BMC Oral Health

RESEARCH

Open Access

Reliability and accuracy of Artificial intelligence-based software for cephalometric diagnosis. A diagnostic study

Jean-Philippe Mercier[1*], Cecilia Rossi[2], Iván Nieto Sanchez[3], Inés Díaz Renovales[4], Patricia Martín-Palomino Sahagún[4] and Laura Templier[5]

Abstract

Background Artificial intelligence (AI) is revolutionizing cephalometric diagnosis in orthodontics, streamlining the patient assessments. This study aimed to assess the reliability, accuracy, and time consumption of artificial intelligence (AI)-based software compared to a conventional digital cephalometric analysis method on 2D lateral cephalogram.

Methods 408 lateral cephalometries were analysed using three methods: manual landmark localization, automatic localization, and semi-automatic localization with AI-based software. On each lateral cephalogram, 15 variables were selected, including skeletal, dental, and soft tissue measurements. The difference between the two AI-based software options (automatic and semi-automatic) was compared with the conventional digital technique. The time required to produce a complete cephalometric tracing was evaluated for each method using Student's t-test.

Results Statistically significant differences in the accuracy of landmark positioning were detected among the three different techniques ($p < 0,01$). However, it is noteworthy that almost all of these differences were not clinically significant. There was a small difference in accuracy between the semi-automatic AI-based option and conventional digital techniques. Regarding the time used for each technique, the automatic version was the fastest, followed by the semi-automatic option and the conventional digital technique. ($p < 0,000$).

Conclusions The study showed a statistical difference in accuracy between the conventional digital technique and two AI-based software alternatives, but these differences were not clinically significant except for specific measurements. The semi-automatic option was more accurate than the automatic one and faster than conventional tracing. Further research is needed to confirm AI's accuracy in cephalometric tracing.

Keywords Artificial intelligence, Cephalometry, Software

*Correspondence:
Jean-Philippe Mercier
jphilippemercier@gmail.com
[1]Department of Orthodontics, University of Alfonso X el Sabio, Avenidad de la universidad,1, Villanueva de la Cañada, Madrid 28691, Spain
[2]Clinica Odontoiatrica Lario, Via Strada Statale dei Giovi, 59, Grandate, Come 22070, Italy
[3]University-Hospital San Rafael, Concha Espina 32, Madrid 28016, Spain
[4]Madrid, Spain
[5]Cabinet Templier, 167 rue Camille Desmoulins, Saint Quentin 02100, France

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Introduction

Background

Orthodontic practice requires a comprehensive diagnosis to develop an accurate treatment plan. This process typically involves gathering various types of data, including photographs, radiographs, and study models. Among these, 2D lateral radiography plays a crucial role in cephalometric analysis, which is fundamental for orthodontists [1, 2].

Cephalometric analysis helps classify dental and skeletal malocclusion, determine the severity of orthodontic issues, assess craniofacial structures for orthodontic, implant, and surgical interventions, and evaluate both growth and treatment progress [3].

Historically, cephalometric tracing was performed manually using acetate paper and a protractor, a time-consuming and error-prone process. The advent of digital software has streamlined this process, allowing for the easier manual placement of reference points, thereby improving efficiency [4–7]. Uysal et al. [8] found that both manual and digital tracings demonstrated similar levels of landmark placement error.

However, with advances in technology, new AI-based software has been developed to further expedite cephalometric analysis. These AI software introduce a fully automated system that detects cephalometric landmarks without the need for operator input, identifying all landmarks and generating measurements in just seconds [9, 10].

An additional advantage of AI-based detection is its consistency and reliability. While different operators may interpret cephalometric analyses differently, AI consistently delivers uniform results, thereby improving diagnostic reliability. Moreover, AI systems are trained using large datasets of annotated cephalometric radiographs, enabling the AI to learn from patterns across numerous cases [11–13].

AI presents several unknown aspects, particularly regarding its generalizability across different populations, as it may not perform equally well across all age groups and ethnicities as the investment in AI research are mainly done in Asia [12]. In managing and interpreting complex cases, nuanced clinical judgment may still be required. Additionally, ethical and legal concerns, such as data privacy and clinician collaboration, remain areas of active investigation [12, 14].

Despite these challenges, AI shows great promise for the future, especially if its accuracy continues to meet acceptable standards. The acceleration of cephalometric analysis could significantly streamline the treatment planning process for orthodontists [10]. Nevertheless, research on this topic is currently limited, with available studies often constrained by small sample sizes. As AI software evolves, further investigations will be essential to validate its effectiveness and broaden its clinical application.

Objectives

Therefore, this study aims to assess the reliability, accuracy, and time consumption of AI-based software compared to a conventional digital cephalometric analysis method on 2D lateral cephalogram.

Materials and methods

Study design

This diagnostic study followed the Standards for Reporting Diagnostic accuracy studies (STARD) [15] protocol and was carried out in compliance with the Declaration of Helsinki and was approved by the ethics committee of the university with the following reference: 2023_03/184.

Participants

The inclusion criteria for this study were limited to lateral cephalogram of patients from the orthodontic clinic of the university with mixed or permanent dentitions. Patients with syndromes, very poor quality, and misdirected lateral cephalogram were excluded to prevent any issues in measurements results. Another exclusion criterion was lateral cephalograms that were not recognized by the AI software.

Patients' lateral cephalogram were randomly collected retrospectively at the university between September 2022 until may 2023 from the radiograph archives by the main operator.

Test methods

Different students performed cephalograms using the same X-ray machine (Satelec X-Mind Pano Ceph D+) and were traced by a single operator. All the eligible participants were assessed with the 3 diagnostic methods (digital manual method, semi-automatic and automatic technique). The digital manual method, regarded as the gold standard for the study, involved manual placement of landmarks. Whereas with AI-based software, the cephalometric points were localized automatically by the software over the cephalometry or semi-automatically, modifying the previous automatic tracing by the operator, allowing the good placement of the landmarks over the studied lateral cephalometry.

The gold standard (the manual method) use for tracing

Digital cephalometric tracing using NemoCeph^{Tm} involved several steps. First, the patient's information and the digital cephalogram were uploaded into the software, ensuring proper orientation and scaling. Next, the user identified and marked key anatomical landmarks, using the program's tools. The software then assisted in automatically generating cephalometric measurements based

on these landmarks. After this, the user could refine the placement of points and adjust the tracing as needed to ensure accuracy. Finally, the software provided an analysis of the traced cephalogram, offering diagnostic measurements and comparisons to normative data [16].

Employment of the AI systems

In the fully automated AI version, the process started with uploading a lateral cephalometric radiograph to the platform. The AI used by Webceph^{Tm} operates through deep learning algorithms that have been trained on large datasets of annotated cephalometric images. These neural networks were designed to recognize anatomical patterns, enabling the AI to detect and trace key cephalometric landmarks. The system used a multi-step process that involved feature extraction, classification, and prediction to locate the landmarks. The AI cross-referenced the radiograph with its vast database, ensuring accuracy by minimizing variability in landmark identification. Once these points were detected, the software automatically calculated a range of cephalometric parameters both angular and linear and produces a detailed report for diagnosis and treatment planning [11, 17].

In the semi-automatic version, the workflow was initially the same as the fully automatic version. However, this version allowed the operator to manually review and adjust the positions of the identified points. In cases where landmarks were misidentified or inaccurately placed, users could intervene by manually repositioning these points, creating a new analysis report [11, 17].

Analysis: the landmarking and tracing procedures along with the sample size and statistical analyses

This study used different cephalometries for the analysis taking relevant measurements. From Steiner, we employed SNA (°), SNB (°), ANB (°), SN-mandibular plane (°), superior incisor position (U1)-NA (mm), superior incisor inclination (U1)-NA (°) inferior incisor position (L1)-NB (mm), inferior incisor position (L1)-NB (°) and Wits (mm).

From Jarabak, the Sum of Björk-Jarabak angles (°) and IMPA (°) were used.

The Ricketts measurements were also evaluated, including the position of the upper incisor (U1)-APg (mm), the position of the lower incisor (L1)-APg (mm), the inclination of the upper incisor (1)-APg (°), and the labial position in relation to the Ricketts aesthetic plane (mm). All the landmarks used were summarised in the Table 1 [1, 18].

The main operator also measured the time using a chronometer from the start to the end of landmark colocation for each technique.

To assess the intra-examiner and inter-examiner reliability, a 10% sample (41 cephalometries) were analysed a second time by the main operator, a 2nd year orthodontic student and a second operator, a University professor with 6 years of experience.

Bias

To avoid bias in the tracing, the operator measured the lateral cephalograms with NemoCeph™ (conventional digital technique), starting with even numbers, while the odd numbers were initially analysed with WebCeph™ (AI-based software). An average of 40 tracings per day were followed for each method under the same conditions to avoid lack of concentration and fatigue.

