UDC 004.93
DOI https://doi.org/10.24144/2616-7700.2025.47(2).289-302

K. S. Khabarlak

Dnipro University of Technology,
Associate Professor of the Department of System Analysis and Control,
PhD in Computer Science
Khabarlak.K.S@nmu.one
ORCID: https://orcid.org/0000-0003-4263-0871

MULTI-MODAL LLM-ASSISTED HEALTHY NUTRITION AND DIET SMARTPHONE APPLICATION

Adhering to a healthy diet is one of the key factors towards improving lifespan and quality of life. However, we do not always know which meals are unhealthy. Existing food tracking applications require manual input of consumed dishes, which is time consuming and inaccurate. Neural networks have been shown to be good at estimating food ingredients and nutrition from a picture. In this work we propose a novel 2-stage approach that includes post-train adaptive neural network for image segmentation into ingredients, assisted by a large language model (LLM) that transforms ingredient information into nutritional value, gives insight into improving nutrition of the food consumed, and has chat interface, where user can clarify any questions. Both vision and language models are inferred fully on the user's device. Thus, no server connection is required, and the user data remains private. We develop several prompts for nutritional value extraction, and analyze Qwen 2.5 0.5, 1.5, 3 billion parameter LLMs. To score nutritional value estimates we propose LLM-judge method. The final prompt improves nutritional value estimation score from 2.66 to 3.68, when using Qwen 2.5 3 billion model. The proposed application is more user friendly and provides nutrition tracking with less effort, when compared to existing applications, enabling faster food tracking with detailed information on improving nutrition.

Keywords: healthy nutrition, nutritional value, large language model, image segmentation, mobile computing, edge computing, computer vision.

1. Introduction. Having an unhealthy diet results in numerous risks to human health, such as cardiovascular diseases, diabetes, and many others. Therefore, tracking consumed nutritional value is important to avoid diseases and to improve quality of life. Existing smartphone applications [1,2] propose to manually enter information about food, which requires significant effort from the user. Recent research direction is to estimate food ingredients or nutritional value using a single photo of the user's meal [3–5].

In this work, we improve existing approaches by developing smartphone application with U-Net+PTA [6] image segmentation neural network that extracts ingredient information from the user's meal image. The information is then augmented using Qwen 2.5 [7] large language model (LLM) that estimates calories, fat, carbohydrates, proteins; gives insight into current meal and how to improve its nutrition. The user can continue the dialog by asking the model more specific questions. Processing is fully performed on the user device. Thus, the application does not require any paid large language model API subscriptions.

To summarize, our main contributions are as follows:

1. We develop a smartphone application for nutritional value estimation from a photo with 2 stages: image segmentation into ingredients (via U-Net+PTA network), nutritional value estimation (via Qwen 2.5 LLM).

2. Evaluating nutritional value estimates from cooked food is a hard task. Therefore, we develop an LLM-judge technique for evaluating nutritional value extraction results without known labels.

2. Related works. Several existing applications propose a solution to the problem, such as Samsung Health [1], MyFitnessPal [2]. These applications propose to enter information about food consumed manually by 1) scanning product bar-code; 2) specifying the number of calories manually; 3) searching for food by name. Each approach has its strength and limitations.

Scanning a bar-code is the easiest for the user. However, this is only useful if cooked food is purchased. Cooking from ingredients might result in significant change in calories (depending on their proportion and type of cooking (frying, boiling, etc.)). Also, bar-code scanning does not support all regions or products. For instance, Samsung Health shows "Invalid bar code error" when scanning products in Ukraine.

Specifying the number of calories manually is the most flexible, but the question is how to compute the number of calories. Recent research suggests that even trained nutritionists have high errors in estimating cooked food calories [4].

Finally, the option to find food by name. Based on our experience this is the most practical approach. For instance, Samsung Health proposes many variations of the dish with different modifications or cooking types. Also, favorite dishes can be saved for easier search, or even custom recipes can be added for future use. Still, the food must be manually weighted, and the system lacks suggestions on improving nutrition with minor taste changes.

To resolve the problem several neural network-based approaches have been proposed. Such systems try to estimate nutritional value from a photograph.

In [8] the authors collect web images and propose a fine-grained food classification dataset with 158,846 images assigned to one of 251 classes. While the dataset is large, classification limits the ability to recognize photos of multiple dishes at once or to distinguish between cooking recipes.

The authors of [3] propose a FoodSeg103 dataset, that contains 9490 images of food. Each image has a segmentation mask with one of 154 ingredient classes. Also, FoodSeg154 dataset is proposed with extra images of Asian food. The authors propose ReLeM method for joint training of a neural network on textual description of cooking recipes (using LSTM or Transformer) and image with segmentation masks, improving over existing approaches.

In [9] the authors fine-tune 7 billion parameter multi-modal LLM on different food-related tasks, including question answering and nutritional value estimation. However, the network takes 2 days of training 4 GPUs and is too large for a smartphone.

