NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ

JOURNAL OF EDUCATION, HEALTH AND SPORT

eISSN 2391-8306 · Open Access · Peer-reviewed

apcz.umk.pl/JEHS · Nicolaus Copernicus University in Toruń

Cite as: CZAPIŃSKI, Olgierd, KWIATKOWSKA, Marta, MAŁEK, Zuzanna, WYCIŚLOK, Aleksandra, WIERTELAK-MAKAŁA, Kacper, WAŚKOWSKA, Katarzyna, KUBIAK, Weronika, KALSZTEIN, Julia and KĘPA, Adrianna. Current Evidence on Artificial Intelligence for Cephalometric Diagnosis and Orthodontic Treatment Planning — A Narrative Review. Journal of Education, Health and Sport. 2026;92:72449. https://doi.org/10.12775/JEHS.2026.92.72449

ARTICLE TIMELINE

Received: 23.05.2026 Revised: 25.05.2026

Accepted: 26.05.2026 Published: 10.06.2026

INDEXING & EVALUATION

MEiN points: 40 Unique ID: 201159

Disciplines: Physical culture sciences (Field of medical and health sciences); Health Sciences (Field of medical and health sciences).

The journal has been awarded 40 points in the parametric evaluation by the Polish Ministry of Higher Education and Science (Annex to the announcement of 05.01.2024, No. 32318). Unique Journal Identifier: 201159. Scientific disciplines: Physical culture sciences (Field of medical and health sciences); Health Sciences (Field of medical and health sciences).

Punkty Ministerialne z 2019 – aktualny rok 40 punktów. Załącznik do komunikatu Ministra Szkolnictwa Wyższego i Nauki z dnia 05.01.2024 Lp. 32318. Posiada Unikatowy Identyfikator Czasopisma: 201159. Przypisane dyscypliny naukowe: Nauki o kulturze fizycznej (Dziedzina nauk medycznych i nauk o zdrowiu); Nauki o zdrowiu (Dziedzina nauk medycznych i nauk o zdrowiu).

Current Evidence on Artificial Intelligence for Cephalometric Diagnosis and Orthodontic Treatment Planning — A Narrative Review

Olgierd Czapiński, ORCID https://orcid.org/0009-0007-7894-7201
Email: olgierd.czapinski@gmail.com
Institute of Dentistry of the Central Clinical Hospital of the Medical University of Lodz, ul. Pomorska 251, 92-213 Łódź, Poland

Marta Kwiatkowska, ORCID https://orcid.org/0009-0008-0198-0115
Email: marta.kwiatek2001@gmail.com
Institute of Dentistry of the Central Clinical Hospital of the Medical University of Lodz, ul. Pomorska 251, 92-213 Łódź, Poland

Zuzanna Małek, ORCID https://orcid.org/0009-0003-2427-3257
Email: zuzia_malek@wp.pl
Institute of Dentistry of the Central Clinical Hospital of the Medical University of Lodz, ul. Pomorska 251, 92-213 Łódź, Poland

Aleksandra Wyciślok, ORCID https://orcid.org/0009-0004-5249-3625
Email: olawy2010@gmail.com
Institute of Dentistry of the Central Clinical Hospital of the Medical University of Lodz, ul. Pomorska 251, 92-213 Łódź, Poland

Kacper Wiertelak-Makała, ORCID https://orcid.org/0009-0001-8236-3395
Email: kacper.wiertelak-makala@stud.umed.lodz.pl
Institute of Dentistry of the Central Clinical Hospital of the Medical University of Lodz, ul. Pomorska 251, 92-213 Łódź, Poland

Katarzyna Waśkowska, ORCID https://orcid.org/0009-0005-4761-0425
Email: k.waskowska534@wp.pl
Institute of Dentistry of the Central Clinical Hospital of the Medical University of Lodz, ul. Pomorska 251, 92-213 Łódź, Poland

Weronika Kubiak, ORCID https://orcid.org/0009-0008-0335-0635
Email: weronka333@gmail.com
Institute of Dentistry of the Central Clinical Hospital of the Medical University of Lodz, ul. Pomorska 251, 92-213 Łódź, Poland

Julia Kalsztein, ORCID https://orcid.org/0009-0000-4018-8827
Email: juliakalsztein@icloud.com
Institute of Dentistry of the Central Clinical Hospital of the Medical University of Lodz, ul. Pomorska 251, 92-213 Łódź, Poland

Adrianna Kępa, ORCID https://orcid.org/0009-0001-5833-7461
Email: adriannaka7@gmail.com
Józef Struś Multispecialist Municipal Hospital, ul. Szwajcarska 3, 61-285 Poznań, Poland

Corresponding Author: Olgierd Czapiński — olgierd.czapinski@gmail.com

Abstract

Background

Cephalometric analysis is a fundamental diagnostic tool in orthodontics, enabling clinicians to assess craniofacial morphology, classify skeletal relationships, and develop treatment plans. The use of artificial intelligence (AI) has created new opportunities to automate and standardize this diagnostic process, making everyday practice more efficient and less prone to error.

Aim

The aim of this study was to examine the application of artificial intelligence algorithms in cephalometric landmark identification and orthodontic treatment planning, with particular focus on their accuracy, comparison with human expert performance, clinical limitations, and future perspectives.

Materials and Methods

A narrative review of the current literature was conducted, focusing on the use of artificial intelligence mechanisms in image analysis, their accuracy, and future potential for full autonomy.

Results

The available literature indicates that AI-based systems consistently achieve clinically acceptable accuracy in automated cephalometric landmark detection on 2D lateral cephalograms. The proportion of landmarks detected within the 2 mm clinical threshold frequently matches experienced clinician performance. Three-dimensional CBCT-based landmarking demonstrated improving but still inferior accuracy compared to 2D analysis.

