Odontology
https://doi.org/10.1007/s10266-026-01409-0

ORIGINAL ARTICLE

Multi-class anatomical landmark detection in periapical radiographs with deep learning

Cansu Buyuk[1] · Alperen Saruhan[2] · Fatma Yuce[3] · Ozer Celik[4] · Elif Bilgir[5] · İbrahim Sevki Bayrakdar[5]

Received: 13 February 2026 / Accepted: 26 April 2026

Abstract

This study aims to develop a deep learning model for the detection and segmentation of multiple anatomical landmarks on periapical radiographs from the maxilla and mandible. A total of 1930 paralleling-technique periapical radiographs with 21 annotated anatomical landmarks were divided into training (80%), validation (10%), and test (10%) sets. Geometry-preserving preprocessing was applied before dataset splitting, while appearance-based augmentation was performed exclusively on the training subset after the split. A YOLOv8x-seg architecture was trained for multi-class detection and instance segmentation. Performance was evaluated using precision, recall, F1-score, Dice coefficient, Intersection-over-Union, mean average precision, and receiver operating characteristic analysis. The model demonstrated stable training and consistent performance. Overall precision, recall, and F1-score were 0.820, 0.725, and 0.769, respectively, with an overall Dice coefficient of 0.621. High detection accuracy was achieved for well-defined structures such as the maxillary sinus, nasal fossa, nasal fossa floor, and nasal septum, whereas low-contrast landmarks showed reduced performance. Confidence-dependent analysis indicated optimal performance at low confidence thresholds (approximately 0.05–0.10). In conclusion, the proposed model effectively detected major anatomical landmarks on periapical radiographs while demonstrating expected limitations for small or low-contrast structures. Despite substantial anatomical variability across maxillary and mandibular regions, anterior–posterior sites, and projection-dependent appearances of similar structures, these findings demonstrate that deep learning can reliably identify key anatomical landmarks, supporting safer, more consistent, and clinically meaningful radiographic interpretation in routine dental practice.

Keywords Periapical radiography · Anatomical landmarks · Deep learning · Image segmentation · Artificial intelligence

Introduction

Intraoral radiographs are fundamental diagnostic tools in dentistry, providing high-resolution visualization of roots, periapical tissues, and surrounding anatomical structures that cannot be assessed clinically. Compared with extraoral imaging, periapical radiographs offer superior spatial resolution, minimal distortion, and nearly true-to-size representation, maintaining their central role in routine diagnosis despite major advances in digital technologies and three-dimensional imaging modalities [1]. A typical periapical image visualizes 3–5 teeth along with trabecular bone patterns and critical anatomical landmarks—including the inferior alveolar canal, maxillary sinus floor, mental foramen, and lateral fossa—together with potential pathologies [2].

Accurate interpretation of these radiographs requires detailed knowledge of maxillofacial anatomy.

✉ Cansu Buyuk
cansu.buyuk@okan.edu.tr

1 Department of Oral and Maxillofacial Radiology, Faculty of Dentistry, Istanbul Okan University, Aydıntepe Mah. Prof. Dr. Necmettin Erbakan Cad. No:2, Istanbul 34947, Turkey

2 Private Dental Practice, Izmir 35390, Turkey

3 Department of Oral and Maxillofacial Radiology, Faculty of Dentistry, Istanbul Kent University, Istanbul 34421, Turkey

4 Department of Mathematics and Computer Science, Eskisehir Osmangazi University, Eskisehir 26040, Turkey

5 Department of Oral and Maxillofacial Radiology, Faculty of Dentistry, Eskisehir Osmangazi University, Eskisehir 26040, Turkey

Published online: 21 May 2026

Misinterpretation may arise from the inherent limitations of two-dimensional projection and the complex superimpositions of craniofacial structures [3]. Clinically relevant challenges include distinguishing the mental foramen from periapical radiolucencies; identifying the relationship between molar roots and the maxillary sinus to prevent oroantral communication, particularly during procedures such as tooth extraction; and assessing the proximity of surgical sites to the mandibular canal to avoid neurovascular injury [4–6].

Panoramic imaging provides a broad overview of the maxillofacial region; however, it is limited by magnification, distortion, and superimposition. In contrast, periapical radiography offers higher spatial resolution and remains a primary tool for detailed local assessment, although it is inherently constrained by its two-dimensional nature [7]. Cone beam computed tomography (CBCT) imaging can overcome these limitations by providing volumetric visualization and improved assessment of complex anatomical relationships; however, due to its higher radiation dose compared to two-dimensional imaging, it cannot be routinely used in all clinical cases [8]. Importantly, the high variability of anatomical structures in the oral and maxillofacial region further contributes to the complexity of radiographic interpretation [9]. Advanced imaging studies have also demonstrated that detailed three-dimensional assessment can significantly improve the understanding of anatomical relationships, particularly in regions such as the maxillary sinus, where vascularization patterns and structural variations directly impact surgical safety [10]. These challenges highlight the need for supportive tools capable of enhancing anatomical awareness in routinely used imaging modalities such as periapical radiographs.

Recent advances in artificial intelligence (AI) and deep learning (DL)—with more than 1000 FDA-cleared AI/ML tools reported across medical fields, and a growing subset specifically developed for dental imaging, including approximately 29 FDA-cleared platforms—have demonstrated excellent performance in classification, detection, and segmentation tasks by learning texture, density, and shape patterns directly from data [11–14]. DL-based models have been widely applied in dentistry, particularly for detecting caries, periodontal bone loss, periapical lesions, and various radiographic abnormalities [14, 15]. Studies focused on anatomical landmark identification have primarily used panoramic radiography or CBCT, reporting promising results for automated localization of structures relevant to orthodontics, surgery, and implantology [16, 17].

Despite their diagnostic importance, periapical radiographs present marked anatomical variability across maxillary and mandibular regions, anterior to posterior sites, and projection-dependent appearances of similar anatomical structures. In addition to limited field of view, this heterogeneity poses a substantial challenge for automated anatomical interpretation, as identical landmarks may appear with different shapes, contrasts, and spatial relationships depending on tooth region and projection geometry.

However, DL applications targeting anatomical landmark recognition specifically on periapical radiographs remain relatively limited. Existing periapical studies have largely focused on disease detection [18–22], while anatomical interpretation—which is essential for surgical planning, endodontics, implant procedures, and avoidance of iatrogenic complications—has been understudied. To the best of our knowledge, no prior study has developed a DL-based program capable of detecting a comprehensive set of anatomical landmarks on periapical images.

