Evaluating OCR, LLM, and VLM Approaches for Nutritional Information Extraction and Nutri-Score Calculation from Packaged Food Labels

Cagatay Onbasioglu, Rizan Hasan, Serkan Saglam
Istanbul University - Cerrahpasa
Istanbul, Türkiye
cagatayonbasioglu@ogr.iuc.edu.tr, rizanhasan@ogr.iuc.edu.tr, serkan.saglam@iuc.edu.tr

***Abstract*—Automated extraction of nutritional information from packaged food images is challenging due to variations in label layout, text size, image quality, and package design. This study compares OCR-based, OCR-LLM, and vision-language model (VLM) approaches for extracting structured nutritional information from Turkish food package images and examines their impact on Nutri-Score classification. A dataset of 1,000 product images was constructed from Open Food Facts. Three OCR systems were evaluated with rule-based parsing and in combination with locally deployed LLMs, while Qwen2.5-VL 7B and Gemini 3.5 Flash-Lite were evaluated as direct VLM-based approaches. All methods extracted eight nutritional fields. The results showed that OCR quality strongly influenced downstream LLM performance and that LLM-based processing substantially improved extraction compared with rule-based OCR alone. Qwen3.8-27B with GLM-OCR achieved the best performance among the OCR-LLM configurations, while Gemini 3.5 Flash-Lite achieved the highest overall field-level accuracy. For Nutri-Score classification, Gemini 3.5 Flash-Lite achieved 89.2% accuracy, compared with 78.5% for Qwen3.8-27B with GLM-OCR. These results highlight the importance of both OCR quality and semantic interpretation in automated food-label understanding.**

I. Introduction

Packaged food labels contain information about the energy and nutrient content of food products. Accurate interpretation of this information is useful for product comparison and dietary assessment [1]. However, differences in package design, font style, language, image quality, and table layout can make the automatic extraction of nutritional information from product images difficult [2], [3]. Therefore, extracting textual and numerical information from food packaging and converting it into structured nutritional data is an important step in automated nutrition analysis systems [3].

Optical Character Recognition (OCR) is commonly used to extract text from product images and documents. Although current OCR systems can provide accurate text recognition in many cases, processing nutritional labels also requires the extracted information to be interpreted correctly [2]. Numerical values must be matched with the corresponding nutrients, measurement units should be interpreted correctly, and values reported per serving must be distinguished from those reported per 100 g or 100 ml. The position of rows, columns, and headers in nutritional tables can also affect how the extracted information is interpreted. Therefore, text that is recognized correctly by OCR may still lead to errors when it is converted into structured nutritional data. These errors can also affect later calculations and the nutritional assessment of the product [4], [5].

Recent developments in large language models (LLMs) and vision-language models (VLMs) have led to their use in extracting nutritional information from food package images [3], [6]. LLMs can process OCR-extracted text to match nutrient names with numerical values, standardize different writing formats and measurement units, and convert unstructured text into structured nutritional data [7], [8]. VLMs can process both textual and visual information directly from an image, allowing nutritional values to be interpreted together with their spatial context, such as the corresponding headers, rows, columns, or table sections [9].

Although OCR, OCR-LLM, and VLM-based approaches can be used for nutritional information extraction, their performance may differ depending on how textual, numerical, and visual information is processed. In addition, errors in extracted nutrient values or nutrient-value associations may affect not only the structured data but also subsequent nutritional assessment [10], [11]. Evaluating these methods only at the text recognition level may therefore provide limited information about their effect on the final result. In this study, we compare OCR-based, OCR-LLM, and VLM-based approaches for extracting structured nutritional information from Turkish food package images and examine how extraction errors affect downstream Nutri-Score classification. In particular, we investigate the contribution of locally deployed LLMs when used to interpret OCR output.

Based on this objective, the following research questions are addressed:

- **RQ1:** How does the performance of OCR-based, OCR-LLM, and VLM-based approaches differ in extracting structured nutritional information from packaged food images?
- **RQ2:** To what extent does the integration of locally deployed LLMs improve nutritional information extraction compared with rule-based processing of OCR output?
- **RQ3:** To what extent do errors in extracted nutritional information affect the resulting Nutri-Score classification?

To address these questions, RQ1 is examined by comparing field-level extraction performance across OCR-based, OCR-LLM, and VLM-based approaches. RQ2 is investigated by comparing rule-based processing of OCR output with LLM-based processing of the same OCR outputs. RQ3 is evaluated by calculating Nutri-Score from the extracted nutritional values and comparing the resulting classes with the corresponding reference classes.

This study's main contribution is a comparative evaluation of OCR-only, OCR-LLM, and direct VLM pipelines on the same Turkish food-label benchmark using a common field-level evaluation protocol. In addition to comparing extraction accuracy, the study examines how errors in structured nutritional information affect downstream Nutri-Score classification.

