Original Paper

AI-Assisted Tooth Detection and Segmentation Workflow on Orthopantomograms Using the FDI Notation

DAN-ADRIAN LUTESCU[1], CLAUDIU CONSTANTIN MANOLE[2], RADU ILINCA[1], REINHARD CHUN WANG CHAU[3], ANDREEA CRISTIANA DIDILESCU[4]

[1]Department of Medical Informatics and Biostatistics, Faculty of Dentistry, "Carol Davila" University of Medicine and Pharmacy, Bucharest, Romania
[2]Department of Biophysics, Faculty of Dentistry, "Carol Davila" University of Medicine and Pharmacy, Bucharest, Romania
[3]Faculty of Dentistry, The University of Hong Kong, Hong Kong
[4]Department of Embryology and Microbiology, Faculty of Dentistry, "Carol Davila" University of Medicine and Pharmacy, Bucharest

ABSTRACT: Background: Automated tooth detection and segmentation on orthopantomograms (OPGs) is a foundational task for AI-based dental diagnostics. Most published studies have focused on subsets of the dentition or employed rectangular bounding boxes that inadequately capture overlapping dental structures. This study developed and evaluated a YOLO26s-seg instance segmentation model for polygon-based detection and FDI-based identification of all 32 permanent tooth classes on a publicly available panoramic dataset. Methodology: A convenience sample of 282 OPGs from the Tufts Dental Database was annotated with polygon-based FDI labels using an in-house browser-based platform. The dataset was partitioned at the patient level into training (n=197), validation (n=42), and test (n=43) subsets. A YOLO26s-seg model was trained using transfer learning at 1280×1280 resolution with laterality-preserving augmentation. Results: On the held-out test set, the model achieved box-level precision of 0.883 and recall of 0.892, mAP50 of 0.896 (95% CI: 0.865-0.923) for detection and 0.890 (95% CI: 0.858-0.919) for segmentation, and mAP50-95 of 0.620 and 0.556, respectively. Twenty-eight of 32 classes exceeded a box-level mAP50 of 0.80, with three of the four lowest-performing classes located in the mandibular left quadrant. Conclusions: YOLO26s-seg achieved strong mAP50 performance for instance segmentation and FDI identification of all 32 permanent tooth classes. The workflow provides a transparently reported research benchmark; clinical effectiveness was not evaluated.

KEYWORDS: *Orthopantomography, instance segmentation, deep learning, YOLO26, tooth detection, FDI numbering.*

Introduction

Panoramic radiography, or orthopantomography (OPG), is a widely used extraoral imaging modality that provides a single two-dimensional view of the dental arches, temporomandibular joints, and adjacent maxillofacial structures at a substantially lower radiation dose than cone-beam computed tomography [1,2].

Owing to its broad coverage, rapid acquisition, low cost, and wide availability, the OPG remains a routine first-line investigation for comprehensive dental assessment, including eruption status, partial edentulism, prosthodontic and orthodontic treatment planning, and evaluation of impacted teeth [1,3].

However, panoramic images are affected by geometric distortion, anatomical superimposition, ghost artifacts, and inhomogeneous contrast, which constrain manual interpretation and contribute to inter- and intra-examiner variability [2,3].

Deep learning models may provide reproducible and scalable assistance for OPG interpretation, with potential applications in standardized reporting and decision-support workflows [3-5].

Applications range from caries detection to recognition of dental restorations, impacted teeth, and periapical lesions [6,7].

Tooth identification in clinical practice relies on the FDI World Dental Federation two-digit notation used internationally for tooth localization and numbering [8-10], and automatic AI-based detection of individual teeth is a foundational step for downstream diagnostic tasks.

The challenge is non-trivial: a complete permanent dentition contains 32 morphologically distinct classes that frequently overlap on panoramic projections, exhibit substantial inter-patient variability, and may be partially obscured by restorations or prostheses, in addition to missing teeth in partially edentulous patients [2,9].

This is an open-access article distributed under the terms of a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Public License, which permits unrestricted use, adaptation, distribution and reproduction in any medium, non-commercially, provided the new creations are licensed under identical terms as the original work and the original work is properly cited.
10.12865/CHSJ.52.03.09

Early work relied on region-based detectors such as Faster R-CNN, achieving F1 scores of 0.96-0.99 for tooth detection on moderate-to-large panoramic datasets [8,11], while subsequent approaches introduced spatial regularization to improve robustness against missing teeth [9].

More recently, instance segmentation architectures have been applied for polygon-level tooth delineation on OPGs [2].

The YOLO family of one-stage detectors has emerged as particularly attractive for OPG analysis owing to its balance between accuracy and speed, with mAP50 values of 0.63-0.74 reported for multiclass detection on reduced panoramic datasets [12].

