

How is artificial intelligence implemented in dentistry, and what are its applications, performance, and benefits compared to traditional methods in recent studies?

¿Cómo se implementa la inteligencia artificial en odontología y cuáles son sus aplicaciones, desempeño y beneficios en comparación con los métodos tradicionales en estudios recientes?

Rohan Sharma, Dewi Lestari, Ahmad Rizky Pratama, Arjun Patel¹ 

¹All India Institute of Medical Sciences. New Delhi, India.

²Faculty of Dentistry, Universitas Indonesia. Jakarta, Indonesia.

³Faculty of Dentistry, Universitas Gadjah Mada. Yogyakarta, Indonesia.

⁴King George's Medical University. Lucknow, India.

Cite as

Sharma R, Lestari D, Pratama AR, Patel A. How is artificial intelligence implemented in dentistry, and what are its applications, performance, and benefits compared to traditional methods in recent studies? SAP Artificial Intelligence in Dentistry. 2025; 1:30. <https://doi.org/10.62486/aid202530>

Publication History

Submitted: 07-04-2025

Revised: 02-06-2025

Accepted: 12-08-2025

Published: 13-08-2025

Corresponding Author

Arjun Patel 

ABSTRACT

Introduction: artificial intelligence has been progressively incorporated into dentistry, particularly in diagnostic imaging, treatment planning, predictive modeling and patient education. The expansion of machine learning, deep learning and language-based models has enabled new approaches to clinical decision support and workflow optimization.

Objective: to synthesize the available evidence on the implementation of artificial intelligence in dentistry, focusing on its clinical applications, performance, comparison with traditional methods and reported limitations.

Method: A systematic review was conducted that included studies published between 2020 and 2025. A comprehensive search was performed in PubMed, Scopus, Web of Science, LILACS and ProQuest. The selection process included identification, duplicate removal, screening by title and abstract, and full-text assessment. Studies were included if they reported applications of artificial intelligence in clinical, diagnostic or educational contexts within dentistry. Data extraction considered the study design, application area, type of artificial intelligence, purpose, performance metrics, comparison with conventional methods and main findings. The evidence was synthesized through a structured narrative approach because of methodological heterogeneity.

Results: artificial intelligence is widely applied in radiological diagnosis, orthodontics, implantology, periodontics and educational settings. Deep learning models, particularly convolutional neural networks, are predominant in image-based tasks and demonstrate high diagnostic accuracy. In structured clinical scenarios, artificial intelligence systems achieve performance comparable to or exceeding that of traditional methods. Improvements are observed in diagnostic precision, standardization of procedures and decision support. Performance variability is identified in complex clinical tasks, where contextual interpretation remains necessary. Most studies present limitations related to sample size, lack of external validation and heterogeneity in methodological approaches.

Conclusions: artificial intelligence represents a relevant tool for supporting clinical and educational processes in dentistry, particularly in structured and data-driven tasks. Its implementation shows consistent benefits in terms of accuracy and standardization. Current evidence indicates that these systems should complement, rather than replace, clinical judgment. Further research with a rigorous design and external validation is needed to determine the generalizability and clinical impact of these findings.

Keywords: Artificial Intelligence; Dentistry; Machine Learning; Deep Learning; Diagnostic Imaging; Clinical Decision Support.

RESUMEN

Introducción: la inteligencia artificial ha sido incorporada progresivamente en la odontología, con aplicaciones en diagnóstico por imágenes, planificación de tratamientos, modelado predictivo y educación del paciente. El desarrollo de técnicas de aprendizaje automático, aprendizaje profundo y modelos basados en lenguaje ha permitido nuevas formas de apoyo a la toma de decisiones clínicas y organización de procesos asistenciales.

Objetivo: sintetizar la evidencia disponible sobre la implementación de la inteligencia artificial en odontología, considerando sus aplicaciones clínicas, su desempeño, la comparación con métodos tradicionales y sus principales limitaciones.

Método: se realizó una revisión sistemática de estudios publicados entre 2020 y 2025. La búsqueda se llevó a cabo en PubMed, Scopus, Web of Science, LILACS and ProQuest. El proceso incluyó identificación de registros, eliminación de duplicados, cribado por título y resumen y evaluación a texto completo. Se incluyeron estudios que abordaran aplicaciones de inteligencia artificial en contextos clínicos, diagnósticos o educativos en odontología. La extracción de datos consideró diseño del estudio, área de aplicación, tipo de inteligencia artificial, propósito, métricas de desempeño, comparación con métodos convencionales y hallazgos principales. La síntesis se realizó mediante un enfoque narrativo estructurado debido a la heterogeneidad metodológica.

Resultados: la inteligencia artificial se aplica principalmente en diagnóstico radiológico, ortodoncia, implantología, periodoncia y educación. Los modelos de aprendizaje profundo, en especial las redes neuronales convolucionales, predominan en tareas basadas en imágenes y presentan alta exactitud diagnóstica. En escenarios estructurados, los sistemas de inteligencia artificial alcanzan un desempeño comparable o superior al de métodos tradicionales. Se observan mejoras en precisión diagnóstica, estandarización de procedimientos y apoyo a la toma de decisiones. Se identifica variabilidad en el rendimiento en tareas clínicas complejas. La mayoría de los estudios presenta limitaciones relacionadas con tamaño de muestra, ausencia de validación externa y heterogeneidad metodológica.

Conclusiones: la inteligencia artificial constituye una herramienta de apoyo relevante en odontología, con impacto en tareas estructuradas y basadas en datos. La evidencia disponible indica beneficios en precisión y estandarización. Su implementación debe considerarse complementaria al juicio clínico. Se requieren estudios con diseños más rigurosos y validación externa para establecer su aplicabilidad clínica y su impacto en resultados en salud oral.

Palabras clave: Inteligencia Artificial; Odontología; Aprendizaje Automático; Aprendizaje Profundo; Diagnóstico por Imágenes; Apoyo a la Decisión Clínica.

INTRODUCTION

The application of artificial intelligence (AI) in dentistry has steadily increased in recent years, driven by advances in machine learning, deep learning, and neural network-based models. These technologies have been incorporated into multiple clinical areas, including diagnostic imaging, treatment planning, clinical outcome prediction, and patient education, shaping a new paradigm in contemporary dental practice.^(1,2,3,4)

In the diagnostic field, deep learning models, particularly convolutional neural networks, have demonstrated high performance on tasks involving radiographic images, such as caries detection, tooth identification, and calcification recognition.^(5,6,7,8) In these contexts, sensitivity and specificity levels comparable to or superior to those of clinical professionals have been reported for geometrically well-defined tasks. Concurrently, AI has been integrated into surgical procedures, particularly in implantology, where computer-assisted systems have demonstrated greater precision in implant placement than manual techniques have.^(2,9,10,11)

The implementation of AI also extends to clinical prediction using machine learning algorithms and the use of language models for patient education and communication. In these scenarios, improvements have been observed in understanding clinical information and in treatment adherence, particularly in orthodontics and periodontal health.^(12,13,14) However, the performance of these systems varies depending on the complexity of the task, being more consistent in standardized processes and less predictable in complex clinical decision-making.^(3,4,15,16)

Despite these advances, the available evidence has significant methodological limitations. High heterogeneity among studies, the frequent use of small samples, and a lack of external validation in many developed models have been documented. Furthermore, the geographical concentration of research and the limited availability of open databases restrict the generalizability of the results. In the case of language models, their performance in formal clinical evaluations remains moderate, suggesting a complementary rather than a substitute role for specialized clinical judgment.^(5,17,18)

In this context, it is necessary to systematically synthesize the available evidence on the implementation of AI in dentistry, considering its clinical applications, performance, benefits, and limitations compared with traditional methods. The present study aims to critically analyze the literature to establish the current state of knowledge and define the actual scope of these technologies in dental practice.

METHOD

Study Design

A systematic literature review was conducted to identify, synthesize, and analyze recent studies on the implementation of artificial intelligence in dentistry. The review focused on publications from 2020 to 2025. The methodological process was structured around a multidatabase search strategy, a phased selection process, and standardized data extraction, with an emphasis on clinical applications, diagnostic performance, treatment planning, monitoring, and the educational uses of artificial intelligence in dentistry.

Research question

This review was based on the following question: How is artificial intelligence implemented in dentistry, and what are its main applications, performance, and benefits compared to traditional methods?

Information sources and search strategy

The literature search was conducted in PubMed, Scopus, Web of Science, LILACS, and ProQuest. Studies published between 2020 and 2025 were considered. The strategy combined controlled and free-text terms related to artificial intelligence and dentistry

and adapted them to each database's syntax. Descriptors linked to artificial intelligence, machine learning, deep learning, neural networks, natural language processing, large language models, and robotics were included, as were terms associated with dentistry, oral medicine, oral surgery, prosthodontics, endodontics, periodontics, orthodontics, and implantology.

The initial record identification yielded 139 results in PubMed, 2372 in Scopus, 978 in Web of Science, 279 in LILACS, and 721 in ProQuest, for a total of 4489 retrieved records.

Eligibility criteria

Original studies and reviews with relevant data on artificial intelligence applications in dentistry were included, provided that they addressed clinical, educational, or diagnostic contexts with quantitative results or directly interpretable findings. Studies that analyzed the performance of artificial intelligence models, compared them with conventional methods, and assessed their diagnostic accuracy, clinical prediction, treatment planning, patient monitoring, or educational tools applied to the dental field were eligible.

Editorials, letters to the editor, opinions, conference abstracts, documents lacking sufficient methodological information, and studies in which artificial intelligence was not substantively applied to the dental context were excluded. Studies focused solely on administrative, management, or billing aspects without relevant clinical or educational application were also excluded.

Study Selection

The selection was conducted in consecutive stages. In the first phase, the results from all the databases were consolidated, and duplicate records were removed. In the second phase, screening was performed on the basis of title and abstract, considering thematic relevance, the presence of artificial intelligence applied to dentistry, and the clinical or educational relevance of the study. In the third phase, the full texts of potentially eligible records were reviewed to confirm their final inclusion.

For the screening, criteria related to the application of artificial intelligence in dentistry, clinical or human relevance, study design quality, the presence of quantitative results, practical implementation of the technology, publication type, and the relevance of the clinical approach were considered. Final inclusion required a comprehensive assessment of these criteria, with priority given to studies involving real-world application, reported results, and a direct relationship to dental practice.

Data Extraction

The following variables were extracted from the included studies in a standardized manner: study design, dental specialty or area of application, artificial intelligence technique used, primary purpose of the application, reported performance metrics, comparison with traditional methods, and main findings. When available, information on the type of input data, specific algorithms, clinical context, system validation, study population, reported advantages, and methodological or implementation limitations was also recorded.

