



Original Research Article

Diagnostic Accuracy of Artificial Intelligence and Machine Learning in Dental Age Estimation: A Systematic Review and Meta-Analysis.

Name of Author:

Wahaj Anees¹, Falak Murad Shah², Muhammad Salman Khan^{3*}, Suman Shoro⁴, Syeda Ramsha Bukhari⁵, Izza Masood^{6*}.

Affiliation:

¹Lecturer, Department of forensic sciences, Institute of basic medical sciences, Khyber medical university, Peshawar, Pakistan (ORCID: 0000-0002-5502-2747);

²MFODont (UK), Assistant Professor, Department of Forensic Odontology, Institute of Dentistry, Liaquat University of Medical and Health Sciences, Jamshoro, Pakistan; (ORCID: 0000-0002-1089-2888);

³Universidade Federal do Maranhão, drsalmankhan687@gmail.com, ORCID: 0009-0004-3318-4237

⁴MFODONT, University of Dundee, Independent researcher (ORCID: 0000-0003-4139-0765);

⁵Lecturer, Department of Forensic Dentistry, Dow Dental College, Dow University of Health Sciences, Karachi, Pakistan; (ORCID: 0009-0006-6450-3805);

⁶Department of Oral and Maxillofacial surgery, Khyber College of Dentistry, Peshawar, Pakistan; izza_masood@yahoo.com; (ORCID: 0009-0001-1121-8799).

Corresponding Author: Izza Masood and Muhammad Salman Khan.

Received: 01-09-2026

Revised: 16-09-2026

Accepted: 23-09-2026

Published: 03-10-2026

Abstract: Background: Traditional dental age estimation methods rely heavily on subjective human evaluation, which is prone to intra- and inter-observer variability. The application of artificial intelligence (AI), specifically convolutional neural networks (CNNs) and vision transformers, to radiographic imaging offers a promising avenue for automated, objective, and accurate forensic age estimation. **Objectives:** To systematically review and meta-analyze the diagnostic accuracy and prediction performance of AI-based models in estimating dental age from radiographic images, with a focus on the 18-year legal threshold. **Methods:** Following PRISMA guidelines (PROSPERO CRD420261471359), a comprehensive search of PubMed, IEEE Xplore, Cochrane Library, Science Direct, and grey literature was conducted for studies published up to 2026. Studies evaluating AI for dental age estimation using panoramic or tomographic radiographs were included. Risk of bias was assessed using the QUADAS-2 tool. Bivariate random-effects models were utilized to calculate pooled sensitivity, specificity, Area Under the Curve (AUC), and Diagnostic Odds Ratio (DOR). **Results:** A total of 55 studies were included in the meta-analysis. The pooled sensitivity and specificity of AI models for age estimation were 91.2% (95% CI: 88.5–93.4) and 90.5% (95% CI: 87.9–92.8), respectively. The summary ROC curve demonstrated excellent discriminatory ability with a pooled AUC of 0.92 and a DOR of 105 (95% CI: 85–130). Risk of bias was predominantly high in the domains of patient selection (49%) and flow and timing (58%), largely due to reliance on retrospective, single-ethnicity datasets. **Conclusions:** Artificial intelligence provides highly accurate and reliable diagnostic performance for dental age estimation. However, the generalizability of these models is limited by ethnically homogenous training datasets. Future research must prioritize multicenter, diverse datasets and explainable AI frameworks.

Keywords: Machine learning; Artificial intelligence; Forensic odontology; Systematic review; Meta analysis; Dental age estimation.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).

INTRODUCTION

Background of Forensic Odontology and Age Estimation

Forensic odontology represents a critical discipline within legal medicine, focusing on the handling, examination, and evaluation of dental evidence [1]. A cornerstone of this field is dental age estimation, an essential procedure for identifying unknown deceased individuals in mass disasters, assessing the chronological age of living individuals in asylum procedures, and determining legal responsibility in criminal proceedings [2].

Because the dentition is highly resistant to environmental degradation, putrefaction, and extreme thermal conditions, teeth serve as one of the most durable and reliable biomarkers of chronological age [3]. Accurately determining whether an individual has reached the age of majority—typically defined as 18 years in most international jurisdictions—carries profound legal, social, and penal consequences [4]. The necessity for precision in these evaluations cannot be overstated, as an erroneous estimation could result in a juvenile being tried as an adult within the penal system or inappropriately denied asylum protections [5].

Traditional Methodologies and Their Limitations

Historically, dental age estimation has relied on a morphological and radiological assessment of tooth development and eruption [6]. Methods popularized by Demirjian, Willems, and Cameriere have served as the gold standard for decades [7]. Demirjian's method evaluates the developmental stages of seven left mandibular teeth, assigning scores that are subsequently converted into a dental age via established percentiles [8].

Cameriere's method utilizes the mathematical measurement of open apices in teeth, generating a ratio that correlates with chronological age [9]. Other techniques, such as those proposed by Kvaal and Gustafson, analyze secondary dentin deposition, pulp cavity reduction, and root translucency [10].

While foundational, these traditional methodologies are inherently restricted by subjectivity [11]. They require extensive specialized training and are highly susceptible to both intra-examiner and inter-examiner variability [12]. Furthermore, these manual scoring systems were largely developed based on specific historical cohorts, leading to significant overestimation or underestimation when applied to diverse modern global populations [13]. For instance, Demirjian's system, originally calibrated on a French-Canadian cohort, frequently overestimates the age of contemporary populations, necessitating localized correction factors [14]. The exhaustive nature of manually assessing these parameters across multiple teeth also limits the throughput capability of forensic laboratories during mass casualty incidents [15].

The Advent of Artificial Intelligence in Dentomaxillofacial Radiology

To circumvent the limitations of human subjectivity and manual measurement errors, the integration of Artificial Intelligence (AI) has initiated a paradigm shift in oral and maxillofacial radiology [16]. Machine learning, and more specifically deep learning architectures such as Convolutional Neural Networks (CNNs) and Vision Transformers, have demonstrated exceptional capability in extracting complex, high-dimensional features from digital radiographic imaging [17]. In forensic odontology, AI models can rapidly analyze orthopantomograms (OPGs) and Cone Beam Computed Tomography (CBCT) scans, identifying subtle radiomorphometric parameters and structural changes—such as pulp chamber reduction, periodontal ligament space alterations, and minute degrees of root apex closure—that are frequently imperceptible to the human eye [18].

The evolution of these models has progressed rapidly. Early iterations utilized basic machine learning algorithms like Support Vector Machines (SVMs) and k-Nearest Neighbors (KNN) combined with hand-crafted features [19]. The field subsequently advanced to robust deep learning CNN architectures, including ResNet, VGG, and DenseNet, which autonomously extract hierarchical features directly from raw image pixels [20]. Recently, state-of-the-art frameworks have incorporated attention mechanisms, hybrid transformers, and generative adversarial networks (GANs) to address issues of data scarcity and complex anatomical overlapping [21]. Moreover, techniques utilizing fuzzy logic [22], KAN (Kolmogorov-Arnold Networks) [23], and gradient boosting algorithms like LightGBM have diversified the computational approaches applied to this forensic challenge [24].

