Human Factors in Artificial Intelligence and Social Computing, Vol. 203, 2026, 41–50
https://doi.org/10.54941/ahfe1007316

Multi-Source Food Names Mapping Using OpenAI Vision, Manual Dictionary and Fuzzy Matching Techniques

Shyam Bhetuwal[1], Lauri Koivunen[1], Rehan Khalil[1], Sanna Koskimäki[2], Hanna Lähde[2], Veera Houttu[2], Kirsi Laitinen[2,3], and Tuomas Mäkilä[1]

[1]Department of Computing, University of Turku, Finland
[2]Nutrition and Food Research Center, University of Turku, Finland
[3]Institute of Biomedicine, University of Turku, Finland

ABSTRACT

Accurate harmonization of food names across heterogeneous and multilingual datasets remains a major challenge in food informatics, dietary assessment systems, and data-driven public health research. Modern AI-based food recognition models such as LogMeal, FoodSAM, and OpenAI Vision can identify multiple components within complex dishes, but they frequently produce inconsistent, culturally specific, and multilingual labels. These inconsistencies complicate downstream tasks including nutritional analysis and cross-dataset integration. In this study, we evaluated practical methods for mapping AI-generated food component names to Finnish menu-based ground truth in a real-world restaurant setting. We collected 320 meal images using an integrated camera–scale system; 167 images containing multi-component dishes were selected for detailed evaluation against Finnish lunch-line menu labels. We compared (i) a segment-aware, menu-constrained mapping approach that uses LogMeal segmentations and prompts OpenAI Vision to select the best-matching item from the daily menu for each segment, and (ii) a hybrid manually curated canonical dictionary and fuzzy string matching pipeline applied separately to labels from different AI sources. Mapping performance is measured using precision, recall, and F1-score. The segment-aware OpenAI Vision approach achieved the best overall results (Precision = 0.90, Recall = 0.70, F1 = 0.79), while the hybrid dictionary+fuzzy method also improved consistency over direct label matching. These results indicate that menu-aware segment-level reasoning and lightweight lexical normalization are effective for food-name harmonization and can support scalable dietary monitoring and menu analytics.

Keywords: Name mapping, Fuzzy matching, Entity reconciliation, Data integration, Name standardization, String similarity, Multilingual datasets

INTRODUCTION

Food recognition systems are increasingly used in dietary assessment, food informatics, and public-health analytics. In practical deployments, however, the main challenge is often not detecting that a dish contains multiple components, but *harmonizing* the produced labels into a consistent vocabulary that supports downstream tasks such as nutrient computation, menu-level reporting, and cross-dataset integration. Modern recognition APIs and vision–language models frequently output free-text food names

Received February 6, 2026; Revised April 8, 2026; Accepted April 22, 2026; Available online July 20, 2026

that vary across languages, regional terminology, and levels of specificity (e.g., "chorizo" vs. "sausage"), which makes direct comparison to structured menu labels difficult, especially for multi-component meals.

This paper studies the problem of mapping heterogeneous AI-generated food component names to Finnish menu item labels used as ground truth in a real-world restaurant setting. We used images collected with an integrated camera–scale system and evaluated name mapping on 167 multicomponent meal images. The goal is to reconcile naming variability caused by multilingual outputs, culturally specific terms, and inconsistent granularity across models. Accurate name mapping is important for data integration, natural language processing, and knowledge graph construction. Current approaches struggle particularly with translation errors, typographical variations, and inconsistent labeling.

To address these challenges, we evaluated two practical mapping pipelines. First, we propose a segment-aware, menu-constrained approach in which **LogMeal**, a commercial deep-learning–based food recognition API that provides food labels and pixel-level component segmentation from images, is used to segment food components (LogMeal, 2024). For each segmented component, **OpenAI Vision**, a multimodal vision–language model capable of joint image understanding and textual reasoning, is prompted to select the best-matching item from the restaurant's daily menu for each segment (OpenAI, 2024). Second, we implement a hybrid baseline that normalizes AI outputs using a manually curated canonical dictionary and maps them to menu labels using fuzzy string matching. This paper presents these approaches in detail and evaluates their effectiveness in producing a standardized naming convention for complex, multi-component dishes.