The authors of [4] note limitations of existing approaches, such as small number of categories and inability to automatically estimate food weight. Also, the authors note that given a photograph even expert nutritionists incorrectly estimate food calories. To resolve the problem, they collect a dataset and propose a novel neural network-based approach. Food has been cooked specifically for this experiment with all ingredients known. All ingredients were weighted before cooking, and their macronutrient information is taken from USDA Food and Nutrient Database. 5066 dishes with unique sets of ingredients have been cooked. Each dish has been captured from multiple angles with an RGB camera. Some dishes include additional depth information. The authors train Inception V2-based neural network to estimate food calorie, mass, fat, carbohydrates, and proteins. The authors show that their approach supersedes by a large margin human experts in estimating food calories. The developed method is suitable for complex nutritional value estimation.

However, for improved usability these approach lack on-device chat interface that would provide smart suggestions on how to improve the user's diet.

Therefore, in this work we propose a novel approach that enriches image-based food segmentation with an LLM-based chat interface. Thus, allowing for a more fine-gradient food information extraction and further smart suggestions. The developed application runs fully on the user's smartphone.

3. Materials and Methods. Large language models have shown to be effective in many domains, such as mathematics, programming, medicine and many others [7]. Existing LLMs either operate with a single modality (text) or can be multi-modal, understanding both text and images. For instance, Qwen 2.5 [7], ChatGPT 4 [10] are single-modal, and ChatGPT 4o [10], Gemma 3 [11] are multi-modal. While multi-modal LLMs contain general information in many domains, they are not finetuned for nutrition-related tasks, and finetuning LLM is a resource-intensive task. Also, they contain larger number of parameters, making them too slow for typical smartphone hardware; some of the models (e.g. ChatGPT 4o) can only be accessed using paid API. All of which makes them inapplicable for on-device nutritional value estimation

To solve the issue we propose a 2-stage approach where a lightweight convolutional neural network is used to extract ingredient information from food, then the information is fed into on-device Qwen2.5 3B LLM [7] for nutritional value estimation. The scheme of the proposed approach is shown in Fig. 1. In blue color user actions are shown, in yellow processing done by the vision model, in green LLM processing.

For food ingredient extraction we propose to use a lightweight U-Net+PTA image segmentation network [6], which combines U-Net [12] neural network and Post-Train Adaptive (PTA) [13] as a backbone. The key feature of the network is that after training the number of active parameters can be changed on user demand. Research [14] has shown, that using the following 3 post-train configurations offers the most flexibility: LLL (light), HHH (heavy), BBB (combines both heavy and light branches). To speed up inference on low-end devices lightweight configuration is suggested (80% of floating-point operations when compared to heavy). While heavy and both configurations improve the network quality on faster devices.

Some prediction noise is present in image segmentation into ingredients as in certain regions the model might not be confident which of the ingredients that is. To remove the noise we ignore ingredient predictions, where the model confidence is below 0.85. Then we compute the area taken by each ingredient relative to the total image area. We remove ingredient predictions if the total area of prediction is less than 4%. Such ingredients will not significantly influence nutritional value results and are typically prediction noise. Thresholding constants are found empirically. Finally, we remove background class, and normalize ingredient area predictions, so that the total sum is 100%.

Extracted information is passed to the large language model using the following

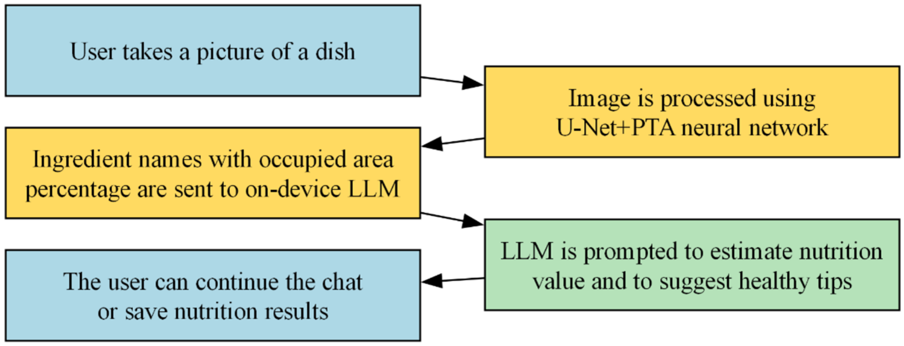


Figure 1. Scheme of the proposed nutrition suggestion app.

prompt: "I have food on a plate with ***ingredients***. Estimate food nutritional value (calories, fat, carbohydrates, proteins). Then give suggestions on improving nutrition. Be brief." ***Ingredients*** are replaced with actual ingredients extracted by the vision model. Analysis of different prompts is shown in the Discussion section.

Overall, 2 separate models are used to perform the inference:

1. U-Net+PTA image segmentation model, that extracts ingredients from an image.
2. Qwen 2.5 LLM to estimate nutritional value from ingredients.

To evaluate the vision model $Dice_{score}$ is used, which is a common image segmentation metric:

$$Dice_{score} = \frac{2\sum_{i=1}^{N} p_i g_i + \varepsilon}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} g_i^2 + \varepsilon}, \qquad (1)$$

where p_i is the predicted probability distribution, g_i is the ground true one-hot vector, N is the number of classes to distinguish between, ε is a small constant.

