Odovtos-International Journal of Dental Sciences (Odovtos-Int. J. Dent. Sc.), Online First, 2026. ISSN: 2215-3411
https://doi.org/10.15517/kc1cbz61

ODOVTOS
International Journal of Dental Sciences

UNIVERSIDAD DE COSTA RICA
FOd Facultad de Odontología

https://revistas.ucr.ac.cr/index.php/rOdontos

CLINICAL RESEARCH:

Deep Learning-Based Tooth Localization and Abnormality Detection on Panoramic Radiographs

Localización dental y detección de anomalías en radiografías panorámicas basadas en aprendizaje profundo

Tung Hoang Pharm, PhD[1] https://orcid.org/0000-0001-6653-3406
Young-Seok Park DDS, PhD[2] https://orcid.org/0000-0002-0148-7848
Khoa Van Pham DDS, PhD[3] https://orcid.org/0000-0002-5709-0719
Huong Thu Vu DDS, PhD[3] https://orcid.org/0000-0001-5222-6042
Khanh Hung Nguyen DDS, MSc[3] https://orcid.org/0009-0003-0948-040X
Soyeon Kim PhD[2] https://orcid.org/0000-0001-6377-9902
Van Mai Truong DDS, PhD[4] https://orcid.org/0000-0002-1473-9485

[1]Faculty of Pharmacy, University of Health Sciences, Vietnam National University, Ho Chi Minh City, Vietnam.
[2]Department of Oral Anatomy and Dental Research Institute, School of Dentistry, Seoul National University, Seoul, Republic of Korea.
[3]Faculty of Dentistry, University of Medicine and Pharmacy at Hochiminh City, Ho Chi Minh City, Vietnam.
[4]Faculty of Odonto-Stomatology, University of Health Sciences, Vietnam National University, Ho Chi Minh City, Vietnam.

Correspondence to: Van Mai Truong DDS, PhD - drtruongmaivan@gmail.com

Received: 23-IV-2026 Accepted: 19-VI-2026

ABSTRACT: Automated analysis of panoramic radiographs remains challenging due to anatomical complexity and image variability. While deep learning has shown strong performance in dental imaging, most studies focus on isolated tasks. This study aimed to propose a hierarchical YOLOv8-based framework aligned for comprehensive analysis of panoramic radiographs using structured dental annotations. A three-stage deep learning pipeline based on YOLOv8 was developed using the DENTEX dataset. The framework includes (1) quadrant classification, (2) tooth enumeration (FDI 11-48), and (3) tooth-level abnormality detection using a two-stage approach (binary screening followed by subtype classification). Panoramic radiographs with hierarchical annotations were used, with an 80:20 train–validation split. Performance was evaluated using mAP50, mAP50-95, precision, recall, and F1-score. The model achieved near-perfect performance for quadrant classification (mAP50=0.994, mAP50-95=0.750, precision=0.994, recall=0.995, and F1=0.994) and strong performance for tooth enumeration (mAP50=0.936, mAP50-95=0.536, precision=0.902, recall=0.897, and F1=0.899). Abnormality detection showed moderate performance (mAP50=0.687, mAP50-95=0.480, precision=0.655, recall=0.742, and F1=0.696). At the class level, impacted teeth (F1=0.904) and caries (F1=0.885) were well detected, whereas periapical lesions (F1=0.568) and deep caries (F1=0.585) showed lower performance. Precision and recall were balanced across tasks. The proposed hierarchical framework enables anatomical localization and integration of detection tasks of panoramic radiographs within a unified pipeline using YOLOv8. While performance is near-ceiling for anatomical tasks, disease detection remains challenging, particularly for low-contrast lesions.

KEYWORDS: Panoramic radiographs; Hierarchical framework; YOLO.

RESUMEN: El análisis automatizado de radiografías panorámicas sigue siendo un desafío debido a la complejidad anatómica y la variabilidad de las imágenes. Aunque el aprendizaje profundo ha demostrado un alto rendimiento en el análisis de imágenes dentales, la mayoría de los estudios se centran en tareas aisladas. El objetivo fue proponer un marco jerárquico basado en YOLOv8 para el análisis integral de radiografías panorámicas mediante anotaciones dentales estructuradas. Se desarrolló una canalización de aprendizaje profundo de tres etapas basada en YOLOv8 utilizando el conjunto de datos DENTEX. El marco comprende: (1) la clasificación por cuadrantes, (2) la enumeración dental según el sistema FDI (11-48) y (3) la detección de anomalías a nivel dental mediante un enfoque de dos etapas (cribado binario seguido de la clasificación por subtipos). Se utilizaron radiografías panorámicas con anotaciones jerárquicas, divididas en conjuntos de entrenamiento y validación en una proporción de 80:20. El rendimiento se evaluó mediante mAP50, mAP50-95, precisión (precision), sensibilidad (recall) y puntuación F1. El modelo alcanzó un rendimiento casi perfecto en la clasificación por cuadrantes (mAP50=0.994, mAP50-95=0.750, precisión=0.994, sensibilidad=0.995 y F1=0.994), un rendimiento elevado en la enumeración dental (mAP50=0.936, mAP50-95=0.536, precisión=0.902, sensibilidad=0.897 y F1=0.899). La detección de anomalías mostró un rendimiento moderado (mAP50=0.687, mAP50-95=0.480, precisión=0.655, sensibilidad=0.742 y F1=0.696). A nivel de clase, los dientes impactados (F1=0.904) y la caries dental (F1=0.885) fueron detectados con alta precisión, mientras que las lesiones periapicales (F1=0.568) y la caries profunda (F1=0.585) presentaron un rendimiento inferior. La precisión y la sensibilidad se mantuvieron equilibradas en todas las tareas. El marco jerárquico propuesto permite la localización anatómica y la integración de múltiples tareas de detección en radiografías panorámicas dentro de una canalización unificada basada en YOLOv8. Aunque el rendimiento fue casi óptimo para las tareas anatómicas, la detección de enfermedades continúa siendo un desafío, especialmente en lesiones de bajo contraste.

