IOSR Journal of Dental and Medical Sciences (IOSR-JDMS)
e-ISSN: 2279-0853, p-ISSN: 2279-0861.Volume 25, Issue 9 Ser. 2 (September. 2026), PP 62-68
www.iosrjournals.org

Artificial Intelligence in Automated Cephalometric Analysis and Orthodontic Diagnosis: Accuracy, Validation, and Clinical Integration — A Narrative Review

Adrielen Henrique Picoli[1]
(Specialist in Orthodontics – FAIPE, Cuiabá, MT, Brazil)
ORCID: 0009-0009-5864-171X
Corresponding Author`s E-mail: paviolba75@gmail.com

Carolina dos Santos Alcantara Oliveira[2]
(Specialist in Orthodontics, IOGP - Instituto Odontológico de Pós-Graduação, Bauru, SP, Brazil)
ORCID: 0009-0006-4214-5955

Hannah Kamilla Azevedo da Silva [3]
(Master`s Degree in Integrated Clinical Dentistrys – Federal University of Pernambuco (UFPE), Recife, PE, Brazil) ORCID: 0000-0002-0543-9210

Gina Maria da Silva[4]
(Specialist in Functional Orthopedics of the Jaws, IVO Institute, São Paulo, SP, Brazil)
ORCID: 0009-0000-7494-3704

Paula Cristina da Costa Alcantara[5]
(Master`s Degree in Pediatric Dentistry – UFRJ; PhD Student in Orthodontics – UERJ, Rio de Janeiro, RJ, Brazil. ORCID: 0009-0001-8288-6479

Noemi Borgas de Góes Cavalcanti[6]
(Specialist in Oral Maxillofacial Surgery And Traumatology – Universidade Sete Lagoas (ABO/FACSETE), Sete Lagoas, MG, Brazil.
ORCID: 0000-0002-9420-4519

Danielle Santos Quindos[7]
(Specialist in Orthodontics – Faculdade Unidas do Norte de Minas/FUNORTE; Specialist in Pediatric Dentistry, - Faculdade Conectada/FACONNECT; Specialist in Patients with Disabilities – Universidade de São Paulo (USP), São Paulo, SP, Brazil)
ORCID: 0009-0004-3159-6182

Aline Marques de Castro Chagas[8]
(Specialist in Orthodontics – Faculdade Unidas do Norte de Minas (FUNORTE) – SOEBRAS – Instituto de Ciências da Saúde – ICS, Manaus, AM, Brazil)
ORCID: 0009-0004-5379-7981

Paulo Vinicius de Oliveira Bastos (Supervisor)
(Dentist, Specialist in Family Health & Specialist in Public Health – UFRJ; Specialist in Periodontology, Specialist in Implantolog, and Specialist in TMD/Orofacial Pain Odontoclínica Central do Exército (OCEx), Rio de Janeiro, RJ; Specialist in Orthodontics – ABO, Master in Orthodontics – UFJF;; INBDE-certified (JCNDE, 2025); RJ, Brazil.
ORCID: 0009-0008-2449-3010. Corresponding Author`s E-mail: paviolba75@gmail.com)

Abstract:
Background: *Cephalometric analysis is a cornerstone of orthodontic diagnosis, yet manual landmark identification is time-consuming and subject to intra- and inter-examiner variability. Over the past decade, deep*

learning (DL) — particularly convolutional neural networks (CNNs) — has been applied to automate landmark detection on two-dimensional (2D) lateral cephalograms and three-dimensional (3D) cone-beam computed tomography (CBCT), and, more recently, to support downstream diagnostic decisions such as skeletal classification and extraction planning.

Objective: *To synthesize the current evidence on the accuracy, validation, and clinical integration of artificial intelligence (AI) in automated cephalometric analysis and orthodontic diagnosis, and to identify the conditions under which these tools can be safely incorporated into clinical workflows.*

Methods: *Narrative review of systematic reviews, meta-analyses, scoping reviews, diagnostic accuracy studies, and validation studies of commercial software published between 2016 and 2026, retrieved from PubMed/MEDLINE and complementary databases.*

Results: *Pooled evidence indicates that DL-based landmark detection on 2D cephalograms achieves mean errors clustered around the 2-mm clinical threshold, with approximately 80% of landmarks detected within 2 mm; 3D landmarking on CBCT shows higher and more heterogeneous errors, although accuracy has improved with each publication year. AI models for sagittal skeletal classification and extraction decision support reach accuracies above 90% in single-center datasets. However, independent validation of commercial platforms reveals statistically significant — and in CBCT-based systems clinically relevant — systematic bias in core sagittal angles, arguing against standalone use.*