Study size

For the calculation of the sample size, the article by Yassir YA. et al. [19] was used, taking as reference the inclination of the lower incisor (L1-MP). Consequently, we conducted a one-sided test with an alpha risk of 0,05 and a beta risk of 0,2 as references. In order to detect a difference equal to or greater than one degree for angular variable or one millimetre for linear variable, we needed a sample size of 408 subjects in the conventional digital group, 408 in the AI-based automatic group, and 408 in the AI-based semi-automatic group. It was assumed that the average standard deviation was 5,74. The follow-up loss rate was estimated to be 0%.

Statistical methods

Statistical analysis was performed using SPSS (version 25, SPSS Inc., Armonk, NY, USA). Quantitative variables were tested for normality using skewness, kurtosis, and the Kolmogorov-Smirnov test. A significant deviation was considered when $p<0.01$.

Quantitative variables were described using mean, median, standard deviation, and interquartile range. Accuracy was assessed by calculating the mean absolute error, with the conventional method as the reference. Differences in measurements were evaluated using the Wilcoxon paired-sample test and Student's t-test for paired measurements. The Friedman test was applied to compare measurement times across techniques.

Intra-examiner and inter-examiner reliability

The inter- and intra-operator reliability was calculated to assess the agreement among replicas performed by the primary operator (intra-operator agreement) and between measurements made by both operators (inter-operator agreement).

Results

Participants

After removing the 25 excluded lateral cephalograms, the final sample of the study consisted of 408 lateral

Table 1 Landmark used for the cephalometric tracing

Sella (S)	Localized in the center of the hypophyseal fossa (Sella Tursica)
Nasion (Na)	Represents the most posterior point on the curvature of the bridge of the nose at the junction of the nasal and frontal bones.
A-point (A)	Defined as the point on the innermost curvature from the maxillary anterior nasal spine of the crest of the maxillary alveolar process.
B-point (B)	It's the point on the anterior bony curvature of the mandible. B point = Innermost point of the curvature from chin to alveolar junction.
Condylion (Co)	Most posterosuperior point on the curvature of the condylar head.
Gonion (Go)	Located at the midpoint between the posterior point and the lower point of the angle of the mandible.
Pterygomaxillary Fissure (Ptm)	Halfway point between the most superior point and the most posterior point of the image of the pterygomaxillary fossa.
Gnathion (Gn)	Point on the profile curvature of the mandibular symphysis, the most external and everted point, localized midway between pogonion and menton.
Articulare (Ar)	Point located on the posterior borders of the mandibular rami at the intersection with the basilar portion of the occipital bone.
Porion (Po)	Point of the upper contour of the external auditory canal (Anatomic Porion).
Orbitale (Or)	The lowest point on the inferior margin of the orbits.
Menton (Me)	The lowest point of the symphysis of the mandible.
Anterior Nasal Spine (ANS)	The most anterior point on the maxilla at the nasal base.
Posterior Nasal Spine (PNS)	Tip of the posterior nasal spine located on the palatine bone, at the junction of the soft and hard palate.
Basion (Ba)	Anterior margin of the foramen magnum. Midpoint of the curvature between upper and lower surfaces of the basilar portion of the occipital bone.
Distal of maxillary first molar (A6)	Horizontal position of the upper first molar.
Distal of mandibular first molar (B6)	Horizontal position of the lower first molar.
Wits (Witwatersrand University, Johannesburg, South Africa)	Assessment of sagittal jaw relationship by relating A and B points to the occlusal plane
Occlusal plane	Line passing through one-half the cusp height of the permanent first molars and one-half the overbite of the incisors.
Upper incisor apex (UIA)	Maxillary central incisor root tip.
Upper incisor incisal edge (UIE)	Maxillary central incisor incisal tip.
Lower incisor apex (LIA)	Mandibular central incisor root tip.
Lower incisor incisal edge (LIE)	Mandibular central incisor incisal tip.
Pronasale (Pn)	The most prominent midline points on the tip of the nose.
Soft tissue pogonion (Pg')	The most anterior point on the chin in the midsagittal plane.
Glabella (G)	Point on the curvature of the bone overlying the frontal sinus.
Protuberance menti (suprapogonion) (Pm)	Point at the anterior border of the symphysis between B-point and pogonion where the curvature changes from concave to convex.

cephalograms (Fig. 1), of which 40.9% were male and 59.1% female, with a mean age of 19.77 years.

To determine the skeletal class of patients, the Steiner's measurements and norms were used (SNA, SNB and ANB). We observed that the majority of patients had malocclusion, as only 34.6% were classified as skeletal class I. Moreover, 56.1% of the sample exhibited skeletal class II, while 9.3% presented skeletal class III.

In the same way, to identify the patients' biotype, the Steiner's mandibular plan (SN- Mandibular plane) and the Björk-Jarabak angle sum were also employed. The most frequent biotype was mesofacial, accounting for 59.5% of the patients, followed by brachyfacial at 34.2%, and finally, dolichofacial at 6.4%.

All of these results were summarized in Table 2.

Test results

Accuracy

By calculating the accuracy, it was possible to determine how closely the measurements realized with each technique were to the true value. In this case, we considered the measurements made with the conventional digital option as the true or reference values.

Therefore, to calculate accuracy, we determined the differences between each measurement of the conventional digital technique with the automatic tracing (Webceph auto) and with the semiautomatic option (Webceph semiauto), where the previous automatic landmarks placement was modified manually. Then, these differences were averaged for each type of measurement, resulting in the accuracy values presented in Table 3.

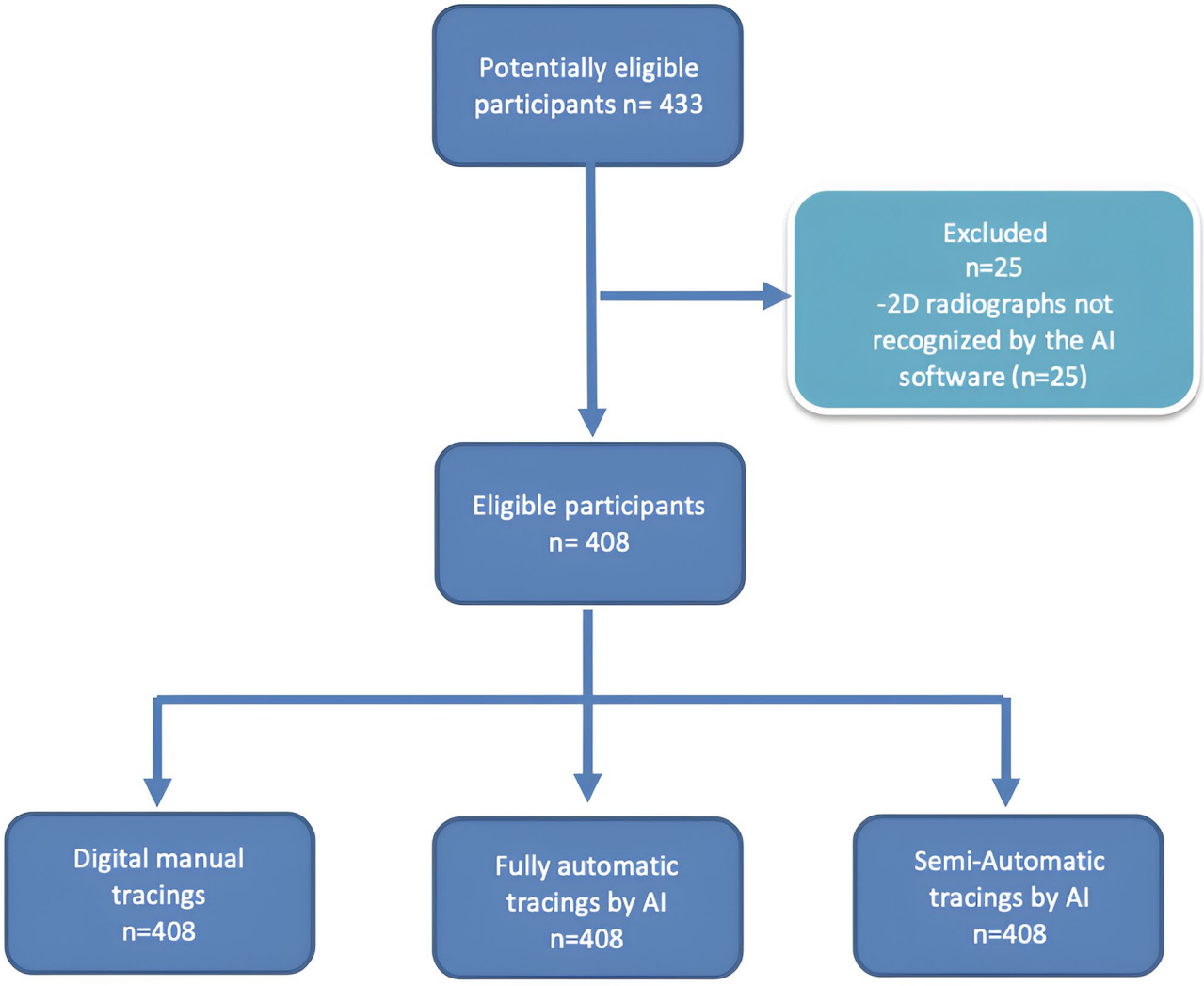


Fig. 1 Flow of participants. Diagram of the flow of participants following the STARD checklist

Table 2 Frequency table. Variables: sex, skeletal class (ANB) and Biotype (sum of angles)

Category	Abs. freq.	Cumulative abs. freq.	Relative freq.	Cumulative rel. freq.
Sex				
1. Male	167	167	40.9%	40.9%
2. Female	241	408	59.1%	100.0%
Skeletal class (ANB)				
1. Skeletal class I	141	141	34.6%	34.6%
2. Skeletal class II	229	370	56.1%	90.7%
3. Skeletal class III	38	408	9.3%	100.0%
Biotype (sum of angles)				
1. Dolichofacial	26	26	6.4%	6.4%
2. Mesofacial	242	268	59.5%	65.8%
3. Brachyfacial	139	407	34.2%	100.0%

Results observation showed that AI-based semi-automatic option exhibited higher accuracy in most variables, surpassing its automatic alternative for the following parameters: SNA (°), SNB (°), ANB (°), Björk-Jarabak angle sum (°), Wits (mm), (U1)-NA (°), (L1)-NB (°), (L1)-APg (mm), (U1)-APg (°), and IMPA (°).