PALABRAS CLAVE: Radiografías panorámicas; Marco jerárquico; YOLO.

INTRODUCTION

Panoramic radiography is widely used in dental practice for comprehensive evaluation of the maxillofacial region (1). It enables assessment of overall jaw anatomy, individual tooth position, and a range of abnormal conditions within a single image. However, automated analysis of panoramic radiographs remains challenging due to projection distortion, overlapping anatomical structures, and substantial variation in tooth morphology and alignment (2).

Deep learning approaches have increasingly been applied to panoramic radiographs for tooth localization, enumeration, and disease detection (3, 4). Among dental public datasets, the DENTEX dataset provides standardized annotations for quadrant identification, tooth enumeration based on the FDI system, and tooth-level abnormality labeling (5). This structured annotation scheme reflects the inherent anatomical hierarchy of panoramic radiographs, progressing from global jaw partitioning to individual tooth identification and diagnostic categorization.

Given this hierarchical labeling structure, a sequential analytical strategy can be implemented to align with the dataset organization. Quadrant-level localization constrains spatial regions, tooth enumeration refines instance-level identification, and abnormality assessment operates at the tooth level. Modeling these tasks in stages allows independent optimization while maintaining anatomical consistency across levels.

In this study, YOLOv8-based architectures were employed as the common detection framework for all tasks. Dedicated models were trained for quadrant classification, tooth enumeration, and abnormality detection, ensuring methodological consistency while accommodating task-specific outputs.

MATERIALS AND METHODS

DATA SOURCES

Panoramic radiographs and corresponding annotations were obtained from the DENTEX dataset (5). The dataset provides bounding box annotations for quadrant identification, tooth enumeration based on the FDI numbering system, and abnormal tooth conditions. The annotated data consist of 693 radiographs with quadrant labels only, 634 radiographs with both quadrant and tooth enumeration labels, and 678 radiographs fully annotated for abnormal tooth detection, including quadrant, enumeration, and diagnostic labels. Detailed information on image acquisition and annotation protocols has been reported elsewhere.

ETHICS APPROVAL

This study involved secondary analysis of a publicly available, fully anonymized dataset; therefore, additional ethical approval was not required. The DENTEX dataset is distributed under a Creative Commons Attribution (CC-BY) license, permitting its use for research purposes.

MODEL DEVELOPMENT STRATEGY

Figure 1 shows three-stage hierarchical framework that was developed for panoramic radiograph analysis. The pipeline follows a vertical anatomical refinement and a horizontal diagnostic branching strategy: quadrant classification (Task 1), tooth enumeration (Task 2), and abnormality detection (Task 3). Task 1 provides coarse spatial partitioning, Task 2 refines detection to individual teeth (FDI 11-48), and Task 3 performs hierarchical disease classification through a binary screening stage followed by subtype classification. All tasks were implemented using YOLOv8-based architectures (6) to maintain structural consistency across stages. For each task, the dataset was randomly split into training and validation sets using an 80:20 ratio.

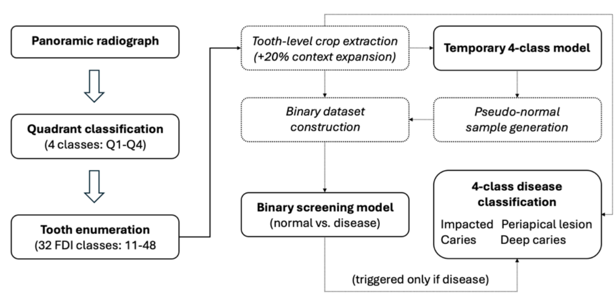

Figure 1. Hierarchical framework for quadrant classification, tooth enumeration, and abnormality detection tasks.