Conclusion: *AI-driven cephalometry is accurate and time-efficient enough to function as a decision-support tool under expert supervision, particularly in semi-automatic workflows. Generalizability across imaging devices and populations, transparent reporting, and external validation of commercial systems remain the main barriers to fully autonomous clinical integration.*

Key Word: *Artificial intelligence; Cephalometry; Deep learning; Orthodontic diagnosis; Convolutional neural networks; Cone-beam computed tomography; Clinical decision support.*

Date of Submission: 04-09-2026 Date of Acceptance: 14-09-2026

I. Introduction

Since Broadbent introduced the standardized lateral cephalogram in 1931, cephalometric analysis has remained one of the most widely used diagnostic instruments in orthodontics and dentofacial orthopedics. Its clinical value rests on the reproducible identification of anatomical landmarks from which angular and linear measurements are derived, allowing clinicians to classify skeletal and dental relationships, estimate growth patterns, plan treatment mechanics, and evaluate outcomes. Yet the reliability of the method has always been constrained by a human bottleneck: landmark identification is time-consuming, requires extensive training, and is subject to substantial intra- and inter-examiner variability, particularly for landmarks located on curved or low-contrast structures such as gonion, orbitale, and porion.

The application of artificial intelligence (AI) — and, more specifically, of machine learning (ML) and deep learning (DL) — to orthodontic imaging has expanded rapidly in the last decade. A scoping review of the literature up to 2020 identified 62 studies on AI and ML in orthodontics, of which more than two-thirds had been published within the preceding ten years [1]. A parallel systematic review documented the scope and performance of AI-based models used for orthodontic diagnosis, treatment planning, and prognosis, concluding that most applications relied on artificial neural networks and convolutional neural networks (CNNs) [2]. The most recent scoping review, covering 71 studies, classified AI applications into three domains — diagnostics (n = 29), landmark identification (n = 20), and treatment planning (n = 22) — and found that anatomical landmark detection remains the most extensively studied area, while emphasizing that human supervision continues to be essential [3].

Automated cephalometric landmark detection was among the first orthodontic tasks addressed by DL, and the field has now matured sufficiently to be summarized by several systematic reviews and meta-analyses [4]. In parallel, a second wave of research has moved “beyond landmarks,” using AI to classify sagittal skeletal patterns directly from images, to predict the need for extractions or orthognathic surgery, and to support treatment planning. A third and more recent stream has begun to validate commercially available AI cephalometric platforms against expert manual analysis under real clinical conditions — with results that are less uniformly optimistic than the algorithm-development literature.

The present narrative review synthesizes these three streams of evidence. It addresses four questions: (1) how accurate is AI-driven landmark detection on 2D and 3D images; (2) how far has AI progressed from landmarking to diagnostic decision support; (3) how reliable are commercial AI cephalometric systems when validated independently; and (4) what conditions must be met for safe clinical integration. The review also situates these findings within the broader context of public oral health systems, where diagnostic efficiency and access to specialist expertise are unevenly distributed.

II. Methodology

This narrative review was based on a structured, non-systematic search of PubMed/MEDLINE, Scopus, and Google Scholar, complemented by manual screening of reference lists, covering publications from January 2016 to September 2026. Search terms combined "artificial intelligence," "deep learning," "convolutional neural network," and "machine learning" with "cephalometric," "cephalometry," "landmark detection," "orthodontic diagnosis," "skeletal classification," and "treatment planning." Priority was given to systematic reviews and meta-analyses, scoping reviews, diagnostic accuracy studies with expert-based reference standards, and independent validation studies of commercial software. Only studies indexed in peer-reviewed journals and with verifiable bibliographic metadata (DOI or PubMed identifier) were retained. Twenty references were selected for citation. Given the narrative design, no formal risk-of-bias assessment or quantitative pooling was performed; where pooled estimates are reported, they are those published by the cited meta-analyses.