This study addresses the following research questions:

RQ1: *How accurately can AI-generated food names be mapped to Finnish menu-based ground truth using automated versus hybrid methods?*
RQ2: *To what extent do manual linguistic resources improve the precision, recall and F1-score of food name harmonization?*
RQ3: *How do semantic reasoning and multi-component segmentation contribute to reducing false positives and disambiguating culturally or linguistically diverse food items?*

BACKGROUND

Food Recognition Systems

AI-based food recognition has seen rapid advancement with models such as LogMeal (LogMeal, 2024), FoodSAM (Lan et al., 2025), and OpenAI (OpenAI, 2024). These systems typically rely on trained data to detect and classify the multiple food items within an image (Hieronimus et al., 2024). Although they achieve high accuracy in object detection, they often produce inconsistent labels due to regional naming differences, multilingual terminology, and multi-component dishes. Previous reviews report that most food detection and classification systems rely on convolutional neural networks, and that performance decreases for complex multi-component dishes (Shonkoff et al., 2023). This inconsistency creates challenges for applications such as dietary tracking, public health analysis, and related applications.

Challenges in Harmonizing Food Labels

Harmonizing AI-detected food names with standardized menus is complicated by several factors:

- **Multilingual labels**: AI models may output names in different languages, requiring translation or mapping.
- **Regional variations**: Different algorithms may detect the same food component under different names (e.g., Chorizo → Sausage).
- **Multi-component dishes**: Dishes often contain multiple ingredients, making consistent labeling more complex.

Label inconsistency has commonly been addressed using techniques such as fuzzy string matching, manually curated dictionaries, and machine translation. While these methods can be effective in constrained settings, they often struggle with ambiguous, incomplete, or multilingual labels.

Knowledge Alignment and Entity Resolution Methods

Beyond food-specific studies, the literature on knowledge graph alignment and entity resolution offers insights for label harmonization.

Transformer-Gather and Fuzzy-Reconsider (Sharifi & Ahmadzadeh, 2025) combine transformer-based semantic embeddings with fuzzy matching for scalable, noise-resilient entity resolution. These methods are efficient but mainly limited to structured data. LLM-based embeddings (Xu et al., 2024) employ contrastive fine-tuning for entity resolution, offering semantically rich representations but requiring labeled data and significant computational resources. Adaptive Fuzzy String Matching (Kaufman & Klevs, 2022) integrates machine learning with human-in-the-loop corrections, improving accuracy when only a single messy identifier exists.

While these methods enhance alignment and semantic similarity detection, they often face limitations such as high computational cost, sensitivity to label quality, and poor performance on heterogeneous or multilingual datasets.

Text Matching and Multilingual Harmonization

General text-matching research provides further guidance. Jiang & Cai, (2024) categorize techniques into string-based, corpus-based, neural-network-based, graph-based and LLM-based approaches. These methods inform strategies for mapping AI-generated food labels, especially when dealing with multilingual outputs and inconsistent naming.

Despite advances in entity resolution and text matching, few studies have addressed the integration of multilingual AI-detected food labels with real-world menu systems. Existing food taxonomies and ontologies, such as FoodOn (Dooley et al., 2018), provide a valuable ontology, but using it directly for consumer-level, cross-lingual menu mapping remains non-trivial due to granularity and lexical variation.

Research Gap

Although food recognition models can detect multiple items in a single plate image, practical deployments require mapping heterogeneous and often multilingual free-text predictions to a structured target vocabulary such as daily menu item labels. Moreover, few studies evaluate menu-constrained, segment-level mapping pipelines in real restaurant conditions or compare vision–language model (VLM) reasoning against lightweight lexical baselines.