The remainder of this paper is organized as follows. Section II reviews related work on food-label information extraction, document understanding, and automated nutritional assessment. Section III describes the dataset, extraction pipelines, Nutri-Score calculation, and evaluation methodology. Section IV presents and discusses the experimental results, including extraction performance and the impact of extraction errors on Nutri-Score classification. Finally, Section V concludes the study and outlines directions for future work.

II. Background and Related Work

A. Nutritional Information Extraction from Food Labels

Automated analysis of food-related information has been studied using computer vision, deep learning, and text-based approaches. Kaushal et al. reviewed recent computer vision and deep learning methods for food recognition and nutrient estimation, highlighting the increasing use of image-based methods for nutritional analysis [12]. In a different setting, Ma et al. used ingredient statements from a large branded food product database to predict food categories and nutrient values with deep learning models [13]. These studies show that nutritional information can be estimated or inferred from different forms of food-related data, including images and textual product information.

More recent studies have focused directly on information available on food package labels. Rosyadi et al. proposed a system that combines Paddle-OCR with ChatGPT to identify ingredients from package labels and retrieve related information [14]. Assiri et al. investigated the extraction of nutritional elements and values from bilingual Arabic - English food labels using large language models and evaluated the models on a manually curated set of food labels [3]. These studies indicate that OCR and language-model-based methods can support the extraction and interpretation of information from food labels.

B. Language and Vision Models for Document Information Extraction

Recent research on document understanding has increasingly focused on extracting structured information from visually complex documents. Zhang et al. [15] reviewed document parsing methods ranging from modular pipelines to end-to-end vision-language models, emphasizing the importance of layout analysis, content extraction, and multimodal information integration. OCR-free approaches have also been proposed to reduce dependence on separate text recognition stages. Hu et al. [16] introduced a structure-aware approach for understanding text-rich documents, tables, and charts directly from images. These studies highlight the role of document structure and spatial information in extracting information beyond plain text recognition.

Language models have also been integrated with OCR-based pipelines for structured information extraction. Khanchandani et al. [17] combined OCR and LLM-based processing for invoice data extraction, using the language model to support semantic interpretation and field-level information extraction from OCR outputs. Shen et al. [18] proposed a multilingual visual information extraction framework designed to improve generalization across languages. Together, these studies show how OCR, language models, and visual information can be combined for structured information extraction from documents with different layouts and languages.

C. Nutri-Score and Automated Nutritional Assessment

Nutri-Score is a front-of-package nutrition labeling system that summarizes the nutritional quality of food products using five classes ranging from A to E. Lower classes indicate a more favorable nutritional profile, while higher classes represent a less favorable profile. Hercberg et al. [19] described Nutri-Score as a public health tool intended to support the interpretation and comparison of the nutritional quality of food products.

Several studies have examined the design and validation of the Nutri-Score system. Van der Bend et al. [20] evaluated the validation process of the algorithm, while Hafner and Pravst [21] assessed the revised Nutri-Score algorithm and its suitability for front-of-package labeling across European food products. Since Nutri-Score is derived from nutritional information, errors in automatically extracted nutrient values may also affect the resulting nutritional assessment.

III. Materials and Methods

Fig. 1 presents the study's experimental workflow. Packaged food images were evaluated using three processing strategies: OCR-only extraction, OCR followed by a locally deployed LLM, and direct VLM-based extraction. Each strategy produced structured nutritional information, which was subsequently used for Nutri-Score calculation and comparative performance evaluation.

A. Dataset Collection and Preparation

Open Food Facts (OFF) was used as the data source in this study. At the time of data collection, the database contained approximately 4.7 million product records. Using the `countries_tags` field, 14,221 products associated with Türkiye were identified. Among these, 3,321 products contained an image of the nutritional information table. The

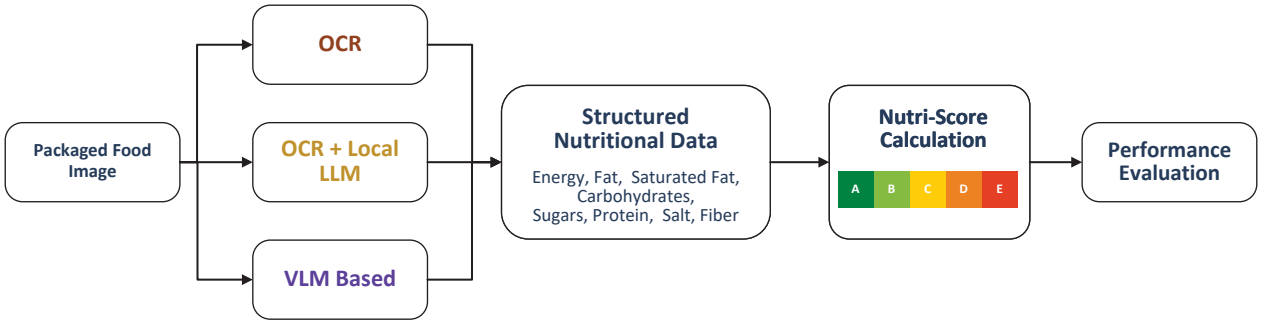


Fig. 1. Overview of the experimental workflow for nutritional information extraction and evaluation.

language of the nutritional labels was determined by manual visual inspection performed by two researchers. This screening identified 1,517 images with Turkish nutritional labels, 1,051 images with labels in other languages, and 753 images for which the language could not be reliably determined. Only the 1,517 images with Turkish nutritional labels were retained for further screening.