Summary of the Evidence

The synthesis of results was conducted narratively and comparatively, organizing the findings by application area. Special attention was given to radiological dentistry, diagnostic imaging, computer-assisted surgery, implantology, periodontics, orthodontics, dental education, and language-model-based tools. Metrics such as sensitivity, specificity, accuracy, area under the curve, positioning errors, concordance measures, and differences compared to conventional procedures were analyzed, depending on the nature of each study.

Methodological considerations

Given that the identified studies varied in design, technology, population, metrics, and clinical context, the synthesis focused primarily on the structured description of the available evidence. This approach acknowledges the expected heterogeneity in reviews on artificial intelligence applied to dentistry. Variability among databases, differences in indexing strategies, and the possibility of selection bias resulting from the exclusion of literature not retrieved by the employed strategy were also considered potential limitations.

If you'd like, I can now adapt this to a more formal journal format, with PRISMA-style subheadings: "Protocol and reporting standard," "Eligibility criteria," "Information sources," "Study selection," "Data extraction," and "Data synthesis."

RESULTS AND DISCUSSION

Characteristics of Included Studies

The 65 included sources span a broad range of AI applications across dental specialties, study designs, and geographic contexts. Sources include systematic reviews and meta-analyses, randomized controlled trials (RCTs), prospective and retrospective observational studies, cross-sectional diagnostic studies, and one study protocol. The table below summarizes key characteristics.

Table 1. General Characteristics of the Included Studies

Study	Study design	Dental specialty/application area	AI technique	Application purpose
Nieri et al. ⁽¹⁹⁾	Systematic review	Oral pathology – oral cancer diagnosis	Deep learning	Diagnosis
Kahya Karaca et al. ⁽²⁰⁾	Observational	Prosthodontics – complete denture impressions	Unspecified AI scan	Diagnosis/treatment planning

Al-Sarem et al. ⁽²¹⁾	Diagnostic study	Oral surgery – implant planning	CNNs (DenseNet169, VGG16, ResNet50, etc.)	Treatment planning
Hunsrisakhun et al. ⁽²²⁾	RCT	Preventive dentistry – childhood caries	Chatbot/NLP	Education
Castro Dos Santos et al. ⁽²³⁾	Post hoc RCT analysis	Periodontics – periodontal therapy response	Machine learning (Random Forest, SVM, etc.)	Prediction/treatment planning
Tanaka et al. ⁽¹⁸⁾	Cross-sectional pilot	Orthodontics – treatment planning	LLM chatbots (ChatGPT, Bard, Bing)	Treatment planning
Adel et al. ⁽²⁴⁾	Prospective clinical trial	Orthodontics – smile simulation	Proprietary AI simulation algorithm	Prediction
Ma et al. ⁽²⁵⁾	RCT	Periodontics – dental education	LLM (GPT-4)	Education
Islam ⁽⁷⁾	RCT	Endodontics/restorative – patient education	LLM chatbot (ChatGPT)	Education
Hendrickx et al. ⁽⁹⁾	Systematic review and meta-analysis	Orthodontics – cephalometric landmark detection	CNNs/deep learning	Treatment planning/diagnostics
Liu, Chaofeng et al. ⁽²⁶⁾	P r o s p e c t i v e observational	Orthodontics – clear aligner therapy evaluation	AI-based 3D fusion software	Treatment planning
Shi et al. ⁽²⁷⁾	RCT	Oral surgery – implant placement	Robotic surgery/navigation	Treatment planning and execution
Li, Yuan et al. ⁽⁶⁾	RCT	Periodontics – oral hygiene monitoring	AI-enabled sensing toothbrush/machine learning	Education/monitoring
Dashti et al. ⁽¹¹⁾	Systematic review and meta-analysis	Endodontics – MB2 canal detection	Deep learning	Diagnosis
Sadeghi et al. ⁽²⁸⁾	Systematic review and meta-analysis	Orthodontics – skeletal maturity (CVM)	CNNs/ML (SVM, logistic regression)	Diagnosis/treatment planning
Hu et al. ⁽²⁹⁾	RCT	Periodontics – patient monitoring	AI-based dental monitoring (unspecified)	Education/monitoring
Stillhart et al. ⁽³⁰⁾	Observational	General dentistry – dental pain screening	Computer vision/automated face coding	Screening/diagnosis
Carvalho et al. ⁽¹⁾	Systematic review and meta-analysis	General dentistry – approximal caries	AI/unspecified	Diagnosis
Peng et al. ⁽³¹⁾	Observational (claims data)	Pediatric dentistry – general anesthesia prediction	Machine learning (lasso logistic regression)	Prediction
Uzar et al. ⁽³²⁾	C r o s s - s e c t i o n a l observational	Periodontics – periodontitis grading	Machine learning (unspecified)	Diagnosis/grading
Tarighatnia et al. ⁽³³⁾	Systematic review and meta-analysis	General – osteoporosis screening via OPG	CNNs/deep learning (AlexNet, ResNet, VGG)	Diagnosis
Schwendicke et al. ⁽¹³⁾	Systematic review and meta-analysis	Orthodontics – cephalometric landmark detection	CNNs/deep learning	Treatment planning
Pul et al. ⁽³⁴⁾	RCT (crossover)	Endodontics – periapical radiolucency detection	CNN (modified ResNet-50)	Diagnosis
Deng et al. ⁽³⁵⁾	C r o s s - s e c t i o n a l diagnostic	Periodontics – nonclinical screening	Machine learning (Random Forest)	Diagnosis
Weingart et al. ⁽³⁶⁾	Diagnostic study	Orthodontics/orthognathic surgery – 3D cephalometric landmarks	Deep learning (deep neural patchworks)	Diagnosis/treatment planning
de Queiroz Tavares Borges Mesquita et al. ⁽¹⁴⁾	Systematic review and meta-analysis	Orthodontics – cephalometric landmark detection	Computer vision/AI	Treatment planning/diagnosis
Hegde et al. ⁽³⁷⁾	Diagnostic study	General dentistry – caries detection on bitewing radiographs	Machine learning	Diagnosis
Bespalez-Neto et al. ⁽³⁸⁾	RCT	Orthodontics – root resorption quantification	AI-based segmentation (Dental Segmentator)	Diagnosis
Rodrigues et al. ⁽³⁹⁾	L o n g i t u d i n a l observational	General dentistry – interproximal caries detection	AI software (unspecified)	Diagnosis
Badr et al. ⁽⁴⁰⁾	Systematic review and meta-analysis	Oral surgery/periodontics/orthodontics – CBCT analysis	CNNs/deep learning (U-Net)	Diagnosis
Mansour et al. ⁽⁴¹⁾	Diagnostic study (retrospective)	Endodontics – MB2 canal detection and segmentation	Deep learning (CNN + U-Net)	Diagnosis/treatment planning

Li, Xin et al. ⁽¹²⁾	Systematic review and meta-analysis	Periodontics – periodontitis classification	CNNs/deep learning	Diagnosis
Liu, Mingxin et al. ⁽¹⁶⁾	Systematic review and meta-analysis	Dental education – licensing examinations	LLMs (GPT-3.5, GPT-4, Bard)	Education
Khubrani et al. ⁽⁴²⁾	Systematic review	Periodontics – periodontal bone loss detection	CNNs/deep learning	Diagnosis
Luke et al. ⁽¹⁰⁾	Systematic review and meta-analysis	General dentistry – caries detection	CNNs/deep learning	Diagnosis
Abbott et al. ⁽⁴³⁾	Systematic review and meta-analysis	General dentistry – caries detection	Machine learning	Diagnosis
Estrella, Navarro-Fraile et al. ⁽⁴⁴⁾	RCT	Orthodontics – root resorption assessment	AI-aided segmentation	Diagnosis/assessment
Khaohoen et al. ⁽⁴⁵⁾	Systematic review and meta-analysis	Oral surgery – computer-aided implant placement	Robotic systems/computer vision	Treatment planning and execution
Vaira et al. ⁽⁴⁶⁾	Observational and valutive	Head and neck/oromaxillofacial surgery	LLM (GPT-4)	Diagnosis/treatment planning/education
Santonocito et al. ⁽⁸⁾	RCT	Orthodontics – patient education	AI chatbot (LLM, Botpress)	Education
Sadr et al. ⁽²⁾	Systematic review and meta-analysis	Dental radiology – tooth identification and numbering	Deep learning	Diagnosis
Bonfanti-Gris et al. ⁽⁴⁷⁾	Systematic review and meta-analysis	Radiology – tooth detection and segmentation in panoramic radiographs	CNNs/deep learning (ResNet, VGG, YOLO)	Diagnosis
Dashti et al. ⁽⁴⁸⁾	Systematic review and meta-analysis	Implantology – implant brand detection	CNNs (ResNet152, Neuro-T)	Diagnosis
Luo et al. ⁽⁴⁾	Systematic review and meta-analysis	Oral surgery – robotic implant placement	Robotic CAIS (R-CAIS)	Treatment planning and execution
Serafin et al. ⁽⁴⁹⁾	Systematic review and meta-analysis	Orthodontics/maxillofacial surgery – 3D cephalometric landmarks	CNNs/deep learning	Diagnosis
Devlin et al. ⁽⁵⁰⁾	RCT	Preventive dentistry – enamel-only proximal caries	AI software (AssistDent, unspecified)	Diagnosis
Digmayer Romero et al. ⁽⁵¹⁾	Secondary analysis of RCT data	Restorative dentistry – restoration failure prediction	Machine learning (Random Forest, XGBoost, CatBoost, etc.)	Prediction
Pozzi et al. ⁽⁵²⁾	Prospective clinical trial	Oral surgery – navigation-guided implant surgery	AI-driven surface anatomy identification	Treatment planning and execution
Dashti et al. ⁽⁵³⁾	Systematic review and meta-analysis	Forensic – gender determination from OPGs	Machine learning/deep learning	Forensic determination
Picoli et al. ⁽⁵⁴⁾	Diagnostic study (within-patient controlled trial)	Oral surgery – IAN injury risk assessment	CNNs (Virtual Patient Creator)	Risk assessment/treatment planning
Arzani et al. ⁽³⁾	Systematic review and meta-analysis	Dental imaging – carotid artery calcification detection	CNNs (Faster R-CNN, U-Net, EfficientNet)	Diagnosis/screening
Ziaei et al. ⁽⁵⁵⁾	Systematic review and meta-analysis	Orthodontics – extraction treatment planning	CNNs (ResNet, VGG), RF, SVM, MLP	Prediction
Nascimento da Silva Mulder et al. ⁽⁵⁶⁾	Systematic review (evidence gap map)	Teledentistry – broad dental applications	Unspecified AI/ICT	Patient care/education/health services
Liu, Chen et al. ⁽⁵⁷⁾	RCT	Oral surgery – robotic extraction of impacted teeth in children	AI-based segmentation + robotic system (Yakebot)	Treatment planning and execution
Rivera Capacho et al. ⁽⁵⁸⁾	Retrospective observational cohort	Orofacial myofunctional therapy – OSA	AI-enhanced telemedicine (Smart Therapy Manager)	Treatment planning/monitoring
Canullo et al. ⁽⁵⁹⁾	RCT	Periodontics – peri-implant soft tissue profiling	Machine learning (decision trees, C4.5)	Prediction
Ghensi et al. ⁽⁶⁰⁾	Observational	Periodontics/oral surgery – peri-implant disease prediction	Machine learning (microbiome profiling)	Prediction/diagnosis
Jin et al. ⁽¹⁷⁾	Systematic review and meta-analysis	Dental education – national licensing examinations	LLMs (ChatGPT-3.5, GPT-4)	Education

