

# Are Vision-Language Models Ready for Dietary Assessment? Exploring the Next Frontier in AI-Powered Food Image Recognition

## Supplementary Material

Table 1. Overview of the food *categories* and *subcategories* considered in this study for the labeling of the FoodNExTDB. The table defines the hierarchical structure used for food classification, where each main *category* encompasses multiple *subcategories*.

<i>Category</i>	<i>Subcategory</i>	<i>Category</i>	<i>Subcategory</i>
Cereals and Legumes	Rice	Beverages	Water
	Pasta		Infusions (Coffee, Tea, etc.)
	Bread		Soft Drinks
	Sliced Bread		Alcoholic Beverages
	Other Cereals (Quinoa, etc.)	Sugary Dairy Shakes	
	Legumes	Juices	
	Breakfast Cereals	Fast Food	Pizzas
	Biscuits		Burgers
	Whole Grains		Kebab
Vegetables and Fruits	Vegetables and Greens		Sandwiches
	Fruits		Battered and Fried Foods
	Mushrooms	Oils and Fats	Olive Oil
	Potatoes		Fats (Butter, etc.)
Processed Fruits	Sauces and Condiments		
Dairy and Plant-Based Drinks	Milk and Plant-Based Drinks		Avocado/Fatty Fruits
	Yogurt and Fresh Cheese	Creams	
	Cheeses	Snacks and Appetizers	Raw or Roasted Nuts
	Processed Dairy		Fried or Salted Nuts
Protein Sources	Poultry		Raw or Roasted Seeds
	Other Meats	Olives	
	Fish	Traditional Mediterranean Dishes	Paella
	Seafood		Spanish Stew ( <i>Cocido</i> )
	Eggs		Lentil Stew
	Cured Meats and Cold Cuts		Asturian Fabada
	Pâtes		Gazpacho/Salmorejo
	Offal		Russian Salad
	Processed Fish		Spanish Ratatouille ( <i>Pisto</i> )
Sweets and Pastries	Candies	Spanish Omelette	
	Chocolates and Pralines	Croquettes	
	Pastries and Baked Goods	Soup	
	Sweet Spreads	Purées and Creams	

Table 2. *Cooking styles* considered in this study.

<i>Cooking Styles</i>	
Fresh	Grilled
Fried	Stewed
Boiled/Steamed	Preserved
Oven-Baked	Fermented
None	

## Structured Prompt for Food Recognition

**Analyze the image and identify all visible foods.**

For each detected food, assign a **category**, **subcategory**, and **cooking style** based on the predefined lists. Most of the food products correspond to a Spanish/Mediterranean diet. Selection is restricted to the predefined options.

**Strictly follow the structured format. Do not provide explanations, assumptions, or additional text. Only return the structured response.**

### **Instructions:**

1. Detect all visible foods in the image.
2. Match each food to the closest **category** and **subcategory** from the list.
3. Identify the **cooking style** from the predefined options.
4. **Return the results in the following structured format:**

Food1: [*Category, Subcategory, Cooking Style*]

Food2: [*Category, Subcategory, Cooking Style*]

...

### **Example Output:**

Food1: [Cereals and Legumes, Bread, Oven-Baked]

Food2: [Protein Sources, Fish, Grilled]

Food3: [Vegetables, Fresh Vegetables and Greens, Fresh]

5. **If a food is detected but not listed, classify it as "Others" and include its name:**

FoodX: [Others, Detected Food Name, *Cooking Style*]

6. **If no foods from the predefined list are detected, return:**

"No food products detected."

7. **Foods that are partially visible or too far away should not be labeled, as they are not considered part of the current diet.**

Supplementary Prompt for Response Formatting (open-source models)

**Analyze the image and identify all visible food products and beverages.**

For each detected item, provide its **category**, **subcategory**, and **cooking style**.

**Predefined Cooking Styles:**

[Fresh, Fried, Boiled or Steamed, Oven-Baked, Grilled, Stewed, Preserved (Canned or Jarred), Fermented, None]

This image likely contains foods aligned with a **Spanish and Mediterranean diet**. **Consider traditional ingredients from both diets when identifying food items.**

## Final Post-Processing Prompt for Output Standardization

**Task:** You will receive a list of detected food products and beverages, each with an associated **category**, **subcategory**, and **cooking style**. Your task is to strictly format the output according to the following structure:

FoodX: [*Category, Subcategory, Cooking Style*]

### Processing Rules:

- Ensure that each detected food item follows this exact format:  
Food Name: [*Category, Subcategory, Cooking Style*]
- Use only the predefined **categories**, **subcategories**, and **cooking styles** from the provided tables.
- **Strictly follow the structured format. Do not provide any explanations, assumptions, or additional text.** Only return the structured response.

### Handling Unlisted Foods:

- If a detected food does not match any *subcategory* from the predefined list:
- Assign "Others" as the **category**.
- Use the detected food name as the **subcategory**.
- Assign an appropriate **cooking style** from the predefined list.

**If no relevant foods are detected, return exactly:**

"No food products detected."

### Example Input:

Chicken (Grilled, Protein)

Olive oil (Raw, Oils and Fats)

Food Product 2: [Meat, Sausage, Fried]

The image shows a pile of white, oval-shaped food items on a wooden surface.

### Expected Output:

Food1: [Protein Sources, Poultry, Grilled]

Food2: [Oils and Fats, Olive Oil, None]

Food3: [Beverages, Infusions (Coffee, Tea, etc.), None]

Food5: [Others, Unknown dish, Stewed]

### Final Requirement:

Ensure that all detected items are formatted correctly.

**Exclude any additional explanations, assumptions, or extra text|only return the structured response.**