Journal of Imaging Informatics in Medicine (2026) 39:2537–2552
https://doi.org/10.1007/s10278-025-01642-z

ORIGINAL PAPER

Detection of Jaw Lesions on Panoramic Radiographs Using Deep Learning Method

Dilek Çoban[1] · Yasin Yaşa[2] · Abdulsamet Aktaş[3] · Hamza Osman İlhan[4]

Received: 7 April 2025 / Revised: 14 July 2025 / Accepted: 3 August 2025 / Published online: 28 August 2025
© The Author(s) under exclusive licence to Society for Imaging Informatics in Medicine 2025

Abstract

This study aimed to evaluate and compare the performance of state-of-the-art deep learning models for detecting and segmenting both radiolucent and radiopaque jaw lesions on panoramic radiographs. A total of 2371 anonymized panoramic radiographs containing jaw lesions were retrospectively collected and categorized into radiolucent and radiopaque datasets. Expert annotation was performed to delineate lesion boundaries and assign anatomical localization (anterior/posterior maxilla and mandible). Four deep learning architectures—YOLOv8, YOLOv11, Mask R-CNN, and RT-DETR—were trained and evaluated under three experimental scenarios: (I) training without spatial labels, (II) data augmentation with unlabeled background images, and (III) inclusion of spatial localization annotations. Performance metrics included precision, recall, F1-score, and mean average precision (mAP@0.5 and mAP@0.5–0.95), with paired t-tests used for statistical comparison. In Scenario I, YOLOv11x-seg and YOLOv8x-seg achieved the highest segmentation performance for radiolucent and radiopaque lesions, respectively. For detection, YOLOv8x performed best on radiolucent lesions, while RT-DETR-L outperformed others on radiopaque lesions. In Scenario II, while YOLOv8x-seg achieved the best segmentation results across both lesion types, RT-DETR-L demonstrated superior detection performance, particularly for radiolucent lesions. In Scenario III, RT-DETR-L consistently outperformed all models across both lesion types. This study demonstrates the potential of state-of-the-art deep learning models for effective detection of lesions in panoramic radiographs. The developed models may offer valuable support to clinicians in lesion evaluation; however, it is recommended that they be employed primarily as decision support tools within clinical workflows, rather than as standalone diagnostic systems.

Keywords Deep learning · Lesions · Panoramic radiography · Artificial intelligence · YOLO

Introduction

The application of artificial intelligence (AI) in the field of medicine began with the development of clinical decision support systems (CDSS). These are interactive computer programs that have been developed with the objective of assisting clinicians in making diagnoses, determining appropriate treatment options, and predicting prognosis. In the present, AI-based models have become a pervasive tool in numerous domains within the medical and dental professions. These models use technologies such as machine learning and deep learning [1].

The compatibility of AI with image processing has led to an increase in studies on radiology. The most commonly used imaging method in dental practice is panoramic radiography. Panoramic radiography is a valuable tool for the detection of lesions in a range of anatomical structures, including teeth, jaws, the temporomandibular joint (TMJ),

✉ Dilek Çoban
dilek.coban@kafkas.edu.tr

Yasin Yaşa
yasayasin@outlook.com

Abdulsamet Aktaş
abdulsamet.aktas@marmara.edu.tr

Hamza Osman İlhan
hoilhan@yildiz.edu.tr

[1] Department of Dentomaxillofacial Radiology, Faculty of Dentistry, Kafkas University, Kars, Turkey

[2] Department of Dentomaxillofacial Radiology, Faculty of Dentistry, Ordu University, Ordu, Turkey

[3] Department of Computer Engineering, Faculty of Technology, Marmara University, Istanbul, Turkey

[4] Department of Computer Engineering, Faculty of Electrical and Electronic, Yildiz Technical University, Istanbul, Turkey

maxillary sinuses, and salivary glands [2]. Lesions of the jaw bones comprise a group of pathological entities that present with a range of symptoms, clinical conditions, and radiographic and histological features. These lesions are often detected incidentally during routine radiographic examinations due to the rarity of symptoms in the early stages. However, some lesions exhibit a destructive potential associated with high recurrence rates and can grow to large sizes before becoming symptomatic, which can have a detrimental impact on the quality of life of affected patients [3, 4].

The majority of jaw lesions are radiolucent (> 80%) and typically include benign entities such as apical granulomas, radicular cysts, and abscesses [5]. However, radiolucent features may also be associated with aggressive or malignant conditions such as odontogenic keratocysts, ameloblastomas, and metastatic lesions [6]. Radiopaque or mixed-density lesions are less common and include conditions such as idiopathic osteosclerosis, condensing osteitis, and dense bone islands [7, 8]. These lesions are often subtle and may be overlooked during manual interpretation.

The automatic detection of lesions by artificial intelligence is clinically useful for busy and/or inexperienced dentists, as it can contribute to reducing the risk of missing lesions other than the patient's main complaint. Similarly, it is thought to directly increase the efficiency of clinical services by supporting dentists' clinical decisions and reducing the time allocated to interpret radiographs. It can also be thought to positively affect the prognosis of diseases by accelerating the diagnostic process [9, 10]. The number of studies on deep learning for lesion detection in panoramic radiographs has been increasing in recent years.

Yang et al. [11] conducted a comprehensive evaluation of the diagnostic performance of the YOLOv2 deep learning algorithm for detecting odontogenic cysts and tumors on panoramic radiographs. The study utilized a retrospective dataset comprising 1602 lesions collected from Yonsei University Dental Hospital between 2010 and 2019, classified into dentigerous cysts, odontogenic keratocysts, ameloblastomas, and normal cases. The performance of YOLOv2 was systematically compared with that of oral and maxillofacial surgeons and general practitioners using precision, recall, accuracy, and F1-score metrics.

The results demonstrated that YOLOv2 achieved superior precision (0.707) and recall (0.680) compared to both clinician groups; however, these differences did not reach statistical significance ($p > 0.05$). This study provides preliminary evidence supporting the potential utility of real-time object detection algorithms as adjunct diagnostic tools in oral and maxillofacial radiology. The findings suggest that convolutional neural networks may facilitate earlier detection of pathological lesions, thereby contributing to reduced morbidity through timely intervention. Nevertheless, further validation using larger, multicenter datasets is warranted to confirm clinical applicability.

In the study conducted by Rasic et al. [12], the YOLOv8 architecture was employed for the automatic detection and segmentation of radiolucent lesions in the mandible. The research analyzed a total of 226 panoramic radiographs obtained from patients treated at the University of Zagreb between 2013 and 2023. The study focused on common lesions such as radicular cysts, ameloblastomas, odontogenic keratocysts (OKC), dentigerous cysts, and residual cysts. Due to the limited size of the dataset, data augmentation techniques were applied to enhance the overall performance of the model. In the detection task, the precision, recall, mAP@50, and mAP@50–95 scores obtained after data augmentation were reported as 95.2%, 94.4%, 97.5%, and 68.7%, respectively. For the segmentation task, these values reached 100%, 94.5%, 96.6%, and 72.2%, respectively. Although the results demonstrate strong model performance, the limited number of samples in the dataset suggests potential constraints regarding the model's generalizability and clinical applicability. Therefore, future studies utilizing larger and more diverse datasets are essential to strengthen the clinical relevance of the approach and enhance the validity of the findings.

Tropčić et al. [13] developed a YOLOv8-based deep learning model using a mandibular lesion panoramic radiograph dataset established at the University of Zagreb. The study evaluated detection, multi-class detection, segmentation, and multi-class segmentation tasks on 200 images containing 226 lesions. Through data augmentation techniques, the model achieved high performance in detection and segmentation tasks, reaching up to 100% precision and 96% mAP@50. However, performance in multi-class segmentation was limited, with an mAP@50–95 value of 59.5%. While the study demonstrates the effectiveness of YOLOv8 for mandibular lesion detection, it is constrained by the limited dataset size, exclusive focus on the mandibular region, and decreased performance in multi-class tasks. These findings highlight the need for validation on larger and more diverse datasets to enhance the model's clinical applicability.

Shoorgashti et al. [14] conducted a systematic review and meta-analysis to evaluate the diagnostic accuracy of AI models in identifying odontogenic keratocysts (OKCs) on panoramic radiographs. Following a comprehensive literature search across five databases, eight studies meeting predefined inclusion criteria were incorporated into the meta-analysis. The pooled results revealed that AI-based models achieved an overall sensitivity of 83.66% (95% CI, 73.75–93.57) and specificity of 82.89% (95% CI, 70.31–95.47) in the diagnosis of OKCs. Notably, models utilizing the YOLO architecture demonstrated significantly higher diagnostic performance, with a sensitivity of 96.4% and specificity of 96.0%, compared to other deep learning

algorithms. Meta-regression analysis identified model architecture as a statistically significant predictor of diagnostic accuracy ($p < 0.05$), while Egger's test ($p = 0.042$) indicated potential publication bias and considerable heterogeneity among studies ($I^2 > 50\%$). These findings support the potential of YOLO-based AI systems to achieve high diagnostic accuracy in detecting OKCs. However, the study emphasizes that while AI performs well in typical cases, it cannot replace expert clinical judgment in atypical or complex scenarios. Instead, AI should be considered a complementary tool. Accordingly, the authors suggest that future studies employing larger sample sizes and standardized evaluation protocols are essential for establishing the clinical validity of AI models in this domain.