Yang et al. ⁽⁵⁾	RCT	Oral surgery – robotic implant placement	Semiactive robotic surgery	Treatment planning and execution
Uribe et al. ⁽¹⁵⁾	Observational	General – publicly available dental imaging datasets	Computer vision/machine learning	Diagnosis/treatment planning
Chau et al. ⁽⁶¹⁾	P r o s p e c t i v e e x p e r i m e n t a l (protocol)	Prosthodontics – single molar prosthesis design	3D GAN/deep learning	Treatment planning
Takács et al. ⁽⁶²⁾	Systematic review and meta-analysis	Oral surgery – computer-assisted implant surgery (in vitro)	Robotic/static/AR-based CAIS	Treatment planning and execution
Jain et al. ⁽⁶³⁾	Systematic review and meta-analysis	Oral surgery – robotic implant placement (in vitro)	Robotic systems	Treatment planning and execution
Pražetina et al. ⁽⁶⁴⁾	RCT (protocol)	Oral surgery (maxillofacial free flap) – hemodynamic optimization	Machine learning (HPI algorithm, HemoSphere)	Treatment planning
Wong et al. ⁽⁶⁵⁾	M u l t i c e n t e r observational	Orthodontics – smile dynamics in unilateral cleft lip	Automated smile feature detection software	Analysis

The 65 sources represent substantial heterogeneity in design and focus. Systematic reviews and meta-analyses constitute the most common design types, covering cephalometric landmark detection,^(9,13,14,49) caries detection,^(1,10,43) periodontal bone loss,⁽⁴²⁾ implant placement accuracy,^(4,45,62,63) and LLM performance in examinations,^(16,17) among other topics. The primary studies included 14 RCTs and multiple prospective or cross-sectional diagnostic investigations.

Full text was unavailable for 29 of the 65 sources, which limits the depth of detail that can be reported for those studies. The body of evidence skews heavily toward image-based AI applications in radiology and surgical guidance, with a smaller but growing cluster addressing patient-facing tools such as chatbots and educational applications. The vast majority of primary studies have been conducted in academic or university-affiliated clinical settings,^(6,7,21,25, 41) with geographic concentration in East Asia (particularly China, South Korea, and Japan)^(4,12) and Europe.^(8,59)

Effects

AI Applications in Dental Imaging and Diagnosis

Caries detection

Multiple systematic reviews with meta-analyses address the use of AI for detecting dental caries, and they collectively cover dozens of primary studies. Carvalho et al.⁽¹⁾ pooled data from 21 studies and reported a sensitivity of 0,94 (95 % CI: 0,78–0,99) and a specificity of 0,91 (95 % CI: 0,84–0,95) for AI detection of approximal caries on bitewing radiographs, with the NPV ranging from 0,79 to 1,00, indicating high reliability for ruling out caries on healthy surfaces. In a review of 21 studies, Luke et al.⁽¹⁰⁾ reported accuracies ranging from 73,3 % to 99 %, sensitivities ranging from 71 % to 98,85 %, and specificities ranging from 0,82 % to 98,19 % across the included primary studies. Abbott et al.⁽⁴³⁾, meta-analyzing 7 studies from a pool of 45 eligible sources, reported a mean sensitivity of 76 % (95 % CI: 65 %–85 %) and a specificity of 91 % (95 % CI: 86 %–95 %), with an AUC of 92 % (95 % CI: 89 %–94 %). Across all three reviews, high heterogeneity was a consistent finding,^(10,43) reflecting variations in imaging modality, AI architecture, and caries classification thresholds.

At the primary study level, Devlin et al.⁽⁵⁰⁾ demonstrated that AI-assisted caries detection (AssistDent software) increased dentists’ sensitivity for detecting enamel-only proximal caries from 44,3 % (control) to 75,8 % (AI-assisted), a 71 % relative improvement that was statistically significant ($p < 0,01$). However, this came at the cost of specificity, which decreased by 11 %, ⁽⁵⁰⁾ with the false positive rate rising from 3,7 % to 14,6 %.

Table 2. Applications of artificial intelligence in dentistry						
Study	Design	Comparison	Sensitivity (AI)	Specificity (AI)	AUC/Accuracy	Key finding
Carvalho et al. ⁽¹⁾	SR & MA	—	0,94	0,91	NPV 0,79–1,00	High reliability for ruling out caries
Luke et al. ⁽¹⁰⁾	SR & MA	Human experts	71 %–98,85 %	0,82 %–98,19 %	OR ~2,72	AI comparable to or better than experts
Abbott et al. ⁽⁴³⁾	SR & MA	B i t e w i n g radiography	76 %	91 %	AUC 92 %	Superior sensitivity vs. bitewing radiography
Devlin et al. ⁽⁵⁰⁾	RCT	Unaided dentists	75,8 % vs. 44,3 %	Decreased by 11 %	—	71 % sensitivity increase; false positive rate rises
Rodrigues et al. ⁽³⁹⁾	Longitudinal observational	Prior software version	80 % vs. 87 %	93 %	AUC 0,794	Shift toward higher specificity over time
Hegde et al. ⁽³⁷⁾	D i a g n o s t i c study	Control + checklist groups	79 % (ML group)	90,9 % (checklist group)	DOR 20,3	ML improved diagnostic accuracy and interrater agreement

Rodrigues et al.⁽³⁹⁾ tracked one commercial AI platform longitudinally between 2021 and 2025 and reported that the updated software version showed lower sensitivity (80 % vs. 87 % in the prior version) but improved specificity (93 %), PPV (56 %), and AUC (0,794); a McNemar test confirmed that the difference between versions was statistically significant ($p < 0,001$). Hegde et al.⁽³⁷⁾ evaluated the effect of ML-assisted diagnosis among dental students and reported that the ML group achieved a sensitivity of 79 %, a diagnostic odds ratio of 20,3, and an interrater agreement of $\kappa = 0,702$, outperforming both an unaided control group and a checklist group on most accuracy metrics.

Cephalometric Landmark Detection

AI for automated cephalometric landmark detection is among the most extensively studied applications, with four systematic reviews and multiple primary studies available. Schwendicke et al.⁽¹³⁾ meta-analyzed 19 studies and reported that the mean landmark prediction error centered around the 2 mm clinical threshold (mean: $-0,581$ mm; 95 % CI: $-1,264$ to $0,102$ mm), with 79,9 % (95 % CI: 77,0 %–82,4 %) of landmarks detected within 2 mm for 2D imagery. Similarly, Hendrickx et al.⁽⁹⁾ reported that the AI error for 2D images was less than the 2 mm clinical threshold (mean 1,39 mm; 95 % CI: 0,85–1,92 mm), with landmark detection taking less than 1 minute compared with approximately 20 minutes for manual identification. For 3D-CBCT images, the same review reported mean errors ranging from 1,0 to 5,8 mm, with heterogeneity precluding meta-analysis. De Queiroz Tavares Borges Mesquita et al.⁽¹⁴⁾ pooled 40 studies and reported AI agreement rates of 79 % (95 % CI: 76 %–82 %) at the 2 mm threshold and 90 % (95 % CI: 87 %–92 %) at 3 mm, with a mean divergence of 2,05 mm (95 % CI: 1,41–2,69). Serafin et al.⁽⁴⁹⁾ focused exclusively on 3D imagery (15 studies) and reported a pooled mean error of 2,44 mm with very high heterogeneity ($I^2 = 98,13$ %) but also reported a significant reduction in error over time as a function of publication year ($p = 0,012$).

Weingart et al.⁽³⁶⁾ applied a deep neural patchwork (DNP) algorithm to full skull CT scans to detect 60 landmarks simultaneously, reporting a mean error of 1,94 mm (SD 1,45 mm) compared with 1,32 mm (SD 1,08 mm) for manual annotation. All four systematic reviews note a high risk of bias in data selection and reference standards in the underlying primary studies.^(9,13,14,49)

Table 3. Artificial Intelligence Techniques Employed

Study	Design	Image type	Mean error (AI)	Clinical threshold	Processing time	Key finding
Schwendicke et al. ⁽¹³⁾	SR & MA	2D (mainly)	$-0,581$ mm (95 % CI: $-1,264$ to $0,102$)	2 mm; 79,9 % within threshold	Not reported	High risk of bias
Hendrickx et al. ⁽⁹⁾	SR & MA	2D and 3D-CBCT	1,39 mm (2D); 1,0–5,8 mm (3D)	2 mm; below threshold for 2D	<1 min	High risk of bias
de Queiroz Tavares Borges Mesquita et al. ⁽¹⁴⁾	SR & MA	2D	2,05 mm (95 % CI: 1,41–2,69)	79 % at 2 mm; 90 % at 3 mm	—	Very low certainty of evidence
Serafin et al. ⁽⁴⁹⁾	SR & MA	3D	2,44 mm ($I^2 = 98,13$ %)	Improved 3,71→1,34 mm over 2 yrs	~1 min	Significant year-of-publication effect
Weingart et al. ⁽³⁶⁾	Diagnostic	3D (CT)	1,94 mm (SD 1,45)	Manual: 1,32 mm (SD 1,08)	—	60 landmarks simultaneously

Periodontal Bone Loss and Periodontitis Classification

Several systematic reviews with meta-analyses address AI for periodontitis. Li, Xin et al.⁽¹²⁾ pooled 13 studies and reported a pooled sensitivity of 0,88 (95 % CI: 0,82–0,92), specificity of 0,82 (95 % CI: 0,72–0,89), and SROC area of 0,92 (95 % CI: 0,89–0,94) for the DL-based classification of periodontitis. Using the APPRAISE-AI quality tool across 30 articles, Khubrani et al.⁽⁴²⁾ reported a pooled sensitivity of 87 % (95 % CI: 80 %–93 %), a specificity of 76 % (95 % CI: 69 %–81 %), and an accuracy of 84 % (95 % CI: 75 %–91 %) for periodontal bone loss detection from panoramic and periapical radiographs. Notably, no included studies achieved the highest quality tier on APPRAISE-AI scoring, underscoring transparency and reporting concerns.

