Original Article

Influence of head positioning errors on the accuracy of fully automated artificial intelligence-based cephalometric software

Alessandro Polizzi[a]; Antonino Lo Giudice[b]; Cristina Conforte[c]; Gaetano Isola[b]; Rosalia Leonardi[d]

ABSTRACT

Objectives: To evaluate the accuracy of three fully automated software systems compared to nonautomated cephalometric analysis software tested using cephalograms featuring correct and incorrect head positions.

Materials and Methods: The study sample consisted of 40 lateral cephalograms retrieved retrospectively from a larger pool of pretreatment orthodontic records. Cephalograms were recruited and divided into correct head posture group (CHP) and incorrect head posture group (IHP). Cephalometric data were obtained by manual landmarking (Dolphin software), which served as a reference, and by fully automated AI software (WebCeph, Ceph Assistant, and AudaxCeph). Intraclass correlation coefficients (ICC) and paired *t*-tests were used for intragroup comparisons, whereas analysis of variance and post-hoc analysis were used to compare performance among artificial intelligence (AI) based software applications.

Results: The tested software exhibited a good level of consistency for angular measurements whereas linear measurements were more error-prone. AudaxCeph demonstrated the most consistent accuracy, achieving excellent agreement (ICC $>$ 0.90) for several skeletal parameters; however, it failed in detecting soft tissue accurately. WebCeph and Ceph Assistant showed greater variability, especially for linear measurements (ICC $<$ 0.50). Positional errors drastically reduced measurement accuracy, with linear parameters such as Go-Me showing the poorest agreement across all software.

Conclusions: AI-based cephalometric software demonstrated variable accuracy depending on the cephalometric measurement, and this pattern was exacerbated under conditions involving positional errors in cephalograms. Accordingly, oversight by expert clinicians is still required to minimize marginal error. (*Angle Orthod*. 2025;00:000–000.)

KEY WORDS: Orthodontics; Cephalometric analysis; Fully automatic cephalometry; WebCeph; AudaxCeph; Ceph Assistant; Dolphin

[a] PhD Student, Department of General Surgery and Surgical-Medical Specialties, University of Catania, Catania, Italy.

[b] Associate Professor, Department of General Surgery and Surgical-Medical Specialties, University of Catania, Catania, Italy.

[c] Private Practice, Augusta, Italy.

[d] Full Professor, Department of General Surgery and Surgical-Medical Specialties, University of Catania, Catania, Italy.

Corresponding author: Dr Rosalia Leonardi, Department of General Surgery and Surgical-Medical Specialties, Section of Orthodontics, University of Catania, Via S. Sofia 68, Catania, 95124, Italy
(e-mail: rleonard@unict.it)

Accepted: April 13, 2025. Submitted: December 31, 2024.
Published Online: July 1, 2025

INTRODUCTION

Cephalometric analysis is an indispensable diagnostic tool in clinical orthodontics. It enables assessment of dentofacial proportions, anatomical basis of malocclusion, growth pattern, and post-treatment changes.[1]

Traditionally, cephalometric analysis has been conducted using a manual drawing technique involving acetate tracing paper, rulers, and protractors. However, this approach is inherently time-consuming and susceptible to inter- and intra-operator variability.[2] Over recent decades, the advent of computer-aided cephalometric analysis software allowed for a more efficient measurement process and reduced error in landmark identification and linear/angular measurements compared to the manual technique.[3] Nonetheless, variability in landmark identification remains a significant source of random error, even with computer-aided systems.[4–6] Thus, manual and computer-aided cephalometric analyses are open to considerable subjectivity and remain time-intensive, although to a different extent.

Given these limitations, the integration of artificial intelligence (AI) has been tested to automate cephalometric analysis. Briefly, these models can be broadly categorized into semi-automatic and fully automatic systems. Semi-automatic AI models assist operators by providing tools for landmark detection, measurement estimation, and image segmentation. In contrast, fully automated systems perform all aspects of cephalometric analysis autonomously, including landmark identification, measurement computation, and reporting.[7,8]

Despite the obvious advantages of the fully automated method,[9] a general consensus for its clinical application has not been reached due to accuracy and reliability concerns.[10,11] Indeed, the literature has shown conflicting results since some studies[12–15] reported good-to-excellent agreement with manual landmarking, whereas other studies[16–18] reported a lower level of accuracy, especially for linear measurements. Additionally, a recent meta-analysis[19] highlighted biases in patient selection, insufficient randomization processes, and absence of standardized protocols.

