OPEN ACCESS

Saudi Journal of Oral and Dental Research
Abbreviated Key Title: Saudi J Oral Dent Res
ISSN 2518-1300 (Print) | ISSN 2518-1297 (Online)
Scholars Middle East Publishers, Dubai, United Arab Emirates
Journal homepage: https://saudijournals.com

Review Article | **Dentistry**

Artificial Intelligence in Dental Radiographic Diagnosis: A Comprehensive Review

Dr. Malik Hina, MDS[1]*, Dr. Sulekha Beevi Jalal, BDS[2], Dr. Shabnam Shardimanaheji, BDS[3], Dr. Nameera Usmani, BDS[4], Dr. Sana Mohmedasif Shaikh, BDS[5], Dr. Shazia Zainab, BDS[6], Dr. Abhirami K., MDS[7], Dr. Khan Sara Akmal, BDS[6], Dr. Mohd Sadique Ali, MDS[1], Dr. Gurkanwal Kaur Aurora, BDS[8]

[1]Saraswati Dental College and Hospital, Lucknow, India
[2]Malabar Dental College, Edappal, Kerala, India
[3]Ajman University, UAE
[4]Bharti Vidyapeeth University, India
[5]Al Badar Rural Dental College and Hospital, Gulbarga, India
[6]Career Post Graduate Institute of Dental Sciences & Hospital, Lucknow, India
[7]RajaRajeswari Dental College, Bengaluru, India
[8]Baba Farid University of Health and Sciences, India

DOI: https://doi.org/10.36348/sjodr.2026.v11i07.003 | **Received:** 27.05.2026 | **Accepted:** 15.07.2026 | **Published:** 21.07.2026

***Corresponding author:** Dr. Malik Hina, MDS
Saraswati Dental College and Hospital, Lucknow, India

Abstract

Background: Artificial intelligence (AI), particularly convolutional neural networks (CNNs) and vision transformers (ViTs), is transforming dental radiographic diagnosis across multiple clinical domains. ***Methods:*** A narrative review searched PubMed, PubMed Central, BMC Oral Health, and Google Scholar for systematic reviews, meta-analyses, and original studies published between 2019 and 2025, covering ten domains: caries detection, periapical lesion identification, alveolar bone loss, peri-implant assessment, tooth detection and segmentation, cone-beam computed tomography (CBCT) analysis, vertical root fracture (VRF) detection, cephalometric landmark analysis, dental anomaly detection, and AI-versus-clinician comparison. ***Results:*** Pooled sensitivity was 0.94 and specificity 0.91 for caries detection on bitewing radiographs. Periapical AI sensitivity was 92.9% with specificity of only 58.6%, yielding a positive predictive value of 15.3%. Peri-implant bone loss assessment achieved AUC of 0.94. Vision transformers outperformed CNNs in 58% of dental imaging studies. VRF detection accuracy ranged from 75% to 97.8%. Cephalometric landmark detection rate reached 98.29%. AI diagnosed radiographs in 1.5 seconds versus 53.8 seconds for clinicians. ANSI/ADA Standard 1110-1:2025 provides the first formal dental AI validation framework. ***Conclusions:*** AI achieves substantial and reproducible diagnostic accuracy across dental radiographic tasks. High false-positive rates, demographic dataset bias, limited multi-centre external validation, and regulatory gaps must be resolved before broad clinical implementation.

Keywords: Artificial intelligence, deep learning, dental radiograph, cone-beam computed tomography, caries detection, periapical lesion, alveolar bone loss, cephalometric analysis, vertical root fracture, diagnostic accuracy.

INTRODUCTION

The integration of artificial intelligence (AI) into clinical medicine has accelerated markedly over the past decade, propelled by advances in computational power, the growing availability of large annotated clinical datasets, and fundamental breakthroughs in deep learning methodology (Topol, 2019). Among the most impactful healthcare applications of AI is image-based diagnostic analysis, where deep learning systems have demonstrated performance comparable to—and in certain domains exceeding—that of trained specialists (Litjens *et al.,* 2017). Dentistry, with its heavy reliance on radiographic imaging for diagnosis and treatment planning, represents a particularly compelling setting for AI-assisted clinical decision support.

Dental radiographs encompass a spectrum of modalities—periapical, bitewing, panoramic, lateral cephalometric, and cone-beam computed tomography (CBCT)—each critical to the identification of specific pathologies and anatomical structures. Interpretation of these images is time-consuming and subject to inter-observer variability (Schwendicke *et al.,* 2020).

Automated analysis systems powered by CNNs, YOLO-based object detectors, vision transformers (ViTs), and emerging multimodal large language models (LLMs) offer the potential to standardize diagnostic workflows, reduce missed diagnoses, and augment clinician performance across a wide range of radiographic tasks.

The published literature in this field has grown substantially in volume and breadth, now encompassing original studies, systematic reviews, meta-analyses, and umbrella reviews across multiple diagnostic tasks and radiographic modalities. The aim of this comprehensive narrative review is to synthesize the evidence across ten key clinical domains, address persistent limitations including dataset bias and regulatory gaps, and evaluate the current and future role of AI as a clinical decision-support tool in dental radiographic diagnosis. A flowchart illustrating the literature search strategy and review scope is presented in Figure 1.

