

The Impact Of Artificial Intelligence (AI) In Various Dental Specialties: A Systematic Review

Noor salah alyousef¹ , Fatmah Almaqboul² , Rawan Numayan Albishry³ , Fahad Shami Alamri⁴ , Dr . Salman Ahmad Maree⁵ , Aljawharah Ali Aqely⁶ , Hanan Ghazi D. Alzayer⁷ , Dr.Sukayna Alawi Alsadah⁸ , Mohrah Mohammed Alharbi⁹ , Alaa Ali Alsubki¹⁰ , Ahmed Hassan Alhamoud¹¹ , Reem Mansour Alabdan¹²

¹Buraidah college General Dentist almntazah dispensary-MOH
Dr.noor.alyousf@gmail.com

²Saudi Board Endodontics Resident MOH 0543366813 fa.almaqboul@gmail.com

³General Dentist-MOHRawan.albishry@gmail.com 0555160182

⁴General Dentist-MOH Alamrident@gmail.com 0535080307

⁵BDS : College of dentistry, King Khalid University . General Dentist-MOH Kingdom of Saudi Arabia alharoobi11@hotmail.com Smaree@moh.gov.sa 0542118802

⁶General Dentist-MOH aljawhara.341@gmail.com 0547922200

⁷Ajman university of science and technology General Dentist – MOH Qatif primary health care Noun_222@hotmail.com

⁸King Saud University General Dentist -MOH Salsadah@hotmail.com 0506810839

⁹Dental assistant Qassim University alqasim bariduh Markaz almntaza 0558455451

¹⁰Vision college General dentist Ministry of Health allooo_1994@hotmail.com 0500327282

¹¹General Dentist Abqaiq - Estern cluster Ministry of health ahmed.1381413@gmail.com 0564423369

¹²General dentist Qassim Armed Forces Hospital reem.alabdan@gmail.com 0555256061

Abstract

Background:

Artificial intelligence (AI), including machine learning and deep learning approaches, is increasingly integrated into dental diagnostics and treatment planning. Across dental specialties, AI has shown potential to enhance diagnostic accuracy, optimize workflows, and support clinical decision-making. However, the breadth of applications and the performance consistency of AI systems remain variable, necessitating a comprehensive synthesis of current evidence.

Objective:

To systematically evaluate the impact, diagnostic performance, and clinical utility of artificial intelligence applications across various dental specialties.

Methods:

A systematic review was conducted following PRISMA guidelines. A comprehensive search of PubMed, ScienceDirect, Google Scholar, and the Cochrane Library was performed to identify studies assessing AI applications in dentistry. Eligible studies included primary research evaluating AI-based diagnostic, predictive, or treatment-planning tools across dental specialties. Data extraction encompassed study design, AI model characteristics, dataset size, performance metrics, and key findings. Quality assessment was performed using QUADAS-2 for diagnostic studies, the Newcastle–Ottawa Scale for observational studies, and the Cochrane

Risk of Bias tool for trials. Due to significant heterogeneity in methodologies, datasets, and performance indices, a narrative synthesis was conducted instead of a meta-analysis.

Results:

Fifteen studies met the inclusion criteria. AI applications were most commonly studied in periodontology (n=5) and orthodontics (n=5), followed by endodontics (n=3), cariology/radiology (n=2), and paediatric dentistry (n=1). Deep learning architectures—particularly CNNs, U-Net models, Faster R-CNN, Mask R-CNN, and YOLO variants—demonstrated strong diagnostic performance in tasks such as detecting periodontal bone loss (AUC up to 0.951), segmenting caries lesions (accuracy up to 0.93), identifying cephalometric landmarks (2 mm success rates >90%), and detecting periapical pathology or fractured instruments. Several AI systems matched or exceeded clinician performance, especially in radiographic interpretation. However, limitations included small datasets, lack of external validation, variability in ground truth labeling, and reduced performance for subtle or complex pathologies.

Conclusion:

AI demonstrates significant promise across multiple dental specialties, showing high diagnostic accuracy and strong potential as a clinical decision-support tool. Despite encouraging results, widespread adoption is limited by methodological heterogeneity, insufficient validation, and inconsistent reporting standards. Future research should emphasize larger, multi-center datasets, external validation, transparency in model development, and real-world clinical trials to strengthen evidence for clinical integration.

Keywords: Artificial intelligence, deep learning, machine learning, dentistry, diagnostic accuracy, radiographic analysis, dental specialties, convolutional neural networks.

Introduction

Artificial intelligence (AI), encompassing machine learning (ML) and deep learning (DL) techniques, is rapidly reshaping the landscape of healthcare — and dentistry is no exception. In recent years, AI applications have proliferated across dental specialties, promising to enhance diagnostic accuracy, streamline treatment planning, and ultimately improve patient outcomes. By leveraging large datasets and advanced algorithms, AI systems can detect subtle pathologies in imaging, predict treatment trajectories, and assist clinicians in making data-driven decisions [1]. The potential of AI in dentistry is broad: from interpreting radiographs and cone-beam computed tomography (CBCT) images, to optimizing prosthodontic designs, to predicting the prognosis of periodontal disease [2,3]. For instance, deep convolutional neural networks have shown strong performance in image-based diagnostics, often matching or even surpassing the accuracy of experienced human experts in tasks such as cephalometric landmark detection or caries segmentation [2]. In implant dentistry specifically, AI has been used not only to plan implant placement, but also to classify bone quality and predict long-term prognosis, underscoring its multifaceted role in this specialty [4]. Moreover, in resource-limited settings, the adoption of AI is being explored as a way to augment clinical capacity and bridge gaps in care — scoping reviews from lower- and middle-income countries have highlighted both successes and persistent challenges in such contexts [5]. Despite these promising developments, several critical issues remain. Methodological heterogeneity, bias in training datasets, limited clinical validation, and ethical concerns such as transparency and accountability continue to hinder the translation of AI models into routine dental practice [6]. Furthermore, variability in regulatory frameworks and the lack of standardized protocols complicate the integration of AI tools into real-world workflows. Given this rapidly evolving yet fragmented landscape, a systematic synthesis of current evidence is needed. While prior narrative reviews have outlined broad trends in AI adoption in dentistry [7,8] a focused, specialty-

wise analysis can provide deeper insight into where AI is most mature, where the research gaps lie, and how future work can address barriers to clinical implementation. Therefore, this systematic review aims to critically evaluate and synthesize the impact of AI across various dental specialties. By mapping the current applications, performance metrics, and real-world implications of AI in fields such as periodontology, orthodontics, implantology, restorative dentistry, and more, we seek to inform both researchers and practitioners about the current state of the art, as well as future directions for the field.

