7	34.4	56.0
Carbohydrates	Baseline	6.1	32.1	45.7
	Stepwise	5.9	28.5	43.7
	Diet app	5.5	27.5	42.2

value of 0 g and was excluded from MAPE, leaving 41) compared with 55 for fat and only 18 for carbohydrates. Instead, it depends on how close those dishes are to zero and how large the model’s errors on them are. The low-protein dishes had a median true value of 0.56 g, and the model’s estimates stayed close (median absolute error 0.18 g; largest 1.7 g). Their MAPE (38.7%) was therefore almost identical to that of other dishes, so removing them left protein MAPE at 38.5%. In contrast, most low-fat dishes contained less than 0.5 g of fat, and the ten largest fat errors (0.8–5.2 g) produced very large percentage errors. Excluding dishes below 2 g, MAPE fell to 33.6% for fat and 33.5% for carbohydrates, similar to protein (38.5%). We therefore rely on MAE, MdAPE, and sMAPE for comparisons across nutrients. On these metrics, we did not observe a difference suggesting that fat is estimated less accurately than protein or carbohydrates, although our analysis does not formally test this.

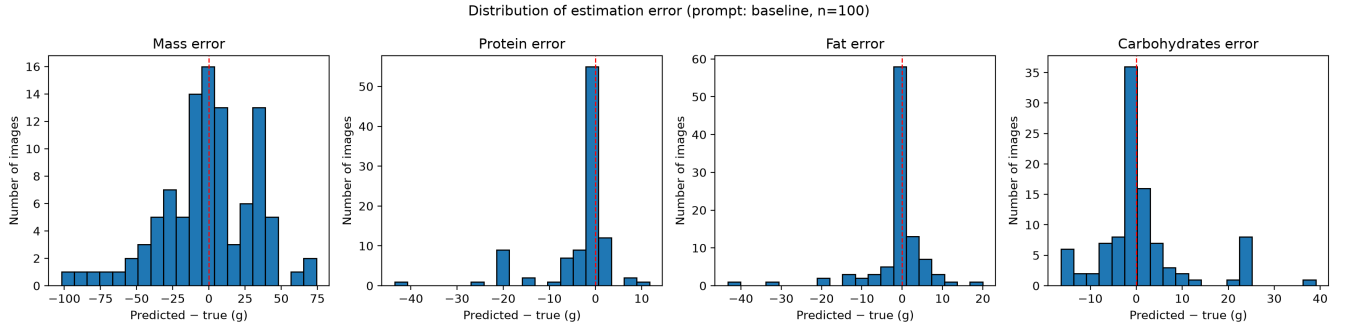


Figure 2: Distribution of signed estimation error (predicted – true, in grams) for total mass, protein, fat, and carbohydrates under the baseline prompt ($n = 100$ dishes).

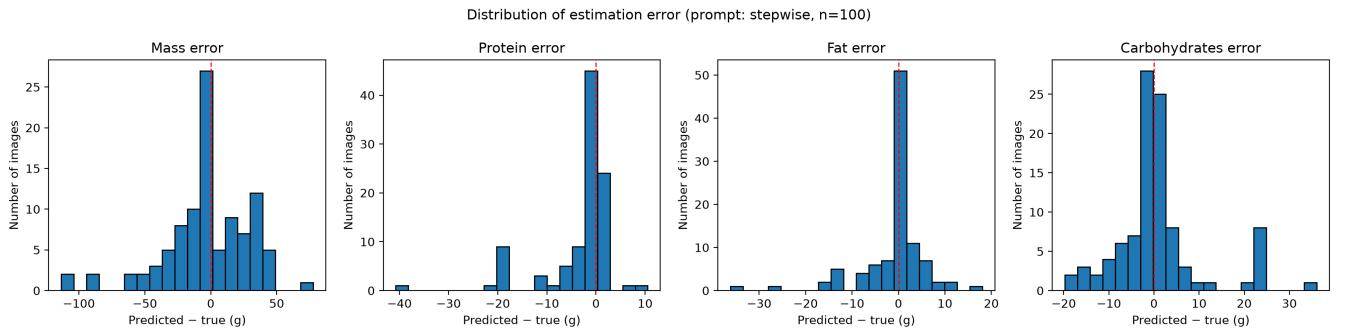


Figure 3: Distribution of signed estimation error (predicted – true, in grams) for total mass, protein, fat, and carbohydrates under the stepwise prompt ($n = 100$ dishes).