A further limitation of existing studies is that none of them considered the influence of head positional error on automated landmark identification. In such cases, when the head is slightly rotated and/or inclined relative to the ideal position, in which the midsagittal plane of the patient's head should align parallel to the detector's plane, it is suggested to average the positions of bilateral structures for landmark identification.[20] This method has been shown to ensure reliable cephalometric analysis while avoiding the need for repeated acquisition of cephalograms. However, no studies in the literature have tested AI models to address this task under the conditions of imprecise patient posture.

The present study aimed to compare the accuracy and the reliability of different fully automated AI-based cephalometric software applications in deriving linear and angular measurements performed on cephalograms featuring correct and incorrect head posture. The null hypothesis was that fully automatic software would achieve excellent agreement in all cephalometric measurements compared to the manual method, under ideal and challenging conditions.

MATERIALS AND METHODS

This retrospective study was approved by the Institutional Review Committee of the University of Catania (IRC n° A.Q.A.M.DI. 119/2020/PO) and was conducted in accordance with the principles of the Declaration of Helsinki. Informed consent was obtained from all the patients and/or their legal guardians.

Inclusion and Exclusion Criteria

The study sample was obtained from a retrospective pool of pretreatment cephalograms retrieved from the archives of the Department of Orthodontics at the University of Catania. During the recruitment process, cephalograms were divided into two groups: correct head posture (CHP) and incorrect head posture (IHP) groups. In the CHP group, the midsagittal plane of the patient's head was parallel with the detector's plane, without generating duplicated anatomical structure profiles or landmarks. In the IHP group, cephalograms were characterized by head positional error, as reflected by ear rod markers and associated with some duplicated landmarks. As described previously,[21] only rotations around the vertical axis and those around the antero-posterior axis were considered positioning errors, as they could affect horizontal measurements and vertical measurements, respectively. On the other hand, rotations around the transverse axis were not considered positioning errors, as they do not cause image distortions since the location of the head is parallel to the central ray. Inclusion criteria were: (1) good quality images, (2) presence of a calibration ruler, and (3) no image artifacts. Exclusion criteria were: difficulty in identifying landmarks due to (1) extra soft tissue on cephalograms, (2) image motion, discrepancy in resolution, or lack of contrast. After, each cephalogram was labeled and a web application was used (www.randomizer.org) to randomly select the radiographs for final inclusion in both groups. In detail, each cephalogram retrieved from the archive received a unique number depending upon correct (CHP group) or incorrect (IHP group) head positioning. The system generated 20 random numbers for each group.

Cephalometric Analysis

Cephalometric measurements considered in the present study were derived from the American Board of Orthodontics (ABO) cephalometric guidelines.[22] Calibration of measurements was performed using a known distance (20 mm) between two ruler points.

Nonautomated computer-aided landmarking was performed in CHP and IHP groups by an experienced orthodontist with more than 10 years of experience, using Dolphin 11.8 software (Patterson Dental Supply) (Figure 1). Measurements of 10 randomly selected cephalograms were repeated in three sessions with an interval of 1 week between each repetition. The midpoint was constructed to make a single landmark for bilateral structures and double images (IHP group). The experienced orthodontist was not aware of the AI-