However, AI-based automatic option was more accurate for the variables: SN-MP (°), (U1)-NA (mm), (L1)-NB (mm), and lower lip to E-plane (mm).

Error type description in cephalometric measurements in AI-based automatic option The AI-based automatic method yielded significantly higher measurements compared to the conventional technique for the

Table 3 Accuracy of Webceph Auto and Semiauto compared to the conventional digital technique

Variable	Webceph auto		Webceph semiauto	
	Trueness	SE	Trueness	SE
SNA (°)	1.037	0.118	0.062	0.075
SNB (°)	0.367	0.097	0.068	0.063
ANB (°)	0.660	0.061	-0.018	0.041
SN-mandibular plane (°)	0.104	0.138	0.612	0.091
Bjork-Jarabak's sum of angles (°)	-0.817	0.140	-0.402	0.086
Wits (mm)	1.169	0.090	0.441	0.070
Position of the upper incisor (U1)-NA (mm)	-0.345	0.094	0.363	0.067
Inclination of the upper incisor (U1)-NA (°)	-1.727	0.205	-0.204	0.150
Position of the lower incisor (L1)-NB (mm)	-0.069	0.044	0.271	0.031
Inclination of the lower incisor (L1)-NB (°)	-2.414	0.189	-1.505	0.164
Protrusion of the upper incisor (U1)-APg (mm)	-0.156	0.046	0.161	0.032
Protrusion of the lower incisor (L1)-APg (mm)	-0.555	0.045	0.217	0.032
Inclination of the upper incisor (U1)-APg (°)	-0.632	0.177	-0.065	0.139
IMPA (°)	-1.615	0.194	-0.912	0.164
Lower lip to E-plane (mm)	0.311	0.062	0.333	0.049

variables SNA (°), SNB (°), ANB (°), Wits (mm), and lower lip to E-plane (mm). Conversely, it produced significantly lower values than the conventional technique for the following variables: Björk-Jarabak angle sum (°), (U1)-NA (mm and °), (L1)-NB (°), (U1)-APg (mm and °), (L1)-APg (mm), and IMPA (°). All of these differences were statistically significant ($p<0.010$) (Table 4).

Examining the 95% confidence interval revealed that the automatic version of the AI-based software tended to underestimate or overestimate the results, depending of the measurements. (Table 5)

Error type description in cephalometric measurements in AI-based semi-automatic option For the semi-automatic version, significantly higher measurements were observed compared to the conventional technique for the following variables: SN-mandibular plane (°), Wits (mm), (U1)-NA (mm), (L1)-NB (mm), (U1)-APg (mm), (L1)-APg (mm), and lower lip to E-plane (mm). However, it showed significantly lower values for the Björk-Jarabak angle sum (°), (L1)-NB (°), and IMPA (°) compared to the conventional digital method. As previously mentioned, all of these differences were statistically significant ($p<0.010$) (Table 4).

Therefore, a tendency to overestimate most measurements was observed in the semi-automatic version of the AI-based software, observing the 95% confidence interval. (Table 5)

Error clinically significant

In cephalometric orthodontics, it is crucial to distinguish between statistical significance and clinical significance. A statistically significant difference indicates that a change is unlikely due to chance, but it does not necessarily mean it has a real impact on patient treatment. Clinical significance, on the other hand, refers to results that have practical implications and truly improve functional or aesthetic outcomes for the patient. Therefore, in orthodontic research and practice, cephalometric changes must be not only statistically significant but also clinically relevant to influence treatment decisions [20, 21].

In the study, measurements were taken using two clinical acceptance criteria (2° or mm and 1° or mm), and results were calculated based on these criteria when statistically significant.

For the AI-based automatic option, a significant clinical difference was observed in L1-NB (°) with a clinical acceptance threshold of 2° or 2 mm. When using a clinical acceptance threshold of 1° or 1 mm, clinically significant differences were found in SNA (°), Wits (mm), (U1)-NA (°), (L1)-NB (°), and IMPA (°). (Table 6)

For the AI-based semi-automatic option, a clinical acceptance threshold of 2° or 2 mm did not reveal any significant clinical differences. In contrast, when using a threshold of 1° or 1 mm, only (L1)-NB (°) demonstrated clinically significant differences. (Table 6)

Time management

The time required to perform the cephalometric tracings was compared for each device.

The results were highly significant and confirmed the existence of a difference in the time consumption of the different tracing methods: AI-based automatic option was the fastest, followed by AI-based semi-automatic option, and finally, the conventional option ($p<0,000$) (Fig. 2).

Table 4 Results of the parametric hypothesis testing. Variables: cephalometric measurements. Segmented by: measurement type. Test: paired-samples T test

Variable	Mean dif.	SE	CI 95%		Statistic	*p*-value	Result
			Lím. Inf.	Lím. sup.			
Nemoceph vs. Webceph auto							
SNA (°)	1.037	0.118	0.806	1.268	**8.816****	**0.000**	**Significant**
SNB (°)	0.367	0.097	0.177	0.558	**3.792****	**0.000**	**Significant**
ANB (°)	0.660	0.061	0.540	0.780	**10.818****	**0.000**	**Significant**
SN-mandibular plane (°)	0.104	0.138	-0.166	0.375	.758^{NS}	0.449	Not significant
Bjork-Jarabak's sum of angles (°)	-0.817	0.140	-1.091	-0.542	**-5.851****	**0.000**	**Significant**
Wits (mm)	1.169	0.090	0.992	1.347	**12.949****	**0.000**	**Significant**
Position of the upper incisor (U1)-NA (mm)	-0.345	0.094	-0.529	-0.160	**-3.666****	**0.000**	**Significant**
Inclination of the upper incisor (U1)-NA (°)	-1.727	0.205	-2.131	-1.323	**-8.411****	**0.000**	**Significant**
Position of the lower incisor (L1)-NB (mm)	-0.069	0.044	-0.156	0.018	-1.567^{NS}	0.118	Not significant
Inclination of the lower incisor (L1)-NB (°)	-2.414	0.189	-2.785	-2.043	**-12.791****	**0.000**	**Significant**
Protrusion of the upper incisor (U1)-APg (mm)	-0.156	0.046	-0.246	-0.066	**-3.395****	**0.001**	**Significant**
Protrusion of the lower incisor (L1)-APg (mm)	-0.555	0.045	-0.644	-0.465	**-12.203****	**0.000**	**Significant**
Inclination of the upper incisor (U1)-APg (°)	-0.632	0.177	-0.980	-0.283	**-3.565****	**0.000**	**Significant**
IMPA (°)	-1.615	0.194	-1.996	-1.234	**-8.325****	**0.000**	**Significant**
Lower lip to E-plane (mm)	0.311	0.062	0.190	0.432	**5.059****	**0.000**	**Significant**
Nemoceph vs. Webceph semiauto							
SNA (°)	0.062	0.075	-0.085	0.209	.828^{NS}	0.408	Not significant
SNB (°)	0.068	0.063	-0.055	0.191	1.081^{NS}	0.281	Not significant
ANB (°)	-0.018	0.041	-0.099	0.062	– .446^{NS}	0.656	Not significant
SN-mandibular plane (°)	0.612	0.091	0.434	0.791	**6.739****	**0.000**	**Significant**
Bjork-Jarabak's sum of angles (°)	-0.402	0.086	-0.572	-0.232	**-4.655****	**0.000**	**Significant**
Wits (mm)	0.441	0.070	0.305	0.578	**6.350****	**0.000**	**Significant**
Position of the upper incisor (U1)-NA (mm)	0.363	0.067	0.231	0.495	**5.409****	**0.000**	**Significant**
Inclination of the upper incisor (U1)-NA (°)	-0.204	0.150	-0.499	0.091	-1.357^{NS}	0.176	Not significant
Position of the lower incisor (L1)-NB (mm)	0.271	0.031	0.210	0.331	**8.752****	**0.000**	**Significant**
Inclination of the lower incisor (L1)-NB (°)	-1.505	0.164	-1.827	-1.182	**-9.180****	**0.000**	**Significant**
Protrusion of the upper incisor (U1)-APg (mm)	0.161	0.032	0.098	0.223	**5.055****	**0.000**	**Significant**
Protrusion of the lower incisor (L1)-APg (mm)	0.217	0.032	0.154	0.280	**6.783****	**0.000**	**Significant**
Inclination of the upper incisor (U1)-APg (°)	-0.065	0.139	-0.338	0.207	– .472^{NS}	0.638	Not significant
IMPA (°)	-0.912	0.164	-1.235	-0.590	**-5.560****	**0.000**	**Significant**
Lower lip to E-plane (mm)	0.333	0.049	0.236	0.430	**6.767****	**0.000**	**Significant**