Table 4. Performance of Artificial Intelligence Systems

Study	Design	Condition	Sensitivity	Specificity	AUC/SROC	Notable
Li, Xin et al. ⁽¹²⁾	SR & MA	Periodontitis classification	0,88	0,82	0,92	DOR 33
Khubrani et al. ⁽⁴²⁾	SR	Periodontal bone loss	87 %	76 %	Accuracy 84 %	No high-quality studies
Uzar et al. ⁽³²⁾	Cross-sectional	Periodontitis grading	—	—	Accuracy 0,967/0,858	IL-1 β key predictor
Deng et al. ⁽³⁵⁾	Cross-sectional diagnostic	Periodontal health status	—	—	AUC >0,94	Nonclinical inputs only
Castro Dos Santos et al. ⁽²³⁾	Post hoc RCT	Therapy response (DM + periodontitis)	64 %	87 %	AUC 0,83	PD and CAL most predictive
Ghensi et al. ⁽⁶⁰⁾	Observational	Peri-implant disease	—	—	AUC 0,78–0,96	Metagenomics-based ML

At the primary study level, Uzar et al.⁽³²⁾ demonstrated that AI classifiers using 23 of 46 biomarkers and clinical attributes achieved a classification accuracy of 0,967 for periodontal status and 0,858 for periodontitis grading, with IL-1 β , HDL, and the neutrophil/lymphocyte ratio emerging as particularly important predictors. Deng et al.⁽³⁵⁾, using a random forest classifier on nonclinical screening inputs (self-administered questionnaires, aMMP-8 point-of-care test, and gingival bleeding), achieved an AUC > 0,94 for discrimination across six periodontal health categories, with misclassification of periodontitis as health or gingivitis below 1–2 %.

Castro Dos Santos et al.⁽²³⁾ applied ML post hoc to an RCT of 75 patients with periodontitis and type 2 diabetes and reported that a random forest model achieved an area under the curve (AUC) of 0,83, an accuracy of 80 %, a sensitivity of 64 %, and a specificity of 87 % for predicting which patients would achieve the HbA1c-defined clinical endpoint at 6 months, with probing depth and clinical attachment level being the most predictive variables. Using shotgun metagenomics with machine learning, Ghensi et al.⁽⁶⁰⁾ distinguished clinical groups of peri-implant disease with AUCs ranging from 0,78–0,96.

Endodontic Applications

Dashti et al.⁽¹¹⁾ synthesized 4 studies on AI for MB2 canal detection in maxillary first molars on CBCT, reporting a pooled sensitivity of 0,82 (95 % CI: 0,75–0,89) and a specificity of 0,74 (95 % CI: 0,58–0,90). Deep learning outperformed traditional ML in this context, with a sensitivity of 0,87 vs. 0,80, specificity of 0,90 vs. 0,68, and accuracy of 0,84 vs. 0,75. Mansour et al.⁽⁴¹⁾ developed and tested a two-step CNN + U-Net model for the same task using 37 CBCT scans, achieving an F1 score of 0,93, a recall of 1,0 (no false negatives in the test set), a precision of 0,87, and a segmentation Dice coefficient of 0,79 for testing.

Pul et al.⁽³⁴⁾, in a crossover RCT involving 30 dentists and 50 panoramic radiographs, reported that AI assistance improved the overall diagnostic accuracy for periapical radiolucencies from 91,6 % (unaided) to 93,3 % (AI-aided; $p < 0,001$), primarily by reducing false positive rates from 4,3 % to 2,0 %. The sensitivity remained stable (46,0 % vs. 45,8 %), and the PPV improved from 48,7 % to 68,6 %. Junior dentists achieved the greatest gains in performance and confidence, whereas senior dentists achieved minimal benefits.

Additional Imaging Diagnostics

Badr et al.⁽⁴⁰⁾ reviewed AI applications for CBCT analysis across oral surgery, periodontics, and orthodontics (13 studies) and reported a pooled DICE/DSC of 0,85 (95 % CI: 0,73–0,92), precision of 0,88 (95 % CI: 0,77–0,94), recall of 0,93 (95 % CI: 0,84–0,97), and accuracy of 0,83 (95 % CI: 0,68–0,91). Bonfanti-Gris et al.⁽⁴⁷⁾ meta-analyzed deep learning for tooth detection and segmentation in panoramic radiographs (20 studies), with a pooled sensitivity of 0,92 (95 % CI: 0,84–0,96), a specificity of 0,94 (95 % CI: 0,89–0,97), and a diagnostic odds ratio of 186, with a moderate evidence recommendation for clinical use. Sadr et al.⁽²⁾ reviewed deep learning for tooth identification and numbering (29 studies), reporting an accuracy of 81,8 %–99 %, an average DOR of 1612, an average sensitivity of 89 %, and a specificity of 99 %. Dashti et al.⁽⁴⁸⁾ reported that CNNs detected dental implant brands from radiographs with a pooled accuracy of 95,63 %, sensitivity of 94,55 %, and specificity of 97,91 %, with CNN Multitask ResNet152 reaching 99,08 % accuracy.

Arzani et al.⁽³⁾ reported that AI (CNNs applied to panoramic and CBCT images) detected carotid artery calcifications with a pooled sensitivity of 0,92 (95 % CI: 0,81–0,97) and a specificity of 0,96 (95 % CI: 0,92–0,97), highlighting a potential systemic health screening application of existing dental imaging infrastructure. Tarighatnia et al.⁽³³⁾ evaluated deep learning for osteoporosis screening from panoramic radiographs and reported a pooled sensitivity of 80 % (95 % CI: 0,74–0,86), a specificity of 92 % (95 % CI: 0,88–0,95), and an AUC of 93 % (95 % CI: 0,91–0,95), with AlexNet achieving 0,89 sensitivity and 0,99 specificity. Dashti et al.⁽⁵³⁾ reported that AI models determined sex from orthopantomograms with a pooled accuracy of 88,66 %, with some models reaching 99,20 %. In a within-patient controlled trial ($n = 25$), Picoli et al.⁽⁵⁴⁾ reported that 3D AI-driven models were used to assess inferior alveolar nerve injury risk with a sensitivity of 0,87, comparable to that of CBCT (0,89) and superior to that of panoramic radiography (0,73), with no statistically significant difference among modalities ($p > 0,05$).

Sadeghi et al.⁽²⁸⁾ reported that the AI for cervical vertebral maturation (CVM) assessment achieved the highest accuracy for CS1 (sensitivity 0,87, specificity 0,97, DOR 213) but the lowest for CS3 (sensitivity 0,64, specificity 0,96, DOR 32), with kappa values between 0,80 and 0,96 reported in the included studies.

Table 5. Comparison between Artificial Intelligence and Traditional Methods

Study	Design	Imaging application	Pooled sensitivity	Pooled specificity	AUC/Accuracy
Badr et al. ⁽⁴⁰⁾	SR & MA	CBCT – detection and segmentation	0,93 (recall)	—	DSC 0,85; accuracy 0,83
Bonfanti-Gris et al. ⁽⁴⁷⁾	SR & MA	OPG – tooth detection/segmentation	0,92	0,94	DOR 186
Sadr et al. ⁽²⁾	SR & MA	Radiographs – tooth numbering	89 %	99 %	AUC 96 %; DOR 1612
Dashti et al. ⁽⁴⁸⁾	SR & MA	Radiographs – implant brand detection	94,55 %	97,91 %	Accuracy 95,63 %
Arzani et al. ⁽³⁾	SR & MA	OPG/CBCT – carotid calcification	0,92	0,96	DOR 241,45

Tarighatnia et al. ⁽³³⁾	SR & MA	OPG – osteoporosis screening	80 %	92 %	AUC 93 %
Dashti et al. ⁽⁵³⁾	SR & MA	OPG – gender determination	Variable	High	Pooled accuracy 88,66 %
Picoli et al. ⁽⁵⁴⁾	Diagnostic (within-patient)	CBCT/AI 3D – IAN injury risk	0,87 (CBCT)	(AI)/0,89 (CBCT)	AUC 0,63 (AI)
Sadeghi et al. ⁽²⁸⁾	SR & MA	Radiographs assessment – CVM	CS1: 0,87; CS3: 0,64	CS1: 0,97; CS3: 0,96	DOR CS1: 213; CS3: 32
Dashti et al. ⁽¹¹⁾	SR & MA	CBCT – MB2 canal detection	0,82 (DL: 0,87)	0,74 (DL: 0,90)	Accuracy 0,84 (DL)
Mansour et al. ⁽⁴¹⁾	Diagnostic	CBCT – MB2 canal detection/segmentation	1,0 (recall)	—	F1 0,93; DCE 0,79
Pul et al. ⁽³⁴⁾	RCT (crossover)	OPG – periapical radiolucency	45,8 % (AI-aided)	98,0 %	Accuracy 93,3 %

AI for Surgical Guidance and Implant Placement

A substantial cluster of studies examines AI-assisted or robotic/navigated dental implant placement. Across study designs ranging from in vitro meta-analyses to clinical RCTs, the consistent finding is that compared with freehand techniques, guided approaches reduce placement error, with robotic systems generally producing the smallest deviations.