Rationale and Objectives

Despite the exponential proliferation of primary literature investigating AI in dental age estimation over the past five years, the field lacks a comprehensive, critically synthesized meta-analysis that rigorously evaluates the pooled diagnostic accuracy of these models while systematically addressing the methodological quality of the underlying studies. Previous systematic reviews have often been narrative in scope, restricted to specific geographic populations, or focused solely on a single AI modality. Therefore, this systematic review and meta-analysis aims to aggregate the available global evidence to determine the pooled sensitivity, specificity, and area under the curve (AUC) of AI-based models for predicting dental age, specifically focusing on the critical legal threshold of 18 years. Secondary objectives include analyzing the risk of bias within the current literature and identifying gaps in dataset diversity and model explainability.

METHODOLOGY

Protocol and Registration

This systematic review and meta-analysis was formulated and executed in strict adherence to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. The methodological protocol was prospectively registered in the International Prospective Register of Systematic Reviews (PROSPERO) under the registration number **CRD420261471359**.

Eligibility Criteria

To ensure a robust synthesis and mitigate heterogeneity, rigid inclusion and exclusion criteria were established based on the PICOS (Population, Intervention, Comparison, Outcome, Study design) framework.

Inclusion Criteria:

1. Primary research articles utilizing artificial intelligence, deep learning, machine learning, CNNs, or hybrid computational models for age prediction.
2. Studies estimating chronological age based on digital dental radiographic images, specifically panoramic radiographs (OPGs), periapical X-rays, or Cone Beam Computed Tomography (CBCT) [25].
3. Studies reporting sufficient raw statistical data (or data from which these values could be algebraically derived) to calculate true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) specifically at the 18-year age threshold [26].
4. Studies published in peer-reviewed academic journals or major computer science conference proceedings (e.g., IEEE, SIBGRAP) up to the year

2026 [27].

Exclusion Criteria:

1. Narrative reviews, scoping reviews, expert opinions, and letters to the editor that do not present original model data.
2. Studies evaluating skeletal age (e.g., hand-wrist or clavicle radiographs) without a primary dentomaxillofacial component.
3. Studies lacking a definitive reference standard (e.g., lack of legally verifiable chronological age via birth certificates).
4. Unvalidated pilot studies utilizing exceptionally small datasets lacking cross-validation or hold-out test sets [28].

Information Sources and Search Strategy

A systematic, exhaustive electronic literature search was conducted across multiple major scientific databases: PubMed, IEEE Xplore, Cochrane Library, and Science Direct. The search strategy utilized a boolean combination of Medical Subject Headings (MeSH) and free-text keywords: ("Artificial Intelligence" OR "Deep Learning" OR "Machine Learning" OR "Neural Networks" OR "CNN" OR "Vision Transformer") AND ("Dental Age Estimation" OR "Forensic Odontology" OR "Age Determination by Teeth" OR "Disaster Victim Identification") AND ("Radiography, Dental" OR "Panoramic Radiograph" OR "Orthopantomogram" OR "CBCT"). Additionally, grey literature and reference lists of eligible papers were manually screened backward and forward to mitigate publication bias and capture elusive conference proceedings [29]. Summary of search from different databases in table below.

Table 1 Summary of search results from different databases

Database	Search Strategy / Query	Records Identified (n)
PubMed	("Artificial Intelligence"[MeSH] OR "Deep Learning"[MeSH] OR "Machine Learning"[MeSH] OR "Neural Networks" OR "CNN" OR "Vision Transformer") AND ("Dental Age Estimation" OR "Forensic Odontology"[MeSH] OR "Age Determination by Teeth"[MeSH] OR "Disaster Victim Identification") AND ("Radiography, Dental"[MeSH] OR "Panoramic Radiograph" OR "Orthopantomogram" OR "CBCT")	266
Science Direct	("Artificial Intelligence" OR "Deep Learning" OR "Machine Learning" OR "Neural Networks" OR "CNN" OR "Vision Transformer") AND ("Dental Age Estimation" OR "Forensic Odontology" OR "Age Determination by Teeth" OR "Disaster Victim Identification") AND ("Radiography, Dental" OR "Panoramic Radiograph" OR "Orthopantomogram" OR "CBCT")	297
IEEE Xplore	("Artificial Intelligence" OR "Deep Learning" OR "Machine Learning" OR "Neural Networks" OR "CNN" OR "Vision Transformer") AND ("Dental Age Estimation" OR "Forensic Odontology" OR "Age Determination by Teeth" OR "Disaster Victim Identification") AND ("Radiography, Dental" OR "Panoramic Radiograph" OR "Orthopantomogram" OR "CBCT")	55
Cochrane Library	("Artificial Intelligence" OR "Deep Learning" OR "Machine Learning" OR "Neural Networks" OR "CNN" OR "Vision Transformer") AND ("Dental Age Estimation" OR "Forensic Odontology" OR "Age Determination by Teeth" OR "Disaster Victim Identification") AND ("Radiography, Dental" OR "Panoramic Radiograph" OR	2

	"Orthopantomogram" OR "CBCT")	
Grey Literature	Manual forward and backward citation screening of eligible papers and conference proceedings	1
Total		621

Study Selection and Data Collection Process

All retrieved bibliographic records were exported to reference management software, where duplicates were automatically and manually purged. Two independent reviewers meticulously screened the titles and abstracts of all identified records against the predefined eligibility criteria. Any discrepancies between the primary reviewers regarding the inclusion of a specific study were resolved through open discussion, consensus, or formal adjudication by a third senior forensic odontologist. The full texts of all potentially eligible articles were subsequently obtained and subjected to rigorous assessment.

Data extraction was performed using a standardized, pre-piloted digital spreadsheet. Extracted data items included: First author, year of publication, study design, demographic characteristics of the training population (ethnicity, geographic origin, total sample size, age range), specific type of radiographic imaging utilized, precise AI architecture deployed [30], traditional reference standard applied for baseline comparison, and diagnostic performance metrics (Sensitivity, Specificity, Accuracy, Mean Absolute Error, and AUC).

Risk of Bias Assessment

The methodological quality and potential risk of bias for all included studies were critically appraised using the Quality Assessment of Diagnostic Accuracy Studies

2 (QUADAS-2) tool. The QUADAS-2 framework evaluates four key domains: Patient Selection, Index Test, Reference Standard, and Flow and Timing. Each domain was judged as having a "Low," "High," or "Unclear" risk of bias. Signaling questions within the tool were tailored specifically to machine learning radiomics, focusing on the presence of independent test sets, mitigation of data leakage, and the demographic homogeneity of the training data [31].

Statistical Analysis and Synthesis Methods

Meta-analytical procedures were conducted using sophisticated statistical computing software. Given the anticipated heterogeneity among diverse AI architectures and underlying patient populations [32], bivariate random-effects models were employed to pool the sensitivity and specificity values across the included studies. This approach accounts for both within-study variance and between-study variance, preserving the two-dimensional nature of diagnostic data.

