



International Journal of Innovative Technologies in Social Science

e-ISSN: 2544-9435

Operating Publisher
SciFormat Publishing Inc.
ISNI: 0000 0005 1449 8214

2734 17 Avenue SW,
Calgary, Alberta, T3E0A7,
Canada
+15878858911
editorial-office@sciformat.ca

ARTICLE TITLE	AI-ASSISTED CEPHALOMETRIC ANALYSIS VERSUS MANUAL TRACING: A LITERATURE REVIEW
----------------------	---

DOI	https://doi.org/10.31435/ijitss.3(51).2026.6342
------------	---

RECEIVED	30 May 2026
-----------------	-------------

ACCEPTED	11 September 2026
-----------------	-------------------

PUBLISHED	16 September 2026
------------------	-------------------

LICENSE



The article is licensed under a **Creative Commons Attribution 4.0 International License**.

© The author(s) 2026.

This article is published as open access under the Creative Commons Attribution 4.0 International License (CC BY 4.0), allowing the author to retain copyright. The CC BY 4.0 License permits the content to be copied, adapted, displayed, distributed, republished, or reused for any purpose, including adaptation and commercial use, as long as proper attribution is provided.

AI-ASSISTED CEPHALOMETRIC ANALYSIS VERSUS MANUAL TRACING: A LITERATURE REVIEW

Agnieszka Ciepiela (Corresponding Author, Email: aaga.ciepiela@gmail.com)

Private Dental Practice, Krakow, Poland

ORCID ID: 0009-0007-7132-9195

Gabriela Pilch

Private Dental Practice, Kraków, Poland

ORCID ID: 0009-0004-9302-5844

Małgorzata Berowska

Private Dental Practice, Bielsko-Biała, Poland

ORCID ID: 0009-0005-0789-6721

Jakub Sagan

Private Dental Practice, Krakow, Poland

ORCID ID: 0009-0000-6105-4860

Michał Zemowski

Private Dental Practice, Krakow, Poland

ORCID ID: 0009-0003-3891-7324

Zuzanna Kucia

Private Dental Practice, Kraków, Poland

ORCID ID: 0009-0005-9639-8571

Daniil Yefimchuk

Private Medical Practice, Krakow, Poland

ORCID ID: 0009-0002-3843-6646

ABSTRACT

Objectives: This review aimed to investigate the accuracy, agreement, and repeatability of artificial intelligence driven cephalometric analysis compared with conventional manual tracing performed by orthodontic experts.

Methods: A structured literature review was conducted using the PubMed database. The search was performed on May 5, 2026, and included English-language studies published within the previous five years. Studies evaluating the accuracy and reliability of AI-based cephalometric landmark identification and/or cephalometric measurements compared with manual tracing were considered eligible. Meta-analyses, systematic reviews, umbrella reviews, critical reviews, and high-quality comparative studies were included. Of the 113 records identified, 30 full-text articles were assessed for eligibility, and 17 studies met the inclusion criteria and were included in the final review.

Results: AI-assisted cephalometric analysis demonstrated high agreement with manual landmark identification performed by experienced orthodontists across the reviewed studies. Contemporary deep learning algorithms achieved clinically acceptable accuracy for most routinely used cephalometric landmarks and markedly reduced analysis time compared with conventional manual tracing. Nevertheless, discrepancies persisted for anatomically challenging landmarks, and all reviewed studies recommended clinician verification of AI-generated analyses.

Conclusions: AI-driven cephalometric analysis provides clinically comparable accuracy to manual tracing and offers enhanced consistency and efficiency. While AI cannot yet replace orthodontists in complex diagnostic scenarios, it serves as a reliable and effective complementary tool for routine cephalometric assessment.

KEYWORDS

Artificial Intelligence, Orthodontists, Cephalometric Analysis, Landmark Detection, Comparison

CITATION

Ciepiela, A., Pilch, G., Berowska, M., Sagan, J., Zemowski, M., Kucia, Z., & Yefimchuk, D. (2026). AI-assisted cephalometric analysis versus manual tracing: A literature review. *International Journal of Innovative Technologies in Social Science*, 3(51). [https://doi.org/10.31435/ijitss.3\(51\).2026.6342](https://doi.org/10.31435/ijitss.3(51).2026.6342)

COPYRIGHT

© The author(s) 2026. This article is published as open access under the Creative Commons Attribution 4.0 International License (CC BY 4.0), allowing the author to retain copyright. The CC BY 4.0 License permits the content to be copied, adapted, displayed, distributed, republished, or reused for any purpose, including adaptation and commercial use, as long as proper attribution is provided.

Introduction

Artificial intelligence (AI) has become one of the most rapidly evolving technologies in healthcare, fundamentally transforming diagnostic imaging, clinical decision-making, and workflow automation. In orthodontics, the integration of digital technologies—including cone-beam computed tomography (CBCT), intraoral scanning, computer-aided design and manufacturing (CAD/CAM), and AI—has accelerated the transition toward fully digital diagnostic and treatment planning workflows. Among these applications, automated cephalometric analysis has emerged as one of the most extensively investigated fields because of its direct relevance to routine orthodontic practice.

Cephalometric analysis remains an indispensable component of orthodontic diagnosis, allowing quantitative assessment of craniofacial morphology, evaluation of skeletal and dentoalveolar relationships, prediction of facial growth, and assessment of treatment outcomes. The reliability of cephalometric analysis depends primarily on precise identification of anatomical landmarks, as even minor localisation errors may alter linear and angular measurements, potentially influencing diagnosis and treatment planning. Despite decades of clinical use, conventional landmark identification remains a labour-intensive procedure requiring considerable experience and anatomical knowledge.