This study addresses this gap by evaluating two practical approaches for mapping AI-generated food component names to Finnish menu-based ground truth: (i) a segment-aware, menu-constrained mapping approach using LogMeal segmentations and OpenAI Vision, and (ii) a hybrid baseline using a canonical dictionary and fuzzy string matching. The focus is on improving semantic consistency under multilingual and culturally variable naming.

DATASET AND METHODOLOGY

This section describes the data collection setting, ground-truth construction, AI label sources, and the two evaluated name-mapping approaches. The work is implemented within the Flavoria Flex platform, which integrates image capture, inference orchestration, and menu-aware post-processing (Bhetuwal, 2025; Bhetuwal et al., 2025).

System Architecture and Implementation

The Flavoria Flex platform integrates hardware, software, and AI-based components into an end-to-end pipeline enabling automated food recognition, menu-aware correction, and nutritional analysis.

Food components are identified using AI based food and image detection APIs e.g. **LogMeal**, **OpenAI Vision**, and **FoodSAM**. These models provide heterogeneous predictions due to differences in training data and taxonomy, thereby improving the coverage and robustness of food recognition.

To standardize naming and address inconsistencies across AI predictions, the system incorporates daily restaurant menus as a structured knowledge base. Detected labels are mapped to Finnish-language menu items, yielding harmonized output suitable for quantitative evaluations and longitudinal analyses.

Data Collection

Data were collected at the Flavoria® research restaurant (Koivunen et al., 2020) using a fixed overhead camera setup integrated with a scale. Images represent real daily lunch-line meals (e.g., salads, main courses, and combined meals) and therefore include substantial variation in component combinations.

In total, 320 meal images were captured. For quantitative evaluation of component-level name mapping, 167 images were selected that contained multi-component meals (i.e., plates with multiple visually distinct items). Images excluded from the evaluation subset were primarily single-component plates or samples where component-level analysis was not applicable (e.g., heavily mixed dishes), or where required metadata was missing.

- Total images collected: 320
- Evaluation subset (multi-component images): 167
- Typical number of components per plate: 3–6

Ground Truth Menu Labels

Ground truth consists of Finnish lunch-line menu component labels indicating what each meal plate actually contained at the time of image capture. These labels were obtained directly from the lunch-line data associated with each captured meal image, i.e., the recorded set of menu components served on that plate. All ground-truth component names are drawn from the restaurant's daily lunch-line menu (the list of items offered on the corresponding day), which defines the target vocabulary for name mapping and evaluation.

AI Label Sources

Food component names were generated using three AI sources:

- **LogMeal**: provides detected food labels and pixel-level component segmentation.
- **FoodSAM**: provides food segmentation and associated labels.
- **OpenAI Vision**: provides image-based textual descriptions and food labels.

The three sources often output labels in different languages or with different naming conventions, motivating the need for mapping to the Finnish menu vocabulary. The goal is to map all detected names to standardized Finnish menu labels.

Preprocessing and Normalization

Before mapping, all predicted labels and menu labels were normalized using a consistent text preprocessing pipeline:

- lowercasing,
- trimming whitespace,
- removing punctuation and redundant separators.

This normalization reduces superficial string differences and improves the stability of downstream fuzzy matching.

Name Mapping Approaches

Approach 1: Segment-Aware, Menu-Constrained OpenAI Vision Mapping

This approach maps components at the segment level by combining (1) visual isolation from LogMeal segmentations and (2) the daily menu as a bounded candidate set.

Input: LogMeal based components with segmented image and the list of Finnish menu component labels available on that day.

Procedure:

1. **Segmentation overlay:** LogMeal component masks are rendered as a numbered overlay on the original image (segments labeled 1, 2, ...).
2. **Menu-constrained prompting:** OpenAI Vision (gpt-4o) is prompted with (a) the daily menu label list and (b) the numbered segmented image, and asked to select the best-matching menu item for each segment.
3. **Structured parsing:** The response is parsed into a segment→menu-label mapping, yielding a mapped Finnish label for each segment.