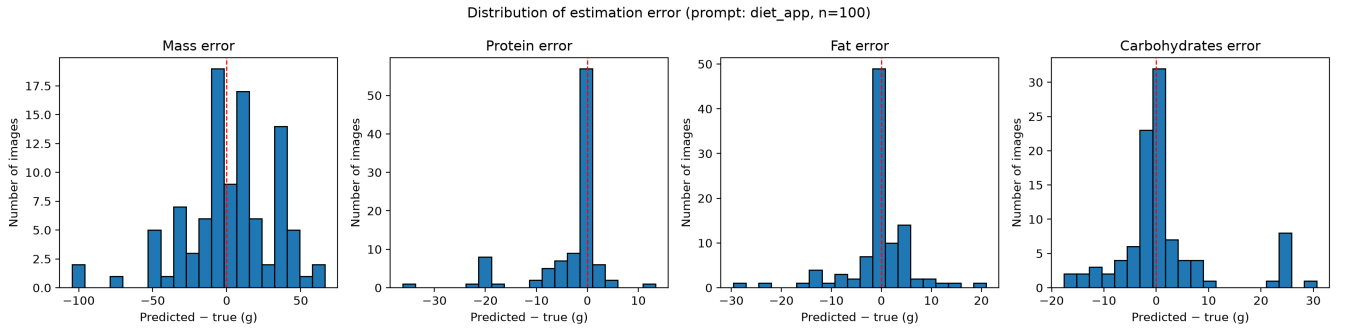


Figure 4: Distribution of signed estimation error (predicted – true, in grams) for total mass, protein, fat, and carbohydrates under the diet-app prompt ($n = 100$ dishes).

Bias. Protein was underestimated on average, with a bias of -3.2 g against an MAE of 4.1 g. This net underestimation is not a uniform shift: Figure 2 shows most protein errors near zero, with a left tail of dishes underestimated by roughly 10 – 40 g that accounts for much of the negative bias. The mean signed errors for fat (-0.9 g), carbohydrates ($+1.2$ g), and total mass (-0.3 g) were close to zero.

Run variability. The mean standard deviation across queries of the same image was 0.71 g for protein, 1.11 g for fat, 1.60 g for carbohydrates, and 8.93 g for mass, or roughly 17 – 37% of the respective MAEs. It was the smallest for protein (17%) and the largest for total

mass (37%). Run-to-run variability is therefore a relatively meaningful share of the total error, which supports averaging multiple queries per image.

Prompt wording. Table 2 compares results from the three prompt phrasings. At most, MAE differed by 0.6 g for each macronutrient across the prompts, and no prompt was the best across all nutrients. For the baseline prompt, Figure 2 shows that most errors for each macronutrient fall close to zero. The protein underestimation appears as a long left tail of dishes underestimated by 10 – 40 g. Additionally, the distribution for the diet-app prompt (Figure 4) closely

Table 3: MAE on Nutrition5k compared with Khlaisamniang et al. The two studies used different test subsets (Nutrition320 vs. our 100 dishes), so the values are not directly comparable.

Study	Model	Approach	MAE (g)		
			Protein	Fat	Carbs
Khlaisamniang [6]	Gemini 1.5 Pro	Two-step	7.38	5.62	7.76
This study	Gemini 2.5 Pro	One-step	4.1	3.6	6.1

resembled that of the baseline prompt. The diet-app framing gave the lowest MdAPE for mass (17.1%), protein (23.0%), and carbohydrates (27.5%), but a higher one for fat (34.4%). Protein bias was very similar for all three prompts (−3.1 to −3.3 g). However, fat MdAPE was higher under both alternative prompts (stepwise 37.0%, diet app 34.4%) than under the baseline (29.2%). Since we did not compute confidence intervals or test these differences, we report them as observed differences on this sample.