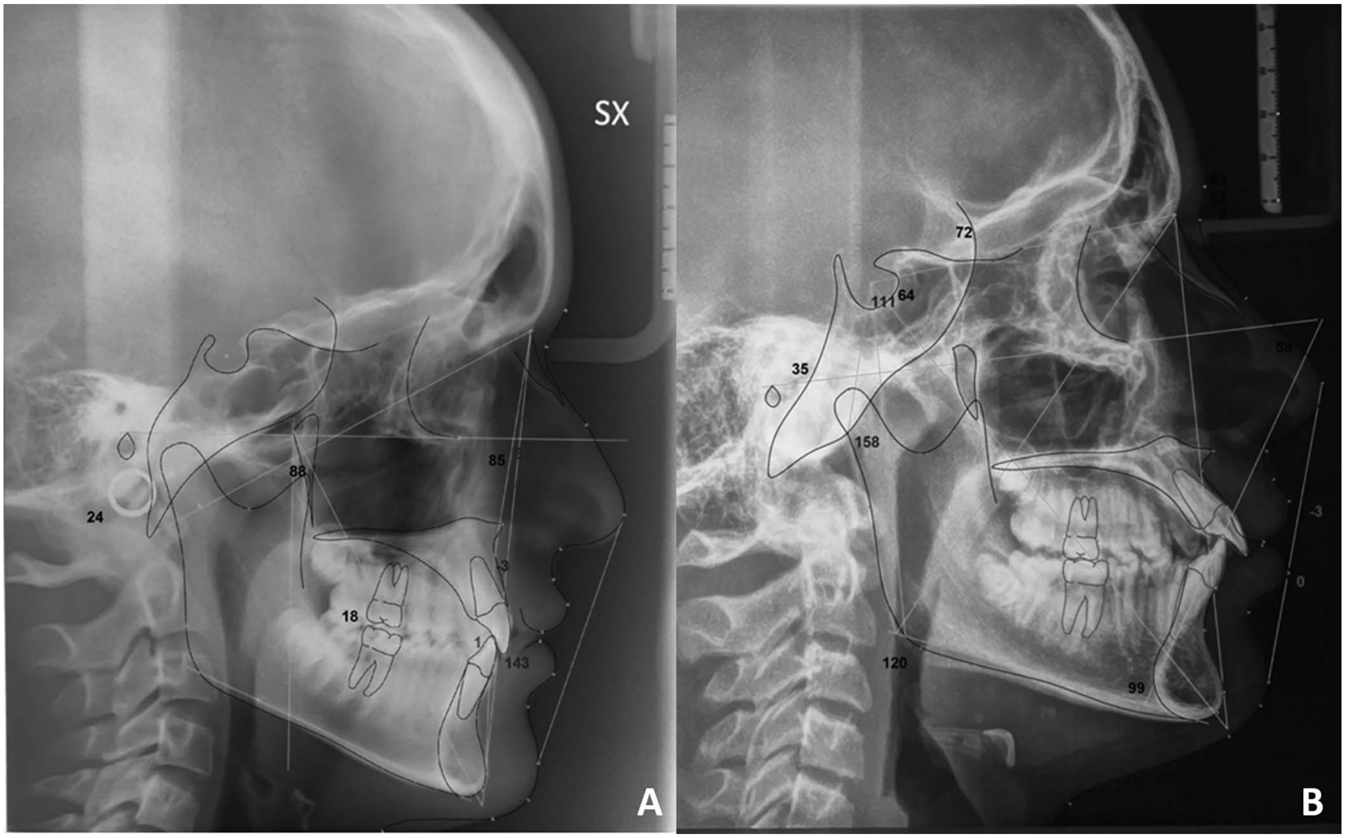

Figure 1. Nonautomated computer-aided landmarking with Dolphin 11.8 software: (A) CHP group, (B) IHP group. CHP indicates correct head posture; IHP, incorrect head posture.

driven cephalometric analysis results before finishing manual landmarking.

Fully automatic AI-driven landmarking was performed in CHP and IHP groups without any human correction and before nonautomated cephalometric analysis. Data were collected by an expert operator author who was not involved in the nonautomated landmarking process. Three software programs were tested: (1) WebCeph (AssembleCircle Corp., Gyeonggi-do, Republic of Korea, https://webceph.com) (Figure 2), (2) Ceph Assistant