Methodology

Study Design

This study was designed as a systematic review aimed at synthesizing current evidence on the applications and impact of artificial intelligence across various dental specialties. The methodology followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to ensure transparency, reproducibility, and scientific rigor throughout the review process.

Eligibility Criteria

Studies were considered eligible if they evaluated the use, performance, or clinical impact of AI technologies in any dental specialty, including but not limited to orthodontics, oral and maxillofacial radiology, restorative dentistry, periodontology, endodontics, prosthodontics, oral surgery, and pediatric dentistry. Primary research articles such as randomized controlled trials, observational studies, diagnostic accuracy studies, machine-learning model development or validation studies, and relevant preclinical research were included. Only peer-reviewed articles published in English were considered. Editorials, conference abstracts, commentaries, case reports, reviews, and studies not directly assessing AI applications in dentistry were excluded.

Information Sources and Search Strategy

A comprehensive electronic search was conducted in major scientific databases, including PubMed, ScienceDirect, Google Scholar, and the Cochrane Library. The search covered all literature available up to the time of review. Keywords and controlled vocabulary terms related to “artificial intelligence,” “machine learning,” “deep learning,” “neural networks,” “dentistry,” and individual dental specialties were used. Boolean operators and database-specific filters were applied to optimize retrieval. Additionally, reference lists of included studies were screened to identify any relevant articles missed by the main search strategy.

Study Selection Process

All search results were imported into a reference management software to remove duplicates. Two independent reviewers screened titles and abstracts to determine initial relevance. Full texts were then retrieved for studies meeting the screening criteria or when eligibility remained unclear. Final inclusion was determined after thorough full-text assessment by the same reviewers. Disagreements between reviewers were resolved through discussion or consultation with a third reviewer to ensure objectivity.

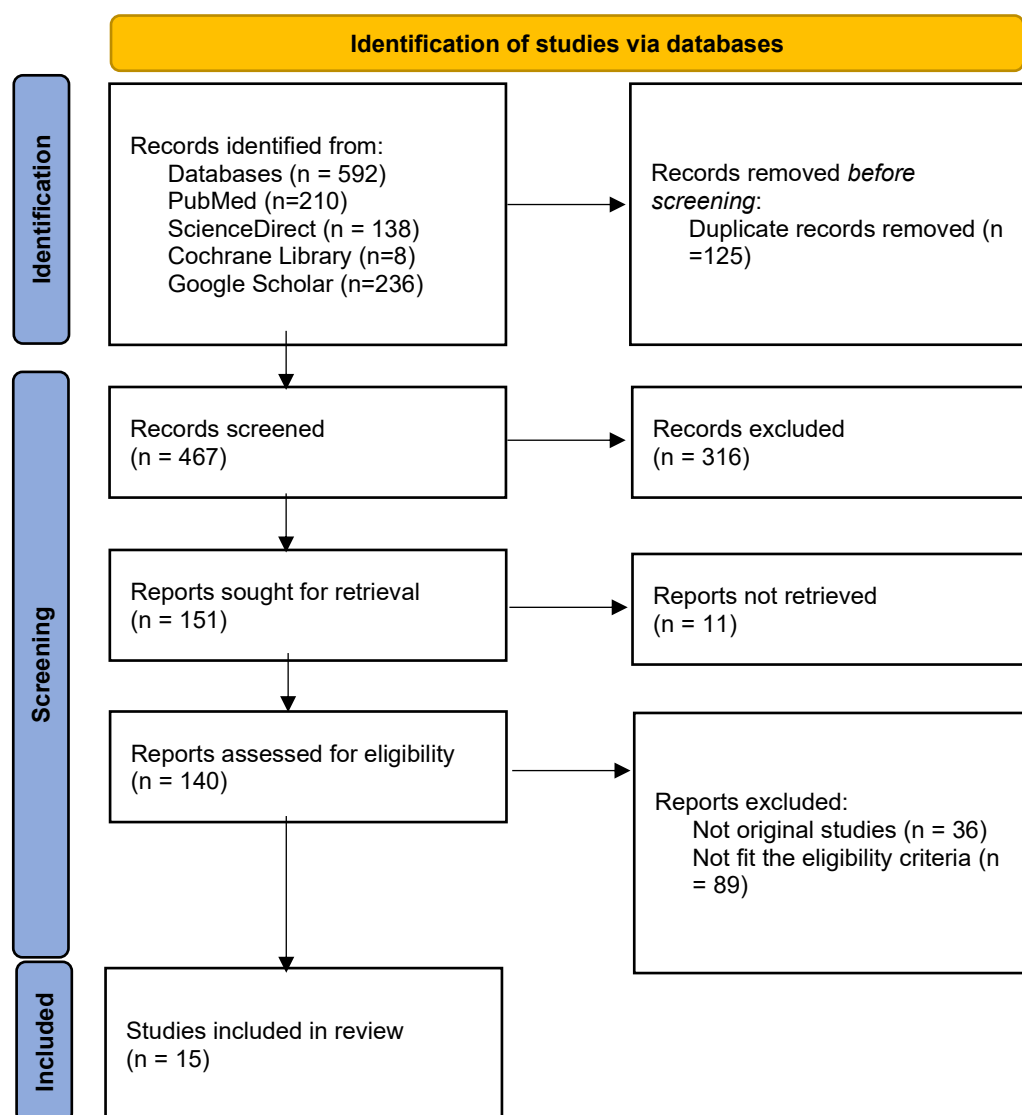


Figure 1: PRISMA flowchart showing the study selection process.