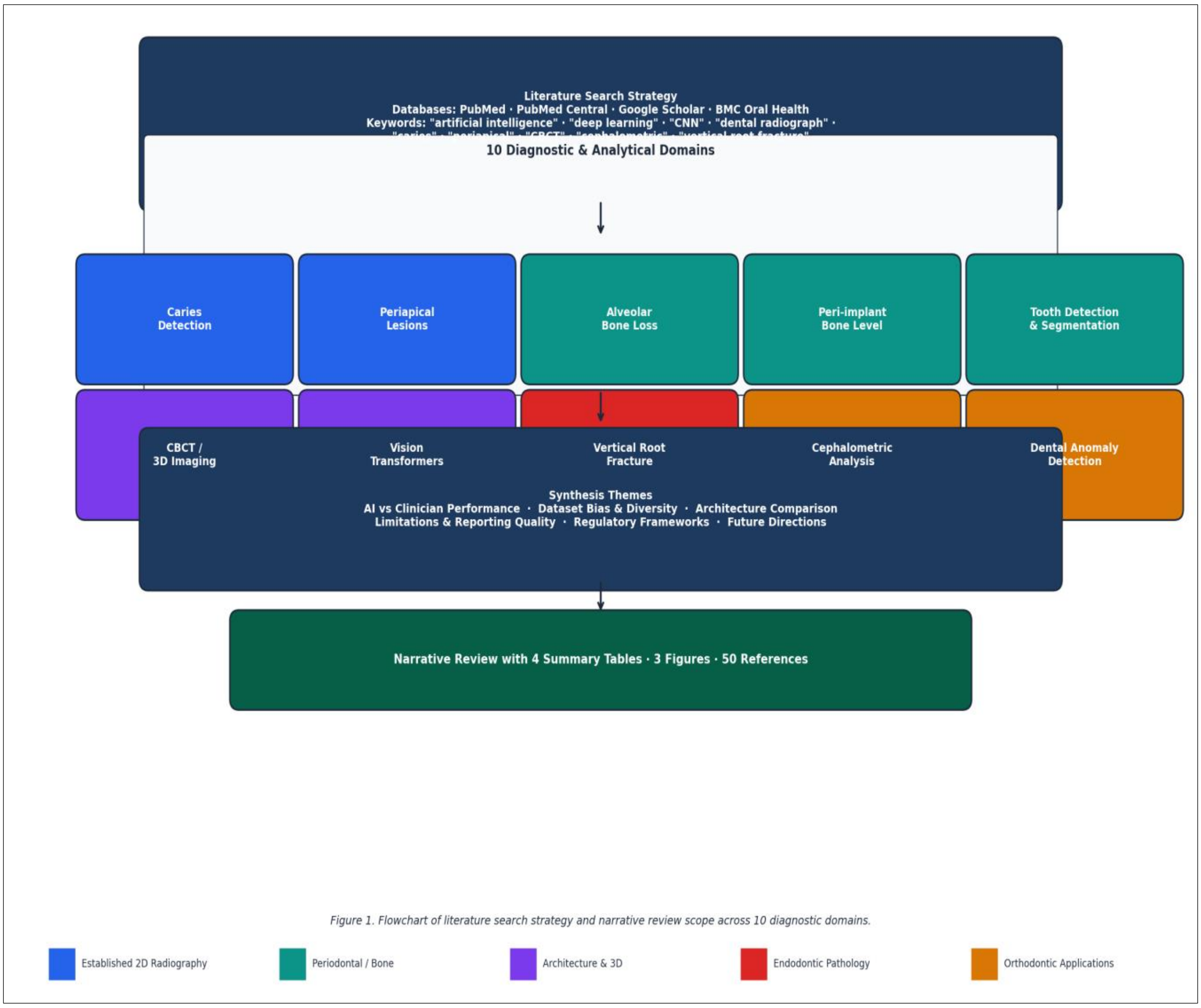

Figure 1: Flowchart illustrating the literature search strategy, databases consulted, and the ten diagnostic domains covered in this narrative review. Domains are colour-coded by clinical specialty

Artificial Intelligence in Dental Caries Detection

Dental caries is the most prevalent non-communicable disease globally, and its radiographic identification—particularly of early interproximal and occlusal lesions—represents a high-volume diagnostic challenge in general dental practice. The majority of published AI studies in dentistry have focused on caries detection, reflecting its clinical significance and the relative availability of annotated radiographic datasets (Negi *et al.,* 2024).

CNNs have emerged as the dominant architecture for radiographic caries detection. A systematic review published in BMC Oral Health analyzed twenty studies across periapical, bitewing, and panoramic radiograph types, reporting consistently high sensitivity and specificity metrics, with CNN-based models outperforming earlier machine learning approaches (Mertens *et al.,* 2024). For approximal caries on bitewing radiographs—the most common radiographic caries presentation—a systematic review and meta-analysis found pooled sensitivity of 0.94 and specificity of 0.91, performance comparable to experienced clinicians (Kühnisch *et al.,* 2024). An umbrella review by Negi *et al.,* (2024) confirmed that CNNs consistently outperformed multilayer perceptrons and support vector machines. YOLO-based architectures have also demonstrated high performance: a YOLOv5s model evaluated on the DENTEX CHALLENGE 2023 dataset achieved precision of 99.6%, recall of 98.9%, and mAP50 of 99.5% for tooth-level caries identification. The clinical application of deep learning for multistage caries detection on panoramic radiographs, evaluated in Scientific Reports, demonstrated feasibility of automated caries staging (Pornprasertsuk-Damrongsri *et al.,* 2025). Despite these results, variability in reported performance across studies is driven by differences in image acquisition protocols, preprocessing, ground-truth labeling, and the radiographic stage of caries being detected.

Artificial Intelligence in Periapical Lesion Detection

Periapical pathology manifests radiographically as radiolucencies at root apices, and its accurate identification is essential for endodontic management. A narrative review published in PubMed Central evaluating studies between 2021 and 2025 confirmed high overall diagnostic accuracy of CNN-based systems for periapical radiolucency identification (Makaremi *et al.,* 2025). A diagnostic accuracy study reviewing twenty-four studies also confirmed high sensitivity in periapical AI detection (Ahmed *et al.,* 2025).

A critical clinical concern is the false positive paradox. One comprehensive evaluation reported AI sensitivity of 92.9% and negative predictive value of 99% for periapical assessment; however, specificity was only 58.6%, yielding a positive predictive value of just 15.3% (Sparnon *et al.,* 2026). In clinical settings where periapical lesion prevalence is modest, the majority of AI-flagged cases will be false positives, risking unnecessary endodontic intervention. Clinicians must understand the statistical relationship between model specificity, disease prevalence, and positive predictive value when interpreting AI-generated alerts (see Figure 2 and Table 1).

Artificial Intelligence in Alveolar Bone Loss and Peri-Implant Bone Level Assessment

Periodontal bone loss, quantified as reduction in alveolar crest height relative to the cementoenamel junction (CEJ), is central to periodontal staging and monitoring. A study employing CNN architectures for automatic tooth recognition and periodontal bone loss measurement on digital radiographs demonstrated the feasibility of end-to-end automated periodontal assessment (Chen *et al.,* 2023). A deep learning model for automated periodontitis staging on panoramic radiographs demonstrated clinically relevant accuracy for AI-based staging (Xue *et al.,* 2024). A YOLOv8x-seg approach achieved intra-class correlation coefficients of up to 0.80 for bone loss severity and 87% accuracy for pattern classification (Wimalasiri *et al.,* 2026). An evaluation of a commercial AI application by dental professionals found reliable bone level estimates suitable for decision support (Lee *et al.,* 2025).

AI has also been applied to peri-implant marginal bone loss detection. A systematic review and meta-analysis (Atieh *et al.,* 2026) including nine studies and 7,284 implants found pooled sensitivity of 0.84 (95% CI: 0.72–0.91), specificity of 0.91 (95% CI: 0.81–0.96), and AUC of 0.94 for AI-based peri-implant bone loss detection on dental radiographs. A YOLOv8-based model for classifying peri-implant bone loss on CBCT images achieved 0.88 three-class accuracy with Kappa of 0.954 (Madani & Fakhar, 2026).