Despite these advances, publicly available well-annotated panoramic datasets remain scarce, and inter-annotator agreement is rarely formally reported, both of which directly influence model reliability [10,12].

The development of models evaluated on standardized, publicly accessible datasets remains an essential prerequisite for reproducible benchmarking [12,13].

Objective

This study aimed to develop and evaluate a YOLO26s-seg instance segmentation model for the polygon-based detection and FDI-based identification of all 32 permanent tooth classes on orthopantomograms from a standardized, publicly available dataset.

The contribution is not a novel architecture but a comprehensive evaluation including same-split model comparison, bootstrap confidence intervals, 5-fold cross-validation, inter-annotator agreement, and analysis of predictions assigned to absent FDI classes.

Methods

Dataset

Orthopantomographic images were sourced from the Tufts Dental Database (TDD), a de-identified multimodal panoramic radiographic dataset publicly available at http://tdd.ece.tufts.edu/.

The TDD contains 1,000 digital panoramic radiographs collected from patients at Tufts University School of Dental Medicine, acquired with OP100 Orthopantomograph (Instrumentarium Imaging/Kavo Kerr) and Planmeca Promax 2D (Henry Schein) radiographic units under automatic exposure control. The dataset contains one image per patient, making patient-level and image-level partitions equivalent.

The working subset comprised a convenience sample of 282 OPGs from the permanent-dentition fraction that met the eligibility criteria and were fully annotated within the available resources. Images were excluded for major radiographic artifacts or insufficient quality, incomplete arch coverage, predominantly deciduous or mixed dentition, complete edentulism, or the presence of supernumerary teeth that could not be mapped onto the 32-class FDI scheme. A per-reason numerical breakdown was not prospectively recorded; this is acknowledged as a limitation.

Annotation protocol

Polygon-based instance segmentation annotations were produced using an in-house, browser-based annotation platform developed as an internal research tool. The platform supported manual delineation of closed polygonal contours, FDI class label assignment, senior-annotator review, and export to the YOLO segmentation format. It is not a commercial product and no author derives commercial benefit from it. Each tooth was delineated by tracing its visible outer boundary, with labels following the FDI notation system (ISO 3950) across all 32 permanent tooth classes (11-18, 21-28, 31-38, 41-48).

Delineations were performed by dental specialists with clinical training in oral and maxillofacial radiology. The original TDD annotations were not used; all annotations were produced de novo to ensure a single, consistently applied protocol. All delineations were reviewed by a senior annotator prior to export.

The annotation set comprised 7,036 polygon-based instances with a mean of 25.0 annotations per image (range: 1-32).

Per-class counts ranged from 93 (tooth 18) to 268 (tooth 33). To quantify annotation reliability, a second dental specialist independently re-annotated 50 randomly selected OPGs (seed=42), blinded to the originals. Inter-annotator agreement was assessed as polygon intersection-over-union (IoU) between annotations matched by shared FDI label, quantifying boundary-delineation agreement rather than label agreement. A formal intra-rater study was not performed.

Dataset partitioning

The dataset was partitioned into training (n=197, 69.9%), validation (n=42, 14.9%), and test (n=43, 15.2%) subsets using a random split with a fixed seed (seed=42), with each patient

assigned to exactly one subset. Annotations were exported in YOLO segmentation format with polygon coordinates pre-normalized to image dimensions.

Model architecture and training

Model training used the YOLO26s-seg architecture (yolo26s-seg.pt checkpoint) implemented in the Ultralytics framework (version 8.4.51, Ultralytics Inc., Los Angeles, CA, USA). Transfer learning was applied by initializing from a pre-trained checkpoint. Training was distributed across two NVIDIA RTX PRO 6000 Blackwell GPUs (NVIDIA Corporation, Santa Clara, CA, USA) with a total batch size of 32, a maximum of 150 epochs, input size of 1280 × 1280 pixels, weight decay of 0.0005, and early stopping with patience of 30 epochs. The optimizer and learning-rate schedule were selected automatically by the framework ('auto' setting).

Data augmentation was constrained to laterality-preserving transformations. Horizontal and vertical flipping were disabled (fliplr=0.0; flipud=0.0) because horizontal flipping transposes left- and right-side FDI labels without exchanging them, and vertical flipping produces non-physiological orientations.

Mosaic augmentation (probability 0.5), mild rotation (±5°), and default photometric (HSV) and geometric (translation, scaling) augmentations were applied.

Two additional models (YOLOv8s-seg and YOLO11s-seg) were trained using the same data split, seed, and user-specified settings to enable a same-split architecture comparison.

Evaluation metrics

Performance was evaluated on the held-out test set (n=43) using Precision, Recall, mAP50 and mAP50-95 for both bounding box and segmentation mask tasks, macro-averaged across the 32 FDI classes. Box-level precision and recall were obtained at the confidence threshold maximizing mean F1 on the test set and are reported as descriptive curve-derived summaries.