Data Extraction

Data extraction was performed independently by two reviewers using a standardized data extraction template designed for this review. Extracted information included study characteristics, dental specialty, type of AI model or algorithm, dataset characteristics, primary objectives, evaluation metrics, and key findings. Whenever relevant, clinical performance measures such as accuracy, sensitivity, specificity, precision, and area under the curve (AUC) were recorded. Any discrepancies in extracted data were resolved by consensus.

Quality Assessment

The methodological quality of the included studies was assessed using appropriate tools based on study design. Diagnostic accuracy studies were evaluated using the QUADAS-2 tool, observational studies were assessed with the Newcastle–Ottawa Scale, and randomized trials were evaluated using the Cochrane Risk of Bias tool. For AI model development or validation studies, additional considerations such as dataset adequacy, risk of overfitting, external validation, and transparent reporting of model architecture were also noted. Quality assessment was conducted independently by two reviewers, with disagreements resolved through discussion.

Data Synthesis

Given the anticipated heterogeneity in AI models, datasets, evaluation methods, and dental specialties, a meta-analysis was not planned. Instead, a narrative synthesis approach was employed to summarize findings across studies. Results were grouped according to dental specialty and AI application type, with emphasis on model performance, clinical relevance, and reported benefits or limitations. Patterns, consistencies, and variations among studies were highlighted to provide an integrated understanding of the current role and impact of AI in dentistry.

Results

The initial database search yielded 592 records. After removing 125 duplicates, 467 titles and abstracts were screened for relevance. Following this screening, 151 full-text articles were assessed for eligibility. A total of 15 studies met the eligibility criteria and were included in this systematic review. The included studies spanned multiple dental specialties, with the highest representation in periodontology (n=5), orthodontics (n=5), followed by endodontics (n=3), cariology/radiology (n=2), and paediatric dentistry (n=1). Study designs included diagnostic accuracy studies, model development and validation studies, and comparative evaluations between AI systems and human clinicians. Table 1 provides a detailed summary of the characteristics of the included studies, while

Table 1: Summary Characteristics of Included Studies

Author (Year)	Dental Specialty	Study Design / AI Task	AI Model / Algorithm	Dataset Characteristics (n)	Primary Objective
Bennasar et al. (2025) [9]	Endodontics	Diagnostic Accuracy Study	ResNet-18 (DL), Logistic Regression, Random Forest (ML)	119 patients (108 periapical radiographs)	To evaluate the efficacy of DL vs. ML models in predicting non-surgical root canal treatment (NSRCT) outcomes and compare them to clinician prognosis.
Boztuna et al. (2024) [10]	Endodontics / Radiology	Model Development & Validation	U ² -Net (for segmentation)	400 panoramic radiographs (780 periapical lesions)	To assess the diagnostic accuracy of a U ² -Net model in segmenting periapical lesions on panoramic radiographs.

Author (Year)	Dental Specialty	Study Design / AI Task	AI Model / Algorithm	Dataset Characteristics (n)	Primary Objective
Çatmabacak & Çetinkaya (2025) [11]	Endodontics	Model Development & Validation	DenseNet201, EfficientNet B0, ResNet-18, VGG-19, MaxVit-T	700 periapical radiographs (381 with fractured instruments)	To evaluate and compare the performance of five deep learning models in detecting fractured endodontic instruments (FEIs) in periapical radiographs.
Chang et al. (2020) [12]	Periodontology	Model Development & Validation	Hybrid: Mask R-CNN (detection) + Conventional CAD (classification)	330-73 panoramic radiographs (for different structures)	To automatically diagnose periodontal bone loss and stage periodontitis using a hybrid deep learning and conventional CAD method.
Davidovitch et al. (2022) [13]	Orthodontics	Diagnostic Accuracy Study	Convolutional Neural Network (CNN) - Algoceph® (CephX)	10 lateral cephalometric radiographs (21 landmarks each)	To compare the accuracy and reliability of AI vs. human observers for cephalometric landmark identification.

Author (Year)	Dental Specialty	Study Design / AI Task	AI Model / Algorithm	Dataset Characteristics (n)	Primary Objective
AlGhaihab et al. (2025) [14]	Periodontology / Radiology	Diagnostic Accuracy Study (Observer Performance)	Denti.AI (Faster R-CNN, ResNet, FPN)	26 intraoral radiographs (14 periapical, 12 bitewing); 10 dentists	To investigate the impact of a DL software (Denti.AI) on general dentists' ability to detect radiographic bone loss (RBL) in intraoral radiographs.
Gao & Tang (2025) [15]	Orthodontics	Model Development & Validation	DeepFuse (Multi-modal CNN with attention-guided fusion)	3 combined datasets (CephNet, OrthoFace, DentalFusion); 1,608 cases for treatment prediction	To develop a multi-modal deep learning framework for simultaneous landmark detection and orthodontic treatment outcome prediction.
Krois et al. (2019) [16]	Periodontology	Diagnostic Accuracy Study	Custom 7-layer Feed-Forward CNN	2,001 panoramic radiograph segments (tooth-level)	To assess the diagnostic performance of a CNN for detecting periodontal bone loss (PBL) on panoramic radiographs compared to dentists.