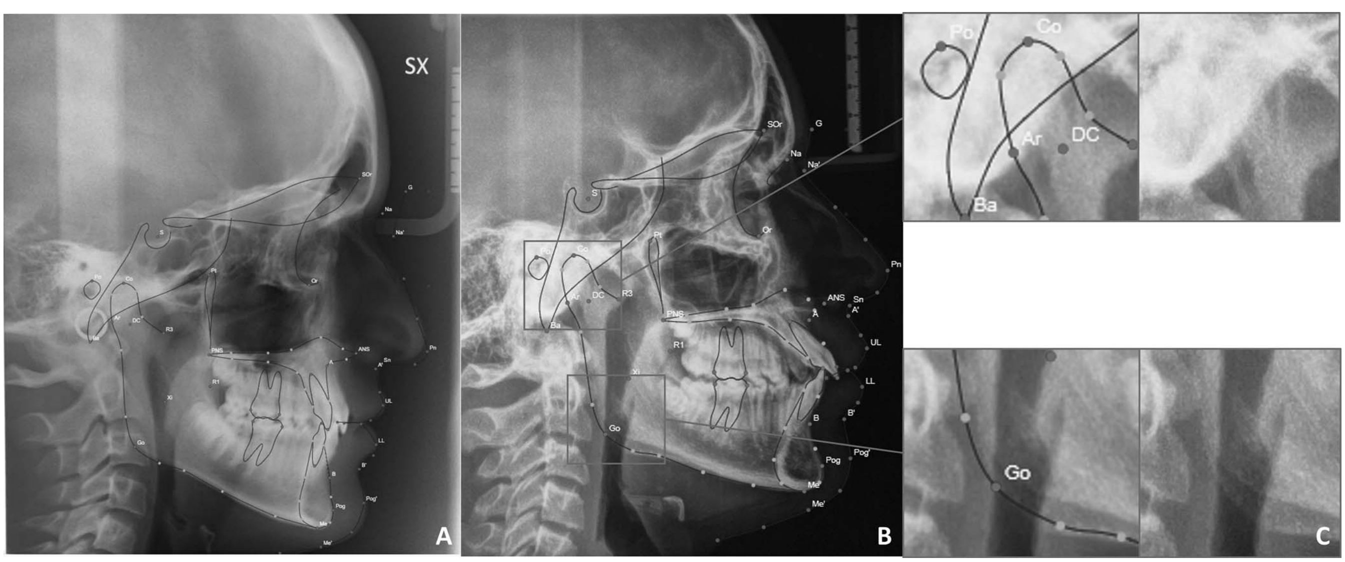

Figure 2. AI-driven automatic landmarking with WebCeph: (A) CHP group, (B) IHP group. (C) Greater magnification of some landmarks that are not well localized, in a patient of the IHP group.

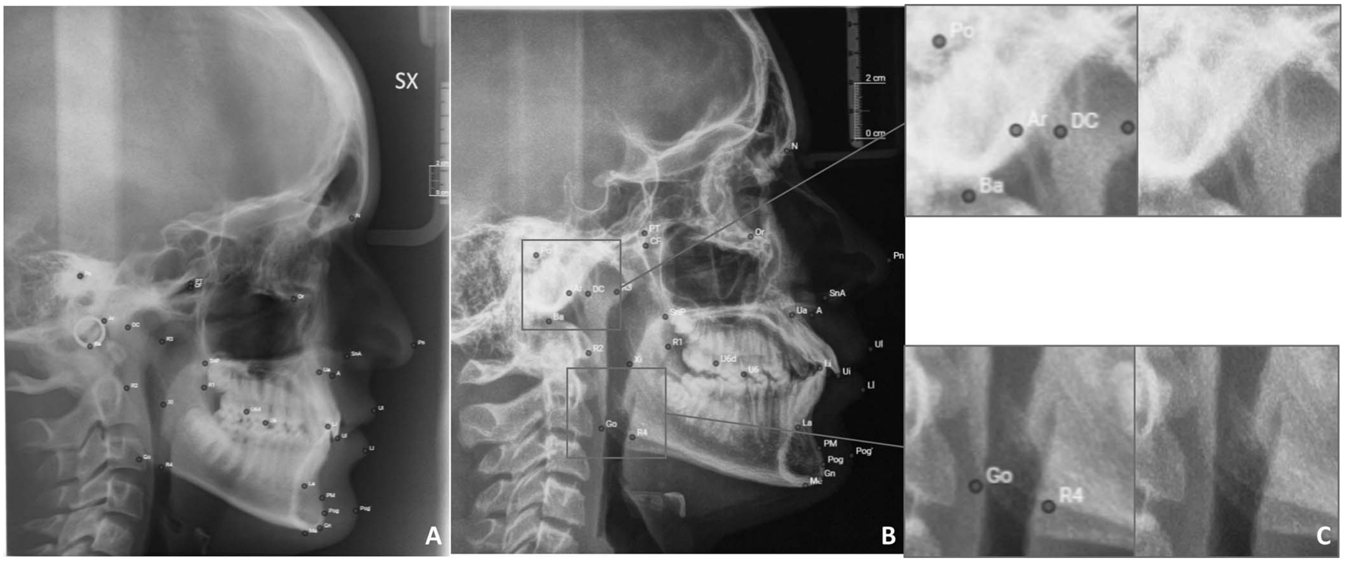

Figure 3. AI-driven automatic landmarking with Ceph Assistant: (A) CHP group, (B) IHP group. (C) Greater magnification of some landmarks that are not well localized, in a patient of the IHP group.

(Budapest, Hungary, https://www.cephassistant.com/) (Figure 3), and (3) AudaxCeph (Audax, d.o.o., Ljubljana, Slovenia, https://www.audaxceph.com/) (Figure 4). Digital lateral cephalograms were uploaded in the systems that automatically identified the landmarks and performed the cephalometric analysis.

In addition, the performance of AI cephalometric analyses was assessed with the success classification rate (SCR) compared to nonautomated cephalometric analysis. This is a metric for the classification of anatomical types established by the Symposium on Biomedical Imaging conferences in 2015.[23] The SCR was applied to the ABO parameters,[22] which were included within this classification: ANB(^) (type 1: 3.2°–5.7° Class I; type 2: >5.7° Class II; type 3: <3.2° Class III), SNA(^) (type 1: 79.4°–83.2° normal maxilla; type 2: >83.2° prognathic maxilla; type 3: <79.4° retrognathic maxilla), SNB(^) (type 1: 76.4°–78.7° normal mandible; type 2: <74.6° retrognathic mandible; type 3: >78.7° prognathic mandible), SN-MP(^) (type 1: 26.8°–31.4°; type 2: >31.4° mandible high angle; type 3: <26.8° mandible low angle).