From a clinical perspective, accurate identification of anatomical landmarks is essential in procedures such as endodontic treatment, apical surgery, and implant planning, where misinterpretation of normal anatomy may lead to diagnostic errors or procedural complications. Therefore, improving accurate recognition of anatomical structures in commonly used intraoral imaging techniques is of significant clinical importance. In this context, automated approaches capable of reliably detecting anatomical structures may support clinical decision-making, reduce interpretation variability, and enhance diagnostic confidence in daily practice.

Therefore, the aim of this study is to develop and validate a deep learning model for the automatic identification of key anatomical structures in periapical radiographs obtained from both anterior and posterior regions of the maxilla and mandible.

Materials and methods

An overview of the dataset preparation, preprocessing, training, and evaluation workflow is illustrated in Fig. 1.

Data collection and selection

This retrospective study was conducted using periapical radiographs obtained from patients who presented to the Department of Oral and Maxillofacial Radiology, Faculty of Dentistry, Istanbul Okan University, between September 2018 and September 2021. Approximately 4000 periapical radiographs archived in the institutional Picture Archiving and Communication System (PACS) were screened by a senior dental student and a dentomaxillofacial radiologist team. Radiographs that fulfilled the inclusion criteria were anonymized and exported in JPEG format (886 × 1171 pixel size and 8-bit) for further analysis. Institutional ethical

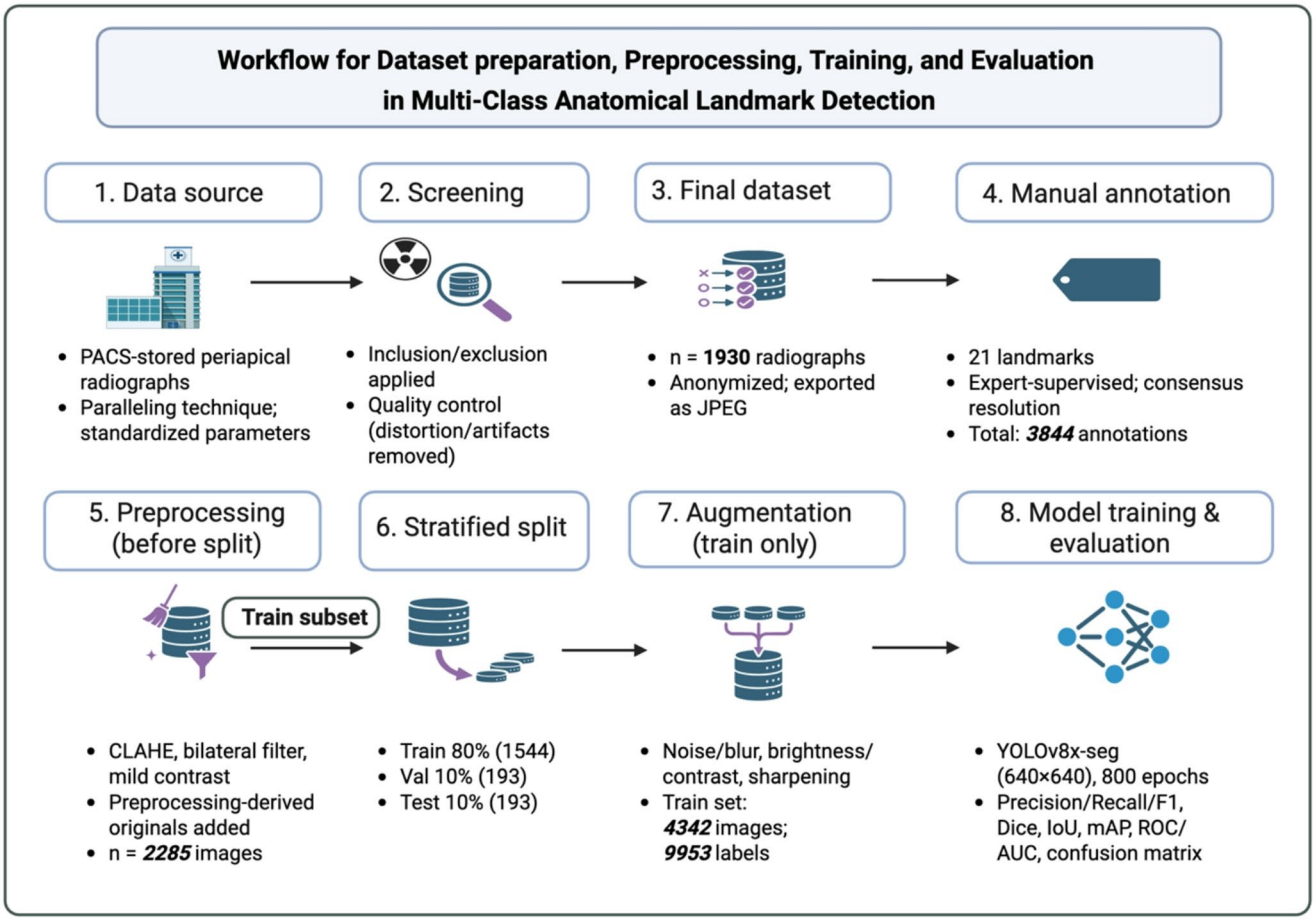

Fig. 1 Workflow illustrating dataset preparation, preprocessing, model training, and evaluation for multi-class anatomical landmark detection on periapical radiographs. Geometry-preserving preprocessing outputs were incorporated prior to stratified dataset splitting to prevent data leakage, whereas appearance-based augmentation was applied exclusively to the training subset

approval was obtained prior to data collection (Ethics Committee approval no: 2024/175-3).

Radiographs were included if they were acquired using the paralleling technique, obtained from patients aged 18 years or older, demonstrated no distortion related to cone misalignment, patient positioning, foreign objects, or motion, and met the standard exposure parameters (60 kV, 8 mA, 0.16 s). Images were excluded if they were not taken with the paralleling technique, belonged to patients younger than 18 years, contained orthodontic brackets, rubber dam, or other materials obscuring anatomical details, exhibited anatomical alterations caused by periapical pathologies such as cysts, resorptive defects, or large lesions, showed compromised diagnostic quality due to superimposition or imaging artifacts, or were acquired with exposure settings outside the defined standards.