Confidence intervals (95%) were estimated by non-parametric image-level bootstrap resampling (1,000 replicates). False positives with respect to absent tooth classes were quantified by comparing predicted FDI classes against the reference standard for each test image, using a confidence threshold of 0.25.

Cross-validation

Five-fold cross-validation was performed on the full dataset (282 OPGs, seed=42). In each fold, the model was trained on four folds with 15% withheld for internal validation, and evaluated on the remaining fold using the same training settings; the optimizer was selected automatically within each fold. The held-out test set was included in this partitioning; cross-validation therefore assesses robustness across alternative splits rather than providing independent replication.

Results

Dataset composition and annotation statistics

The final dataset comprised 282 OPGs with 7,036 polygon-based annotations across 32 FDI tooth classes. The mean number of annotated teeth per image was 25.0 (range: 1-32), with per-class instance counts ranging from 93 (tooth 18) to 268 (tooth 33). Third molars consistently showed the lowest representation. Per-class instance counts and the distribution of annotations per OPG are presented in Figure 1.

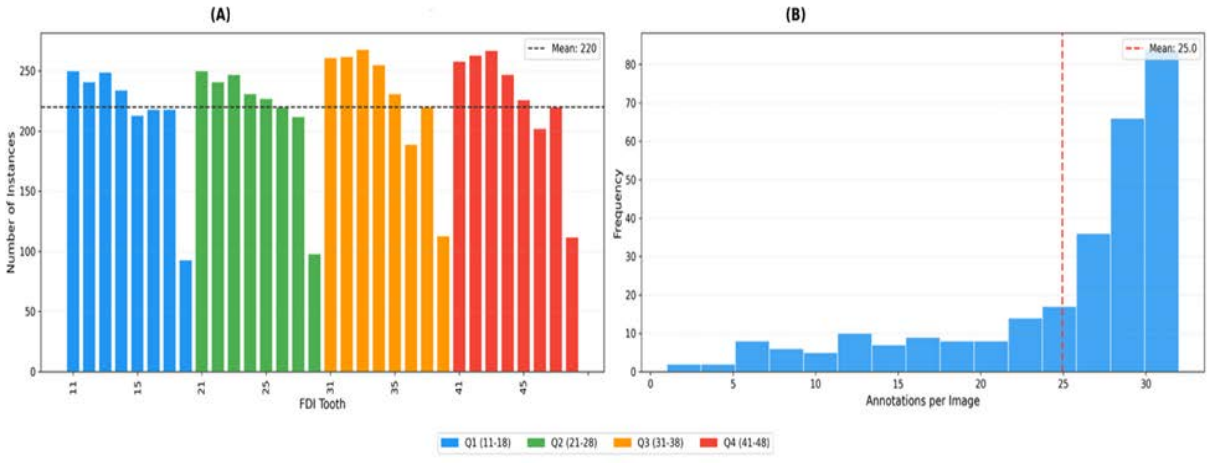


Figure 1. Dataset annotation statistics. (A) Per-class instance count across the 32 FDI tooth classes. (B) Distribution of the number of annotations per OPG.

10.12865/CHSJ.52.03.09

Inter-annotator agreement

Boundary-delineation agreement between the finalized reference annotations and an independent reannotation of 50 OPGs (1,260 annotation pairs matched by shared FDI label) yielded a mean polygon IoU of 0.863 (median 0.869, SD 0.053; range 0.582-1.000).

The lowest-IoU pairs were concentrated in mandibular posterior and third-molar regions.

These values indicate consistent boundary delineation across annotators, although variability in anatomically complex regions warrants consideration when interpreting per-class segmentation performance.

Global model performance

All metrics were computed exclusively on the held-out test set (n=43 OPGs, 1,044 reference-standard tooth instances).

Mean box-level precision and recall were 0.883 and 0.892, respectively. The mAP50 was 0.896 (95% CI: 0.865-0.923) for detection and 0.890 (95% CI: 0.858-0.919) for segmentation, with mAP50-95 of 0.620 (95% CI: 0.591-0.659) and 0.556 (95% CI: 0.530-0.595), respectively (Table 1, Figure 2A).

Table 1. Global performance metrics of the YOLO26s-seg model on the held-out test set, with 95% bootstrap confidence intervals (1,000 replicates).

Metric	Value	95% CI
Precision (box)	0.883	0.841-0.917
Recall (box)	0.892	0.865-0.928
mAP50 (box)	0.896	0.865-0.923
mAP50-95 (box)	0.620	0.591-0.659
mAP50 (seg)	0.890	0.858-0.919
mAP50-95 (seg)	0.556	0.530-0.595

Precision and recall are macro-averaged across the 32 FDI classes and were calculated at the confidence threshold that maximized mean F1 on the test set. CI, confidence interval; mAP50, mean average precision at IoU 0.50; mAP50-95, mean average precision across IoU thresholds from 0.50 to 0.95 in increments of 0.05.