Author (Year)	Dental Specialty	Study Design / AI Task	AI Model / Algorithm	Dataset Characteristics (n)	Primary Objective
Kurt-Bayrakdar et al. (2024) [17]	Periodontology	Model Development & Validation	U-Net Architecture (for segmentation)	1,121 panoramic radiographs (30,369 total labels)	To develop a deep learning algorithm for detecting and segmenting periodontal bone loss patterns (horizontal, vertical) and furcation defects.
Lee et al. (2020) [18]	Orthodontics	Model Development & Validation	Bayesian Convolutional Neural Network (BCNN) with a two-stage framework (LRS & HRS)	400 lateral cephalograms (150 train, 250 test)	To develop an automated framework for locating cephalometric landmarks with confidence regions based on uncertainty.
Makrygiannakis et al. (2025) [19]	Orthodontics	Diagnostic Accuracy Study	Diagnocat™ (Proprietary AI model)	150 panoramic radiographs (50 control, 50 missing teeth, 50 supernumerary)	To assess the diagnostic accuracy of an AI-based model (Diagnocat™) in detecting congenitally missing and supernumerary teeth on panoramic radiographs.

Author (Year)	Dental Specialty	Study Design / AI Task	AI Model / Algorithm	Dataset Characteristics (n)	Primary Objective
Akgül & Topal (2025) [20]	Paediatric Dentistry	Model Development & Validation	YOLOv5, YOLOv8 (Detection & Segmentation)	656 panoramic radiographs (patients aged 8-18)	To evaluate the success of YOLOv5 and YOLOv8 models with two labeling methods for detecting dens invaginatus in panoramic radiographs.
Pérez de Frutos et al. (2024) [21]	Cariology / Radiology	Diagnostic Accuracy Study	RetinaNet, YOLOv5, EfficientDet (Object Detection)	13,887 bitewing radiographs (from HUNT4 Study)	To compare AI object detection models against dental clinicians for the task of proximal caries detection on bitewing radiographs.
Pornprasertsuk-Damrongsri et al. (2025) [22]	Cariology / Radiology	Model Development & Validation	YOLOv5 (tooth detection), Attention U-Net (caries segmentation)	500 panoramic radiographs (14,997 teeth; confirmed with bitewings)	To identify and segment multistage caries on panoramic radiographs using a reliable ground truth.
Ryu et al. (2023) [23]	Periodontology	Model Development & Validation	Faster R-CNN (with ResNet-101 backbone)	4083 panoramic radiographs (48,996	To automatically detect periodontal bone loss on

Author (Year)	Dental Specialty	Study Design / AI Task	AI Model / Algorithm	Dataset Characteristics (n)	Primary Objective
				Regions of Interest)	panoramic radiographs and evaluate its potential for practical use.

Table 2 summarizes key findings and performance outcomes.

Table 2: Summary of Key Findings and Outcomes

Author (Year)	Key Performance Metrics	Main Findings / Clinical Relevance	Reported Limitations
Bennasar et al. (2025) [9]	- DL vs. Clinician: Significantly higher sensitivity (0.90 vs. 0.61), NPV (0.80 vs. 0.52), accuracy (0.77 vs. 0.60). - Image-based models (DL) showed superior predictive capability over categorical data-based ML models.	Image-based AI models outperform clinician prognosis and categorical data models in predicting NSRCT outcomes, highlighting the rich predictive information in radiographs.	- Single-institution, single-device dataset limits generalizability. - Small dataset size, risk of overfitting despite LOOCV. - Scarcity of comparable studies for validation.
Boztuna et al. (2024) [10]	- Dice Score: 0.788 (Test), 0.83 (Validation). - Precision: 0.776, Recall: 0.854, F1-score: 0.81 (Test set).	The U ² -Net model can accurately segment periapical lesions on panoramic radiographs, serving as a promising adjunct tool to minimize misdiagnosis and save time in clinical workflow.	- False positives from overlapping anatomical structures; false negatives with smaller lesions. - Single-source dataset (Cyprus), limiting generalizability. - No clinical data (e.g., pulp tests) used for diagnosis.
Çatmabacak & Çetinkaya (2025) [11]	- DenseNet201: AUC: 0.900, MCC: 0.810.	DenseNet201 emerged as the most effective model for	- Single-institution dataset limits external validity. - Dataset size may have

Author (Year)	Key Performance Metrics	Main Findings / Clinical Relevance	Reported Limitations
	- ResNet-18: AUC: 0.890, MCC: 0.709. - MaxVit-T performed poorly (AUC: 0.560).	FEI detection, highlighting its clinical potential. ResNet-18 offers a good balance of accuracy and computational efficiency.	- limited statistical power for pairwise comparisons. - Lack of pre-trained models specific to dental imaging.
Chang et al. (2020) [12]	- Detection (Dice): PBL: 0.93, CEJ: 0.91, Teeth: 0.91. - Staging: PCC: 0.73, ICC: 0.91 with radiologists. - MAD: 0.25 stages.	The hybrid framework combining deep learning (for detection) and conventional CAD (for classification) achieved high accuracy and excellent reliability in automatically staging periodontitis.	- Cannot provide absolute bone loss values due to inherent magnification and distortion in panoramic radiography. - Performance needs evaluation through inter-organizational collaboration.
Davidovitch et al. (2022) [13]	- 97.6% of AI landmark coordinates within 2 mm acceptable error. - Only 1/42 coordinates statistically different and >2mm. - High correlation with human observers ($r=0.99$).	The AI system was highly accurate and comparable to human experts. It can serve as a reliable aid to reduce time and human error in orthodontic diagnosis.	- Small sample size (10 radiographs). - Single-center study. - Lack of external validation.
AlGhaihab et al. (2025) [14]	- Overall AUC difference (AI-assisted vs. unaided): 0.008 (not significant). - Periapical AUC difference: 0.031; Bitewing AUC difference: -0.009. - No statistically	The Denti.AI DL technology did not significantly improve the diagnostic accuracy of experienced general dentists for detecting bone loss. AI's role may be more pronounced for less	- Small sample size of observers (n=10). - Absence of time constraints may not reflect real clinical pressure. - Experienced participants may have had a high baseline accuracy, leaving little room for improvement.