Artificial Intelligence in Tooth Detection and Segmentation on Panoramic Radiographs

Automated tooth detection and segmentation on panoramic radiographs is a prerequisite for downstream AI diagnostic tasks. A systematic review and meta-analysis in Bonfanti-Gris *et al.,* (2025) identified twenty relevant studies from 2,207 records, confirming consistently high pooled accuracy across deep learning architectures (Bonfanti-Gris *et al.,* 2025). A study comparing YOLO-V4 with Faster R-CNN on 1,200 panoramic radiographs reported YOLO-V4 mean precision of 99.90%, recall of 99.18%, and F1 score of 99.54%, outperforming Faster R-CNN (Yilmaz *et al.,* 2023). AI has also demonstrated utility in impacted tooth identification and detection of multiple treatment features including restorations, endodontic posts, and prostheses on panoramic radiographs in a single inference pass (Khurshid *et al.,* 2025).

Artificial Intelligence in Cone-Beam Computed Tomography and Three-Dimensional Dental Imaging

Cone-beam computed tomography (CBCT) provides three-dimensional radiographic data with high spatial resolution and is indispensable in implantology, orthodontics, endodontics, and oral surgery. A landmark study published in Nature Communications demonstrated a fully automatic AI system capable of simultaneously segmenting individual teeth and alveolar bone from CBCT images, achieving average Dice scores of 96.5% for tooth segmentation, 95.4% for alveolar bone, 93.6% for maxillary sinus, and 94.8% for mandibular canal segmentation (Chen *et al.,* 2022). A systematic review of AI for tooth segmentation in CBCT images, published in Applied Sciences in 2024, confirmed that U-Net and V-Net architectures represent the dominant and most effective approaches (Tarce *et al.,* 2024). A systematic review of alveolar bone segmentation in CBCT for periodontal regeneration assessment confirmed that deep learning methods enable reliable automated measurement of periodontal defects (Mohammed *et al.,* 2025). A deep learning framework for automated dental tissue segmentation and diagnostic report generation from CBCT was published in PubMed

Central in 2025 (Zhou *et al.,* 2025). For surgical planning, deep learning models based on nnU-Net v2 have been developed for automatic segmentation of impacted canines in CBCT scans (Unal *et al.,* 2025). YOLOv8-based architectures for peri-implant bone loss detection and classification on CBCT have further demonstrated that 3D imaging can enhance implant monitoring accuracy beyond 2D radiographs (PMC, 2025d).

Vision Transformers and Emerging Architectures in Dental Radiographic AI

A systematic review evaluating ViTs and CNNs in dental image processing, published in BMC Oral Health in 2025 (Felek *et al.,* 2025), identified 21 eligible studies from 2,596 screened records. ViT-based models demonstrated the highest performance in 58% of the included studies, a finding with immediate implications for model selection in new AI development (Felek *et al.,* 2025). The architectural basis for this advantage lies in the global attention mechanism of transformers, which captures long-range spatial dependencies across entire radiographic images without the local receptive field constraints of CNNs. A study evaluating ViT for furcation involvement classification on panoramic radiographs demonstrated improved accuracy over CNN-based alternatives (Zhang *et al.,* 2025). A transformer-based model for early dental caries detection on panoramic radiographs confirmed superior detection performance compared to CNN benchmarks (Wang & Li, 2026). A systematic review of hybrid CNN-ViT architectures for radiological image analysis identified hybrid models as demonstrating best overall performance across medical imaging tasks (Kim *et al.,* 2024). Future dental AI development should systematically evaluate and incorporate ViT and hybrid architectures alongside established CNN baselines.

Artificial Intelligence in Vertical Root Fracture Detection

Vertical root fracture (VRF) is one of the most diagnostically challenging and clinically consequential pathologies in endodontics, frequently detected late due to subtle and variable radiographic presentation, and typically necessitating tooth extraction upon confirmed diagnosis. A systematic review published in Diagnostics in 2026 (PMC12896978) searched PubMed, Scopus, Web of Science, and the Cochrane Library through January 2025, identifying nine eligible studies. CNN-based models demonstrated accuracy ranging from 75% to 97.8%, sensitivity of 75% to 98%, and specificity of 60% to 99% (Alshahrani *et al.,* 2026). A second systematic review in the Journal of Endodontics (2026) identified seven eligible studies and confirmed this performance range, with study quality assessed using QUADAS-2 (Patil *et al.,* 2026). Diagnostic performance was substantially influenced by study design: ex vivo studies consistently reported higher metrics than in vivo studies. An ensemble model combining VGG16, VGG19, and ResNet50 via a voting technique demonstrated improved accuracy over single-model approaches for root fracture detection on periapical radiographs (Abdelazim & Fouad, 2024). A systematic review specifically evaluating AI on periapical radiographs (Tasdemir *et al.,* 2026) confirmed that CBCT-based AI achieves the highest in vivo diagnostic metrics, while periapical AI performs best in controlled ex vivo environments.

Artificial Intelligence in Cephalometric Landmark Detection

Lateral cephalometric radiographs are among the most frequently used dental radiograph types in orthodontic diagnosis and treatment planning. A systematic review and meta-analysis published in Bioengineering in December 2024 (Ribas-Sabartes *et al.,* 2024) evaluated CNN-based AI algorithms for cephalometric point detection, reporting localization accuracy of 64.3–97.3%, with mean radial errors of 1.04–3.40 mm—within or near the clinically accepted threshold of 2 mm for most landmarks. A complementary systematic review and meta-analysis by Kunz *et al.,* (2021) (Schwendicke *et al.,* 2021) covering 19 studies found that deep learning models detected a mean of 30 cephalometric landmarks per image, with performance approaching expert-level accuracy for many but not all landmarks. A 2024 pilot study using Mask R-CNN (Jiao *et al.,* 2024) achieved a detection rate of 98.29% with average radial errors ranging from 0.56 to 9.51 mm (mean 2.19 mm), indicating that anatomically variable landmarks remain challenging. A critical review in Artificial Intelligence Review (Neeraja & Jani Anbarasi, 2025) concluded that while AI cephalometric analysis generates results significantly faster and more reproducibly than manual tracing, accuracy for complex landmarks still does not consistently match experienced orthodontists, and AI should currently function as a time-saving first-pass tool subject to expert review.

Artificial Intelligence in Detection of Dental Anomalies

Dental anomalies including supernumerary teeth (hyperdontia) and congenitally missing teeth (hypodontia) have clinically significant implications for orthodontic treatment planning. A study published in the European Journal of Orthodontics in 2025 assessed the diagnostic accuracy of Diagnocat™ AI software in detecting congenitally missing and supernumerary teeth on panoramic radiographs, representing one of the first evaluations of a commercial AI system for these specific anomalies (Makrygiannakis *et al.,* 2025). A PLOS One study assessed a deep learning-based automated detection system for supernumerary teeth in pediatric panoramic radiographs, with a DetectNet-based model achieving 96% accuracy in identifying impacted supernumerary teeth (Uzel *et al.,* 2025). U-Net-based deep learning has additionally been evaluated for simultaneous tooth segmentation and agenesis detection in panoramic radiographs (Tunc *et al.,* 2025), demonstrating that segmentation architectures can identify absence of expected tooth germs in mixed dentition panoramic images for automated hypodontia screening.