In the existing literature, deep learning models—particularly those based on the YOLO architecture and Mask R-CNN—have been widely applied for the detection and segmentation of odontogenic lesions in panoramic radiographs. For instance, Yang et al. [11] demonstrated that YOLOv2 could perform comparably to human experts in detecting cysts and tumors, while Rasic et al. [12] and Tropčić et al. [13] showed promising results using YOLOv8 on small datasets focused mainly on radiolucent lesions. A recent meta-analysis by Shoorgashti et al. [14] further confirmed the superior diagnostic performance of YOLO-based models, though it also highlighted significant heterogeneity and publication bias across studies.

Despite these advances, the current body of research faces several key limitations: (i) reliance on relatively small, homogeneous datasets; (ii) limited lesion diversity, with most studies focusing exclusively on radiolucent lesions; and (iii) the lack of statistical comparisons between deep learning models, which hinders robust model selection for clinical deployment. To address these gaps, our study utilizes a newly constructed and larger dataset that includes both radiolucent and radiopaque lesions, a novel addition rarely seen in previous research.

The objective of this study is to evaluate and compare the performance of advanced deep learning architectures—including YOLOv8, YOLOv11, Mask R-CNN, and the transformer-based RT-DETR—for the detection and segmentation of radiolucent and radiopaque jaw lesions on panoramic radiographs.

Methods

This study was approved by the Ordu University Faculty of Medicine Clinical Research Ethics Committee (Decision No. 2023/285) and was conducted in accordance with the Declaration of Helsinki. The requirement to obtain written informed consent was waived by the Ordu University Faculty of Medicine Clinical Research Ethics Committee due to the retrospective nature of this study.

Patient Selection and Data Preparation

In this study, the panoramic radiographs of patients over 18 years of age who applied to Ordu University, Faculty of Dentistry, Department of Dentomaxillofacial Radiology, for various reasons were retrospectively analyzed. Panoramic radiographs were obtained using Planmeca Promax 2D S3 (66 kVp; 8 mA; 15.8 s).

Images with at least one jaw lesion in the upper and/or lower jaw were selected and saved in JPEG format by two dentomaxillofacial radiology specialists. Radiographs were collected anonymously without considering personal information such as name, gender, and address. Only preoperative radiographs were included, and postoperative radiographs were excluded. The recorded radiographs were divided into two separate datasets according to radiopaque and radiolucent lesion content.

The radiolucent lesion dataset included radicular cysts, dentigerous cysts, odontogenic keratocysts, ameloblastomas, residual cysts, and lesions thought to be Stafne bone cavities, while the radiopaque lesion dataset included lesions thought to be idiopathic osteosclerosis, bone dysplasia, and odontomas. As the definitive diagnosis was based on a histopathological assessment, the lesions were not classified according to their specific type.

A total of 2371 panoramic radiographs with lesions were divided into two separate datasets as radiolucent and radiopaque. There were 1325 panoramic radiographs and 1561 lesions in the radiolucent lesion dataset and 1046 panoramic radiographs and 1323 lesions in the radiopaque lesion dataset.

The lesions were classified according to their localization as maxilla anterior, maxilla posterior, mandible anterior, and mandible posterior. The anterior region was determined to include the canine. If the lesion was located in both anterior and posterior regions, the estimated center of the lesion was taken into consideration. The number of panoramic radiographs and labels used in the study are given in Table 1.

Each radiograph was manually labeled by a radiology specialist by drawing the lesion borders and identifying the lesion site. Colabeler software was used for labeling. Colabeler is a free, open-source application designed for the labeling of positioning and classification.

Deep Learning Models and Training Process

YOLOv8 is a cutting-edge real-time object detection model that belongs to the YOLO model family. It represents an extension of YOLOv5, designed to enhance performance through the integration of advanced features and

Table 1 Number of panoramic radiographs and labels in this study

Lesion	Panoramic radiographs	Labels	Number of labels by localization			
			Anterior maxilla	Posterior maxilla	Anterior mandible	Posterior mandible
Radiolucent	1325	1561	393	142	189	837
Radiopaque	1046	1323	35	139	111	1038

optimizations. YOLOv8 offers a range of model variants, including 8n, 8s, 8m, 8l, and 8x. While these models share the same functional structure, they differ in terms of model size, computational speed, and the number of trainable parameters, increasing progressively toward the 8x variant [15]. In this study, the xlarge-sized 8x version was selected for analysis. The model and its corresponding configuration were retrieved from the Ultralytics GitHub repository and deployed on the Google Colab platform (available at https://github.com/ultralytics/ultralytics).

The YOLOv8 framework supports multiple tasks, including object detection, segmentation, and image classification. Object detection is performed through the prediction of bounding boxes, whereas segmentation enables the localization of object boundaries at the pixel level. Notably, YOLOv8 is capable of executing both tasks simultaneously [16]. In this study, YOLOv8 is employed for the detection and segmentation of both radiolucent and radiopaque lesions. After performing the necessary preprocessing steps—such as label format adjustments and index alignment—the training procedure is initiated using predefined parameters, including batch size, learning rate, optimization method, and number of epochs.

To assess the model's adaptability across different clinical conditions, three distinct experimental scenarios were established for both radiolucent and radiopaque lesion categories, resulting in six independent training and evaluation pipelines.

- In the first scenario, the model was trained using lesion images without localization labels.
- In the second, the dataset from the first scenario was extended by including additional background-only images from a separate lesion dataset, enabling the evaluation of the model under more generalized conditions.
- The third scenario introduced four spatial localization labels during training.

The dataset was partitioned into 70% for training, 15% for validation, and 15% for testing.

In addition to the YOLOv8 model, which was used for both object detection and segmentation tasks, this study incorporated three additional deep learning architectures to expand the evaluation framework and provide a more comprehensive comparison across different model families and design paradigms.

Specifically, YOLOv11, an advanced iteration of the YOLO series, was integrated for both detection and segmentation tasks. Compared to earlier versions, YOLOv11 introduces enhancements such as refined spatial attention mechanisms, deeper convolutional backbones, and more efficient learning strategies. These innovations contribute to improved performance in complex clinical images characterized by overlapping structures or subtle lesion margins. The YOLOv11x variant was selected for its favorable balance between accuracy and computational cost.

To introduce architectural diversity and examine region-based approaches, Mask R-CNN was also implemented. Unlike the YOLO series, Mask R-CNN employs a two-stage process that first proposes candidate object regions and then performs instance-level segmentation. In this study, Mask R-CNN was used solely for segmentation, providing detailed pixel-wise delineation of lesion shapes for both radiolucent and radiopaque classes.

Additionally, to represent transformer-based architectures, RT-DETR-L (Real-Time Detection Transformer–Large) was adopted for object detection only. By leveraging attention mechanisms and global context modeling, RT-DETR offers advantages in handling complex spatial dependencies and detecting visually similar structures. Its inclusion adds a novel architectural dimension that complements the CNN-based models.

By incorporating models with distinct functionalities—dual-task models (YOLOv8, YOLOv11), segmentation-only (Mask R-CNN), and detection-only (RT-DETR-L)—this study significantly broadens its analytical scope. This diversified model ensemble enables a more robust evaluation of detection and segmentation performance across architectures rooted in CNN and transformer paradigms.

Despite their strengths, each model has limitations. For example, YOLO-based models are highly efficient but may underperform in cases with occluded or overlapping objects, whereas transformer-based models like RT-DETR require substantial computational resources and longer training times. Mask R-CNN, while accurate, is slower and more complex to deploy in real-time applications. Therefore, selecting an appropriate model should be guided by the specific requirements of the clinical

task, dataset characteristics, and available computational resources [17].

Furthermore, to ensure the statistical soundness of performance comparisons, pairwise evaluations were conducted using the paired *t*-test, with a significance threshold set at $p < 0.05$. This approach allowed for objective interpretation of differences in model performance and reinforced the reliability of the reported findings.

Performance Evaluation Method

In this study, model performance was systematically evaluated using widely accepted object detection and segmentation metrics. For object detection tasks, metrics such as precision, recall, F1-score, and mean average precision (mAP) were employed. These evaluations were conducted under an intersection over union (IoU) threshold of 0.7, which defines the minimum overlap required between predicted and ground truth bounding boxes for a detection to be considered correct. Precision, recall, and F1-score were calculated based on true positive (TP), false positive (FP), and false negative (FN) detections, whereas true negatives are typically excluded in object detection evaluations. Additionally, mAP@0.5 and mAP@0.5:0.95 metrics were reported to provide a robust assessment of detection accuracy across a range of IoU thresholds (from 0.5 to 0.95 in 0.05 increments), reflecting both detection and localization performance comprehensively.

In the context of lesion segmentation, particularly in dental radiographs, it is crucial to quantitatively assess the alignment between predicted masks and ground truth annotations. To this end, precision, recall, and IoU were adopted as the primary performance metrics. These metrics offer not only a numerical summary of model performance but also meaningful insights into different types of segmentation errors. Figure 1 visualizes several representative segmentation scenarios along with their corresponding metric values, enhancing interpretability.

Precision, in the segmentation context, indicates the proportion of correctly identified lesion pixels among all pixels predicted as lesions. A high precision suggests a low rate of false positives. For example, in the radiopaque lesion of Patient 76, precision reaches 99.04%, reflecting an almost perfect overlap between the predicted and actual lesion regions, characterized by a dominant yellow intersection and minimal false positives. Conversely, in Patient 14, precision drops to 14.21% due to the model erroneously labeling large non-lesion areas as lesions, resulting in widespread red (false positive) regions.

Recall measures the proportion of actual lesion pixels that are successfully detected by the model. A high recall indicates comprehensive coverage of the lesion area, even in the presence of over-segmentation. For instance, although Patient 14 exhibits low precision, its recall is relatively high (95.22%), demonstrating that most of the true lesion area is captured. On the other hand, in the radiolucent lesion of Patient 169, recall is low (20.01%), implying that a significant portion of the lesion remains undetected, despite achieving relatively high Precision.