Khaohoen et al.⁽⁴⁵⁾ synthesized data from 67 clinical trials (5673 implants) and reported an overall mean coronal deviation of 1,11 mm (95 % CI: 1,02–1,19), an apical deviation of 1,40 mm (95 % CI: 1,31–1,49), and an angular deviation of 3,51° (95 % CI: 3,27–3,75). Among computer-assisted modalities, the robotic system resulted in the lowest deviation (0,81 mm coronal, 0,77 mm apical, and 1,71° angular). Luo et al.⁽⁴⁾, focusing exclusively on robotic CAIS (16 studies, 908 implants), reported even lower deviations: a global platform of 0,69 mm (95 % CI: 0,61–0,77), a global apex of 0,72 mm (95 % CI: 0,64–0,79), and an angular angle of 1,62° (95 % CI: 1,34°–1,89°). Jain et al.⁽⁶³⁾, in an in vitro meta-analysis (5 studies), reported that compared with dynamic CAIS, robotic systems produced significantly lower angular deviations (mean difference –1,22°; 95 % CI: –1,06 to –1,39; $p < 0,00001$), although the differences in the coronal and apical angles were not statistically significant. Takács et al.⁽⁶²⁾ similarly reported that robotic and fully guided static CAIS had the most accurate and clinically acceptable in vitro outcomes and that freehand placement produced the greatest number of errors.

At the clinical level, Yang et al.⁽⁵⁾ conducted a multicenter RCT (140 patients, 140 implants) that compared semiactive robotic surgery to free-hand surgery and reported that the RAIS significantly outperformed the FHIS in terms of platform deviation ($0,76 \pm 0,36$ mm vs. $1,48 \pm 0,93$ mm), apex deviation ($0,85 \pm 0,48$ mm vs. $2,14 \pm 1,25$ mm), and angular deviation ($2,05 \pm 1,33^\circ$ vs. $7,36 \pm 4,67^\circ$), all $p < 0,001$. In an RCT comparing robotic surgery (RS), dynamic navigation (DN), and static guides (SG) in 45 patients, Shi et al.⁽²⁷⁾ reported global platform deviations of $1,1 \pm 0,4$ mm (RS), $1,3 \pm 0,6$ mm (DN), and $1,1 \pm 0,6$ mm (SG), respectively, with RS producing near-zero systematic error. The duration of surgery was significantly shorter with static guides ($p < 0,001$), and patients in the SG group reported better oral health-related quality of life at 3 days. Pozzi et al.⁽⁵²⁾, in a prospective trial of navigation-guided complete-arch implant placement (25 patients, 161 implants), reported a mean angular deviation of 2,19° (SD 1,26°) and linear deviations of approximately 1,17–1,30 mm; implant site characteristics (healed vs. postextractive) were the only variable with a statistically significant influence on accuracy.

Table 6. Reported Benefits of Artificial Intelligence in Dentistry

Study	Design	Modality	Coronal deviation	Apical deviation	Angular deviation	Key finding
Khaohoen et al. ⁽⁴⁵⁾	SR & MA (clinical)	Static/dynamic/robotic	1,11 mm (overall)	1,40 mm	3,51°	Robotic lowest: 0,81 mm
Luo et al. ⁽⁴⁾	SR & MA	Robotic (R-CAIS)	0,69 mm	0,72 mm	1,62°	16 studies, 908 implants
Jain et al. ⁽⁶³⁾	SR & MA (in vitro)	Robotic vs. DCAIS/SCAIS	–0,15 mm (ns)	–0,19 mm (ns)	–1,22° (p<0,00001)	Significant angular advantage for robotics
Takács et al. ⁽⁶²⁾	SR & MA (in vitro)	All modalities	Robotic/static best	Robotic/static best	Freehand worst	Fully guided static most predictable
Yang et al. ⁽⁵⁾	RCT	RAIS vs. FHIS	0,76 vs. 1,48 mm (p<0,001)	0,85 vs. 2,14 mm (p<0,001)	2,05° vs. 7,36° (p<0,001)	Multicenter, 140 implants
Shi et al. ⁽²⁷⁾	RCT	RS vs. DN vs. SG	1,1/1,3/1,1 mm	1,5/1,9/2,0 mm	4,7°/5,5°/6,2°	RS near-zero systematic error
Pozzi et al. ⁽⁵²⁾	Prospective trial	Navigation (DN)	1,17 mm	1,30 mm	2,19°	Site type only significant variable
Liu, Chen et al. ⁽⁵⁷⁾	RCT	Autonomous robot vs. conventional	—	—	—	35 % less time; 42 % less bone resection
Al-Sarem et al. ⁽²¹⁾	Diagnostic study	CBCT – implant site detection	—	—	Precision 0,98 (DenseNet169)	Segmentation accuracy 93,3 %

Liu, Chen *et al.*⁽⁵⁷⁾ applied an autonomous robotic platform to the extraction of completely impacted supernumerary teeth in 10 children and reported that the robotic arm reduced the total operative time by 35 % (30,7 vs. 39,3 min, $p < 0,05$), bone resection time by 65 % (8,6 vs. 24,2 min, $p < 0,001$), and bone resection volume by 42 % (56,8 % vs. 98,4 %, $p < 0,001$), while nerve injuries were eliminated (0 vs. 1 case in the conventional group).

Al-Sarem *et al.*⁽²¹⁾ addressed the implant planning step and used CNN-based models on 500 CBCT images to classify missing tooth regions for implant site identification. DenseNet169 achieved the best performance, with a precision of 0,98, a segmentation accuracy of 93,3 %, and a classification accuracy of 89 %.

AI for Patient Monitoring and Periodontal Therapy

Two RCTs examined AI as an adjunct to periodontal care delivery. Li, Yuan *et al.*⁽⁶⁾ evaluated an AI-enabled multimodal-sensing toothbrush (AI-MST) combined with targeted micromessaging from clinicians ($n = 100$). At 6 months, the proportion of inflamed periodontal pockets (≥ 4 mm with bleeding on probing) decreased from 81,4 % to 44,4 % in the intervention group vs. 80,7 % to 52,3 % in the control group, yielding an intergroup difference of 7,9 % (95 % CI: 1,6–14,6; $p < 0,05$). The intervention group also achieved significantly better oral hygiene levels ($p < 0,001$), with no adverse events.

Hu *et al.*⁽²⁹⁾ randomized patients with periodontitis into three groups: AI monitoring alone ($n = 32$), AI-assisted health counseling (AIHC; $n = 33$), and control ($n = 33$). At 6 months, the AI group showed greater improvements in self-care behavior ($a\beta = 0,32$) and plaque control ($a\beta = -13,28$) than the control group did. Probing pocket depth was more reduced in both AI groups at 3 months (AI: $a\beta = -0,43$; AIHC: $a\beta = -0,55$) than in controls. The protection motivation construct improved similarly across all three groups.

Canullo *et al.*⁽⁵⁹⁾, in an RCT of 30 participants examining argon plasma-treated vs. nontreated healing abutments, used ML (C4,5 decision trees) for soft tissue profiling and achieved an accuracy of 0,868, a precision of 0,920, a recall of 0,850, and an AUC of 0,938. Ghensi *et al.*⁽⁶⁰⁾ applied ML to shotgun metagenomic data to predict peri-implant disease status, achieving AUCs of 0,78–0,96 for group differentiation and 0,79–0,87 for correlation with peri-implant clinical parameters.

AI for Treatment Planning and Orthodontics

Liu, Chaofeng *et al.*⁽²⁶⁾ applied a 3D fusion model (integrating CBCT and intraoral scan data) to evaluate clear aligner therapy outcomes in 136 patients and reported no significant differences in total score according to sex (mean difference 0,01; $p = 0,85$), age ($p = 0,60$), or treatment method ($p = 0,32$). Compared with adults, adolescents showed smaller discrepancies in arch width measures but greater differences in anteroposterior and curve-of-Spee parameters ($p < 0,001$). Bspalez-Neto *et al.*⁽³⁸⁾, in an RCT comparing root resorption in clear aligner vs. fixed appliance patients ($n = 32$), used the Dental Segmentator AI tool for CBCT-based surface analysis. The overall median EARR was $-0,72$ mm, with no significant difference between appliance types. The AI tool yielded an ICC of 0,808 for measurement reproducibility, with a Bland–Altman bias of $-0,048$ mm (95 % limits of agreement: $-0,597$ to $0,501$). Compared manual vs. AI-aided segmentation for root resorption assessment in an RCT, Estrella, Navarro-Fraile *et al.*⁽⁴⁴⁾ reported that while AI detected similar root length loss (mean: $0,20 \pm 0,23$ mm control vs. $0,42 \pm 0,43$ mm test; $p = 0,045$), manual segmentation detected greater volume loss at most levels, whereas AI detected more volume loss within 1 mm of the apex ($p = 0,011$).

Ziaei *et al.*⁽⁵⁵⁾, in a meta-analysis of 7 studies (6261 patients), reported a pooled AI sensitivity of 70 % (95 % CI: 61–78) and a specificity of 90 % (95 % CI: 87–92) for predicting extraction needs in orthodontics, with CNN-based models (ResNet, VGG) achieving the highest performance with no heterogeneity. Disease incidence significantly influenced sensitivity ($p = 0,050$).

In a prospective clinical trial ($n = 24$), Adel *et al.*⁽²⁴⁾ evaluated the Invisalign SmileView AI simulation tool and reported that five quantitative parameters (philtrum height, commissure height, smile width, buccal corridor, and smile index) were reliable predictors of actual posttreatment outcomes, whereas the maxillary intercanine width significantly differed between the simulation and actual outcomes ($p = 0,05$).

Tanaka *et al.*⁽¹⁸⁾ compared AI-generated treatment plans (ChatGPT-3,5, Microsoft Bing Copilot, and Google Bard Gemini) to a human orthodontist's plan and reported that orthodontist evaluators significantly preferred human expert plans over all AI platforms ($p < 0,001$), with no statistically significant differences among the three AI platforms.

Large Language Models in Dental Education and Clinical Assistance

A distinct cluster of studies examines LLM chatbot performance across educational and clinical assistance contexts. In a meta-analysis of 23 studies covering four health professions, Jin *et al.*⁽¹⁷⁾ reported that ChatGPT models achieved an overall accuracy of 70,1 % (95 % CI: 65 %–74,8 %) across national licensing examinations. Compared with medicine and pharmacy, GPT-4 substantially outperformed GPT-3,5, with accuracies ranging from 64,4 % to 100 %, and its performance in dentistry specifically was 63,2 % (95 % CI: 54,6 %–71,1 %). Liu, Mingxin *et al.*⁽¹⁶⁾, in a separate meta-analysis (11 studies, 8 countries), reported integrated accuracy rates of 54 % (GPT-3,5), 72 % (GPT-4), and 56 % (Bard), respectively; GPT-3,5 performed significantly better in English-speaking contexts, whereas GPT-4 performance was consistent across regions.