Evaluating nutritional value, generated by the LLM, is more complicated as FoodSeg103 dataset doesn't propose nutritional value information. Also, as is shown in [4] people have high error in evaluating nutritional value information from images. Therefore, an automated approach should be developed.

In [15] the authors propose a method to evaluate LLM outputs by using another larger LLM as a judge. This is useful in cases when true labels are not available or might be presented in different ways. Reasoning LLMs (e.g. Qwen 3) require more compute, but in exchange are superior to non-reasoning LLMs (e.g. Qwen 2.5) in complex tasks, that require calculations, logical derivations, etc. [16].

In this work we develop a method that uses reasoning LLM as a judge evaluating nutritional value information correctness. We use Qwen3 [17] 30 billion parameter reasoning model. The authors show its superiority over Qwen 2.5 and many other models. Obviously, the model is too large for inference on a smartphone.

For the model to act as judge the prompt that is shown in Table 1 has been developed. The prompt has several parameters: **food_area** parameter is replaced by the vision model output, **model_estimates** is a JSON document with outputs for a single Qwen 2.5 LLM model produced by different prompts. During experiments we have found that detailed judging information (e.g. "+1 if nutritional value estimate is within 10% of real estimate") results in scoring more aligned to the end user expectations, than in case when the model is asked simply to score from 1 to 5.

Table 1.

LLM judge prompt and its parameters

Prompt	**Parameters**
You are a nutrition expert. You evaluate LLM nutritional value estimates. The model is given different prompts with the following information about food: ***food_area*** Replies for each prompt are the following: **model_estimates**. Score model replies to the prompt from 1 to 5. Always give a score of at least 1. Give +1 if nutritional value estimate (calories, fat, carbohydrates, proteins) is present and it mentions relevant food. Give +1 if nutritional value estimate is within 10% of real estimate. Give +1 if suggestions on improving nutrition are relevant to the food given. Give +1 if the prompt doesn't include unnecessary information. Reply in json format: [{"prompt": prompt_name, "score": score}]. Array should contain scores for each prompt.	food_area, model_estimates

Scores are given from 1 to 5. Example reply of the LLM judge is shown below:

```
[
    {
        "prompt": "01-basic",
        "score": 3
    },
    {
        "prompt": "02-specify-exactly-what-to-estimate",
        "score": 5
    },
    {
        "prompt": "03-grams-instead-of-percents",
        "score": 4
    },
    {
        "prompt": "04-be-brief",
        "score": 4
    }
]
```

4. Experiments. To train U-Net+PTA neural network Stochastic Gradient Descent optimizer has been used with initial learning rate of 0.1 for 90 epochs. Learning rate is multiplied by 0.1 every 30 epochs. Batch size is set to 32. Images of size 384×384 are used during training. Random resized crop and horizontal flip augmentations are used.

$Dice_{loss}$ is used as training loss:

$$Dice_{loss} = 1 - \frac{2\sum_{i=1}^{N} p_i g_i + \varepsilon}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} g_i^2 + \varepsilon}. \tag{2}$$

We use FoodSeg103 [3] dataset for training and evaluation of the U-Net+PTA neural network. The dataset contains 2 subsets: training (4983 images) and test (2135 images). We randomly split the original training set into actual train and validation in proportion 80/20. The final FoodSeg103 dataset split is shown in Table 2. To prevent overfitting the best model is selected on the validation set. Results are reported on the test set.

Table 2.

FoodSeg103 dataset split used for U-Net+PTA model training and evaluation

Subset	**Number of images**
Train	3986
Validation	997
Test	2135

To process ingredient information and extract nutritional value data Qwen 2.5 open-source non-reasoning LLM is used in 0.5, 1.5, 3 billion parameter variants. Fine-tuning the model is resource intensive. LLMs have been shown to solve many tasks in zero-shot mode (without pretraining) [18] by using prompting. Therefore, we develop and analyze several prompting strategies for nutritional value extraction.

Finally, we develop smartphone application called "Food, Nutrition and Diet App". The custom-trained U-Net+PTA neural network is inferred on-device using ONNX runtime, while Qwen 2.5 is run via MediaPipe library [19]. The application targets Android 10 or newer devices with a potential of expanding to other platforms.