Dataset composition and annotation

A total of 1930 periapical radiographs meeting the predefined inclusion criteria were retrieved from the institutional PACS archive after anonymization. All radiographs were obtained using standardized periapical imaging parameters and paralleling technique. Images were independently reviewed by a senior student to verify diagnostic adequacy and exclude suboptimal scans.

Each radiograph was manually annotated using DentalAi software (Osmangazi University, Eskişehir, Turkey). Annotation was performed by a trained operator under the supervision of an experienced dentomaxillofacial radiologist to ensure consistency in anatomical boundaries and labelling definitions. Annotation guidelines were predefined based on anatomical reference standards, and ambiguous cases were resolved through consensus meetings between the annotator and the supervising radiologist to maintain label uniformity. To further reduce annotation variability, landmark boundary definitions and shape criteria were reviewed in pilot

sessions before the full dataset was labelled. In total, 3844 anatomical structures were annotated across maxillary and mandibular regions.

Annotated maxillary structures included: incisive foramen, anterior nasal spine, median palatal suture, nasal septum, nasal fossa, nasal fossa floor, inferior concha, lateral fossa, lip soft tissue, soft tissue of the nose, maxillary sinus, maxillary sinus septum, zygomatic process, hamular notch, pterygoid process, and maxillary tuberosity. ***Annotated mandibular structures included***: coronoid process, external oblique ridge, mental foramen, genial tubercle, mandibular canal, mental fossa, mental ridge, mylohyoid ridge, nutrient canal, and submandibular fossa. Representative examples of annotated landmarks are shown in Fig. 2a–d.

The dataset was divided into training (80%, $n=1544$), test (10%, $n=193$), and validation (10%, $n=193$) subsets using stratified random sampling to preserve the distribution of anatomical regions and label frequencies across all partitions.

Preprocessing and data augmentation

A total of 1930 periapical radiographs containing 3844 labelled anatomical structures were initially collected. During preprocessing, geometry-preserving transformations—including contrast-limited adaptive histogram equalization (CLAHE), bilateral filtering, and mild contrast adjustments—generated additional image files without altering landmark locations. These outputs were exported as preprocessing-derived originals, increasing the dataset to 2285 images. Importantly, all preprocessing-derived originals were added before splitting the dataset into training, validation, and test subsets to prevent any leakage between partitions.

Subsequently, appearance-based augmentation techniques (Gaussian blur, Gaussian noise, ISO noise, multiplicative noise, brightness/contrast variation, snow effect, sharpening, and sepia filtering) were applied to enhance image diversity. In total, 4342 radiographs and 9953 labelled structures were used for training.

The increase from 1930 to 2285 images resulted from preprocessing-derived original images generated through geometry-preserving transformations. The subsequent increase to 4342 images reflects appearance-based augmented images applied exclusively to the training set and was not used to expand the validation or test datasets. To avoid data leakage, all geometry-preserving preprocessing steps were applied before dataset splitting, whereas appearance-based augmentations were applied exclusively to the training subset after the split. Validation and test sets remained unaugmented.

Model development

A deep learning model was implemented using the YOLOv8x-seg architecture (Ultralytics, USA), optimized

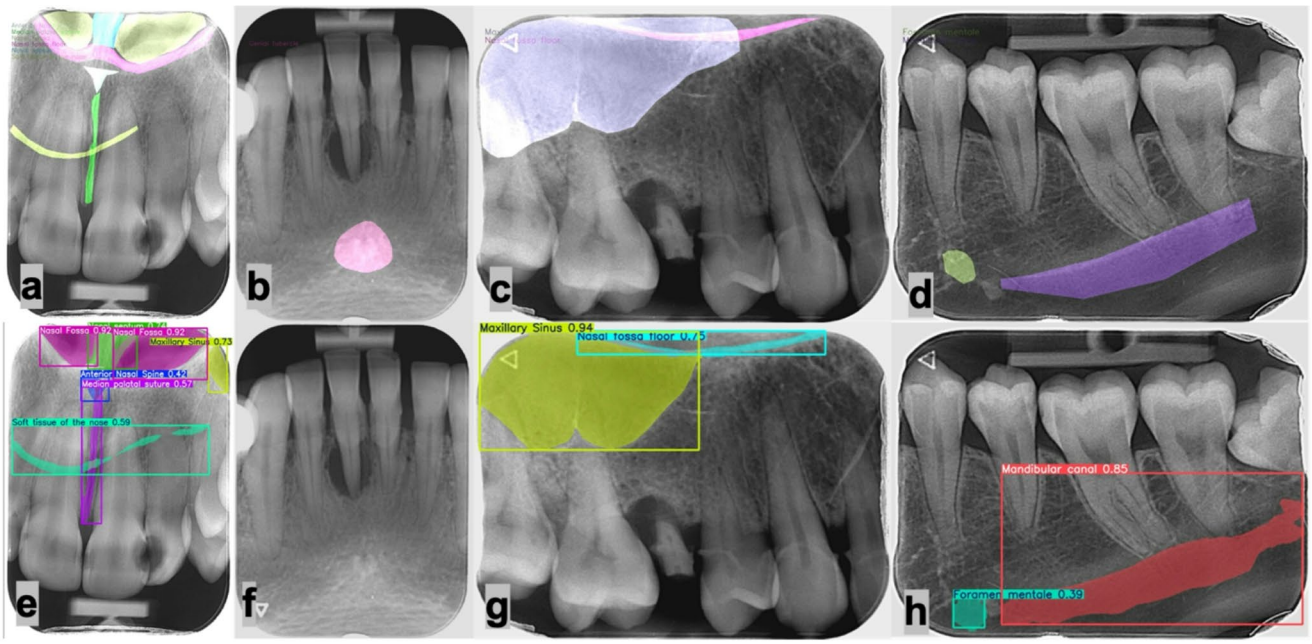

Fig. 2 Representative examples of ground-truth annotations (top row) and corresponding model predictions (bottom row) on periapical radiographs. Panels depict anterior maxilla (a/e), mandibular anterior (b/f), maxillary posterior (c/g), and mandibular posterior (d/h) regions. Confidence values correspond to model prediction scores. Colored overlays indicate instance-level segmentations, while bounding boxes represent detected anatomical landmarks. Numerical values denote model confidence scores for each prediction

for multi-class object detection and instance-level anatomical segmentation. Training was performed for 800 epochs at an image resolution of 640 × 640 pixels with a batch size of 4, leveraging pretrained weights. All images were originally exported as 8-bit JPEG files with a resolution of 886 × 1171 pixels. During both training and inference, images were resized by the YOLOv8 pipeline to 640 × 640 pixels to match the network input resolution. Key hyperparameters included: patience = 100, optimizer = auto, seed = 0, deterministic = True, pretrained = True. Training and visualization utilized OpenCV, PyTorch, NumPy, Pandas, TorchVision, TensorBoard, and Seaborn.