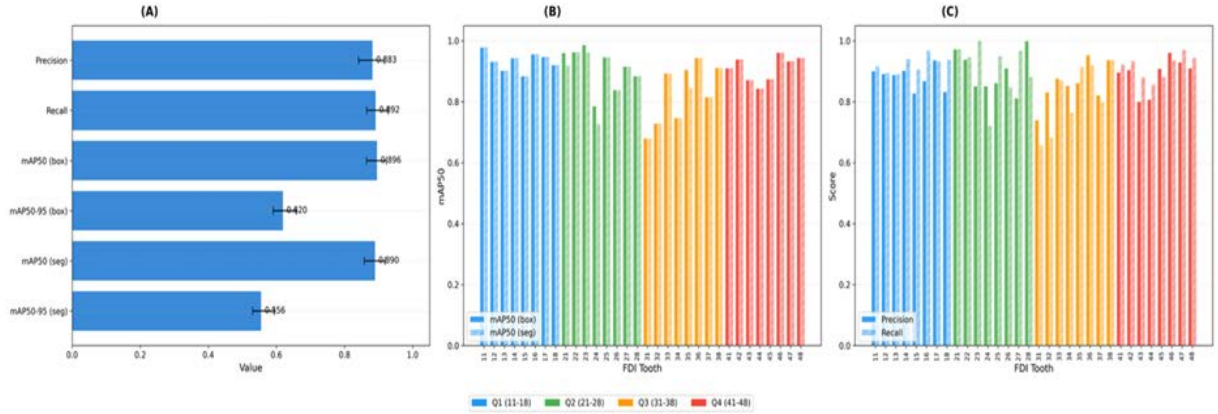


Figure 2. Performance of the YOLO26s-seg model on the held-out test set. (A) Global metrics with 95% bootstrap confidence intervals. (B) Per-class mAP50 for bounding box and instance segmentation, color-coded by dental quadrant. (C) Per-class box-level precision and recall, calculated at the test-set confidence threshold maximizing mean F1, color-coded by dental quadrant.

The gap between mAP50 and mAP50-95 indicates that while detection and identification were strong at a lenient overlap threshold, precise mask localization under stricter thresholds was more modest.

Of 1,124 predictions, 27 (2.40%) were assigned to an FDI class absent from the corresponding image, distributed across 15 of 43 images, suggesting that false positives in edentulous positions are infrequent but not negligible.

Cross-validation

Five-fold cross-validation yielded a mean mAP50 of 0.845±0.032 for detection and 0.839±0.032 for segmentation, with mAP50-95 of 0.592±0.019 and 0.546±0.021 (Table 2).

Table 2. Comparison of held-out test-set performance and 5-fold cross-validation results (mean±SD across folds).

Metric	Held-out test set	5-fold CV (mean±SD)
mAP50 (box)	0.896	0.845±0.032
mAP50-95 (box)	0.620	0.592±0.019
mAP50 (seg)	0.890	0.839±0.032
mAP50-95 (seg)	0.556	0.546±0.021
Precision (box)	0.883	0.853±0.029
Recall (box)	0.892	0.863±0.019

The low between-fold standard deviations suggest limited performance variability across alternative data partitions. Cross-validation estimates were lower than the held-out values, likely reflecting the smaller effective training subsets within each fold.

Architecture comparison

Under identical data split, seed, and training settings, YOLO11s-seg achieved the highest performance (mAP50 segmentation=0.949), followed by YOLOv8s-seg (0.928) and YOLO26s-seg (0.890) (Table 3).

Table 3. Same-split comparison of instance-segmentation architectures (identical data split, random seed, and user-specified training settings).

Model	mAP50 (box)	mAP50-95 (box)	mAP50 (seg)	mAP50-95 (seg)	Precision (box)	Recall (box)
YOLOv8s-seg	0.930	0.645	0.928	0.576	0.921	0.933
YOLO11s-seg	0.950	0.668	0.949	0.596	0.934	0.952
YOLO26s-seg	0.896	0.620	0.890	0.556	0.883	0.892

Comparisons are descriptive; each architecture was trained once with a single random seed and no paired statistical analysis was performed.

The differences were within six percentage points, though no formal paired comparison was performed. These results indicate that YOLO26s-seg did not outperform its predecessors on this dataset.

Per-class performance

Per-class mAP50 values exceeded 0.80 for 28 of 32 classes. The highest box-level mAP50 values were recorded for teeth 23 (0.987), 11 (0.979), and 22 (0.964).

Table 4. Per-class performance metrics for all 32 FDI tooth classes on the held-out test set (n=43 OPGs).