IoU, which integrates both precision and recall, provides a stricter measure of segmentation accuracy by quantifying the ratio of the intersection to the union of predicted

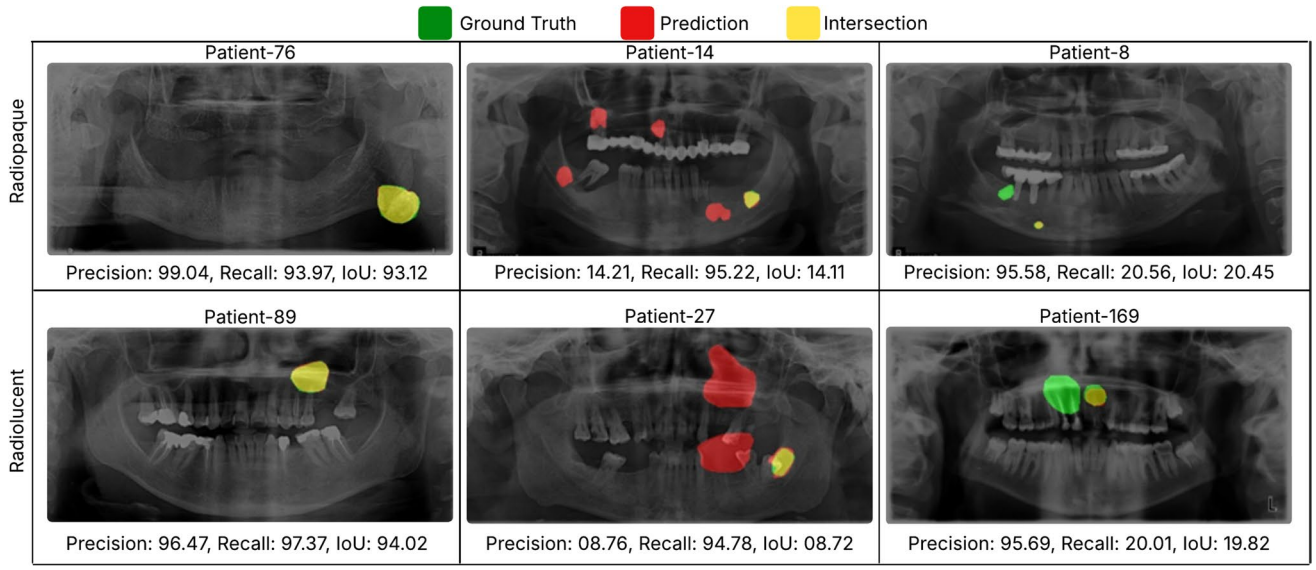


Fig. 1 Segmentation results with precision, recall, and IoU for radiopaque and radiolucent lesions, illustrating accurate, over-, and under-segmentation cases

and ground truth masks. High IoU values, such as those observed in Patients 76 and 89 (both exceeding 93%), reflect high-quality segmentation with minimal errors. In contrast, low IoU values, as seen in Patients 14 and 27 (e.g., 8.72%), arise from the combined effects of false positives and missed lesion regions.

Overall, these findings underscore the importance of jointly interpreting precision, recall, and IoU when evaluating segmentation performance. High values across all three metrics typically signify accurate and reliable segmentation, whereas imbalances may point to over- or under-segmentation tendencies. The visual examples in Fig. 1 provide additional clarity, especially in clinical contexts where precision in lesion delineation is of utmost importance.

To assess whether the observed differences in performance across models were statistically significant, paired *t*-tests were conducted for each evaluation metric. A significance threshold of $p < 0.05$ was applied.

Results

To evaluate the effectiveness of the constructed dataset, experiments were conducted under three different scenarios using state-of-the-art detection and segmentation architectures. The results of these experiments are presented in the following subsections, organized by scenario.

Scenario 1

In the first scenario, the models were trained separately on the radiolucent and radiopaque datasets based solely on lesion type, without any spatial localization annotations. At the end of training, the resulting segmentation models were evaluated on the corresponding test sets, and their performance metrics are presented in detail in Table 2, categorized by lesion type.

According to the results shown in the right section of Table 2, the YOLOv11x-seg model achieved the best performance on the radiolucent lesion test set, with 90% precision, 74% recall, 81% F1-score, 83% mAP@50, and 57% mAP@50–95. Similarly, for the radiopaque lesions presented in the left section of the table, the YOLOv8x-seg model outperformed the others, achieving 92% precision, 69% recall, 79% F1-score, 82% mAP@50, and 49% mAP@50–95.

The results of the detection models are summarized in Table 3. For radiolucent lesions, the YOLOv8x model yielded the highest detection performance with 96% precision, 75% recall, 84% F1-score, 87% mAP@50, and 62% mAP@50–95. On the other hand, for radiopaque lesions, the RT-DETR-L model provided the best results, with 86% precision, 67% recall, 75% F1-score, 79% mAP@50, and 51% mAP@50–95.

The confusion matrices and precision–recall curves corresponding to the segmentation and detection tasks are illustrated in Figs. 2 and 3, respectively. These visualizations represent the best-performing results as reported in Tables 2 and 3, providing deeper insight into the model's classification and localization capabilities.

Overall, the radiolucent lesion dataset demonstrated superior performance compared to the radiopaque dataset under Scenario I. In the segmentation task, the best-performing model on the radiolucent dataset was YOLOv11x-seg, achieving an mAP@50 of 83%, while YOLOv8x-seg attained 82% on the radiopaque dataset. For detection, the highest mAP@50 was obtained by YOLOv8x on the

Table 2 Segmentation performance metrics for Scenario I

Models	Radiopaque					Radiolucent				
	P	R	F1	mAP@**50**	mAP@50–95	P	R	F1	mAP@50	mAP@50–95
YOLOv8x-seg	**0.92**	**0.69**	**0.79**	**0.82**	**0.49**	0.93	0.69	0.79	0.82	0.59
YOLOv11x-seg	0.90	0.66	0.76	0.79	0.48	**0.90**	**0.74**	**0.81**	**0.83**	**0.57**
Mask R-CNN	0.70	0.74	0.72	0.68	0.45	0.66	0.67	0.66	0.70	0.54

The bold values indicate the most successful model for each scenario, considering overall performance

Table 3 Detection performance metrics for Scenario I

Models	Radiopaque					Radiolucent				
	P	R	F1	mAP@50	mAP@50–95	P	R	F1	mAP@50	mAP@50–95
YOLOv8x	0.83	0.68	0.75	0.78	0.46	**0.96**	**0.75**	**0.84**	**0.87**	**0.62**
YOLOv11x	0.84	0.65	0.74	0.77	0.45	0.93	0.72	0.81	0.84	0.58
RT-DETR-L	**0.86**	**0.67**	**0.75**	**0.79**	**0.51**	0.82	0.71	0.76	0.81	0.55

The bold values indicate the most successful model for each scenario, considering overall performance

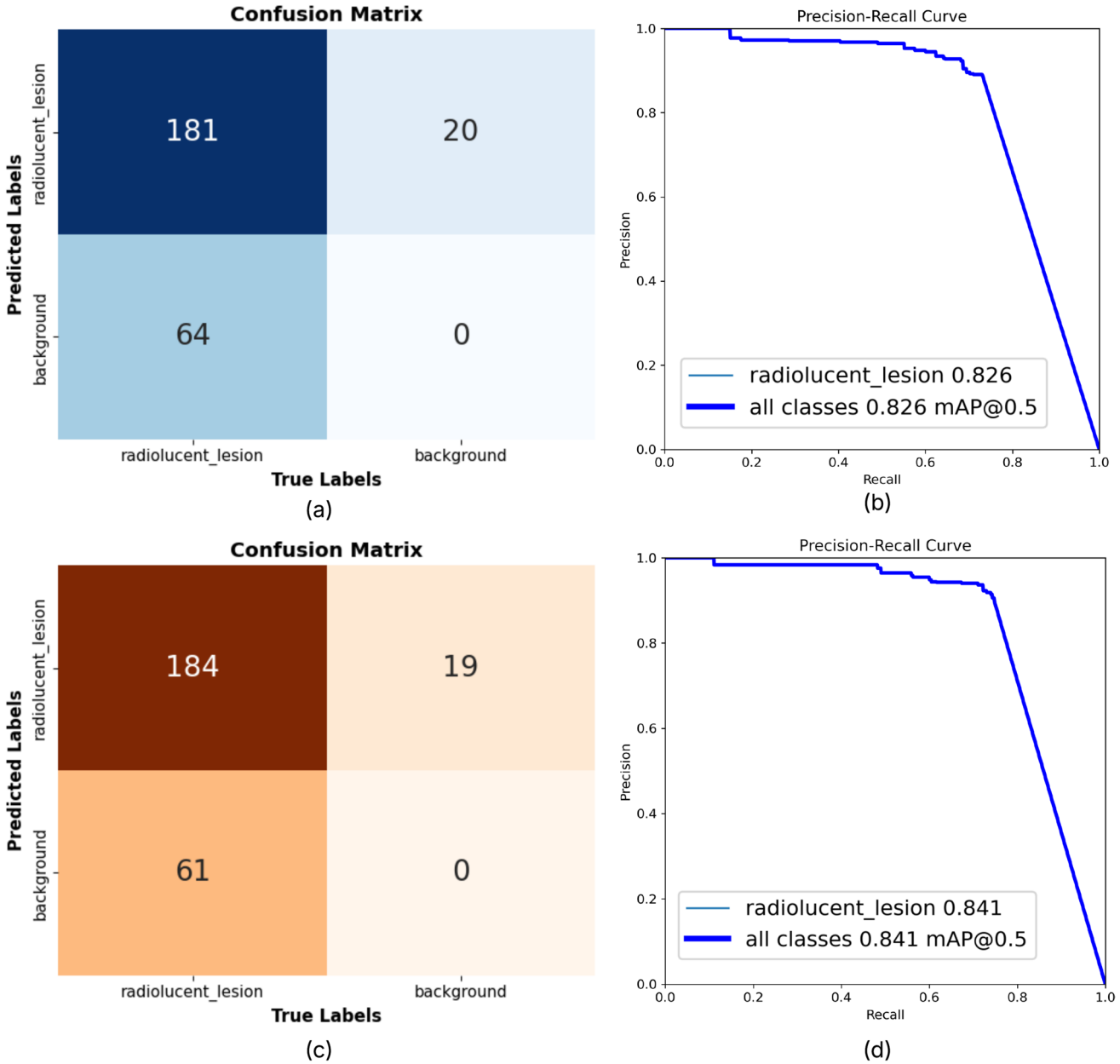


Fig. 2 Comparative visualization of radiolucent lesion analysis results: **a** confusion matrix and **b** precision–recall curve of the YOLOv11x-seg segmentation model; **c** confusion matrix and **d** precision–recall curve of the YOLOv8x-based detection model