Comparison of Artificial Intelligence Performance with Dental Clinicians

A systematic review and meta-analysis published in Clinical and Experimental Dental Research in 2026 (Alabdulkareem *et al.,* 2026) found that deep learning algorithms improved average clinician sensitivity from 60.7% to 85.9% when used as a decision-support tool. Mean AI diagnosis time per periapical radiograph was 1.5 ± 0.3 seconds, compared with a clinician mean of 53.8 ± 46.0 seconds—a nearly 35-fold speed advantage. For caries detection specifically, AI demonstrated sensitivity of 88% and specificity of 91%, compared with human sensitivity of 84% and specificity of 88%. A diagnostic accuracy study evaluating panoramic radiographs by AI compared with human reference readers (Turosz *et al.,* 2024) confirmed that AI consistently demonstrated higher average sensitivity than clinicians across a range of pathological findings. A systematic review and meta-analysis for peri-implant marginal bone loss (Atieh *et al.,* 2026) found AI specificity of 0.91 versus clinician specificity of 0.70, representing the most distinctive AI advantage for bone loss tasks. These findings are summarized in Table 3 and Figure 3.

These results collectively suggest that AI functions most effectively as a complementary tool augmenting clinician capability rather than replacing clinical judgment. The greatest performance gains from AI adoption are expected for generalist practitioners managing high radiograph volumes and for expanding diagnostic access in settings where specialist expertise is limited.

Summary Tables

Table 1: Summary of AI Diagnostic Performance Across Dental Radiographic Tasks

Diagnostic Task	Modality	Architecture	Sensitivity	Specificity	Additional Metric	Key Reference
Caries Detection	Bitewing / Periapical / OPG	CNN, YOLO	0.94	0.91	mAP50: 99.5% (YOLO)	Kühnisch *et al.,* 2024; Negi *et al.,* 2024
Periapical Lesion Detection	Periapical radiograph	CNN	0.929	0.586 ⚠	PPV: 15.3% (low prev.)	Ahmed *et al.,* 2025; Sparnon *et al.,* 2026
Alveolar Bone Loss	Panoramic radiograph	YOLOv8, Faster R-CNN	~0.85	~0.87	ICC: 0.80 (severity)	Xue *et al.,* 2024; Wimalasiri *et al.,* 2026
Peri-implant Bone Loss	Periapical / CBCT	CNN, YOLOv8	0.84	0.91	AUC: 0.94; Kappa 0.954	Atieh *et al.,* 2026
Tooth Detection	Panoramic radiograph	YOLO-V4, YOLOv5	0.991 (prec.)	0.992 (rec.)	F1: 99.54%; mAP50: 99.5%	Yilmaz *et al.,* 2023
CBCT Tooth Segmentation	CBCT (3D)	U-Net, V-Net	Dice: 96.5%	—	Mandibular canal: 94.8%	Chen *et al.,* 2022
Vertical Root Fracture	Periapical / CBCT	CNN (ensemble)	75–98%	60–99%	Accuracy: 75–97.8%	Patil *et al.,* 2026; Alshahrani *et al.,* 2026
Cephalometric Landmarks	Lateral cephalogram	CNN, Mask R-CNN	DR: 98.29%	—	Mean error: 1.04–3.40 mm	Ribas-Sabartes *et al.,* 2024
Dental Anomaly Detection	Panoramic radiograph	DetectNet, YOLO	96% accuracy	—	Supernumerary & hypodontia	Makrygiannakis *et al.,* 2025; Uzel *et al.,* 2025

⚠ Low specificity despite high sensitivity — clinicians should be aware of elevated false-positive rates in periapical AI diagnosis. OPG = orthopantomogram; ICC = intra-class correlation coefficient; DR = detection rate; Prec. = precision; Rec. = recall; Prev. = prevalence.

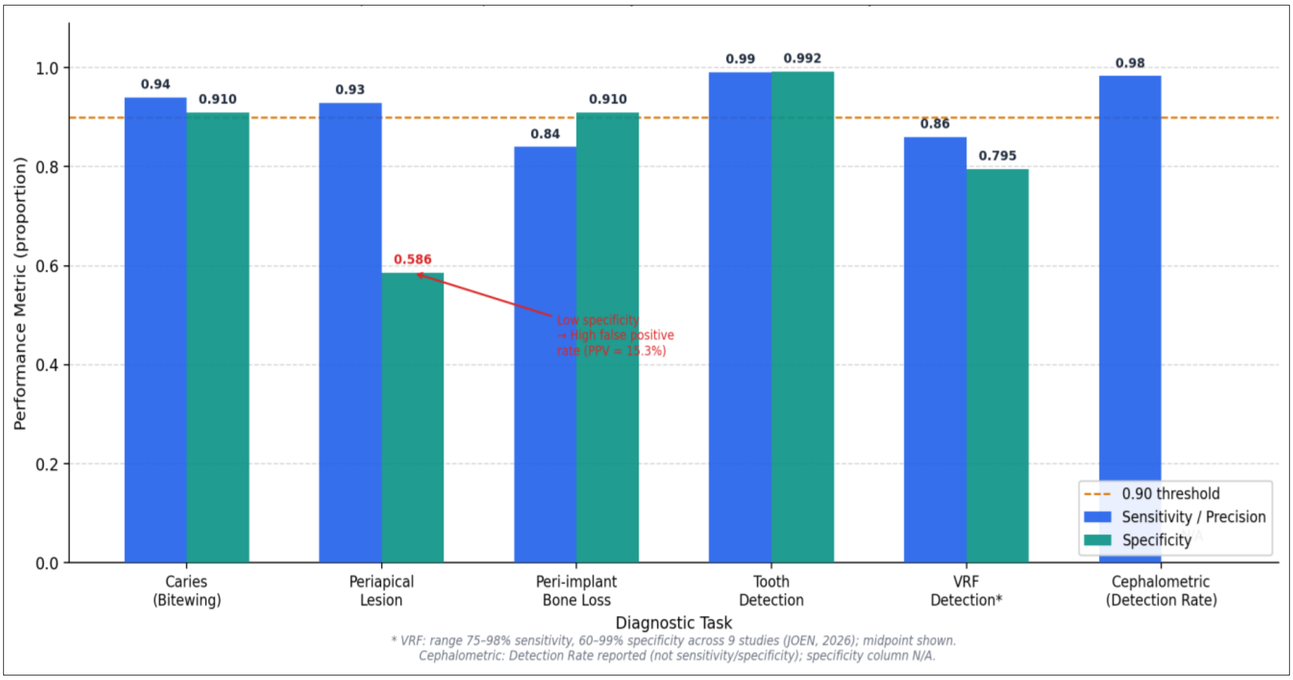

Figure 2: Grouped bar chart showing pooled AI sensitivity and specificity across six diagnostic tasks. The amber dashed line marks the 0.90 performance threshold. The low specificity in periapical lesion detection (0.586) is highlighted in red, reflecting the clinically significant false-positive paradox reported by Sparnon *et al.*, (2026). VRF values represent midpoints of the reported range (sensitivity 75–98%, specificity 60–99%)