TASK 1: QUADRANT CLASSIFICATION

Quadrant segmentation was formulated as a four-class instance segmentation problem (upper-right: Q1, upper-left: Q2, lower-left: Q3, and lower-right: Q4). A YOLOv8 segmentation model was trained with mixed precision optimization. Standard geometric augmentation (mosaic, rotation, scaling, translation) was applied. The model with the highest validation mAP was selected and used to guide downstream localization.

TASK 2: TOOTH ENUMERATION

Tooth enumeration was formulated as a 32-class instance segmentation problem corresponding to FDI codes 11-48. The model was initialized using pretrained weights from Task 1 to retain anatomical feature representations. Early backbone layers were partially frozen to stabilize transfer learning while allowing higher-level adaptation to tooth-specific features.

Standard geometric augmentation was applied to improve generalization, while mask-distorting strategies were avoided to preserve segmentation fidelity. The model outputs tooth masks, bounding boxes, FDI labels, and confidence scores, which serve as structured inputs for the downstream disease classification stage.

TASK 3: ABNORMALITY DETECTION

Abnormality detection was implemented as a hierarchical tooth-level classification framework to address class imbalance and inter-class ambiguity. Rather than performing direct multi-class classification, a two-stage design was adopted: a binary screening model (normal vs. abnormal) followed by a disease subtype classifier applied only to abnormal teeth. Details of the process are described as follows:

(1) Tooth-level crop extraction: Tooth instances detected in Task 2 were converted into standardized crops with contextual expansion to preserve peri-tooth anatomical information. Disease labels were assigned based on spatial overlap between detected teeth and annotated lesions. This step transforms segmentation outputs into structured inputs for classification while maintaining anatomical consistency.

(2) Temporary disease model training: Since annotated data were predominantly disease-centered and no verified normal samples were available, a temporary four-class disease classifier was trained on confirmed disease crops extracted. This model was not intended for deployment but served exclusively as a scoring mechanism to assess disease likelihood for unannotated tooth instances.

(3) Pseudo-normal sample construction: The temporary model was applied to all detected teeth that did not spatially overlap with any annotated lesion. A conservative three-criterion filtering strategy was imposed: maximum disease prediction confidence below 0.25, top-two prediction gap below 0.10, and tooth detection confidence above 0.60. Only teeth satisfying all three criteria simultaneously were retained as pseudo-normal samples, minimizing label noise without relying on external normal datasets.

(4) Binary dataset construction: Disease samples and pseudo-normal samples from previous steps were merged into a unified binary dataset. Class labels were remapped such that all disease subtypes (Impacted, Caries, Periapical, Deep Caries) were consolidated into a single abnormal class, while pseudo-normal samples retained the normal class label. Balanced sampling with a 1:1 ratio was applied to prevent class dominance during binary classifier training.

(3) Binary screening model: A binary classifier was trained using confirmed disease samples and pseudo-normal samples with balanced sampling. Noise-robust optimization strategies were applied to mitigate potential pseudo-label uncertainty.

(4) Multi-class abnormality classification: For teeth predicted as abnormal, a dedicated multi-class classifier was trained using only confirmed disease samples. Class imbalance was addressed through weighted optimization.

EVALUATION METRICS

Model performance was evaluated using mean Average Precision at an Intersection over Union threshold of 0.5 (mAP50) as the primary metric. Mean Average Precision averaged across IoU thresholds from 0.5 to 0.95 (mAP50-95) was also reported. Precision and recall were calculated based on true positives, false positives, and false negatives. A detection was considered a true positive if the predicted bounding box overlapped with the ground truth by an IoU ≥ 0.5 and the predicted class matched the reference label; otherwise, detections were classified as false positives, and missed ground-truth objects were considered false negatives. All metrics were computed on the validation set using a fixed confidence threshold of 0.25 and an IoU threshold of 0.5 for non-maximum suppression. Per-class metrics were reported to account for class imbalance.

RESULTS

DATASET ANNOTATION SUMMARY

Table 1 summarizes the number of annotations of the panoramic dental dataset. Quadrant classification was available for 693 images, yielding 2,772 quadrant annotations with balanced distribution (693 per quadrant). Quadrant classification and tooth enumeration were jointly annotated in 634 images, producing 18,095 quadrant annotations (Quadrant 1: 4,506; Quadrant 2: 4,490; Quadrant 3: 4,542; Quadrant 4: 4,557) and 18,095 tooth annotations across eight tooth positions, with moderate imbalance (range: 1,504-2,508, lowest in Tooth 8 and highest in Tooth 1). Full quadrant, tooth, and abnormality annotations were available for all 678 images, comprising 3,529 quadrant annotations (823-965 per quadrant), 4,529 tooth annotations with pronounced skew toward posterior teeth (Tooth 6-8: 2,594 annotations), and 3,529 abnormality annotations showing strong class imbalance (Caries: 2,189 *vs*. Periapical lesion: 158).