FDI	Instances	Precision (box)	Recall (box)	mAP50 (box)	mAP50-95 (box)	mAP50 (seg)	mAP50-95 (seg)
11	36	0.901	0.917	0.979	0.723	0.979	0.631
12	38	0.892	0.895	0.932	0.661	0.932	0.564
13	36	0.889	0.891	0.903	0.523	0.903	0.420
14	34	0.903	0.941	0.944	0.584	0.944	0.442
15	32	0.828	0.906	0.885	0.593	0.885	0.565
16	31	0.869	0.968	0.957	0.680	0.957	0.640
17	32	0.937	0.933	0.948	0.688	0.948	0.649
18	16	0.833	0.938	0.921	0.599	0.921	0.567
21	37	0.973	0.972	0.961	0.663	0.920	0.614
22	38	0.939	0.947	0.964	0.653	0.964	0.599
23	37	0.852	1.000	0.987	0.680	0.962	0.560
24	32	0.852	0.722	0.786	0.487	0.728	0.361
25	33	0.862	0.950	0.947	0.580	0.947	0.548
26	33	0.911	0.848	0.839	0.548	0.839	0.543
27	30	0.812	0.967	0.916	0.722	0.916	0.633
28	17	1.000	0.882	0.885	0.635	0.885	0.541
31	39	0.740	0.658	0.680	0.410	0.680	0.404
32	41	0.832	0.683	0.729	0.462	0.729	0.443
33	41	0.877	0.871	0.894	0.653	0.894	0.597
34	38	0.853	0.765	0.748	0.544	0.748	0.482
35	35	0.862	0.914	0.906	0.602	0.846	0.550
36	25	0.954	0.920	0.945	0.738	0.945	0.605
37	29	0.822	0.799	0.816	0.682	0.816	0.639
38	16	0.938	0.938	0.912	0.641	0.912	0.631
41	39	0.897	0.923	0.911	0.578	0.911	0.500
42	41	0.905	0.934	0.940	0.562	0.940	0.484
43	41	0.801	0.881	0.872	0.566	0.872	0.529
44	35	0.808	0.857	0.844	0.561	0.844	0.518
45	34	0.909	0.882	0.875	0.631	0.875	0.588
46	27	0.962	0.937	0.962	0.774	0.962	0.607
47	33	0.930	0.970	0.934	0.718	0.934	0.697
48	18	0.911	0.944	0.945	0.702	0.945	0.640

Instances indicate the number of reference-standard tooth instances in the held-out test set.
Precision and recall are box-level metrics calculated at a test-set-optimized confidence threshold.
Per-class values are descriptive point estimates; class-specific confidence intervals were not calculated.