(a)

(b)

(c)

(d)

Fig. 2. Example packaged food images included in the final dataset.

The remaining Turkish-label images were further screened using the Open Food Facts reference metadata. Eight fields were considered: energy, fat, saturated fat, carbohydrates, sugars, protein, salt, and fiber. Images were retained when a nutritional information panel was present and reference values were available for at least seven of the eight target nutritional fields. This filtering process resulted in a final dataset of 1,000 product images for the experiments.

Example images from the final dataset are shown in Fig. 2. The examples illustrate several visual challenges encountered during nutritional information extraction. Fig. 2(a) shows text embedded within the overall package design, while Fig. 2(b) contains dense information presented in a small font. Fig. 2(c) illustrates the effect of reflections and surface curvature, whereas Fig. 2(d) shows reduced readability due to glare, low contrast, and perspective variation. These characteristics can make both text recognition and the correct association of nutritional values with their corresponding fields more difficult.

B. Nutritional Information Extraction Pipelines

In this study, three different approaches were evaluated for extracting structured nutritional information from packaged food images: OCR-only extraction, a hybrid approach in which OCR output is processed by a locally deployed LLM, and direct VLM-based extraction. All approaches targeted the same benchmark of 1,000 product images and the same eight nutritional fields: energy, fat, saturated fat, carbohydrates, sugars, protein, salt, and fiber. This enabled the methods to be compared in terms of both nutritional information extraction performance and their impact on the subsequent Nutri-Score calculation.

1) OCR-Based Extraction: Three OCR systems were evaluated for extracting textual information from the nutritional label images: GLM-OCR (`zai-org/GLM-OCR`), GOT-OCR-2.0 (`stepfun-ai/GOT-OCR-2.0-hf`), and Paddle-OCR. Each OCR system was applied independently to the product images, producing a separate raw textual output for each image. The raw outputs of all three OCR systems were processed using the same shared rule-based parser, without OCR-engine-specific tuning. The parser used regular-expression matching to

TABLE I. LANGUAGE AND VISION MODELS EVALUATED IN THE OCR-LLM AND VLM PIPELINES

Approach	Model	Size	Deployment
OCR-LLM	Qwen 2.5	3B	Local
OCR-LLM	Qwen 3.5	9B	Local
OCR-LLM	Qwen 2.5	14B	Local
OCR-LLM	Qwen 3.8	27B	Local
OCR-LLM	DeepSeek-R1	8B	Local
OCR-LLM	DeepSeek-R1	14B	Local
OCR-LLM	Llama 3.2	3B	Local
OCR-LLM	Llama 3.1	8B	Local
VLM	Qwen 2.5-VL	7B	Local
VLM	Gemini 3.5 Flash-Lite	-	API

identify Turkish and English nutrient labels and associate them with nearby numerical values within a label-anchored search window. To reduce ambiguity between related fields, specific matching rules were applied, including ordered handling of fat and saturated-fat labels. Extracted values were additionally checked against predefined per-100 g plausibility bounds before being accepted. The same matching rules, search procedure, and plausibility checks were applied identically to all OCR outputs.

The parser extracted eight nutritional fields: energy, fat, saturated fat, carbohydrates, sugars, protein, salt, and fiber. Energy values reported only in kJ were converted to kcal, and serving-based values were converted to a per-100 g or per-100 ml basis when the serving size could be identified. Fields that could not be identified reliably were treated as missing. No LLM-based correction or semantic post-processing was applied in the OCR-only pipeline.

2) OCR-LLM Based Extraction: In the OCR-LLM pipeline, the textual outputs produced by the OCR systems were further processed using locally deployed language models. Models from three different families, Qwen, DeepSeek-R1, and Llama, were evaluated at different parameter scales. The language and vision models used in the study are summarized in Table I.

The selected models cover a broad range of parameter sizes, from compact 3B models to larger models with up to 27B parameters. Multiple scales were evaluated within the Qwen and DeepSeek-R1 groups, while Llama models provided additional comparisons at the 3B and 8B scales. This setup makes it possible to examine whether increasing model size improves structured nutritional information extraction and whether models from different families show different performance at similar parameter scales.