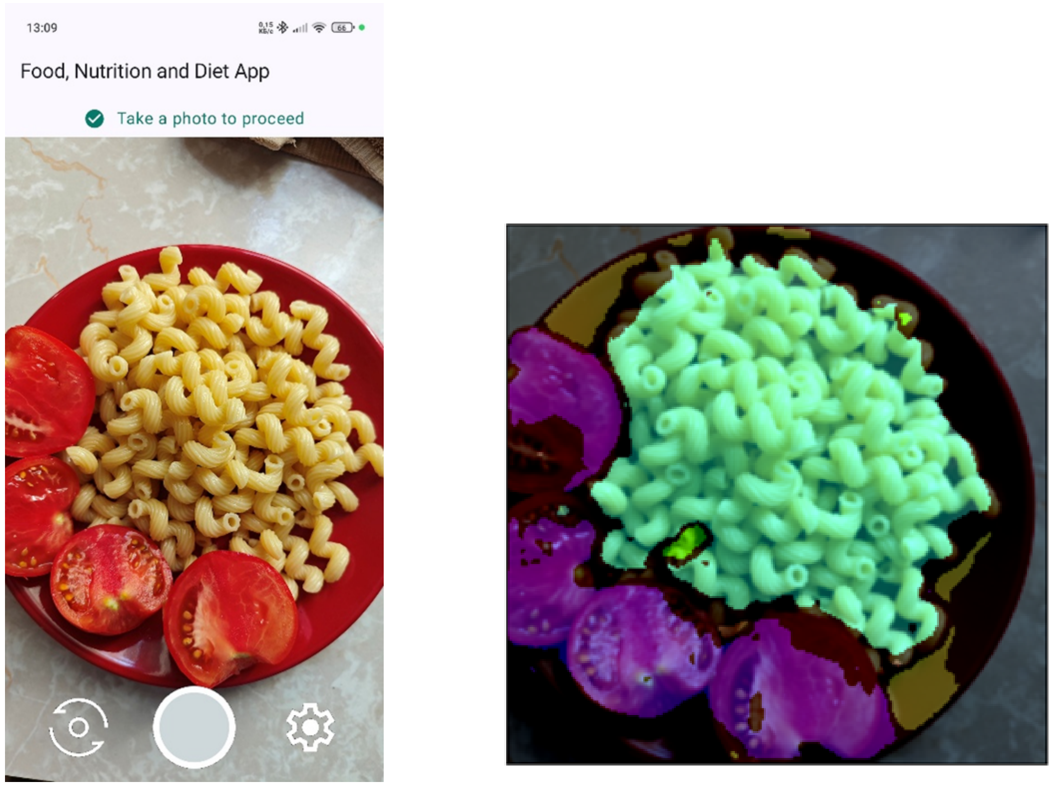
5. Results. $Dice_{score}$ of the trained U-Net+PTA neural network is shown in Table 3. The largest version (BBB) has the highest score, while the smallest (LLL) offers a compromise in terms of speed to quality ratio.

Table 3.

Image segmentation performance of the U-Net+PTA network

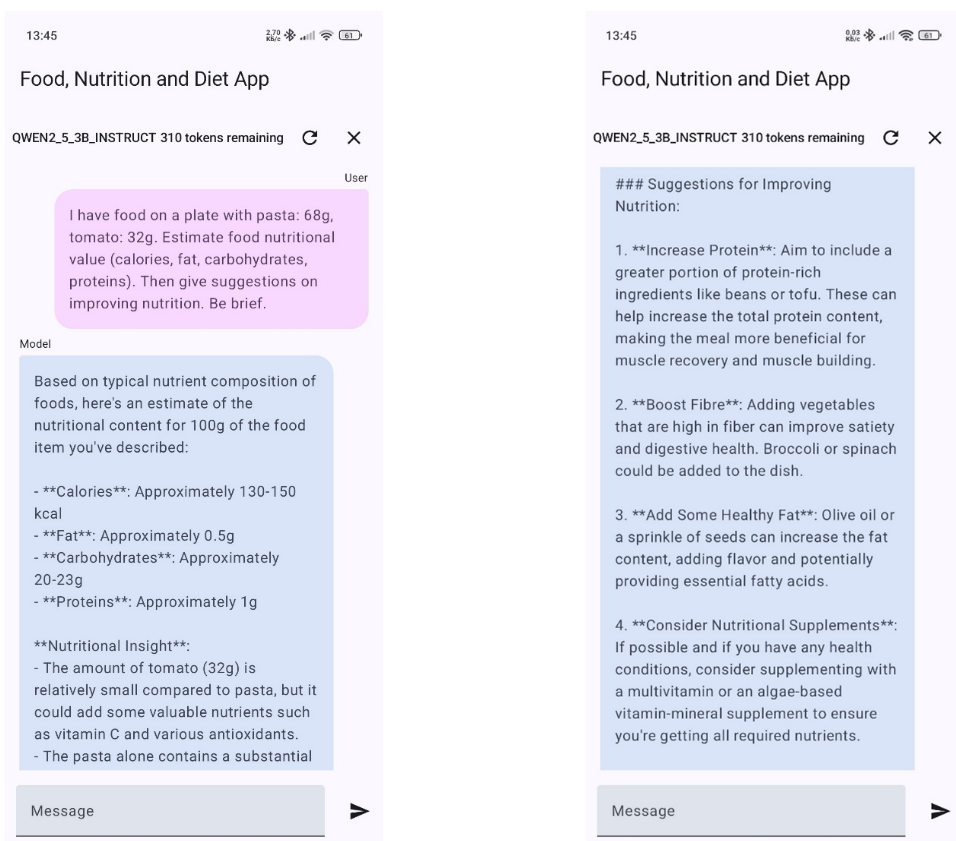
Configuration	**Dice score**
BBB	0.5165
HHH	0.5143
LLL	0.4987

Example of dish processing by the vision model is shown in Fig. 2. Fig. 2 (a) shows interface of the developed "Food, Nutrition and Diet App" with a plate, capturing tomatoes and pasta. Fig. 2 (b) shows segmentation, where tomatoes and pasta are highlighted in color. After thresholding with 0.85 confidence the computed area for each of the ingredients is: background: 43.86%; shrimp: 0.04%; pasta: 35.68%; rice: 0.30%; tomato: 17.03%; onion: 0.12%; pepper: 2.97%. As can be seen, most of the dish area is correctly recognized as pasta and tomatoes. Other components are filtered out by the 4% area threshold. After thresholding and normalization, we get pasta: 68%, tomato: 32%.