III. From Manual Tracing to Automated Landmarking: Technical Foundations

A. Evolution of automated landmark detection

Early attempts at computer-assisted landmarking relied on knowledge-based rules, edge detection, pixel-intensity thresholds, and template matching; their performance was inconsistent and highly sensitive to image quality [5]. The decisive shift came with the adoption of CNNs, which learn hierarchical image features directly from annotated training data rather than from hand-crafted rules. Comparative studies within the systematic literature consistently show that DL-based methods outperform earlier knowledge-based, atlas-based, and shallow machine learning approaches [5]. Contemporary architectures include heatmap-regression CNNs, cascaded or two-stage CNNs that first localize a region of interest and then refine the landmark position, single-shot detectors such as YOLO (You Only Look Once), Bayesian CNNs that produce confidence regions around each prediction, and, most recently, vision transformers.

B. Fully automated analysis on 2D lateral cephalograms

An early and influential example of a complete DL pipeline was the customized CNN developed by Kunz et al., which was trained to perform a fully automated cephalometric analysis on lateral cephalograms and evaluated against a panel of experienced examiners. The authors reported that the algorithm's landmark positions and derived measurements were within the range of variability observed among human experts for most parameters, which established the plausibility of fully automated 2D analysis [6]. Subsequent work addressed generalizability, a recurrent weakness of single-center models. Kim et al. trained cascade CNNs on lateral cephalograms collected from multiple centers nationwide, deliberately exposing the model to variation in devices, exposure settings, and patient characteristics, and reported clinically acceptable accuracy for automated identification of lateral cephalometric landmarks across this heterogeneous sample [7].

C. Three-dimensional landmarking on CT and CBCT

The extension of DL landmarking to 3D volumetric data is technically more demanding: the search space increases by an order of magnitude, annotated 3D datasets are scarce, and clinical CBCT/CT images frequently contain metal artifacts from restorations and appliances. Dot et al. demonstrated automatic 3D cephalometric landmarking via DL on a clinically relevant test dataset randomly drawn from presurgical CT scans, of which more than 90% contained metal artifacts, and reported accuracy compatible with clinical use in semi-automatic workflows [8]. A meta-analysis restricted to 3D automated landmarking found an overall mean error of 2.44 mm with very high heterogeneity between studies, but also a statistically significant improvement in accuracy with each successive year of publication, indicating rapid algorithmic progress [9].

IV. Diagnostic Accuracy of AI Landmark Detection: Evidence from Systematic Reviews and Meta-Analyses

The most rigorous synthesis of the first generation of DL landmarking studies is the meta-analysis by Schwendicke et al., which included 19 diagnostic accuracy studies published between 2017 and 2020, all employing CNNs and mostly using 2D lateral radiographs, with a mean of 30 landmarks tested per study. Landmark prediction error centered around the 2-mm threshold conventionally accepted as clinically tolerable, and approximately 80% of landmarks were detected within 2 mm. The authors nevertheless judged the risk of bias to be high, principally because of data selection and reference-standard conduct, and cautioned that the evidence was of limited generalizability and robustness [4].

More recent reviews have confirmed and refined these estimates. A 2022 systematic review following PRISMA-DTA concluded that AI identifies cephalometric landmarks with accuracy comparable to human examiners and superior to earlier machine learning methods [5]. A 2023 meta-analysis reported that automated methods and expert examiners did not differ significantly for most landmarks and measurements, while noting persistent heterogeneity in study design and reference standards [12]. A systematic review of algorithms published between 2013 and 2023 found that CNN-based systems achieved point-localization accuracy ranging from 64.3% to 97.3%, with a mean error of approximately 1.04 mm, although six of the thirteen included studies were judged to be at high risk of bias [11].

The meta-analysis by Hendrickx et al., with a search extending to January 2024, is the first to pool 2D and 3D evidence side by side. AI-driven landmark detection on 2D cephalograms and 3D CBCT images exhibited high accuracy and marked time efficiency, with the majority of 2D studies indicating superior automated performance; however, the error rates reported by 3D studies were inconsistent, and the authors advised clinicians to remain vigilant regarding the risk of inaccurate landmark identification and stressed that generalizability and robustness still require improvement [10].

Head-to-head diagnostic studies illustrate how these pooled estimates translate to individual algorithms. Bulatova et al. compared a YOLOv3-based commercial system with manual tracing on 110 cephalograms from the American Association of Orthodontists Foundation Legacy collection, evaluating 16 landmarks; agreement was clinically acceptable for most points, but accuracy varied by landmark, reflecting the well-known difficulty of structures with poorly defined edges [13]. Jeon and Lee compared measurements generated by an automatic CNN-based program with those of a conventional cephalometric program in 35 patients and found that most angular and linear parameters did not differ to a clinically meaningful degree, although selected measurements dependent on hard-to-locate landmarks showed larger discrepancies [14]. Taken together, the accuracy literature supports a consistent conclusion: DL landmarking is at, or very near, expert level for 2D images on average, but performance is landmark-specific, dataset-dependent, and less mature in 3D.