^{NS}*Not significant, * Significant (p < 0.05) and ** Significant (p < 0.01)*

Table 5 Error type of AI-based Automatic and semi-automatic options

	Overestimation	Subestimation
AI-based Automatic version	SNA (°) SNB (°) ANB (°) Wits (mm) Lower lip to E-plane (mm)	Bjork-Jarabak's sum of angles (°) (U1)-NA (mm) (U1)-NA (°) (L1)-NB (°) (U1)-APg (mm) (L1)-APg (mm) (U1)-APg (°) IMPA (°)
AI-based Semi-Automatic version	SN-MP (°) Wits (mm) (U1)-NA (mm) (L1)-NB (mm) (U1)-APg (mm) Lower lip to E-plane (mm)	Bjork-Jarabak's sum of angles (°) (L1)-NB (°) IMPA (°)

Reliability analysis

Intra-operator reliability The main operator repeated 10% of the total sample (41 radiographs for each technique) in order to determine the intra-operator reliability.

For all the variables, an Intra class correlation coefficient (ICC) greater than 0,940 was obtained. This value corresponded to very high level of agreement, indicating that the measurements by the primary operator were highly reliable (Table 7).

Inter-operator reliability Similarly, in order to analyse the inter-operator reliability of the study, an independent operator also conducted measurements on the same 10% of the sample.

Inter class correlation coefficient was calculated as a measure of agreement between the main and the

Table 6 Accuracy of AI-based software – automatic and semi-automatic option

	Statistically significant differences	Clinical acceptance Within 2° or 2 mm	Clinical acceptance Within 1° or 1 mm
AI-based Automatic method			
SNA (°)	Significant	Acceptable	None
SNB (°)			Acceptable
ANB (°)			
SN-mandibular plane (°)	Not significant	Not available	Not available
Bjork-Jarabak's sum of angles (°)	Significant	Acceptable	Acceptable
Wits (mm)			None
Position of the upper incisor (U1)-NA (mm)			Acceptable
Inclination of the upper incisor (U1)-NA (°)			None
Inclination of the lower incisor (L1)-NB (°)		None	None
Position of the lower incisor (L1)-NB (mm)	Not significant	Not available	Not available
Protrusion of the upper incisor (U1)-APg (mm)	Significant	Acceptable	Acceptable
Protrusion of the lower incisor (L1)-APg (mm)			
Inclination of the upper incisor (U1)-APg (°)			
IMPA (°)			None
Lower lip to E-plane (mm)			Acceptable
AI-based Semi-Automatic method			
SNA (°)	Not significant	Not available	Not available
SNB (°)	Not significant	Not available	Not available
ANB (°)	Not significant	Not available	Not available
SN-mandibular plane (°)	Significant	Acceptable	Acceptable
Bjork-Jarabak's sum of angles (°)			
Wits (mm)			
Position of the upper incisor (U1)-NA (mm)			
Inclination of the upper incisor (U1)-NA (°)	Not significant	Not available	Not available
Inclination of the lower incisor (L1)-NB (°)	Significant	Acceptable	None
Position of the lower incisor (L1)-NB (mm)			Acceptable
Protrusion of the upper incisor (U1)-APg (mm)			
Protrusion of the lower incisor (L1)-APg (mm)			
Inclination of the upper incisor (U1)-APg (°)	Not significant	Not available	Not available
IMPA (°)	Significant	Acceptable	Acceptable
Lower lip to E-plane (mm)			

independent operator results. The results were showed in Table 8.

In all measured variables, the Inter class correlation coefficient was above 0,900, suggesting excellent inter-operator reliability for the experiment.

Discussion

Key results and interpretation

Previous studies have conducted a comparative analysis between digital and manual tracing techniques in cephalometry. The findings of these studies have indicated that the use of technology, such as digital software for cephalometric tracing, reduces landmark localization errors [22].

In this current study, the conventional digital option was taken as the gold standard for the results, as other studies have previously demonstrated its accuracy for cephalometric tracing [23, 24]. Additionally, it is important to mention that WebCeph™ was chosen for the study as it has been recently introduced and accepted for orthodontic diagnosis by the Food and Drug Administration (FDA) and the Korean Food and Drug Administration (KFAD) [25]. This software is based on machine learning AI, trained on dataset of craniofacial images, records, and associated diagnoses or treatments [26].

The results of this study showed that AI-based semi-automatic (Fig. 3) and AI-based automatic options (Fig. 4) presented statistically significant differences with the conventional digital option (Fig. 5), but most of them were not clinically significant. This is in accordance with other studies from Tsolakis et al., Ye et al., and Hwang et al. [27–29] that demonstrated that AI led to reliable cephalometric measurements. However, Schwendicke et al. [30] pointed out that landmark detection based on deep learning still lacks robustness. They have noted that there is limited evidence due to a high risk of bias, although the method has shown high accuracy. The systematic review

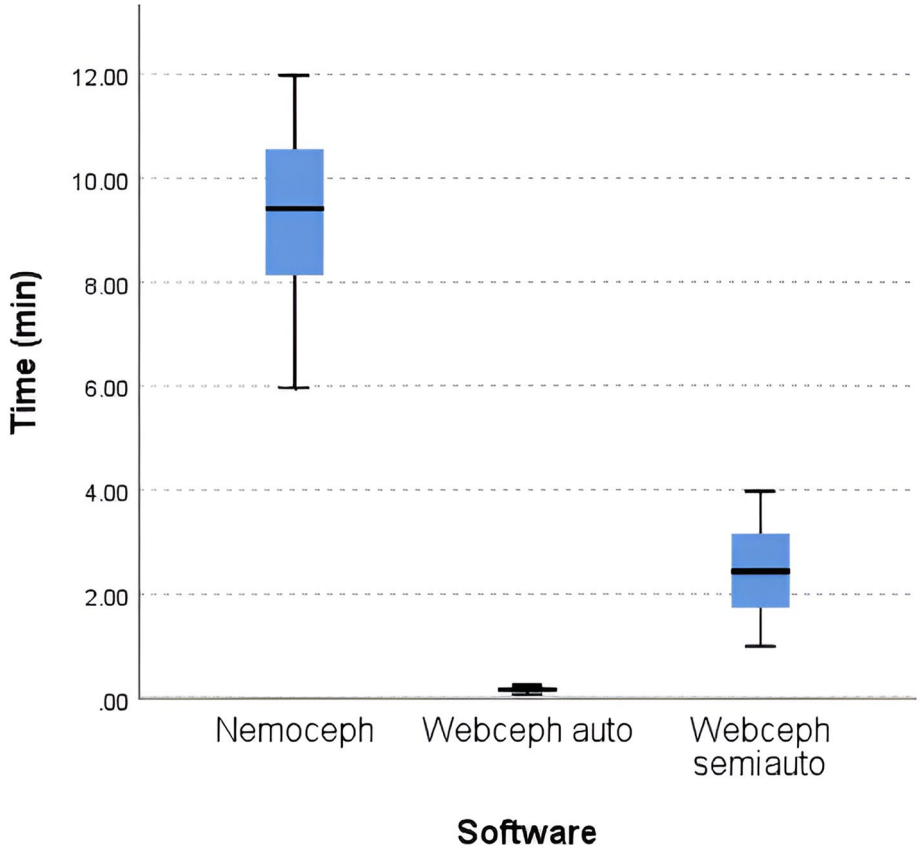


Fig. 2 Boxplots. Variable: Time. Segmented by: Software. A comparison was made of the tracing time required among the three different software options

Table 7 Intraclass correlation coefficients (ICCs). Variables: cephalometric measurements. Segmented by: measurement type. Groups based on: replicate

Variable	Single measures		Average measures	
	ICC	Agreement	ICC	Agreement
SNA (°)	0.971	High	0.985	High
SNB (°)	0.978	High	0.989	High
ANB (°)	0.976	High	0.988	High
SN-mandibular plane (°)	0.985	High	0.992	High
Bjork-Jarabak's sum of angles (°)	0.977	High	0.989	High
Wits (mm)	0.961	High	0.980	High
Position of the upper incisor (U1)-NA (mm)	0.961	High	0.980	High
Inclination of the upper incisor (U1)-NA (°)	0.933	High	0.965	High
Position of the lower incisor (L1)-NB (mm)	0.988	High	0.994	High
Inclination of the lower incisor (L1)-NB (°)	0.962	High	0.980	High
Protrusion of the upper incisor (U1)-APg (mm)	0.993	High	0.996	High
Protrusion of the lower incisor (L1)-APg (mm)	0.984	High	0.992	High
Inclination of the upper incisor (U1)-APg (°)	0.957	High	0.978	High
IMPA (°)	0.944	High	0.971	High
Lower lip to E-plane (mm)	0.973	High	0.986	High

by Junaid et al. [31] and the study of Le et al. [32] indicated that the deep anatomical context feature learning (DACFL) model could even help students detect anatomical landmarks.