Manual cephalometric tracing is also inherently subjective. Variability between and within observers has been consistently reported, particularly for landmarks located in anatomically complex regions or affected by overlapping structures on two-dimensional radiographs. Additional limitations arise from image magnification, projection errors, and superimposition of bilateral anatomical structures, all of which may compromise landmark visibility and measurement accuracy. These limitations have stimulated the development of automated approaches aimed at improving the efficiency, reproducibility, and standardisation of cephalometric analysis.

Recent advances in deep learning, particularly convolutional neural networks (CNNs), have markedly improved the performance of automated landmark detection systems. Modern AI algorithms are capable of identifying cephalometric landmarks and performing complete cephalometric analyses within seconds,

achieving levels of accuracy that increasingly approach those obtained by experienced orthodontists. Consequently, several commercial AI-assisted cephalometric software platforms have been introduced into clinical practice, where they promise to reduce clinician workload, shorten analysis time, and improve workflow efficiency.

Despite these technological advances, the clinical role of AI remains a matter of ongoing debate. Most published studies have focused on technical performance indicators, including mean radial error (MRE), success detection rate (SDR), or agreement within predefined error thresholds. Although these metrics provide valuable information regarding localisation accuracy, they do not necessarily reflect clinical performance. Accurate landmark detection alone does not guarantee equivalent cephalometric measurements, diagnostic interpretation, or treatment planning. Moreover, virtually all AI systems are evaluated against manual landmark identification performed by experienced orthodontists, which itself is affected by intra- and inter-observer variability. Consequently, the absence of an objective reference standard represents one of the principal methodological challenges in assessing the true diagnostic capability of AI-assisted cephalometric analysis.

Over the past few years, several systematic reviews and meta-analyses have summarised the rapidly expanding evidence regarding AI-based cephalometric landmark detection, while recent comparative clinical studies have directly evaluated the performance of commercially available AI software against experienced orthodontists. Collectively, these studies suggest that AI has achieved clinically acceptable accuracy for routine landmark identification and substantially reduces analysis time. Nevertheless, uncertainty remains regarding its reliability for anatomically challenging landmarks and, more importantly, whether current AI systems can independently support orthodontic diagnosis without expert supervision.

Therefore, the aim of this literature review was to critically evaluate the current evidence regarding the accuracy, reproducibility, and clinical applicability of AI-assisted cephalometric analysis compared with conventional analysis performed by experienced orthodontists. Particular emphasis was placed on the agreement between AI and human examiners, factors influencing landmark localisation accuracy, clinically challenging landmarks, workflow efficiency, and the current and future role of AI as a decision-support tool in contemporary orthodontic practice.

Methodology

A structured literature search was conducted on May 5, 2026 in the PubMed database to identify clinically relevant studies evaluating artificial intelligence (AI)-assisted cephalometric analysis in comparison with manual tracing performed by orthodontists. The search strategy was designed to capture studies directly addressing diagnostic accuracy, reliability, and usability in orthodontic practice. The following query was applied:

(((Artificial intelligence) AND Orthodontics) AND (Cephalometry OR cephalometric analysis OR cephalometric landmarks)) AND (Humans OR orthodontists)) AND (Comparison OR evaluation)).

PubMed automatically broadened the search using equivalent MeSH terms. Eligibility filters were applied to include English-language studies in human populations with full-text availability in open access, published within the last five years to ensure recency and clinical applicability.

Studies were included if they evaluated AI for the detection of cephalometric landmarks or measurement of skeletal and dental parameters and reported a direct comparison with manual tracing performed by trained orthodontists. Studies were required to present quantitative outcomes relevant to clinical interpretation, such as mean landmark error, linear and angular deviations, or reproducibility metrics. Systematic reviews, meta-analyses, umbrella reviews, critical reviews, and primary comparative studies were eligible.

Exclusion criteria included: studies without manual comparison, purely technical algorithm-development research without clinical validation, prediction-only models unrelated to cephalometric measurement accuracy, and studies involving non-cephalometric diagnostics.

The initial literature search identified a total of 113 potentially relevant studies. After the removal of duplicate records and the screening of titles and abstracts, 83 articles were excluded because they did not meet the predefined clinical or methodological inclusion criteria or were considered irrelevant to the scope of this review. The full texts of the remaining 30 publications were subsequently retrieved and assessed in detail for eligibility. Following comprehensive evaluation, 17 studies fulfilled all inclusion criteria and were included in the qualitative synthesis. These comprised three systematic reviews with meta-analysis [1–3], one umbrella review [4], one critical review [5], and twelve primary comparative studies investigating the diagnostic

performance of artificial intelligence-based cephalometric analysis in comparison with conventional manual tracing performed by orthodontic specialists [6–17].

For the final clinical synthesis, particular emphasis was placed on seven high-quality publications, including three meta-analyses [1–3], one umbrella review [4], one critical review [5], and two primary comparative clinical studies [6,7]. These studies were prioritized because they provided the highest level of available evidence, demonstrated robust methodological quality, and directly addressed the primary objective of this review by comparing the diagnostic performance of AI-assisted cephalometric analysis with conventional manual methods. Furthermore, these investigations included comprehensive statistical analyses and clinically relevant outcome measures, making them particularly valuable for evaluating the current evidence regarding the accuracy and reliability of AI-based systems. The remaining ten studies [8–17] were included as supporting evidence, providing additional information on the performance of different commercial and research-based AI platforms, landmark-specific diagnostic accuracy, measurement reproducibility, operator-independent consistency, and the potential clinical applicability of automated cephalometric analysis in contemporary orthodontic practice.

Results

Characteristics of the included studies

Seventeen studies published between 2021 and 2025 met the inclusion criteria and were included in this review. The evidence comprised four systematic reviews with meta-analyses [1–4], one critical review [5], two comparative evaluations of commercially available AI-assisted cephalometric software [6,7], and ten comparative clinical or diagnostic studies directly comparing AI-assisted cephalometric analysis with manual tracing performed by experienced orthodontists [8–17].