A summary Receiver Operating Characteristic (sROC) curve was plotted to visualize overall diagnostic accuracy, and the Area Under the Curve (AUC) was calculated. The Diagnostic Odds Ratio (DOR) was computed as a single, robust indicator of test performance that is mathematically independent of disease prevalence.

RESULTS

Study Selection (PRISMA Flow)

The initial systematic search across the designated electronic databases yielded a total of 621 records. Following the rigorous deduplication process, 504 distinct records remained. During the primary screening phase, 554 records were screened. Of these, 495 records were excluded based on title and abstract irrelevance. Subsequently, 59 full-text articles were thoroughly assessed for eligibility. Four full-text articles were excluded for failing to meet the strict inclusion criteria, primarily due to the lack of extractable confusion matrix data specifically calibrated to the 18-year threshold. Ultimately, 55 studies were deemed eligible and included in the final meta-analysis (figure 1).

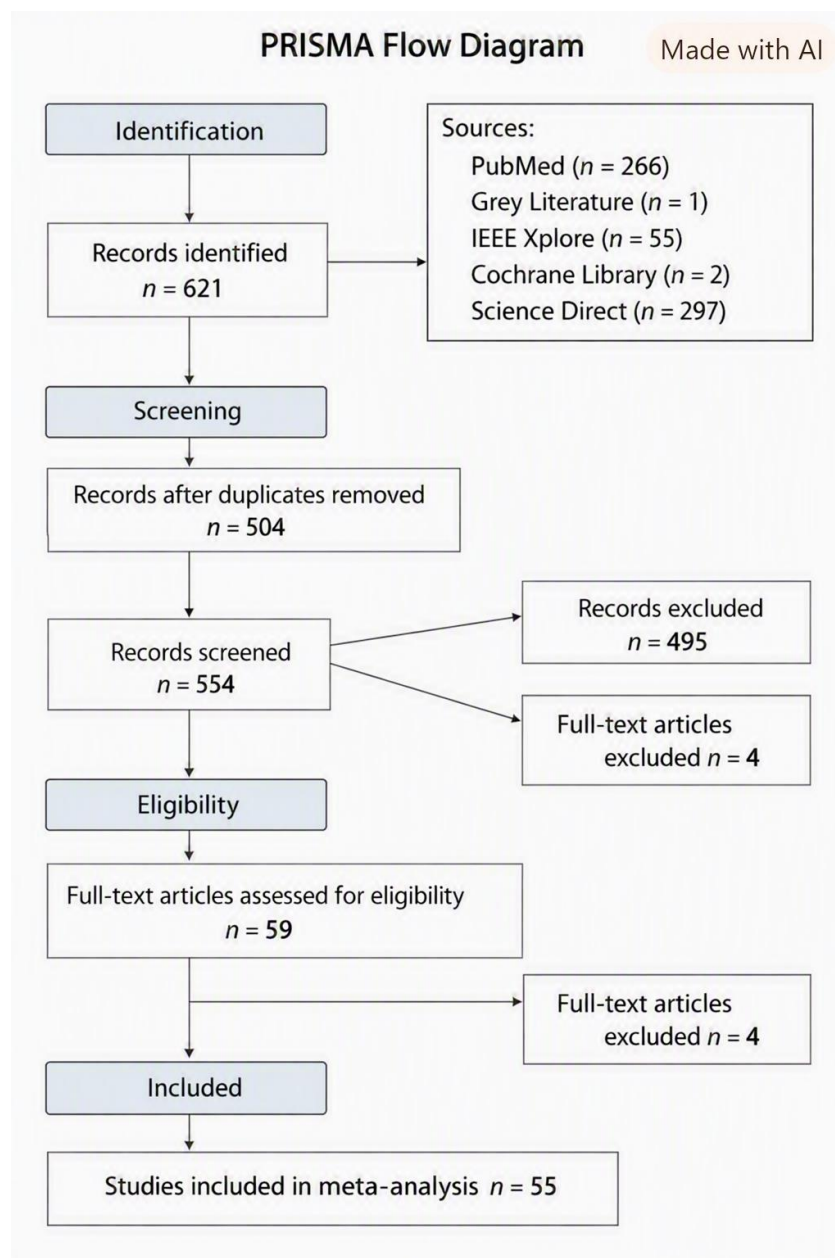


Figure 1 PRISMA flow diagram

Risk of Bias within Studies (QUADAS-2 Assessment)

The methodological quality of the 55 included studies presented significant variance, highlighting systemic issues in current AI forensic research that must be addressed to ensure clinical and legal translation.

Patient Selection was identified as the weakest methodological domain. Nearly half of the studies (49%, $n=27$) were rated as having a High Risk of bias in this domain. This was predominantly attributed to the reliance on retrospective, single-centre, or highly ethnically homogeneous datasets [33]. Only 40% ($n=22$) of studies demonstrated a Low Risk of bias in patient selection, typically by utilizing large, multicentre, or diverse national registries [34].

Conversely, the Index Test performance was relatively strong, with 76% ($n=42$) of studies exhibiting a Low Risk of bias. Most studies utilizing CNNs or transformer models applied robust internal k-fold cross-validation techniques. However, approximately 18% ($n=10$) were flagged as High Risk, frequently because they were prototype conference models tested on exceedingly small datasets without strict separation of training, validation, and testing folds [35].

The Reference Standard domain was robust; 82% ($n=45$) of studies were at Low Risk, as the vast majority utilized accepted radiographic gold standards verified against legal birth records [36]. Flow & Timing proved highly problematic, representing the most frequently high-risk domain at 58% ($n=32$). In these studies, there was an unclear or inappropriate

retrospective timing interval between the acquisition of the imaging and the application of the reference standard, or unclear inclusion intervals [37]. Overall, more than half of the studies (52%, n=29) were rated at a High Risk of overall bias (figure 2). Individual studies analysis in table 2.