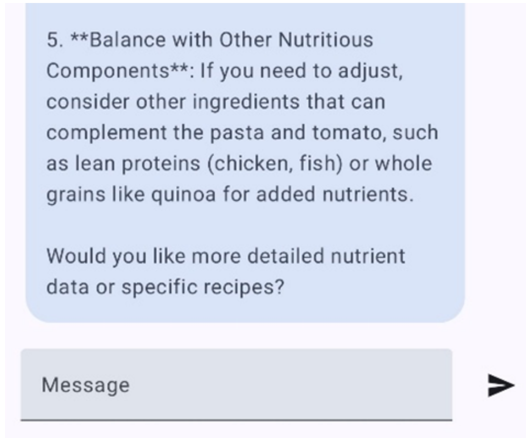


(a) (b)

Figure 2. Sample image and its segmentation by U-Net+PTA network.



(a) (b)



(c)

Figure 3. Generated prompt (a) message in pink, and LLM reply (a)–(c) in blue.

The extracted information is passed to the Qwen 2.5 3B LLM with a prompt "I have food on a plate with pasta: 68g, tomato: 32g. Estimate food nutritional value (calories, fat, carbohydrates, proteins). Then give suggestions on improving nutrition. Be brief.". We have found out that replacing percents with grams results in better LLM reply (the prompt analysis is given in the Discussion section). The generated prompt and full LLM reply are shown in Fig. 3. The prompt is in pink, the reply is in blue. As is demonstrated, the model gives an insight into nutritional value of the current meal, followed by suggestions on improving nutrition. The user can continue the discussion using the chat interface at the bottom of the screen.

By analyzing several dishes, we have found that the model is quite useful for quick extraction of information about nutritional value and giving suggestions to improve it. The process is much quicker and more informative than trying to scan barcodes or finding each of the food components manually.

Prompts given to the LLM have significant influence on the quality of the generated output. Thus, in this work we develop several prompts, which are shown in Table 4. Parameter is shown in bold, changes between prompts are underlined. The goal is to get model output that contains correct and specific information about food nutritional value. Also, the answer should not contain unnecessary information.

We use Qwen3:30b thinking model as a judge as described in the Materials and Methods section. We evaluate each of the prompts for Qwen2.5 0.5, 1.5, and 3 billion parameter models, that can be run on smartphones. Ingredients is a parameter, filled with information about ingredients either in percents (prompts 1 and 2) or in grams (prompts 3, 4).

Evaluation results are shown in Table 5 and Fig. 4. Qwen2.5:3b with prompt 1 has average score of 2.66, by using prompt 4 the score significantly improves to 3.68. Similarly, smaller models are also improved from prompt 1 to 4. Qwen2.5:3b has the best results for each prompt. We analyze scores given to the models and prompts on a representative example in the Discussion section.

6. Discussion. In this section we analyze LLM-based nutritional value extraction in the developed application. First, we show how the final prompt has been built. We have analyzed model generations for several ingredient inputs and show a representative sample in Table 6, where the prompt is shown along with

Table 4.

Prompts developed for nutritional value extraction

ID	Name	Sample Prompt
1	Basic	I have food on a plate with ***ingredients (%)***. Estimate food nutritional value. Then give suggestions on improving nutrition.
2	Specify exactly what to estimate	I have food on a plate with ***ingredients (%)***. Estimate food nutritional value (calories, fat, carbohydrates, proteins). Then give suggestions on improving nutrition.
3	Grams instead of percents	I have food on a plate with ***ingredients (grams)***. Estimate food nutritional value (calories, fat, carbohydrates, proteins). Then give suggestions on improving nutrition.
4	Be brief	I have food on a plate with ***ingredients (grams)***. Estimate food nutritional value (calories, fat, carbohydrates, proteins). Then give suggestions on improving nutrition. Be brief.

Table 5.

LLM judge scores for each model and prompt

Model	Prompt	Score
qwen2.5:0.5b	01 basic	1.78
	02 specify exactly what to estimate	2.48
	03 grams instead of percents	2.67
	04 be brief	3.30
qwen2.5:1.5b	01 basic	2.43
	02 specify exactly what to estimate	2.96
	03 grams instead of percents	3.36
	04 be brief	3.57
qwen2.5:3b	01 basic	2.66
	02 specify exactly what to estimate	3.51
	03 grams instead of percents	3,57
	04 be brief	3.68

analysis of the model output with information about what can be improved. We do not show model outputs here as they are quite long.