Sample Size and Statistical Analysis

Preliminary evaluation of sample size power was performed using 20 cephalograms (10 in the CHP

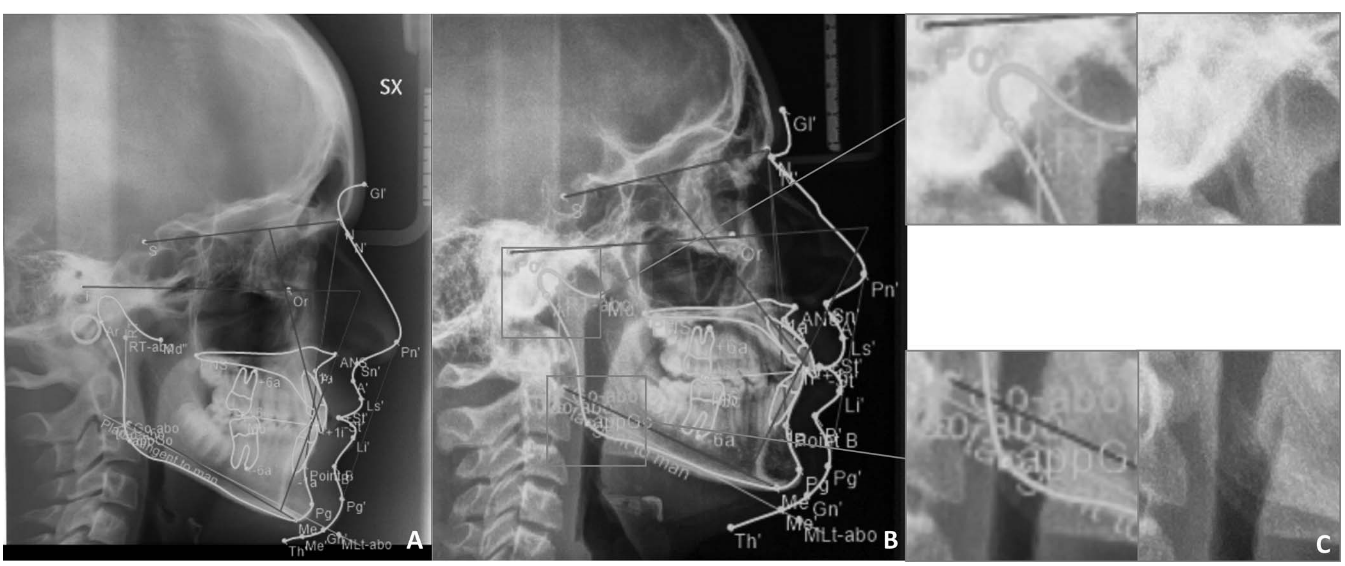

Figure 4. AI-driven automatic landmarking with AudaxCeph: (A) CHP group, (B) IHP group. (C) Greater magnification of some landmarks that are not well localized, in a patient of the IHP group.

group and 10 in the IHP group). Assuming ANB (^) as the primary outcome, the analysis suggested that 16 patients per group were required to reach the 80% power to detect a mean difference of 0.7° and a standard deviation of 0.5° between groups, considering a two-sided significance level of 5% and a 1:1 enrollment ratio. However, according to the inclusion criteria, 20 subjects were able to be included in each group, which increased robustness of the data.

Numerical variables were expressed as mean and standard deviation (SD). A parametric approach was used due to the normal distribution of most of the variables as verified with the Shapiro–Wilk test. Agreement between the fully automatic AI-driven cephalometric algorithms and the manual measurements was evaluated with the intraclass correlation coefficient (ICC). According to previous studies,[15] ICC values were classified as follows:

- ICC $<$ 0.75: poor to moderate agreement
- ICC 0.75–0.90: good agreement
- ICC $>$ 0.90: excellent agreement

A paired *t*-test was used to compare the linear and angular measurements within the same group (ie, manual vs AI-driven cephalometric analysis in the CHP and IHP group). One-way analysis of variance (ANOVA) analysis and post-hoc testing were applied for AI software comparisons. $P < .05$ was considered statistically significant using IBM SPSS Statistics for Windows, version 26.0 (IBM Corp.) software for statistical analysis.

RESULTS

Measurement Reliability

Intra-examiner ICC for repeated measurements was $>$0.90 indicating high reliability. Table 1 reports the results of manual tracing with Dolphin and the comparisons with AI software.