10.12865/CHSJ.52.03.09

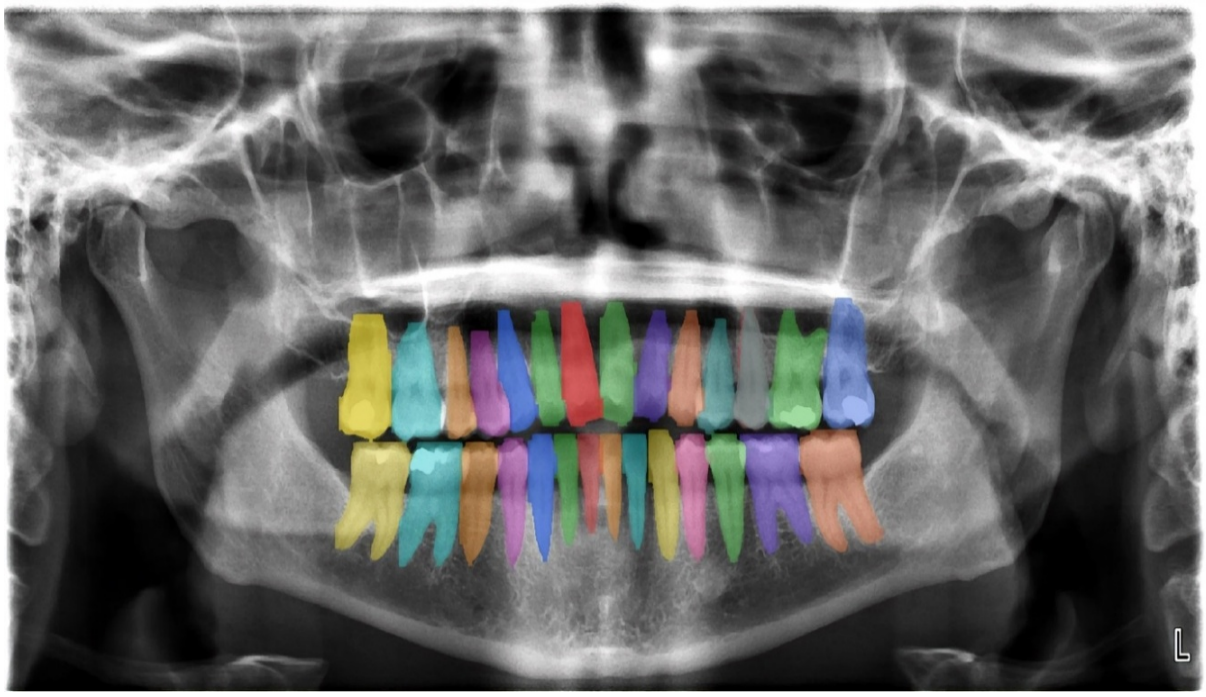


Figure 3. Representative inference result of the YOLO26s-seg model on a test orthopantomogram, showing polygon-based instance segmentation masks with individual color coding per detected tooth.

The four lowest were teeth 31 (0.680), 32 (0.729), 34 (0.748), and 24 (0.786), three of which were located in the mandibular left quadrant, suggesting a regional pattern that may reflect anatomical overlap, projection distortion, or annotation variability.

Third molars achieved mAP50 values of 0.921, 0.885, 0.912, and 0.945 despite having only 16-18 test instances each.

The lowest mAP50-95(seg) values were observed for tooth 24 (0.361), tooth 31 (0.404), and tooth 13 (0.420), confirming that boundary-level precision was weakest in regions of high anatomical complexity.

Complete per-class results are presented in Figure 2B, Figure 2C, and Table 4.

A representative inference result is shown in Figure 3.

Discussion

Principal findings

The YOLO26s-seg model trained on 282 orthopantomograms achieved mAP50 of 0.896 for bounding-box detection and 0.890 for polygon-based segmentation across all 32 permanent FDI classes on the held-out test set.

Twenty-eight of 32 classes exceeded a box-level mAP50 of 0.80, while mask localization under stricter overlap thresholds was more modest (mAP50-95 of 0.556).

On the same-split comparison, YOLO26s-seg did not outperform YOLOv8s-seg or YOLO11s-seg and is therefore presented as a transparently reported reference baseline rather than an architecture-specific advance.

Five-fold cross-validation (mAP50=0.845±0.032) confirmed limited performance variability across alternative data partitions.

Comparison with previous studies

Direct comparison with prior work is limited by differences in dataset size, annotation format, and evaluation protocol.

Bilgir et al. trained a Faster R-CNN model on 1,984 panoramic radiographs covering all 32 FDI classes and reported precision of 0.965 and F1-score of 0.961 [11], on a dataset approximately seven times larger.

Mahdi et al. reported mAP of 0.96-0.98 using Faster R-CNN on 1,000 radiographs with 32-class labeling [8].

The present model achieves lower but strong precision (0.883) and recall (0.892), suggesting that a one-stage segmentation architecture can achieve robust detection even on a substantially smaller training set.

For YOLO-based segmentation, Hua et al. reported mAP50 of 0.872 (box) and 0.855 (seg) using YOLO-DentSeg on an oral disease dataset [14], values comparable to those obtained here.

Singh et al. reported metrics approaching 0.998 using YOLOv11 on 670 pediatric radiographs [15], though under substantially

different conditions (56 classes, pediatric population).

The meta-analysis by Sadr et al. reported pooled sensitivity of 89% and specificity of 99% for deep learning-based tooth identification [16]; these classification-based metrics are not directly comparable with the object-detection framework used here but provide broader context.

The most directly comparable study is Eftimie et al. [17], who applied YOLOv11 models to the same Tufts Dental Database [18] for 32-class FDI instance segmentation. Using the full 1,000-image dataset, their best configuration achieved mask mAP50 of 0.986 and mAP50-95 of 0.622, exceeding the present results.

However, the studies differed in dataset size, model scale (YOLO11x vs. YOLO26s), partitioning, and annotation workflow. The present study contributes formally reported inter-annotator agreement and bootstrap confidence intervals, which Eftimie et al. did not report.

Per-class performance and segmentation quality

Three of the four lowest-performing classes (teeth 31, 32, 34) were located in the mandibular left quadrant. This regional pattern may reflect anatomical overlap, projection distortion, or annotation variability, consistent with difficulties reported for machine learning models on panoramic radiographs [12].

Third molars, despite only 16-18 test instances each, achieved mAP50 values of 0.921-0.945, which is noteworthy given that third molar detection is among the more challenging automated dental tasks [19].

The gap between mAP50 (0.890) and mAP50-95 (0.556) for segmentation warrants attention: since the rationale for polygon-based annotation is precise boundary delineation, mAP50-95 is the more informative quality measure. The weakest per-class values (tooth 24: 0.361; tooth 31: 0.404) may partly reflect intrinsic boundary ambiguity, consistent with the inter-annotator IoU variability observed in anatomically complex regions. The annotation quality assurance workflow, comprising senior-annotator review and independent reannotation with polygon IoU assessment, follows practices recommended for diagnostic AI development on dental radiographs [1,20].