The included studies evaluated AI performance using both two-dimensional (2D) lateral cephalograms and three-dimensional (3D) CBCT images. Most investigations employed deep learning-based algorithms, predominantly convolutional neural networks (CNNs), integrated into proprietary or commercially available software platforms. The reference standard across all comparative studies was manual landmark identification performed by experienced orthodontists. The most frequently reported outcome measures included mean radial error (MRE), success detection rate (SDR), landmark localisation accuracy, agreement with manual tracing, reproducibility of cephalometric measurements, and time required for analysis.

Accuracy of AI-assisted cephalometric landmark identification

Overall, the reviewed evidence demonstrated high accuracy of AI-assisted cephalometric landmark identification. The systematic review and meta-analysis by de Queiroz Tavares Borges Mesquita et al. [2] reported agreement rates between AI and manual landmark identification ranging from 79% within a 2-mm error threshold to 90% within a 3-mm threshold, with a pooled mean discrepancy of 2.05 mm. Similarly, Hendrickx et al. [1] demonstrated that AI-based landmark localisation generally remained below the clinically accepted 2-mm threshold, with a pooled mean localisation error of 1.43 mm (95% CI: 0.95–1.91).

Studies evaluating three-dimensional landmark detection also reported encouraging results. Serafin et al. [3] observed a pooled mean localisation error of 2.44 mm (95% CI: 1.83–3.05), although considerable heterogeneity was present among studies due to differences in CBCT datasets, landmark definitions, and deep learning architectures. Despite these variations, more recent investigations demonstrated a progressive reduction in localisation error, indicating continuous improvements in AI performance.

Across the included studies, deep learning algorithms consistently outperformed conventional machine learning methods, reflecting advances in network architecture, computational capacity, and optimisation strategies [1–5].

Comparison between AI-assisted analysis and experienced orthodontists

All comparative studies reported a high level of agreement between AI-assisted and manual cephalometric analysis performed by experienced orthodontists. Most commercially available AI software generated linear and angular measurements that closely approximated manual tracing, with discrepancies generally remaining within clinically acceptable limits [6–17].

Ye et al. [6] compared three commercially available AI-assisted software platforms and demonstrated mean radial errors of 1.11 ± 0.48 mm for MyOrthoX, 0.80 ± 0.26 mm for Angelalign, and 0.97 ± 0.51 mm for Digident. Corresponding success detection rates within a 2-mm threshold reached 89.99%, 93.09%, and 87.53%, respectively, supporting the clinical applicability of these systems.

Similarly, Bor et al. [8], Zaheer et al. [12], Mercier et al. [13], Çoban et al. [14], Prince et al. [10], Mahto et al. [11], Jeon and Lee [17], and Kunz et al. [9] reported that AI-assisted cephalometric measurements demonstrated reproducibility and reliability comparable to manual analyses performed by orthodontists. Although statistically significant differences were observed for selected cephalometric measurements or individual landmarks, these discrepancies were generally considered clinically acceptable. Nevertheless, all authors recommended expert verification before using AI-generated analyses for diagnosis or treatment planning.

Challenging cephalometric landmarks

Landmark localisation accuracy varied considerably among anatomical structures. Across multiple systematic reviews and comparative studies, the gonion was consistently identified as one of the most difficult landmarks for both AI systems and human observers [1–4]. Reduced localisation accuracy was also reported for the posterior nasal spine, articulare, pterygomaxillary fissure, porion, orbitale, basion, and the upper and lower incisor landmarks [1,2,6,16].

These localisation difficulties were mainly attributed to anatomical superimposition, bilateral projection of craniofacial structures on two-dimensional radiographs, poorly defined anatomical boundaries, image quality, and variations in craniofacial morphology. In contrast, landmarks such as the sella, nasion, subnasale, and soft tissue pogonion generally demonstrated higher localisation accuracy across different AI systems [1,2,6].

Factors influencing AI performance

Several factors influencing AI performance were consistently identified throughout the reviewed literature. Algorithm architecture, image quality, imaging modality, anatomical variability, and the size and diversity of training datasets all affected landmark localisation accuracy [1–5].

Studies using standardised public datasets, particularly the IEEE ISBI 2015 dataset, generally reported higher localisation accuracy than investigations based on independent clinical datasets [1]. Conversely, studies performing external validation using heterogeneous patient populations often demonstrated lower accuracy, highlighting the importance of diverse multicentre training datasets for improving model generalisability [1,4].

Deep learning algorithms consistently demonstrated superior performance compared with earlier machine learning approaches, whereas three-dimensional applications remain limited by the relatively small number of available CBCT datasets and the greater complexity of volumetric landmark identification [3,4].

Clinical applicability and workflow efficiency

One of the most consistent findings across the reviewed literature was the substantial reduction in analysis time achieved by AI-assisted cephalometric analysis. Systematic reviews reported that automated landmark localisation was typically completed in less than one minute [1,2]. In contrast, manual cephalometric tracing performed by experienced orthodontists generally required approximately 10–20 minutes, depending on examiner experience, image quality, and the number of landmarks analysed [6].

Comparative clinical studies confirmed these observations, demonstrating that AI-assisted software substantially improved workflow efficiency while maintaining measurement accuracy comparable to manual analysis [6–17]. Nevertheless, all reviewed studies emphasised that automated landmark detection should be considered a decision-support tool rather than a replacement for orthodontists. Expert review of automatically identified landmarks remained recommended before diagnostic interpretation and treatment planning.

Discussion

AI-assisted cephalometric analysis has reached performance comparable to experienced orthodontists

The present review demonstrates that contemporary AI-assisted cephalometric analysis has achieved a level of accuracy that is, in most clinical scenarios, comparable to manual analysis performed by experienced orthodontists. Evidence from systematic reviews, meta-analyses, and recent comparative clinical studies consistently indicates high agreement between automated and manual landmark localisation, with pooled success rates ranging from 79% to 90% depending on the accepted error threshold and mean localisation errors generally remaining below the clinically accepted limit of 2 mm [1,2]. Likewise, recent comparative investigations evaluating commercially available software confirmed that AI-generated cephalometric measurements closely resemble those obtained by experienced clinicians, while substantially reducing the time required for analysis [6–17].