Error type description in cephalometric measurements

AI-based automatic version

Observing the 95% confidence interval, the study indicated that the automatic AI-based software tended to

Table 8 Interclass correlation coefficients (ICCs). Variables: cephalometric measurements. Segmented by: measurement type. Groups based on: Operator

Variable	Single measures		Average measures	
	ICC	Agreement	ICC	Agreement
SNA (°)	0.952	High	0.975	High
SNB (°)	0.966	High	0.983	High
ANB (°)	0.959	High	0.979	High
SN-mandibular plane (°)	0.970	High	0.985	High
Bjork-Jarabak's sum of angles (°)	0.955	High	0.977	High
Wits (mm)	0.924	High	0.960	High
Position of the upper incisor (U1)-NA (mm)	0.940	High	0.969	High
Inclination of the upper incisor (U1)-NA (°)	0.917	High	0.957	High
Position of the lower incisor (L1)-NB (mm)	0.981	High	0.990	High
Inclination of the lower incisor (L1)-NB (°)	0.948	High	0.973	High
Protrusion of the upper incisor (U1)-APg (mm)	0.989	High	0.994	High
Protrusion of the lower incisor (L1)-APg (mm)	0.983	High	0.991	High
Inclination of the upper incisor (U1)-APg (°)	0.953	High	0.976	High
IMPA (°)	0.906	High	0.951	High
Lower lip to E-plane (mm)	0.970	High	0.985	High

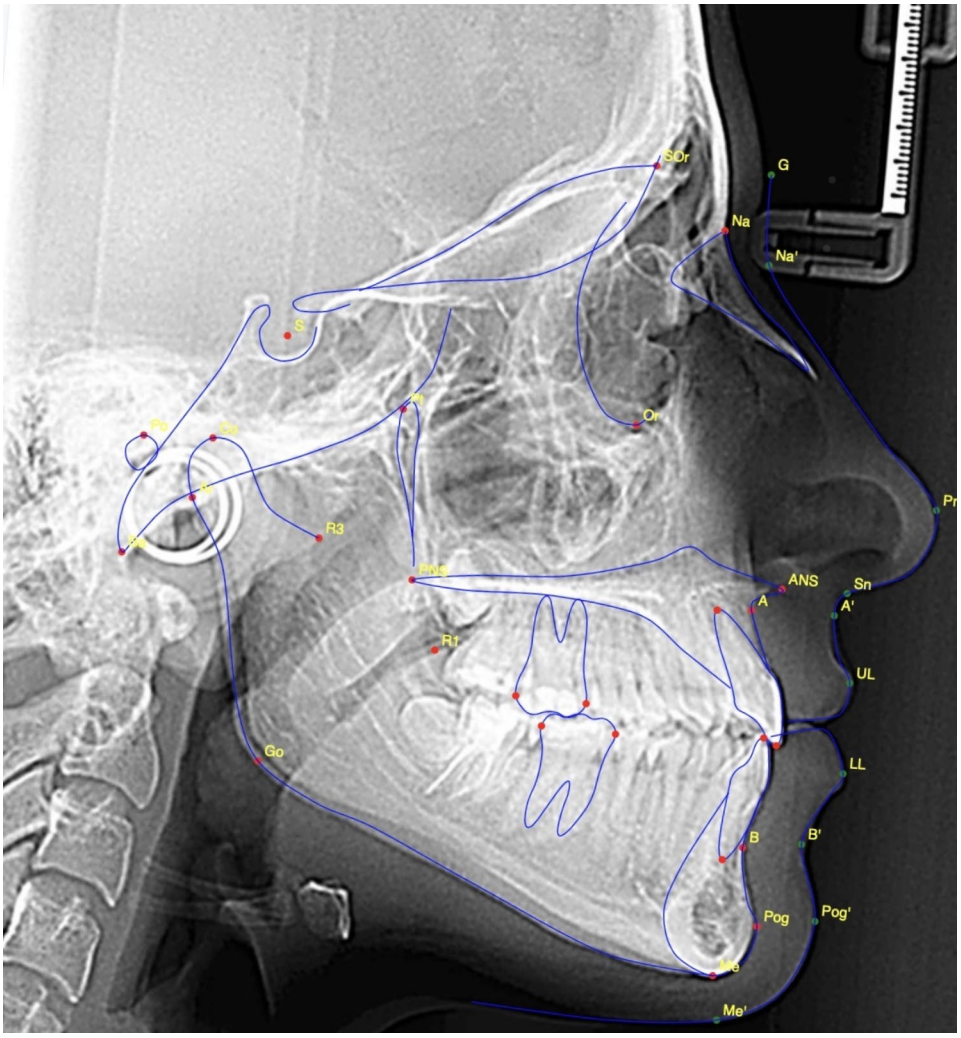


Fig. 3 Cephalometric tracing with AI-based software semi-Automatic option. Cephalometric tracing following manual adjustment of the automatic tracing

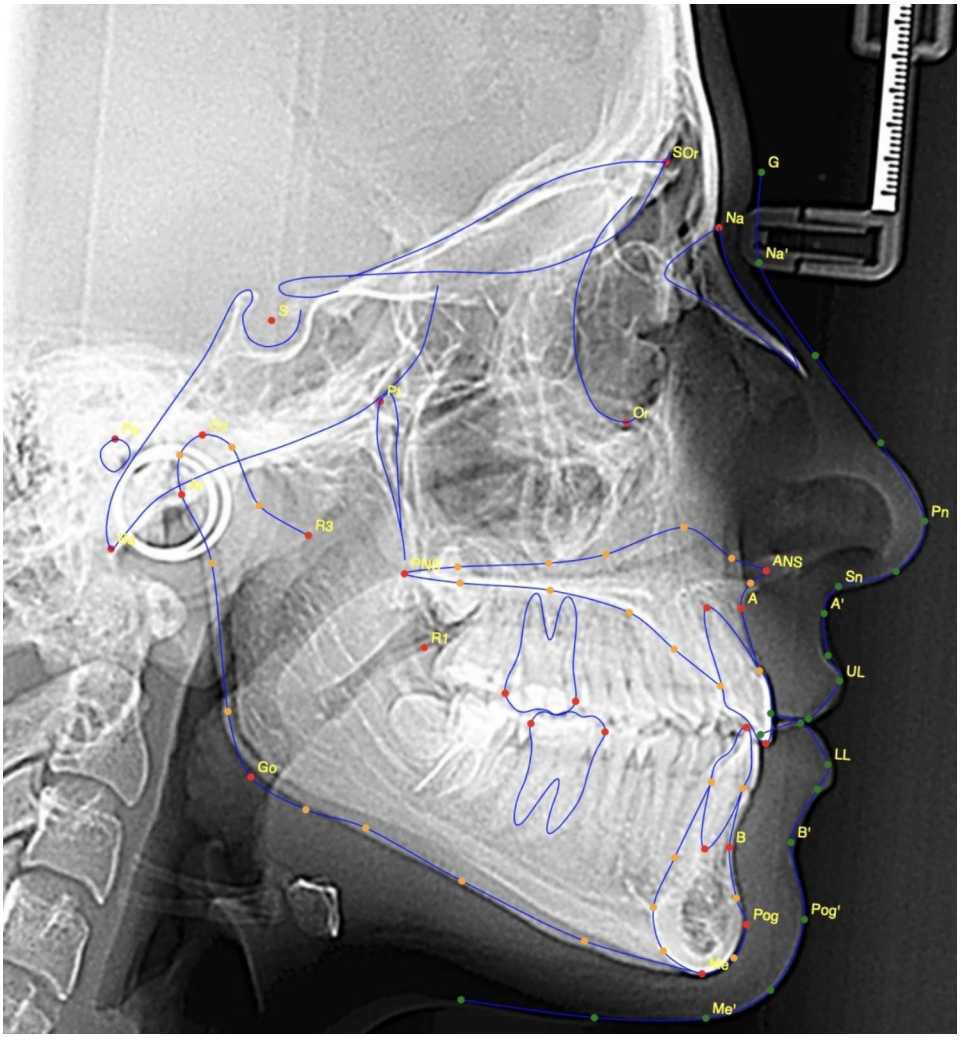


Fig. 4 Cephalometric tracing with AI-based software automatic option. Automatic cephalometric tracing without manual adjustment

either under- or overestimate the results, depending on the measurement. In contrast, El-Dawlatly et al. [33] found an overestimation for SNA (°), SNB (°), and ANB (°), and an underestimation for SN-MP (°). However, (U1)-NA (°) and (L1)-NB (°) were overestimated in their study, while our results showed them to be underestimated (Table 5).