Each local LLM received OCR-derived textual input and was instructed to organize the extracted information into the same eight nutritional fields: energy, fat, saturated fat, carbohydrates, sugars, protein, salt, and fiber. The same structured output format was used across all models to support a consistent comparison. The resulting nutritional values were subsequently used for extraction performance evaluation and Nutri-Score calculation.

3) VLM-Based Extraction: For the VLM-based approach, nutritional information was extracted directly from the product images without a separate OCR stage. Two vision-language models were evaluated: Gemini 3.5 Flash-Lite, accessed through an API, and the locally deployed Qwen2.5-VL 7B. The evaluated models and their deployment settings are summarized in Table I.

Both VLMs received the product images directly and were instructed to identify the same eight nutritional fields: energy, fat, saturated fat, carbohydrates, sugars, protein, salt, and fiber. By processing the images directly, the models could use both textual content and visual layout when associating nutrient names with their corresponding values.

For the locally deployed VLM, images were first processed at their original resolution. If an input request exceeded the model's available context window, the image was resized while preserving its aspect ratio, first limiting the longest side to 1280 pixels and, if necessary, to 1024 pixels, after which inference was repeated. The context window was set to 8192 tokens. This adaptive resizing procedure was applied only when required by the local model.

Including both an API-based and a locally deployed VLM allowed direct image-based extraction to be evaluated under two different deployment settings. The extracted nutritional values were subsequently used for performance evaluation and Nutri-Score calculation.

The two VLMs were evaluated as representative direct image-based extraction approaches under different deployment settings. They differ in model size, training, and deployment. Therefore, the results should be interpreted only as a comparison of these two specific models.

C. Nutri-Score Calculation

Nutri-Score was calculated using the updated 2023 algorithm. Nutritional values were evaluated per 100 g or 100 ml, depending on the product type. For general food products, the negative component (N) was calculated from energy, sugars, saturated fat, and salt, while the positive component (P) was based on fiber, protein, and the proportion of fruits, vegetables, and legumes. The final numerical score was obtained according to the Nutri-Score 2023 calculation rules [21]. The extraction pipelines in this study focused on the nutritional information table and did not extract the proportion of fruits, vegetables, and legumes (FVL) from ingredient lists. Therefore, the FVL proportion was set to 0% for all products during Nutri-Score calculation.

For general foods, numerical scores were mapped to Nutri-Score classes as follows: A for scores ≤ 0, B for scores from 1 to 2, C for scores from 3 to 10, D for scores from 11 to 18, and E for scores ≥ 19. Category-specific rules were applied where required for beverages and for fats, oils, nuts, and seeds, as well as for products subject to specific scoring rules [21]. The product categories required for these category-specific Nutri-Score rules were obtained from the product category metadata provided by Open Food Facts and were not inferred by the OCR, LLM, or VLM extraction pipelines.

Nutri-Score was calculated from the nutritional values produced by each extraction pipeline. For downstream classification evaluation, the Nutri-Score grades available in Open Food Facts were used as the reference labels. The resulting scores and A-E classes were then compared to determine how errors in nutritional information extraction affected the final nutritional assessment. When an extracted nutritional value was missing, it was treated as zero during the Nutri-Score calculation. Of the 1,000 products in the extraction benchmark, 693 had a reference Nutri-Score grade available in Open Food Facts and were therefore eligible for the downstream Nutri-Score evaluation. The reference-grade distribution was A: 73, B: 80, C: 144, D: 175, and E: 221. The remaining 307 products comprised 283 products with an unknown grade and 24 products for which Nutri-Score was not applicable.

D. Evaluation Metrics

Extraction performance was evaluated at the field level using fixed absolute-error tolerances. Predictions were considered correct when the error was within ± 5 kcal/100 g for energy and ± 0.5 g/100 g for the remaining nutritional fields. Missing predictions were treated as incorrect when calculating field-level accuracy. The evaluation did not separately distinguish between a nutrient that was genuinely absent from the visible label and a nutrient that the extraction pipeline failed to recover. When a reference value was available in Open Food Facts but the pipeline returned no value for that field, the prediction was counted as incorrect.

To assess the impact of extraction errors on Nutri-Score, the class calculated from each method was compared with the corresponding reference class. Overall accuracy represents the proportion of products assigned to the correct A-E class, while class-specific accuracy represents the proportion of correctly classified products within each reference class. Confusion matrices were used to examine how misclassified products were distributed across the Nutri-Score classes.

E. Experimental Setup

All local experiments were conducted on a system equipped with an AMD Ryzen 9 8940HX processor, an NVIDIA GeForce RTX 5070 Laptop GPU, and 32 GB DDR5 memory. The OCR systems, all locally deployed LLMs, and Qwen2.5-VL 7B were executed on this system, whereas Gemini 3.5 Flash-Lite was accessed through its API and therefore did not use the local hardware for inference.