V. Beyond Landmarks: AI in Orthodontic Diagnosis and Decision Support

A. Automated skeletal classification

A conceptually different approach bypasses explicit landmarking altogether and trains a network to output a diagnostic category directly from the image. Yu et al. developed an AI system for automated sagittal and vertical skeletal classification from lateral cephalograms and reported classification accuracy exceeding 90% when compared with expert diagnosis, establishing end-to-end image-to-diagnosis as a viable paradigm [15]. Kim et al. subsequently trained a deep CNN on 1,574 cephalometric images classified by ANB angle into Class I, II, and III, and compared its performance with an automated-tracing AI software that derives the classification from three detected landmarks; the direct DCNN classifier produced an immediate diagnosis and performed at least as well as the landmark-based pipeline, suggesting that the two strategies may be complementary [16].

B. Extraction and treatment-planning decisions

The use of neural networks for treatment decisions predates the DL era. Jung and Kim trained a neural network on cephalometric and model measurements to determine whether a patient required extraction and, if so, which extraction pattern was indicated, reporting high agreement with expert orthodontists and thereby demonstrating that AI could support one of the most consequential decisions in orthodontic planning [17]. The scoping review by Gracea et al. identified 22 treatment-planning studies covering extraction decisions, orthognathic surgery need, growth prediction, and soft-tissue outcome prediction; the authors concluded that AI shows potential in improving time efficiency and reducing operator variability, but that its accuracy and reliability have not yet consistently surpassed those of expert clinicians, so that human supervision remains essential at all moments [3].

VI. Independent Validation of Commercial AI Cephalometric Platforms

A critical distinction must be drawn between algorithm-development studies, in which a model is evaluated on a held-out subset of the same dataset used for training, and independent validation of commercial products in a clinical setting. The latter is far scarcer and considerably more sobering. Mercier et al. analyzed 408 lateral cephalograms with three methods — manual landmark localization, fully automatic AI localization, and semi-automatic AI localization with clinician correction — across 15 skeletal, dental, and soft-tissue variables. Statistically significant differences in landmark positioning were detected among the three techniques, although almost all were not clinically significant; the semi-automatic option was more accurate than the fully automatic one and faster than conventional tracing, whereas the fully automatic version tended to systematically over- or underestimate specific measurements [18].

The most demanding test to date concerns AI cephalometry performed on CBCT volumes. Kazimierczak et al. evaluated two commercial platforms in 135 patients who had undergone concurrent CBCT and lateral cephalometric imaging, using manual 2D Steiner and Ricketts analyses as the reference standard. Both systems showed systematic bias for key skeletal and dental measures: for one platform, SNA differed by +3.40°, SNB by +5.88°, and ANB by −2.48°, with 95% confidence intervals for sentinel angles commonly exceeding ±2°. The authors concluded that, in their current form, the evaluated AI CBCT cephalometry tools do not match manual 2D cephalometry closely enough for standalone clinical use, because differences of 3–6° in core sagittal angles are large enough to change diagnosis and treatment plans [19]. This finding does not contradict the favorable meta-analytic evidence on 2D landmarking; rather, it exposes the gap between landmark-level accuracy reported in research settings and measurement-level reliability of proprietary systems whose training data, architectures, and validation procedures are undisclosed.

VII. Clinical Integration: Workflow, Time Efficiency, Education, and Human Oversight

A. Time efficiency and workflow

The most robust and consistent advantage of AI cephalometry is speed. Across the systematic literature, automated landmarking is completed in seconds rather than the 10–20 minutes typically required for manual digital tracing, and the meta-analysis by Hendrickx et al. identified time efficiency as a benefit common to both 2D and 3D systems [10]. In the diagnostic study by Mercier et al., the fully automatic mode was the fastest, followed by the semi-automatic mode, with conventional digital tracing the slowest, and the semi-automatic mode combined this speed advantage with the highest accuracy among the AI options [18]. This favors a workflow in which AI produces an initial landmark set that the clinician reviews and corrects — a model that preserves the efficiency gain while retaining professional accountability.