radiolucent dataset (87%), followed by RT-DETR-L on the radiopaque dataset (79%). These results highlight the potential clinical applicability of the proposed models, particularly for radiolucent lesion analysis.

Scenario II

In the second scenario, additional background data were incorporated into the training phase to investigate whether the inclusion of such information could enhance the model's recall performance. Specifically, unlabeled images from the radiolucent dataset were added to the radiopaque training set, and vice versa, without providing lesion annotations. This strategy aimed to improve the model's ability to learn inter-class distinctions more effectively and reduce false negatives.

The segmentation results for this scenario are presented in Table 4. As seen in the table, the YOLOv8x-seg model achieved the best performance for both lesion types. For radiolucent lesions, the model reached 92% precision, 64% recall, 76% F1-score, 78% mAP@50, and 56% mAP@50–95, while for radiopaque lesions, it achieved 89%

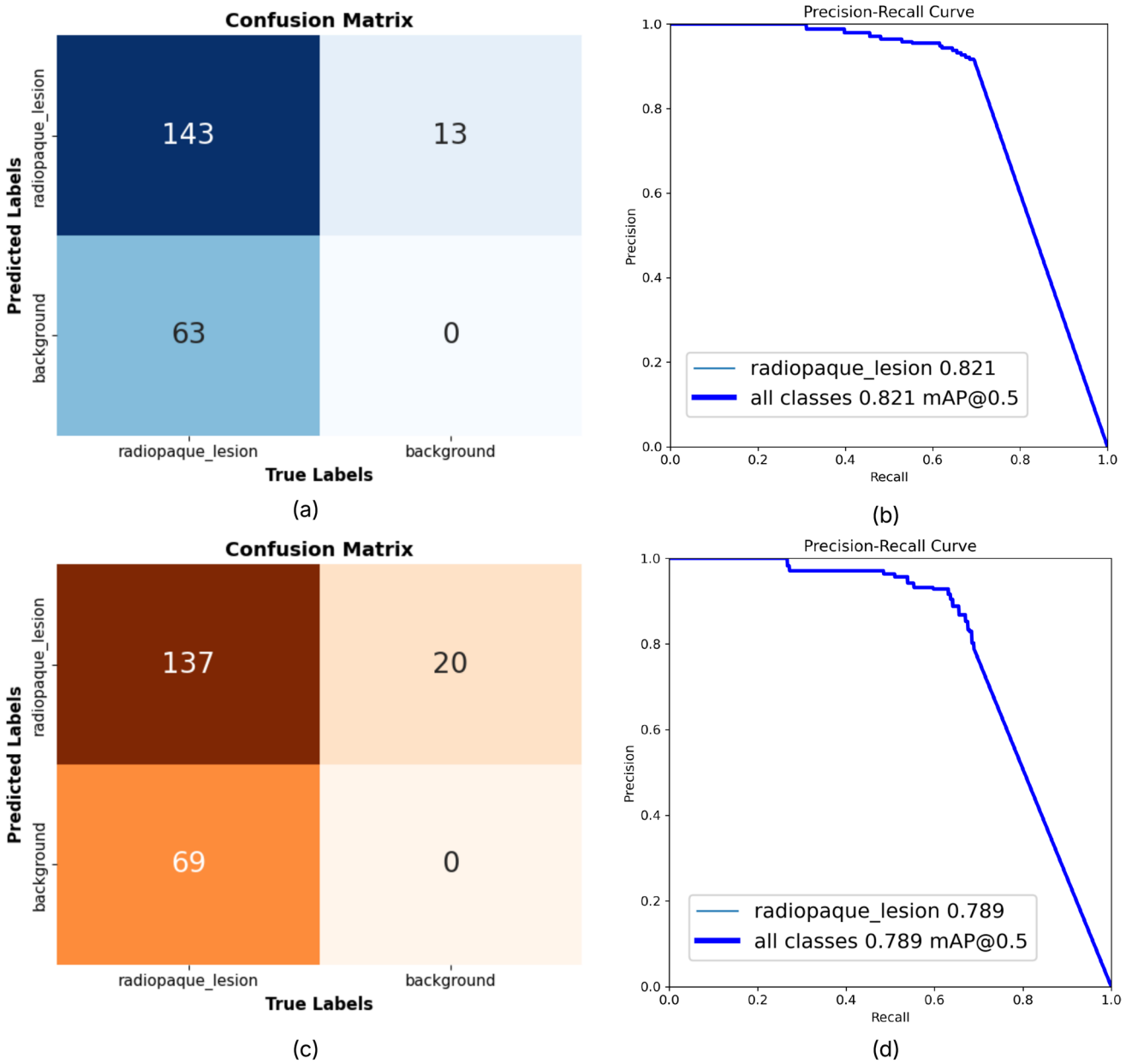


Fig. 3 Comparative visualization of radiopaque lesion analysis results: **a** confusion matrix and **b** precision–recall curve of the YOLOv8x-seg segmentation model; **c** confusion matrix and **d** precision–recall curve of the RT-DETR-L-based detection model

Table 4 Segmentation performance metrics for Scenario II

Models	Radiopaque					Radiolucent				
	P	R	F1	mAP@50	mAP@50–95	P	R	F1	mAP@50	mAP@50–95
YOLOv8x-seg	**0.89**	**0.65**	**0.75**	**0.79**	**0.46**	**0.92**	**0.64**	**0.76**	**0.78**	**0.56**
YOLOv11x-seg	0.79	0.52	0.63	0.68	0.40	0.91	0.48	0.63	0.71	0.49
Mask R-CNN	0.55	0.64	0.60	0.60	0.38	0.63	0.67	0.65	0.63	0.42

The bold values indicate the most successful model for each scenario, considering overall performance

Table 5 Detection performance metrics for Scenario II

Models	Radiopaque					Radiolucent				
	P	R	F1	mAP@50	mAP@50–95	P	R	F1	mAP@50	mAP@50–95
YOLOv8x	**0.91**	**0.57**	**0.70**	**0.76**	**0.45**	0.91	0.74	0.82	0.84	0.59
YOLOv11x	0.89	0.45	0.60	0.68	0.41	0.93	0.72	0.81	0.84	0.58
RT-DETR-L	0.63	0.47	0.54	0.54	0.29	**0.87**	**0.84**	**0.85**	**0.88**	**0.58**

The bold values indicate the most successful model for each scenario, considering overall performance

precision, 65% recall, 75% F1-score, 79% mAP@50, and 46% mAP@50–95.

The corresponding detection results are summarized in Table 5. On the radiolucent lesion test set, the RT-DETR-L model achieved the highest performance, with 87% precision, 84% recall, 85% F1-score, 88% mAP@50, and 58% mAP@50–95. For radiopaque lesions, the YOLOv8x model again outperformed the others, yielding 91% precision, 57% recall, 70% F1-score, 76% mAP@50, and 45% mAP@50–95.

A comparison of the results reveals that, relative to Scenario I, the RT-DETR-L model demonstrated a consistent improvement across all evaluation metrics for the radiolucent lesion dataset.

Figure 4 presents a representative radiolucent lesion example from Scenario II using YOLOv8x-seg, the best-performing segmentation model across both lesion types. In this case, the model successfully segmented a well-defined radiolucent lesion. Similarly, Fig. 5 illustrates a radiopaque lesion that was accurately identified by the same model.

Scenario III and Statistical Analysis

In the third scenario, four spatial localization annotations were added to the datasets used in the first scenario to provide detailed labeling of the lesions. These annotated datasets were then independently used to train each model, and the resulting models were evaluated for statistical comparison.

The models trained under each scenario were tested on the radiolucent and radiopaque test sets, and IoU values were calculated for each detected object. The test sets consisted of 200 images for radiolucent lesions and 157 images for radiopaque lesions. The collected IoU values were statistically compared across scenarios using a paired *t*-test, evaluated at a 95% confidence level ($p < 0.05$).