AI-based semi-automatic version

In contrast, we observed a tendency for the semi-automatic AI-based software to overestimate most measurements. These findings align with the study by El-Dawlatly et al. [33], which also showed positive values for similar measurements, such as (L1)-NB (mm) within the 95% confidence interval. However, (L1)-NB (°) was overestimated in their study, whereas it was underestimated in ours (Table 5).

Error clinically significant

AI-based automatic option

For AI-based automatic version, the differences with the conventional option were statistically significant for all measurements except for SN-MP (°) ($p=0.449$) and L1-NB (mm) ($p=0.118$).

In the literature, the different studies were realized with two criteria of clinical acceptance. Some of them used 2° or mm or 1° or mm. Our results were calculated with these two acceptance criteria.

With a clinical acceptance threshold of 2° or 2 mm, a significant clinical difference was observed in L1-NB (°). Similar findings were reported by Yassir et al. [19], who compared AI-based software with AutoCAD™ in a sample of 50 patients. The study revealed a clinically significant difference in the inclination of the lower incisor (L1-MP). This discrepancy to localize lower incisor, may be attributed to the difficulty and subjectivity involved in accurately defining its shape.

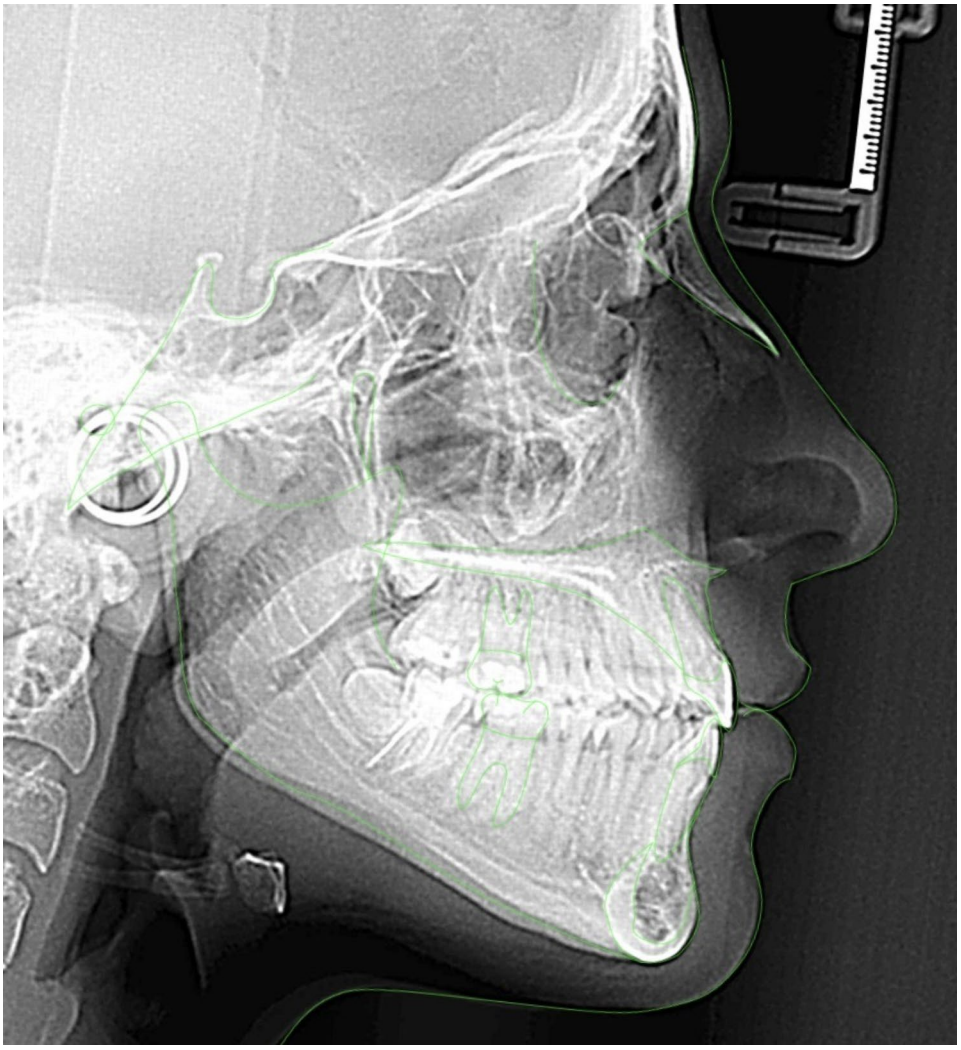


Fig. 5 Cephalometric tracing with Conventional digital tracing software. Traditional tracing involving manual placement of the cephalometric points

In contrast, with a clinical acceptance threshold of 1° or 1 mm, SNA (°), Wits (mm), (U1)-NA (°), (L1)-NB (°), and IMPA (°) exhibited clinically significant differences. Similar results were reported by El-Dawlatly et al. [33], who compared AI-based software with Onyx-Ceph™ in a sample of 200 patients. The study found that SNA (°), (U1)-NA (°), and ANB (°) exceeded the clinical acceptance threshold of 1° or 1 mm. Additionally, Duran et al. [34] compared WebCeph and Dolphin and identified statistically significant differences in several cephalometric measurements, including SNA, ANB, Occlusal plane angle, articular angle, gonial angle, anterior cranial base, corpus length (Go-Me), and ramus height.

All of these results were presented in Table 6.

AI-based software semi-automatic option

For the AI-based software semi-automatic version, the differences with the conventional option were statistically significant for every measurement except for SNA (°) ($p=0.408$), SNB (°) ($p=0.281$), ANB (°) ($p=0.656$), (U1)-NA (°) ($p=0.176$), and (U1)-APg (°) ($p=0.638$).

However, with a clinical acceptance threshold of 2° or 2 mm, no significant clinical differences were found. Similar results were reported by Yassir et al. [19], who identified only (L1)-MP (°) as having a clinically significant difference. In contrast, Azeez et al. [35], who compared the semi-automatic AI-based software with manual tracing, observed a lack of agreement in the following measurements: FMA, IMPA, UI to NA, and LI to NB.

However, using a clinical acceptance of 2° or 2 mm, no significant clinical differences were found. Similar results were reported by Yassir et al. [19], but they found only (L1)-MP (°) with a clinically significant difference. Whereas, Azeez et al. [35] who compared AI-based software semi-automatic option with manual tracing, observed a lack of agreement in the following measurements: FMA, IMPA, UI to NA, and LI to NB.

In contrast, when considering a clinical acceptance of 1° or 1 mm, only (L1)-NB (°) exhibited clinically significant differences. These findings were consistent with the El-Dawlatly et al. [33] study's, where (U1)-NA (°) and (L1)-NB (°) presented results above the 1° or 1 mm setting.

All of these results were indicated in Table 6.

Time-saving

Time is important in orthodontic practice. From the analysed results, AI-based automatic option proved to be the quickest, with an average time of 10.2 s, while the conventional one took 9 min and 25 s. The semi-automatic alternative was also a good time-saving solution, with a mean time of 2 min and 26 s.

In the study conducted by Bao et al. [36], the average time was two seconds for the automatic tracing and two minutes and 37 s for the conventional method with 2 operator of 10 and 17 years of experiences. The difference in time between our study and the study conducted by Bao et al. [36], for conventional tracing could be explain by the operator's experience. In fact, the main operator in our study was a 2nd year orthodontic student and used more time to place the landmarks. This comparison showed that the more experienced the doctor, the faster the landmarks will be recognized and placed. While for the AI-based, the small difference found could be explained by the quality of the internet network. In the same way, El-Dawlatly et al. [33] reported comparable time consumption. Ye et al. [28] showed a significant reduction in time for automatic tracing using three AI systems: MyOrthoX™, Angelalign™, and Digident™.

Reliability analysis: Intra- and inter-operator reliability

This study exhibited high intra-observer reliability with an ICC greater than 0,940 (0,933 min to 0,993 max). Similar results were obtained in the study by Bao et al. [36], where the ICC ranged from 0,91 to 0,99. The study conducted by El-Dawlatly et al. [33] reported an ICC of 0,81 to 0,91.

In the same way, a high level of inter-observer reliability was achieved with an ICC superior to 0,900 (0,906 min to 0,989 max). This significant degree of reliability was demonstrated in both studies conducted by El-Dawlatly et al. and Bao et al. [33, 36], indicating consistent results between operators. Nevertheless, Duran et al. [34] found that the inter-operator ICC was usually over 0.7. However, for the following measurements: articular angle, anterior cranial base, ramus height, ICC dropped between 0.5 and 0.7 (moderate reliability). The occlusal angle showed a poor reliability with an ICC of 0,294, which requires further research in the future.