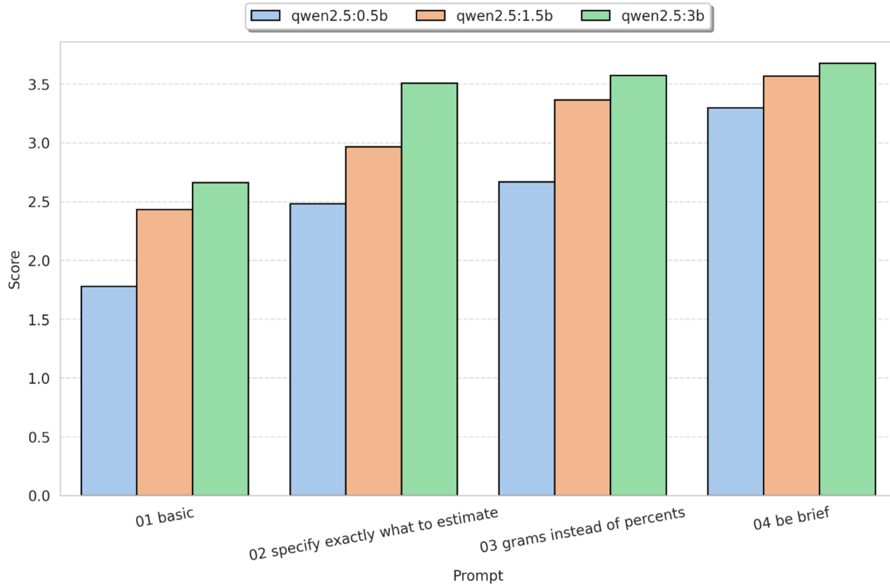


Figure 4. Nutritional value extraction scores for Qwen2.5 model variants and different prompts.

Table 6.

Analysis of the influence of different prompts on nutritional value extraction quality

ID	Sample Prompt	What can be improved?
1	I have food on a plate with pasta: 68%, tomato: 32%. Estimate food nutritional value. Then give suggestions on improving nutrition.	Food nutritional value is quite vague and changes significantly from query to query. Sometimes the model generates information about vitamins and fats, sometimes about carbohydrates.
2	I have food on a plate with pasta: 68%, tomato: 32%. Estimate food nutritional value <u>(calories, fat, carbohydrates, proteins)</u>. Then give suggestions on improving nutrition.	By specifying list of nutritional value types (underlined), the model is much more consistent and specifies approximate number of the specified nutrition items. However, the model mixes food percentages and weights, which should be avoided.
3	I have food on a plate with pasta: <u>68g</u>, tomato: <u>32g</u>. Estimate food nutritional value (calories, fat, carbohydrates, proteins). Then give suggestions on improving nutrition.	By using grams instead of percents, the model does not mix percents and grams. While the percentage of area taken does not directly equal to grams, this is a sufficiently fair assumption to get nutritional value estimate per 100g of food. The model output is quite long still.
4	I have food on a plate with pasta: 68g, tomato: 32g. Estimate food nutritional value (calories, fat, carbohydrates, proteins). Then give suggestions on improving nutrition. <u>Be brief.</u>	By adding “Be brief” the model's output is shortened, without losing much information.

Qwen2.5 model is available in multiple sizes: 0.5B, 1.5B, 3B, 7B, 14B, 32B, 72B, where B stands for billions of parameters. While models with 7B and above are

targeted for desktop computers and servers. The 0.5B, 1.5B, 3B can be considered mobile-friendly. For instance, the example prompt that we use takes 32 seconds for the LLM to generate full reply on Snapdragon 8 Gen 3 processor. While it might seem a long wait, in the developed application the reply generation is interactive, so text generation is faster than the average user's reading speed.

Analysis of replies of 0.5B, 1.5B, and 3B models is shown in Table 7. The reply of Qwen2.5 3B can be seen in Fig. 3. As can be seen, 0.5B model reply is very short. Also, the model mixes units when generating nutritional value. For instance, calories have no unit, while others are in percents. Nutritional value suggestions are very generic and are only slightly related to the ingredients detected by the vision model. The reply of 1.5B model is much more structured, with suggestions being more reasonable.

Table 7.

LLM processing of the same prompt depending on the number of parameters

LLM	**Generated output**
qwen2.5:0.5b	Calories: 410, Fat: 7%, Carbohydrates: 55%, Proteins: 3% Tips: Try adding more vegetables and fruits to increase protein intake. Tips: Choose whole grains and legumes for better fiber content and a balanced diet. Remember to consume less saturated fats.
qwen2.5:1.5b	Food Composition: - Pasta: 68 grams - Tomato: 32 grams Nutrition Analysis: Calories: About 910 kcal (depending on the type of pasta) Fat: Approximately 14g Carbohydrates: Around 75g Proteins: Roughly 8g Improvement Suggestions: 1. **Increase Vegetables and Fiber**: Include more vegetables like spinach, broccoli, or carrots to increase fiber intake which can lead to better digestion and satiety. 2. **Boost Protein Source**: Incorporate lean proteins such as chicken thighs, turkey, or fish for additional muscle building benefits and energy levels. 3. **Add Healthy Fats**: Consider adding a small amount of olive oil, avocado, or nuts (like almonds) which can contribute healthy fats to your diet while lowering saturated fat content. 4. **Reduce Saturated Fat Sources**: Opt for cooking methods like boiling, steaming, baking, grilling, and broiling rather than frying with butter or margarine to minimize the risk of unhealthy cholesterol production. Remember, nutrition is a complex mix of various foods tailored to individual preferences and dietary restrictions. Adjusting these suggestions can help balance your diet towards better overall health benefits.