Model evaluation

Performance was evaluated on the reserved test set (10%, $n = 193$). Metrics included precision, recall, F1-score, and area under the ROC curve (AUC). A confusion matrix quantified true positives (TP), false positives (FP), and false negatives (FN) for each class. Segmentation performance was assessed via Intersection-over-Union (IoU / Jaccard index) and Dice coefficient.

Results

A total of 193 periapical radiographs and 403 annotated anatomical structures in the held-out test set were used for performance evaluation. The YOLOv8x-seg model demonstrated a stable training profile and convergent learning dynamics, supported by progressively decreasing training losses and consistent validation loss behavior across 800 epochs. Table 1 shows the overall performance of the model achieved across all anatomical classes.

Confidence-dependent analysis was performed using the test set. The maximum aggregated F1-score was obtained at a confidence threshold of ~0.09 (F1 ≈ 0.58), precision peaked at high thresholds (1.00 at confidence ≈ 0.96), and recall was highest at near-zero thresholds (≈ 0.70). These test-set patterns closely matched the trends observed during training, demonstrating consistent confidence-dependent model behavior. The corresponding curves are shown in Fig. 3.

Table 1 Summary of overall classification and segmentation performance metrics

Classification metrics	Value	Segmentation metrics	Value
Overall accuracy	0.625	Overall dice coefficient	0.6211
Overall precision	0.820	Overall IoU (Jaccard index)	0.5175
Overall recall	0.725	mAP@0.5	0.4895
Overall F1-score	0.769	mAP@0.5–0.95	0.2294

Class-wise performance

Class-specific metrics demonstrated substantial variability depending on anatomical complexity and representation frequency in the dataset (Table 2). For interpretability, anatomical structures were grouped into high-, moderate-, and low-performing categories using F1-score thresholds: high (> 0.70), moderate (0.40–0.70), and low (< 0.40).

High-performing structures

High-performing structures included the maxillary sinus, nasal fossa floor, nasal fossa, and nasal septum, all of which showed high precision–recall balance and stable performance across confidence levels. These findings align with the visual consistency of these structures across radiographs and their comparatively larger annotated areas.

Moderate-performing structures

Structures with smaller size or greater variability—such as mandibular canal (F1 = 0.642), the median palatal suture (F1 = 0.591), zygomatic process of the maxilla (F1 = 0.582), external oblique ridge (F1 = 0.444), and genial tubercle (F1 = 0.522)—demonstrated moderate performance. Their errors mainly stemmed from higher false-negative rates and partial or incomplete segmentations.

Low-performing structures

Small, low-contrast, or infrequently represented structures—such as soft tissue of the nose, the incisive foramen, mylohyoid ridge, tuber maxilla, inferior concha, lateral fossa, mental foramen, pterygoid process, and nutrient canal—exhibited limited detection performance. Many of these classes demonstrated F1-scores below 0.30, with high false-negative rates and frequent misclassification as background, reflecting both class imbalance and the intrinsic radiographic difficulty of visualizing these landmarks on periapical images. In particular, the absence of true-positive detections for the mental foramen likely reflects its small size, variable morphology, overlap with surrounding trabecular patterns, and natural underrepresentation in the dataset.

Receiver operating characteristic (ROC) analysis yielded class-wise AUC values ranging approximately from 0.74 to 0.98 across anatomical structures, with the highest separability observed for maxillary sinus and mandibular canal.

Confidence-based behavior

Confidence–precision–recall–F1 curves (Fig. 3) provided additional insights: Precision consistently increased with