The dataset split proportions (70/15/15) are aligned with conventions in the dental deep learning literature [15,21], and patient-level separation ensured no data leakage, in accordance with standards for diagnostic accuracy studies [16].

Limitations and future directions

This study has several limitations. The 282 OPGs represented a pragmatically selected convenience subset, and a per-reason exclusion registry was not maintained, introducing a potential risk of selection bias. The test set contained only 43 OPGs, and no external dataset was used for validation. Patient demographic characteristics and acquisition-device distributions were not analyzed; subgroup performance could not be assessed.

The reliability assessment was limited to inter-annotator polygon IoU without formal intra-rater analysis. The architecture comparison relied on a single random seed without repeated runs or paired testing.

Per-class confidence intervals were not calculated; class rankings should be interpreted as exploratory. Precision and recall were reported at a test-set-optimized confidence threshold and should be interpreted descriptively. No downstream clinical task or clinician-AI interaction was evaluated.

Future work should prioritize external multi-institutional validation, per-class uncertainty quantification, and prospective evaluation of downstream applications such as automated dental charting and AI-assisted radiographic screening.

The documented workflow, from polygon annotation to one-stage model training, may support future benchmarking on publicly available panoramic datasets, addressing the recognized gap where methodological standardization remains a barrier to clinical translation [10,22,23].

Conclusions

YOLO26s-seg achieved mAP50 values of 0.896 for detection and 0.890 for instance segmentation across all 32 permanent FDI classes on a publicly available panoramic dataset.

Mask localization under stricter overlap thresholds was lower (mAP50-95=0.556). The model did not outperform same-split YOLOv8s-seg and YOLO11s-seg baselines and should be interpreted as a reference benchmark rather than an architectural improvement.

External validation and prospective clinical evaluation are required before clinical application.

Acknowledgement

None to declare.

Author Contributions

Conceptualization, D-A.L.; Methodology, D-A.L.; Dataset curation and annotation, D-A.L.; Software development and model training, D-A.L.; Formal analysis and visualization, D-A.L.; Investigation, D-A.L.; Manuscript writing and initial draft preparation, D-A.L.; Manuscript review and editing, C.C.M., R.I., R.C.W.C., A.C.D.; Supervision, A.C.D.

All authors read and approved the final manuscript.

Funding

This research received no specific funding

Conflicts of interest

The authors declare no competing interests. The in-house annotation platform used in this study is an internal research tool; no author derives financial or commercial benefit from it.

Institutional Review Board

This study constituted a secondary analysis of publicly available, de-identified panoramic radiographs from the Tufts Dental Database, which was originally collected under approval from the Tufts University Institutional Review Board (IRB ID MODCR-01-12631).

Consent Statement

Not applicable

Data availability

The source radiographs are publicly available through the Tufts Dental Database (http://tdd.ece.tufts.edu/).

The FDI-labeled polygon annotations, dataset partitioning, and training code are available from the corresponding author upon reasonable request.