Mahto et al. [37] found that the values of ICC were over 0,75 when comparing AI-based software and manual tracing. They noted that the ICC values were over 0,90 for the following measurements: ANB, FMA, IMPA/L1 to MP (°), LL to E-line, L1 to NB (mm), L1 to NB (°), and S-N to Go-Gn. These results are also similar to our findings.

Limitations of AI

The study of Liu et al. [38] explained, "If the landmark was on a well-delineated border of the craniofacial structure, like the points S and M, the error was smaller. However, if the point was in an area that was not well delineated, like points P and O, the error was bigger." The study by Kang et al. [39] also supported this idea, stating that the cephalometric structures with high contrast exhibited less mean error than the low contrast section of the cephalometric radiograph. In fact, in this study, the operator had to modify many landmark localizations, such as the basion, posterior nasal spine, condyle, articular, lower incisor, and porion (Figs. 6 and 7). In contrast to this study, Indermun et al. [40] and Jiang et al. [41] had difficulties to localize Gonion and other landmarks as nasion, anterior nasal spine [39] and soft-tissue pogonion [40, 41]. Those inaccurate localizations were also attributed to inadequate image quality and anatomical variations (incisal shape). Duran et al. [34] also pointed out that AI is less accurate in soft tissue measurements.

Study limitations

Patient head position

The 408 lateral cephalograms were taken by different students, which was the main limitation of this study. Every student did not always place the patients in the correct position. They sometimes seated or positioned the patients straight and did not always ensure that the Frankfort plane was perpendicular to the ground. That is why only patients with the correct head position in their lateral cephalometry were selected. In fact, even a slight modification of the patient's head position can produce errors during the tracing process, thereby reducing the accuracy of landmark placement (Fig. 8).

Radiographs quality

The quality of the 2D lateral cephalograms plays an important role in the detection of landmarks. In this research, 25 lateral cephalometries were excluded initially because AI-based software was unable to accurately place cephalometric points, unlike conventional digital tracing (Fig. 9).

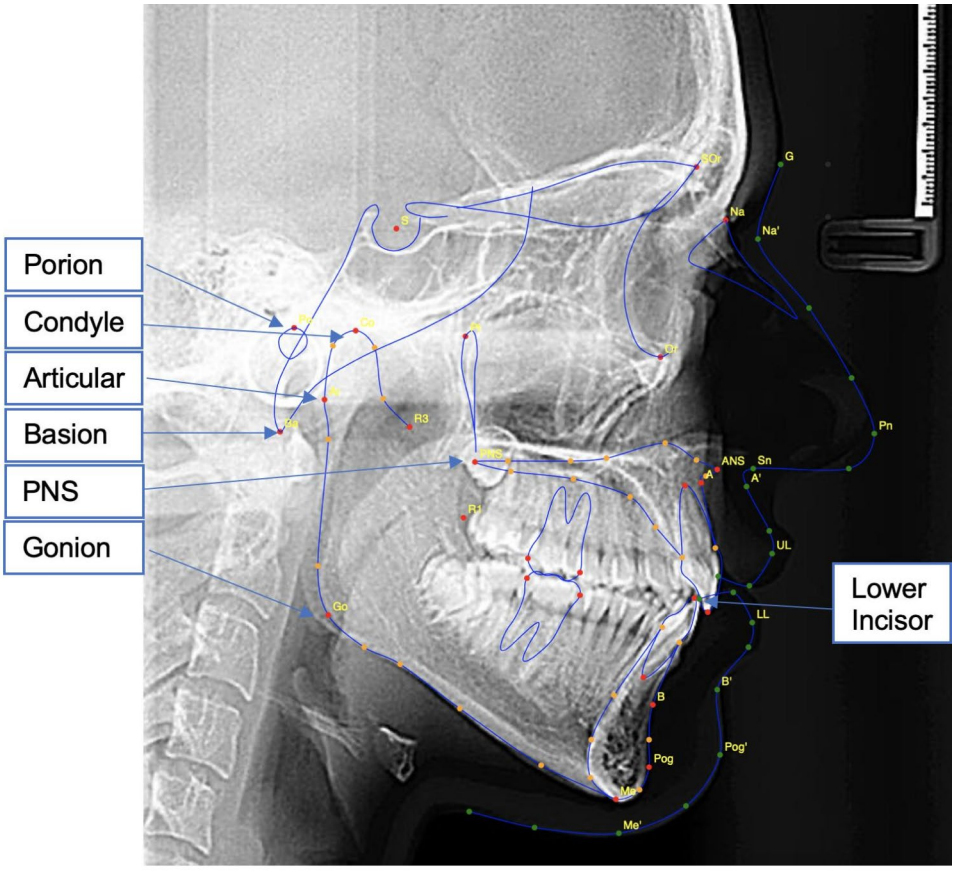


Fig. 6 Error of point placement with Automatic AI-based tracing software. The AI-based automatic option inaccurately places several points such as basion, posterior nasal spine, condyle, articular, incisive, and porion

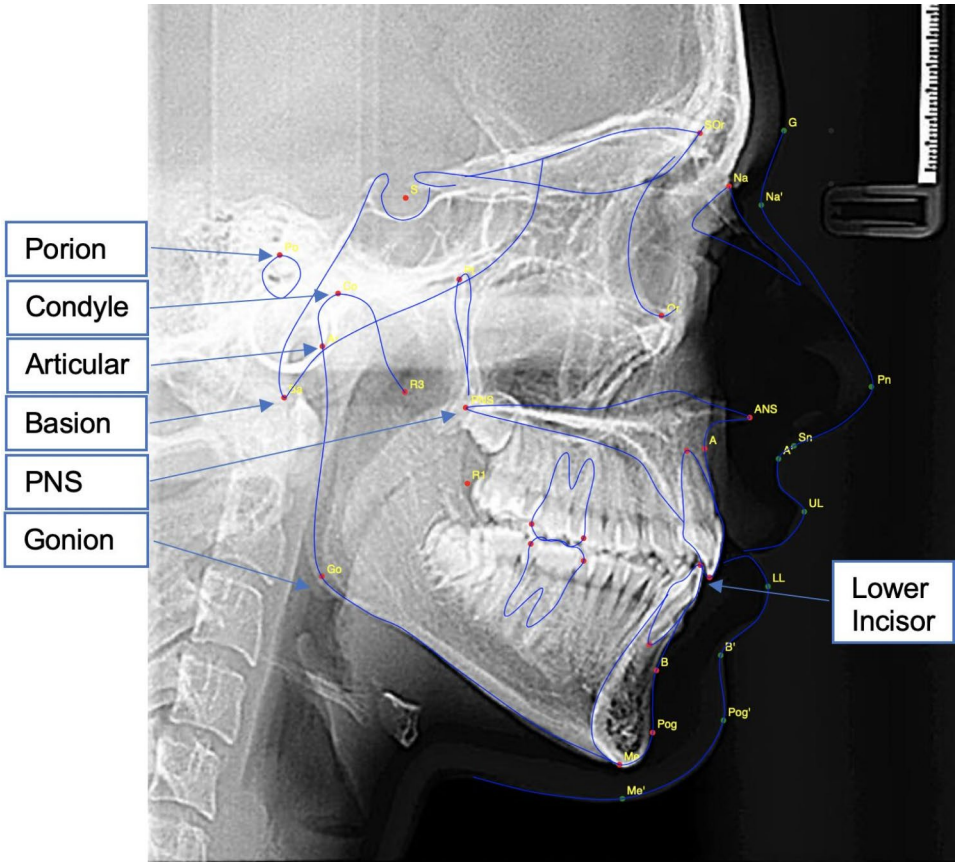


Fig. 7 Adjustment of point placement error with AI-based semi-automatic software. Manual refinement of initially inaccurately placed points using AI-based semi-automatic software



Fig. 8 2D lateral cephalogram discarded for wrong head position. This is an example of a discarded lateral cephalogram due to incorrect head positioning (not parallel to the Frankfurt plan)



Fig. 9 2D lateral cephalogram discarded for AI-based software analysis due to bad quality. The image was rejected because the AI-based software failed to precisely position cephalometric points, unlike traditional digital tracing methods, due to the bad quality of the image

Conclusion

Overall, the study results exhibited a statistical difference in accuracy between the conventional digital technique and the two alternatives to AI-based software. However, these differences were not clinically significant, except for SNA (°), Wits (mm), (U1)-NA (°), (L1)-NB (°), and IMPA (°) at the 1° or 1 mm setting for AI-based automatic option, and (L1)-NB (°) for AI-based semi-automatic option.

Nevertheless, automatic landmark placement followed by operator modification could be a good option in 2D cephalometric analysis. While automatic AI alone is not completely accurate and reliable for cephalometric tracing, it shows promising results.

Significant time savings were demonstrated using the AI-based automatic and semi-automatic methods compared to the conventional digital technique. However, the automatic method was the fastest.