B. Human oversight as a design requirement

Every systematic review and scoping review included in this synthesis, from 2021 to 2025, converges on the same recommendation: AI cephalometry should be implemented as a decision-support tool under expert supervision, not as an autonomous diagnostic system [3,4,10]. The commercial validation studies transform this from a precautionary principle into an empirical necessity, since systematic biases of several degrees in SNA, SNB, and ANB would be invisible to a clinician who did not verify the output [19]. Landmark-specific verification is particularly important for points with historically low success detection rates, such as gonion, orbitale, porion, and the soft-tissue profile.

C. Education and expertise

AI landmarking offers educational value beyond clinical efficiency. For orthodontic residents and general dentists, automated tracing with confidence estimates can serve as an immediate reference against which manual identification is compared, accelerating skill acquisition. Conversely, over-reliance on automated outputs risks the erosion of landmarking competence, which remains essential precisely because supervision of AI presupposes the ability to recognize its errors. The systematic review by Bichu et al. noted that translation of AI into clinical practice has lagged behind the volume of publications, in part because of such workforce and training considerations [1].

VIII. Limitations of the Current Evidence and Research Gaps

Several limitations recur across the literature. First, most algorithm-development studies use a single imaging device, a single institution, and a single population, so that reported accuracy overestimates real-world performance; multi-center, multi-device, and multi-ethnic datasets such as those used by Kim et al. and Dot et al. remain the exception [7,8]. Second, reference standards are inconsistent: some studies rely on one annotator, others on the consensus of several, and few report the intra- and inter-examiner reliability of the reference itself, even though this ceiling limits the maximum achievable accuracy of any model [4]. Third, risk of bias is high in the majority of included studies according to QUADAS-2 assessments [4,11]. Fourth, commercial systems are rarely subjected to external validation, and when they are, systematic biases emerge that are absent from vendor-reported figures [18,19]. Fifth, 3D landmarking on CBCT remains markedly less mature than 2D, with mean errors above 2 mm and inconsistent results across studies [9,10]. Finally, outcome measures are almost exclusively technical (mean radial error, success detection rate); no study has yet demonstrated that AI cephalometry improves treatment decisions, treatment outcomes, or patient-reported outcomes — the clinical endpoints that ultimately justify adoption.

Priority research directions therefore include: prospective, multi-center validation studies of commercial platforms with pre-registered protocols; standardized reporting of AI diagnostic accuracy studies in dentistry; open benchmark datasets that reflect ethnic, age, and device diversity; uncertainty quantification that flags low-confidence landmarks for mandatory review; and, above all, trials linking AI-assisted diagnosis to treatment decisions and outcomes.

IX. National Interest and Significance

The relevance of AI-assisted cephalometry extends beyond the specialist practice. In large public and mixed health systems — including the United States, where orthodontic specialists are concentrated in urban areas and Medicaid orthodontic eligibility depends on standardized severity indices, and Brazil, where the Unified Health System (SUS) delivers primary oral health care through the Family Health Strategy with limited access to specialist referral — a substantial share of malocclusion screening is performed by general dentists without immediate access to an orthodontist. In these settings, an accurate, supervised AI cephalometric tool can standardize the initial skeletal assessment, reduce the variability of eligibility determinations, shorten the time between screening and specialist referral, and make interceptive treatment windows in growing patients more likely to be captured.

The value of capturing these interceptive windows has been quantified in a complementary narrative review published in this journal by Temponi et al., which examined the use of AI for the prediction of skeletal growth and the optimization of orthopedic and orthodontic treatment timing [20]. Reviewing literature from 1982 to 2026, the authors reported that deep-learning models achieve pooled accuracies of approximately 80–95% for staging skeletal maturation from cervical vertebral maturation on lateral cephalograms and from hand-wrist radiographs, frequently matching or exceeding inter-observer agreement among clinicians, and highlighted that sex acts as both a biological and an algorithmic modifier of performance, with several systems showing lower accuracy in male patients during the pubertal window most relevant to treatment timing [20]. Of particular relevance to the present discussion, Temponi et al. framed timing accuracy as a cost variable within the U.S. fee-for-service dental economy: a missed or mistimed growth window may convert an interceptive-treatment candidate (approximately US$2,000–4,500) into a combined orthodontic-orthognathic surgical candidate (approximately US$20,000–40,000), a differential borne by families and public payers; they further argued that, in Texas — where a large Medicaid-enrolled pediatric population coincides with extensive dental Health Professional Shortage Area designations — AI-based growth staging derived from the single cephalogram already obtained for routine diagnosis could offer disproportionate value where serial specialist monitoring is least feasible, and could improve the reproducibility of documentation submitted for Medicaid prior authorization under the Handicapping Labio-Lingual Deviation Index [20]. Automated landmark detection, the subject of the present review, is the technical foundation on which such growth-stage prediction rests, since the same cephalometric image and landmark set feed both the diagnostic and the timing algorithms; the two lines of evidence therefore converge on a single conclusion, namely that supervised AI cephalometry can extend specialist-level assessment to access-constrained, publicly funded settings, provided that sex-stratified calibration and independent validation precede deployment [20].