Vaira *et al.*⁽⁴⁶⁾, evaluating ChatGPT-4 for head and neck surgical questions (144 questions, 15 clinical scenarios, 18 surgeons from 14 Italian centers), reported an overall median accuracy of 6/6 and a completeness of 3/3 for open-ended questions, with correct responses in 84,7 % of closed-ended questions and correct diagnoses in 81,7 % of clinical scenarios. However, bibliographic references were nonexistent in 46,4 % of the answers, and the accuracy ($p = 0,001$) and completeness ($p = 0,011$) of the AI performance were significantly inferior to those of a second-year resident.

In an RCT of 126 dental students, Ma *et al.*⁽²⁵⁾ compared GPT-4-generated periodontology examinations to human-designed examinations. The AI examination achieved higher content coverage (81,3 % vs. 72,4 %; $p = 0,027$) and generated higher student total scores ($79,34 \pm 6,93$ vs. $73,17 \pm 9,57$; $p = 0,027$) but had a lower discrimination index (0,35 vs. 0,49; $p = 0,004$) and received lower

student ratings for appropriateness, knowledge coverage, and learning inspiration (all $p \leq 0,001$). Islam⁽⁷⁾ reported that ChatGPT-generated patient educational materials for endodontic and restorative procedures outperformed traditional verbal explanations and pamphlets across all measured dimensions, including clarity (4,42 vs. 3,25), usefulness (4,63 vs. 3,50), comprehensiveness (4,50 vs. 3,29), and anxiety reduction (mean score 2,63 vs. 3,38; $p < 0,001$), with interrater reliability among expert evaluators of $\kappa = 0,75$ ($p < 0,001$). In an RCT of 100 orthodontic patients in which an AI chatbot (Botpress) was compared to information leaflets, Santonocito et al.⁽⁸⁾ reported a statistically lower increase in the modified gingival index (MGI) at 5 weeks ($p < 0,001$), with better outcomes in aligner-wearing patients than in fixed-appliance patients. In an RCT of 320 caregiver–child pairs, Hunsrisakhun et al.⁽²²⁾ reported that a standalone chatbot (21-Day FunDee Plus) produced outcomes comparable to those of a chatbot combined with in-person toothbrushing training across dmft, plaque scores, and most behavioral variables, demonstrating standalone chatbot viability for oral health promotion in young children.

Table 7. Limitations and challenges of artificial intelligence implementation in dentistry						
Study	Design	LLM tested	Context	Accuracy/Key metric	Comparator	
Jin et al. ⁽¹⁷⁾	SR & MA	GPT-3,5 and GPT-4	Health licensing exams	70,1 % overall; dentistry 63,2 %	—	
Liu, Mingxin et al. ⁽¹⁶⁾	SR & MA	GPT-3,5 (54 %), GPT-4 (72 %), Bard (56 %)	Dental licensing exams	GPT-4: 72 %	Passing threshold	
Vaira et al. ⁽⁴⁶⁾	Observational	GPT-4	Head and neck surgery scenarios	84,7 % closed-ended; 81,7 % clinical scenarios	Second-year resident (superior)	
Ma et al. ⁽²⁵⁾	RCT	GPT-4	Periodontology exam generation	Content coverage 81,3 % vs. 72,4 % (human)	Human-designed exam	
Islam ⁽⁷⁾	RCT	ChatGPT	Patient education (endodontics / restorative)	Clarity 4,42 vs. 3,25; anxiety 2,63 vs. 3,38	Verbal + pamphlets	
Santonocito et al. ⁽⁸⁾	RCT	Botpress LLM chatbot	Orthodontic patient education	Lower MGI increase ($p < 0,001$)	Information leaflets	
Hunsrisakhun et al. ⁽²²⁾	RCT	Standalone chatbot	Childhood caries prevention	Comparable dmft and plaque reduction	Chatbot + in-person training	
Tanaka et al. ⁽¹⁸⁾	Cross-sectional	ChatGPT-3,5, Bing, Bard	Orthodontic treatment planning	Significantly less accepted than human plan ($p < 0,001$)	Experienced orthodontist	

Miscellaneous AI Applications

Peng et al.⁽³¹⁾ applied lasso-regularized logistic regression to predict general dental anesthesia use among 12 410 children with behavioral health diagnoses and reported that models incorporating electronic medical records (lift = 2,59) or the Ohio Child Opportunity Index (lift = 2,56) independently achieved satisfactory performance above the 2,5 threshold, whereas combining both sources actually decreased performance (lift = 2,34). Digmayer Romero et al.⁽⁵¹⁾ developed ML models for predicting posterior restoration failure, achieving AUC values of 0,67–0,75 for primary teeth but only 0,53–0,62 for permanent teeth. Rivera Capacho et al.⁽⁵⁸⁾ applied an AI-enhanced telemedicine framework (Smart Therapy Manager) to orofacial myofunctional therapy for obstructive sleep apnea in 87 adults and reported improvements in Epworth Sleepiness Scale scores from $9,41 \pm 5,65$ to $7,26 \pm 5,43$. Stillhart et al.⁽³⁰⁾ evaluated automated facial coding (AFC) software for dental pain screening in 57 patients and reported that while pain VAS scores significantly decreased from baseline ($5,65 \pm 2,08$) to posttreatment ($0,40 \pm 0,80$; $p < 0,001$), the software did not detect statistically significant differences in facial expression between time points ($p > 0,05$), indicating the current limitations of this approach. Wong et al.⁽⁶⁵⁾ used automated smile feature detection software in an observational study of 42 individuals with unilateral cleft lip and reported that smile features and personality traits did not significantly differ between the UCL $\pm$ P and control groups. Uribe et al.⁽¹⁵⁾ cataloged 16 publicly available dental imaging datasets and reported that 75 % contained annotations, that tooth segmentation was the dominant focus (62,5 %), and that panoramic radiography was the most common modality (58,8 %), but only 31,2 % reported ethical approval and 56,25 % specified no license, highlighting the fragmented data infrastructure underpinning the development of dental AI.

Synthesis

Apparent Contradictions in AI Diagnostic Performance

Across the included studies, AI diagnostic performance metrics vary substantially even within the same application domain. For caries detection, the pooled sensitivity ranged from 76 %⁽⁴³⁾ to 94 %,⁽¹⁾ depending on the review, whereas Devlin et al.⁽⁵⁰⁾ reported an AI-assisted sensitivity of 75,8 % paired with a sharp decrease in specificity. A longitudinal study by Rodrigues et al.⁽³⁹⁾ revealed that the same platform increased specificity while reducing sensitivity over a four-year software update cycle) suggesting that individual AI tools undergo deliberate recalibration that alters the sensitivity–specificity tradeoff over time rather than uniformly improving all the metrics. Clinicians adopting AI tools therefore cannot assume static performance characteristics and should rely on version-specific validation data.

The Role of Experience Level in AI Benefit

A pattern across multiple application areas is that AI provides the most measurable benefit to less experienced practitioners. Pul et al.⁽³⁴⁾ reported that junior dentists gained the most from AI assistance in periapical radiolucency detection, whereas senior dentists derived minimal additional benefit. Hegde et al.⁽³⁷⁾ confirmed that dental students aided by ML algorithms achieved higher sensitivity and interrater reliability than those working unaided or with checklists did. Similarly, Schwartz et al.⁽¹³⁾ reported that the performance of DL for cephalometric landmark detection is comparable to that of experienced clinicians but may exceed the accuracy of inexperienced clinicians. Conversely, Tanaka et al.⁽⁴⁶⁾ reported that experienced orthodontists substantially preferred human expert treatment plans over LLM-generated plans, and Vaira et al.⁽³⁴⁾ reported that GPT-4 was inferior to even a second-year resident for clinical scenarios. This pattern suggests that LLMs in their current form are more useful as entry-level educational or patient-facing tools than as clinical decision support for expert practitioners, whereas image-based AI diagnostics may genuinely augment specialist performance by reducing false positives.

Technical Task Difficulty and AI Performance Gradients

The performance is not uniform across imaging tasks even when the same broad technique (deep learning on radiographs) is applied. For tasks with well-defined geometric features—tooth numbering, implant brand detection, and carotid calcification detection—AI achieves consistently high accuracy (88 %–95 %+).^(2,3,48) For tasks requiring nuanced clinical judgment—carious lesion grading with ICDAS, periodontitis staging, and cervical vertebral maturation—the performance is more variable and lower.^(28,42) The CVM assessment data from Sadeghi et al.⁽²⁸⁾ illustrate this gradient precisely: the accuracy is highest for CS1 (DOR 213) and lowest for CS3 (DOR 32), reflecting the greater morphological overlap among intermediate stages that also challenges human raters. This suggests that AI applications are currently most reliable for categorical or geometrically defined tasks and require more development for continuous or ordinal clinical classifications.

Surgical Guidelines: Consistent Accuracy Hierarchy

Across the implant surgery literature, there is a consistent hierarchy: robotic systems < dynamic navigation < static guides in angular accuracy, but static guides are faster and associated with better short-term patient-reported recovery.⁽²⁷⁾ The pooled data from Luo et al.⁽⁴⁾ and Khaohoen et al.⁽⁴⁵⁾ converge on robotic CAIS deviations of approximately 0.7 mm coronal and 1.6–1.7° angular, substantially lower than freehand placement. An RCT by Yang et al.⁽⁵⁾ confirms this finding in a clinical setting with statistical rigor. Liu, Chen et al.⁽⁵⁷⁾ extended this finding to pediatric oral surgery, where an autonomous robotic platform additionally eliminated nerve injury and reduced bone removal. However, the cost and complexity of robotic platforms, combined with evidence that static guides already provide clinically acceptable accuracy and require shorter operative times, indicate that the incremental benefit of robotics must be weighed against practical constraints in most clinical settings.

LLM Accuracy Ceiling in Dental Applications

Across licensing examination benchmarks, GPT-4 achieves approximately 63–72 % accuracy in dentistry,^(16,17) which is consistently below the passing thresholds in most jurisdictions and below GPT-4's own performance in medicine and pharmacy.⁽¹⁷⁾ The RCT by Ma et al.⁽²⁵⁾ provides a mechanistic explanation: AI-generated content has broader topical coverage but lower discrimination between easy and difficult items, likely because LLMs are trained on broad textual corpora rather than case-graded clinical datasets. Dentistry's reliance on image-based diagnostics⁽¹⁶⁾ and the relative scarcity of high-quality dental text in LLM training data⁽¹⁶⁾ further constrain performance. In contrast, patient-facing educational applications—where the bar for accuracy is lower and the benefit of accessibility and anxiety reduction is high—show consistent advantages for AI-generated content over conventional pamphlets and verbal explanations.^(7,8) This bifurcation suggests that compared with professional decision support or credentialing contexts, LLMs currently occupy a more defensible niche in patient communication.