Output: A set of mapped Finnish menu labels per image (one label per segment).

Rationale: Segment-level visual grounding reduces ambiguity in multicomponent plates, while restricting outputs to the menu reduces open-vocabulary variability.

The following prompt illustrates how the OpenAI Vision language model was instructed to associate numbered food segments in the image with items from the provided menu.

```
Here is the provided menu: {menu_names}

An image is attached below. It contains clearly numbered
    segments of food items as 1, 2, 3, etc.
Using your visual reasoning capabilities, logically infer
    and match each segment number to an item listed in the
    provided menu.
If a segment cannot be confidently matched, choose the most
    likely item from the menu.
You must include all segment numbers in your output, even if
    unsure.
```

Approach 2: Hybrid Canonical Dictionary and Fuzzy Matching

This approach maps AI-predicted labels to menu labels using two stages: (i) canonicalization via a manually curated dictionary and (ii) fuzzy string matching to the closest menu label.

Input: AI-predicted labels from one source at a time (LogMeal, FoodSAM, and OpenAI Vision) and the menu label list.

Stage 1 – Canonicalization: A manual canonical dictionary maps frequent variants, multilingual terms, and culturally specific names to a canonical English form (e.g., *chorizo*, *bratwurst* → *sausage*). This step reduces vocabulary fragmentation across sources.

Stage 2 – Fuzzy mapping to menu: Canonicalized labels are matched to menu labels using a fuzzy string similarity score computed as the average of token-set ratio and weighted ratio. For each label, the menu item with the highest similarity score is selected. Matches with similarity below a tuned threshold are treated as unmatched.

The fuzzy similarity threshold was tuned on a held-out validation subset and set to **0.3**, which provided the best precision–recall trade-off for this dataset after dictionary-based normalization.

For example, AI-detected labels such as *chorizo*, *bratwurst*, or *longaniza* were first normalized to the category *sausage* using a manual dictionary and subsequently matched via fuzzy mapping to the closest best matched menu item *cheese-coated oven-baked sausage*.

Table 1: Canonical mapping dictionary example.

Canonical Key	Additions (Synonyms, Variants)
pasta	alfredo pasta, pasta with bolognese, pasta with cheese sauce, penne pasta, pennepasta, pesto pasta salad, baked feta pasta, pasta with vegetables, pasta with chicken, cooked pasta
chicken	chicken croquette, chicken curry, chicken roulade, grilled chicken, grilled chicken patty, grilled chicken breast, fried chicken, chicken wrapped in bacon and cheese, cooked chicken
potato	baked potatoes, baked sweet potatoes with spices, boiled potatoes, mashed potato, mashed potatoes, oven baked potato, potatoes with garlic and parsley, potato casserole, cooked potato, roasted potato, sweet potato, potato salad, potatoes (boiled), boiled potatoes, potatoes

Table 1 shows the example canonical dictionary which was used in hybrid approach for correction of names before analysis. For example, pasta was detected from AI models in different names based on the trained data. So the differently detected pasta names were converted to pasta for mapping and analysis.

Mapping Accuracy Comparison

We evaluated mapping quality by comparing the predicted mapped menu labels against the ground truth menu labels for each image. Performance is reported using precision, recall, and F1-score (Bishop & Nasrabadi 2006) computed over component labels. A predicted menu label is counted as a *true positive* if it exactly matches a ground-truth menu component label present on the same plate. A *false positive* occurs when a predicted label does not correspond to any ground-truth component for that plate, while a *false negative* corresponds to a ground-truth menu component that was not predicted. Because meals are multi-component, evaluation is performed at the label-set level per image and aggregated over the evaluation subset.

RESULTS

Food-name mapping performance was evaluated on the **167** multi-component meal images. Mapping quality is measured against the Finnish lunch-line ground-truth component labels using precision, recall, and F1-score.