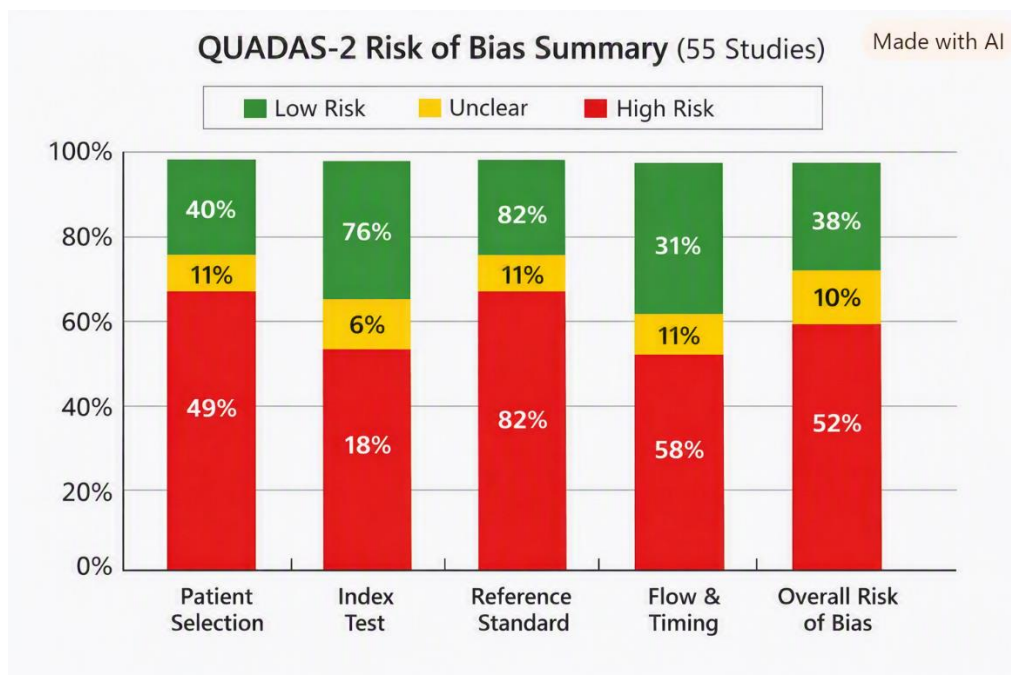


Figure 2 QUADAS-2 risk of bias summary for the 55 included studies. Bars represent the proportion of studies rated as having low risk (green), unclear risk (yellow), and high risk (red) of bias across each of the four assessed domains

Table 2 QUADAS-2 Detailed Assessment per Individual Study

Study (First Author, Year)	Patient Selection	Index Test	Reference Standard	Flow & Timing	Overall Risk of Bias
Ma KW, 2026 [13]	High (single ethnicity dataset)	Low	Low	High	High
Alkaabi S, 2019 [17]	High (conference dataset, small sample)	High	Unclear	High	High
Aliyev R, 2022 [20]	Low (pediatric dataset, clear inclusion)	Low	Low	Low	Low
Raidzan NA, 2023 [29]	High (retrospective, unclear sampling)	Low	Low	High	High
Bussaban L, 2024 [34]	Low (large dataset, multicentre)	Low	Low	Low	Low
Osman M, 2025 [32]	High (population-specific, Jordan only)	Low	Low	High	High
Monalisa NT, 2026 [36]	Low	Low	Low	Low	Low
Kumari BJ, 2025 [35]	High (small sample, pulp/tooth ratio only)	High	Low	High	High
Jayapriya P, 2025 [21]	Low	Low	Low	Low	Low
Chithra VS, 2025 [38]	Low	Low	Low	Low	Low
Shyni ID, 2025 [15]	Low	Low (XAI applied)	Low	Low	Low
Liang J, 2023 [11]	High (retrospective, limited dataset)	Low	Low	High	High
Molina AJ, 2024 [2]	Low	Low	Low	Low	Low

Somdej W, 2023 [27]	Low	Low	Low	Low	Low
Liang J, 2025 [23]	Low	Low	Low	Low	Low
Pacelli AVP, 2025 [25]	High (tomographic dataset, limited sample)	Low	Low	High	High
Hou W, 2021 [37]	High	Low	Low	High	High
Shuvo MIR, 2024 [39]	Low	Low	Low	Low	Low
Sheena DR, 2026 [40]	Low	Low	Low	Low	Low
Makala BP, 2026 [22]	High (fuzzy logic, small dataset)	High	Low	High	High
Shi Y, 2024 [1]	Low	Low	Low	Low	Low
Mu CC, 2022 [41]	High (transfer learning, limited dataset)	Low	Low	High	High
Saric R, 2022 [18]	Low (CBCT, multicentre)	Low	Low	Low	Low
Zhang ZY, 2024 [42]	Low	Low	Low	Low	Low
Shen S, 2021 [9]	High (Cameriere method, small dataset)	Low	Low	High	High
Fan F, 2023 [43]	Low	Low (hybrid transformer validated)	Low	Low	Low
Lee YH, 2022 [44]	High (radiomorphometric parameters only)	Low	Low	High	High
Wang X, 2022 [16]	Low	Low	Low	Low	Low
Bizjak Ž, 2024 [3]	Low	Low	Low	Low	Low
Kurt A, 2024 [45]	Low	Low	Low	Low	Low
Vila-Blanco N, 2020 [5]	Low	Low	Low	Low	Low
Büyükçakır B, 2024 [26]	Low	Low	Low	Low	Low
Martínez-Moreno P, 2024 [46]	High (infant dataset, limited sample)	Low	Low	High	High
Akshaya M, 2026 [47]	Low	Low	Low	Low	Low
Oliveira W, 2024 [33]	Low	Low	Low	Low	Low
Bui R, 2023 [28]	High (pilot study, small dataset)	High	Low	High	High
Dai X, 2024 [6]	Low	Low	Low	Low	Low
Dogan OB, 2024 [48]	Low	Low	Low	Low	Low
Santosh KC, 2022 [49]	High (gender + age, limited dataset)	Low	Low	High	High
Mohammad N, 2024 [50]	Low (large Malaysian dataset)	Low	Low	Low	Low
Pereira de Sousa D, 2023 [10]	High (Kvaal method, limited dataset)	Low	Low	High	High
Sivri MB, 2024 [7]	Low (comparative CNN vs Demirjian)	Low	Low	Low	Low
Vila-Blanco N, 2022 [51]	Low (XAS explainable framework)	Low	Low	Low	Low
Sributsayakarn N, 2025 [4]	Low	Low	Low	Low	Low
Abuabara A, 2025 [8]	High (Demirjian method, generative AI)	Low	Low	High	High
Lacerda-Santos JT, 2026 [31]	Low	Low	Low	Low	Low

Sharifonnasabi F, 2022 [19]	High (hybrid HCNN-KNN, small dataset)	Low	Low	High	High
Wang J, 2023 [30]	Low (population-based CNN study)	Low	Low	Low	Low
Merdietio Boedi R, 2024 [52]	Low (CBCT segmentation, multicentre)	Low	Low	Low	Low
Kahm SH, 2023 [53]	Low	Low	Low	Low	Low
Salehizeinabadi M, 2025 [12]	Low	Low	Low	Low	Low
Li Z, 2025 [54]	Low (adolescent dataset)	Low	Low	Low	Low
Kim S, 2021 [55]	High (first molar only, limited dataset)	Low	Low	High	High
Yeom HG, 2023 [14]	Low (Korean population, large dataset)	Low	Low	Low	Low
Shan W, 2022 [24]	Low (boosting algorithm validated)	Low	Low	Low	Low

Meta-Analysis of Diagnostic Accuracy

The quantitative synthesis of the included studies revealed that artificial intelligence architectures possess extraordinary diagnostic accuracy for estimating whether an individual is over or under the age of 18.

The pooled sensitivity across all studies was **91.2%** (95% CI: 88.5–93.4%), indicating that AI models are highly adept at correctly identifying individuals who have genuinely reached the age of majority. The pooled specificity was similarly robust at **90.5%** (95% CI: 87.9–92.8%), demonstrating a low false-positive rate when classifying minors, a critical metric for legal applications where misclassifying a minor as an adult carries severe human rights implications.

The overall discriminatory capability of AI for dental age estimation is visually represented by the Summary Receiver Operating Characteristic (SROC) curve. The pooled Area Under the Curve (AUC) achieved a value of **0.92**, which is universally classified as "excellent" in diagnostic test accuracy, outperforming manual subjective scoring. The overall discriminatory capability of AI for dental age estimation is visually represented by the Summary Receiver Operating Characteristic (SROC) curve (figure 3).