References

1. Beser B, Reis T, Berber MN, Topaloglu E, Gungor E, Kilic MC, Duman S, Celik O, Kuran A, Bayrakdar IS. YOLO-V5 based deep learning approach for tooth detection and segmentation on pediatric panoramic radiographs in mixed dentition. BMC Med Imaging, 2024, 24(1):172.
2. Meine H, Metzger MC, Weingart P, Wuster J, Schmelzeisen R, Rorich A, Georgii J, Brandenburg LS. Determination of Kennedy's classification in panoramic X-rays by automated tooth labeling. Int J Comput Assist Radiol Surg, 2025, 20(9):1835-1843.
3. Bulut A, Askin MB, Cinarer G. Comparative Analysis of Third Molar Segmentation Performance Between Sexes Using Deep Learning Models. Diagnostics (Basel), 2026, 16(7):977.
4. Lutescu DA, Manole CC, Tancu AMC, Ilinca R, Chau RCW, Felszeghy S, Didilescu AC. Can artificial intelligence in orthopantomography advance dental diagnostics through automated image analysis? Front Radiol, 2026, 6:1701356.
5. Samaranayake L, Tuygunov N, Schwendicke F, Osathanon T, Khurshid Z, Boymuradov SA, Cahyanto A. The Transformative Role of Artificial Intelligence in Dentistry: A Comprehensive Overview. Part 1: Fundamentals of AI, and its Contemporary Applications in Dentistry. Int Dent J, 2025, 75(2):383-396.
6. Jedlinski M, Jedlinski A, Rostkowski G, Janiszewska-Olszowska J, Mazur M. Artificial Intelligence in the Radiological Diagnosis of Impacted Maxillary Canines: A Systematic Review. J Clin Med, 2026, 15(9)
7. Lin S-Y, Chang H-Y. Tooth Numbering and Condition Recognition on Dental Panoramic Radiograph Images Using CNNs. IEEE Access, 2021, 9:166008-166026.
8. Mahdi FP, Motoki K, Kobashi S. Optimization technique combined with deep learning method for teeth recognition in dental panoramic radiographs. Sci Rep, 2020, 10(1):19261.
9. Chung M, Lee J, Park S, Lee M, Lee CE, Lee J, Shin YG. Individual tooth detection and identification from dental panoramic X-ray images via point-wise localization and distance regularization. Artif Intell Med, 2021, 111:101996.
10. Delamare E, Fu X, Huang Z, Kim J. Panoramic imaging errors in machine learning model development: a systematic review. Dentomaxillofac Radiol, 2024, 53(3):165-172.
11. Bilgir E, Bayrakdar IS, Celik O, Orhan K, Akkoca F, Saglam H, Odabas A, Aslan AF, Ozcetin C, Killi M, Rozylo-Kalinowska I. An artificial intelligence approach to automatic tooth detection and numbering in panoramic radiographs. BMC Med Imaging, 2021, 21(1):124.
12. Bonfanti-Gris M, Herrera A, Paraiso-Medina S, Alonso-Calvo R, Martinez-Rus F, Pradies G. Performance evaluation of three versions of a convolutional neural network for object detection and segmentation using a multiclass and reduced panoramic radiograph dataset. J Dent, 2024, 144:104891.
13. Adnan N, Ahmed SMF, Umer F. Developing Synthetic Orthopantomogram Datasets Through Generative Models. Int Dent J, 2026, 76(3):109465.
14. Hua Y, Chen R, Qin H. YOLO-DentSeg: A Lightweight Real-Time Model for Accurate Detection and Segmentation of Oral Diseases in Panoramic Radiographs. Electronics, 2025, 14(4)
15. Singh RK, Singh S, Awasthi N. Automated detection and federation dentaire internationale-based identification of teeth in mixed dentition panoramic radiographs using deep learning. J Indian Soc Pedod Prev Dent, 2026, 44(2):218-222.
16. Sadr S, Rokhshad R, Daghighi Y, Golkar M, Tolooie Kheybari F, Gorjinejad F, Mataji Kojori A, Rahimirad P, Shobeiri P, Mahdian M, Mohammad-Rahimi H. Deep learning for tooth identification and numbering on dental radiography: a systematic review and meta-analysis. Dentomaxillofac Radiol, 2024, 53(1):5-21.

17. Eftimie S, Ileni T, Iacob L, Hedesiu M, Diosan L. Tooth-level detection and mapping of dental pathologies on panoramic radiographs using YOLOv11 and RT-DETR. MethodsX, 2025, 15:103696.
18. Panetta K, Rajendran R, Ramesh A, Rao S, Agaian S. Tufts Dental Database: A Multimodal Panoramic X-Ray Dataset for Benchmarking Diagnostic Systems. IEEE J Biomed Health Inform, 2022, 26(4):1650-1659.
19. Celik ME. Deep Learning Based Detection Tool for Impacted Mandibular Third Molar Teeth. Diagnostics (Basel), 2022, 12(4)
20. Costa ED, Gaeta-Araujo H, Carneiro JA, Zancan BAG, Baranauskas JA, Macedo AA, Tirapelli C. Development of a dental digital data set for research in artificial intelligence: the importance of labeling performed by radiologists. Oral Surg Oral Med Oral Pathol Oral Radiol, 2024, 138(1):205-213.
21. Peker RB, Kurtoglu CO. Evaluation of the Performance of a YOLOv10-Based Deep Learning Model for Tooth Detection and Numbering on Panoramic Radiographs of Patients in the Mixed Dentition Period. Diagnostics (Basel), 2025, 15(4):405.
22. Khurshid Z, Osathanon T, Shire MA, Schwendicke F, Samaranayake L. Artificial Intelligence in Dentistry: A Concise Review of Reporting Checklists and Guidelines. Int Dent J, 2026, 76(1):109322.
23. Tuygunov N, Samaranayake L, Khurshid Z, Rewthamrongsris P, Schwendicke F, Osathanon T, Yahya NA. The Transformative Role of Artificial Intelligence in Dentistry: A Comprehensive Overview Part 2: The Promise and Perils, and the International Dental Federation Communique. Int Dent J, 2025, 75(2):397-404.

Corresponding Author: Dan-Adrian Lutescu, Department of Medical Informatics and Biostatistics, Faculty of Dentistry, "Carol Davila" University of Medicine and Pharmacy, Bucharest, Romania, e-mail: dan-adrian.lutescu@drd.umfcd.ro

 10.12865/CHSJ.52.03.09