MODEL PERFORMANCE

Table 2 summarizes model parameters used in this study, including training, loss, and augmentation parameters. Table 3 and Figure 2.A-2.C display model performance metrics for each task. Quadrant classification achieved the highest performance, with mAP50 of 0.994 and F1-score of 0.994. Tooth enumeration maintained strong detection performance across 32 FDI classes, reaching mAP50 of 0.936 and F1-score of 0.899. Abnormality detection showed lower overall performance compared to anatomical tasks, with mAP50 of 0.687 and F1-score of 0.696. At the class level, impacted teeth and caries achieved F1-scores of 0.904 and 0.885, respectively. Periapical lesions and deep caries demonstrated lower F1-scores of 0.568 and 0.585. Across tasks, precision and recall remained balanced, with no substantial discrepancy between the two metrics.

Table 1. Summary of datasets and annotation statistics used for quadrant classification, tooth enumeration, and abnormality detection.

Task	Category level	Category value	Number of annotations
Quadrant classification (693 images)	Quadrant	Quadrant 1	693
		Quadrant 2	693
		Quadrant 3	693
		Quadrant 4	693
Quadrant classification + Tooth enumeration (634 images)	Quadrant	Quadrant 1	4,506
		Quadrant 2	4,490
		Quadrant 3	4,542
		Quadrant 4	4,557
	Tooth	Tooth 1	2,508
		Tooth 2	2,497
		Tooth 3	2,499
		Tooth 4	2,404
		Tooth 5	2,299
		Tooth 6	2,070
		Tooth 7	2,314
		Tooth 8	1,504
Quadrant classification + Tooth enumeration + Abnormality detection (678 images)	Quadrant	Quadrant 1	823
		Quadrant 2	799
		Quadrant 3	965
		Quadrant 4	942
	Tooth	Tooth 1	78
		Tooth 2	97
		Tooth 3	81
		Tooth 4	249
		Tooth 5	430
		Tooth 6	894
		Tooth 7	768
		Tooth 8	932
	Abnormality	Impacted	604
		Caries	2,189
		Periapical lesion	158
		Deep caries	578

Table 2. Summary of model parameters.

Parameter	Task 1	Task 2	Task 3 (Temporary 4-class disease model)	Task 3 (Binary disease detection model)	Task 3 (Class disease detection model)
Model information					
Model base	yolov8n-seg.pt	best.pt (Task 1)	yolov8n.pt	yolov8n.pt	yolov8n.pt
Task type	Segmentation	Detection	Detection (4-class)	Detection (Binary)	Detection (4-class)
Classes	4 quadrants	32 FDI teeth	4 diseases	2 (Normal/Disease)	4 diseases
Training parameter					
epochs	100	100	100	120	150
imgsz	640	512	320	320	320
batch	16	4	32	32	32
device	'0'	0	'0'	'0'	'0'
optimizer	'auto'	'AdamW'	'AdamW'	'AdamW'	'AdamW'
lr0	—	0.001	0.001	0.001	0.001
lrf	—	0.01	0.01	0.005	0.01
weight_decay	—	0.0005	—	—	—
freeze	—	10	—	—	—
patience	50	20	20	15	30
save	TRUE	TRUE	TRUE	TRUE	TRUE
save_period	10	20	—	20	25
val	TRUE	TRUE	—	—	—
workers	8	2	8	8	8
amp	TRUE	TRUE	TRUE	TRUE	TRUE
seed	42	—	—	—	—
deterministic	TRUE	—	—	—	—
verbose	TRUE	TRUE	TRUE	TRUE	TRUE
plots	TRUE	TRUE	TRUE	TRUE	TRUE
Loss parameter					
box	—	—	7.5	7.5	7.5
cls	—	—	0.5	0.6	dynamic
dfl	—	—	1.5	1.5	1.5
conf	—	—	0.25	0.25	0.25
iou	—	—	0.7	0.7	0.7
label_smoothing	—	—	—	0.1	—
Augmentation parameter					
hsv_h	0.015	—	—	0.02	0.015
hsv_s	0.7	—	—	0.8	0.7
hsv_v	0.4	—	—	0.5	0.4
degrees	0	10	15	20	15
translate	0.1	0.1	0.1	0.15	0.1
scale	0.5	0.2	0.3	0.4	0.3
shear	0	—	—	—	—

Parameter	Task 1	Task 2	Task 3 (Temporary 4-class disease model)	Task 3 (Binary disease detection model)	Task 3 (Class disease detection model)
perspective	0	—	—	—	—
flipud	0	0.5	0.5	0.5	0.5
fliplr	0.5	0	0	0	0
mosaic	1	0.5	0.5	0.6	0.5
mixup	0	0	—	—	—
copy_paste	—	0	—	—	—

Table 3. Performance of quadrant classification, tooth enumeration, and abnormality detection models.