The local language and vision models were deployed through Ollama. Detailed reproducibility information, including the exact model tags, model identifiers, parameter counts, context lengths, and quantization settings, is provided in the supplementary reproducibility file [22]. All evaluated local models used Q4_K_M quantization. For the local OCR-LLM experiments, the temperature was explicitly set to 0.0 to reduce stochasticity and promote consistent structured outputs. No explicit overrides were applied to top-p, top-k, seed, or maximum generation length; therefore, the corresponding model defaults were used. Thinking mode was disabled, and structured JSON output was requested for all local OCR-LLM models. Each model-input pair was evaluated using a single inference run. Repeated inference runs and run-to-run variance estimation were not performed.

The outputs generated by all pipelines were stored for subsequent performance evaluation and Nutri-Score calculation. The prompts were developed during a preliminary study using a separate small set of samples that was not included in the final evaluation dataset. The final prompt formulation was fixed before evaluation on the 1,000-image benchmark. To support reproducibility and facilitate comparison with future studies, the prompts used in the experiments are publicly available in our GitHub repository [22].

IV. Results and Discussion

A. Nutritional Information Extraction Performance

Figure 3 compares the field-level accuracy of the three OCR-based extraction pipelines. The outputs of GLM-OCR, GOT-OCR, and Paddle-OCR were processed using the same rule-based regex parser. GLM-OCR achieved the highest accuracy across all eight fields, with particularly strong performance for saturated fat, protein, salt, and carbohydrates.

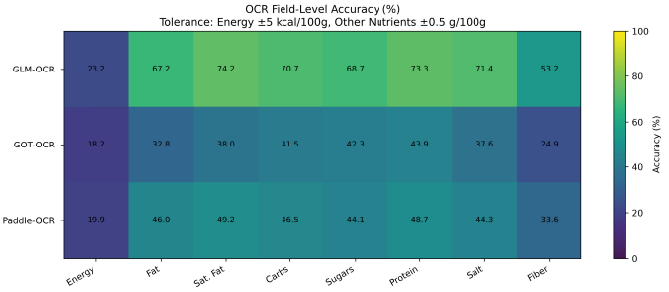


Fig. 3. Field-level accuracy of the OCR-based extraction pipelines

Figure 4 shows the field-level extraction accuracy of the OCR-LLM pipelines. Across nearly all model configurations, GLM-OCR produced substantially higher accuracy than GOT-OCR and Paddle-OCR, indicating that OCR quality strongly affects downstream LLM-based extraction. Qwen3.8-27B with GLM-OCR achieved the strongest overall performance, reaching over 79% accuracy for most nutritional fields and 83.4% for both fat and saturated fat. Accuracy generally increased with model size, particularly within the Qwen and DeepSeek model families. However, the performance differences between OCR sources were often larger than those between model sizes. Fiber remained the most difficult field across most configurations, with noticeably lower accuracy than the other nutritional fields.

B. Effect of Local LLM Integration

Figure 5 compares the field-level extraction accuracy of the best-performing OCR-LLM pipelines and the two VLM-based approaches. Gemini 3.5 Flash-Lite achieved the highest accuracy across all eight nutritional fields. Qwen2.5-VL 7B

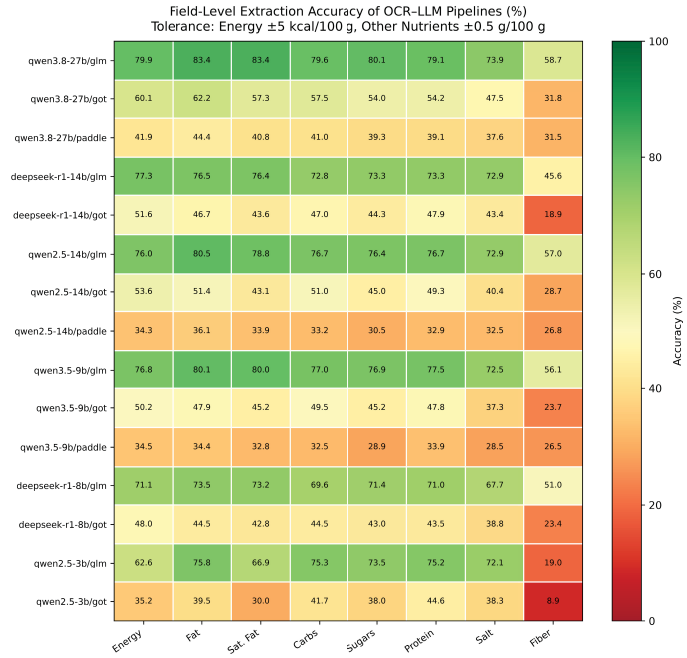


Fig. 4. Field-level extraction accuracy of the OCR-LLM pipelines

also showed strong performance, outperforming the best OCR-LLM configuration in seven of the eight fields, with energy being the only exception.