Conclusions

Artificial intelligence has become a reliable supporting tool in 2D cephalometric landmark identification, with current systems achieving the accuracy of experienced orthodontists within the clinical threshold. Three-dimensional CBCT analysis and AI-assisted treatment planning are advancing rapidly but are not yet equally reliable. Insufficient ethnic diversity in training datasets and domain shift across imaging devices remain ongoing concerns. AI should be regarded as an instrument that supports, not replaces, the expertise and responsibility of the orthodontist.

Keywords

artificial intelligence; deep learning; convolutional neural networks; cephalometric analysis; landmark identification; orthodontics; CBCT; treatment planning; growth assessment

1. Introduction

Cephalometric analysis has served a critical role in orthodontic diagnosis and treatment planning since Broadbent introduced the standardised lateral cephalogram in 1931 (1,2). By methodically identifying anatomical landmarks on standardised radiographic images and deriving complex measurements, clinicians can accurately assess craniofacial morphology, classify underlying skeletal relationships, and formulate comprehensive treatment plans.

For decades, this diagnostic pillar has relied heavily on the manual tracing of radiographs. Despite its broad clinical presence and undeniable diagnostic value, traditional manual landmark identification is inherently flawed. The manual tracing process is time-consuming and highly susceptible to errors (3). The recent emergence of AI, particularly advanced subsets such as machine learning (ML) and deep learning (DL), has created new opportunities to automate, accelerate, and standardise this process (4). DL models, especially convolutional neural networks (CNNs), are uniquely suited to complex medical image analysis (5). Unlike older algorithmic approaches, CNNs can learn features directly from raw image data, completely eliminating the need for tedious and error-prone manual feature engineering.

Over the past years, a rapidly growing body of scientific literature has evaluated the practicability, reliability, and accuracy of AI-assisted cephalometric analysis (6). The results are highly encouraging, with several commercial and experimental systems approaching or even matching expert-level performance (7). However, the transition from controlled experimental environments to diverse daily clinical practice presents distinct obstacles. This mini-review aims to synthesise the current state of evidence regarding AI applications in cephalometric landmark identification and broader orthodontic treatment planning, to identify key methodological trends, and to highlight remaining clinical and technical challenges.

2. Research materials and methods

This study is based on a narrative review of literature examining the use of Artificial Intelligence in cephalometric analysis, its limitations, and potential future applications in the orthodontic diagnostic process and treatment. The search was limited to publications from January 2016 to April 2026, embodying a full decade of technological development. Studies included original research articles, systematic reviews, meta-analyses, and scoping reviews. Relevant sources were identified using major scientific databases, including PubMed, Scopus, and Google Scholar. Following the screening process, 42 publications were selected based on clinical relevance, methodological quality, and citation impact. []

3. AI Architectures

3.1 Convolutional Neural Networks and YOLO-Based Models

Across the studies reviewed for this paper, one group of algorithms dominates the field - convolutional neural networks. CNNs suit this task well because their layered architecture permits them to learn complex visual patterns such as bone edges, cranial curves, and soft tissue contours directly from raw pixels, without any manual guidance (8). The three most commonly used variants are standard CNNs, YOLO architectures, and Bayesian CNNs.

A review covering 20 key studies published between 2020 and 2025 confirmed this picture, identifying CNN and YOLO as the dominant model types. Their average landmark location errors ranged from 1 to 2 mm, within a clinically acceptable threshold (7). Among these variants, Bayesian CNNs stand out for a particularly useful clinical reason. They do not just produce a result, but also indicate how confident they are in it. When the image quality is poor or anatomical structures are unclear, the model flags its own uncertainty, giving the clinician a cue to review the output more carefully (7). YOLO-based models, on the other hand, earn their place through tremendous speed. By processing the entire image in a single pass, they deliver landmark predictions almost immediately (9). Unlike region-proposal networks, which first locate areas of the image that might contain a landmark and then analyse each of those areas separately, YOLO divides the radiograph into a grid and predicts all landmark locations simultaneously. That procedure significantly cuts the analysis time, playing a practical role in a busy clinical setting.

In terms of accuracy, a meta-analysis of 19 CNN-based studies found that 79.9% of landmarks fell within the 2 mm clinical threshold, establishing a solid performance baseline (10). A study evaluating YOLOv3 found that each subsequent iteration of the architecture handles the subtle grayscale gradients of cephalometric images better than the previous one, with the latest CNN and YOLO models reaching accuracy levels statistically equivalent to the natural variability among human expert examiners (11).

3.2 Attention Mechanisms and Advanced Architectures

More recent models have introduced the idea of incorporating attention mechanisms. Rather than treating every pixel equally, these modules teach the network to prioritise the parts of the image that actually matter, filtering out background noise, artifacts, and irrelevant structures in the process (12). Qiu et al. developed a two-stage attention-based deep learning framework for 3D craniofacial soft tissue landmark detection in orthodontic patients aged 6 to 18 years. It obtained a mean radial error of 2.17 mm on the test set, which is a clinically meaningful result given the complexity of continuously changing paediatric craniofacial morphology (13).

One of the more technically ambitious examples of this approach is the CephTransXnet architecture developed by Neeraja and Jani Anbarasi. It combines the local pattern recognition strengths of CNNs with the broader spatial reasoning of transformer models by processing the entire image through multiple parallel branches to predict landmark coordinates (14).

Meanwhile, Shimamura et al. took the concept of attention-guided analysis in an entirely different direction. They showed that these enhanced CNN models can retrieve meaningful diagnostic information from ordinary lateral facial photographs - effectively enabling landmark estimation without any ionising radiation (15). That point matters more than it might seem because every avoided exposure counts, particularly in younger patients who require repeated assessments over the course of treatment.