Further studies are needed to confirm the ability of AI to produce accurate results for cephalometric tracing.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12903-024-05097-6.

Supplementary Material 1

Acknowledgements
Not applicable.

Author contributions
JPM and LT contributed to conceptualization, data curation, formal analysis, investigation, methodology, visualization, writing—original draft, and writing—review and editing. LT also contributed to project administration and supervisionCR, INS, PMP and IDR contributed to supervision, visualization, formal analysis, writing—original draft, and writing—review and editing. This manuscript is original, it has not been published previously and is not under consideration for publication elsewhere. The final version has been seen and approved by all authors. All authors discussed and contributed to the final manuscript.

Funding
This study was carried out without any external funding.

Data availability
The datasets generated and/or analysed during the current study are not publicly available due to University Data Protections Policies but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
The study was approved by the Institutional Review Board (IRB) of the University Alfonso X el Sabio Bioethics committee (Resolution 2023_03/184) which belongs to the dentistry faculty at the Alfonso X el Sabio University. And since the study used pseudonymised patient's data, a patients additional informed consent was not deemed necessary. All patients under treatment sign a generic consent at the start of the treatment stating that all data could be used for research purposes, with pseudonymised data.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 20 April 2024 / Accepted: 23 October 2024
Published online: 28 October 2024

References

1. Proffit WR, Fields HW, Larson B. Contemporary orthodontics. 6th ed. Philadelphia: Elsevier Health Sciences; 2018.
2. Nielsen IL. A comprehensive diagnostic system for orthodontists-Beyond Angle's classification. J Taiwan Assoc Orthod. 2019;31(3):153–65.
3. Tanna NK, AlMuzaini AAAY, Mupparapu M. Imaging in orthodontics. Dent Clin North Am. 2021;65:623–41.
4. Shrestha R, Kandel S. A comparative study on use of manual versus digital method using mobile application for cephalometric measurements. Orthod J Nepal. 2020;10:11–6.
5. Hlongwa P. Cephalometric analysis: manual tracing of a lateral cephalogram. S Afr Dent J. 2019;74:389–93.
6. Yoon HJ, Kim DR, Gwon E. Fully automated identification of cephalometric landmarks for upper airway assessment using cascaded convolutional neural networks. Eur J Orthod. 2022;44:66–77.
7. Gupta AR, Jha MS, Saurav S. Hand tracing verses digital tracing with PACS method. Int J Contemp Med Res. 2019;6:33–7.
8. Uysal T, Baysal A, Yagci A. Evaluation of speed, repeatability, and reproducibility of digital radiography with manual versus computer-assisted cephalometric analyses. Eur J Orthod. 2009;31(5):523–8.
9. Nathasha MM, Chakravarthi NCS, Srinivasan D. Orthodontics in the era of digital innovation – a review. J Evol Med Dent Sci. 2021;10:2114–21.
10. Monill-González A, Rovira-Calatayud L, d'Oliveira NG. Artificial intelligence in orthodontics: where are we now? A scoping review. Orthod Craniofac Res. 2021;24:6–15.
11. Kiełczykowski M, Kamiński K, Perkowski K, Zadurska M, Czochrowska E. Application of Artificial Intelligence (AI) in a cephalometric analysis: a narrative review. Diagnostics. 2023;13(16).
12. 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).
13. Narkhede S, Rao P, Sawant V, Sachdev SS, Arora S, Pawar AM et al. Digital versus Manual tracing in Cephalometric Analysis: a systematic review and Meta-analysis. J Pers Med. 2024;14(6).
14. Butul B, Sharab L. Obstacles behind the innovation- a peek into Artificial intelligence in the field of orthodontics – A literature review. Saudi Dent J. 2024;36(6):830–4.
15. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015;351:h5527.
16. Tikku T, Khanna R, Maurya RP, Srivastava K, Bhushan R. Comparative evaluation of cephalometric measurements of monitor-displayed images by Nemoceph software and its hard copy by manual tracing. J Oral Biol Craniofacial Res. 2014;4(1):35–41. https://doi.org/10.1016/j.jobcr.2013.11.002
17. Kim YH, Park JB, Chang MS, Ryu JJ, Lim WH, Jung SK. Influence of the depth of the convolutional neural networks on an artificial intelligence model for diagnosis of orthognathic surgery. J Pers Med. 2021;11(5).
18. Kula K, Ghoneima A. In: Huffman L, editor. Cephalometry in Orthodontics: 2D and 3D. Quintessence Publishing; 2019.
19. Yassir YA, Salman AR, Nabbat SA. The accuracy and reliability of WebCEPH for cephalometric analysis. J Taibah Univ Med Sci. 2022;17:57–66.
20. Ranganathan P, Pramesh C, Buyse M. Common pitfalls in statistical analysis: clinical versus statistical significance. Perspect Clin Res. 2015;6(3):169.
21. Sharma H. Statistical significance or clinical significance? A researcher's dilemma for appropriate interpretation of research results. Saudi J Anaesth. 2021;15(4):431–4.
22. Santoro M, Jarjoura K, Cangialosi TJ. Accuracy of digital and analogue cephalometric measurements assessed with the sandwich technique. Am J Orthod Dentofac Orthop. 2006;129:345–51.
23. Erkan M, Gurel HG, Nur M. Reliability of four different computerized cephalometric analysis programs. Eur J Orthod. 2012;34:318–21.
24. Segura FJE, Valverde AS, Ocampo AM. Comparative study between digital and manual cephalometry with digital radiographs. Rev Mex De Ortod. 2014;2:e93–6.
25. Alfawzan AA. Assessment of Skeletal Maturity in a sample of the Saudi Population using cervical vertebrae and frontal sinus index: a cephalometric study using Artificial Intelligence. Cureus. 2023;15(7):e41811.
26. Kunz F, Stellzig-Eisenhauer A, Zeman F, Boldt J. Artificial intelligence in orthodontics. J Orofac Orthop / Fortschr Der Kieferorthopädie. 2020;81(1):52–68.
27. Tsolakis IA, Tsolakis AI, Elshebiny T. Comparing a fully automated cephalometric tracing method to a manual tracing method for orthodontic diagnosis. J Clin Med. 2022;11:6854.
28. Ye H, Cheng Z, Ungvijanpunya N. Is automatic cephalometric software using artificial intelligence better than orthodontist experts in landmark identification? BMC Oral Health. 2023;23:467.
29. Hwang HW, Park JH, Moon JH. Automated identification of cephalometric landmarks: part 2 – might it be better than human? Angle Orthod. 2019;90:69–76.
30. Schwendicke F, Chaurasia A, Arsiwala L. Deep learning for cephalometric landmark detection: systematic review and meta-analysis. Clin Oral Investig. 2021;25:4299–309.
31. Junaid N, Khan N, Ahmed N. Development, application, and performance of artificial intelligence in cephalometric landmark identification and diagnosis: a systematic review. Healthcare. 2022;10:2454.
32. Le VNT, Kang J, Oh IS. Effectiveness of human–Artificial intelligence collaboration in cephalometric landmark detection. J Pers Med. 2022;12:387.
33. El-Dawlatly MM, Attia KH, AbdelGhaffar AY. Preciseness of artificial intelligence for lateral cephalometric measurements. J Orofac Orthop. 2023():1–7.
34. Duran GS, Gökmen Ş, Topsakal KG, Görgülü S. Evaluation of the accuracy of fully automatic cephalometric analysis software with artificial intelligence algorithm. Orthod Craniofac Res. 2023;26(3):481–90.
35. Azeez SM, Surji FF, Kadir SO, Karim R. Accuracy and reliability of WebCeph Digital Cephalometric Analysis in Comparison with conventional cephalometric analysis. World J Dent. 2023;14:727–32.
36. Bao H, Zhang K, Yu C. Evaluating the accuracy of automated cephalometric analysis based on artificial intelligence. BMC Oral Health. 2023;23:191.
37. Mahto RK, Kafle D, Giri A. Evaluation of fully automated cephalometric measurements obtained from web-based artificial intelligence driven platform. BMC Oral Health. 2022;22:1–8.
38. Liu J, Chen YT, Cheng K. Accuracy of computerized automatic identification of cephalometric landmarks. Am J Orthod Dentofac Orthop. 2000;118:535–40.
39. Kang S, Kim I, Kim YJ, Kim N, Baek SH, Sung SJ. Accuracy and clinical validity of automated cephalometric analysis using convolutional neural networks. Orthod Craniofac Res. 2024;27(1):64–77.

40. Indermun S, Shaik S, Nyirenda C, Johannes K, Mulder R. Human examination and artificial intelligence in cephalometric landmark detection-is AI ready to take over? Dentomaxillofacial Radiol. 2023;52(6):1–14.
41. Jiang F, Guo Y, Yang C, Zhou Y, Lin Y, Cheng F, et al. Artificial intelligence system for automated landmark localization and analysis of cephalometry. Dentomaxillofacial Radiol. 2023;52(1):20220081.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.