These findings reflect the remarkable progress achieved in deep learning over recent years. Earlier machine learning approaches relied primarily on handcrafted image features, whereas current convolutional neural networks automatically learn hierarchical image representations directly from large training datasets. Consequently, modern algorithms demonstrate substantially improved robustness and generalisability, resulting in progressively higher localisation accuracy than that reported in earlier generations of AI systems [1–5].

Importantly, recent comparative studies extend the conclusions of previous systematic reviews by evaluating commercially available software under routine clinical conditions rather than laboratory settings alone. Studies comparing systems such as MyOrthoX, Angelalign, Digident, WebCeph, AutoCEPH, and other commercially available platforms consistently demonstrated that AI-assisted analysis produces measurements comparable to manual tracing performed by experienced orthodontists [6–17]. Although statistically significant differences were occasionally identified for individual measurements, these discrepancies were generally small and remained within clinically acceptable limits. Collectively, these findings suggest that automated cephalometric analysis has matured from an experimental technology into a clinically applicable decision-support tool.

However, comparable landmark localisation accuracy should not be interpreted as evidence that AI has become equivalent to an experienced orthodontist. Most studies evaluated technical agreement between AI-generated and manually identified landmarks rather than the overall diagnostic process. Orthodontic diagnosis extends beyond landmark identification and requires integration of cephalometric findings with clinical examination, facial analysis, patient history, growth potential, occlusal characteristics, and treatment objectives. Therefore, although AI increasingly reproduces the technical component of cephalometric analysis, it cannot currently replace the comprehensive clinical reasoning performed by experienced orthodontists [5,13,15].

Furthermore, the comparison between AI and human performance is inherently complex because manual landmark identification itself is not perfectly reproducible. Multiple studies included in the reviewed systematic reviews demonstrated clinically relevant intra- and inter-observer variability even among experienced orthodontists [1,2]. Consequently, the differences observed between AI and manual tracing should be interpreted within the context of normal human variability rather than assuming manual analysis represents an absolute reference standard. This issue is particularly important because current AI algorithms are trained using manually annotated datasets. As a result, AI performance inevitably reflects the strengths and limitations of the human-generated reference data on which it was developed.

Taken together, the available evidence suggests that the central question is no longer whether AI can accurately identify cephalometric landmarks. Instead, the current challenge is determining how AI should be integrated into orthodontic practice to maximise efficiency while maintaining diagnostic reliability and appropriate clinical oversight. Current evidence consistently supports a collaborative model in which AI performs automated landmark detection and cephalometric measurements, whereas experienced orthodontists remain responsible for verification, interpretation, diagnosis, and treatment planning.

Why does AI still differ from experienced orthodontists?

Although recent studies consistently demonstrate that AI-assisted cephalometric analysis achieves accuracy comparable to experienced orthodontists for most routinely used landmarks, complete equivalence between automated and human assessment has not yet been achieved. The remaining discrepancies arise not only from technical limitations of AI algorithms but also from the inherent complexity of cephalometric analysis itself.

One of the most important methodological challenges is the absence of a true reference standard for cephalometric landmark identification. Nearly all studies included in this review evaluated AI performance by comparing automated landmark localisation with manual tracing performed by experienced orthodontists [1–17]. However, manual landmark identification is inherently subjective and has long been recognised as being affected by both intra- and inter-observer variability. Even experienced clinicians may identify the same landmark differently, particularly when anatomical boundaries are poorly defined or obscured by overlapping structures [1–4]. Consequently, discrepancies observed between AI and manual tracing do not necessarily indicate inferior AI performance but may instead reflect normal variability among human observers.

This limitation is particularly relevant because current deep learning algorithms are trained using manually annotated datasets. The quality of AI predictions therefore depends directly on the consistency and accuracy of the reference annotations used during model development. If the training data contain inconsistencies or observer-dependent errors, these may be learned and subsequently reproduced by the algorithm. Several systematic reviews identified variability in annotation protocols, insufficient reporting of examiner calibration, and inadequate description of reference standards as important sources of bias within the existing literature [1–4]. Consequently, improving the quality and standardisation of manually annotated datasets may be as important as further optimisation of AI algorithms themselves.

Another factor contributing to differences between AI and orthodontists is the variability of clinical datasets. Several studies highlighted that algorithm performance depends strongly on the characteristics of the training population, including age, sex, craniofacial morphology, skeletal pattern, ethnicity, image quality, and the presence of craniofacial anomalies [1,4]. AI systems trained on relatively homogeneous datasets often perform well under controlled validation conditions but demonstrate reduced accuracy when applied to more heterogeneous clinical populations. For example, lower localisation accuracy has been reported in patients with craniofacial deformities, cleft lip and palate, or severe skeletal discrepancies, indicating that algorithm generalisability remains an important challenge [1,4].

The reviewed literature also demonstrated that differences between AI systems themselves should not be overlooked. Comparative studies evaluating commercially available software reported variations in localisation accuracy despite analysing identical radiographs [6,7,9]. In the study by Ye et al. [6], Angelalign achieved the lowest mean radial error (0.80 ± 0.26 mm) and the highest success detection rate within a 2-mm threshold (93.09%), whereas MyOrthoX and Digident demonstrated slightly lower performance. These findings suggest that algorithm architecture, training strategy, and dataset composition continue to influence clinical performance, emphasising that conclusions regarding AI accuracy cannot be generalised to all available software.