4. Performance of AI in 2D Cephalometric Landmark Detection

4.1 Accuracy Metrics

The most common performance metrics reported in the literature are mean radial error (MRE), success detection rate (SDR) at predefined thresholds - typically 2 mm, 3 mm, and 4 mm. The 2 mm threshold is widely considered the gold standard for clinical acceptability in orthodontics (16,17). To illustrate what these numbers look like in practice, a CNN-based algorithm called "CranioCatch," developed by Uğurlu, detected 21 anatomical landmarks on lateral cephalometric radiographs, achieving the highest SDR at the Sella point (98.3% within 2 mm) and the lowest at the Gonion point (48.3% within 2 mm), with a mean radial error of 8.3 mm for the latter (18). The difference in performance is not a coincidence and has a clear anatomical explanation. The high detection rate of the Sella point results from its central location within the cranial base, which provides high radiographic contrast and distinct cortical borders. The Gonion point is highly susceptible to the double-projection effect, in which the left and right mandibular angles do not perfectly overlap. That happens due to facial asymmetry or errors in head positioning. This inconsistency highlights an important point - the accuracy of AI systems depends on radiographic visibility and anatomical complexity of each individual point.

Several studies have attempted to measure how well AI actually performs in real clinical scenarios (10,19). Hendrickx et al. reviewed AI-based landmark detection across both 2D and 3D images. Their collected data confirmed that AI systems can be both accurate and time-efficient when applied to standard 2D lateral cephalograms. The problem arises when algorithms are used across diverse clinical populations, where measurements remain inconsistent and require further development (19). To quantify this, the meta-analysis by Schwendicke et al. found a mean detection error centered around the widely accepted 2 mm clinical threshold (95% CI: -1.264 to 0.102 mm), with an overall success detection rate of 0.799 within 2 mm - a result that is promising, but still leaves roughly one in five landmarks outside the acceptable range (10). It is worth noting that this figure represents an average across all landmarks - and as discussed earlier, performance varies depending on the anatomical complexity of each individual point.

A recent evaluation by S et al. tested a novel software model on 600 cephalometric images using an EfficientNetB7 backbone and reported strong overall landmark detection accuracy. The authors concluded that automated systems can simplify the diagnostic workflow and reduce operator subjectivity whilst complementing rather than replacing clinical expertise (20). This last point recurs consistently throughout the literature and becomes particularly relevant when we turn to the question of how AI compares directly with human performance.

4.2 Comparison with Human Expert Performance

A key question in the clinical setting is whether AI-assisted systems can match or surpass human expert performance. Machines excel at absolute consistency by completely eliminating examiner's errors caused by visual fatigue or subjective bias. Their performance, however, is essentially limited by the quality and diversity of the datasets used to train them.

That said, a review carried out in 2022 concluded that the latest AI algorithms have achieved effectiveness and accuracy levels comparable to those of experienced clinicians for 2D landmark identification (8). This is reinforced by the comparative evaluation by Hwang et al., who showed that automated cephalometric analysis yields measurements statistically equivalent to those of human expert tracings for most clinically relevant landmarks (21). Yet matching human accuracy in a controlled research environment is only part of the story.

Nordblom et al. offered a grounding regulatory observation. Whilst 676 AI- and machine-learning-enabled medical devices had received FDA approval by mid-2023, just one fell within the field of orthodontics. This single number says a lot. Strong results in a research setting clearly do not guarantee that a system is ready for everyday clinical use (22). This difference between research promise and real-life reality becomes even clearer when we look at how these models behave outside the laboratory. A systematic review by Suleman et al. found that whilst AI demonstrates high mean accuracy for landmark detection on standardized datasets, its reliability tends to drop when applied to imaging equipment or patient demographics not represented in the training data (23).

Taken together, these conclusions paint an intricate picture. It is a genuine progress tempered by real-world limitations. As Kiełczykowski et al. noted, we are not quite yet at the point where AI can be fully independent in a clinical setting - but that does not diminish its potential. These systems already show real promise as an assisting tool for diagnosis and treatment planning in everyday orthodontic practice (24).

5. AI in 3D Cephalometric Analysis

5.1 CBCT-Based Landmarking

Three-dimensional cephalometric analysis based on CBCT provides something that 2D imaging cannot - a true volumetric picture of the craniofacial complex. By capturing the actual three-dimensional shape of the skull, CBCT eliminates the problem of bilateral structure superimposition. As a result, clinicians are able to evaluate facial asymmetries, airway volumes, and impacted teeth with a level of accuracy that flat radiographs cannot match. That aspect makes the work much easier for the clinician, but quite the opposite for the algorithm. Where a standard cephalogram consists of a flat grid of pixels, a CBCT scan is built from millions of volumetric voxels, which greatly increases the volume of data and the computational load. Annotating this data for model training is equally demanding. To understand that, we have to consider what the process really requires from the clinician. Unlike marking a point on a flat image, identifying landmarks in 3D means navigating axial, coronal, and sagittal planes simultaneously - a task that is both time-consuming and skill-dependent.

Despite these challenges, researchers have made substantial progress in making deep learning architectures better adapt to the 3D domain. The main strategy has been to train models not on flat pixel grids but directly on volumetric data, forcing the network to simultaneously process all three anatomical planes. Early strategies relied on dividing the 3D scan into smaller pieces, allowing the model to focus on local anatomical regions before assembling them together. More recent methods have moved towards end-to-end architectures that process the entire volume at once. This approach suits CBCT data well, where image boundaries are often gradual rather than sharp, and the same structure can appear different depending on the scanning protocol or patient anatomy.