The summary of these comparisons is presented in Table 6. The table includes the lesion type, the scenario being compared, the models (Model 1 and Model 2), the mean IoU ± standard deviation for each model, the *p*-value (for hypothesis testing), the *t*-statistic (indicating

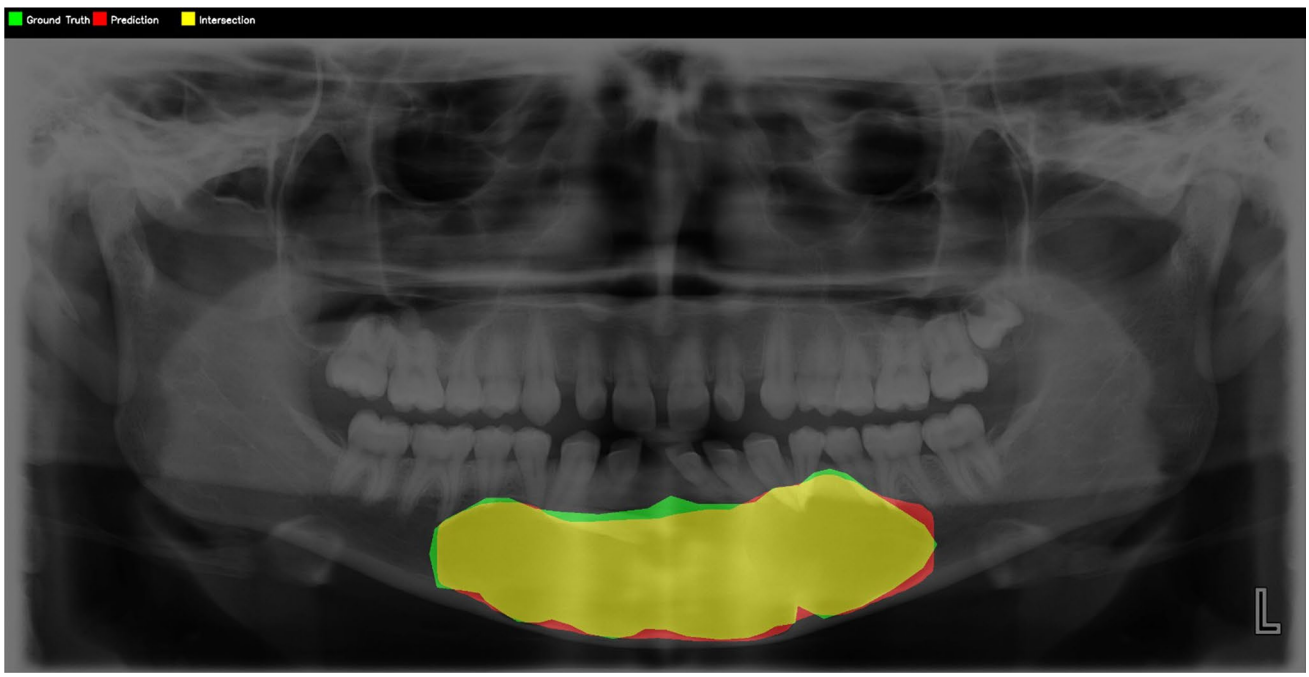


Fig. 4 Examples with success of detecting radiolucent lesion

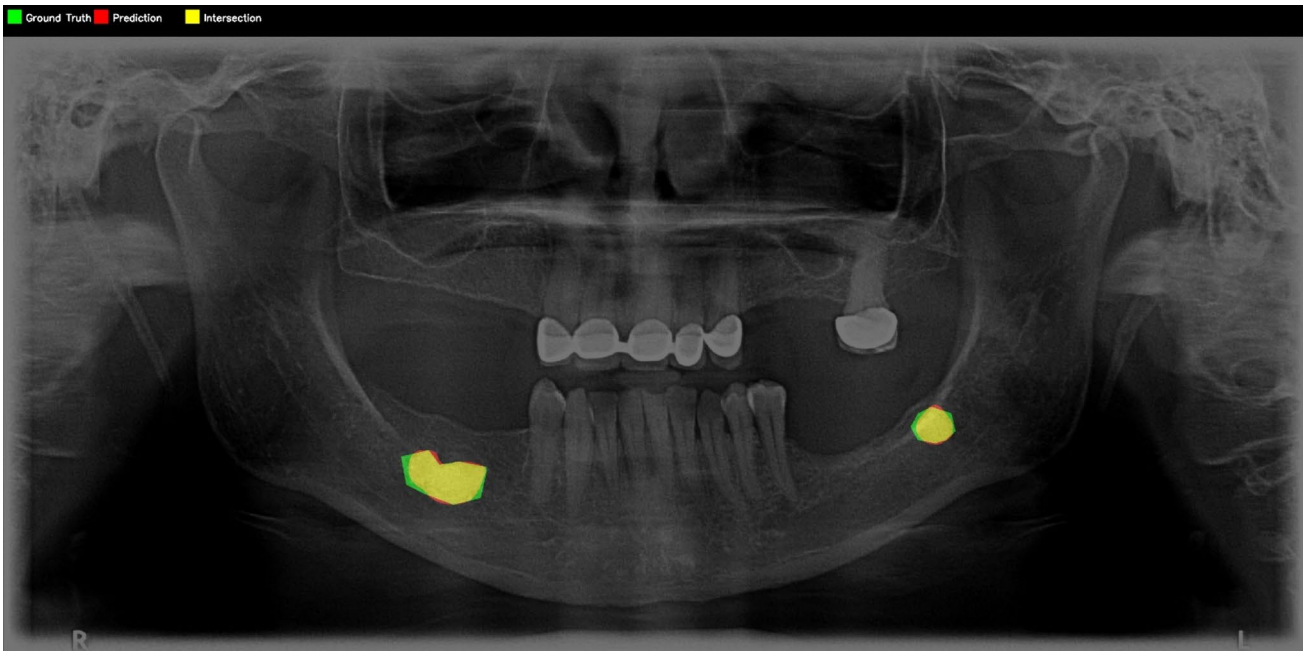


Fig. 5 Examples with success of detecting radiopaque lesions

Table 6 Statistical comparison of detection performance metrics across models and scenarios

Type	Scenario	Model 1	Mean ± std	Model 2	Mean ± std	*t*-Stat	*p*-value	Best
Radiopaque	Scenario I	RT-DETR-L	0.7000 ± 0.24	YOLOv8x	0.6662 ± 0.27	1.7923	0.07503	No
		RT-DETR-L	0.7000 ± 0.24	YOLOv11x	0.6389 ± 0.30	3.1071	0.00225	**RT-DETR-L**
		YOLOv8x	0.6662 ± 0.27	YOLOv11x	0.6389 ± 0.30	1.6029	0.11098	No
	Scenario II	RT-DETR-L	0.4978 ± 0.37	YOLOv8x	0.5772 ± 0.34	− 2.7497	0.00667	**YOLOv8x**
		RT-DETR-L	0.4978 ± 0.37	YOLOv11x	0.5756 ± 0.33	− 3.0431	0.00275	**YOLOv11x**
		YOLOv8	0.5772 ± 0.34	YOLOv11x	0.5756 ± 0.33	0.0786	0.93749	No
	Scenario III	RT-DETR-L	0.7135 ± 0.21	YOLOv8x	0.6318 ± 0.30	3.8628	0.00016	**RT-DETR-L**
		RT-DETR-L	0.7135 ± 0.21	YOLOv11x	0.6236 ± 0.30	4.3915	2e − 05	**RT-DETR-L**
		YOLOv8x	0.6318 ± 0.30	YOLOv11x	0.6236 ± 0.30	0.4342	0.66472	No
Radiolucent	Scenario I	RT-DETR-L	0.7578 ± 0.18	YOLOv8x	0.7281 ± 0.27	1.7359	0.08413	No
		RT-DETR-L	0.7578 ± 0.18	YOLOv11x	0.6959 ± 0.30	3.1589	0.00183	**RT-DETR-L**
		YOLOv8x	0.7281 ± 0.27	YOLOv11x	0.6959 ± 0.30	1.9803	0.04904	**YOLOv8x**
	Scenario II	RT-DETR-L	0.7694 ± 0.18	YOLOv8x	0.7281 ± 0.27	2.5695	0.01092	**RT-DETR-L**
		RT-DETR-L	0.7694 ± 0.18	YOLOv11x	0.6959 ± 0.30	3.8172	0.00018	**RT-DETR-L**
		YOLOv8x	0.7281 ± 0.27	YOLOv11x	0.6959 ± 0.30	1.9803	0.04904	**YOLOv8x**
	Scenario III	RT-DETR-L	0.7708 ± 0.19	YOLOv8x	0.7143 ± 0.25	4.0908	6e − 05	**RT-DETR-L**
		RT-DETR-L	0.7708 ± 0.19	YOLOv11x	0.7066 ± 0.27	4.0344	8e − 05	**RT-DETR-L**
		YOLOv8x	0.7143 ± 0.25	YOLOv11x	0.7066 ± 0.27	0.4602	0.64584	No

performance difference), and the "Best" column indicating the superior model in each comparison.

Based on the results of all scenario-wise comparisons, the RT-DETR-L model demonstrated the best performance in eight different cases, followed by YOLOv8x in three cases and YOLOv11x in one case. Notably, in Scenario III, RT-DETR-L consistently outperformed the other models across both lesion types.

The class-wise performance metrics obtained using the RT-DETR-L model in Scenario III are presented in Table 7 for radiolucent lesions and Table 8 for radiopaque lesions. The corresponding confusion matrices and

Table 7 Class-wise performance metrics of RT-DETR-L on radiolucent lesions in Scenario III

	Anterior maxilla	Posterior maxilla	Anterior mandible	Posterior mandible
Precision	0.91	0.88	0.54	0.74
Recall	0.42	0.70	0.61	0.80
F1-Score	0.57	0.78	0.57	0.77

Table 8 Class-wise performance metrics of RT-DETR-L on radiopaque lesions in Scenario III

	Anterior maxilla	Posterior maxilla	Anterior mandible	Posterior mandible
Precision	0.98	0.73	0.64	0.78
Recall	0.75	0.46	0.42	0.71
F1-Score	0.85	0.57	0.51	0.74

precision–recall curves are visualized in Figs. 6 (radiolucent) and 7 (radiopaque).