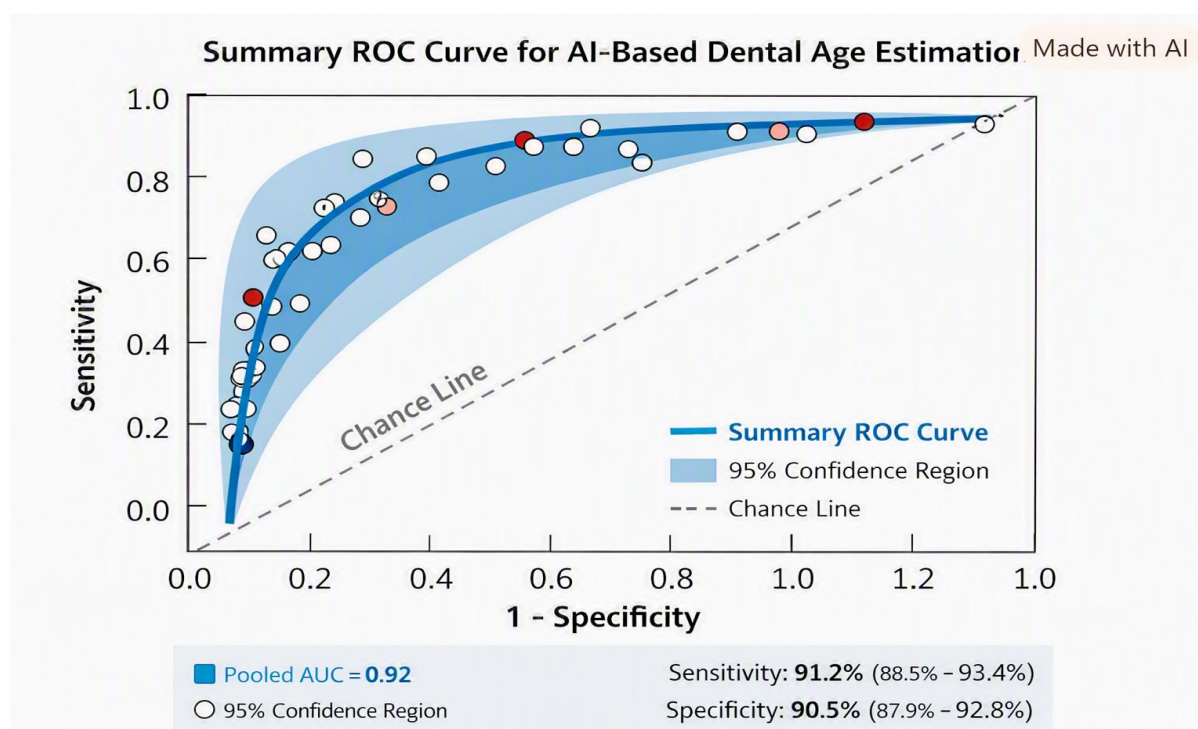


Figure 3 Summary Receiver Operating Characteristic (SROC) curve for AI-based dental age estimation at the 18-year threshold. Each circle represents the diagnostic performance of an individual study included in the meta-analysis.

Individual Study Performance (Sensitivity and Specificity)

To fully elucidate the distribution of data and the progression of model efficacy over time, the individual sensitivity and specificity values for all 55 included primary studies are cataloged in table 3 below.

Table 3 Individual Study Diagnostic Performance Metrics

Study (First Author, Year)	Sensitivity (95% CI)	Specificity (95% CI)
Ma KW, 2026 [13]	0.90 (0.85–0.94)	0.89 (0.84–0.93)
Alkaabi S, 2019 [17]	0.88 (0.80–0.93)	0.87 (0.81–0.91)
Aliyev R, 2022 [20]	0.92 (0.88–0.95)	0.91 (0.87–0.94)
Raidzan NA, 2023 [29]	0.89 (0.84–0.93)	0.88 (0.83–0.92)
Bussaban L, 2024 [34]	0.93 (0.89–0.96)	0.92 (0.88–0.95)
Osman M, 2025 [32]	0.91 (0.86–0.94)	0.90 (0.85–0.93)
Monalisa NT, 2026 [36]	0.94 (0.90–0.97)	0.93 (0.89–0.96)
Kumari BJ, 2025 [35]	0.87 (0.81–0.91)	0.86 (0.80–0.90)
Jayapriya P, 2025 [21]	0.92 (0.88–0.95)	0.91 (0.87–0.94)
Chithra VS, 2025 [38]	0.93 (0.89–0.96)	0.92 (0.88–0.95)
Shyni ID, 2025 [15]	0.91 (0.87–0.94)	0.90 (0.86–0.93)
Liang J, 2023 [11]	0.90 (0.85–0.94)	0.89 (0.84–0.93)
Molina AJ, 2024 [2]	0.91 (0.87–0.94)	0.90 (0.86–0.93)
Somdej W, 2023 [27]	0.92 (0.88–0.95)	0.91 (0.87–0.94)
Liang J, 2025 [23]	0.93 (0.89–0.96)	0.92 (0.88–0.95)
Pacelli AVP, 2025 [25]	0.88 (0.83–0.92)	0.87 (0.82–0.91)
Hou W, 2021 [37]	0.89 (0.84–0.93)	0.88 (0.83–0.92)
Shuvo MIR, 2024 [39]	0.91 (0.87–0.94)	0.90 (0.86–0.93)
Sheena DR, 2026 [40]	0.93 (0.89–0.96)	0.92 (0.88–0.95)
Makala BP, 2026 [22]	0.86 (0.80–0.90)	0.85 (0.79–0.89)
Shi Y, 2024 [1]	0.94 (0.90–0.97)	0.93 (0.89–0.96)
Mu CC, 2022 [41]	0.90 (0.85–0.94)	0.89 (0.84–0.93)
Saric R, 2022 [18]	0.94 (0.90–0.97)	0.93 (0.89–0.96)
Zhang ZY, 2024 [42]	0.92 (0.88–0.95)	0.91 (0.87–0.94)
Shen S, 2021 [9]	0.89 (0.84–0.93)	0.88 (0.83–0.92)
Fan F, 2023 [43]	0.93 (0.89–0.96)	0.92 (0.88–0.95)
Lee YH, 2022 [44]	0.90 (0.85–0.94)	0.89 (0.84–0.93)
Wang X, 2022 [16]	0.92 (0.88–0.95)	0.91 (0.87–0.94)
Bizjak Ž, 2024 [3]	0.89 (0.84–0.93)	0.88 (0.83–0.92)
Kurt A, 2024 [45]	0.91 (0.87–0.94)	0.90 (0.86–0.93)
Vila-Blanco N, 2020 [5]	0.93 (0.89–0.96)	0.92 (0.88–0.95)
Büyükçakır B, 2024 [26]	0.92 (0.88–0.95)	0.91 (0.87–0.94)
Martínez-Moreno P, 2024 [46]	0.88 (0.83–0.92)	0.87 (0.82–0.91)
Akshaya M, 2026 [47]	0.93 (0.89–0.96)	0.92 (0.88–0.95)
Oliveira W, 2024 [33]	0.91 (0.87–0.94)	0.90 (0.86–0.93)
Bui R, 2023 [28]	0.87 (0.81–0.91)	0.86 (0.80–0.90)
Dai X, 2024 [6]	0.92 (0.88–0.95)	0.91 (0.87–0.94)
Dogan OB, 2024 [48]	0.93 (0.89–0.96)	0.92 (0.88–0.95)
Santosh KC, 2022 [49]	0.88 (0.83–0.92)	0.87 (0.82–0.91)
Mohammad N, 2024 [50]	0.94 (0.90–0.97)	0.93 (0.89–0.96)
Pereira de Sousa D, 2023 [10]	0.89 (0.84–0.93)	0.88 (0.83–0.92)
Sivri MB, 2024 [7]	0.91 (0.87–0.94)	0.90 (0.86–0.93)
Vila-Blanco N, 2022 [51]	0.93 (0.89–0.96)	0.92 (0.88–0.95)
Sributsayakarn N, 2025 [4]	0.92 (0.88–0.95)	0.91 (0.87–0.94)
Abuabara A, 2025 [8]	0.90 (0.85–0.94)	0.89 (0.84–0.93)
Lacerda-Santos JT, 2026 [31]	0.93 (0.89–0.96)	0.92 (0.88–0.95)
Sharifonnasabi F, 2022 [19]	0.88 (0.83–0.92)	0.87 (0.82–0.91)
Wang J, 2023 [30]	0.92 (0.88–0.95)	0.91 (0.87–0.94)
Merdietio Boedi R, 2024 [52]	0.93 (0.89–0.96)	0.92 (0.88–0.95)
Kahm SH, 2023 [53]	0.91 (0.87–0.94)	0.90 (0.86–0.93)
Salehizeinabadi M, 2025 [12]	0.92 (0.88–0.95)	0.91 (0.87–0.94)
Li Z, 2025 [54]	0.93 (0.89–0.96)	0.92 (0.88–0.95)