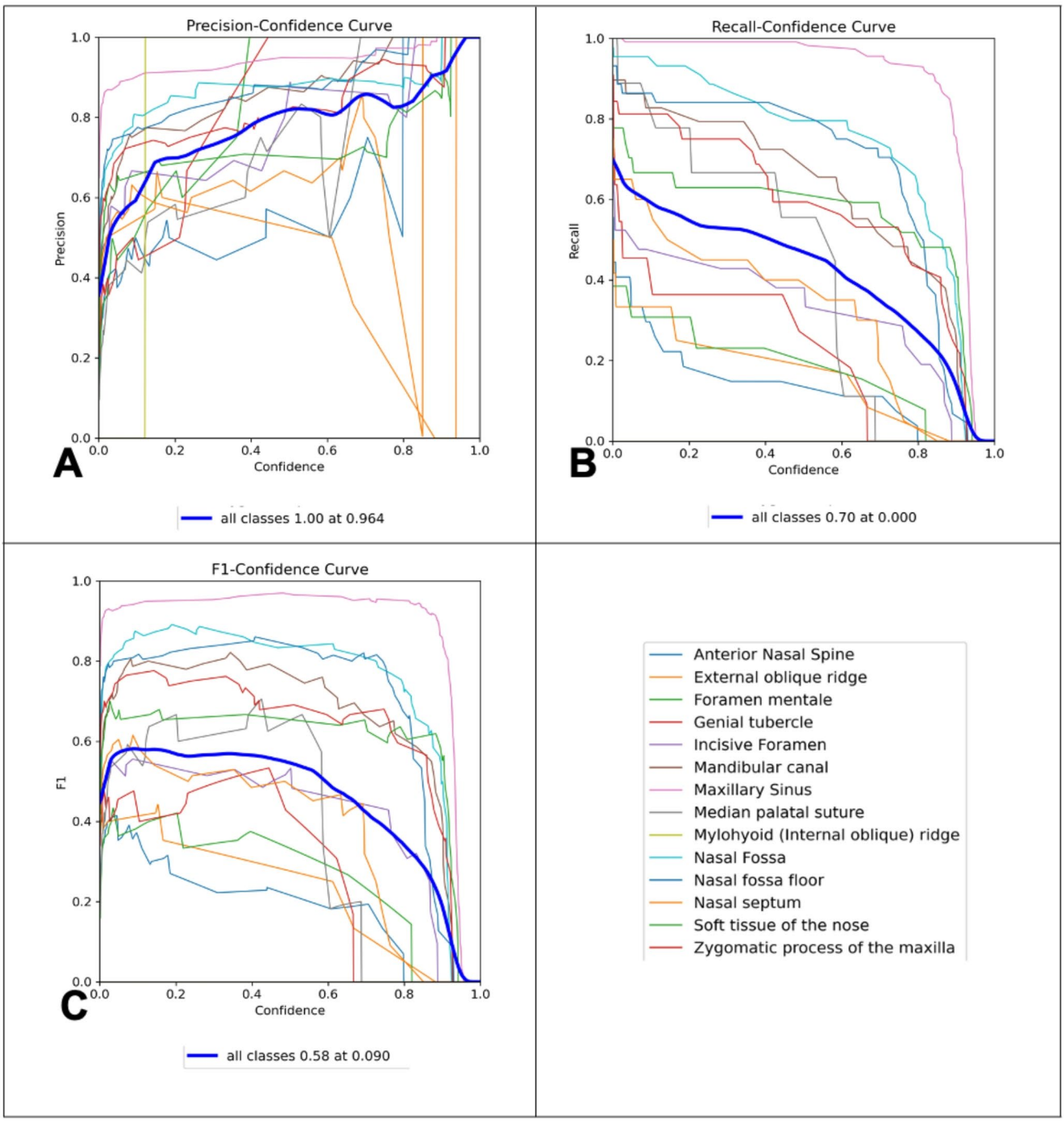

Fig. 3 Confidence-dependent performance of the proposed model across all anatomical landmark classes. Precision increased with higher confidence thresholds (**A**), whereas recall decreased accordingly (**B**). The resulting F1–confidence curve demonstrated a global maximum at approximately 0.09 (**C**), indicating optimal balance between precision and recall at low confidence thresholds. Thin lines represent class-wise trends, and the bold line denotes aggregated performance across all classes

confidence and approached 1.0 at high thresholds. Recall sharply declined as confidence increased, indicating that low-confidence predictions captured more true structures at the cost of precision. F1-score peaked early and declined with increasing thresholds, confirming 0.05–0.10 as the optimal operational range for the model. Class-specific curves showed that the maxillary sinus, nasal fossa floor, and mandibular canal exhibited the most stable performance across varying confidence thresholds, which can be attributed to their larger size, clearer radiographic boundaries, and greater anatomical consistency across images.

Table 2 Class-wise detection and segmentation performance metrics for 21 periapical anatomical structures, including TP, FP, FN, precision, recall, F1-score, Dice similarity coefficient, and IoU

Structure	TP	FP	FN	Precision	Recall	F1	Dice	IoU
High-performing structures								
Maxillary sinus	97	7	4	0.933	0.960	0.946	0.946	0.899
Nasal fossa floor	46	7	12	0.868	0.793	0.829	0.829	0.708
Nasal fossa	23	6	9	0.793	0.719	0.754	0.754	0.605
Nasal septum	13	1	8	0.929	0.619	0.742	0.742	0.591
Moderate-performing structures								
Mandibular canal	16	3	15	0.842	0.516	0.642	0.642	0.471
Median palatal suture	13	6	12	0.684	0.520	0.591	0.591	0.419
Zygomatic process of maxilla	16	15	8	0.516	0.667	0.582	0.582	0.410
Genial tubercle	6	0	11	1.000	0.353	0.522	0.522	0.353
External oblique ridge	6	3	12	0.667	0.333	0.444	0.444	0.286
Anterior nasal spine	12	9	22	0.571	0.353	0.435	0.435	0.279
Maxillary sinus septum	4	0	11	1.000	0.267	0.421	0.421	0.267
Low-performing structures								
Soft tissue of the nose	7	14	13	0.333	0.350	0.342	0.342	0.206
Tuber maxilla	3	2	8	0.600	0.273	0.375	0.375	0.231
Pterygoid process	1	0	4	1.000	0.200	0.333	0.333	0.200
Mylohyoid ridge	1	5	5	0.167	0.167	0.167	0.167	0.091
Incisive foramen	1	5	16	0.167	0.059	0.088	0.088	0.045
Mental foramen	0	1	25	0.000	0.000	0.000	0.000	0.000
Lateral fossa	0	2	18	0.000	0.000	0.000	0.000	0.000
Coronoid process	0	1	1	0.000	0.000	0.000	0.000	0.000
Inferior concha	0	2	0	0.000	0.000	0.000	0.000	0.000
Nutrient canal	0	1	0	0.000	0.000	0.000	0.000	0.000

Structures are grouped as high-, moderate-, and low-performing based on F1-score

Confusion matrix analysis

The normalized and raw confusion matrices (Fig. 4A, B) demonstrated that the model achieved its highest correct classification rates for large, well-defined structures such as the maxillary sinus (0.99), nasal fossa (0.89), nasal fossa floor (0.77), and the mandibular canal (0.79). These findings reflect the radiographic clarity and anatomical consistency of these regions. In contrast, small or low-contrast structures—including the mental foramen, incisive foramen, mylohyoid ridge, tuber maxilla, and the zygomatic process—were frequently misclassified as background, largely due to class imbalance and limited visual distinctiveness. Nutrient canal and inferior concha did not appear on the predicted axis of the matrix because the model produced no true-positive detections for these classes, a pattern fully consistent with their class-wise metrics. Importantly, confusion among morphologically adjacent nasal structures (nasal fossa, nasal fossa floor, nasal septum) remained low, indicating that the model was able to differentiate these closely related regions despite spatial proximity. Overall, the confusion matrix patterns highlight the model's clinical strength in detecting robust, high-contrast landmarks while underscoring the inherent difficulty of identifying subtle anatomical features on periapical radiographs.

Visual evaluation

Visual inspection of representative periapical radiographs (Fig. 2) showed that the model reproduced most annotated anatomical structures across different regions. In the anterior maxilla, most landmarks were detected, whereas the nasal fossa floor was missed and a small unannotated maxillary sinus region was additionally predicted. In the mandibular anterior region, the genial tubercle was not detected. Posterior maxillary images showed correct identification of the maxillary sinus and nasal fossa floor with high confidence. In the posterior mandible, the mandibular canal was consistently detected, while mental foramen predictions remained variable.