Label Variability Across Sources

The ground truth dataset contained **58** unique menu component names. In contrast, the AI sources produced substantially more unique label strings due to open-vocabulary predictions, different naming conventions, and varying specificity levels. Table 2 summarizes the number of unique food-item name strings produced by each source.

For illustration, frequently detected labels included *carrot*, *mashed potato*, and *lettuce* (LogMeal); *lettuce*, *carrot*, *sauce*, and *potato* (FoodSAM); and a broader set of free-text variants such as *lettuce*, *carrots*, and *mashed potatoes* (OpenAI Vision). In the ground truth annotations, common menu items included *green salad*, *grated carrot*, *tomato*, and *mashed potatoes*. Overall, OpenAI Vision produced the highest lexical variability, motivating the need for mapping to a bounded menu vocabulary.

Table 2: Unique food item name strings produced by different sources.

Source	Unique Food Items
LogMeal API	98
FoodSAM	54
OpenAI Vision	264
Ground truth (Lunch Line)	58

Mapping Performance

Table 3 compares the two evaluated mapping approaches. **Approach 1** is the segment-aware OpenAI Vision mapping constrained to the daily menu using labeled segments. **Approach 2** (dictionary+fuzzy mapping) is applied separately to labels from each AI source.

Table 3: Performance comparison of food name mapping approaches.

Method	Precision	Recall	F1-Score
Approach 1: OpenAI Vision automated mapping using labeled segments (menu-constrained)	**0.90**	**0.70**	**0.79**
Approach 2: Dictionary+Fuzzy using OpenAI-based names	0.72	0.69	0.70
Approach 2: Dictionary+Fuzzy using FoodSAM-based names	0.70	0.67	0.69
Approach 2: Dictionary+Fuzzy using LogMeal-based names	0.75	0.56	0.63
Approach 2: Dictionary+Fuzzy using all 3 AI sources corrected names	**0.88**	**0.63**	**0.73**

Approach 1: Segment-Aware OpenAI Vision Mapping

The segment-aware OpenAI Vision mapping constrained to the daily menu achieved the best overall performance (Precision = 0.90, Recall = 0.70, F1 = 0.79). The improvement in recall relative to dictionary+fuzzy variants indicates that segment-level visual grounding and menu-constrained selection help disambiguate components in multi-item plates and reduce missed ground truth items.

Approach 2: Dictionary+Fuzzy Mapping

Using a dictionary with fuzzy matching on corrected names from all three AI sources achieved strong performance (Precision = 0.88, Recall = 0.63, F1 = 0.73). The same approach showed balanced results for OpenAI-based names (F1 = 0.70) and FoodSAM-based names (F1 = 0.69). For LogMeal-based names, it achieved higher precision (0.75) but lower recall, suggesting that LogMeal outputs remain relatively conservative or incomplete even after lexical normalization.

DISCUSSION

This study evaluated two practical approaches for food-name harmonization and addressed the research questions.

RQ1: Accuracy of automated versus hybrid methods. The segment-aware, menu-constrained OpenAI Vision mapping achieved the best overall performance, outperforming the dictionary+fuzzy approach applied to direct AI labels. This suggests that combining component-level visual grounding with constrained selection from the daily menu can align detected items to menu labels more reliably than relying on lexical similarity alone.

RQ2: Impact of manual linguistic resources. The canonical dictionary improved lexical consistency by normalizing multilingual variants and culturally specific terms before fuzzy matching, which mainly benefits precision. However, recall improvements were limited when upstream predictions were generic (e.g., *sauce*, *salad*) or when menu labels differed in granularity from predicted names, indicating that dictionaries are helpful but cannot fully replace menu-constrained semantic mapping.

RQ3: Role of semantic reasoning and multi-component segmentation. Processing food at the segment level reduced ambiguity in multi-component plates and helped disambiguate visually similar items. Restricting outputs to menu labels reduced open-vocabulary variation and lowered false positives, particularly in dishes with multiple components and culturally variable naming.