Table 2: Comparison of AI Architectures Used in Dental Radiographic Diagnosis

Architecture	Type	Key Dental Application	Best Reported Performance	Key Advantage
CNN (ResNet, VGG, AlexNet)	Deep learning	Caries detection, periapical lesions, VRF detection	Sensitivity 0.94 / Specificity 0.91 (caries, bitewing)	Proven accuracy; widely validated; transfer learning from ImageNet
YOLO (v4, v5, v8)	Object detection	Tooth detection, caries staging, implant classification	Precision 99.90%; mAP50 99.5%; F1 99.54%	Real-time inference; simultaneous multi-class detection
Faster R-CNN	Object detection	Tooth localization, bone loss measurement	Precision 93.67%; recall 90.79%	Strong localization; anchor-based precision
U-Net / V-Net	Segmentation	CBCT tooth and bone segmentation	Dice score 93.6–96.5% across dental structures	Pixel-level precision; ideal for 3D volumetric data
Vision Transformer (ViT)	Transformer	Furcation classification, panoramic caries, multi-label tasks	Highest performance in 58% of dental imaging studies (n=21)	Global attention captures long-range spatial dependencies
Hybrid CNN + ViT	Hybrid	Multi-task dental imaging; panoramic analysis	Best overall performance across radiological image tasks	Combines local texture (CNN) with global context (ViT)
Multimodal LLM (GPT-4o, BLIP-2)	Foundation model	Diagnostic report generation, oral pathology Q&A	Preliminary; 123 oral pathology cases assessed (BMC OralH 2025)	Natural language output; integrates image + clinical context

CNN = convolutional neural network; YOLO = You Only Look Once; R-CNN = region-based CNN; ViT = vision transformer; LLM = large language model; CBCT = cone-beam computed tomography.

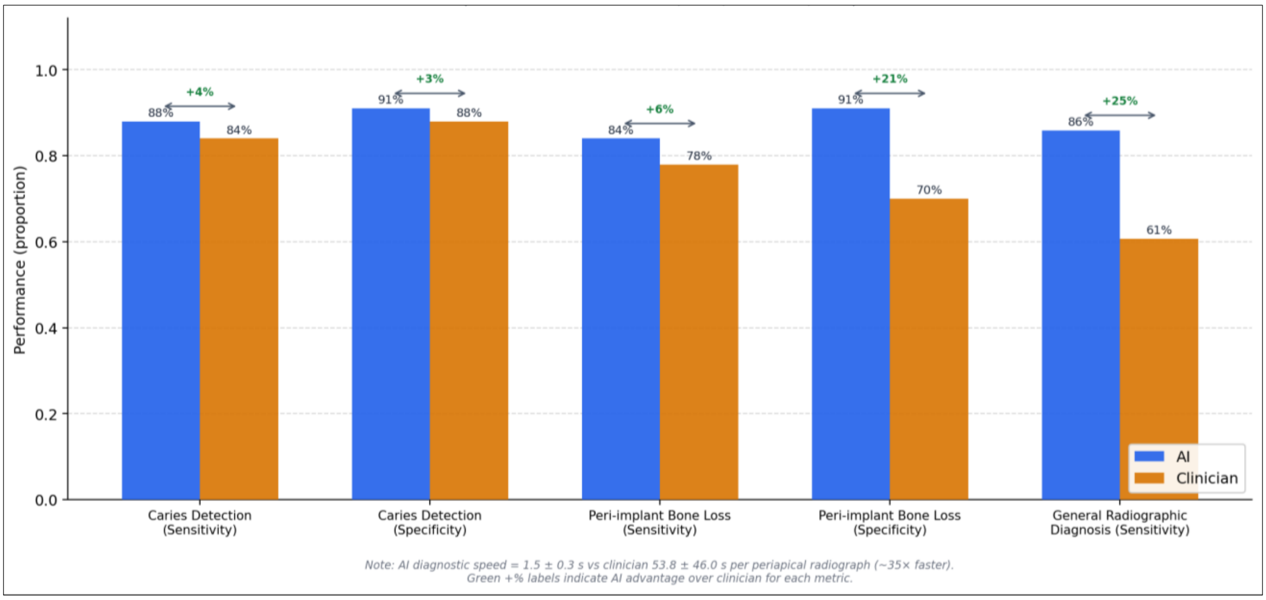

Figure 3: Comparative bar chart of AI versus dental clinician diagnostic performance for three tasks. Green labels indicate the percentage advantage of AI over the clinician. AI diagnostic speed was 1.5 ± 0.3 s versus clinician 53.8 ± 46.0 s per periapical radiograph (~35× faster). Sources: Alabdulkareem *et al.*, (2026); Atieh *et al.*, (2026).

Table 3: AI vs. Dental Clinician Performance Comparison Across Diagnostic Tasks

Diagnostic Task	AI Sensitivity	Clinician Sensitivity	AI Specificity	Clinician Specificity	Additional Finding	Source
Dental Caries Detection	88%	84%	91%	88%	AI: modest advantage on both metrics	Alabdulkareem *et al.*, 2026
General Radiographic Diagnosis	85.9%	60.7%	NR	NR	AI improved clinician sensitivity by +25.2%; speed: 1.5 s vs 53.8 s	Alabdulkareem *et al.*, 2026
Peri-implant Bone Loss	84%	78%	91%	70%	AI specificity advantage: +21%; AUC 0.94	Atieh *et al.*, 2026
Periapical Lesion Detection	92.9%	~85%*	58.6%	~75%*	AI sensitivity high but low specificity → elevated FP rate	Ahmed *et al.*, 2025; Sparnon *et al.*, 2026
Panoramic Radiograph Assessment	Higher sensitivity (overall)	Lower sensitivity	Variable	Variable	AI consistently higher avg. sensitivity across findings	Turosz *et al.*, 2024

*NR = not reported; FP = false positive; * = approximate figures based on comparable studies. Clinician performance varies by experience level; values represent general practitioners unless otherwise noted.*

Table 4: Key Limitations, Evidence Gaps, and Recommended Actions in Dental AI Research