The same evidence base also defines the boundaries of responsible deployment. The demonstrated systematic biases of commercial CBCT-based systems [19] imply that public payers and health authorities should require independent validation before AI outputs are accepted as the basis for coverage decisions or treatment authorization. From a regulatory standpoint, the transition of AI cephalometry from research tool to medical device demands the same standards of external validation, transparency of training data, and post-market surveillance that apply to other diagnostic technologies. Investment in these safeguards is in the national interest of any health system seeking to expand equitable access to orthodontic diagnosis without sacrificing diagnostic quality.

X. Conclusion

Artificial intelligence has transformed cephalometric analysis from a manual, examiner-dependent procedure into a task that can be automated in seconds with accuracy approaching expert level on 2D lateral cephalograms. Pooled evidence places DL landmark error at or near the 2-mm clinical threshold, and end-to-end models for skeletal classification and extraction decision support achieve accuracies above 90% in controlled datasets. Nevertheless, three caveats define the current state of the art: accuracy is landmark-specific and dataset-dependent; 3D landmarking on CBCT remains less mature and more heterogeneous; and independent validation of commercial platforms reveals systematic biases — clinically relevant in CBCT-based systems — that vendor-reported figures do not disclose. The evidence therefore supports the clinical integration of AI cephalometry as a supervised decision-support tool, preferably in semi-automatic workflows in which the clinician reviews and corrects automated landmarks. Fully autonomous diagnostic use is not yet justified. Closing this gap will require multi-center external validation, transparent reporting, uncertainty-aware software design, and studies that connect AI-assisted diagnosis to treatment decisions and outcomes.

References

[1]. Bichu YM, Hansa I, Bichu AY, Premjani P, Flores-Mir C, Vaid NR. Applications of artificial intelligence and machine learning in orthodontics: a scoping review. Prog Orthod. 2021;22(1):18. doi:10.1186/s40510-021-00361-9

[2]. Khanagar SB, Al-Ehaideb A, Vishwanathaiah S, Maganur PC, Patil S, Naik S, Baeshen HA, Sarode SS. Scope and performance of artificial intelligence technology in orthodontic diagnosis, treatment planning, and clinical decision-making — a systematic review. J Dent Sci. 2021;16(1):482-492. doi:10.1016/j.jds.2020.05.022

[3]. Gracea RS, Winderickx N, Vanheers M, Hendrickx J, Preda F, Shujaat S, Cadenas de Llano-Pérula M, Jacobs R. Artificial intelligence for orthodontic diagnosis and treatment planning: a scoping review. J Dent. 2025;152:105442. doi:10.1016/j.jdent.2024.105442

[4]. Schwendicke F, Chaurasia A, Arsiwala L, Lee JH, Elhennawy K, Jost-Brinkmann PG, Demarco F, Krois J. Deep learning for cephalometric landmark detection: systematic review and meta-analysis. Clin Oral Investig. 2021;25(7):4299-4309. doi:10.1007/s00784-021-03990-w

