# Supplementary Materials

Calibration, explainability and spatial uncertainty for YOLO-based detection in panoramic dental radiography

Hanan Alaskar and Nikos Nikolaou
Department of Physics and Astronomy, University College London (UCL), London, United Kingdom
*Correspondence:* hanan.alaskar.24@ucl.ac.uk

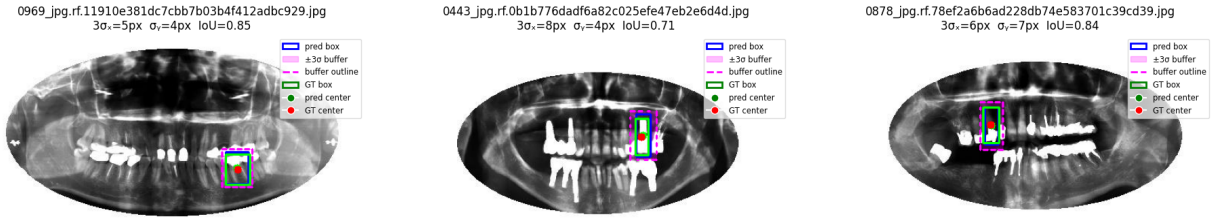
## Supplementary Tables and Figures

Table S1: Calibration performance of YOLOv11s at IoU thresholds 0.3, 0.5, and 0.7. Metrics include Adaptive Expected Calibration Error (ECE) and Negative Log-Likelihood (NLL) for pooled (overall) and per-class results. Best values per IoU/method are in bold.

| IoU | Method | Overall | | Cavity | | Implant | | Fillings | | Impacted Tooth | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | ECE | NLL | ECE | NLL | ECE | NLL | ECE | NLL | ECE | NLL |
| 0.3 | Uncal. | 0.0329 | 0.2653 | 0.0435 | 0.2279 | 0.0580 | 0.1897 | 0.0350 | 0.2915 | 0.0518 | 0.2963 |
| | Temp. | 0.0345 | 0.2594 | 0.0493 | 0.2143 | **0.0264** | 0.1764 | 0.0393 | 0.2897 | 0.0561 | 0.2898 |
| | Platt | **0.0325** | **0.2545** | **0.0235** | **0.2019** | 0.0313 | **0.1647** | 0.0404 | **0.2882** | 0.0625 | **0.2882** |
| 0.5 | Uncal. | 0.0333 | 0.2636 | 0.0459 | 0.2259 | 0.0552 | 0.1948 | 0.0337 | 0.2918 | 0.0447 | 0.2529 |
| | Temp. | 0.0394 | 0.2585 | 0.0515 | 0.2128 | **0.0264** | 0.1836 | 0.0394 | 0.2901 | 0.0456 | 0.2524 |
| | Platt | **0.0339** | **0.2536** | **0.0259** | **0.1990** | 0.0287 | **0.1726** | 0.0404 | **0.2887** | 0.0511 | **0.2519** |
| 0.7 | Uncal. | 0.0357 | 0.2686 | 0.0580 | 0.2100 | 0.0422 | 0.2114 | 0.0386 | 0.3022 | 0.0556 | 0.2314 |
| | Temp. | 0.0469 | 0.2643 | 0.0632 | 0.2010 | **0.0330** | 0.2059 | 0.0438 | 0.2991 | 0.0549 | 0.2313 |
| | Platt | **0.0316** | **0.2590** | **0.0209** | **0.1744** | 0.0365 | **0.2051** | 0.0390 | **0.2984** | **0.0436** | **0.2295** |

Table S2: Sample counts per class–case pair in the Pixel-Level Explanation Dataset. Each sample corresponds to one detection used for coverage and validity analysis. TP = True Positive, FP = False Positive, FN = False Negative.

| **Class** | **TP** | **FP** | **FN** | **Total** |
|---|---|---|---|---|
| Cavity | 26 | 35 | 17 | 78 |
| Impacted Tooth | 31 | 13 | 7 | 51 |
| Implant | 156 | 16 | 3 | 175 |
| Fillings | 330 | 129 | 42 | 501 |
| **All classes** | 543 | 193 | 69 | **805** |




(a) Cluster 0 example ($\sigma_x \approx 2.25$, $\sigma_y \approx 1.35$).

(b) Cluster 1 example ($\sigma_x \approx 2.9$, $\sigma_y \approx 1.5$).

(c) Cluster 2 example ($\sigma_x \approx 2.45$, $\sigma_y \approx 2.6$).

Figure S1: Representative detections from different uncertainty clusters, showing predicted boxes (solid blue), ground-truth boxes (solid green), and $\pm 3\sigma$ buffers (dashed purple). Smaller uncertainties correspond to confident, high-contrast detections, whereas larger ones occur for ambiguous or overlapping regions.