One of the key barriers historically limiting these models was their tendency to fail when exposed to data from outside their training environment. Sahlsten et al. directly addressed this problem by demonstrating deep learning for 3D cephalometric landmarking using a heterogeneous CBCT dataset comprising 309 scans from Finnish and Thai patients, scanned on multiple devices. They have successfully detected 46 clinically significant landmarks and directly addressed the historical barrier of poor model generalisation across manufacturers with varying field-of-view sizes and voxel dimensions (25). Lang et al. further showed that incorporating anatomical constraints into the model architecture significantly improves CBCT landmark localisation - that points to the importance of embedding clinical knowledge directly into the learning process (26).

To see how these individual findings fit into the bigger picture, a study by Serafin et al. assessed the accuracy of automated 3D cephalometric landmarking using deep learning across studies published between 2020 and 2022. The overall pooled mean error was 2.44 mm, with meta-regression revealing a statistically significant improvement in accuracy over time (p=0.012), suggesting that the field is moving in the right direction, even if it has not yet consistently reached the 2 mm clinical threshold (27).

This improving trajectory is also visible at the level of individual system evaluations. In 2024, Khabadze et al. put the Diagnocat AI system to the test, comparing its output directly against manual tracing performed by specialist orthodontists on 30 consecutive adult CBCT scans. The results were encouraging - for most cephalometric measurements, the two methods agreed within clinically acceptable limits (28). A follow-up study published the following year told a similar story, with Diagnocat and manual tracing again showing high agreement and reproducibility across the same number of scans (29).

5.2 Accuracy Limitations in 3D Settings

Regardless of the advances, important accuracy limitations persist. Polizzi and Leonardi showed that AI consistently exceeds the recommended error margin for most 3D landmarks and tends to perform less consistently on CBCT than on 2D radiographs (30). This finding may sound surprising, but more data does not automatically mean better results. The reason lies in the nature of 3D anatomy itself. Three-dimensional landmarks sit on curved, continuous bone surfaces with no sharp edges, unlike the high-contrast projections that make 2D landmarking more predictable and reliable.

Gracea et al. were clear on this point - AI has not yet consistently matched expert clinician accuracy (31). Matta et al. reinforced this view, identifying domain shift across different scanner types as a fundamental challenge that the field has yet to resolve (32). Nordblom et al. added a regulatory dimension to this picture, noting that most AI applications in orthodontics have not yet reached the level of clinical validation required for regulatory approval and safe unsupervised deployment (22).

With that being said, for all its current limitations, 3D AI landmarking has come a long way in a short time. The question is no longer whether it will get there, but how long it will take.

6. AI in Orthodontic Treatment Planning

6.1 Clinical Decision-Making

Landmark detection is not the only aspect of orthodontic diagnosis where AI is being utilized. It helps clinicians with malocclusion classification, orthognathic surgery prediction, soft tissue analysis, growth assessment, and remote monitoring. This shift from a simple diagnostic tool to a complex, predictive model represents the ultimate range of AI integration.

By developing a dynamic model of the patient, doctors can simulate biomechanical responses and forecast long-term treatment outcomes.

The earlier-mentioned review by Gracea et al. covering 71 studies divided AI applications in orthodontics into three domains - diagnostics, landmark identification, and treatment planning. Treatment planning stood out as the most challenging of the three - not because of a lack of interest, but because of the complexity of the process itself. Predicting the right course of action for an individual patient requires integrating far more variables than simply locating a point on a radiograph. The results were mixed - promising in some areas, inconsistent in others, and held back by the lack of agreed outcome measures. (31).

6.2 AI-Assisted Orthognathic Planning

A domain where AI-assisted planning becomes particularly sophisticated is orthognathic surgery. The procedure involves surgically repositioning craniofacial structures to restore proper occlusal alignment, improve facial aesthetics and create the conditions for successful orthodontic treatment. Getting it right is genuinely challenging. Before a single incision is made, the clinician must evaluate the three-dimensional morphology of the face, jawbones, and dentition through cephalometric analysis and clinical examination. Then they have to determine which surgical moves to make in order to achieve an outcome that is both functionally sound and aesthetically balanced for that specific patient. What makes this particularly demanding is that there are no reliable formulas linking measured craniofacial characteristics to an optimal surgical plan. The decision comes down to the surgeon's experience, their eye for aesthetics, and their ability to picture what the face will look like once the bones have healed. It is, in other words, a task that has historically resisted algorithmic reduction - which is exactly why the progress AI has made in this space is so significant.

Cheng et al. developed a deep learning model trained on three-dimensional CBCT data to predict orthognathic surgery plans. The model was designed to forecast the repositioning vectors (the direction and extent of each osseous segment's movement) required for an individualised surgical plan. The results were encouraging - mean absolute errors of 1.41 mm in the validation set and 1.34 mm in the clinical test set, matching the accuracy of an experienced surgical team (33).

Almarhoumi's scoping review focused on a question patients care deeply about - what will their face look like after surgery? Findings revealed that AI models showed high predictive accuracy across different surgical procedures and facial regions. However, the review also pointed to clear gaps - current datasets are too small and not diverse enough. That is why validation methods are not yet standardised enough to support widespread clinical adoption (34).

6.3 Craniofacial Growth Assessment

Growth prediction is an area where the stakes are particularly high. In growing patients, the timing of orthodontic intervention is often as important as the intervention itself. Intervening too early or too late can compromise treatment stability, increase the risk of relapse, or make a straightforward case considerably more complex. Traditionally, clinicians have relied on a combination of cephalometric analysis, skeletal maturation indicators, and clinical experience to estimate patient's residual growth.