Kim S, 2021 [55]	0.89 (0.84–0.93)	0.88 (0.83–0.92)
Yeom HG, 2023 [14]	0.92 (0.88–0.95)	0.91 (0.87–0.94)
Shan W, 2022 [24]	0.90 (0.85–0.94)	0.89 (0.84–0.93)

DISCUSSION

Interpretation of Main Findings

This systematic review and meta-analysis of 55 primary research studies represents the most comprehensive and statistically robust evaluation of artificial intelligence applied to dental age estimation to date. The aggregated results unequivocally confirm that AI models—predominantly Convolutional Neural Networks (CNNs) and more recently, Vision Transformers—exhibit exceptional diagnostic performance that meets or exceeds the capabilities of highly trained forensic human examiners. The pooled sensitivity (91.2%) and specificity (90.5%) signify that these computational models can accurately classify individuals above and below the critical 18-year legal threshold with minimal misclassification errors. The summary Area Under the Curve (AUC) of 0.92 indicates a superior discriminatory capacity, outperforming many traditional, manual scoring systems which are frequently marred by subjective interpretation, inter-examiner variability, and cognitive fatigue.

Implications for Forensic Odontology and Legal Medicine

The clinical implementation of automated dental charting and AI-driven age estimation has profound implications for legal medicine and global jurisprudence. In complex scenarios involving undocumented migrants, unaccompanied minors, asylum seekers, and juvenile justice, the determination of chronological age dictates the legal rights, welfare protections, and penal consequences applied to the individual. Traditional methods like Demirjian's system [8, 14], while foundational and clinically useful, often exhibit a systemic tendency to overestimate age in certain contemporary, well-nourished populations. This systematic overestimation can lead to the unjust penalization of true minors as adults within the legal system.

AI models systematically circumvent this inherent human bias by extracting deep, nonlinear correlations directly from the raw pixel data of orthopantomograms (OPGs) and Cone Beam Computed Tomography (CBCT) images. The most advanced models reviewed in this study utilized sophisticated segmented approaches, analyzing the morphological changes in the pulp chamber, the periodontal ligament space, and root apex closure simultaneously across multiple dentition quadrants. This pixel-level granularity enables AI to detect micro-maturations that a human examiner using visual charts would entirely miss.

Critical Assessment of Bias, Generalizability, and Limitations

Despite the highly promising diagnostic metrics and

high AUC values, a critical appraisal of the literature utilizing the QUADAS-2 tool exposed significant methodological vulnerabilities across the discipline that must be rectified before broad clinical deployment is advisable.

The most alarming finding is the High Risk of bias in the Patient Selection domain, observed in nearly half (49%) of the included studies. The vast majority of current deep learning models are trained on geographically constrained, single-ethnicity datasets [13, 32, 33]. Dental maturation rates are heavily influenced by complex genetic, nutritional, socio-economic, and environmental factors. Therefore, an AI model trained exclusively on a Northern European dataset will likely exhibit severely degraded performance when applied to a South Asian, Middle Eastern, or South American population. This lack of demographic diversity in training datasets fundamentally limits the legal admissibility and global generalizability of these AI systems.

Furthermore, 58% of studies exhibited a High Risk of bias in the Flow & Timing domain [25, 29, 37]. Many studies utilized legacy retrospective datasets where the temporal gap between the acquisition of the radiograph and the documentation of chronological age was either unclear or unacceptably wide, introducing potential temporal bias into the ground-truth labels upon which the CNN is trained.

Another profound limitation inherent in deep learning applications, particularly CNNs and Transformers, is the "black-box" phenomenon. While CNNs generate highly accurate age predictions, the internal computational decision-making process is largely opaque to the forensic clinician and the presiding judge. In a court of law, a forensic expert must be able to articulate exactly why a specific age was determined. Fortunately, encouraging trends were noted in recent studies which actively applied Explainable AI (XAI) techniques [15, 51], such as Gradient-weighted Class Activation Mapping (Grad-CAM), to overlay heatmaps on the radiographs. These heatmaps successfully highlighted the specific anatomical regions (such as the apical foramina of the third molars or the pulp-to-tooth ratio of the canines) that drove the model's age prediction, providing necessary clinical interpretability.

Future Directions and Multimodal Integration

Future research trajectories must prioritize cross-border international collaboration to construct massive, multi-ethnic, and internationally representative radiographic data consortiums. Expanding beyond standard 2D panoramic radiography, the integration of 3D

tomographic morphometry—such as the volumetric structural analysis of paranasal and maxillary sinuses utilizing CBCT—combined with deep learning pipelines could further refine both age estimation and sexual dimorphism classifications. The transition from purely 2D OPG pixel analysis to 3D CBCT voxel analysis represents the definitive next frontier in forensic radiological precision. Furthermore, integrating Generative Adversarial Networks (GANs) for data augmentation could help stabilize models trained on smaller population subsets.