Discussion

This study developed and validated a deep learning model for the automatic identification of 21 anatomical landmarks on periapical radiographs representing both maxillary and mandibular regions. To the best of our knowledge, this is the first investigation focusing on comprehensive multi-class anatomical landmark detection on periapical radiographs using an instance segmentation approach. The

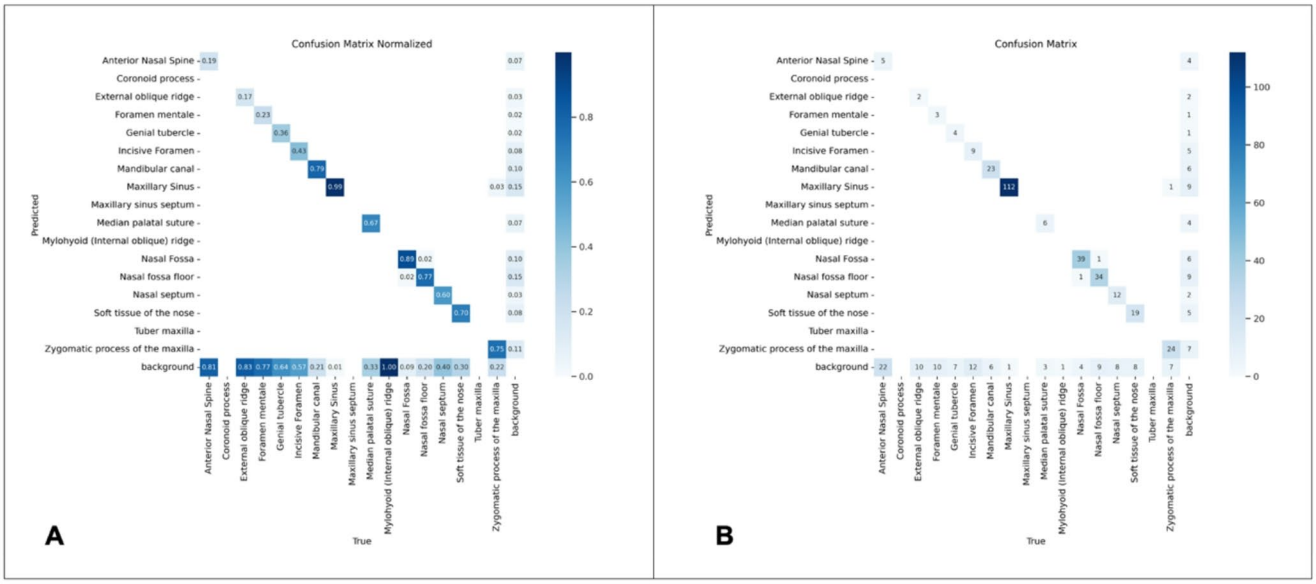

Fig. 4 Confusion matrices illustrating class-wise performance of the proposed anatomical landmark detection model on the test set. **A** Normalized confusion matrix showing the proportion of predictions per true class. **B** Raw confusion matrix displaying absolute prediction counts. The matrices highlight that misclassifications predominantly involve anatomical landmarks being predicted as background rather than confusion between neighboring structures

model demonstrated high accuracy for large, well-defined structures, whereas performance for small, low-contrast, or infrequently represented landmarks remained limited. These findings highlight both the promise and inherent constraints of applying deep learning to detailed anatomical interpretation on conventional intraoral radiographs.

Previous studies have reported encouraging results in anatomical landmark detection using panoramic radiographs and CBCT, particularly for specific structures such as the mandibular canal and maxillary sinus [23, 24]. However, these approaches have generally been limited to a small number of anatomical classes and have predominantly relied on detection or classification frameworks, without performing detailed instance-level segmentation of anatomical structures. In contrast, the present study focuses on periapical radiographs, which present unique challenges due to their limited field of view and projection-dependent anatomical variability. Furthermore, our model performs multi-class segmentation across multiple anatomical landmarks, enabling not only localization but also structural delineation. This approach contributes to the existing literature by addressing a wider range of anatomical structures within a unified framework and by focusing on a clinically relevant imaging modality that has been comparatively less explored in the context of multi-class anatomical landmark segmentation.

High-performing structures, including the maxillary sinus, nasal fossa, nasal fossa floor, and nasal septum, benefited from their comparatively large anatomical footprint and clearly defined cortical margins. These structures also demonstrate low inter-individual variability, leading to more stable radiographic patterns that facilitate model learning. Prior studies on panoramic radiographs and CBCT similarly reported high segmentation accuracy for large or distinct anatomical regions, particularly the maxillary sinus [16, 17]. Comparable findings have been documented in comparative radiographic analyses, where the sinus floor and surrounding cortical boundaries in the posterior maxilla were more consistently visualized on CBCT than periapical radiographs [25]. Our results confirm that despite the smaller field of view of intraoral radiographs, high-contrast anatomical structures retain sufficient radiographic cues for robust automated detection.

Moderate-performing structures, including the mandibular canal, median palatal suture, zygomatic process of the maxilla, nasal soft tissue, and genial tubercle, exhibited variable accuracy with higher false-negative rates. Although the mandibular canal represents a relatively large anatomical structure, its elongated course, variable cortical definition, and partial visualization within the limited field of view of periapical radiographs likely contributed to reduced recall and moderate overall performance in this study. Many of the remaining moderate-performing structures are thin, anatomically variable, or demonstrate low contrast relative to surrounding bone. Similar limitations have been described in previous intraoral deep learning studies, where lamina dura, trabecular bone architecture, and subtle cortical ridges produced inconsistent segmentation outputs due to indistinct boundaries and radiodensity variability [26, 27]. These findings suggest that thin or elongated structures on periapical

images may require higher-resolution data, more robust sampling strategies, or architectures specifically designed to better capture fine-scale variations.

Low-performing structures, such as the incisive foramen, mental foramen, mylohyoid ridge, pterygoid process, inferior concha, and nutrient canal, demonstrated limited or absent true-positive detections. Importantly, although the overall dataset was substantial, these specific landmarks naturally appear less frequently in periapical radiographs due to the limited field of view and projection geometry inherent to the imaging modality. This low representation is consistent with clinical studies demonstrating the limited visualization of key anatomical structures, particularly in the posterior mandible and premaxillary region, when periapical radiographs are compared with CBCT [6, 28]. In addition, many of these structures blend with surrounding trabecular bone, lack distinct cortical margins, and can only be differentiated from adjacent anatomy by slight reductions in gray-value intensity. This subtle contrast is a well-recognized source of interpretive difficulty and contributes to both human and algorithmic misclassification. The mental foramen, in particular, is widely documented as a frequent source of diagnostic confusion because of its variable position and indistinct borders.