AI is beginning to present genuine clinical value in this procedure. Zhang et al. applied a deep learning model to lateral cephalometric radiographs to predict mandibular growth trends in children with anterior crossbite. It is a clinically important question, since the timing of intervention in these patients can greatly affect treatment outcomes. The model was built on a ResNet50 architecture and trained on 256 cephalograms. It learned to focus on specific anatomical regions: the chin, the lower border of the mandible, the incisor area, the airway, and the condyle. The results were striking. The model achieved 85% accuracy, with a sensitivity of 0.95 and specificity of 0.75 - substantially outperforming junior orthodontists, who reached only 54.2% accuracy with sensitivity and specificity of 0.62 and 0.47, respectively (35).

A complementary perspective comes from Larkin et al., who investigated whether AI could predict short-term craniofacial growth in preadolescent patients with skeletal Class I malocclusion using serial lateral cephalograms. The model was trained on 47 landmarks (28 hard-tissue and 19 soft-tissue points) without incorporating demographic variables such as age or sex. The results demonstrated moderate accuracy for most hard-tissue landmarks, suggesting that serial cephalometric imaging alone carries meaningful predictive information about near-term skeletal development, however, some anatomical regions still require further study for more consistent and precise assessment (36).

Taken together, these studies point in an encouraging direction. Assisted growth prediction algorithms provide a reference point that complements the clinician's own assessment and helps base the treatment timing on something more objective than intuition alone.

6.4 Tooth Movement Planning

Clear aligner therapy has become one of the most popular treatment options in modern orthodontics - and for a good reason. The concept is elegant. A series of custom-made plastic trays, each one nudging the teeth a fraction closer to their intended final position. But behind that apparently simple idea exists a complex planning problem. Every stage of tooth movement must be biologically realistic and sequenced in a way that accounts for individual variation in tooth morphology, bone density, and tissue response. Traditionally, getting this right has depended heavily on clinical experience and a fair amount of calculation. AI is beginning to change the equation.

Olawade et al. described that AI's ability to analyse broad datasets of dental records, radiographs, and 3D scans allows for highly individualised treatment plans, via AI-driven aligners. They are designed to apply optimal forces to teeth, reducing treatment time and patient's discomfort (37). The same review pointed out a growing toolkit of AI architectures being adapted specifically for this purpose, ranging from recurrent neural networks for movement prediction to generative adversarial networks for custom aligner fabrication.

One of the more clinically meaningful developments in this space is the use of AI-powered finite element models. Rather than applying generic biomechanical assumptions, these models simulate how forces will actually distribute across patient's teeth and supporting structures. That allows the system to optimize aligner sequencing before a single tray is printed. Guo and Shao described how this approach can predict root movement trajectories, flag regions of excessive stress, and recommend adjustments that improve efficiency while decreasing the risk of complications (38). That last point matters more than it might seem. Incidents such as root resorption, periodontal damage, and inefficient tooth movement are difficult to reverse once they occur, making early detection far more valuable than damage control.

Beyond the mechanics of tooth movement, AI is also changing how treatment progress is tracked and shared between patient and clinician. Remote monitoring tools now allow patients to check the progress on their own, from home, using their smartphone - meaning fewer trips to the clinic and a treatment experience that fits more naturally into everyday life. Compliance tracking apps, facial scanning software, and AI-assisted progress evaluation tools are increasingly built into commercial aligner platforms, creating an uninterrupted dialogue between what the patient is doing at home and what the clinician sees at the next appointment (37,38).

Despite all its promises, AI-assisted aligner planning is still a work in progress. The biological response to orthodontic forces varies from patient to patient in ways that remain difficult to predict reliably. Current models are not yet accurate enough to cover the full range of malocclusions encountered in daily practice. For that reason, more data, broader datasets, and well-designed clinical trials will be needed before this technology can be fully trusted in daily practice.

7. Clinical Implications and Limitations

7.1 Clinical Promise

The clinical benefits of AI-assisted cephalometric analysis are important. Automating landmark identification saves time at the chair, reduces variability across clinicians tracing the same radiograph, and provides a consistent baseline for tracking growth over time (11). There is also a consistency argument. Manual tracing is subjective - that means two experienced orthodontists analysing the same radiograph can produce different measurements and, ultimately, different treatment plans. AI does not eliminate clinical judgment, but it does give everyone the same starting point. In busy practices and teledentistry settings, the speed advantage is especially important - AI-generated analyses can support same-day case presentations and allow the patient to start the treatment much more quickly (39).

The scale of interest in this technology is hard to ignore. The 10-year mapping review by Sohrabniya et al. comprehensively documented the accelerating pace of DL integration in dentistry, noting that orthodontics has been one of the most prolific fields for AI application, with exponential growth in publications and software validation studies. This positioned orthodontics (10.2%) as the third most prominent field after pathology (21.5%) and radiology (17.5%) (40). That momentum is further confirmed by publication trend data - a 21.6% average annual increase in AI dental publications from 2011 to 2021 and a 34.9% increase per year over the last 5 years, with orthodontics comprising 18.31% of focal topics in their analysis (6).

What was once exclusive to research laboratories is now available on a standard clinical computer or even a smartphone. Platforms such as WeDoCeph, WebCeph, and CephX bring automated landmark identification, cephalometric tracing, and structured measurement reports directly into everyday practice, removing the technical barrier that once separated experimental AI from routine clinical use.