Author (Year)	Key Performance Metrics	Main Findings / Clinical Relevance	Reported Limitations
	significant improvement in diagnostic accuracy.	experienced clinicians or under time constraints.	
Gao & Tang (2025) [15]	- Landmark Detection: MRE: 1.21 mm; 2mm success rate: 92.4%. - Treatment Prediction: Overall Accuracy: 85.6%.	The multi-modal (DeepFuse) framework significantly outperformed single-modality and existing methods. It provides a comprehensive tool for enhanced diagnostic precision and predictive treatment planning.	- High resource requirements for multi-modal data (CBCT cost/radiation). - Performance degrades with partial modality data. - Computational complexity.
Krois et al. (2019) [16]	- CNN Accuracy: 0.81 (± 0.02). - CNN vs. Dentists: No statistically significant superiority ($p=0.067$). - Dentists had higher sensitivity (0.92) but lower specificity (0.63).	The CNN demonstrated at least similar diagnostic performance to experienced dentists. It shows promise as an automated assistance system to reduce diagnostic effort in PBL screening.	- Manually cropped image segments, not full radiographs. - Relatively small dataset for deep learning. - No fully independent hold-out test set.
Kurt-Bayrakdar et al. (2024) [17]	- AUC: Total Bone Loss: 0.951; Horizontal: 0.910; Furcation: 0.868; Vertical: 0.733. - F1-Score: Vertical Bone Loss: 0.673.	The AI system was highly effective, especially for detecting total alveolar and horizontal bone loss. It offers detailed, segmented insights for periodontal diagnosis and treatment planning.	- Lower performance for vertical bone loss detection. - No direct comparison with clinician performance on the same test set. - 2D radiography limitations (superimposition, no 3D assessment).
Lee et al. (2020) [18]	- Mean Landmark Error (LE): 1.53 ± 1.74 mm.	The Bayesian CNN framework provided high accuracy for	- Does not consider spatial relationships between landmarks, leading to

Author (Year)	Key Performance Metrics	Main Findings / Clinical Relevance	Reported Limitations
	- Successful Detection Rate (SDR): 82.11% (2mm), 92.28% (3mm), 95.95% (4mm). - Provided 95% confidence regions for landmarks.	landmark detection and, uniquely, quantified uncertainty via confidence regions. This enhances clinical trust and aids in decision-making, especially for difficult landmarks like Gonion.	- possible aberrant predictions. - Trained on a relatively small dataset (400 images) with inherent expert variability. - Limited generalizability due to single-dataset training.
Makrygiannakis et al. (2025) [19]	- Missing Teeth: Sensitivity: 84.7%, Specificity: 100%, PPV: 100%, NPV: 99.4%. - Supernumerary Teeth: Sensitivity: 43.9%, Specificity: 100%, PPV: 100%, NPV: 98.9%. - Kappa: 0.91 (missing), 0.60 (supernumerary).	The AI model shows promise for detecting missing teeth but is not yet reliable for supernumerary teeth due to low sensitivity. It can act as a useful supplementary tool but cannot replace human assessment.	- Convenience sampling and retrospective design. - Proprietary "black box" model limits insight into algorithm and training data. - Performance likely affected by mixed dentition and overlapping structures.
Akgül & Topal (2025) [20]	- YOLOv8 (Detection): Precision: 0.950, Sensitivity: 1.0, F1-score: 0.974. - YOLOv8 (Segmentation): Precision: 0.940, Sensitivity: 0.994, F1-score: 0.966. - YOLOv5 also showed high performance.	YOLO models, particularly YOLOv8, are highly successful in detecting dens invaginatus on panoramic radiographs. They can be integrated as a decision support mechanism in paediatric dentistry for early diagnosis.	- All types of dens invaginatus were labeled together without differentiation due to imbalanced class distribution. - Difficulties in labeling rotated/crowded teeth or those with fractures/restorations. - Single-center dataset from one panoramic device.
Pérez de Frutos et al. (2024) [21]	- Best Model (YOLOv5): mAP: 0.647, mF1: 0.548, mFNR: 0.149.	Deep learning models, particularly YOLOv5, can perform equal to or	- Task remains challenging due to natural artifacts in bitewing images. - Models were not tailored

Author (Year)	Key Performance Metrics	Main Findings / Clinical Relevance	Reported Limitations
	- Outperformed best human annotators (mAP: 0.299, mF1: 0.495, mFNR: 0.164).	better than dental clinicians in detecting proximal caries on bitewings, showing potential as an effective assistive tool.	- for the task; an ensemble approach could improve robustness. - No inference runtime information provided.
Pornprasertsuk-Damrongsri et al. (2025) [22]	- Recall: 0.96, F1-score: 0.85, Accuracy: 0.93 (posterior teeth). - Strong agreement with radiologists (Kappa ≥ 0.9 for caries stages). - High performance in segmenting enamel, dentine, and pulp caries.	The two-model AI system (YOLOv5 + Attention U-Net) is highly effective at detecting various caries stages on panoramic radiographs, using bitewings for robust ground truth. Its high recall minimizes missed lesions, making it a valuable screening aid.	- Tends to overpredict caries (false positives) in healthy teeth, leading to moderate precision (0.77). - Performance challenges with enamel caries and some misclassification between stages. - Anterior teeth caries were not validated with bitewings.
Ryu et al. (2023) [23]	- AUC: 0.91 (overall), 0.88 (periodontitis). - Precision: 0.86, Recall: 0.84 for periodontitis. - High inter-/intra-examiner reliability (ICC=0.94).	Using a large dataset and regional tooth grouping, the Faster R-CNN model reliably automates the detection of periodontal bone loss on panoramic radiographs. This allows for efficient, large-scale screening and patient awareness.	- Inherent limitations of panoramic radiography (low sharpness, superimpositions) affect diagnostic detail. - Algorithm trained only on panoramic images, lacking data from CBCT or clinical attachment levels for higher precision. - Misclassifications occurred, particularly in anterior regions.