Success Classification Rate (SCR)

In the CHP group, WebCeph reached a mean SCR = 65%, whereas AudaxCeph and Ceph Assistant values were 81.25% and 76.25%, respectively. The worst results were obtained for SNA(^) classification (WebCeph = 50%, AudaxCeph = 65%, Ceph Assistant = 55%), whereas SNB(^) (WebCeph = 65%, AudaxCeph = 85%, Ceph Assistant = 85%), ANB(^) (WebCeph = 80%, AudaxCeph = 85%, Ceph Assistant = 85%) and SN-MP(^) (WebCeph = 65%, AudaxCeph = 90%, Ceph Assistant = 80%) demonstrated higher classification accuracy.

In the IHP group, WebCeph and Ceph Assistant showed similar mean classification accuracy (SCR = 73.75% and 71.25%, respectively), whereas AudaxCeph reached the highest SCR = 88.75%. Even in this case, the worst results were obtained for SNA (^) classification (WebCeph = 60%, AudaxCeph = 80%, Ceph Assistant = 60%), whereas SNB (^) (WebCeph = 75%, AudaxCeph = 95%, Ceph Assistant = 70%), ANB(^) (WebCeph = 80%, AudaxCeph = 90%, Ceph Assistant = 60%) and SN-MP (^) (WebCeph = 80%, AudaxCeph = 90%, Ceph Assistant = 95%) demonstrated higher classification accuracy.

Intragroup Assessments

For WebCeph software, none of the cephalometric parameters showed excellent agreement in the CHP group. Indeed, good agreement was achieved only for SNB (^), U1 to SN, L1 to NB (^), and LL to E-line, whereas the other values showed moderate-to-poor agreement. Go-Me (mm), U1 to NA (mm), and L1 to NB (mm) showed the worst agreement (ICC $<$ 0.50). In the IHP group, SN-MP (^) demonstrated excellent agreement (ICC = 0.94), whereas good agreement was achieved for SNB (^), U1 to SN (^), and LL to E-line (mm). All the other measurements showed moderate-to-poor agreement. Go-Me (mm), U1 to NA (mm), L1 to NB (mm), and L1 to NB (^) showed the worst agreement (ICC $<$ 0.50) (Table 2).

For Ceph Assistant software, none of the cephalometric parameters showed excellent agreement in the CHP group. Indeed, good agreement was achieved for SNA (^), SNB (^), SN-MP (^), U1 to NA (mm), U1 to NA (^), L1 to NB (mm), and LL to E-line (mm), whereas the other values showed moderate-to-poor agreement. Go-Me (mm) demonstrated the worst ICC agreement (ICC = 0.12). In the IHP group, SN-MP (^) showed excellent agreement (ICC = 0.96), whereas good agreement (ICC: 0.75–0.90) was achieved for SNB (^), FMA (^), U1 to SN (^), U1 to NA (^), and UL to E-line (mm). All the other measurements had moderate-to-poor agreement. Go-Me (mm) showed the worst result (ICC = −0.21).

Regarding AudaxCeph software, SNA (^), SNB (^), and L1 to NB (^) showed excellent agreement in the CHP group. Good agreement was achieved for SN-MP (^), U1 to SN (^), U1 to NA (^), U1 to NA (mm), L1 to MP (^), L1 to NB (^), and L1 to NB (mm), whereas the other values showed moderate-to-poor agreement, with FMA (^) having the worst ICC agreement (ICC = 0.44). In the IHP group, excellent agreement was found for SNB (^), SN-MP (^), U1 to NA (^) and U1 to NA (mm), whereas good agreement was achieved for SNA (^), U1 to SN (^), L1 to NB (^) and L1 to NB (mm). All the other measurements showed moderate-to-poor agreement, with FMA (^), Go-Me (mm), L1 to MP (^), UL to E-line (mm), and LL to E-line (mm) having ICC values $<$0.50.

Table 1. Comparison of Manual Measurements and AI Fully Automated Measurements Based on WebCeph, Ceph Assistant, and AudaxCeph[a]