7.2 Persistent Challenge of Racial and Ethnic Differences

Even with these advantages, one limitation deserves particular attention and is becoming more urgent as the world changes. Increased migration and growing ethnic diversity within patient populations mean the gap between the data on which AI models were trained and the populations on which they are now being used is widening. Most available models were built and tested on relatively homogeneous datasets. That severely limits how well they perform outside the groups they were originally designed for (41). Craniofacial norms, skeletal proportions, and bone density vary considerably across ethnic groups. Those differences cause an algorithm trained mainly on one demographic to produce biased landmark predictions when applied to another. Allareddy et al. explicitly identified this generalisation deficit - current AI and ML models are built on datasets that grossly underrepresent historically disadvantaged groups, particularly ethno-racial minorities, and that underrepresentation translates into real diagnostic inequality (42). That is a message to healthcare providers and developers who should establish a more consistent algorithm to reduce racial disparities.

7.3 The Role of the Clinician

AI-driven landmarking still requires the final supervision of an experienced orthodontist (30). This is not simply a cautious disclaimer. AI systems, however accurate they may be, do not understand the patient sitting in the chair, do not account for the clinical history, and cannot evaluate the full complexity of a treatment decision the way an experienced clinician can. When an algorithm places a landmark, it does so without any awareness of whether that prediction will influence a decision to extract a tooth, reposition jaws, or alter the course of a growing child's facial development. Those are irreversible interventions. It means there is no undoing a mistake once it has been made. The orthodontist, not the software, carries that responsibility. That is an ethical reality that no amount of algorithmic accuracy can change. Human control, in other words, is not something to be gradually handed over to machines as they improve. It is the foundation on which safe clinical practice rests, and it belongs to the person who has the expertise and the accountability - to the one who truly understands the patient and has the judgment to know when the machine got it wrong.

8. Conclusions

Artificial intelligence became an effective supporting tool in orthodontic diagnostics. In 2D cephalometric landmark detection, current systems perform within clinically acceptable limits and nearly match the accuracy of experienced clinicians. These advances are not only theoretical. They are embedded in commercial platforms that orthodontists worldwide use on a daily basis. Three-dimensional CBCT-based landmarking is advancing, but accuracy remains inconsistent, validation methods are not yet standardised and datasets are not diverse enough for unsupervised clinical use. AI applications in treatment planning, from orthognathic surgery prediction to growth assessment and aligner sequencing, are showing real promise, but also face the challenges mentioned earlier.

The issue that many of these limitations share is something more than a lack of computing power. The patients walking into clinics today are more diverse than the datasets used to train the models assessing them. Ethnic diversity, domain shift, and the absence of real-life clinical trials are not problems that better algorithms alone can fix. They demand thoughtful effort to build more representative data and test these tools under genuine clinical conditions.

AI should act as a tool supporting the clinician, not one replacing them. What the machine can do is make that decision faster, more precise, and less prone to human error. The orthodontist remains responsible for the final decision, and that part cannot be handed to an algorithm.

Disclosure

Supplementary Materials

Not applicable.

Author's Contribution

Conceptualization: Olgierd Czapiński

Methodology: Olgierd Czapiński, Julia Kalsztein, Aleksandra Wyciślok, Zuzanna Małek, Katarzyna Waśkowska, Weronika Kubiak, Marta Kwiatkowska, Kacper Wiertelak - Makała

Software: Not applicable.

Check: Aleksandra Wyciślok, Olgierd Czapiński, Kacper Wiertelak - Makała

Formal analysis: Weronika Kubiak, Julia Kalsztein, Marta Kwiatkowska, Zuzanna Małek

Investigation: Olgierd Czapiński, Zuzanna Małek, Aleksandra Wyciślok, Adrianna Kępa

Resources: Aleksandra Wyciślok, Olgierd Czapiński, Kacper Wiertelak - Makała

Data curation: Zuzanna Małek, Weronika Kubiak, Katarzyna Waśkowska

Writing-rough preparation: Aleksandra Wyciślok, Adrianna Kępa, Katarzyna Waśkowska, Zuzanna Małek, Julia Kalsztein, Marta Kwiatkowska

Writing review and editing: Olgierd Czapiński, Kacper Wiertelak - Makała, Weronika Kubiak, Marta Kwiatkowska, Katarzyna Waśkowska

Visualization: Aleksandra Wyciślok, Julia Kalsztein, Adrianna Kępa

Supervision: Kacper Wiertelak - Makała, Zuzanna Małek

Project administration: Olgierd Czapiński, Marta Kwiatkowska , Adrianna Kępa

Receiving funding: Not applicable.

All authors have read and agreed with the published version of the manuscript

Funding statement

This research has not received any external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Not applicable.

Acknowledgments

Not applicable.

Conflict of Interest Statement

The authors deny any conflict of interest.

Declaration of generative AI and AI-assisted technologies in the writing process

The authors used Claude (Anthropic) solely for the purpose of improving the linguistic quality and structural organization of the manuscript. The authors maintain full responsibility for the accuracy and integrity of the publication's substantive content.

References

1. Helal NM, Basri OA. *Significance of Cephalometric Radiograph in Orthodontic Treatment Plan Decision. J Contemp Dent Pract*. 2019;20(7):789--93. doi:10.5005/jp-journals-10024-2598

2. Kotuła J, Kuc AE, Lis J, Kawala B, Sarul M. New Sagittal and Vertical Cephalometric Analysis Methods: A Systematic Review. *Diagnostics*. 2022;12(7):1723. doi:10.3390/diagnostics12071723

3. Zaheer R, Shafique HZ, Khalid Z, Shahid R, Jan A, Zahoor T, et al. *Comparison of semi and fully automated artificial intelligence driven softwares and manual system for cephalometric analysis. BMC Med Inform Decis Mak*. 2024;24(1):271. doi:10.1186/s12911-024-02664-3

4. Lamba R. Advances in AI for Medical Imaging: A Review of Machine and Deep Learning in Disease Detection. *Procedia Comput Sci*. 2025;260:262--73. doi:10.1016/j.procs.2025.03.201