Task	mAP50	mAP50-95	Precision	Recall	F1-score
Quadrant classification	0.994	0.750	0.994	0.995	0.994
Tooth enumeration	0.936	0.536	0.902	0.897	0.899
Abnormality detection	0.687	0.480	0.655	0.742	0.696
Impacted	0.957	0.608	0.933	0.876	0.904
Caries	0.933	0.675	0.818	0.963	0.885
Periapical lesion	0.519	0.322	0.555	0.581	0.568
Deep caries	0.612	0.389	0.558	0.615	0.585

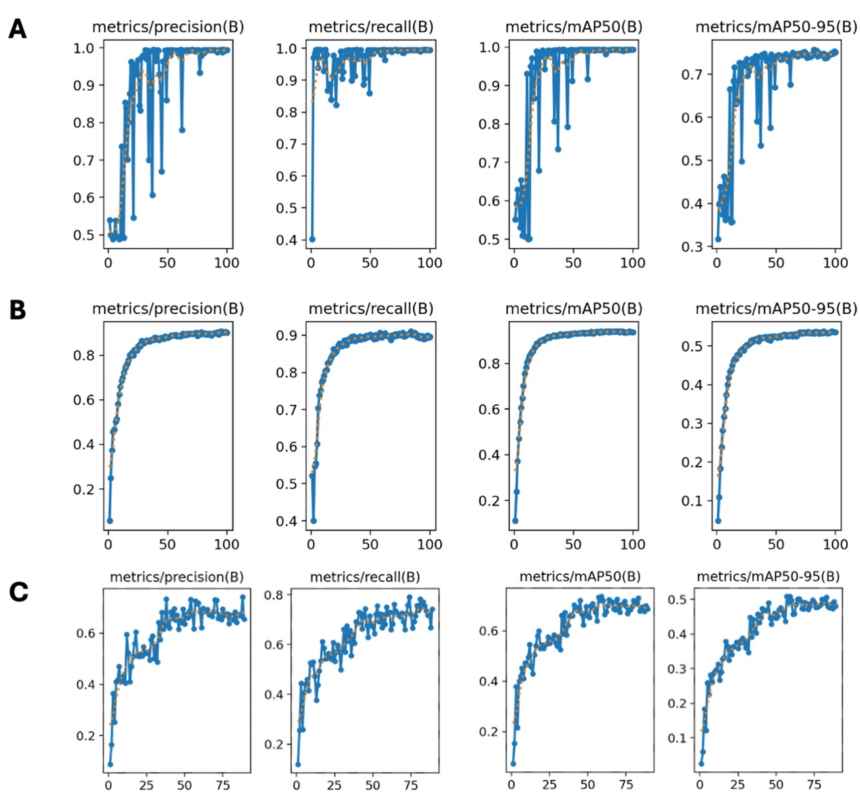

Figure 2. Model performance for (A) quadrant classification, (B) tooth enumeration, and (C) abnormality detection tasks.

CONFUSION MATRIX ANALYSIS

Confusion matrix analysis shows that quadrant classification (Figure 3.A) shows almost exclusive diagonal dominance, reflecting the near-perfect performance (F1=0.994). Tooth enumeration (Figure 3.B) maintains a strong diagonal pattern across 32 classes, with limited confusion mainly between neighboring FDI labels, corresponding to an overall F1-score of 0.899. For abnormality detection (Figure 3.C), diagonal concentration is evident for impacted teeth and caries, while periapical lesions and deep caries exhibit greater off-diagonal distribution.

QUALITATIVE RESULTS

Figure 4.A-4.D illustrates representative outputs across the hierarchical pipeline. Quadrant segmentation accurately delineates anatomical regions, followed by consistent tooth-level localization and FDI labeling. The final abnormality detection stage identifies affected teeth while preserving alignment with upstream anatomical predictions.

