

Additional file 1

Original Chinese prompts and four-shot demonstrations

An evaluation of large language models for panoramic radiographs

This file reproduces the request text used in the benchmark. The Chinese prompts below were sent unchanged to all seven models under both prompting conditions; the English translations are provided for review only. Field names in the JSON schema were sent in English. In the JSON field names, `wisdom_tooth` corresponds to the third-molar label.

S1. Original Chinese prompt for Dataset A (treatment and restoration)

这是一张牙科全景片。请只根据图像完成口腔修复/治疗相关评估。

请判断以下图像级标签是否存在：

1. `oral_implant`: 种植体
2. `crown`: 牙冠或冠修复体
3. `wisdom_tooth`: 智齿
4. `root_canal_therapy`: 根管治疗
5. `filling`: 充填体

同时请判断是否有缺失牙，并尽量给出缺失牙FDI编号。注意：全景片图像左侧对应患者右侧，FDI
↪ 1区在图像左上，2区在图像右上，3区在图像右下，4区在图像左下。

请严格输出一个 JSON 对象，不要输出 Markdown，不要解释 JSON 以外内容。字段必须为：

```
{
  "oral_implant": 0 或 1 或 null,
  "crown": 0 或 1 或 null,
  "wisdom_tooth": 0 或 1 或 null,
  "root_canal_therapy": 0 或 1 或 null,
  "filling": 0 或 1 或 null,
  "missing_tooth_fdi": " 用下划线连接的 FDI 编号，若无或无法判断则为 null",
  "brief_reason": " 一句话说明主要依据"
}
```

S2. English translation of the Dataset A prompt

This is a dental panoramic radiograph. Based only on the image, assess findings related to restoration and treatment.

Determine whether each of the following image-level labels is present:

1. `oral_implant`: dental implant
2. `crown`: crown or crown restoration
3. `wisdom_tooth`: third molar (wisdom tooth)
4. `root_canal_therapy`: root-canal treatment

5. filling: filling

Also determine whether any teeth are missing and, as far as possible, give the FDI numbers of the missing teeth. Note: the left side of the panoramic image corresponds to the patient's right side; FDI quadrant 1 is at the upper left of the image, quadrant 2 at the upper right, quadrant 3 at the lower right, and quadrant 4 at the lower left.

Output exactly one JSON object, without Markdown and without any explanation outside the JSON. The fields must be: **oral_implant**, **crown**, **wisdom_tooth**, **root_canal_therapy**, and **filling**, each 0, 1, or null; **missing_tooth_fdi**, the FDI numbers joined by underscores, or null if there are none or they cannot be determined; and **brief_reason**, one sentence stating the main basis.

S3. Original Chinese prompt for Dataset B (disease)

这是一张牙科全景片。请只根据图像完成口腔疾病相关评估。

请判断以下图像级标签是否存在：

1. **apical_periodontitis**: 根尖周炎。影像表现通常是在牙根根尖周围出现边界清晰或模糊的透射影，
→ 可能伴随深龋、大面积充填、根管治疗史、牙冠修复或根尖吸收。注意不要把颞孔、上颌窦、鼻腔、
→ 下颌管等正常解剖结构误判为根尖周炎。
2. **caries**: 龋坏。龋坏区域因牙体脱矿密度降低，常表现为牙冠咬合面、邻面、
→ 颈部或根面的局部低密度透射影。注意鉴别牙颈部正常低密度区、磨损、楔状缺损和投影重叠。

同时请尽量给出涉及牙齿的FDI编号。注意：全景片图像左侧对应患者右侧，FDI 1区在图像左上，
→ 2区在图像右上，3区在图像右下，4区在图像左下。

请严格输出一个 JSON 对象，不要输出 Markdown，不要解释 JSON 以外内容。字段必须为：

```
{
  "apical_periodontitis": 0 或 1 或 null,
  "apical_periodontitis_fdi": " 用下划线连接的 FDI 编号，若无或无法判断则为 null",
  "caries": 0 或 1 或 null,
  "caries_fdi": " 用下划线连接的 FDI 编号，若无或无法判断则为 null",
  "brief_reason": " 一句话说明主要依据"
}
```