Challenge	Current Status	Supporting Evidence	Recommended Action
Demographic & geographic dataset bias	Only 31.2% of dental AI datasets report ethical approval; marked geographic skew toward high-income countries	Uribe *et al.*, 2024 (JDR) — characterization of 100+ publicly available dental AI datasets	Mandate FAIR data reporting; recruit datasets from LMICs; require demographic metadata in all published AI studies
High false-positive rates (periapical)	AI periapical PPV: 15.3% despite 92.9% sensitivity; specificity 58.6%	Sparnon *et al.*, 2026 — false positive paradox analysis	Report prevalence-adjusted metrics (PPV, FDR); set clinical decision thresholds based on local prevalence; never deploy AI periapical screening without clinician verification

Challenge	Current Status	Supporting Evidence	Recommended Action
Limited external validation	Most published models validated on single-centre or single-country datasets; prospective multi-centre validation studies are sparse	Schwendicke *et al.,* 2020 (JDR); reporting checklist review PMC 2025	Require prospective multi-centre external validation before any clinical approval recommendation; register validation studies on ClinicalTrials.gov
Inconsistent reporting quality	Majority of studies do not report FDR, FOR, or confidence intervals; only minority achieve high APPRAISE-AI scores	Khurshid *et al.,* 2025 — AI reporting checklists review	Mandate APPRAISE-AI, STARD, and TRIPOD-AI reporting; journals should require reporting checklist as supplementary material
Model interpretability ("black box")	Most CNN/YOLO models provide probability outputs without visual explanation; Grad-CAM used inconsistently	Schwendicke *et al.,* 2020; general XAI literature	Standardize use of Grad-CAM or SHAP visualizations in dental AI publications; require XAI output for clinical deployment
Regulatory fragmentation	ANSI/ADA 1110-1:2025 is the first U.S. standard; covers 2D radiograph annotation only; CBCT and multimodal AI not yet addressed	American Dental Association, 2025 — ANSI/ADA Standard 1110-1:2025	Extend standardization to CBCT and multimodal AI inputs; align with FDA SaMD framework; pursue ISO/TC 106 harmonization

FAIR = Findable, Accessible, Interoperable, Reusable; LMICs = Low- and Middle-Income Countries; PPV = Positive Predictive Value; FDR = False Discovery Rate; FOR = False Omission Rate; Grad-CAM = Gradient-weighted Class Activation Mapping; SHAP = SHapley Additive exPlanations; SaMD = Software as a Medical Device; ADA = American Dental Association; ISO/TC 106 = ISO Technical Committee for Dentistry.

Limitations and Challenges in AI-Based Dental Radiographic Diagnosis

Despite the considerable diagnostic accuracy reported across multiple domains, AI in dental radiographic diagnosis faces substantive and interrelated challenges summarized in Table 4.

Dataset quality, geographic diversity, and demographic representativeness are primary concerns. A landmark study by Uribe *et al.,* (2024), published in the Journal of Dental Research, characterized all publicly available dental imaging datasets for AI, examining resources from 2011 to 2023. Only 31.2% of datasets reported ethical approval and 56.25% did not specify a data use license. Geographic representation was markedly skewed toward high-income countries, raising serious concerns about algorithmic bias and the potential for AI systems to perform inequitably when deployed in diverse global clinical settings (Uribe *et al.,* 2024).

The false positive paradox, particularly in periapical lesion assessment, represents a significant clinical concern. As demonstrated above, a model with high sensitivity but low specificity will generate clinically unacceptable numbers of false positives when disease prevalence is modest, and reporting of positive predictive value alongside sensitivity and specificity should be mandated in all dental AI studies (Sparnon *et al.,* 2026).

Reporting quality is inconsistent: a review of AI reporting standards in dentistry (Khurshid *et al.,* 2025) found that the majority of published studies did not achieve high-quality scores across established frameworks, and that clinically important metrics such as false discovery rate and false omission rate were seldom reported. The interpretability of deep learning models—the "black box" problem—remains a barrier to clinical trust. ANSI/ADA Standard No. 1110-1:2025, the first formal U.S. standard for dental AI, now provides guidance on validation dataset annotation and data collection for 2D radiographs, representing a critical first step toward regulatory harmonization (American Dental Association, 2025).

Future Directions

Multimodal LLMs represent the most transformative emerging direction. A study in the International Journal of Oral Science explored the potential of multi-modal ChatGPT for dentistry, identifying radiograph interpretation and structured report generation as high-priority applications (Huang *et al.,* 2023). A study evaluating multimodal LLM diagnostic performance in oral pathology, published in PubMed Central in 2025 (Suarez *et al.,* 2025), demonstrated promising accuracy for oral lesion identification. A BLIP-2 and Llama 3 8B-based pipeline for automated OPG report generation was evaluated in 2025, representing the first proof-of-concept for end-to-end multimodal dental AI reporting (Dasanayaka *et al.,* 2025).

Federated learning frameworks, in which AI models are trained collaboratively across multiple institutions without sharing patient-identifiable data, offer a principled solution to the dataset diversity and privacy challenges currently constraining model generalizability. ANSI/ADA Standard No. 1110-1:2025 and the accompanying ANSI/ADA 1109:2025 technical report provide foundational regulatory guidance for dental AI validation. The broader integration of AI into dental practice management software, combined with clinician training and transparent AI limitation disclosures, will be essential for responsible real-world deployment.

CONCLUSION

This comprehensive narrative review demonstrates that AI has achieved substantial and reproducible diagnostic accuracy across ten dental radiographic domains. Performance metrics consistently approach or match experienced clinicians in controlled settings, and AI provides a nearly 35-fold advantage in diagnostic speed. The emergence of vision transformers outperforming CNNs in 58% of dental imaging tasks, the maturation of CBCT-based AI systems achieving Dice scores above 93%, the development of ensemble AI for VRF detection (accuracy 75–97.8%), and early evidence for multimodal LLMs in dental report generation collectively indicate that the field is entering a period of rapid architectural and functional advancement. However, high false-positive rates, marked demographic and geographic bias in training datasets, limited multi-centre external validation, inconsistent reporting quality, and regulatory complexity remain unresolved challenges. The publication of ANSI/ADA Standard 1110-1:2025 represents a critical first step toward regulatory and methodological standardization. Rigorous prospective validation across diverse patient populations, adoption of standardized reporting frameworks, development of interpretable AI tools, and greater regulatory engagement are essential prerequisites for the safe and beneficial integration of AI into dental radiographic diagnosis.

Conflict of Interest: The authors declare no conflicts of interest.