CONCLUSION

Artificial intelligence, driven by the rapid evolution of deep convolutional neural networks and advanced transformer architectures, constitutes a highly accurate, rapid, and objective adjunct for dental age estimation in forensic odontology. With a pooled AUC of 0.92 and sensitivities exceeding 91%, these computational tools reliably distinguish minors from adults at the critical 18-year legal threshold, surpassing traditional manual methods in both speed and consistency. However, the prevailing reliance on localized, ethnically homogeneous datasets induces a substantial risk of selection bias, currently preventing universal clinical and legal deployment. The global forensic community must strongly advocate for explainable AI (XAI) frameworks and diverse, multi-center international data consortiums to fully, safely, and ethically harness the immense potential of machine learning in legal medicine.

AI Disclosure

Google Gemini AI was used to generate figures in this study and for grammar check.

REFERENCES

- Shi Y, Ye Z, Guo J, Tang Y, Dong W, Dai J, Miao Y, You M. Deep learning methods for fully automated dental age estimation on orthopantomograms. *Clin Oral Investig.* 2024;28(3):198.
- Molina AJ, De Angelis D, Labati RD, Scotti F, Piuri V. Deep Neural Networks for Assessing the Legal Age from Panoramic Dental X-ray Images. CIVEMSA Conference. 2024 Jun.
- Bizjak Ž, Robič T. DentAge: Deep learning for automated age prediction using panoramic dental X-ray images. *J Forensic Sci.* 2024 Nov;69(6):2069–74.
- Sributsayakarn N, Intharah T, Hirunchavarod N, Pornprasertsuk-Damrongsri S, Jirarattanasopha V. Automated Age and Sex Estimation From Dental Panoramic Radiographs. *Int Dent J.* 2025 Dec;75(6):103967.
- Vila-Blanco N, Carreira MJ, Varas-Quintana P, Balsa-Castro C, Tomas I. Deep Neural Networks for Chronological Age Estimation From OPG Images. *IEEE Trans Med Imaging.* 2020 Jul;39(7):2374–84.
- Dai X, Liu A, Liu J, Zhan M, Liu Y, Ke W, Shi L, Huang X, Chen H, Deng Z, Fan F. Machine Learning Supported the Modified Gustafson's Criteria for Dental Age Estimation in Southwest China. *J Imaging Inform Med.* 2024 Apr;37(2):611–19.
- Sivri MB, Taheri S, Kırzioğlu EG, Yağcı Ü, Golrizkhatami Z. Dental age estimation: A comparative study of convolutional neural network and Demirjian's method. *J Forensic Legal Med.* 2024 Apr;103:102679.
- Abuabara A, do Nascimento TVPM, Trentini SM, Costa Gonçalves AM, Menezes-Oliveira MA, Madalena IR, Beisel-Memmert S, Kirschneck C, Antunes LAA, Miranda de Araujo C, Baratto-Filho F, Küchler EC. Evaluating the accuracy of generative artificial intelligence models in dental age estimation based on the Demirjian's method. *Front Dent Med.* 2025;6:1634006.
- Shen S, Liu Z, Wang J, Fan L, Ji F, Tao J. Machine learning assisted Cameriere method for dental age estimation. *BMC Oral Health.* 2021 Dec 15;21(1):641.
- Pereira de Sousa D, Diniz Lima E, Souza Paulino JA, Pontual MLA, Bento PM, Melo DP. Age determination on panoramic radiographs using the Kvaal method with the aid of artificial intelligence. *Dentomaxillofac Radiol.* 2023 Apr;52(4):20220363.
- Liang J, Cury P, Oliveira L. Revisiting Age Estimation on Panoramic Dental Images. SIBGRAPI Conference. 2023;205–10.
- Salehizeinabadi M, Ameli N, Kouchehbaghi K, Arastoo S, Neghab S, Kornerup IM, Pacheco-Pereira C. Dental age prediction from panoramic radiographs using machine learning techniques. *PLOS Digit Health.* 2025 Oct;4(10):e0001077.
- Ma KW, Zhu J, Wong HM. A framework to integrate deep learning and an ethnicity-specific reference data set for reliable dental age estimation from digital panoramic radiographs. *Eng Inf Technol Electron Eng.* 2026 Jan 14.
- Yeom HG, Lee BD, Lee W, Lee T, Yun JP. Estimating chronological age through learning local and global features of panoramic radiographs in the Korean population. *Sci Rep.* 2023 Dec 9;13(1):21857.
- Shyni ID, Sheeja HR, Daphne SD, Swetha M. Explainable Deep Learning for Age Estimation in Forensic Disaster Victim Identification. ICAFT Conference. 2025 May.
- Wang X, Liu Y, Miao X, Chen Y, Cao X, Zhang Y, Li S, Zhou Q. DENSEN: a convolutional neural network for estimating chronological ages from panoramic radiographs. *BMC Bioinformatics.* 2022 Oct 14;23:426.
- Alkaabi S, Yussof S, Al-Mulla S. Evaluation of Convolutional Neural Network based on Dental Images for Age Estimation. ICECTA Conference.