Data Infrastructure as a Systemic Limitation

Uribe et al.⁽¹⁵⁾ reported that only 16 publicly available dental imaging datasets exist, with inconsistent annotation methods, minimal ethical documentation, and low FAIRness scores for CBCT data. This infrastructure gap helps explain why many individual AI studies rely on small, single-center training sets—Mansour et al.⁽⁴¹⁾ trained on 37 CBCT scans, and Weingart et al.⁽³⁶⁾ on 30 CT scans—and why high heterogeneity is a near-universal finding in meta-analyses.^(4,9,10,13) Studies examining AI performance across multiple institutions or populations remain rare, and almost no studies have reported formal external validation in an independent clinical population. The geographic concentration of primary studies in East Asia^(4,12) further limits the generalizability of trained models to diverse patient populations, since dental morphology, disease prevalence, and imaging equipment differ systematically across regions.^(12,28) Until larger, ethically approved, openly accessible, and demographically diverse dental imaging repositories become available, claims about the generalizability of any specific AI model must be treated with substantial caution.

CONCLUSIONS

The analyzed evidence indicates that artificial intelligence is integrated into multiple areas of dentistry, with established applications in diagnostic imaging, treatment planning, clinical prediction, and patient education. Systems based on deep learning and machine learning demonstrate high performance on structured tasks, particularly in radiographic analysis and procedures with strong geometric dependencies, achieving accuracy levels comparable to those of human operators.

The observed benefits focus on improving diagnostic accuracy, reducing interobserver variability, and supporting decision-making in specific clinical contexts. Advantages include standardizing processes, supporting less experienced professionals, and

communicating with patients using tools based on language models.

However, the evidence has significant limitations. Heterogeneity prevails across study designs, datasets, and metrics, hindering direct comparisons between studies. Many studies are based on small samples and lack external validation, limiting the generalizability of their results. Risks of bias associated with data selection, model training, and the geographic concentration of research are identified. System performance declines in complex clinical scenarios that require contextual interpretation and specialized judgment.

In this context, artificial intelligence should be considered a supportive tool within the clinical process, without replacing professional assessment. Its effective implementation requires validation in real-world settings, integration with existing clinical workflows, and the development of standards for performance evaluation and reproducibility.

Further research with more rigorous methodological designs, representative samples, and multicenter validation is needed to more accurately establish the clinical scope of these technologies and their impact on oral health outcomes.

FINANCING

None.

CONFLICT OF INTEREST

None.

AUTHORSHIP CONTRIBUTION

Conceptualization: Rohan Sharma, Dewi Lestari, Ahmad Rizky Pratama, Arjun Patel.

Data curation: Rohan Sharma, Dewi Lestari, Ahmad Rizky Pratama, Arjun Patel.

Formal analysis: Rohan Sharma, Dewi Lestari, Ahmad Rizky Pratama, Arjun Patel.

Drafting - original draft: Rohan Sharma, Dewi Lestari, Ahmad Rizky Pratama, Arjun Patel.

Writing - proofreading and editing: Rohan Sharma, Dewi Lestari, Ahmad Rizky Pratama, Arjun Patel.

BIBLIOGRAPHIC REFERENCES

1. Carvalho BKG, Nolden E-L, Wenning AS, et al. Diagnostic accuracy of artificial intelligence for approximal caries on bitewing radiographs: A systematic review and meta-analysis. *Journal of dentistry*. 2024; <https://doi.org/10.1016/j.jdent.2024.105388>
2. Sadr S, Rokhshad R, Daghighi Y, et al. Deep learning for tooth identification and numbering on dental radiography: a systematic review and meta-analysis. *Dento maxillo facial radiology*. 2024; <https://doi.org/10.1093/dmfr/twad001>
3. Arzani S, Soltani P, Karimi A, et al. Detection of carotid artery calcifications using artificial intelligence in dental radiographs: a systematic review and meta-analysis. *BMC medical imaging*. 2025; <https://doi.org/10.1186/s12880-025-01719-9>
4. Luo Z, Li A, Unkovskiy A, et al. Accuracy of robotic computer-assisted implant surgery in clinical studies: a systematic review and meta-analysis. *BMC oral health*. 2025; <https://doi.org/10.1186/s12903-025-05837-2>
5. Yang F, Chen J, Cao R, et al. Comparative analysis of dental implant placement accuracy: Semi-active robotic versus free-hand techniques: A randomized controlled clinical trial. *Clinical implant dentistry and related research*. 2024; <https://doi.org/10.1111/cid.13375>
6. Li Y, Wu X, Liu M, et al. Enhanced control of periodontitis by an artificial intelligence-enabled multimodal-sensing toothbrush and targeted mHealth micromessages: A randomized trial. *Journal of clinical periodontology*. 2024; <https://doi.org/10.1111/jcpe.13987>
7. Islam S. Evaluating the impact of AI-generated educational content on patient understanding and anxiety in endodontics and restorative dentistry: a comparative study. *BMC oral health*. 2025; <https://doi.org/10.1186/s12903-025-06069-0>
8. Santonocito S, Cicciù M, Ronsivalle V. Evaluation of the impact of AI-based chatbot on orthodontic patient education: a preliminary randomised controlled trial. *Clinical oral investigations*. 2025; <https://doi.org/10.1007/s00784-025-06356-8>
9. Hendrickx J, Gracea RS, Vanheers M, et al. Can artificial intelligence-driven cephalometric analysis replace manual tracing? A systematic review and meta-analysis. *European journal of orthodontics*. 2024; <https://doi.org/10.1093/ejo/cjae029>
10. Luke AM, Rezallah NNF. Accuracy of artificial intelligence in caries detection: a systematic review and meta-analysis. *Head & face medicine*. 2025; <https://doi.org/10.1186/s13005-025-00496-8>
11. Dashti M, Khosraviani F, Ghadimi N, et al. Use of artificial intelligence for detection of MB2 canals in maxillary first molars on CBCT: a systematic review and meta-analysis. *BMC oral health*. 2025; <https://doi.org/10.1186/s12903-025-07254-x>

12. Li X, Zhao D, Xie J, et al. Deep learning for classifying the stages of periodontitis on dental images: a systematic review and meta-analysis. *BMC oral health*. 2023; <https://doi.org/10.1186/s12903-023-03751-z>
13. Schwendicke F, Chaurasia A, Arsiwala L, et al. Deep learning for cephalometric landmark detection: systematic review and meta-analysis. *Clinical oral investigations*. 2021; <https://doi.org/10.1007/s00784-021-03990-w>
14. de Queiroz Tavares Borges Mesquita G, Vieira WA, Vidigal MTC, et al. Artificial Intelligence for Detecting Cephalometric Landmarks: A Systematic Review and Meta-analysis. *Journal of digital imaging*. 2023; <https://doi.org/10.1007/s10278-022-00766-w>
15. Uribe SE, Issa J, Sohrabniya F, et al. Publicly Available Dental Image Datasets for Artificial Intelligence. *Journal of dental research*. 2024; <https://doi.org/10.1177/00220345241272052>
16. Liu M, Okuhara T, Huang W, et al. Large Language Models in Dental Licensing Examinations: Systematic Review and Meta-Analysis. *International dental journal*. 2025; <https://doi.org/10.1016/j.identj.2024.10.014>
17. Jin HK, Lee HE, Kim E. Performance of ChatGPT-3.5 and GPT-4 in national licensing examinations for medicine, pharmacy, dentistry, and nursing: a systematic review and meta-analysis. *BMC medical education*. 2024; <https://doi.org/10.1186/s12909-024-05944-8>
18. Tanaka OM, Gasparello GG, Mota-Júnior SL, et al. Effectiveness of AI-generated orthodontic treatment plans compared to expert orthodontist recommendations: a cross-sectional pilot study. *Dental press journal of orthodontics*. 2025; <https://doi.org/10.1590/2177-6709.30.1.e2524186.oar>
19. Nieri M, Serni L, Clauser T, et al. Diagnosis of Oral Cancer With Deep Learning. A Comparative Test Accuracy Systematic Review. *Oral diseases*. 2025; <https://doi.org/10.1111/odi.15330>
20. Kahya Karaca S, Akca K. Comparison of conventional and digital impression approaches for edentulous maxilla: clinical study. *BMC oral health*. 2024; <https://doi.org/10.1186/s12903-024-05151-3>
21. Al-Sarem M, Al-Asali M, Alqutaibi AY, Saeed F. Enhanced Tooth Region Detection Using Pretrained Deep Learning Models. *International journal of environmental research and public health*. 2022; <https://doi.org/10.3390/ijerph192215414>
22. Hunsrisakhun J, Naorungroj S, Tangkuptanon W, et al. Impact of Oral Health Chatbot With and Without Toothbrushing Training on Childhood Caries. *International dental journal*. 2025; <https://doi.org/10.1016/j.identj.2024.09.028>
23. Castro Dos Santos N, Mangussi A, Ribeiro T, et al. Factors influencing the response to periodontal therapy in patients with diabetes: post hoc analysis of a randomized clinical trial using machine learning. *Journal of applied oral science : revista FOB*. 2025; <https://doi.org/10.1590/1678-7757-2025-0211>
24. Adel SM, Bichu YM, Pandian SM, et al. Clinical audit of an artificial intelligence (AI) empowered smile simulation system: a prospective clinical trial. *Scientific reports*. 2024; <https://doi.org/10.1038/s41598-024-69314-6>
25. Ma X, Pan W, Yu X-N. Evaluating AI-generated examination papers in periodontology: a comparative study with human-designed counterparts. *BMC medical education*. 2025; <https://doi.org/10.1186/s12909-025-07706-6>
26. Liu C, Liu Y, Yi C, et al. Application of a 3D Fusion Model to Evaluate the Efficacy of Clear Aligner Therapy in Malocclusion Patients: Prospective Observational Study. *Journal of medical Internet research*. 2025; <https://doi.org/10.2196/67378>
27. Shi J-Y, Wu X-Y, Lv X-L, et al. Comparison of Implant Precision with Robots, Navigation, or Static Guides. *Journal of dental research*. 2025; <https://doi.org/10.1177/00220345241285566>
28. Sadeghi TS, Ourang SA, Sohrabniya F, et al. Performance of artificial intelligence on cervical vertebral maturation assessment: a systematic review and meta-analysis. *BMC oral health*. 2025; <https://doi.org/10.1186/s12903-025-05482-9>
29. Hu Y-J, Lin P-C, Lin P-C, et al. Effects of Mobile-Based AI-Enhanced Health Counseling on Protection Motivation, Self-Care Behaviors, and Periodontal Treatment Outcomes in Patients With Periodontitis: Randomized Controlled Trial. *Journal of medical Internet research*. 2025; <https://doi.org/10.2196/78211>
30. Stillhart A, Häfliger R, Takeshita L, et al. Screening for dental pain using an automated face coding (AFC) software. *Journal of dentistry*. 2025; <https://doi.org/10.1016/j.jdent.2025.105647>