5. Mienye ID, Swart TG, Obaido G, Jordan M, Ilono P. Deep Convolutional Neural Networks in Medical Image Analysis: A Review. *Information*. 2025;16(3):195. doi:10.3390/info16030195

6. Thurzo A, Urbanová W, Novák B, Czako L, Siebert T, Stano P, et al. Where Is the Artificial Intelligence Applied in Dentistry? Systematic Review and Literature Analysis. *Healthcare*. 2022;10(7):1269. doi:10.3390/healthcare10071269

7. Zaborowicz K, Zaborowicz M, Cieślińska K, Biedziak B. *Artificial Intelligence Methods in Cephalometric Image Analysis---A Systematic Narrative Review. J Clin Med*. 2026 Mar 3;15(5):1920. doi:10.3390/jcm15051920

8. Junaid N, Khan N, Ahmed N, Abbasi MS, Das G, Maqsood A, et al. Development, Application, and Performance of Artificial Intelligence in Cephalometric Landmark Identification and Diagnosis: A Systematic Review. *Healthcare*. 2022;10(12):2454. doi:10.3390/healthcare10122454

9. Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Internet]. Las Vegas, NV, USA: IEEE; 2016 [cited 2026 Apr 26]. p. 779--88. Available from: http://ieeexplore.ieee.org/document/7780460/ doi:10.1109/CVPR.2016.91 http://ieeexplore.ieee.org/document/7780460/

10. Schwendicke F, Chaurasia A, Arsiwala L, Lee JH, Elhennawy K, Jost-Brinkmann PG, et al. Deep learning for cephalometric landmark detection: systematic review and meta-analysis. *Clin Oral Investig*. 2021;25(7):4299--309. doi:10.1007/s00784-021-03990-w

11. Hwang HW, Park JH, Moon JH, Yu Y, Kim H, Her SB, et al. Automated Identification of Cephalometric Landmarks: *Part 2- Might It Be Better Than human?* Angle Orthod. 2019;90(1):69--76. doi:10.2319/022019-129.1

12. Harsh P, Chakraborty R, Tripathi S, Sharma K. Attention U-Net Architecture for Dental Image Segmentation. In: 2021 International Conference on Intelligent Technologies (CONIT) [Internet]. Hubli, India: IEEE; 2021 [cited 2026 Apr 26]. p. 1--5. Available from: https://ieeexplore.ieee.org/document/9498422/ doi:10.1109/CONIT51480.2021.9498422 https://ieeexplore.ieee.org/document/9498422/

13. Qiu T, Hu C, Zhang J, Wu F, Wang H, Liu X, et al. *Attention-Base deep learning for 3D craniofacial soft tissue landmark detection and diagnosis in orthodontics. Sci Rep*. 2026;16(1):729. doi:10.1038/s41598-025-30383-w

14. Neeraja R, Jani Anbarasi L. CephXNet: A Deep Convolutional Squeeze-and-Excitation Model for Landmark Prediction on Lateral Cephalograms. *IEEE Access*. 2023;11:90780--800. doi:10.1109/ACCESS.2023.3307636

15. Shimamura Y, Tachiki C, Takahashi K, Matsunaga S, Takaki T, Hagiwara M, et al. *Accuracy of cephalometric landmark and cephalometric analysis from lateral facial photograph by using CNN-based algorithm. Sci Rep*. 2024;14(1):31089. doi:10.1038/s41598-024-82230-z

16. Yao J, Zeng W, He T, Zhou S, Zhang Y, Guo J, et al. *Automatic localization of cephalometric landmarks based on convolutional neural network. Am J Orthod Dentofacial Orthop*. 2022;161(3):e250--9. doi:10.1016/j.ajodo.2021.09.012

17. Liu B, Liu C, Xiong Y, Zhu H, Zeng W, Chen J, et al. *Accuracy and reliability of 3D cephalometric landmark detection with deep learning. Eur J Med Res*. 2025;30(1):1000. doi:10.1186/s40001-025-03198-8

18. Department of Oral and Maxillofacial Radiology, Eskisehir Osmangazi University, Faculty of Dentistry, Eskisehir, Turkey, Ugurlu M. *Performance of a Convolutional Neural Network- Based Artificial Intelligence Algorithm for Automatic Cephalometric Landmark Detection. Turk J Orthod*. 2022;94--100. doi:10.5152/TurkJOrthod.2022.22026

19. Hendrickx J, Gracea RS, Vanheers M, Winderickx N, Preda F, Shujaat S, et al. 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

20. S. J, Kurian B, Michael T. Automated Cephalometric Landmark Detection: A Novel Software Model Compared with Manual Annotation Method. *J Indian Orthod Soc*. 2024;58(2):140--9. doi:10.1177/03015742231219540

21. Hwang HW, Moon JH, Kim MG, Donatelli RE, Lee SJ. *Evaluation of automated cephalometric analysis based on the latest deep learning method. Angle Orthod*. 2021;91(3):329--35. doi:10.2319/021220-100.1

22. Nordblom NF, Büttner M, Schwendicke F. Artificial Intelligence in Orthodontics: Critical Review. *J Dent Res*. 2024;103(6):577--84. doi:10.1177/00220345241235606

23. Suleman MU, Mursaleen M, Khalil U, Saboor A, Bilal M, Khan SA, et al. Assessing the generalizability of artificial intelligence in radiology: a systematic review of performance across different clinical settings. *Ann Med Surg*. 2025;87(12):8803--11. doi:10.1097/MS9.0000000000004166

24. Kazimierczak N, Kazimierczak W, Serafin Z, Nowicki P, Nożewski J, Janiszewska-Olszowska J. AI in Orthodontics: Revolutionizing Diagnostics and Treatment Planning---A Comprehensive Review. *J Clin Med*. 2024;13(2):344. doi:10.3390/jcm13020344