The included studies collectively evaluated a variety of artificial intelligence models, predominantly deep learning CNNs, U-Net architectures, Mask R-CNN, Faster R-CNN, and YOLO-based models. Dataset sizes varied widely, ranging from as few as 10 cephalometric radiographs to over 13,000 bitewing images. Most studies utilized retrospective datasets; only a minority included observer performance evaluations involving human clinicians. Model performance was typically assessed using accuracy, sensitivity, specificity, precision, recall, F1-scores, Dice coefficients, and AUC values. Across specialties, most AI systems demonstrated promising diagnostic performance, often comparable to or surpassing that of human clinicians.

AI in Endodontics

Three studies assessed AI applications in endodontics, focusing primarily on radiographic diagnosis [9–11]. Bennasar et al. [9] demonstrated that deep learning models significantly outperformed both clinician judgment and machine learning models based on categorical data in predicting non-surgical root canal treatment outcomes. Image-based models achieved high sensitivity (0.90) and accuracy (0.77), suggesting that periapical radiographs contain clinically rich predictive features that AI can effectively utilize. Boztuna et al. [10] reported strong performance of a U²-Net model for automated segmentation of periapical lesions on panoramic radiographs, with Dice scores exceeding 0.78. While the model showed promise as an adjunctive tool, false positives due to anatomical overlap and poor generalizability from a single-source dataset were noted. Çatmabacak & Çetinkaya [11] compared five deep learning architectures for identifying fractured instruments in root canals. DenseNet201 achieved the highest performance (AUC 0.900), highlighting its potential usefulness in clinical decision-making. However, performance variability among models underscored the influence of dataset size, quality, and model architecture.

AI in Periodontology

AI applications within periodontology primarily focused on diagnosing periodontal bone loss and staging periodontitis using panoramic or intraoral radiographs. Chang et al. [12] showed that a hybrid deep learning and CAD approach achieved high accuracy in detecting and staging periodontal bone loss, with Dice scores above 0.90 for anatomical structures. Similarly, Krois et al. [16] and Ryu et al. [23] reported strong diagnostic performance of CNN and Faster R-CNN models, with AUC values of 0.81 and 0.91, respectively. Kurt-Bayrakdar et al. [17] developed a U-Net model capable of distinguishing between horizontal, vertical, and furcation defects, achieving an AUC of 0.951 for total bone loss. However, detection of vertical defects remained more challenging, with an AUC of 0.733. AlGhaihab et al. [14] assessed the impact of an AI-based system (Denti.AI) on dentists' diagnostic accuracy. The tool did not significantly improve clinicians' performance, suggesting limited benefit when used by experienced practitioners under non-pressured conditions.

AI in Orthodontics

Four studies evaluated the use of AI in orthodontic diagnostics, particularly cephalometric landmark detection and treatment prediction [13,15,18,19]. Davidovitch et al. [13] and Lee et al. [18] demonstrated high accuracy of CNN and Bayesian CNN models in identifying cephalometric landmarks. More than 95% of coordinates fell within acceptable error thresholds. The Bayesian model additionally offered uncertainty quantification, providing confidence regions around predicted landmarks. Gao & Tang [15] introduced a multi-modal deep learning framework (DeepFuse) utilizing cephalograms, facial photographs, and dental models. The model achieved strong performance in both landmark detection (2mm success rate: 92.4%) and treatment outcome prediction (accuracy: 85.6%). Makrygiannakis et al. [19] evaluated Diagnocat™, reporting high accuracy in detecting congenitally missing teeth (specificity 100%). However, sensitivity for supernumerary teeth was limited (43.9%), indicating the model's selective diagnostic reliability.

AI in Cariology and Dental Radiology

Two major studies explored AI applications in detecting proximal and multistage dental caries [21,22]. Pérez de Frutos et al. [21], using a dataset of over 13,000 bitewings, found that YOLOv5-based AI models outperformed human annotators in detecting proximal caries (mAP: 0.647). This suggests strong potential for AI-assisted large-scale screening. Pornprasertsuk-Damrongsri et al. [22] reported high recall (0.96) and accuracy (0.93) for a combined YOLOv5 and Attention U-Net pipeline in detecting enamel, dentine, and pulp caries. Over-prediction and moderate precision remained limitations, but the system demonstrated robust agreement with radiologist assessments.

AI in Paediatric Dentistry

Akgül & Topal [20] evaluated YOLO models for detecting dens invaginatus in pediatric panoramic radiographs. YOLOv8 achieved outstanding performance (precision 0.950; F1-score 0.974), supporting its potential role in early diagnosis of this anomaly. Challenges included class imbalance and image

artifacts affecting labeling. Across dental specialties, AI systems demonstrated consistently strong diagnostic performance, often matching or exceeding clinician accuracy. Deep learning models—particularly CNN-based architectures, U-Net segmentation frameworks, and YOLO detection systems—showed the strongest results across imaging-based tasks.

Discussion

Our systematic review underscores that AI, particularly deep learning models such as CNNs, U-Net, and YOLO variants, has made remarkable progress in dental diagnostics. Across endodontics, periodontology, orthodontics, and caries detection, these AI systems often matched or even outperformed human clinicians in accuracy, sensitivity, and other performance metrics. The high diagnostic performance reported in many of the studies aligns with broader trends in dental AI research. For instance, in caries detection, AI systems consistently demonstrated strong sensitivity and specificity, reducing false negatives and false positives, which has been corroborated by a recent meta-analysis showing that AI often outperforms traditional diagnostic methods on radiographic images [24]. Similarly, in periodontics, advanced AI models have demonstrated high accuracy in detecting bone loss and staging disease. In one recent study using a YOLOv8 model on panoramic radiographs, AI achieved 94.4% accuracy and 100% sensitivity—outperforming both periodontists and general practitioners [25]. This suggests that AI could improve early diagnosis, reduce inter-examiner variability, and provide more consistent assessments, especially in settings where access to specialists is limited.