Interestingly, all 3 models have generated significantly different nutritional value numbers for the given food. So, which of the models is the closest to the actual results?

In Table 8 we compare extracted nutritional value with the one computed manually based on Internet sources. The final line of the table shows overall information for the meal of 68 grams of pasta and 32 grams of tomato. Apparently, the 0.5B model result is far from the truth, with even the units being incorrectly specified. In contrast, while the 1.5B model has proper units for nutritional value, it has significantly overestimated most components in the meal. 3B model is the closest to the actual computation and can be used for food nutritional value estimation.

Table 8.

Comparison of nutritional value extraction by different LLMs vs ground truth

Data Source	**Calories**	**Fat**	**Carbohydrates**	**Proteins**
qwen2.5:0.5b	410	7%	55%	3%
qwen2.5:1.5b	About 910 kcal (depending on the type of pasta)	Approximately 14g	Around 75g	Roughly 8g
qwen2.5:3b	Approximately 130-150 kcal	Approximately 0.5g	Approximately 20-23g	Approximately 1g
Actual data for 100g of pasta	131 kcal	1.5 g	25g	5g
Actual data for 100g of tomatoes	18 kcal	0.2 g	3.9g	0.9g
Computed value for 100g of dish taken on the photo	94.84 kcal	1.08 g	18.25 g	3.69 g

7. Conclusions and prospects for further research. In this work we have developed Food, Nutrition and Diet Application, that combines U-Net+PTA image segmentation model with Qwen 2.5 3B large language model. The vision model efficiently extracts ingredients from the user-provided photograph of food, while LLM is prompted to write down nutritional value and give suggestions. We develop several prompts for nutritional value extraction, and analyze Qwen 2.5 0.5, 1.5, 3 billion models. To score nutritional value estimates we propose LLM-judge method, based on Qwen 3 30 billion parameter reasoning model. The final prompt improves nutritional value estimation score from 2.66 to 3.68, using Qwen 2.5 3 billion model. The proposed application is more user friendly and provides nutrition tracking with less effort, when compared to existing applications, that require barcode scanning or manual input of the dish by name.

The prospect for further research is to include estimation of actual meal weight using neural networks. Thus, the user would get information about nutrition in their portion, and not general information per 100g of food.

Acknowledgements. The work is supported by the state budget scientific research project of Dnipro University of Technology "Development of new adaptive information technologies for recognizing objects in the surrounding environment" (state registration number 0123U100012).