S4. English translation of the Dataset B prompt

This is a dental panoramic radiograph. Based only on the image, assess findings related to oral disease. Determine whether each of the following image-level labels is present:

1. **apical_periodontitis**: apical periodontitis. The usual radiographic appearance is a well- or ill-defined radiolucency around the root apex, which may be accompanied by deep caries, a large filling, previous root-canal treatment, a crown restoration, or apical root resorption. Do not misinterpret normal anatomical structures such as the mental foramen, maxillary sinus, nasal cavity, or mandibular canal as apical periodontitis.
2. **caries**: caries. Because demineralisation lowers tooth density, a carious area usually appears as a focal radiolucency on the occlusal, proximal, cervical, or root surface. Distinguish it from the normal cervical low-density area, wear, wedge-shaped defects, and projection overlap.

Also give, as far as possible, the FDI numbers of the teeth involved. Note: the left side of the panoramic image corresponds to the patient’s right side; FDI quadrant 1 is at the upper left of the image, quadrant 2 at the upper right, quadrant 3 at the lower right, and quadrant 4 at the lower left.

Output exactly one JSON object, without Markdown and without any explanation outside the JSON. The fields must be: `apical_periodontitis` and `caries`, each 0, 1, or null; `apical_periodontitis_fdi` and `caries_fdi`, the FDI numbers joined by underscores, or null if there are none or they cannot be determined; and `brief_reason`, one sentence stating the main basis.

S5. Four-shot request structure

In the zero-shot condition, each request contained the dataset-specific prompt followed by the query image. In the four-shot condition, the following content was placed before the same prompt and query image. The wrapper text was sent in English, exactly as shown; the four demonstrations were taken from the same dataset as the query image and were presented in the fixed order listed in Table S1.

```
Use the following annotated examples as demonstrations.
Demonstration 1 image:
[demonstration image 1]
Demonstration 1 reference JSON: [reference JSON 1]
... (repeated for demonstrations 2-4) ...
Now assess the query image independently using the same annotation rules.
[dataset-specific Chinese prompt]
[query image]
```

In each reference JSON, `brief_reason` was set to “Reference annotation.”

S6. Fixed demonstration sets

Set 1 was used for the primary analysis and the repeated runs; Sets 2 and 3 were used only in the demonstration-set sensitivity analysis. Every set covered both binary states of each Dataset A label and all four combinations of apical-periodontitis and caries status in Dataset B. The same images and order were used for every model.

Table S1. Demonstration images in presentation order. Identifiers refer to the benchmark image identifiers.

Set	Dataset A	Dataset B
1	label1_046, label1_026, label1_011, label1_061	label2_102, label2_103, label2_101, label2_106
2	label1_015, label1_003, label1_023, label1_065	label2_104, label2_111, label2_107, label2_108
3	label1_027, label1_004, label1_099, label1_029	label2_105, label2_115, label2_162, label2_109

S7. Representative demonstrations (Set 1)

The final Set 1 demonstration image from each dataset is shown below with its reference JSON, as supplied in the four-shot requests. Images were resized for API requests as described in the Methods; the reference JSON is reproduced exactly as sent.

Dataset A, demonstration 4 (label1_061)



```
{"oral_implant": 0, "crown": 1, "wisdom_tooth": 1, "root_canal_therapy": 1, "filling": 1,  
  ↳ "missing_tooth_fdi": "14_24_27_36_45_46_47", "brief_reason": "Reference annotation."}
```

Dataset B, demonstration 4 (label2_106)



```
{"apical_periodontitis": 1, "apical_periodontitis_fdi": "13_14", "caries": 1, "caries_fdi":  
  ↳ "37", "brief_reason": "Reference annotation."}
```