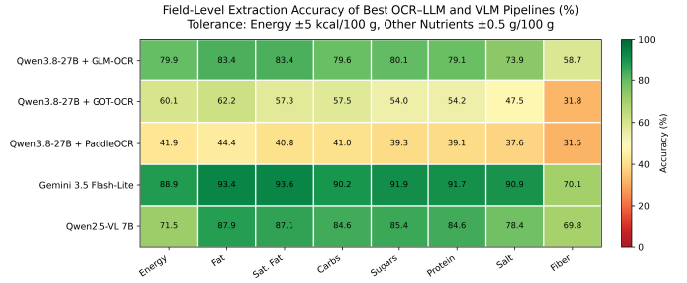


Fig. 5. Field-level accuracy of the best OCR-LLM and VLM pipelines under fixed error tolerances

Compared with the OCR-only results in Fig. 3, integrating Qwen3.8-27B with GLM-OCR substantially improved extraction accuracy. The largest improvement was observed for energy, increasing from 23.2% with the rule-based OCR pipeline to 79.9% with the OCR-LLM pipeline. Improvements were also observed for fat, saturated fat, carbohydrates, sugars, protein, salt, and fiber. These results indicate that LLM-based post-processing can substantially improve the interpretation of OCR output, although the quality of the underlying OCR input remains important. Fiber showed lower extraction accuracy than most other nutritional fields. One possible explanation is that fiber may not be consistently reported on all product labels, which can increase the frequency of missing predictions. Since missing predictions were treated as incorrect in the field-level evaluation, this may have contributed to the lower observed accuracy for fiber.

C. Impact on Nutri-Score Calculation

To evaluate the downstream effect of nutritional information extraction errors, Nutri-Score was calculated using the outputs of the best-performing OCR-LLM configuration, Qwen3.8-27B with GLM-OCR, and the best-performing VLM, Gemini 3.5 Flash-Lite. The calculated classes were compared with the reference Nutri-Score classes. When an extracted nutritional field was missing, its value was treated as zero during the Nutri-Score calculation. Products with pipeline-level processing errors, for which no model output was available, were excluded from this comparison.

Table II summarizes the overall and class-specific Nutri-Score classification accuracy. Qwen3.8-27B with GLM-OCR correctly classified 540 of 688 evaluated products, corresponding to an overall accuracy of 78.5%. Gemini 3.5 Flash-Lite correctly classified 618 of 693 products and achieved a higher overall accuracy of 89.2%. Accuracy for Qwen3.8-27B + GLM-OCR is reported over 688 successfully processed products. Five pipeline-level failures were excluded, while Gemini 3.5 Flash-Lite produced evaluable outputs for all 693 products.

TABLE II. NUTRI-SCORE CLASSIFICATION ACCURACY. MISSING NUTRITIONAL VALUES WERE TREATED AS ZERO

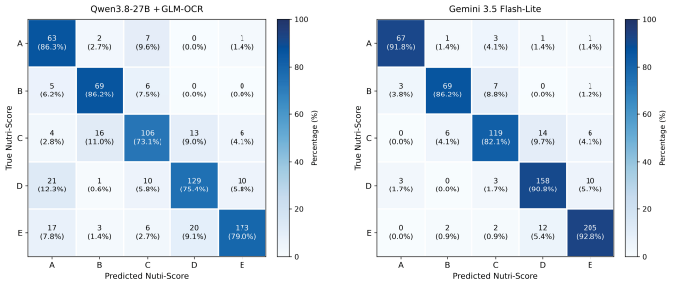
Metric	Qwen3.8-27B + GLM-OCR	Gemini 3.5 Flash-Lite
Accuracy	78.5%	89.2%
Class A	86.3%	91.8%
Class B	86.2%	86.2%
Class C	73.1%	82.1%
Class D	75.4%	90.8%
Class E	79.0%	92.8%

The class-specific results show that Gemini achieved higher accuracy for four of the five Nutri-Score classes. The two approaches showed the same accuracy for class B at 86.2%, while the largest differences were observed for classes D and E. Gemini reached 90.8% accuracy for class D and 92.8% for class E, compared with 75.4% and 79.0%, respectively, for Qwen3.8-27B with GLM-OCR. Class C was the most difficult category for both approaches, with accuracies of 73.1% and 82.1%, respectively.

Figure 6 provides a more detailed view of these classification errors. For Qwen3.8-27B with GLM-OCR, misclassifications were more widely distributed across the classes. Among reference class C products, 11.0% were predicted as B and 9.0% as D. More substantial class shifts were also observed for reference class D, where 12.3% of the products were predicted as class A. Similarly, 7.8% of reference class E products were classified as A.