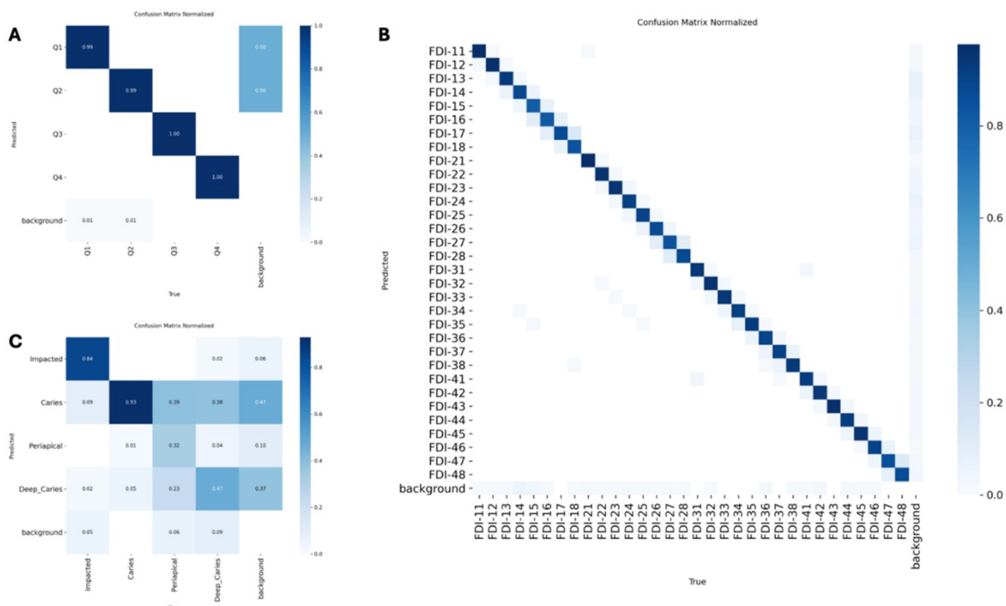

Figure 3. Confusion matrix heatmap for (A) quadrant classification, (B) tooth enumeration, and (C) abnormality detection tasks.

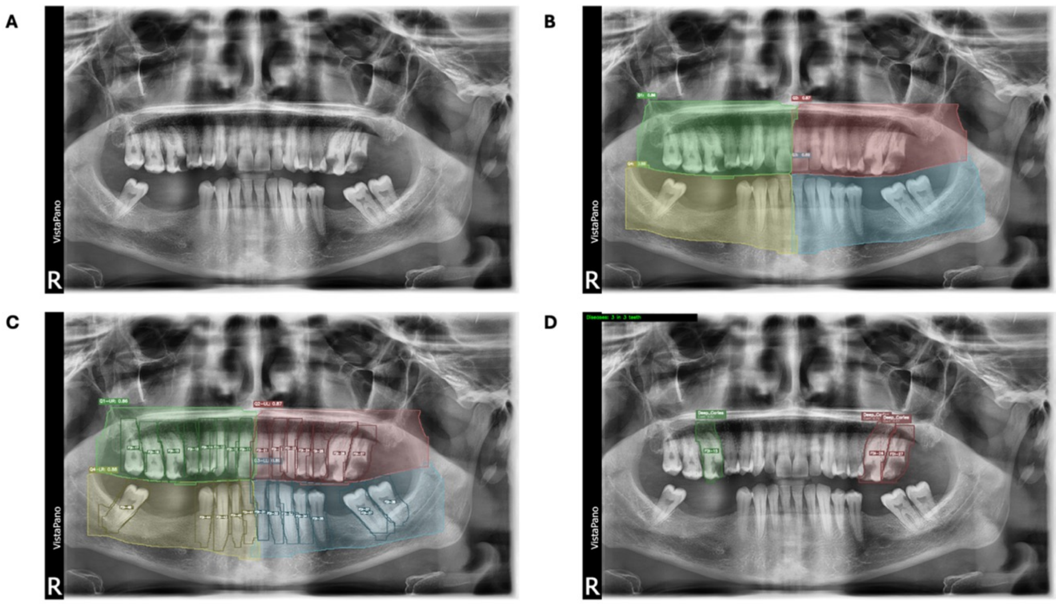

Figure 4. Example of hierarchical annotation and model outputs across the proposed pipeline: (A) Original panoramic radiograph without annotations; (B) Quadrant segmentation results (Task 1); (C) Tooth enumeration with FDI labels (Task 2); (D) Tooth-level abnormality detection (Task 3).

DISCUSSION

In this study, we developed and evaluated a three-stage hierarchical framework for panoramic radiograph analysis using the DENTEX dataset, encompassing quadrant classification, tooth enumeration, and tooth-level abnormality detection. Using a unified YOLOv8-based architecture across tasks, the model achieved near-perfect performance for quadrant classification, strong performance for tooth enumeration, and moderate performance for abnormality detection. These results reflect a clear performance gradient from coarse anatomical localization to fine-grained disease classification.

Our findings are consistent with prior evidence demonstrating strong performance of deep learning in dental radiograph analysis. A previous systematic review of 29 individual studies reported accuracy ranging from 81.8% to 99%, sensitivity from 75.5% to 98%, specificity from 79.9% to 99%, and an AUC of 96% for tooth identification and numbering (3). However, most studies addressed individual tasks such as tooth detection, numbering, or disease classification separately, rather than adopting a unified or hierarchical framework (3).

Recent studies illustrate this task-specific focus. For disease detection, a dual-stage model for caries segmentation achieved Dice=0.926, precision=0.921, recall=0.931, and IoU=0.862, demonstrating strong performance in well-defined lesion segmentation (7). For multi-class anomaly detection, a two-stage ESRGAN-YOLOv8 pipeline reported mAP50 of 56.9% and mAP50-95 of 41.6%, with high sensitivity (0.942) for crown detection but substantially lower sensitivity for caries (0.174) and severely decayed teeth (0.355), indicating variability across classes (8). For impacted teeth detection, a MedSAM-based approach achieved accuracy of 86.73% but relatively low F1-score (0.535) and IoU (0.365), highlighting limitations in fine-grained localization (9). For periapical lesion detection, an integrated YoCNET model (YOLOv5 + ConvNeXt) demonstrated strong performance, with accuracy of 90.93%, precision of 98.88%, sensitivity of 85.30%, F1-score of 0.9159, and AUC of 0.9757, outperforming alternative integrated models such as YOLOv5 + ResNet34 (10). These results indicate that combining detection and classification architectures can improve performance in multi-lesion detection tasks, particularly when tooth-level segmentation is incorporated (10).