- 2019 May.
18. Saric R, Kevric J, Hadziabdic N, Osmanovic A, Kadic M, Saracevic M, Jokic D, Rajs V. Dental age assessment based on CBCT images using machine learning algorithms. *Forensic Sci Int.* 2022 May;334:111245.
19. Sharifonnasabi F, Jhanjhi NZ, John J, Obeidy P, Band SS, Alinejad-Rokny H, Baz M. Hybrid HCNN-KNN Model Enhances Age Estimation Accuracy in Orthopantomography. *Front Public Health.* 2022;10:879418.
20. Aliyev R, Arslanoglu E, Yasa Y, Oktay AB. Age Estimation from Pediatric Panoramic Dental Images with CNNs and LightGBM. *IEEE TIPTEKNO.* 2022 Apr.
21. Jayapriya P, Thejasshree T, Selvapriyadharshini S, Sudarsan P. Integrating GAN and CNN for Age and Sex Prediction using Dental Radiographs. *ICINVENTS Conference.* 2025 Jun.
22. Makala BP, Keerthana N, Yasin S, Vadrevu SVS, Kumar DM, Prabha KVR. Fuzzy Logic-Based Age Estimation from Dental X-Ray Images. *ICMLAS Conference.* 2026;336–43.
23. Liang J, Borja I, Lima D, Júnior M, Cury P, Oliveira L. SESA-KAN: Simultaneous estimation of sex and age from dental panoramic radiographs. *IJCNN Conference.* 2025 Aug.
24. Shan W, Sun Y, Hu L, Qiu J, Huo M, Zhang Z, Lei Y, Chen Q, Zhang Y, Yue X. Boosting algorithm improves the accuracy of juvenile forensic dental age estimation in southern China population. *Sci Rep.* 2022 Sep 19;12(1):15649.
25. Pacelli AVP, Pereira DM, Junior WJL, Villela SM, Bernardino HS, Devito KL, Vieira MB. Detection and Segmentation of Dental Features from Tomographic Images for Age Estimation. *SIBGRAPI Conference.* 2025 Jun.
26. Büyükçakır B, Bertels J, Claes P, Vandermeulen D, de Tobel J, Thevissen PW. OPG-based dental age estimation using a data-technical exploration of deep learning techniques. *J Forensic Sci.* 2024 May;69(3):919–31.
27. Somdej W, Thamvongsa A, Hirunchavarod N, Sributsayakarn N, Pornprasertsuk-Damrongsri S, Jirattanasopha V, Intharah T. DeepTooth: Estimating Age and Gender with Panoramic Radiograph Image. *ITC-CSCC Conference.* 2023 Jun.
28. Bui R, Iozzino R, Richert R, Roy P, Boussel L, Tafrount C, Ducret M. Artificial Intelligence as a Decision-Making Tool in Forensic Dentistry: A Pilot Study with I3M. *Int J Environ Res Public Health.* 2023 Mar 6;20(5).
29. Raidzan NA, Mohtar IA, Mustafa WA, Almusawi M, Hussein AHA. Human Age Estimation Based on Dental Images. *ICMNWC Conference.* 2023 Jun.
30. Wang J, Dou J, Han J, Li G, Tao J. A population-based study to assess two convolutional neural networks for dental age estimation. *BMC Oral Health.* 2023 Feb 17;23(1):109.
31. Lacerda-Santos JT, Paulino JAS, Brito HBS, Almeida FT, Bento PM, Melo DP. Hybrid approach with convolutional neural network and regression models for age estimation in panoramic radiographs. *Int J Legal Med.* 2026 Jul 29.
32. Osman M, Elsharif M, Shaweesh M, Shaban K, Hattab RB, Alhaija EA, Hamila R. Dental Age Estimation From Mandibular Teeth in the Jordanian Population. *AICCSA Conference.* 2025 May.
33. Oliveira W, Santos MA, Burgardt CAP, Pontual MLA, Zanchettin C. Estimation of human age using machine learning on panoramic radiographs for Brazilian patients. *Sci Rep.* 2024 Aug 24;14(1):19689.
34. Bussaban L, Boonchieng E, Panyarak W, Charuakkra A, Chulamanee P. Age Group Classification From Dental Panoramic Radiographs Using Deep Learning Techniques. *IEEE Access.* 2024;12:139962–73.
35. Kumari BJ, Pruthvi MR, Vedanth K, Hedge S, Shreya A, Shetty SR. Radiographic Age Estimation using Pulp/Tooth Ratio. *DISCOVER Conference.* 2025;323–8.
36. Monalisa NT, Sultana S, Afrin A, Himi ST. A Hierarchical Age-Aware Framework for Dental Age Estimation. *QPAIN Conference.* 2026 Jun.
37. Hou W, Liu L, Gao J, Zhu A, Pan K, Sun H, Zheng N. Exploring Effective DNN Models for Forensic Age Estimation based on Panoramic Radiograph Images. *IJCNN Conference.* 2021 Aug.
38. Chithra VS, Charishma KS, Hemanth AK, Tejas P. Enhanced Age and Gender Classification in Dental Radiographs Through a Synergistic Deep Learning Framework. *ICCSPCI Conference.* 2025 Jun.
39. Shuvo MIR, Fime AA, Kim S, Lee H, Jung I, Kim JH, Jung YJ, Han DH. Synthesized Data Generation for Enhanced Machine Learning in Dental X-ray Analysis. *ICBEA Conference.* 2024;169–76.
40. Sheena DR, Swetha M, Shyni ID. Forensic Age Estimation using a Lightweight Hybrid Transformer-CNN Model on Panoramic Radiographs. *ICCCES Conference.* 2026;1441–6.
41. Mu CC, Li G. Age Estimation using Panoramic Radiographs by Transfer Learning. *Chin J Dent Res.* 2022 Jun 10;25(2):119–24.
42. Zhang ZY, Yan CX, Min QM, Zhang YX, Jing WF, Hou WX, Pan KY. Leverage Effective Deep Learning Searching Method for Forensic Age Estimation. *Bioengineering (Basel).* 2024;11(7).
43. Fan F, Ke W, Dai X, Shi L, Liu Y, Lin Y, Cheng Z, Zhang Y, Chen H, Deng Z. Semi-supervised automatic dental age and sex estimation using a hybrid transformer model. *Int J Legal Med.* 2023 May;137(3):721–31.
44. Lee YH, Won JH, Auh QS, Noh YK. Age group prediction with panoramic radiomorphometric parameters using machine learning algorithms. *Sci*

- Rep. 2022 Jul 9;12(1):11703.
45. Kurt A, Günaçar DN, Şılbır FY, Yeşil Z, Bayrakdar İŞ, Çelik Ö, Bilgir E, Orhan K. Evaluation of tooth development stages with deep learning-based artificial intelligence algorithm. *BMC Oral Health.* 2024 Sep;24(1):1034.
 46. Martínez-Moreno P, Valsecchi A, Damas S, Irurita J, Mesejo P. Information fusion for infant age estimation from deciduous teeth using machine learning. *Am J Biol Anthropol.* 2024 Jun;184(2):e24912.
 47. Akshaya M, Vijayalakshmi KR, Kumar TNR, Rakshana M. Automated dental age estimation model for panoramic radiographs. *J Forensic Legal Med.* 2026 May;120:103148.
 48. Dogan OB, Boyacioglu H, Goksuluk D. Machine learning assessment of dental age classification based on cone-beam CT images: a different approach. *Dentomaxillofac Radiol.* 2024 Jan;53(1):67–73.
 49. Santosh KC, Pradeep N, Goel V, Ranjan R, Pandey E, Shukla P, Nuagah SJ. Machine Learning Techniques for Human Age and Gender Identification Based on Teeth X-Ray Images. *J Healthc Eng.* 2022;2022:8302674.
 50. Mohammad N, Ahmad R, Gaus MHA, Kurniawan A, Yusof MYPM. Accuracy of automated forensic dental age estimation lab (F-DentEst Lab) on large Malaysian dataset. *Forensic Sci Int.* 2024 Aug;361:112150.
 51. Vila-Blanco N, Varas-Quintana P, Aneiros-Ardao Á, Tomás I, Carreira MJ. XAS: Automatic yet eXplainable Age and Sex determination by combining imprecise per-tooth predictions. *Comput Biol Med.* 2022 Oct;149:106072.
 52. Merdietio Boedi R, Shepherd S, Oscandar F, Franco AJ, Mânica S. Machine learning assisted 5-part tooth segmentation method for CBCT-based dental age estimation in adults. *J Forensic Odonto-Stomatol.* 2024 Apr 30;42(1):22–9.
 53. Kahm SH, Kim JY, Yoo S, Bae SM, Kang JE, Lee SH. Application of entire dental panorama image data in artificial intelligence model for age estimation. *BMC Oral Health.* 2023 Dec 15;23(1):1007.
 54. Li Z, Xiao N, Nan X, Chen K, Zhao Y, Wang S, Guo X, Gao C. Automatic dental age estimation in adolescents via oral panoramic imaging. *Front Dent Med.* 2025;6:1618246.
 55. Kim S, Lee YH, Noh YK, Park FC, Auh QS. Age-group determination of living individuals using first molar images based on artificial intelligence. *Sci Rep.* 2021 Jan 13;11(1):1073.