[5]. Junaid N, Khan N, Ahmed N, Abbasi MS, Das G, Maqsood A, Ahmed AR, Marya A, Alam MK, Heboyan A. Development, application, and performance of artificial intelligence in cephalometric landmark identification and diagnosis: a systematic review. Healthcare (Basel). 2022;10(12):2454. doi:10.3390/healthcare10122454
[6]. Kunz F, Stellzig-Eisenhauer A, Zeman F, Boldt J. Artificial intelligence in orthodontics: evaluation of a fully automated cephalometric analysis using a customized convolutional neural network. J Orofac Orthop. 2020;81(1):52-68. doi:10.1007/s00056-019-00203-8
[7]. Kim J, Kim I, Kim YJ, Kim M, Cho JH, Hong M, Kang KH, Lim SH, Kim SJ, Kim YH, et al. Accuracy of automated identification of lateral cephalometric landmarks using cascade convolutional neural networks on lateral cephalograms from nationwide multi-centres. Orthod Craniofac Res. 2021;24(Suppl 2):59-67.
[8]. Dot G, Schouman T, Chang S, Rafflenbeul F, Kerbrat A, Rouch P, Gajny L. Automatic 3-dimensional cephalometric landmarking via deep learning. J Dent Res. 2022;101(11):1380-1387. doi:10.1177/00220345221112333
[9]. Serafin M, Baldini B, Cabitza F, Carrafiello G, Baselli G, Del Fabbro M, Sforza C, Caprioglio A, Tartaglia GM. Accuracy of automated 3D cephalometric landmarks by deep learning algorithms: systematic review and meta-analysis. Radiol Med. 2023;128(5):544-555. doi:10.1007/s11547-023-01629-2
[10]. Hendrickx J, Gracea RS, Vanheers M, Winderickx N, Preda F, Shujaat S, Jacobs R. Can artificial intelligence-driven cephalometric analysis replace manual tracing? A systematic review and meta-analysis. Eur J Orthod. 2024;46(4):cjae029. doi:10.1093/ejo/cjae029
[11]. Ribas-Sabartés J, Sánchez-Molins M, d'Oliveira NG. The accuracy of algorithms used by artificial intelligence in cephalometric points detection: a systematic review. Bioengineering (Basel). 2024;11(12):1286. doi:10.3390/bioengineering11121286
[12]. de Queiroz Tavares Borges Mesquita G, Vieira WA, Vidigal MTC, Travençolo BAN, Beaini TL, Spin-Neto R, Paranhos LR, de Brito Júnior RB. Artificial intelligence for detecting cephalometric landmarks: a systematic review and meta-analysis. J Digit Imaging. 2023;36(3):1158-1179. doi:10.1007/s10278-022-00766-w
[13]. Bulatova G, Kusnoto B, Grace V, Tsay TP, Avenetti DM, Sanchez FJC. Assessment of automatic cephalometric landmark identification using artificial intelligence. Orthod Craniofac Res. 2021;24(Suppl 2):37-42. doi:10.1111/ocr.12542
[14]. Jeon S, Lee KC. Comparison of cephalometric measurements between conventional and automatic cephalometric analysis using convolutional neural network. Prog Orthod. 2021;22(1):14. doi:10.1186/s40510-021-00358-4
[15]. Yu HJ, Cho SR, Kim MJ, Kim WH, Kim JW, Choi J. Automated skeletal classification with lateral cephalometry based on artificial intelligence. J Dent Res. 2020;99(3):249-256. doi:10.1177/0022034520901715
[16]. Kim HJ, Kim KD, Kim DH. Deep convolutional neural network-based skeletal classification of cephalometric image compared with automated-tracing software. Sci Rep. 2022;12:11659. doi:10.1038/s41598-022-15856-6
[17]. Jung SK, Kim TW. New approach for the diagnosis of extractions with neural network machine learning. Am J Orthod Dentofacial Orthop. 2016;149(1):127-133. doi:10.1016/j.ajodo.2015.07.030
[18]. Mercier JP, Rossi C, Sanchez IN, Renovales ID, Sahagún PM, Templier L. Reliability and accuracy of artificial intelligence-based software for cephalometric diagnosis. A diagnostic study. BMC Oral Health. 2024;24(1):1309. doi:10.1186/s12903-024-05097-6
[19]. Kazimierczak N, Chwarścianek N, Sultani N, Janiszewska-Olszowska J, et al. Reliability of AI-driven cephalometric analysis on CBCT: comparison with manual 2D cephalometry. Prog Orthod. 2026;27(1):8. doi:10.1186/s40510-026-00613-6
[20]. Temponi NCG, Petegrosso M, Salles IGV, Rodrigues AC, Girotto AL, Alcantara PCC, Peres RRH, Soares SSC. Artificial intelligence in the prediction of skeletal growth and the optimization of orthopedic and orthodontic treatment timing: closing the Texas Medicaid access gap — a narrative review. IOSR J Dent Med Sci. 2026;25(8 Ser 3):46-50. doi:10.9790/0853-2508034650