While many of the included studies focus on image-based diagnostics, recent reviews emphasize that AI's potential in dentistry extends beyond radiography. For example, narrative reviews highlight applications in treatment planning, outcome prediction, and workflow optimization [26,27]. A particularly promising area is restorative dentistry, where AI is beginning to aid in caries detection, treatment planning, and material selection [28]. The ability of AI to integrate multimodal data—from clinical records to imaging—could help realize more personalized and precise dental care.

In addition to clinical diagnostics, AI could reshape dental education. A recent systematic review examined generative AI in dental and orthodontic education, discussing how it can support curriculum development, simulation-based training, and decision-making exercises [29]. Integrating AI into education may help students build diagnostic confidence and improve their interpretative skills—especially for complex radiographic patterns. Beyond conventional radiographs, AI is also being applied to more sophisticated imaging modalities. For instance, a recent systematic review evaluated AI's diagnostic accuracy in identifying maxillary sinus pathologies on CT/CBCT scans; models like ResNet and DenseNet achieved accuracies of up to ~97% [30]. This highlights the versatility of AI in 3D imaging and its potential in more advanced diagnostic workflows (e.g., implant planning, surgical navigation).

This systematic review has several limitations that should be considered when interpreting the findings. The number of included studies was limited and unevenly distributed across dental specialties, with most research concentrated in orthodontics, periodontology, and endodontics, leaving gaps in areas such as prosthodontics and oral surgery. Most studies used retrospective, single-center datasets with variable sample sizes and inconsistent imaging standards, reducing generalizability. Considerable methodological heterogeneity—ranging from differences in AI architectures to annotation protocols—also limited direct comparison between studies and prevented meta-analysis. Additionally, many models lacked external validation and were evaluated only under controlled research conditions, which may not translate effectively into real-world clinical workflows. The reliance on imaging-only datasets without integration of clinical parameters further constrains the applicability of current AI systems. Publication bias toward studies reporting positive results cannot be ruled out. Looking ahead, future research should focus on developing large, diverse, multi-center datasets and conducting prospective clinical trials that evaluate AI performance in routine practice. Standardizing reporting frameworks, enhancing model interpretability, and integrating multimodal clinical data will be essential to improving the robustness, transparency, and clinical usefulness of AI in dentistry.

Conclusion

This systematic review demonstrates that artificial intelligence is rapidly transforming diagnostic and decision-making processes across multiple dental specialties. Deep learning architectures—particularly convolutional neural networks, U-Net–based segmentation models, Faster R-CNN, Mask R-CNN, and YOLO variants—consistently exhibited high diagnostic accuracy in tasks such as detecting periodontal bone loss, identifying cephalometric landmarks, diagnosing caries, segmenting periapical lesions, and predicting treatment outcomes. In several instances, AI models performed on par with or surpassed experienced clinicians, highlighting their potential to enhance efficiency, reduce diagnostic variability, and support early detection of dental diseases. Despite these promising outcomes, the translation of AI tools into routine dental practice remains limited by several persistent challenges. Many included studies were constrained by small or single-center datasets, limited external validation, and variable annotation quality—all factors that reduce generalizability. Additionally, heterogeneity in model architectures, evaluation metrics, and reporting standards complicates direct comparison across studies. Ethical considerations, including algorithmic transparency, data privacy, and clinical accountability, further underscore the need for cautious implementation. Overall, AI holds significant promise as an adjunctive tool capable of augmenting, rather than replacing, clinical expertise across dental specialties. To fully realize its clinical value, future research should prioritize large, multicenter datasets, rigorous external validation, standardized reporting, and prospective clinical evaluation. Strengthening methodological robustness and addressing existing limitations will be essential for ensuring that AI systems can be safely and effectively integrated into everyday dental practice.

References

1. Tyagi M, Jain S, Ranjan M, et al.: Artificial Intelligence Tools in Dentistry: A Systematic Review on Their Application and Outcomes. *Cureus*. Published Online First: 29 May 2025. 10.7759/cureus.85062
2. Araidy S, Batshon G, Mirochnik R: Artificial Intelligence Applications in Dentistry: A Systematic Review. *Oral*. 2025, 5:90. 10.3390/oral5040090
3. Semerci ZM, Yardımcı S: Empowering Modern Dentistry: The Impact of Artificial Intelligence on Patient Care and Clinical Decision Making. *Diagnostics*. 2024, 14:1260. 10.3390/diagnostics14121260
4. Vázquez-Sebrango G, Anitua E, Macía I, Arganda-Carreras I: The role of artificial intelligence in implant dentistry: a systematic review. *Int J Oral Maxillofac Surg*. 2025, 54:1098–122. 10.1016/j.ijom.2025.04.005
5. Umer F, Adnan S, Lal A: Research and application of artificial intelligence in dentistry from lower-middle income countries – a scoping review. *BMC Oral Health*. 2024, 24:220. 10.1186/s12903-024-03970-y
6. Sitaras S, Tsolakis IA, Gelsini M, Tsolakis AI, Schwendicke F, Wolf TG, Perlea P: Applications of Artificial Intelligence in Dental Medicine: A Critical Review. *Int Dent J*. 2025, 75:474–86. 10.1016/j.identj.2024.11.009
7. Baig I, Azam S, Mushtaq TB: Artificial Intelligence in Dentistry: Literature Review. *J Pharm Res Int*. 2022, 7–14. 10.9734/jpri/2022/v34i53B7228
8. Sachdeva A: Artificial Intelligence in Dentistry: A Literature Review. *Biomed J Sci Tech Res*. 2023, 51:. 10.26717/BJSTR.2023.51.008050
9. Bennasar C, Nadal-Martínez A, Arroyo S, Gonzalez-Cid Y, López-González ÁA, Tárraga PJ: Integrating Machine Learning and Deep Learning for Predicting Non-Surgical Root Canal Treatment Outcomes Using Two-Dimensional Periapical Radiographs. *Diagnostics*. 2025, 15:1009. 10.3390/diagnostics15081009