CONCLUSION

This paper addressed the problem of harmonizing heterogeneous AI-generated food component names to Finnish lunch-line menu labels in a real restaurant setting. We compared a segment-aware, menu-constrained OpenAI Vision mapping approach with a hybrid dictionary+fuzzy matching approach applied to direct AI outputs.

Overall, the segment-aware OpenAI Vision mapping achieved the best performance demonstrating the benefit of combining component-level visual grounding with selection from a bounded daily menu. The dictionary+fuzzy approach provided a lightweight alternative with moderate performance and improved lexical consistency through manual normalization, but remained sensitive to generic upstream labels and menu–prediction granularity differences.

These findings suggest that menu-aware, segment-level mapping is a practical and scalable foundation for improving food-name standardization in dietary monitoring and menu analytics systems.

Future Works

Future work could explore the approaches across multiple restaurants and cuisines, refine segmentation and tie-breaking for multi-component plates, and explore adaptive lexical resources (e.g., semi-automated dictionary expansion) to improve recall for generic or ambiguous component labels.

ACKNOWLEDGMENT

This research was supported by Business Finland (2022/31/2023). We gratefully acknowledge the Flavoria® multidisciplinary research platform and our colleagues at the Nutrition and Food Research Center of the University of Turku for their continued support.

REFERENCES

Bhetuwal, S. (2025), 'Comparing the accuracy and efficiency of existing AI based food detection tools'. ResearchGate, 10.13140/RG.2.2.16221.27366.

Bhetuwal, S. et al. (2025), Automated image recognition system for determining energy composition of meals by AI-powered detection and identification of food items – a study utilizing flavoria flex, *in* '2025 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT)', pp. 1–8.

Bishop, C. M. & Nasrabadi, N. M. (2006), *Pattern recognition and machine learning*, Vol. 4, Springer.

Dooley, D. M. et al. (2018), 'Foodon: A harmonized food ontology to increase global food traceability, quality control and data integration', *NPJ Science of Food* **2**(1), 23.

Hieronimus, B., Hammann, S. & Podszun, M. C. (2024), 'Can the AI tools ChatGPT and Bard generate energy, macro- and micro-nutrient sufficient meal plans for different dietary patterns?', *Nutrition Research* **128**, 105–114.

Jiang, P. & Cai, X. (2024), 'A survey of text-matching techniques', *Information* **15**(6), 332.

Kaufman, A. & Klevs, A. (2022), 'Adaptive fuzzy string matching: How to merge datasets with only one (messy) identifying field', *Political Analysis* **30**(4), 590–596.

Koivunen, L., Laato, S., Rauti, S., Naskali, J., Nissilä, P., Ojansivu, P., Mäkilä, T. & Norrdal, M. (2020), Increasing customer awareness on food waste at university cafeterias with a sensor-based intelligent self-service lunch line, *in* 'Conference Paper'.

Lan, X., Lyu, J., Jiang, H., Dong, K., Niu, Z., Zhang, Y. & Xue, J. (2025), 'Foodsam: Any food segmentation', *IEEE Transactions on Multimedia* **27**, 2795–2808.

LogMeal (2024), 'Logmeal api: AI-powered food recognition and nutritional analysis', https://logmeal.com/api/services/. Accessed: 2026-01-15.

OpenAI (2024), 'Gpt-4 technical report'. Available at: https://arxiv.org/abs/2303.08774.

Sharifi, M. & Ahmadzadeh, D. (2025), 'Transformer-gather, fuzzy-reconsider: A scalable hybrid framework for entity resolution'. arXiv:2509.17470.

Shonkoff, E., Cara, K. C., Pei, X. A., Chung, M., Kamath, S., Panetta, K. & Hennessy, E. (2023), 'Ai-based digital image dietary assessment methods compared to humans and ground truth: A systematic review', *Annals of Medicine* **55**(2), 2273497.

Xu, X., Foua, B. T., Wang, X., Gunasekaran, V. & Talburt, J. R. (2024), 'Leveraging large language models for efficient representation learning for entity resolution'.