For the radiolucent dataset, the highest metric values were observed in the posterior maxilla class, with 88% precision, 70% recall, and 78% F1-score. The posterior mandible class, which contained the highest number of samples, achieved 74% precision, 80% recall, and 77% F1-score.

In the radiopaque dataset, the best performance was recorded for the anterior maxilla class, with 98% precision, 75% recall, and 85% F1-score. For the posterior mandible class, which had the largest number of instances, the performance metrics were 78% precision, 71% recall, and 74% F1-score.

Discussion

In the fields of medicine and dentistry, deep learning models have demonstrated noteworthy capabilities in a range of radiological tasks, including image segmentation, detection, and classification [1]. Algorithms with different models have been used in the studies. YOLO (You Only Look Once) is a deep learning tool for image detection. It is a real-time artificial intelligence model that simultaneously detects and classifies targeted objects within a single network [18]. In this study, both YOLOv8 and the recently proposed YOLOv11 models were employed to perform segmentation and detection tasks.

Additionally, Mask R-CNN, a two-stage detection framework well known for its strong instance segmentation performance, was also included due to its accuracy in delineating object boundaries. To further enhance detection capabilities, RT-DETR (Real-Time Detection Transformer), a transformer-based end-to-end object detector, was used. RT-DETR integrates spatial awareness and

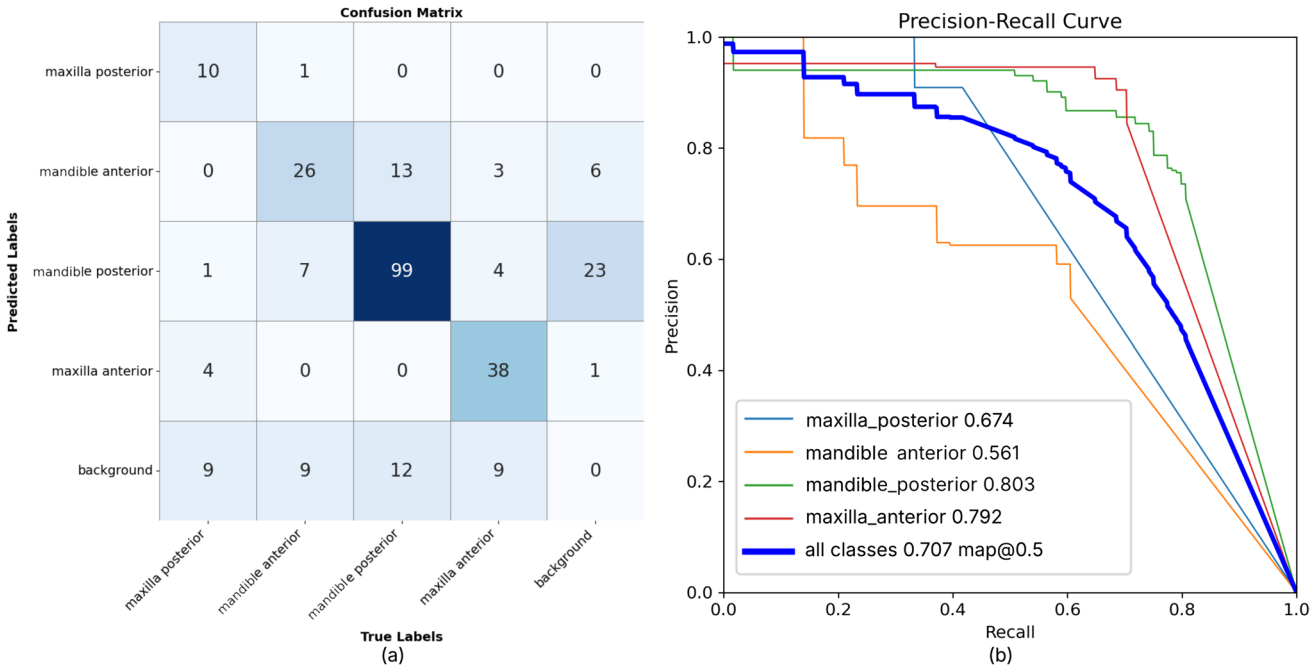


Fig. 6 Confusion matrix (**a**) and precision–recall curve (**b**) for the localization of radiolucent lesions using the RT-DETR-L model

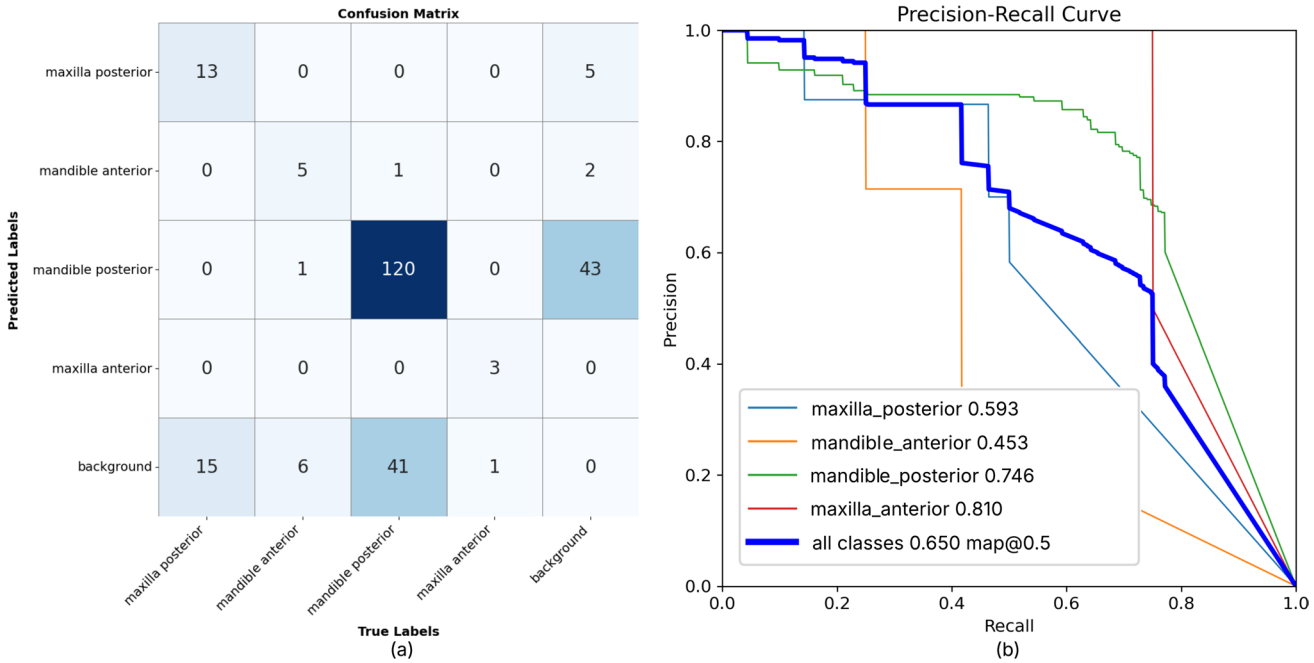


Fig. 7 Confusion matrix (**a**) and precision–recall curve (**b**) for the localization of radiopaque lesions using the RT-DETR-L model

contextual reasoning, making it suitable for complex medical imaging scenarios.

Recent advancements in object detection and segmentation architectures have led to significant improvements in both accuracy and computational efficiency. YOLOv8, for instance, offers notable enhancements in detection speed and precision, making it a strong candidate for real-time applications. The more recent YOLOv11 model builds upon this foundation with deeper architectural refinements, further improving performance in complex environments.

On the other hand, RT-DETR, a transformer-based architecture, provides end-to-end object detection capabilities with enhanced contextual reasoning and spatial awareness, which is particularly advantageous in scenarios involving subtle anatomical variations. Likewise, Mask R-CNN remains a widely adopted two-stage approach known for its high accuracy in instance segmentation, especially when precise object delineation is critical.

In recent years, studies evaluating the performance of artificial intelligence models for the detection of jaw lesions on panoramic radiographs have increased. These studies vary significantly in terms of dataset size, diversity of lesion classes, inclusion of only limited anatomical regions such as the maxilla or mandible, combinations of different deep learning architectures, and varying annotation strategies [9–12, 19, 20]. Despite achieving acceptable results, most existing studies have not adequately addressed radiopaque lesions and are often restricted to specific lesion types or jaw regions. This limitation constrains the generalizability and applicability of the models in clinical practice.

In this study, when evaluating the detection performance of radiolucent lesions, the YOLOv8x model achieved the highest performance with 96% precision, 75% recall, and an F1-score of 84%. The highest segmentation performance was demonstrated by the YOLOv11x model, with precision, recall, and F1-score values of 90%, 74%, and 81%, respectively. Several studies in the current literature have employed the YOLO algorithm. Yang et al. [11] evaluated the performance of the YOLOv2 algorithm for the early detection of odontogenic cysts and tumors on panoramic radiographs, using a dataset containing 1602 lesions. The YOLOv2 model (precision = 0.707, recall = 0.680) achieved accuracy scores close to human observers, with no statistically significant difference reported. Tajima et al. [10] developed an AI model using YOLOv3 to automatically detect cyst-like radiolucent lesions in the jaws on panoramic radiographs. In their study, 358 radiographic images were expanded 20-fold using data augmentation methods, resulting in 7160 training images. The maxillary molar region was excluded due to difficulties in distinguishing cyst-like lesions from the maxillary sinus. The reported results were 99% precision, 94.4% recall, and an F1-score of 0.966. In other studies using YOLOv3, Kwon et al. [9] applied data augmentation techniques and achieved 88.9% precision, 97.2% recall, and a 95.6% F1-score. Data augmentation methods such as rotation, scaling, shifting, and flipping are commonly used to enlarge training datasets and improve model generalizability.