10. Boztuna M, Firincioglu M, Akkaya N, Orhan K: Segmentation of periapical lesions with automatic deep learning on panoramic radiographs: an artificial intelligence study. *BMC Oral Health*. 2024, 24:1332. 10.1186/s12903-024-05126-4
11. Çatmabacak ED, Çetinkaya İ: Deep learning algorithms for detecting fractured instruments in root canals. *BMC Oral Health*. 2025, 25:293. 10.1186/s12903-025-05652-9
12. Chang H-J, Lee S-J, Yong T-H, et al.: Deep Learning Hybrid Method to Automatically Diagnose Periodontal Bone Loss and Stage Periodontitis. *Sci Rep*. 2020, 10:7531. 10.1038/s41598-020-64509-z
13. Davidovitch M, Sella-Tunis T, Abramovicz L, Reiter S, Matalon S, Shpack N: Verification of Convolutional Neural Network Cephalometric Landmark Identification. *Appl Sci*. 2022, 12:12784. 10.3390/app122412784
14. AlGhaihab A, Moretti AJ, Reside J, et al.: An Assessment of Deep Learning's Impact on General Dentists' Ability to Detect Alveolar Bone Loss in 2D Intraoral Radiographs. *Diagnostics*. 2025, 15:. 10.3390/diagnostics15040467
15. Gao F, Tang Y: Multimodal deep learning for cephalometric landmark detection and treatment prediction. *Sci Rep*. 2025, 15:25205. 10.1038/s41598-025-06229-w
16. Krois J, Ekert T, Meinhold L, et al.: Deep Learning for the Radiographic Detection of Periodontal Bone Loss. *Sci Rep*. 2019, 9:8495. 10.1038/s41598-019-44839-3
17. Kurt-Bayrakdar S, Bayrakdar İŞ, Yavuz MB, et al.: Detection of periodontal bone loss patterns and furcation defects from panoramic radiographs using deep learning algorithm: a retrospective study. *BMC Oral Health*. 2024, 24:155. 10.1186/s12903-024-03896-5
18. Lee J-H, Yu H-J, Kim M, Kim J-W, Choi J: Automated cephalometric landmark detection with confidence regions using Bayesian convolutional neural networks. *BMC Oral Health*. 2020, 20:270. 10.1186/s12903-020-01256-7
19. Makrygiannakis MA, Giannakopoulos K, Kavadella A, Paraskevis D, Kaklamanos EG: Diagnostic accuracy of an artificial intelligence-based software in detecting supernumerary and congenitally missing teeth in panoramic radiographs. *Eur J Orthod*. 2025, 47:cjaf054. 10.1093/ejo/cjaf054
20. Akgül EN, Gucyetmez Topal B: Detection of Dens Invaginatus on Panoramic Radiographs Using Deep Learning Algorithms. *Int J Paediatr Dent*. n/a: 10.1111/ipd.70027
21. Pérez de Frutos J, Holden Helland R, Desai S, Nymoen LC, Langø T, Remman T, Sen A: AI-Dentify: deep learning for proximal caries detection on bitewing x-ray - HUNT4 Oral Health Study. *BMC Oral Health*. 2024, 24:344. 10.1186/s12903-024-04120-0
22. Pornprasertsuk-Damrongsri S, Vachmanus S, Papasratorn D, Kitisubkanchana J, Chaikantha S, Arayasantiparb R, Mongkolwat P: Clinical application of deep learning for enhanced multistage caries detection in panoramic radiographs. *Sci Rep*. 2025, 15:33491. 10.1038/s41598-025-16591-4
23. Ryu J, Lee D-M, Jung Y-H, Kwon O, Park S, Hwang J, Lee J-Y: Automated Detection of Periodontal Bone Loss Using Deep Learning and Panoramic Radiographs: A Convolutional Neural Network Approach. *Appl Sci*. 2023, 13:5261. 10.3390/app13095261
24. Luke AM, Rezallah NNF: Accuracy of artificial intelligence in caries detection: a systematic review and meta-analysis. *Head Face Med*. 2025, 21:24. 10.1186/s13005-025-00496-8

25. Jundaeng J, Chamchong R, Nithikathkul C: Artificial intelligence-powered innovations in periodontal diagnosis: a new era in dental healthcare. *Front Med Technol.* 2025, 6:. 10.3389/fmedt.2024.1469852
26. Surlari Z, Budală DG, Lupu CI, Stelea CG, Butnaru OM, Luchian I: Current Progress and Challenges of Using Artificial Intelligence in Clinical Dentistry—A Narrative Review. *J Clin Med.* 2023, 12:7378. 10.3390/jcm12237378
27. Khan MA, Ansari AS, Anwar M: Role Of Artificial Intelligence in The Dental Practice -A Narrative Review. *Pak J Med Dent.* Published Online First: 24 October 2024. 10.36283/ziun-pjmd13-4/023
28. Najeeb M, Islam S: Artificial intelligence (AI) in restorative dentistry: current trends and future prospects. *BMC Oral Health.* 2025, 25:592. 10.1186/s12903-025-05989-1
29. Baxmann M, Kárpáti K, Baráth Z: The potentials and challenges of integrating generative artificial intelligence (AI) in dental and orthodontic education: a systematic review. *BMC Oral Health.* 2025, 25:905. 10.1186/s12903-025-06070-7
30. Uthman AT, Abouelenen H, Khan S, Bseiso O, Al-Rawi N: Diagnostic accuracy of artificial intelligence in the detection of maxillary sinus pathology using computed tomography: A concise systematic review. *Imaging Sci Dent.* 2025, 55:1. 10.5624/isd.20240139