Finally, it is important to recognise that cephalometric analysis represents only one component of orthodontic diagnosis. Unlike AI systems, experienced orthodontists integrate cephalometric findings with clinical examination, facial aesthetics, occlusion, growth potential, patient expectations, and treatment objectives. Current AI algorithms are designed to automate landmark detection and measurement rather than clinical reasoning. Therefore, even when landmark localisation accuracy approaches human performance, AI cannot currently replicate the comprehensive decision-making process required for orthodontic diagnosis and treatment planning. This distinction explains why all reviewed studies recommended clinician supervision despite the high technical performance achieved by contemporary AI systems [5,13,15].

Challenging cephalometric landmarks: why do errors persist?

Although the overall accuracy of AI-assisted cephalometric analysis has improved substantially, landmark localisation remains inconsistent for specific anatomical structures. The reviewed studies consistently demonstrated that localisation errors are not uniformly distributed across all landmarks but are concentrated in anatomically complex regions. Interestingly, many of these landmarks are also recognised as difficult to identify manually by experienced orthodontists, suggesting that the observed limitations are related to anatomical complexity rather than deficiencies unique to AI [1–4,6].

Among all evaluated landmarks, the gonion was most frequently reported as the most challenging for both automated systems and human observers [1–4]. Unlike anatomically distinct landmarks, the gonion is a constructed point defined by the intersection of the posterior border of the mandibular ramus and the inferior mandibular border. On two-dimensional lateral cephalograms, bilateral mandibular contours are superimposed, making precise localisation inherently difficult. Similar challenges have been described for articulare, posterior nasal spine (PNS), pterygomaxillary fissure, porion, orbitale, and basion, all of which demonstrated greater localisation errors than other routinely analysed landmarks [1,2,6].

Several anatomical and technical factors explain these findings. Two-dimensional cephalograms represent projections of three-dimensional craniofacial structures, resulting in image distortion, magnification, and superimposition of bilateral anatomical components. Consequently, structures located within regions of increased anatomical complexity often lack clearly distinguishable radiographic borders, reducing localisation precision for both AI systems and clinicians [1,2,4]. Variations in patient positioning during image acquisition, differences in radiographic equipment, and image quality may further increase localisation uncertainty.

Dental landmarks also present unique challenges. Ye et al. [6] reported increased localisation errors for the upper and lower incisors, particularly in patients presenting with severe crowding, root morphology variations, or open root apices. Similar observations have been reported for posterior nasal spine and pterygomaxillary fissure, where overlapping anatomical structures and reduced radiographic contrast complicate precise landmark identification [2]. These findings indicate that landmark-specific anatomical characteristics remain an important determinant of AI performance.

Three-dimensional imaging has been proposed as a potential solution to overcome several limitations associated with conventional lateral cephalograms. CBCT eliminates anatomical superimposition and allows direct visualisation of craniofacial structures, theoretically facilitating more precise landmark localisation. However, current evidence does not demonstrate a consistent superiority of AI performance on three-dimensional images. Although CBCT provides more detailed anatomical information, automated localisation in volumetric datasets remains computationally more demanding and is currently limited by the relatively small number of available annotated datasets [1,3,4]. Consequently, localisation accuracy on three-dimensional images has not yet consistently exceeded that achieved on conventional two-dimensional cephalograms.

Importantly, landmark-specific performance should be considered when interpreting the clinical applicability of AI-assisted cephalometric analysis. Rather than assuming uniform algorithm performance across all anatomical structures, clinicians should recognise that certain landmarks remain inherently more susceptible to localisation errors regardless of whether they are identified manually or automatically. This observation supports a selective verification strategy in which orthodontists pay particular attention to landmarks known to demonstrate reduced reproducibility, while allowing AI to perform routine localisation of anatomically well-defined structures. Such collaboration may maximise the efficiency of AI-assisted workflows without compromising diagnostic accuracy.

Clinical relevance of AI-assisted cephalometric analysis

While numerous studies have demonstrated that AI-assisted cephalometric analysis achieves landmark localisation accuracy comparable to experienced orthodontists, technical performance alone does not determine clinical usefulness. Orthodontic diagnosis is a multifactorial process that extends beyond the identification of cephalometric landmarks and requires comprehensive interpretation of radiographic findings within the broader clinical context. Consequently, the clinical value of AI should be assessed not only by localisation accuracy but also by its ability to support diagnostic decision-making without compromising patient safety.

One of the most consistent findings across the reviewed literature is that AI substantially improves the efficiency and standardisation of cephalometric analysis while reducing operator-dependent variability [1,2,5]. Unlike manual tracing, which is influenced by examiner experience, fatigue, and subjective landmark interpretation, AI algorithms generate highly reproducible results under identical conditions. This increased consistency may contribute to more standardised orthodontic records and facilitate longitudinal patient follow-up, particularly in large clinical practices and academic institutions.

However, reproducibility should not be interpreted as synonymous with diagnostic accuracy. AI systems identify anatomical landmarks according to patterns learned during training but remain unable to integrate additional clinical information routinely considered by orthodontists. Facial aesthetics, soft tissue evaluation, occlusal relationships, patient age, growth status, skeletal maturation, functional assessment, and treatment

objectives all contribute to orthodontic diagnosis but remain outside the scope of current automated cephalometric software [5,15]. Consequently, AI currently supports only one component of the diagnostic pathway rather than replacing comprehensive clinical reasoning.

Another important consideration concerns the clinical significance of localisation errors. Most studies evaluated AI performance using a 2-mm success threshold, which has traditionally been considered clinically acceptable for cephalometric landmark identification [1,2]. Nevertheless, Polizzi and Leonardi [4] questioned whether this threshold adequately reflects clinical reality, highlighting earlier evidence suggesting that substantially smaller localisation errors—approximately 0.81 mm—may be required to avoid influencing certain cephalometric measurements. Because small deviations in landmark position may propagate into larger errors in linear or angular measurements, particularly when multiple landmarks contribute to a single analysis, statistical agreement with manual tracing does not necessarily guarantee equivalent diagnostic outcomes. This observation emphasises that future validation studies should increasingly focus on the influence of AI-generated landmark localisation on diagnostic conclusions and treatment planning rather than on landmark accuracy alone.