Measurements	Dolphin (Mean + SD)	WebCeph (Mean + SD, md, *P* Value)	Ceph Assistant (Mean + SD, md, *P* Value)	AudaxCeph (Mean + SD, md, *P* Value)
CHP Group (n = 20)				
SNA, degree	79.6 ± 4.1	77.5 ± 7.2 md: 2.1 *P* = .147	78.9 ± 3.7 md: 0.8 *P* = .234	79.6 ± 4.5 md: 0.0 *P* = .944
SNB, degree	77.0 ± 4.7	76.1 ± 4.0 md: 1.0 *P* = .178	75.9 ± 4.2 md: 1.1 *P* = .099	76.7 ± 4.7 md: 0.3 *P* = .472
ANB, degree	36.6 ± 5.0	34.6 ± 6.2 md: 2.0 *P* = .066	36.8 ± 6.1 md: −0.1 *P* = .861	35.2 ± 5.6 md: 1.4 **_P_ = .029**
FMA, degree	27.7 ± 4.6	25.2 ± 4.7 md: 2.5 **_P_ = .01**	22.0 ± 4.8 md: 5.7 **_P_ < .001**	30.4 ± 5.4 md: −2.8 **_P_ = .031**
Go-Me, mm	74.5 ± 8.2	55.1 ± 13.7 md: 19.4 **_P_ < .001**	70.4 ± 17.8 md: 4.0 *P* = .345	66.1 ± 6.6 md: 8.4 **_P_ < .001**
U1 to SN, degree	102.5 ± 8.6	102.8 ± 7.4 md: −0.4 *P* = .757	100.8 ± 7.2 md: 1.6 *P* = .187	100.5 ± 7.9 md: 2.0 **_P_ = .031**
U1 to NA, degree	23.6 ± 5.8	25.3 ± 10.4 md: −1.7 *P* = .387	22.0 ± 5.5 md: 1.6 *P* = .099	21.2 ± 6.2 md: 2.4 **_P_ = .003**
U1 to NA, mm	3.6 ± 2.2	3.8 ± 3.3 md: −0.2 *P* = .441	3.6 ± 2.6 md: 0.0 *P* = .964	2.5 ± 2.1 md: 0.2 *P* = .522
L1 to MP, degree	89.7 ± 5.4	90.8 ± 6.2 md: −1.1 *P* = .41	91.3 ± 4.5 md: −1.6 *P* = .152	90.9 ± 5.9 md: −1.2 *P* = .169
L1 to NB, degree	23.9 ± 7.3	21.2 ± 9.2 md: 2.7 *P* = .063	23.9 ± 5.6 md: 0.0 *P* = .977	22.8 ± 7.5 md: 1.1 *P* = .155
L1 to NB, mm	3.9 ± 2.2	3.5 ± 2.0 md: 0.3 *P* = .54	3.5 ± 1.7 md: 0.4 *P* = .197	3.7 ± 2.4 md: 0.2 *P* = .219
UL to E-line, mm	−3.3 ± 2.3	−2.3 ± 2.5 md: −1.0 **_P_ = .022**	−2.9 ± 3.0 md: −0.5 *P* = .334	2.7 ± 2.6 md: −6.1 **_P_ < .001**
LL to E-line, mm	−1.4 ± 2.3	−0.8 ± 2.3 md: −0.6 *P* = .082	−1.1 ± 3.0 md: −0.3 *P* = .381	0.7 ± 2.9 md: −2.1 *P* = 0.051
IHP Group (n = 20)				
SNA, degree	80.2 ± 3.3	79.5 ± 7.5 md: 0.7 *P* = .653	81.6 ± 3.7 md: −1.4 **_P_ = .023**	80.5 ± 3.2 md: −0.3 *P* = .334
SNB, degree	76.7 ± 3.5	76.5 ± 4.6 md: 0.1 *P* = .825	77.3 ± 3.6 md: −0.6 *P* = .109	77.0 ± 3.1 md: −0.3 *P* = .353
ANB, degree	35.9 ± 7.1	32.5 ± 8.7 md: 3.4 **_P_ < .001**	35.0 ± 8.0 md: 0.9 *P* = .118	35.0 ± 7.4 md: 0.9 *P* = .079
FMA, degree	24.7 ± 5.1	23.7 ± 7.1 md: 1.0 *P* = .391	21.9 ± 6.9 md: 2.7 **_P_ = .008**	28.9 ± 4.3 md: −4.3 **_P_ = .003**
Go-Me, mm	72.1 ± 6.6	48.5 ± 19.6 md: 23.7 *P* = .069	70.2 ± 30.4 md: 2.0 *P* = .786	63.6 ± 5.3 md: 8.5 **_P_ < .001**
U1 to SN, degree	104.3 ± 9.0	106.2 ± 8.9 md: −1.9 **_P_ = .048**	103.4 ± 8.8 md: 0.9 *P* = 0.436	102.1 ± 8.6 md: 2.3 **_P_ = .048**