Confidence–precision–recall analysis revealed a characteristic pattern, with precision increasing at higher confidence thresholds and recall declining accordingly. The F1-score peaked at a low confidence range (approximately 0.05 to 0.10), indicating that low-confidence predictions may convey clinically relevant information for certain anatomical structures. This behavior suggests that the model is best suited for assistive use, where lower thresholds can help highlight candidate regions for clinician review, particularly for subtle landmarks that are easily overlooked on intraoral radiographs. Accordingly, the proposed approach should be regarded as a foundational anatomical awareness module designed to support clinical interpretation rather than as a fully autonomous diagnostic system. In practical terms, such a module could be incorporated into clinical software to provide an additional layer of anatomical reference during image interpretation, with potential to support more advanced diagnostic systems as they evolve.

Confusion matrix evaluation supported the class-wise performance trends. Misclassifications primarily involved small structures being labelled as background rather than incorrectly classified as neighboring anatomy. Of note, despite the spatial proximity of nasal structures, the nasal fossa, nasal fossa floor, and nasal septum were rarely confused with one another, indicating that the model was able to differentiate these adjacent regions based on local contextual information. Conversely, structures such as the incisive foramen, mental foramen, and nutrient canal were frequently undetected or blended with background due to radiographic overlap with trabecular bone and lack of clear contour definition. These findings underscore the need for higher-resolution datasets and more balanced representation of small landmarks to improve detection performance.

The visual evaluation findings were consistent with numerical results. Large, well-bounded structures such as the maxillary sinus, and to a lesser extent the mandibular canal, were consistently reproduced, while small structures with subtle or indistinct boundaries showed greater variability. These observations emphasize that automated detection performance is strongly influenced by the intrinsic radiographic characteristics of the underlying anatomy. Beyond these intrinsic factors, the findings should also be interpreted in the context of domain shift, a well-recognized limitation of deep learning models [29]. Because supervised learning relies on patterns learned from training data, models implicitly assume that new data will follow a similar distribution; however, this assumption is frequently violated in clinical practice [30]. Variations in imaging conditions, including institution- and operator-dependent factors, device-specific characteristics, and the two-dimensional projection of three-dimensional anatomical structures, can alter the statistical properties of images and lead to distributional discrepancies between training and deployment data. In projection-based imaging, variations in projection geometry and patient positioning may systematically modify the superimposition and representation of anatomical structures, effectively giving rise to distinct data domains even within clinically similar settings. Such discrepancies, often referred to as domain shift, have been shown to adversely affect model generalization, particularly in projection-based imaging modalities. In line with recent domain generalization literature, these variations may arise even within a single institution and can persist despite standardization efforts, highlighting the challenges associated with learning domain-invariant representations [31]. In this context, the observed performance characteristics of the present model should be interpreted with consideration of these distributional limitations.

The proposed model may support a range of diagnostic and treatment-planning procedures that rely on accurate anatomical interpretation. In endodontics and apical surgery, precise identification of anatomical landmarks such as the mental foramen, mandibular canal, and maxillary sinus floor is essential for avoiding procedural complications. Similarly, in implant planning, accurate delineation of critical anatomical structures may contribute to safer implant positioning and improved risk assessment. Furthermore, automated recognition of normal anatomical landmarks may help reduce the risk of misinterpreting anatomical variations as pathological findings, particularly in regions where radiographic superimposition and structural overlap

are common. In this context, the model may serve as a supportive tool to improve consistency in anatomical interpretation and enhance diagnostic confidence in routine clinical practice.

Despite these promising applications, several limitations of the present study should be acknowledged. Although the dataset was relatively large, some structures were naturally underrepresented because of the restricted field of view of periapical radiography. Annotation was carried out collaboratively by three annotators, with periodic consensus meetings supervised by a dentomaxillofacial radiologist. Nevertheless, formal interobserver agreement analysis was not performed and would further strengthen label reliability. All images were acquired at a single institution using standardized exposure parameters, which ensures internal consistency but may limit generalizability. External multicenter validation remains necessary to confirm the robustness of the model across different radiographic systems and settings. Finally, although preprocessing-derived originals were added prior to dataset splitting to avoid leakage, geometry-preserving transformations cannot fully account for inter-institutional variability.

Future studies should address class imbalance by expanding multi-center datasets, incorporate resolution-adaptive or geometry-aware architectures to better capture small or elongated structures, and explore targeted augmentation techniques tailored to low-contrast anatomy. Integrating anatomical landmark detection with pathology identification may further enhance clinical relevance, particularly for assessing lesion boundaries, proximity to vital structures, and surgical planning pathways. Incorporating cross-modal validation with CBCT and clinical metadata such as tooth region, surgical history, or patient-level factors also holds promise for improving detection performance for structures that are inherently difficult to visualize on intraoral radiographs.

Conclusion

In conclusion, the proposed deep learning model demonstrated promising performance in detecting anatomical landmarks with well-defined, high-contrast radiographic boundaries, while exhibiting expected limitations for small, low-contrast, or underrepresented structures. These findings highlight both the potential and current constraints of AI-assisted anatomical interpretation on periapical radiographs. With further methodological refinements and multi-center validation, such approaches may support safer and more consistent radiographic interpretation.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). This research project was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) of (2209-A) Research Project Support Programme for Undergraduate Students (project no: 2024/1-1919B012426974).

Data availability The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no competing interests.

Ethical approval This study was approved by the Institutional Ethics Committee of Istanbul Okan University (Approval No: 2024/175-3).

Informed consent The requirement for informed consent was waived due to the retrospective nature of the study.

Generative AI and AI-assisted technologies in the manuscript preparation process During the preparation of this work, the authors used ChatGPT (OpenAI) to assist with language editing, clarity improvement, and refinement of academic writing. After using this tool, the authors critically reviewed and edited all content and take full responsibility for the accuracy, originality, and integrity of the published article.