The reviewed comparative studies consistently reached a similar conclusion regarding the role of AI in contemporary orthodontic practice. Although commercially available systems demonstrated measurement accuracy comparable to experienced orthodontists, none of the studies recommended replacing clinicians with fully automated analysis [6–17]. Instead, AI was consistently described as a decision-support tool capable of accelerating routine workflows, reducing repetitive manual tasks, and improving measurement consistency while maintaining clinician oversight. Studies by Indermun et al. [15], Mercier et al. [13], and Kunz et al. [9] particularly emphasised that expert review remains essential for identifying potential landmark localisation errors, interpreting complex craniofacial anatomy, and integrating cephalometric findings into comprehensive treatment planning.

From a practical perspective, the greatest clinical benefit of AI may therefore lie in the redistribution rather than replacement of professional responsibilities. Automated landmark detection allows orthodontists to spend less time performing repetitive measurements and more time on diagnostic interpretation, patient communication, and treatment planning. Such a collaborative workflow combines the computational speed and consistency of AI with the clinical judgement, experience, and contextual reasoning that currently remain uniquely human. Accordingly, the available evidence supports a model of augmented intelligence, in which AI enhances clinical performance without replacing the orthodontist's responsibility for diagnosis and therapeutic decision-making.

Current limitations of AI implementation in clinical orthodontics

Despite the encouraging diagnostic performance reported across the reviewed studies, several challenges continue to limit the widespread clinical implementation of AI-assisted cephalometric analysis. These limitations relate not only to the algorithms themselves but also to methodological inconsistencies in the existing literature and the characteristics of the datasets used for model development.

One of the major limitations identified across systematic reviews is the lack of standardisation in dataset construction and annotation protocols [1–4]. Most AI models have been trained using independently collected datasets that differ substantially in sample size, imaging protocols, patient demographics, landmark definitions, and annotation procedures. Consequently, direct comparison between algorithms remains difficult, and the reported accuracy may not be fully generalisable to other clinical settings. Hendrickx et al. [1] and Mesquita et al. [2] emphasised that many studies relied on the public IEEE ISBI 2015 dataset, which facilitates benchmarking but may not adequately represent the anatomical variability encountered in everyday orthodontic practice. Models trained predominantly on homogeneous datasets may therefore demonstrate reduced performance when applied to external populations.

Another important limitation concerns the availability of annotated three-dimensional datasets. Although CBCT has the potential to overcome several shortcomings of conventional two-dimensional cephalometry by eliminating anatomical superimposition, the development of reliable AI models for three-dimensional landmark localisation remains constrained by the relatively small number of high-quality annotated volumetric images [1,3,4]. The greater complexity of 3D landmark identification and the lack of standardised three-dimensional cephalometric protocols further contribute to variability among published studies.

The reviewed literature also highlights the methodological quality of the available evidence. Most systematic reviews identified a high risk of bias in the included studies, primarily related to patient selection, insufficient reporting of reference standards, inadequate description of examiner calibration, and limited external validation [1–4]. Furthermore, the majority of comparative studies were retrospective and conducted in single-centre settings, reducing the generalisability of their findings. Prospective multicentre studies using standardised reporting guidelines, such as CONSORT-AI and SPIRIT-AI, remain relatively scarce.

Another important consideration is algorithm transparency. Contemporary deep learning models often function as "black-box" systems, providing highly accurate predictions without explaining how individual decisions are reached [5]. From a clinical perspective, limited interpretability may reduce clinicians' confidence in AI-generated outputs and complicate the identification of erroneous landmark localisation. Consequently, future AI systems should incorporate explainable AI approaches capable of indicating prediction confidence and highlighting landmarks that require manual verification.

Table 1 summarises the key findings of this review and illustrates the current role of artificial intelligence in cephalometric analysis. Collectively, the available evidence suggests that AI has evolved into a clinically valuable tool capable of improving the efficiency, consistency, and standardisation of cephalometric assessment. However, its performance remains influenced by the complexity of anatomical structures and the quality of radiographic images, indicating that further technological development is still required. Finally, despite their high technical performance, current AI systems remain dependent on human expertise. All comparative studies included in this review concluded that automated landmark localisation should be verified by an experienced orthodontist before clinical interpretation and treatment planning [6–17]. Therefore, the current evidence supports AI as an adjunctive technology that enhances clinical workflow rather than an autonomous diagnostic system capable of replacing orthodontists.

Table 1. Comparison of AI performance according to clinically relevant domains

Domain	Evidence from included studies	Overall conclusion
Landmark localisation	High agreement (79–90%)	Clinically acceptable
Linear/angular measurements	Comparable to orthodontists	Minor differences
Difficult landmarks	Lower accuracy for gonion, PNS, articulare, incisors	Further improvement needed
Analysis time	<1 min vs 10–20 min	AI substantially faster
Reproducibility	Higher standardisation	Less operator-dependent
Clinical applicability	Suitable for routine assistance	Orthodontist supervision still required
Independent diagnosis	Limited evidence	Not recommended
Replacement of orthodontists	No study supported complete replacement	AI complements, not replaces

Future perspectives

The rapid evolution of artificial intelligence suggests that its role in orthodontics will continue to expand beyond automated landmark localisation. Future research should therefore focus not only on improving localisation accuracy but also on increasing the clinical applicability, robustness, and transparency of AI-assisted diagnostic systems.

One of the highest priorities is the development of large multicentre datasets representing diverse patient populations. Future training databases should include variations in age, sex, ethnicity, skeletal patterns, craniofacial anomalies, and image acquisition protocols to improve algorithm generalisability and minimise dataset bias [1,2,4]. Standardisation of landmark definitions and annotation protocols would further enhance reproducibility and facilitate meaningful comparisons between different AI models.