YOLO-based architectures have been widely applied across these tasks, including tooth detection, enumeration, and disease classification (8, 11-15). Prior studies reported high performance for structured tasks, such as mAP up to 0.98 and sensitivity and precision approaching 0.99 in tooth detection and numbering, as well as F1-scores above 0.92 in caries classification (8, 11-15). However, these approaches are typically implemented as single-stage or task-specific models, limiting their ability to capture hierarchical anatomical relationships (8, 11-15).

In comparison, our framework integrates quadrant classification, tooth enumeration, and abnormality detection within a unified hierarchical pipeline. This design preserves strong performance in anatomical tasks (F1=0.994 and 0.899) while achieving competitive results in disease detection, particularly for impacted teeth (F1=0.904) and caries (F1=0.885). Lower performance for periapical lesions (F1=0.568) and deep caries (F1=0.585) remains consistent with prior studies, reflecting persistent challenges related to low contrast and anatomical overlap. Prior studies using transfer learning on cropped panoramic tooth images have also demonstrated improved performance for localized abnormality detection, suggesting that reducing background anatomical complexity may facilitate lesion recognition (16). However, such approaches may suffer from limited generalizability to full panoramic radiographs and real-world clinical settings, where anatomical overlap and contextual variability are substantially greater (16).

Overall, these findings suggest that while YOLO-based and other deep learning models achieve high accuracy in isolated tasks, integrating these tasks within a hierarchical framework may provide a more consistent and clinically interpretable approach to panoramic radiograph analysis.

One strength of our study is the alignment between dataset structure and model architecture. The DENTEX dataset provides hierarchical annotations (quadrant→tooth→abnormality), and our pipeline explicitly preserves this anatomical progression. This design improves interpretability: each stage can be independently evaluated, and downstream errors can be traced to upstream localization performance. Using YOLOv8 across all tasks ensured architectural consistency while allowing task-specific adaptation. Transfer learning from quadrant segmentation to tooth enumeration helped retain global anatomical context. Moreover, the two-stage abnormality framework mitigated extreme class imbalance without introducing external datasets. The pseudo-normal sample construction, although conservative, enabled balanced binary training and improved screening robustness. Another strength is the balanced precision-recall profile observed across tasks and stable trade-offs, suggesting consistent confidence calibration under a fixed threshold.

Despite strong performance in anatomical tasks, abnormality detection remains the main bottleneck. Several factors may explain this gap. First, panoramic radiographs inherently contain distortion, overlapping structures, and limited spatial resolution, which particularly affect periapical and deep carious lesions. Second, the dataset exhibits pronounced class imbalance, with caries dominating abnormal annotations. Although weighted optimization and hierarchical classification were applied, minority classes still showed reduced performance. Third, while YOLOv8 provides strong real-time performance and structural simplicity, transformer-based or hybrid architectures have recently demonstrated advantages in modeling global context in medical imaging (17-19). Fourth, external validation using independent datasets or comparison with expert performance was not performed, which may limit the generalizability of the proposed framework across different clinical settings. In addition, abnormality interpretation in panoramic radiographs may inherently involve inter-observer variability, particularly for subtle lesions with ambiguous radiographic boundaries. Similar concerns regarding diagnostic subjectivity have been reported in AI-based imaging studies for temporomandibular joint osteoarthritis, where AI-assisted interpretation demonstrated higher agreement with reference standards than human evaluation alone (20). Finally, from an ethical and clinical perspective, false-negative predictions in abnormality screening may potentially delay diagnosis or referral, highlighting the need for careful human oversight before clinical application. Therefore, the proposed framework should be considered as a decision-support tool rather than a fully autonomous diagnostic system.

Overall, this study demonstrates that a YOLOv8-based hierarchical framework can effectively leverage structured annotations in the DENTEX dataset, achieving robust anatomical localization and moderate disease detection performance. While anatomical tasks approach near-ceiling performance, further methodological and data-level improvements are needed to close the gap in abnormality classification.

DATA AVAILABILITY: The data used in this study are publicly available as part of the DENTEX dataset.

FUNDING/SUPPORT: This study was supported by University of Health Sciences, Vietnam National University Ho Chi Minh City.

CONFLICT OF INTERESTS DISCLOSURES: The authors have declared that no competing interests exist.

ETHICS APPROVAL: This study used publicly available data; therefore, ethics approval was not required.