References

1. White SC, Pharoah MJ. Oral radiology: principles and interpretation. 7th ed. St Louis, MO: Elsevier; 2014. p. 86.
2. White SC, Pharoah MJ. Oral radiology: principles and interpretation. 7th ed. St Louis, MO: Elsevier; 2014. p. 81.
3. Kaji K, Kotaki S, Maesoma A, et al. Hands-on periapical radiography training for dental hygiene students. J Osaka Dent Univ. 2023;57(1):147–52.
4. Shahrour R, Shah P, Withana T, Jung J, Syed AZ. Oroantral communication: causes, complications, treatments, and radiographic features—a pictorial review. Imaging Sci Dent. 2021;51(3):307.
5. Laher AE, Wells M, Motara F, Kramer E, Moolla M, Mahomed Z. Finding the mental foramen. Surg Radiol Anat. 2016;38:469–76.
6. Bornstein MM, Lauber R, Sendi P, von Arx T. Comparison of periapical radiography and limited cone-beam computed tomography in mandibular molars for analysis of anatomical landmarks before apical surgery. J Endod. 2011;37(2):151–7.
7. Doğramaci EJ, Decurcio DDA, Chen CJ, Estrela C, Rossi-Fedele G. Accuracy of periapical and panoramic radiography for

detection of root morphologies: a comparative study. Braz Oral Res. 2021;35:e065. https://doi.org/10.1590/1807-3107bor-2021.vol35.0065.
8. Kadesjö N, Lynds R, Nilsson M, Shi XQ. Radiation dose from X-ray examinations of impacted canines: cone beam CT vs two-dimensional imaging. Dentomaxillofac Radiol. 2018;47(3):20170305.
9. Gerardi D, Angiolani F, Kërpi B, et al. Clinical oral anatomy in clinical dental practice: a scoping review. Int J Morp. 2025;43(2):640–50.
10. Bernardi S, Bianchi S, Gerardi D, et al. Anatomy of maxillary sinus: Focus on vascularization and underwood septa via 3D imaging. Tomography Cureus. 2024;10(4):444–58.
11. Shujaat S, Aljadaan H, Alrashid H, Aboalela AA, Riaz M. FDA-approved AI solutions in dental imaging: a narrative review of applications, evidence, and outlook. Int Dent J. 2025;76:109315.
12. Singh V, Cheng S, Kwan AC, Ebinger J. United States Food and Drug Administration regulation of clinical software in the era of artificial intelligence and machine learning. Mayo Clin Proc Digit Health. 2025;3(3):100231.
13. Liu W, Li X, Liu C, Gao G, Xiong Y, Zhu T, et al. Automatic classification and segmentation of multiclass jaw lesions in cone-beam CT using deep learning. Dentomaxillofac Radiol. 2024;53(7):439–46.
14. Bayrakdar IS, Orhan K, Akarsu S, Çelik Ö, Atasoy S, Pekince A, et al. Deep-learning approach for caries detection and segmentation on dental bitewing radiographs. Oral Radiol. 2022;38(4):468–79.
15. Huang C, Wang J, Wang S, Zhang Y. A review of deep learning in dentistry. Neurocomputing. 2023;554:126629.
16. Cha JY, Yoon HI, Yeo IS, Huh KH, Han JS. Panoptic segmentation on panoramic radiographs: deep learning-based segmentation of various structures including maxillary sinus and mandibular canal. J Clin Med. 2021;10(12):2577.
17. Yuce F, Buyuk C, Bilgir E, Çelik Ö, Bayrakdar İŞ. Deploying a novel deep learning framework for segmentation of specific anatomical structures on cone-beam CT. Oral Radiol. 2025;41(4):562–70.
18. Chen H, Li H, Zhao Y, Zhao J, Wang Y. Dental disease detection on periapical radiographs based on deep convolutional neural networks. Int J Comput Assist Radiol Surg. 2021;16:649–61.
19. Ekert T, Krois J, Meinhold L, Elhennawy K, Emara R, Golla T, Schwendicke F. Deep learning for the radiographic detection of apical lesions. J Endod. 2019;45(7):917–22.
20. Sajad M, Shafi I, Ahmad J. Automatic lesion detection in periapical X-rays. In: Proceedings of the 2019 international conference on electrical, communication, and computer engineering (ICECCE). IEEE; 2019:1–6.
21. Shafi I, Sajad M, Fatima A, Aray DG, Lipari V, Diez IDLT, Ashraf I. Teeth lesion detection using deep learning and the Internet of Things post–COVID-19. Sens (Basel). 2023;23(15):6837.
22. Pul U, Schwendicke F. Artificial intelligence for detecting periapical radiolucencies: a systematic review and meta-analysis. J Dent. 2024;105:105104.
23. Bağ İ, Bilgir E, Bayrakdar IS, et al. An artificial intelligence study: automatic description of anatomic landmarks on panoramic radiographs in the pediatric population. BMC Oral Health. 2023;23(1):764.
24. Tang H, Liu S, Shi Y, et al. Automatic segmentation and landmark detection of 3D CBCT images using semi supervised learning for assisting orthognathic surgery planning. Sci Rep. 2025;15(1):8814.
25. Shahbazian M, Vandewoude C, Wyatt J, Jacobs R. Comparative assessment of periapical radiography and CBCT imaging for radiodiagnostics in the posterior maxilla. Odontology. 2015;103(1):97–104. https://doi.org/10.1007/s10266-013-0144-z.
26. Khan HA, Haider MA, Ansari HA, Ishaq H, Kiyani A, Sohail K, Khurram SA. Automated feature detection in dental periapical radiographs using deep learning. Oral Surg Oral Med Oral Pathol Oral Radiol. 2021;131(6):711–20.
27. Jagtap R, Samata Y, Parekh A, Tretto P, Roach MD, Sethumanjusha S, Tejaswi C, Jaju P, Friedel A, Briner Garrido M, et al. Clinical Validation of Deep Learning for Segmentation of Multiple Dental Features in Periapical Radiographs. Bioengineering. 2024;11(10):1001. https://doi.org/10.3390/bioengineering11101001.
28. Sheth K, Kapoor S, Daveshwar S. Comparison of cone-beam computed tomography and periapical radiography to determine the proximity of periapical lesions to anatomical structures in the premaxillary area prior to surgical endodontics: a clinical study. Int J Clin Pediatr Dent. 2020;13(4):322.
29. Roland T, Böck C, Tschoellitsch T, Maletzky A, Hochreiter S, Meier J, Klambauer G. Domain shifts in machine learning based Covid-19 diagnosis from blood tests. J Med Syst. 2022;46(5):23.
30. Kilim O, Olar A, Joó T, Palicz T, Pollner P, Csabai I. Physical imaging parameter variation drives domain shift. Sci Rep. 2022;12(1):21302.
31. Yoon JS, Oh K, Shin Y, Mazurowski MA, Suk HI. Domain generalization for medical image analysis: A review. Proc IEEE. 2024;112(10):1583–609.