References

1. Samsung Health | Fitness & Health Tracking App. (2025). Retrieved from https://www.samsung.com/us/apps/samsung-health
2. Calorie Tracker & BMR Calculator to Reach Your Goal | MyFitnessPal. (2025). Retrieved from https://www.myfitnesspal.com/en
3. Wu, X., Fu, X., Liu, Y., Lim, E.-P., Hoi, S. C. H., & Sun, Q. (October 20–24, 2021). A large-scale benchmark for food image segmentation. *MM '21: ACM Multimedia Conference.* Virtual Event: China, 506–515. https://doi.org/10.1145/3474085.3475201
4. Thames, Q., Karpur, A., Norris, W., Xia, F., Panait, L., Weyand, T., & Sim, J. (June 19–25, 2021). Nutrition5k: Towards automatic nutritional understanding of generic food. *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021.* Virtual, 8903–8911. https://doi.org/10.1109/CVPR46437.2021.00879
5. Salvador, A., Drozdzal, M., Giró-i-Nieto, X., & Romero, A. (June 16–20, 2019). Inverse cooking: Recipe generation from food images. *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019.* Long Beach, CA: USA, 10453–10462. https://doi.org/10.1109/CVPR.2019.01070
6. Khabarlak, K. (2022). Post-train adaptive U-Net for image segmentation. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 2, 73–78. https://doi.org/10.32782/IT/2022-2-8
7. Yang, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Li, C., Liu, D., Huang, F., Wei, H., Lin, H., Yang, J., Tu, J., Zhang, J., Yang, J., Yang, J., Zhou, J., Lin, J., Dang, K., . . . , & Qiu, Z. (2024). Qwen2.5 technical report. *CoRR, abs/2412.15115.* https://doi.org/10.48550/ARXIV.2412.15115
8. Kaur, P., Sikka, K., Wang, W., Belongie, S. J., & Divakaran, A. (2019). FoodX-251: A dataset for fine-grained food classification. *CoRR, abs/1907.06167.* https://doi.org/10.48550/arXiv.1907.06167
9. Yin, Y., Qi, H., Zhu, B., Chen, J., Jiang, Y.-G., & Ngo, C.-W. (2023). FoodLMM: A versatile food assistant using large multi-modal model. *CoRR, abs/2312.14991.* https://doi.org/10.48550/ARXIV.2312.14991
10. Introducing ChatGPT | OpenAI. (2025). Retrieved from https://openai.com/index/chatgpt
11. Kamath, A., Ferret, J., Pathak, S., Vieillard, N., Merhej, R., Perrin, S., Matejovicova, T., Ramé, A., Rivière, M., Rouillard, L., Mesnard, T., Cideron, G., Grill, J.-B., Ramos, S., Yvinec, E., Casbon, M., Pot, E., Penchev, I., Liu, G., . . . , & Nardini, I. (2025). Gemma 3 technical report. *CoRR, abs/2503.19786.* https://doi.org/10.48550/ARXIV.2503.19786
12. Ronneberger, O., Fischer, P., & Brox, T. (October 5–9, 2015). U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015 – 18th International Conference Munich, Proceedings, Part III.* Germany, 9351, 234–241. https://doi.org/10.1007/978-3-319-24574-4%255C_28
13. Khabarlak, K. (March 23–25, 2022). Post-train adaptive MobileNet for fast anti-spoofing. *Proceedings of the 3rd International Workshop on Intelligent Information Technologies & Systems of Information Security.* Khmelnytskyi: Ukraine, 3156, 44–53. Retrieved from https://ceur-ws.org/Vol-3156/keynote5.pdf
14. Khabarlak, K. (September 17–19, 2025). Improved post-trained adaptive network with texture bias robustness. *2025 14th International Conference on Advanced Computer Information Technologies (ACIT).* Sibernik: Croatia.
15. Khabarlak, K., Koriashkina, L., Laktionov, I., & Diachenko, G. (2025). Zero-Shot Plant Disease Recognition Using Open Large Vision-Language Models. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 3.
16. Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E. H., Le, Q. V., & Zhou, D. (November 28 – December 9, 2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022.* New Orleans, LA: USA. Retrieved from https://papers.nips.cc/paper_files/paper/2022/hash/9d5609613524ecf4f15af0f7b31abca4-Abstract-Conference.html

17. Yang, A., Li, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Gao, C., Huang, C., Lv, C., Zheng, C., Liu, D., Zhou, F., Huang, F., Hu, F., Ge, H., Wei, H., Lin, H., Tang, J., ..., & Qiu, Z. (2025). Qwen3 technical report. *CoRR, abs/2505.09388.* `https://doi.org/10.48550/ARXIV.2505.09388`
18. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ..., & Amodei, D. (December 6–12, 2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020.* NeurIPS 2020, Virtual. Retrieved from `https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html`
19. Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M. G., Lee, J., Chang, W.-T., Hua, W., Georg, M., & Grundmann, M. (2019). MediaPipe: A framework for building perception pipelines. *CoRR, abs/1906.08172.* `https://doi.org/10.48550/arXiv.1906.08172`

Хабарлак К. С. Мультимодальний додаток здорового харчування та дієти на основі великої мовної моделі.

Дотримання здорового харчування є одним з ключових факторів, що впливають на тривалість і якість життя. Однак ми не завжди знаємо, які страви є шкідливими. Існуючі додатки для відстеження харчування вимагають ручного введення спожитих страв, що є трудомістким і неточним. Нейронні мережі добре справляються з оцінкою інгредієнтів і поживної цінності їжі на основі фотографій. У цій роботі ми пропонуємо новий двоетапний підхід, який поєднує адаптивну після навчання нейронну мережу для сегментації інгредієнтів на зображенні, доповнену великою мовною моделлю (LLM), яка перетворює інформацію про інгредієнти в дані про поживну цінність, дає відомості про поліпшення харчової цінності та має чат-інтерфейс, де користувач може уточнити будь-які питання. Як візуальні, так і мовні моделі повністю працюють на пристрої користувача. Таким чином, підключення до сервера не потрібне, а дані користувача залишаються конфіденційними. В роботі розроблено кілька підказок для моделі вилучення харчової цінності та проаналізовано Qwen 2.5 LLM з 0.5, 1.5, 3 мільярдами параметрів. Для оцінки харчової цінності ми пропонуємо метод LLM-судді. Розроблена LLM-підказка покращує оцінку харчової цінності з 2.66 до 3.68 при використанні моделі Qwen 2.5 на 3 мільярди параметрів. Запропонований додаток є більш зручним для користувача і забезпечує відстеження поживної цінності з меншими зусиллями в порівнянні з існуючими додатками, забезпечуючи швидше відстеження їжі з детальною інформацією про поліпшення харчової цінності.

Ключові слова: здорове харчування, харчова цінність, велика мовна модель, сегментація зображень, мобільні обчислення, крайові обчислення, комп’ютерний зір.

Recived 30.07.2025