In comparison, Gemini produced a stronger concentration along the main diagonal of the confusion matrix. For reference class D, 90.8% of products were correctly classified, while only 1.7% were predicted as class A. For class E, 92.8% were correctly classified and no products were assigned to class

A. Most of the remaining Gemini errors occurred between neighboring Nutri-Score classes, such as C-D and D-E.



(a) Qwen3.8-27B + GLM-OCR (b) Gemini 3.5 Flash-Lite

Fig. 6. Nutri-Score confusion matrices. Rows show reference classes and columns show predicted classes. Pipeline-error records were excluded.

Overall, the results show that errors in nutritional information extraction can propagate to the final Nutri-Score classification, but their effect differs between the two approaches. Gemini 3.5 Flash-Lite not only achieved higher field-level extraction accuracy, but also preserved the final Nutri-Score class more reliably. The difference was especially pronounced for the lower nutritional-quality classes D and E, where the OCR-LLM pipeline produced more cross-class errors. These findings indicate that improvements in structured nutritional information extraction can directly translate into more reliable downstream nutritional assessment.

V. Conclusion

This study compared OCR-based, OCR-LLM, and VLM-based approaches for extracting structured nutritional information from Turkish food package images. Regarding RQ1, the results showed clear performance differences between the evaluated approaches. Among the evaluated configurations, Gemini 3.5 Flash-Lite achieved the highest overall field-level extraction accuracy, while Qwen3.8-27B with GLM-OCR was the best-performing OCR-LLM configuration. OCR quality also had a strong influence on downstream performance, and fiber remained one of the most challenging nutritional fields across the evaluated methods.

Regarding RQ2, integrating a locally deployed LLM with OCR substantially improved structured nutritional information extraction compared with rule-based processing of OCR output alone. The improvement was particularly evident when high-quality OCR output was used, indicating that the effectiveness of the OCR-LLM approach depends on both text recognition quality and semantic interpretation.

Regarding RQ3, errors in extracted nutritional information were found to propagate to the final Nutri-Score classification. Gemini 3.5 Flash-Lite achieved an overall Nutri-Score classification accuracy of 89.2%, compared with 78.5% for Qwen3.8-27B with GLM-OCR. These results indicate that improvements in field-level extraction can also lead to more reliable downstream nutritional assessment.

Future work will address limitations related to both the dataset and practical deployment. Images obtained from Open Food Facts vary considerably in resolution and quality, and some nutritional labels are difficult to read. Reference metadata is also incomplete for a subset of the products. Treating missing extracted nutritional values as zero may introduce bias in the downstream Nutri-Score calculation, particularly when a missing value reflects an extraction failure rather than a true zero value. This represents an additional limitation of the present evaluation. Therefore, a purpose-built dataset with more consistent image quality and complete reference annotations is planned.

A further limitation is that the fruit, vegetable, and legume (FVL) proportion was not extracted from ingredient lists and was assumed to be 0% for all products during Nutri-Score calculation. This assumption may affect the calculated scores of products with substantial FVL content. Therefore, future work will extend the extraction pipeline to ingredient-list information so that FVL content can be incorporated into the Nutri-Score calculation. The present study did not include commercial OCR services as external baselines and focused on the evaluated locally deployable OCR systems. Future work will extend the comparison to commercial OCR services to assess their relative extraction performance under the same dataset and evaluation protocol.

Future studies will also examine the trade-off between extraction accuracy, inference time, and deployment constraints. Although VLMs provided strong extraction performance, image-based inference and API-based access may introduce additional computational and usage limitations for large-scale processing. Locally deployed LLMs may therefore offer a practical alternative in scenarios where efficiency and deployment flexibility are important. In particular, LLMs specifically trained or fine-tuned for nutritional information extraction will be investigated to determine whether they can provide a better balance between accuracy and computational efficiency.

In addition to Nutri-Score, future work will consider the NOVA food classification system to investigate how errors in automatically extracted food information affect the assessment of food processing level. This will allow the impact of LLM- and VLM-based extraction methods to be examined across different types of downstream food assessment.

References

[1] I. Ikonen, F. Sotgiu, A. Aydinli, and P. W. Verlegh, "Consumer effects of front-of-package nutrition labeling: An interdisciplinary meta-analysis," *Journal of the academy of marketing science*, vol. 48, no. 3, pp. 360–383, 2020.

[2] G. Kim, T. Hong, M. Yim, J. Nam, J. Park, J. Yim, W. Hwang, S. Yun, D. Han, and S. Park, "Ocr-free document understanding transformer," in *European Conference on Computer Vision*. Springer, 2022, pp. 498–517.

[3] F. Y. Assiri, M. D. Alahmadi, M. A. Almuashi, and A. M. Almansour, "Extract nutritional information from bilingual food labels using large language models," *Journal of Imaging*, vol. 11, no. 8, p. 271, 2025.

[4] C. Romero Ferreiro, D. Lora Pablos, and A. Gómez de la Cámara, "Two dimensions of nutritional value: Nutri-score and nova," *Nutrients*, vol. 13, no. 8, p. 2783, 2021.