Applying such techniques enhances model robustness and its ability to generalize to unseen data. Additionally, data augmentation plays a critical role in improving key accuracy metrics, including precision, recall, and IoU. Rasic et al. [12] evaluated 226 panoramic radiographs containing mandibular cysts using the YOLOv8 model. They reported that the resolution of their radiographs was higher than in previous studies. The precision rate increased from 0.918 before augmentation to 0.952 after applying augmentation methods. These findings indicate that the application of data augmentation and enhancement techniques can improve the performance outcomes of models in such studies.

In performance evaluations involving the inclusion of spatial localization annotations for radiolucent lesions, the highest accuracy was observed in the posterior maxilla region, followed by the posterior mandible, anterior maxilla, and anterior mandible regions, respectively. Watanabe et al. [20] reported that, in their study including only maxillary radiolucent lesions, detection performance in the anterior region was higher than in the posterior region. In addition, there are studies in the literature reporting that mandibular lesions have higher detection accuracy compared to maxillary lesions [10, 11]. In the current study, the fact that the posterior maxilla yielded the highest results may be explained by a model overfitting issue resulting from the relatively small number of samples in that region. In particular, in classes with a low number of samples, the model may exhibit overfitting tendencies, leading to deceptively high accuracy metrics. Additionally, the presence of well-defined lesion boundaries in this region may have contributed to the improved detection performance. Therefore, to evaluate the generalizability of the findings, further studies using more balanced, multicenter, and large-scale datasets are needed.

Although there are many studies evaluating radiolucent lesions with deep learning models, there are a limited number of studies evaluating radiopaque lesions. In the existing studies, only idiopathic osteosclerosis was included as radiopaque lesions. In the first study to our knowledge, Yesiltepe et al. [21] used GoogleNet architecture on 493 anonymized panoramic radiographs and found the recall, precision, and F-score for idiopathic osteosclerosis detection to be 0.88, 0.83, and 0.86, respectively. In another study, Tassoker et al. [22] attempted to detect idiopathic osteosclerosis using the YOLOv5 model on panoramic radiographs. In their study, they found the F1-score to be 0.954 and stated that AI algorithms were successful in detecting idiopathic osteosclerosis.

In this study, the highest detection performance for radiopaque lesions was achieved by the RT-DETR-L model, with precision, recall, and F1-scores of 86%, 67%, and 75%, respectively, while the highest segmentation performance was obtained by the YOLOv8x model, with values of 92% precision, 69% recall, and 79% F1-score. The relatively lower performance of radiopaque lesions compared to radiolucent lesions can be attributed to their similar contrast characteristics with the surrounding jaw bones. Although the radiopaque dataset was created without differentiating between radiopaque lesions such as cemento-osseous dysplasia, odontoma, and idiopathic osteosclerosis, idiopathic osteosclerosis constituted the majority due to its higher incidence. Idiopathic osteosclerosis lesions are continuous with the surrounding trabecular bone without a capsule and often present as opacities with indistinct borders. This may make them difficult to detect. In contrast, radiolucent lesions usually have good cortical borders.

In performance evaluations utilizing spatial localization annotations for radiopaque lesions, the highest scores were obtained in the anterior maxilla and posterior mandible regions. Although high scores were achieved in the anterior maxilla, the markedly low number of samples in this region may not accurately reflect the model's true performance. The lowest scores were observed in the anterior mandible, which may be attributed to the radiopaque appearance of the superimposed cervical vertebrae in this region.

To date, only two studies have simultaneously evaluated both radiolucent and radiopaque lesions on panoramic radiographs: the study by Ebata et al. [23] and a recent study by Tosun et al. [24]. Ebata et al. [23] included only the anterior region of the maxilla and used panoramic radiographic images of 400 patients (100 nasopalatine canal cysts, 100 radicular cysts, 100 impacted supernumerary teeth, 100 radiographs without lesions). Two deep learning systems (Systems 1 and 2) were created for the study, each with five models. For System 1, models were built and tested using radiolucent lesions (nasopalatine canal cyst and radicular cyst) and lesion-free datasets. To develop System 2, data of radiopaque lesions (impacted supernumerary tooth) were added to the dataset used in System 1. These systems used the YOLOv7 architecture. The recall, precision, and F1-score for System 1 were 0.76, 0.76, and 0.75, respectively. For System 2, these values were reported as 0.66, 0.68, and 0.66, respectively, and a decrease in performance was observed. They reported that this may be due to the difference in the image category (radiolucent/radioopaque) of the added data, while the inclusion of radiopaque lesions contributed to an increase in model performance in our study.

In a recent study by Tosun et al. [24], the effectiveness of deep learning models in detecting radiolucent and radiopaque lesions of the maxillofacial region on panoramic radiographs was evaluated. Their study included a total of 650 annotated lesions, comprising 415 radiolucent and 235 radiopaque lesions. For lesion classification, they employed AlexNet, VGG16, and GoogleNet architectures, while lesion localization and segmentation were performed using the YOLOv8 model. Although similar to our study in that radiolucent lesions were detected with higher accuracy than

radiopaque lesions, their comparatively lower YOLOv8 performance demonstrates the improved performance of our model.

Although Tosun et al. [24] evaluated the classification, detection, and segmentation performance of both radiolucent and radiopaque lesions, our study offers substantial contributions through the use of a larger dataset and methodological diversity. In present study, a total of 2884 lesions obtained from 2371 panoramic radiographs were annotated according to detailed anatomical regions, and four state-of-the-art deep learning architectures with distinct designs (YOLOv8, YOLOv11, Mask R-CNN, and RT-DETR) were comparatively evaluated under three different experimental scenarios. In contrast, Tosun et al.'s study did not include anatomical region-specific analysis and was limited to the evaluation of 650 lesions. Furthermore, our study involved a comprehensive analysis of model performance using a wide range of evaluation metrics and included statistical assessments. These methodological advantages enhance the potential generalizability of our study and provide preliminary insights into the clinical relevance of AI-based diagnostic systems.

The findings obtained in the second scenario indicate that the inclusion of additional background data during training positively contributed to model performance. Notably, the consistent improvement in recall and F1-score of the RT-DETR-L model for radiolucent lesions suggests that unlabeled images helped the model learn distinctions between classes and reduced false negatives. This implies that diversifying out-of-class samples may enhance the model's generalization ability and provide more balanced learning. Furthermore, the superior segmentation performance of the YOLOv8x model in Scenario II for both lesion types indicates that this architecture could be a robust option for detecting and segmenting both radiolucent and radiopaque lesions.

This study has several limitations. First, although a radiology archive spanning approximately 10 years was reviewed, the dataset remained limited due to the low prevalence of lesions, and no data augmentation methods were applied. Additionally, all radiographs were acquired from a single center using a single imaging device and were utilized as raw data without any enhancement. The literature indicates that contrast enhancement techniques can improve model performance on medical images [25]. Another limitation of our study is the binary classification of lesions solely as radiolucent or radiopaque. Due to the retrospective design and limited histopathological confirmation, more detailed subclass annotations (e.g., cyst, tumor, fibro-osseous lesion) were not feasible. While this simplified classification provided a balanced basis for comparing lesions with different radiodensity profiles, it may have restricted the clinical specificity in distinguishing complex or borderline cases. Therefore, future studies should employ larger, multicenter datasets with histopathologically confirmed cysts, tumors, and other fibro-osseous subtypes and evaluate multi-class or hierarchical model architectures.

In addition, no size restrictions were imposed when lesions were included in the study. YOLO-based models face challenges when detecting objects with varying aspect ratios and sizes. YOLO divides the input into a fixed-size grid and attempts to detect a limited number of objects within each grid cell. Consequently, small lesions may be split across multiple grid cells, complicating accurate detection. Moreover, fine details of small objects can be lost in the deeper layers of YOLO architectures. Large lesions, on the other hand, may trigger multiple grid cells to incorrectly detect non-existent lesions, contributing to an elevated false negative rate. Although YOLOv8 and the more recent YOLOv11 models introduce several architectural and training improvements to mitigate these limitations, difficulties in detecting very small or very large lesions can still persist [26].

Transformer-based models like RT-DETR offer advantages by leveraging global contextual information and spatial attention mechanisms, enabling better detection of objects across scales and complex backgrounds. However, these models require substantial computational resources and longer training times.

Mask R-CNN, as a two-stage detector with a dedicated segmentation branch, excels in accurately delineating object boundaries, particularly for irregularly shaped lesions. Nevertheless, it tends to be computationally heavier and slower than single-stage detectors, which may limit its usability in real-time settings.

These challenges across different architectures can be alleviated through strategies such as data augmentation, multi-scale training, architectural optimizations, and advanced post-processing techniques. Addressing these limitations may support the development of models suitable for future clinical applications.

The findings highlight both the strengths and limitations of current models in handling varying lesion types and spatial complexities. This work also provides critical insights into dataset design and model selection, offering a foundation for future research.

Further enhancements could be realized by expanding the dataset to include a broader range of lesion types, improving class balance, and incorporating additional imaging modalities such as CBCT. These steps would likely improve diagnostic accuracy and broaden the applicability of these models across diverse clinical environments.