Another important direction involves strengthening external validation. Most currently available algorithms have been evaluated using retrospective datasets originating from the same institutions in which they were developed. Independent validation across multiple clinical centres and imaging systems is necessary before widespread implementation in routine orthodontic practice can be recommended [1–4].

Future AI systems should also evolve from simple landmark detection tools towards intelligent clinical assistants. Rather than merely identifying anatomical landmarks, next-generation algorithms could automatically assess landmark confidence, detect potentially inaccurate localisations, and recommend manual verification when uncertainty exceeds predefined thresholds. Such confidence-based decision support would enable a more efficient collaboration between AI and orthodontists while maintaining diagnostic safety.

Equally important is the integration of explainable AI into orthodontic diagnostics. Improving the transparency of AI decision-making would facilitate clinician acceptance, strengthen trust in automated analyses, and simplify the identification of algorithmic errors [5]. In parallel, future studies should evaluate not only landmark localisation accuracy but also the influence of AI-assisted cephalometric analysis on diagnostic decisions, treatment planning, treatment outcomes, and patient care.

Taken together, the available evidence indicates that the future of cephalometric analysis is unlikely to involve replacing orthodontists with fully autonomous AI systems. Instead, the most realistic and clinically appropriate model appears to be one of augmented intelligence, in which AI performs repetitive analytical tasks while orthodontists retain responsibility for clinical judgement, diagnosis, and therapeutic decision-making. Such collaboration has the potential to improve efficiency, increase standardisation, and optimise patient care without compromising diagnostic quality.

Limitations

This review has several limitations that should be considered when interpreting its findings. First, the literature search was restricted to the PubMed database and included only English-language articles with full-text availability published within the last five years. Although these criteria were selected to ensure the inclusion of recent and accessible evidence, potentially relevant studies indexed exclusively in other databases or published before the selected time frame may have been excluded. Consequently, publication and selection bias cannot be completely ruled out.

Second, although the available evidence includes recent systematic reviews, meta-analyses, and comparative clinical studies, the overall number of eligible publications remains limited [1–17]. Furthermore, considerable methodological heterogeneity was observed among the included studies. Differences in imaging modalities (2D lateral cephalograms versus 3D CBCT), AI algorithms, commercial software platforms, training datasets, reference standards, and reported outcome measures limited direct comparison of the reported findings [1–5]. In addition, studies evaluated different numbers and types of cephalometric landmarks, making quantitative comparisons across publications challenging.

Another important limitation concerns the quality of the available evidence. Most systematic reviews reported a moderate to high risk of bias among the primary studies, mainly due to non-random patient selection, inadequate description of reference standards, variability in annotation protocols, and limited external validation [1–4]. Although recent comparative clinical studies strengthened the available evidence by directly comparing AI-assisted analysis with experienced orthodontists [8–17], most were conducted in single-centre settings using relatively small patient cohorts and different commercial software platforms, limiting the generalisability of their findings.

Moreover, AI performance was almost exclusively evaluated against manual landmark identification performed by experienced orthodontists, which itself is influenced by examiner experience, training, and intra- and inter-observer variability [1–4,6,8,9,13,16]. Consequently, discrepancies between AI and manual tracing may reflect not only limitations of AI algorithms but also natural variability among human examiners. The absence of an objective gold standard for cephalometric landmark identification therefore remains one of the major methodological challenges when evaluating the true diagnostic performance of AI systems.

Finally, most studies focused primarily on landmark localisation accuracy and measurement reproducibility rather than on clinically relevant outcomes such as diagnostic agreement, treatment planning, or patient outcomes [4,5,13]. Consequently, further prospective multicentre studies using standardised protocols are required to determine whether AI-assisted cephalometric analysis can reliably support clinical decision-making across different patient populations and clinical settings.

Conclusions

The findings of this review demonstrate that artificial intelligence has achieved a clinically relevant level of accuracy in cephalometric analysis, with landmark identification and linear and angular measurements showing close agreement with manual tracings performed by experienced orthodontists. Across the highest-quality studies, AI-based systems consistently produced results within accepted diagnostic thresholds, typically defined by clinically acceptable error margins, while simultaneously offering superior repeatability and reduced observer-dependent variability. These characteristics collectively highlight the potential of artificial intelligence to enhance diagnostic efficiency, standardize cephalometric workflows, and support clinicians in routine orthodontic assessment.

In addition to diagnostic accuracy, the included evidence indicates that one of the most significant advantages of AI-based systems lies in their ability to reduce intra- and inter-examiner variability, which is a well-recognized limitation of traditional manual tracing. By minimizing subjective interpretation and fatigue-related inconsistencies, AI has the potential to improve the overall reproducibility of cephalometric measurements, particularly in high-volume clinical environments or large-scale diagnostic settings.

Despite these strengths, current evidence also indicates that AI systems do not yet fully replace the clinical judgement, contextual interpretation, and expertise required in complex or atypical cases. Variations in algorithm design, training datasets, and imaging quality continue to influence performance outcomes, and anatomically challenging landmarks remain a persistent source of error across multiple studies. In particular, cases involving overlapping structures, low radiographic contrast, or anatomical variability still require manual verification to ensure diagnostic accuracy and clinical safety. Furthermore, the observed differences in performance between commercial platforms highlight the ongoing lack of standardization in algorithm development and validation protocols.

From a broader perspective, these limitations emphasize that the current role of AI in cephalometric analysis is best understood as supportive rather than substitutive. While performance levels are increasingly comparable to human experts in many routine tasks, full clinical autonomy is not yet supported by the available evidence. Instead, optimal outcomes are achieved when AI systems are integrated into clinical workflows as decision-support tools, assisting clinicians rather than replacing them.