31. Peng J, Gorham TJ, Meyer BD. Predicting Dental General Anesthesia Use among Children with Behavioral Health Conditions. JDR clinical and translational research. 2025; <https://doi.org/10.1177/23800844241252817>
32. Uzar E, Pence I, Casmeli MS, Yetkin Ay Z. Classification success of salivary interleukin-1 β in periodontitis grading with artificial intelligence models: a cross-sectional observational study. Journal of applied oral science : revista FOB. 2025; <https://doi.org/10.1590/1678-7757-2024-0580>
33. Tarighatnia A, Amanzadeh M, Hamedan M, et al. Deep learning-based evaluation of panoramic radiographs for osteoporosis screening: a systematic review and meta-analysis. BMC medical imaging. 2025; <https://doi.org/10.1186/s12880-025-01626-z>
34. Pul U, Tichy A, Pitchika V, Schwendicke F. Impact of artificial intelligence assistance on diagnosing periapical radiolucencies: A randomized controlled trial. Journal of dentistry. 2025; <https://doi.org/10.1016/j.jdent.2025.105868>
35. Deng K, Zonta F, Yang H, et al. Development of a machine learning multiclass screening tool for periodontal health status based on non-clinical parameters and salivary biomarkers. Journal of clinical periodontology. 2024; <https://doi.org/10.1111/jcpe.13856>
36. Weingart JV, Schlager S, Metzger MC, et al. Automated detection of cephalometric landmarks using deep neural patchworks. Dento maxillo facial radiology. 2023; <https://doi.org/10.1259/dmfr.20230059>
37. Hegde S, Gao J, Cox S, et al. Machine learning algorithms enhance the accuracy of radiographic diagnosis of dental caries: a comparative study. Dento maxillo facial radiology. 2025; <https://doi.org/10.1093/dmfr/twaf053>
38. Bspalez-Neto R, Mattos CT, Cevidanes LHS, et al. Quantifying Root Resorption on the Incisors After Clear Aligner and Fixed Appliance Therapy Using Artificial Intelligence Tool Based CBCT Surface Models: Randomized Clinical Trial. Orthodontics & craniofacial research. 2026; <https://doi.org/10.1111/ocr.70082>
39. Rodrigues N, Bonfanti-Gris M, Bai S, et al. Longitudinal assessment of an AI-based software for interproximal caries detection in bitewing radiographs. Journal of dentistry. 2026; <https://doi.org/10.1016/j.jdent.2025.106252>
40. Badr FF, Jadu FM. Performance of artificial intelligence using oral and maxillofacial CBCT images: A systematic review and meta-analysis. Nigerian journal of clinical practice. 2022; https://doi.org/10.4103/njcp.njcp_394_22
41. Mansour S, Anter E, Mohamed AK, et al. Two step approach for detecting and segmenting the second mesiobuccal canal of maxillary first molars on cone beam computed tomography (CBCT) images via artificial intelligence. BMC oral health. 2025; <https://doi.org/10.1186/s12903-025-06796-4>
42. Khubrani YH, Thomas D, Slator PJ, et al. Detection of periodontal bone loss and periodontitis from 2D dental radiographs via machine learning and deep learning: systematic review employing APPRAISE-AI and meta-analysis. Dento maxillo facial radiology. 2025; <https://doi.org/10.1093/dmfr/twae070>
43. Abbott LP, Saikia A, Anthonappa RP. ARTIFICIAL INTELLIGENCE PLATFORMS IN DENTAL CARIES DETECTION: A SYSTEMATIC REVIEW AND META-ANALYSIS. The journal of evidence-based dental practice. 2025; <https://doi.org/10.1016/j.jebdp.2024.102077>
44. Estrella N-F, Alexandra D-S, Yun C, et al. AI-AIDED VOLUMETRIC ROOT RESORPTION ASSESSMENT FOLLOWING PERSONALIZED FORCES IN ORTHODONTICS: PRELIMINARY RESULTS OF A RANDOMIZED CLINICAL TRIAL. The journal of evidence-based dental practice. 2025; <https://doi.org/10.1016/j.jebdp.2025.102095>
45. Khaohoen A, Powcharoen W, Sornsuwan T, et al. Accuracy of implant placement with computer-aided static, dynamic, and robot-assisted surgery: a systematic review and meta-analysis of clinical trials. BMC oral health. 2024; <https://doi.org/10.1186/s12903-024-04033-y>
46. Vaira LA, Lechien JR, Abbate V, et al. Accuracy of ChatGPT-Generated Information on Head and Neck and Oromaxillofacial Surgery: A Multicenter Collaborative Analysis. Otolaryngology--head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery. 2024; <https://doi.org/10.1002/ohn.489>
47. Bonfanti-Gris M, Herrera A, Salido Rodríguez-Manzaneque MP, et al. Deep learning for tooth detection and segmentation in panoramic radiographs: a systematic review and meta-analysis. BMC oral health. 2025; <https://doi.org/10.1186/s12903-025-06349-9>

48. Dashti M, Londono J, Ghasemi S, et al. Evaluation of accuracy of deep learning and conventional neural network algorithms in detection of dental implant type using intraoral radiographic images: A systematic review and meta-analysis. *The Journal of prosthetic dentistry*. 2025; <https://doi.org/10.1016/j.prosdent.2023.11.030>
49. Serafin M, Baldini B, Cabitza F, et al. Accuracy of automated 3D cephalometric landmarks by deep learning algorithms: systematic review and meta-analysis. *La Radiologia medica*. 2023; <https://doi.org/10.1007/s11547-023-01629-2>
50. Devlin H, Williams T, Graham J, Ashley M. The ADEPT study: a comparative study of dentists' ability to detect enamel-only proximal caries in bitewing radiographs with and without the use of AssistDent artificial intelligence software. *British dental journal*. 2021; <https://doi.org/10.1038/s41415-021-3526-6>
51. Digmayer Romero VH, Chaves ET, Vinayahalingam S, et al. Predicting restoration failures in primary and permanent teeth - A machine learning approach. *Dental materials : official publication of the Academy of Dental Materials*. 2026; <https://doi.org/10.1016/j.dental.2025.09.009>
52. Pozzi A, Carosi P, Laureti A, et al. Accuracy of navigation guided implant surgery for immediate loading complete arch restorations: Prospective clinical trial. *Clinical implant dentistry and related research*. 2024; <https://doi.org/10.1111/cid.13360>
53. Dashti M, Azimi T, Khosraviani F, et al. Systematic Review and Meta-Analysis on the Accuracy of Artificial Intelligence Algorithms in Individuals Gender Detection Using Orthopantomograms. *International dental journal*. 2025; <https://doi.org/10.1016/j.identj.2024.12.018>
54. Picoli FF, Fontenele RC, Van der Cruyssen F, et al. Risk assessment of inferior alveolar nerve injury after wisdom tooth removal using 3D AI-driven models: A within-patient study. *Journal of dentistry*. 2023; <https://doi.org/10.1016/j.jdent.2023.104765>
55. Ziaei S, Samani D, Behjati M, et al. Accuracy of artificial intelligence in orthodontic extraction treatment planning: a systematic review and meta analysis. *BMC oral health*. 2025; <https://doi.org/10.1186/s12903-025-06880-9>
56. Nascimento da Silva Mulder J, Ramos Pinto M, Aníbal I, et al. Teledentistry Applied to Health and Education Outcomes: Evidence Gap Map. *Journal of medical Internet research*. 2024; <https://doi.org/10.2196/60590>
57. Liu C, Di T, Chen Y, et al. Efficiency for robotic-assisted extraction of completely impacted supernumerary teeth in children: a randomized controlled trial. *BMC oral health*. 2025; <https://doi.org/10.1186/s12903-025-07167-9>
58. Rivera Capacho EE, Bossa CPD, Campos MDC, et al. Telemedicine-supported structured Orofacial Myofunctional Therapy model for Obstructive Sleep Apnea: Patients' report outcomes measurements. *Respiratory medicine*. 2025; <https://doi.org/10.1016/j.rmed.2025.108460>
59. Canullo L, Donato A, Savadori P, et al. Effect of argon plasma abutment activation on soft tissue healing: RCT with histological assessment. *Clinical implant dentistry and related research*. 2024; <https://doi.org/10.1111/cid.13286>
60. Ghensi P, Heidrich V, Bazzani D, et al. Shotgun Metagenomics Identifies in a Cross-Sectional Setting Improved Plaque Microbiome Biomarkers for Peri-Implant Diseases. *Journal of clinical periodontology*. 2025; <https://doi.org/10.1111/jcpe.14121>
61. Chau RCW, Chong M, Thu KM, et al. Artificial intelligence-designed single molar dental prostheses: A protocol of prospective experimental study. *PloS one*. 2022; <https://doi.org/10.1371/journal.pone.0268535>
62. Takács A, Hardi E, Cavalcante BGN, et al. Advancing accuracy in guided implant placement: A comprehensive meta-analysis: Meta-Analysis evaluation of the accuracy of available implant placement Methods. *Journal of dentistry*. 2023; <https://doi.org/10.1016/j.jdent.2023.104748>
63. Jain S, Sayed ME, Ibraheem WI, et al. Accuracy Comparison between Robot-Assisted Dental Implant Placement and Static/Dynamic Computer-Assisted Implant Surgery: A Systematic Review and Meta-Analysis of In Vitro Studies. *Medicina (Kaunas, Lithuania)*. 2023; <https://doi.org/10.3390/medicina60010011>
64. Pražetina M, Šribar A, Sokolović Jurinjak I, et al. Effect of machine learning-guided haemodynamic optimization on postoperative free flap perfusion in reconstructive maxillofacial surgery: A study protocol. *British journal of clinical pharmacology*. 2024; <https://doi.org/10.1111/bcp.15942>
65. Wong L, Firth F, Fowler P, et al. Association of Clinical, Biomechanical, and Psychosocial Factors with Smile Dynamics in

Unilateral Cleft Lip: A Multicenter Observational Study. The Cleft palate-craniofacial journal : official publication of the American Cleft Palate-Craniofacial Association. 2026; <https://doi.org/10.1177/10556656241291649>