Comparison with prior work. Table 3 compares these results with that of Khlaisamniang et al. [6], who used Gemini 1.5 Pro with a two-step approach on Nutrition320, a different 320-dish subset of Nutrition5k. Our MAEs were lower for all three macronutrients. However, the two studies differ in model version, prompting approach, and test dishes, so the comparison is indicative only.

5 Discussion

Our hypothesis was that Gemini 2.5 Pro would have similar accuracy across all three macronutrients. The observed errors are consistent with this, although we did not test it statistically. For measures that were not sensitive to small true values, the three nutrients were similar: MAE was 3.6–6.1 g, MdAPE was 29.2–32.1%, and sMAPE was 43.5–46.6%.

The high MAPE values for fat and carbohydrates are driven largely by dishes with near-zero true values. MAPE becomes very large when a dish’s true value is less than a gram and the estimate is off by a few grams. Most low-fat dishes had less than 0.5 g of fat, and seven low-carbohydrate dishes were estimated to contain about 25 g instead of 0.43 g; these seven dishes account for a large share of the carbohydrate MAPE. Errors this large on dishes with low carbohydrate content could reflect model error, but they could also reflect errors in ground-truth labels. On inspection, the images show plain protein dishes with no visible carbohydrate source, so these appear to be genuine model overestimates. Low-protein dishes, by contrast, were mainly estimated within less than a gram, so the percentage errors stayed moderate. Therefore, MAPE should be interpreted carefully when used with nutrition datasets with many near-zero values, and should be reported alongside MdAPE and sMAPE values.

Protein was underestimated on average under all three prompt wordings (bias of −3.2 g for the baseline prompt and −3.1 to −3.3 g across prompts). The histograms show that this bias is concentrated in a tail of dishes underestimated by 10–40 g rather than a uniform shift across all dishes. Total mass had a mean signed error near zero (−0.3 g), which indicates that over- and underestimates of portion size roughly cancel on average; it does not by itself show

that individual portions are judged accurately, since mass MAE was 24.2 g (MdAPE 20.4%). Therefore, future work could involve examining the types of dishes that drive this bias (e.g., whether dishes contain ingredients like meat or eggs).

The three prompt wordings gave similar MAEs, differing by at most 0.6 g per nutrient. Fat MdAPE, however, was higher under both alternative prompts than under the baseline, so prompt wording may affect some nutrients more than others; a larger sample with significance testing would be needed to tell. Run-to-run variability was present but less than the overall error. The mean spread across repeats was 0.7–1.6 g for the macronutrients, or 17–31% of their MAEs (37% for total mass). This supports averaging three queries per image.

Our MAEs using Gemini 2.5 Pro were lower than those reported by Khlaisamniang et al. with Gemini 1.5 Pro on the same dataset. However, the studies differ in model version, prompt design, and test subset (their Nutrition320 versus our 100 dishes), and any of these could account for the difference.

One limitation of this study was the sample size: we only sampled 100 dishes from a single subset due to time constraints. In addition, the 100 dishes used came from a 400-dish subset whose selection cannot be reproduced. Additionally, only one model (Gemini 2.5 Pro) was tested. The 2 g threshold for fat and carbohydrates was a practical choice, not a standard one. We also did not test whether the differences between prompts were statistically significant.

6 Conclusion

In this 100-dish sample, Gemini 2.5 Pro’s estimation error was of similar size across the three macronutrients; protein was underestimated on average, with most of that error concentrated in a subset of dishes, and changing the prompt wording changed MAE by at most 0.6 g, though fat MdAPE rose under both alternative prompts. These are observations rather than statistically tested conclusions. In a further continuation of the study, we could analyze how effective the Gemini 2.5 Pro model is in predicting the content of more types of nutrients in food images, as well as identify the types of foods on which Gemini had poor estimation performance. To do this, we could choose a dataset that has information on more types of nutrients. Additionally, we could use pattern-recognition methods to identify the types of food that the Gemini 2.5 Pro model analyzed poorly.