Overall, artificial intelligence should be regarded not as a substitute for orthodontic expertise but as a valuable complementary tool capable of improving the precision, reproducibility, and efficiency of cephalometric analysis. As future research continues to advance model transparency, expand dataset diversity, and incorporate multicenter external validation, AI is likely to play an increasingly significant and standardized role in orthodontic diagnostics and treatment planning. Continued methodological refinement and harmonization of evaluation protocols will be essential to fully realize its clinical potential.

Funding: the research received no external funding.

Conflict of Interest: No conflicts of interest to declare.

Use of AI: Artificial intelligence tools were used exclusively for linguistic editing and language refinement of the manuscript.

REFERENCES

1. Hendrickx, J., Gracea, R. S., Vanheers, M., Winderickx, N., Preda, F., Shujaat, S., & Jacobs, R. (2024). Can artificial intelligence-driven cephalometric analysis replace manual tracing? A systematic review and meta-analysis. *European Journal of Orthodontics*, 46(4), cjae029. <https://doi.org/10.1093/ejo/cjae029>
2. de Queiroz Tavares Borges Mesquita, G., Vieira, W. A., Vidigal, M. T. C., et al. (2023). Artificial intelligence for detecting cephalometric landmarks: A systematic review and meta-analysis. *J Digit Imaging*, 36, 1158–1179. <https://doi.org/10.1007/s10278-022-00766-w>
3. Serafin, M., Baldini, B., Cabitza, F., et al. (2023). Accuracy of automated 3D cephalometric landmarks by deep learning algorithms: systematic review and meta-analysis. *Radiol Med*, 128, 544–555. <https://doi.org/10.1007/s11547-023-01629-2>
4. Polizzi, A., & Leonardi, R. (2024). Automatic cephalometric landmark identification with artificial intelligence: An umbrella review of systematic reviews. *Journal of Dentistry*, 146, 105056. <https://doi.org/10.1016/j.jdent.2024.105056>
5. Nordblom, N. F., Büttner, M., & Schwendicke, F. (2024). Artificial intelligence in orthodontics: Critical review. *Journal of Dental Research*, 103(6), 577–584. <https://doi.org/10.1177/00220345241235606>
6. Ye, H., Cheng, Z., Ungvijanpunya, N., et al. (2023). Is automatic cephalometric software using artificial intelligence better than orthodontist experts in landmark identification? *BMC Oral Health*, 23, 467. <https://doi.org/10.1186/s12903-023-03188-4>
7. Baig, N., Gyasudeen, K. S., Bhattacharjee, T., et al. (2024). Comparative evaluation of commercially available AI-based cephalometric tracing programs. *BMC Oral Health*, 24, 1241. <https://doi.org/10.1186/s12903-024-05032-9>
8. Bor, S., Cigirim, S. Ç., & Kotan, S. (2024). Comparison of AI-assisted cephalometric analysis and orthodontist-performed digital tracing analysis. *Prog Orthod*, 25, 41. <https://doi.org/10.1186/s40510-024-00539-x>
9. Kunz, F., Stellzig-Eisenhauer, A., Widmaier, L., et al. (2025). Assessment of the quality of different commercial providers using artificial intelligence for automated cephalometric analysis compared to human orthodontic experts. *J Orofac Orthop*, 86, 145–160. <https://doi.org/10.1007/s00056-023-00491-1>
10. Prince, S. T. T., Srinivasan, D., Duraisamy, S., Kannan, R., & Rajaram, K. (2023). Reproducibility of linear and angular cephalometric measurements obtained by an artificial-intelligence assisted software (WebCeph) in comparison with digital software (AutoCEPH) and manual tracing method. *Dental Press Journal of Orthodontics*, 28(1), e2321214. <https://doi.org/10.1590/2177-6709.28.1.e2321214.oar>
11. Mahto, R. K., Kafle, D., Giri, A., Luintel, S., & Karki, A. (2022). Evaluation of fully automated cephalometric measurements obtained from web-based artificial intelligence driven platform. *BMC Oral Health*, 22(1), 132. <https://doi.org/10.1186/s12903-022-02170-w>
12. Zaheer, R., Shafique, H. Z., Khalid, Z., Shahid, R., Jan, A., Zahoor, T., Nawaz, R., & Ul Hassan, M. (2024). Comparison of semi and fully automated artificial intelligence driven softwares and manual system for cephalometric analysis. *BMC Medical Informatics and Decision Making*, 24(1), 271. <https://doi.org/10.1186/s12911-024-02664-3>
13. Mercier, J. P., Rossi, C., Sanchez, I. N., Renovales, I. D., Sahagún, P. M., & Templier, L. (2024). Reliability and accuracy of Artificial intelligence-based software for cephalometric diagnosis. A diagnostic study. *BMC Oral Health*, 24(1), 1309. <https://doi.org/10.1186/s12903-024-05097-6>
14. Çoban, G., Öztürk, T., Hashimli, N., & Yağci, A. (2022). Comparison between cephalometric measurements using digital manual and web-based artificial intelligence cephalometric tracing software. *Dental Press Journal of Orthodontics*, 27(4), e222112. <https://doi.org/10.1590/2177-6709.27.4.e222112.oar>
15. Indermun, S., Shaik, S., Nyirenda, C., Johannes, K., & Mulder, R. (2023). Human examination and artificial intelligence in cephalometric landmark detection-is AI ready to take over? *Dento Maxillo Facial Radiology*, 52(6), 20220362. <https://doi.org/10.1259/dmfr.20220362>
16. Lee, H., Cho, J. M., Ryu, S., et al. (2023). Automatic identification of posteroanterior cephalometric landmarks using a novel deep learning algorithm: a comparative study with human experts. *Sci Rep*, 13, 15506. <https://doi.org/10.1038/s41598-023-42870-z>
17. Jeon, S., & Lee, K. C. (2021). Comparison of cephalometric measurements between conventional and automatic cephalometric analysis using convolutional neural network. *Progress in Orthodontics*, 22(1), 14. <https://doi.org/10.1186/s40510-021-00358-4>