REFERENCES

- Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005
- Schwendicke, F., Samek, W., & Krois, J. (2020). Artificial intelligence in dentistry: chances and challenges. Journal of Dental Research, 99(7), 769–774. https://doi.org/10.1177/0022034520915714
- Negi, A., Goyal, S., & Singh, M. (2024). Artificial intelligence in dental caries diagnosis and detection: an umbrella review. Clinical and Experimental Dental Research, 10(5), e70004. https://doi.org/10.1002/cre2.70004
- Mertens, S., Krois, J., & Schwendicke, F. (2024). Artificial intelligence for radiographic imaging detection of caries lesions: a systematic review. BMC Oral Health, 24, 283. https://doi.org/10.1186/s12903-024-04046-7
- Carvalho, B. K. G., Nolden, E.-L., Wenning, A. S., Kiss-Dala, S., Agócs, G., Róth, I., Kerémi, B., Géczi, Z., Hegyi, P., & Kivovics, M. (2024). Diagnostic accuracy of artificial intelligence for approximal caries on bitewing radiographs: a systematic review and meta-analysis. Journal of Dentistry, 151, 105388. https://doi.org/10.1016/j.jdent.2024.105388
- Pornprasertsuk-Damrongsri, S., Vachmanus, S., Papasratorn, D., Kitisubkanchana, J., Chaikantha, S., Arayasantiparb, R., & Mongkolwat, P. (2025). Clinical application of deep learning for enhanced multistage caries detection in panoramic radiographs. Scientific Reports, 15(1), 33491. https://doi.org/10.1038/s41598-025-16591-4
- Makaremi, M., Lacaule, C., & Mohammad-Djafari, A. (2025). Detection of periapical lesions using artificial intelligence: a narrative review. PubMed Central. https://pmc.ncbi.nlm.nih.gov/articles/PMC12839604/
- Ahmed S, Masood R, Rashid I, Jamil Haq A, Asif R, Sajjad ST, Tasnim N, Ullah Khan W. (2025). Diagnostic accuracy of AI in radiographic detection of dental caries and periapical lesions. Link Medical Journal of Health and Community Research, 3(1). https://doi.org/10.61919/4w0nfp45
- Sparnon E, Stevens K, Song EC, Harris RJ, Strong BW, Bruno MA, *et al.,* (2026). The false positive paradox: examining real-world clinical implications of AI diagnostic accuracy in dental imaging. medRxiv. https://doi.org/10.64898/2026.03.25.26349197
- Chen, C.-C., Wu, Y.-F., Aung, L. M., Lin, J. C.-Y., Ngo, S. T., Su, J.-N., Lin, Y.-M., & Chang, W.-J. (2023). Automatic recognition of teeth and periodontal bone loss measurement in digital radiographs using deep-learning artificial intelligence. Journal of Dental Sciences, 18(3), 1301–1309. https://doi.org/10.1016/j.jds.2023.03.020
- Xue, T., Chen, L., & Sun, Q. (2024). Deep learning method to automatically diagnose periodontal bone loss and periodontitis stage in dental panoramic radiograph. Journal of Dentistry, 150, 105373. https://doi.org/10.1016/j.jdent.2024.105373
- Wimalasiri C, Rathnayake P, Wijerathne S, Rasnayaka S, Leuke Bandara D, *et al.,* (2026). AI-assisted radiographic analysis in detecting alveolar bone-loss severity and patterns. Scientific Reports. https://doi.org/10.1038/s41598-026-38061-1

- Lee SW, Huz K, Gorelick K, Li J, Bina T, Matsumura S, Yin N, Zhang N, Anang YNA, McMahon DJ, Lu HH, Yin MT, Wadhwa S. (2025). Evaluation by dental professionals of an artificial intelligence-based application to measure alveolar bone loss. BMC Oral Health. https://doi.org/10.1186/s12903-025-05677-0
- Atieh, M. A., & co-authors. (2026). Diagnostic accuracy of deep learning models in detecting peri-implant marginal bone loss: a systematic review and meta-analysis. Clinical Oral Implants Research. https://doi.org/10.1111/clr.70090
- Madani Z, Fakhar HB. (2026). Detection and classification of peri-implant marginal bone loss in cone-beam computed tomography using a deep learning approach. Clinical and Experimental Dental Research. https://doi.org/10.1002/cre2.70308
- Bonfanti-Gris M, Herrera A, Salido Rodriguez-Manzaneque MP, Martinez-Rus F, Pradies G. (2025). Deep learning for tooth detection and segmentation in panoramic radiographs: a systematic review and meta-analysis. BMC Oral Health. https://doi.org/10.1186/s12903-025-06349-9
- Yilmaz, E., Guvenc, M. T., & Kayipmaz, S. (2023). Developing deep learning methods for classification of teeth in dental panoramic radiography. Seminars in Orthodontics. https://doi.org/10.1016/j.sodo.2023.01.003
- He Z, Wang Y, Li X. (2024). Deep learning-based detection of impacted teeth on panoramic radiographs. Biomedical Engineering and Computational Biology. https://doi.org/10.1177/11795972241288319
- Khurshid Z, Faridoon F, Silkosessak OC, Trachoo V, Waqas M, Hasan S, Porntaveetus T. (2025). Multi-regional deep learning models for identifying dental restorations and prosthesis in panoramic radiographs. BMC Oral Health. https://doi.org/10.1186/s12903-025-07138-0
- Chen, H., Zhang, K., Lyu, P., Li, H., Zhang, L., Wu, J., & Lee, C. H. (2022). A fully automatic AI system for tooth and alveolar bone segmentation from cone-beam CT images. Nature Communications, 13, 2096. https://doi.org/10.1038/s41467-022-29637-2
- Tarce M, Zhou Y, Antonelli A, Becker K. (2024). The application of artificial intelligence for tooth segmentation in CBCT images: a systematic review. Applied Sciences, 14(14), 6298. https://doi.org/10.3390/app14146298
- Mohammed M, Fernandez-Medina T, Rajashekhar M, Baker S, Jennings E. (2025). Alveolar bone segmentation methods in assessing the effectiveness of periodontal defect regeneration through machine learning of CBCT data: a systematic review. International Journal of Biomedical Imaging. https://doi.org/10.1155/ijbi/9065572
- Zhou W, Zhao D, Zhao B, Shen M, Liu X. (2025). A deep learning framework for automated dental segmentation and diagnostic report generation from cone-beam computed tomography. Head and Face Medicine. https://doi.org/10.1186/s13005-025-00555-0
- Unal T, Kuran A, Gulsen IT, Kizilay FN, Gulsen E, Ozudogru S, Gordeli K, Celik O, Ugurlu M, Bayrakdar IS, Orhan K. (2025). Deep learning-based 3D automatic segmentation of impacted canines in CBCT scans. BMC Oral Health. https://doi.org/10.1186/s12903-025-07117-5
- Felek, T., Tercanli, H., & Gök, R. Ş. (2025). Evaluating vision transformers and convolutional neural networks in the context of dental image processing: a systematic review. BMC Oral Health, 25(1), 1626. https://doi.org/10.1186/s12903-025-07036-5
- Zhang X, Guo E, Liu X, Zhao H, Yang J, Li W, Wu W, Sun W. (2025). Enhancing furcation involvement classification on panoramic radiographs with vision transformers. BMC Oral Health. https://doi.org/10.1186/s12903-025-05431-6
- Wang L, Li Z. (2026). Transformer-based intelligent detection model for early dental caries in panoramic radiographs. Scientific Reports. https://doi.org/10.1038/s41598-025-33391-y
- Kim JW, Urooj Khan A, Banerjee I. (2024). Systematic review of hybrid vision transformer architectures for radiological image analysis. medRxiv. https://doi.org/10.1101/2024.06.21.24309265
- Alshahrani, A. S., Alelyani, A. A., Jabali, A., Al Malwi, A. A., Alroomy, R., Shaiban, A. S., Almnea, R. A., Mehta, V., & Al Moaleem, M. M. (2026). Use of artificial intelligence in diagnosing vertical root fractures: a systematic review. Diagnostics, 16(3), 406. https://doi.org/10.3390/diagnostics16030406
- Patil S, Licari FW, Bennion C, Deshpande A, Nasseh AA, Bhandi S. (2026). Effectiveness of artificial intelligence models for detection of vertical root fractures: a systematic review. Journal of Endodontics, 52(6), 861-870. https://doi.org/10.1016/j.joen.2026.02.012
- Abdelazim R, Fouad EM. (2024). Artificial intelligent-driven decision-making for automating root fracture detection in periapical radiographs. BDJ Open. https://doi.org/10.1038/s41405-024-00260-1
- Tasdemir T, Tassoker M, Arpaci A, *et al.,* (2026). Diagnostic performance of artificial intelligence in periapical radiography: a systematic review. Odontology. https://doi.org/10.1007/s10266-026-01401-8
- Ribas-Sabartes, J., Sánchez-Molins, M., & d'Oliveira, N. G. (2024). The accuracy of algorithms used by artificial intelligence in cephalometric points detection: a systematic review.