Acknowledgments

We would like to thank Catherine Tissot, Matthew Leingang, Wynnian Yang, and the entire NYU GSTEM Program for supporting us on this project.

Data Availability

The Nutrition5k images and dish metadata used in this study are publicly available at <https://github.com/google-research-datasets/Nutrition5k> [4]. The dish IDs of the 400-dish subset and of the 100 dishes evaluated are unavailable to readers.

References

- [1] Carrillo-Larco, R.M.: LLMs for energy and macronutrients estimation using only text data from 24-hour dietary recalls: A parameter-efficient

- fine-tuning experiment using a 10-shot prompt. arXiv:2509.13268 (2025). <https://doi.org/10.48550/arXiv.2509.13268>
- [2] Dhanush, R., Abhilash, G.M., M, E., C, A.S., G, S.H.: NutriVision: YOLO-SAM-driven dietary app for personalized food recommendation and meal tracking. In: 2025 International Conference on Innovative Trends in Information Technology (ICITIIT). pp. 1–6 (2025). <https://doi.org/10.1109/ICITIIT64777.2025.11040294>
 - [3] Fridolfsson, J., Sjöberg, E., Thiwång, M., Pettersson, S.: Performance evaluation of 3 large language models for nutritional content estimation from food images. *Current Developments in Nutrition* **9**(10), 107556 (2025). <https://doi.org/10.1016/j.cdnut.2025.107556>
 - [4] Google Research: Nutrition5k dataset. GitHub repository, <https://github.com/google-research-datasets/Nutrition5k> (2021)
 - [5] Kalaimani, M., D, A.S., M, D.S., Santhoshkumar, S.P.: Smart food recognition and nutrition estimation using computer vision. In: 2026 9th International Conference on Inventive Computation Technologies (ICICT). pp. 1198–1204 (2026). <https://doi.org/10.1109/ICICT68280.2026.11510899>
 - [6] Khlaisamniang, P., Kerdthaisong, K., Vorathammathorn, S., Yongsatianchot, N., Phimsiri, H., Chinkamol, A., Thitseeaeng, T., Veerakanjana, K., Kachai, K., Itichaiwong, P., Saengja, T.: Decomposing food images for better nutrition analysis: A nutritionist-inspired two-step multimodal LLM approach. In: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 482–491 (2025). <https://doi.org/10.1109/CVPRW67362.2025.00053>
 - [7] Lawabni, R., Solano, O.R., Kampen, L.L., Kurasch, J., Alshurafa, N., Chang, J.S., Evans, R., Feng, D., Lin, A.W.: Evaluation of energy and nutrient estimates from large language models using text-based queries. *The Journal of Nutrition* **156**(9), 101712 (2026). <https://doi.org/10.1016/j.tjnut.2026.101712>
 - [8] Matan, P., Thulasiram, M., Vardhan, D.H., Reddy, V.R.V.K.: Precision dietary recommendation using YOLOv8 food detection and optimization based nutritional planning. In: 2026 Third International Conference on Innovations in Cybersecurity and Data Science (ICICDS). pp. 1334–1339 (2026). <https://doi.org/10.1109/ICICDS70526.2026.11604686>
 - [9] Selvaraju, T., Dakshinamurthi, V., Badurudeen, G., Prabhakaran, N., Gopi, P., Hussain, M.A.N.: Food detection and estimation of calories and other macronutrients and features to support healthy lifestyle. In: AIP Conference Proceedings. vol. 3279, p. 020129 (2025). <https://doi.org/10.1063/5.0263063>
 - [10] Thames, Q., Karpur, A., Norris, W., Xia, F., Panait, L., Weyand, T., Sim, J.: Nutrition5k: Towards automatic nutritional understanding of generic food. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8903–8911 (2021)
 - [11] Yan, R., Luo, H., Lu, J., Liu, D., Posluszny, H., Dhaliwal, M.P., MacLeod, J., Qin, Y., Yang, C., Hartman, T.J., Hu, X.: DietAI24 as a framework for comprehensive nutrition estimation using multimodal large language models. *Communications Medicine* **5**(1), 458 (2025). <https://doi.org/10.1038/s43856-025-01159-0>