AUTHOR CONTRIBUTION STATEMENT: Conceptualization and Design, V.M.T., T.H.P., Y.S.P., K.V.P., H.T.V., K.H.N., S.K.; Literature Review, V.M.T., T.H.P.; Methodology and Validation, V.M.T., T.H.P., Y.S.P.; Formal Analysis, T.H.P.; Resources, V.M.T., T.H.P.; Data Analysis and Interpretation, T.H.P., V.M.T.; Writing – Original Draft Preparation, T.H.P.; Writing – Review & Editing, Y.S.P., V.M.T., K.V.P., H.T.V., K.H.N., S.K..

REFERENCES

1. Różyło-Kalinowska I. Panoramic radiography in dentistry. Clin Dent Rev. 2021; 5: 26.
2. Ozden S., Kula B., Tankus M. Automated deep learning detection of orthodontically induced external apical root resorption in maxillary incisors on panoramic radiographs. Prog Orthod. 2026; 27 (1).
3. Sadr S., Rokhshad R., Daghighi Y., Golkar M., Tolooie Kheybari F., Gorjinejad F., et al. Deep learning for tooth identification and numbering on dental radiography: a systematic review and meta-analysis. Dentomaxillofac Radiol. 2024; 53 (1): 5-21.
4. Zhou Z., Zhu J., Zhang Y., Guan X., Wang P., Li T. Deep learning in dental image analysis: A systematic review of datasets, methodologies, and emerging challenges 2025 [Available from: https://arxiv.org/abs/2510.20634].
5. Hamamci I.E., Er S., Durugol O.F., Cakmak G.R., Rosa E.D.L., Simsar E., et al. DENTEX: Dental Enumeration and Tooth Pathosis Detection Benchmark for Panoramic X-ray 2023 [Available from: https://arxiv.org/abs/2305.19112].
6. Varghese R., M. S. YOLOv8: A novel object detection algorithm with enhanced performance and robustness. 2024 International

Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). 2024: 1-6.
7. Vachmanus S., Pornprasertsuk-Damrongsri S., Mongkolwat P., Phinklao N, Kitisubkanchana J, Chaikantha S, et al. Dual-stage deep neural network for tooth localization and caries segmentation in panoramic dental imaging. Neural Comput Appl. 2026; 38 (147).
8. Chaudhari A.Y., Birwadkar P., Joshi S., Verma Y., Sindgi R. Classification of periapical dental X-ray using the YOLOv8 deep learning model. MethodsX. 2025; 15: 103721.
9. Zhicheng H., Yipeng W., Xiao L. Deep learning-based detection of impacted teeth on panoramic radiographs. Biomed Eng Comput Biol. 2024; 15: 11795972241288319
10. Liu J., Liu X., Shao Y., Gao Y., Pan K., Jin C., et al. Periapical lesion detection in periapical radiographs using the latest convolutional neural network ConvNeXt and its integrated models. Sci Rep. 2024; 14 (1): 25429.
11. Bayati M., Alizadeh Savareh B., Ahmadinejad H., Mosavat F. Advanced AI-driven detection of interproximal caries in bitewing radiographs using YOLOv8. Sci Rep. 2025; 15 (1): 4641.
12. Karakus R., Ozic M.U., Tassoker M. AI-assisted detection of interproximal, occlusal, and secondary caries on bite-wing radiographs: A single-shot deep learning approach. J Imaging Inform Med. 2024; 37 (6): 3146-59.
13. Panyarak W., Wantanajittikul K., Charuakkra A., Prapayasatok S., Suttapak W. Enhancing caries detection in bitewing radiographs using YOLOv7. J Digit Imaging. 2023; 36 (6): 2635-47.
14. Lee S., Oh S.I., Jo J., Kang S., Shin Y., Park J.W. Deep learning for early dental caries detection in bitewing radiographs. Sci Rep. 2021; 11 (1): 16807.
15. Chen H., Zhang K., Lyu P., Li H., Zhang L., Wu J., et al. A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. Sci Rep. 2019; 9 (1): 3840.
16. Shetty S., Talaat W., AlKawas S., Al-Rawi N., Sadek M., Elayyan L., et al. Transfer learning models in the detection of pulp calcifications- A preliminary study. Journal of Oral Biology and Craniofacial Research. 2026; 16 (3): 101462.
17. Shi Y., Li F., Zhao S., Yu H., Chen X., Liu Q. IAP-TransUNet: integration of the attention mechanism and pyramid pooling for medical image segmentation. Front Neurorobot. 2025; 19: 1706626.
18. Chong Y., Xie N., Liu X., Pan S. P-TransUNet: an improved parallel network for medical image segmentation. BMC Bioinformatics. 2023; 24 (1): 285.
19. Pan S., Liu X., Xie N., Chong Y. EG-TransUNet: a transformer-based U-Net with enhanced and guided models for biomedical image segmentation. BMC Bioinformatics. 2023; 24 (1): 85.
20. Talaat W.M., Shetty S., Al Bayatti S., Talaat S., Mourad L., Shetty S., et al. An artificial intelligence model for the radiographic diagnosis of osteoarthritis of the temporomandibular joint. Sci Rep. 2023; 13 (1): 15972.