Bioengineering, 11(12), 1286. https://doi.org/10.3390/bioengineering11121286
- Schwendicke, F., Chaurasia, A., Arsiwala, L., Lee, J.-H., Elhennawy, K., Jost-Brinkmann, P.-G., Demarco, F., & Krois, J. (2021). Deep learning for cephalometric landmark detection: systematic review and meta-analysis. Clinical Oral Investigations, 25(7), 4299–4309. https://doi.org/10.1007/s00784-021-03990-w
- Jiao, Z., Liang, Z., Liao, Q., Chen, S., Yang, H., Hong, G., & Gui, H. (2024). Deep learning for automatic detection of cephalometric landmarks on lateral cephalometric radiographs using the Mask Region-based Convolutional Neural Network: a pilot study. Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology, 137(5), 554–562. https://doi.org/10.1016/j.oooo.2024.02.003
- Neeraja R, Jani Anbarasi L. (2025). A critical review of artificial intelligence based techniques for automatic prediction of cephalometric landmarks. Artificial Intelligence Review. https://doi.org/10.1007/s10462-025-11135-8
- Makrygiannakis, M. A., Giannakopoulos, K., Kavadella, A., Paraskevis, D., & Kaklamanos, E. G. (2025). Diagnostic accuracy of an artificial intelligence-based software in detecting supernumerary and congenitally missing teeth in panoramic radiographs. European Journal of Orthodontics, 47(4), cjaf054. https://doi.org/10.1093/ejo/cjaf054
- Uzel I, Ghabchi B, Cogulu D. (2025). Deep learning-based automated detection of supernumerary teeth in pediatric panoramic radiographs. PLOS One. https://doi.org/10.1371/journal.pone.0335845
- Tunc H, Akkaya N, Aykanat B, Unsal G. (2025). U-Net-based deep learning for simultaneous segmentation and agenesis detection of primary and permanent teeth in panoramic radiographs. Diagnostics, 15(20). https://doi.org/10.3390/diagnostics15202577
- Alabdulkareem, M., Atieh, M., AbuMostafa, A., Aldalaan, K., & Alturki, N. (2026). Artificial intelligence in dental treatment planning and diagnostic decision-making: a systematic review and meta-analysis. Clinical and Experimental Dental Research, 12(2), e70343. https://doi.org/10.1002/cre2.70343
- Turosz, N., Chęcińska, K., Chęciński, M., Sielski, M., & Sikora, M. (2024). Evaluation of dental panoramic radiographs by artificial intelligence compared to human reference: a diagnostic accuracy study. Journal of Clinical Medicine, 13(22), 6859. https://doi.org/10.3390/jcm13226859
- Ali, M., Irfan, M., Ali, T., Wei, C. R., & Akilimali, A. (2025). Artificial intelligence in dental radiology: a narrative review. Annals of Medicine and Surgery, 87(4), 2212–2217. https://doi.org/10.1097/MS9.0000000000003127
- Uribe, S. E., Issa, J., Sohrabniya, F., Denny, A., Kim, N. N., Dayo, A. F., Chaurasia, A., Sofi-Mahmudi, A., Büttner, M., & Schwendicke, F. (2024). Publicly available dental image datasets for artificial intelligence. Journal of Dental Research, 103(12), 1195–1205. https://doi.org/10.1177/00220345241272052
- Khurshid, Z., Osathanon, T., Shire, M. A., Schwendicke, F., & Samaranayake, L. (2025). Artificial intelligence in dentistry: a concise review of reporting checklists and guidelines. International Dental Journal, 76(1), 109322. https://doi.org/10.1016/j.identj.2025.109322
- American Dental Association. (2025). ANSI/ADA Standard No. 1110-1:2025. Dentistry – Validation Dataset Guidance for Image Analysis Systems Using Artificial Intelligence, Part 1: Image Annotation and Data Collection. American Dental Association. https://engage.ada.org/p/eg/ansiada-standard-no-1110-1
- Huang, H., Zheng, O., Wang, D., Yin, J., Wang, Z., Ding, S., Yin, H., Xu, C., Yang, R., Zheng, Q., *et al.,* (2023). ChatGPT for shaping the future of dentistry: the potential of multi-modal large language model. International Journal of Oral Science, 15, 29. https://doi.org/10.1038/s41368-023-00239-y
- Suarez A, Freire Y, Suarez M, Diaz-Flores Garcia V, Andreu-Vazquez C, Thuissard Vasallo IJ, Castillo Varon AI, Martin C. (2025). Diagnostic performance of multimodal large language models in the analysis of oral pathology. Oral Diseases, 31(12), 3344-3354. https://doi.org/10.1111/odi.70009
- Dasanayaka C, *et al.,* (2025). Multimodal AI and large language models for orthopantomography radiology report generation and Q&A. BioMedInformatics, 8(2), 39. https://www.mdpi.com/2571-5577/8/2/39