Conclusion

This study demonstrates the potential of state-of-the-art deep learning models, including YOLOv8, YOLOv11, Mask R-CNN, and RT-DETR, for effective detection of lesions in

panoramic radiographs. While these models offer valuable support to clinicians in lesion evaluation, it is recommended that they be employed primarily as decision support tools within clinical workflows, rather than as standalone diagnostic systems.

Future integration of deep learning–based medical image analysis systems into hospital Picture Archiving and Communication Systems (PACS) holds promise for streamlining diagnostic workflows and enhancing accessibility. However, the widespread clinical adoption of such technologies will necessitate clear regulatory frameworks addressing data privacy, model validation, and accountability to ensure patient safety and reliable performance.

Acknowledgements The study was produced from a thesis.

Author Contribution The authors confirm contribution to the paper as follows: study conception and design: Dilek Çoban, Yasin Yaşa, Abdulsamet Aktas, Hamza Osman İlhan; data collection: Dilek Çoban, Yasin Yaşa; analysis and interpretation of results: Abdulsamet Aktas, Hamza Osman İlhan; draft manuscript preparation: Dilek Çoban. All authors reviewed the results and approved the final version of the manuscript.

Data Availability The data can be obtained upon request to the authors.

Declarations

Ethics Approval This study was approved by the Ordu University Faculty of Medicine Clinical Research Ethics Committee (Decision No. 2023/285) and was conducted in accordance with the Declaration of Helsinki.

Consent to Participate Not applicable. The requirement to obtain written informed consent was waived by the Ethics Committee due to the retrospective nature of this study.

Consent for Publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. Pierce RL, Van Biesen W, Van Cauwenberge D, Decruyenaere J, Sterckx S: Explainability in medicine in an era of AI-based clinical decision support systems. Front Genet, 13, 2023. https://doi.org/10.3389/fgene.2022.903600
2. Lim LZ, Padilla RJ, Reside GJ, Tyndall DA: Comparing panoramic radiographs and cone beam computed tomography: Impact on radiographic features and differential diagnoses. Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology, 126(1): 63-71, 2018. https://doi.org/10.1016/j.oooo.2018.03.019
3. Bispo, M. S., Pierre Júnior, M. L. G. D. Q., Apolinário Jr, A. L., Dos Santos, J. N., Junior, B. C., Neves, F. S., & Crusoé-Rebello, I: Computer tomographic differential diagnosis of ameloblastoma and odontogenic keratocyst: classification using a convolutional neural network. Dentomaxillofacial Radiology, 50(7): 20210002, 2021. https://doi.org/10.1259/dmfr.20210002
4. MacDonald D: Lesions of the jaws presenting as radiolucencies on cone-beam CT . Clin Radiol, 71: 972 – 85, 2016. https://doi.org/10.1016/j.crad.2016.05.018
5. Neyaz Z, Gadodia A, Gamanagatti S, Mukhopadhyay S: Radiographical approach to jaw lesions. Singapore Med J, *49*(2): 165-76, 2008.
6. Koivisto T, Bowles WR, Rohrer M: Frequency and Distribution of Radiolucent Jaw Lesions: A Retrospective Analysis of 9,723 Cases. Journal of Endodontics, 38(6): 729-732, 2012. https://doi.org/10.1016/j.joen.2012.02.028
7. Jain S: DeepSeaNet: Improving Underwater Object Detection using EfficientDet. In: 2024 4th International Conference on Applied Artificial Intelligence (ICAPAI). IEEE, p. 1–11, 2024. https://doi.org/10.1109/ICAPAI61893.2024.10541265.
8. Jiang P, Ergu D, Liu F, Cai Y, Ma B: A Review of Yolo algorithm developments. Procedia computer science, 199: 1066-1073, 2022. https://doi.org/10.1016/j.procs.2022.01.135
9. Kwon O, Yong TH, Kang SR, Kim JE, Huh KH, Heo MS, Lee SS, Choi SC, Yi WJ: Automatic diagnosis for cysts and tumors of both jaws on panoramic radiographs using a deep convolution neural network. Dentomaxillofacial Radiology, 49(8): 20200185, 2020. https://doi.org/10.1259/dmfr.20200185
10. Tajima S, Okamoto Y, Kobayashi T, Kiwaki M, Sonoda C, Tomie K, Saito H, Ishikawa Y, Takayoshi S: Development of an automatic detection model using artificial intelligence for the detection of cyst-like radiolucent lesions of the jaws on panoramic radiographs with small training datasets. Journal of Oral and Maxillofacial Surgery, Medicine, and Pathology, 34(5): 553-560, 2022. https://doi.org/10.1016/j.ajoms.2022.02.004
11. Yang H, Jo E, Kim HJ, Cha IH, Jung YS, Nam W, Kim D: Deep learning for automated detection of cyst and tumors of the jaw in panoramic radiographs. J. Clin. Med, 9(6): 1839, 2020. https://doi.org/10.3390/jcm9061839
12. Rašić M, Tropčić M, Karlović P, Gabrić D, Subašić M, Knežević P: Detection and Segmentation of Radiolucent Lesions in the Lower Jaw on Panoramic Radiographs Using Deep Neural Networks. Medicina, 59(12): 2138, 2023. https://doi.org/10.3390/medicina59122138
13. Tropčić M, Rašić M, Subašić M: YOLOv8 unleashed on orthopantomograms: deep learning approach for mandibular cyst diagnosis. In: 2024 47th MIPRO ICT and Electronics Convention (MIPRO), p. 37–42, 2024. https://doi.org/10.1109/MIPRO60963.2024.10569410
14. Shoorgashti R, Alimohammadi M, Baghizadeh S, Radmard B, Ebrahimi H, Lesan S: Artificial intelligence models accuracy for odontogenic keratocyst detection from panoramic view radiographs: a systematic review and meta-analysis. Health Sci Rep, 8(4): e70614, 2025. https://doi.org/10.1002/hsr2.70614
15. Terven J, Córdova-Esparza DM, Romero-González JA: A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. Machine Learning and Knowledge Extraction, 5(4): 1680-1716, 2023. https://doi.org/10.3390/make5040083
16. Bonfanti-Gris M, Herrera A, Paraíso-Medina S, Alonso-Calvo R, Martínez-Rus F, Pradíes G: Performance evaluation of three versions of a convolutional neural network for object detection and segmentation using a multiclass and reduced panoramic radiograph dataset. Journal of Dentistry, 144:104891, 2024. https://doi.org/10.1016/j.jdent.2024.104891
17. Avramidou FM, Markou E, Lambrianidis T: Cross-sectional study of the radiographic appearance of radiopaque lesions of the jawbones in a sample of Greek dental patients. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology, 106(3): e38-e43, 2008. https://doi.org/10.1016/j.tripleo.2008.04.020

18. Tassoker M, Öziç MÜ, Yuce F: Performance evaluation of a deep learning model for automatic detection and localization of idiopathic osteosclerosis on dental panoramic radiographs. Sci Rep, 14(1): 4437, 2021. https://doi.org/10.1038/s41598-024-55109-2
19. Ariji Y, Yanashita Y, Kutsuna S, Muramatsu C, Fukuda M, Kise Y, Nozawa M, Kuwada C, Fujita H, Katsumata A, Ariji E: Automatic detection and classification of radiolucent lesions in the mandible on panoramic radiographs using a deep learning object detection technique. Oral Surg Oral Med Oral Pathol Oral Radiol, 128(4): 424-430, 2019. https://doi.org/10.1016/j.oooo.2019.05.014
20. Watanabe H, Ariji Y, Fukuda M, Kuwada C, Kise Y, Nozawa M, Sugita Y, Ariji E: Deep learning object detection of maxillary cyst-like lesions on panoramic radiographs: preliminary study. Oral Radiology, 37(3): 487-493, 2021. https://doi.org/10.1007/s11282-020-00485-4
21. Yesiltepe S, Bayrakdar IS, Orhan K, Çelik Ö, Bilgir E, Aslan AF, Odabas A, Costa ALF, Jagtap R : A Deep Learning Model for Idiopathic Osteosclerosis Detection on Panoramic Radiographs. Med Princ Pract, 31(6): 555-561, 2022. https://doi.org/10.1159/000527145
22. Curé JK, Vattoth S, Shah R: Radiopaque Jaw Lesions: An Approach to the Differential Diagnosis. RadioGraphics, 32(7): 1909-1925, 2012. https://doi.org/10.1148/rg.327125003
23. Ebata K, Kise Y, Morotomi T, Ariji E: Performance of a deep learning system for simultaneously diagnosing radiolucent and radiopaque lesions in the anterior maxilla on panoramic radiographs. Oral Science International, 21(3): 385-392, 2024. https://doi.org/10.1002/osi2.1233
24. Tosun ZT, Kumbasar N, Sumbullu MA, Miloglu O: Evaluation of the effectiveness of artificial intelligence models in radiopaque and radiolucent lesions of the maxillofacial region on panoramic radiographs. Oral Radiology, 1–15, 2025. https://doi.org/10.1007/s11282-025-00838-x
25. Thomas J, Ulagamuthalvi V: Automatic Detection of Dental Cysts in Panoramic Radiography Images using Preprocessing Techniques and Convolutional Neural Networks. In: 2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT), p. 1–6, 2022. https://doi.org/10.1109/ICERECT56837.2022.10060124
26. Zhiyuan W, Jian H, Song Z: A Feature Selection and Enhanced Reuse Framework for Detection and Classification of Dental Diseases in Panoramic Dental X-Rays Images. IEEE Access, p. 70741–70751, 2025. https://doi.org/10.1109/ACCESS.2025.3561362

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.