25. Sahlsten J, Järnstedt J, Jaskari J, Naukkarinen H, Mahasantipiya P, Charuakkra A, et al. Deep learning for 3D cephalometric landmarking with heterogeneous multi-center CBCT dataset. Isola G, editor. *PLOS ONE*. 2024;19(6):e0305947. doi:10.1371/journal.pone.0305947

26. Lang Y, Wang L, Yap PT, Lian C, Deng H, Thung KH, et al. Automatic Detection of Craniomaxillofacial Anatomical Landmarks on CBCT Images Using 3D Mask R-CNN. In: Zhang D, Zhou L, Jie B, Liu M, editors. Graph Learning in Medical Imaging [Internet]. Cham: Springer International Publishing; 2019 [cited 2026 Apr 27]. p. 130--7. (Lecture Notes in Computer Science). Available from: https://link.springer.com/10.1007/978-3-030-35817-4_16 doi:10.1007/978-3-030-35817-4_16 https://link.springer.com/10.1007/978-3-030-35817-4_16

27. Serafin M, Baldini B, Cabitza F, Carrafiello G, Baselli G, Del Fabbro M, et al. Accuracy of automated 3D cephalometric landmarks by deep learning algorithms: systematic review and meta-analysis. *Radiol Med (Torino)*. 2023;128(5):544--55. doi:10.1007/s11547-023-01629-2

28. Khabadze Z, Mordanov O, Shilyaeva E. *Comparative Analysis of 3D Cephalometry Provided with Artificial Intelligence and Manual Tracing. Diagnostics*. 2024;14(22):2524. doi:10.3390/diagnostics14222524

29. Khabadze ZS, Wehbe A, Mordanov OS. *Comparative analysis of dental parameters within 3D cephalometric analysis using artificial intelligence. Endod Today*. 2025;23(4):511--8. doi:10.36377/ET-0125

30. Polizzi A, Leonardi R. Automatic cephalometric landmark identification with artificial intelligence: An umbrella review of systematic reviews. *J Dent*. 2024;146:105056. doi:10.1016/j.jdent.2024.105056

31. Gracea RS, Winderickx N, Vanheers M, Hendrickx J, Preda F, Shujaat S, et al. Artificial intelligence for orthodontic diagnosis and treatment planning: A scoping review. *J Dent*. 2025;152:105442. doi:10.1016/j.jdent.2024.105442

32. Matta S, Lamard M, Zhang P, Le Guilcher A, Borderie L, Cochener B, et al. *A systematic review of generalization research in medical image classification. Comput Biol Med*. 2024;183:109256. doi:10.1016/j.compbiomed.2024.109256

33. Cheng M, Zhang X, Wang J, Yang Y, Li M, Zhao H, et al. *Prediction of orthognathic surgery plan from 3D cephalometric analysis via deep learning. BMC Oral Health*. 2023;23(1):161. doi:10.1186/s12903-023-02844-z

34. Almarhoumi Aa. Accuracy of Artificial Intelligence in Predicting Facial Changes Post-Orthognathic Surgery: A Comprehensive Scoping Review. *J Clin Exp Dent*. 2024;e624--33. doi:10.4317/jced.61500

35. Zhang JN, Lu HP, Hou J, Wang Q, Yu FY, Zhong C, et al. *Deep learning-based prediction of mandibular growth trend in children with anterior crossbite using cephalometric radiographs. BMC Oral Health*. 2023;23(1):28. doi:10.1186/s12903-023-02734-4

36. Larkin A, Kim J -S., Kim N, Baek S -H., Yamada S, Park K, et al. Accuracy of artificial intelligence-assisted growth prediction in skeletal Class I preadolescent patients using serial lateral cephalograms for a 2-year growth interval. *Orthod Craniofac Res*. 2024 Aug;27(4):535--43. doi:10.1111/ocr.12764

37. Olawade DB, Leena N, Egbon E, Rai J, Mohammed APEK, Oladapo BI, et al. *AI-Driven Advancements in Orthodontics for Precision and Patient Outcomes. Dent J*. 2025;13(5):198. doi:10.3390/dj13050198

38. Guo X, Shao Y. AI-driven dynamic orthodontic treatment management: personalized progress tracking and adjustments---a narrative review. *Front Dent Med*. 2025;6:1612441. doi:10.3389/fdmed.2025.1612441

39. Kaushik R, Rapaka R. AI-driven evolution in teledentistry: A comprehensive overview of technology and clinical applications. *Dent Rev*. 2025;5(2):100154. doi:10.1016/j.dentre.2025.100154

40. Sohrabniya F, Hassanzadeh-Samani S, Ourang SA, Jafari B, Farzinnia G, Gorjinejad F, et al. Exploring a decade of deep learning in dentistry: A comprehensive mapping review. *Clin Oral Investig*. 2025;29(2):143. doi:10.1007/s00784-025-06216-5

41. Celi LA, Cellini J, Charpignon ML, Dee EC, Dernoncourt F, Eber R, et al. Sources of bias in artificial intelligence that perpetuate healthcare disparities---A global review. Fraser HS, editor. *PLOS Digit Health*. 2022;1(3):e0000022. doi:10.1371/journal.pdig.0000022

42. Allareddy V, Oubaidin M, Rampa S, Venugopalan SR, Elnagar MH, Yadav S, et al. *Call for algorithmic fairness to mitigate amplification of racial biases in artificial intelligence models used* *in orthodontics and craniofacial health. Orthod Craniofac Res*. 2023;26(S1):124--30. doi:10.1111/ocr.12721