[5] D. Van Strien, K. Beelen, M. Ardanuy, K. Hosseini, B. McGillivray, and G. Colavizza, "Assessing the impact of ocr quality on downstream nlp tasks," 2020.

[6] P. Ma, Y. Wu, N. Yu, X. Jia, Y. He, Y. Zhang, M. Backes, Q. Wang, and C.-I. Wei, "Integrating vision-language models for accelerated high-throughput nutrition screening," *Advanced Science*, vol. 11, no. 34, p. 2403578, 2024.

[7] V. Perot, K. Kang, F. Luisier, G. Su, X. Sun, R. S. Boppana, Z. Wang, Z. Wang, J. Mu, H. Zhang *et al.*, "Lmdx: Language model-based document information extraction and localization," in *Findings of the Association for Computational Linguistics: ACL 2024*, 2024, pp. 15 140–15 168.

[8] J. Dagdelen, A. Dunn, S. Lee, N. Walker, A. S. Rosen, G. Ceder, K. A. Persson, and A. Jain, "Structured information extraction from scientific text with large language models," *Nature communications*, vol. 15, no. 1, p. 1418, 2024.

[9] K. Lee, M. Joshi, I. R. Turc, H. Hu, F. Liu, J. M. Eisenschlos, U. Khandelwal, P. Shaw, M.-W. Chang, and K. Toutanova, "Pix2struct: Screenshot parsing as pretraining for visual language understanding," in *International Conference on Machine Learning*. PMLR, 2023, pp. 18 893–18 912.

[10] T. Silva, T. Carvalho, V. Filipe, L. Gonçalves, and A. Sousa, "Nutritional insight: Using ocr to decode food labels for better health," in *2024 international conference on graphics and interaction (ICGI)*. IEEE, 2024, pp. 1–8.

[11] B. Merz, E. Temme, H. Alexiou, J. W. J. Beulens, A. E. Buyken, T. Bohn, P. Ducrot, M.-N. Falquet, M. G. Solano, H. Haidar *et al.*, "Nutri-score 2023 update," *Nature Food*, vol. 5, no. 2, pp. 102–110, 2024.

[12] S. Kaushal, D. K. Tammineni, P. Rana, M. Sharma, K. Sridhar, and H.-H. Chen, "Computer vision and deep learning-based approaches for detection of food nutrients/nutrition: New insights and advances," *Trends in Food Science & Technology*, vol. 146, p. 104408, 2024.

[13] P. Ma, Z. Zhang, Y. Li, N. Yu, J. Sheng, H. K. McGinty, Q. Wang, and J. K. Ahuja, "Deep learning accurately predicts food categories and nutrients based on ingredient statements," *Food Chemistry*, vol. 391, p. 133243, 2022.

[14] A. W. Rosyadi, S. Ma'shumah, M. Q. Zaman, M. R. Fajar *et al.*, "Ingredients identification through label scanning using paddleocr and chatgpt for information retrieval," *J. RESTI*, vol. 8, no. 6, pp. 758–767, 2024.

[15] Q. Zhang, B. Wang, V. S.-J. Huang, J. Zhang, Z. Wang, H. Liang, C. He, and W. Zhang, "Document parsing unveiled: Techniques, challenges, and prospects for structured information extraction," *arXiv preprint arXiv:2410.21169*, 2024.

[16] A. Hu, H. Xu, J. Ye, M. Yan, L. Zhang, B. Zhang, J. Zhang, Q. Jin, F. Huang, and J. Zhou, "mplug-docowl 1.5: Unified structure learning for ocr-free document understanding," in *Findings of the Association for Computational Linguistics: EMNLP 2024*, 2024, pp. 3096–3120.

[17] K. Khanchandani, A. Thakur, A. Shetty, C. Reddy, and R. Behera, "Automated invoice data extraction: Using llm and ocr," *arXiv preprint arXiv:2511.05547*, 2025.

[18] H. Shen, G. Li, J. Zhong, and Y. Zhou, "Ldp: Generalizing to multilingual visual information extraction by language decoupled pretraining," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 7, 2025, pp. 6805–6813.

[19] S. Hercberg, M. Touvier, and J. Salas-Salvado, "The nutri-score nutrition label," 2021.

[20] D. L. van der Bend, M. van Eijsden, M. H. van Roost, K. de Graaf, and A. J. Roodenburg, "The nutri-score algorithm: Evaluation of its validation process," *Frontiers in Nutrition*, vol. 9, p. 974003, 2022.

[21] E. Hafner and I. Pravst, "A systematic assessment of the revised nutri-score algorithm: Potentials for the implementation of front-of-package nutrition labeling across europe," *Food Frontiers*, vol. 5, no. 3, pp. 947–963, 2024.

[22] C. Onbasioglu, "nutriscore-calculation-prompts," https://github.com/cagatay384/nutriscore-calculation-prompts, 2026, gitHub repository.