Table 1. Continued

Measurements	Dolphin (Mean + SD)	WebCeph (Mean + SD, md, *P* Value)	Ceph Assistant (Mean + SD, md, *P* Value)	AudaxCeph (Mean + SD, md, *P* Value)
U1 to NA, degree	24.0 ± 8.9	26.7 ± 11.3 md: −2.6 *P* = .202	21.9 ± 9.0 md: 2.1 *P* = .076	21.6 ± 8.5 md: 2.5 ***P* = .007**
U1 to NA, mm	4.3 ± 2.6	3.9 ± 3.3 md: 0.5 *P* = .560	3.1 ± 3.4 md: 1.3 ***P* = .022**	3.8 ± 2.7 md: 0.5 ***P* = .032**
L1 to MP, degree	93.9 ± 8.2	95.0 ± 9.7 md: −1.2 *P* = .674	92.9 ± 7.5 md: 1.0 *P* = .556	89.5 ± 20.1 md: 4.4 *P* = .370
L1 to NB, degree	26.6 ± 8.3	24.3 ± 9.3 md: 2.4 *P* = .305	25.6 ± 6.1 md: 1.0 *P* = .532	25.9 ± 8.1 md: 0.7 *P* = .539
L1 to NB, mm	4.6 ± 2.7	3.4 ± 2.3 md: 1.2 ***P* = .09**	4.0 ± 2.2 md: 0.6 *P* = .283	4.5 ± 2.3 md: 0.1 *P* = .849
UL to E-line, mm	−2.2 ± 2.7	−1.7 ± 3.2 md: −0.6 *P* = .282	−1.5 ± 1.7 md: −0.7 ***P* = .04**	3.3 ± 5.8 md: −5.6 ***P* = .007**
LL to E-line, mm	0.1 ± 2.6	0.0 ± 2.6 md: 0.1 *P* = .767	0.7 ± 2.2 md: −0.6 *P* = .217	0.8 ± 7.1 md: −0.7 *P* = .718

[a] md indicates mean difference compared to Dolphin; *P* = *P* values obtained with paired *t*-test; SD, standard deviation.

Intersoftware Comparisons

ANOVA analysis showed statistically significant differences among the AI software programs for most of the cephalometric measurements ($P < .05$), except for SNB (^), U1 to SN (^), and L1 to MP (^) in the CHP group, and Go-Me (mm), U1 to SN (^), and L1 to MP (^) in the IHP group. In two-by-two comparisons, WebCeph and Ceph Assistant did not show significant differences for all measurements, except for Go-Me (mm) ($P < .001$) in the IHP group. However, AudaxCeph showed statistically significant differences in several cephalometric measurements ($P < .05$) except for SNB (^), U1 to SN (^), L1 to MP (^) in the CHP group and Go-Me (mm), U1 to SN (^) and L1 to MP (^) in the IHP group ($P > .05$).

DISCUSSION

Gradual refinement of AI-driven automatic cephalometric analysis accuracy has been achieved over the years thanks to the development of increasingly high-performance algorithms.[1,24,25] However, the key question is whether an orthodontist could simply load a lateral cephalogram into a fully automated system and expect reliable and predictable cephalometric analysis without having to intervene. In this context, it is important to assess AI-based system accuracy in common scenarios in which cephalograms are taken with the head slightly rotated or inclined. In this regard, this study tested the accuracy of three fully automated AI-based cephalometric analysis software applications compared to the computer-aided tracing with Dolphin Imaging software, performed using cephalograms with correct and incorrect head posture.

Based on the current findings, the fully automated systems tested exhibited only partial consistency in the CHP group. Linear parameters, such as Go-Me (mm), U1 to NA (mm), and L1 to NB (mm) showed greater variability. In contrast, angular measurements such as SN-MP (^) and SNB (^) showed higher accuracy across all software platforms, likely due to their dependence on the relative positioning of landmarks rather than absolute distances. This finding was consistent with prior research, which also identified linear measurements as more error-prone in fully automated systems.[17,18] Indeed, to bridge the gap between manual and fully automated approaches, software for semi-automated cephalometric analysis[14,16] has been developed. Semi-automatic AI-based software allows human correction of AI-generated landmark identification, with the potential to improve accuracy by leveraging the strengths of human expertise and AI-driven precision.[26] Obviously, the trade-off is reduced efficiency compared to a fully automated system.

An intriguing factor influencing the performance of fully automated methods was the position of the cephalometric landmarks. Among the three AI-based systems evaluated, AudaxCeph exhibited the most consistent performance for measurements derived from landmarks located within bony structures, such as SNA (^) and SNB (^) angles, as well as U1 to NA (mm) and U1 to NA (^). This indicated that AudaxCeph's algorithm

Table 2. Comparisons of AI Fully Automated Cephalometric Analysis Based on WebCeph, Ceph Assistant, and AudaxCeph

Measurements	ANOVA Test *P* Value	Post-hoc WebCeph vs Ceph Assistant *P* Value	Post-hoc WebCeph vs AudaxCeph *P* Value	Post-hoc Ceph Assistant vs Audax Ceph *P* Value
CHP Group (n = 20)				
SNA, degree	**.032**	.345	**.023**	.412
SNB, degree	.066	.944	.072	.144
ANB, degree	**.001**	.833	**.010**	**.002**
FMA, degree	<**.001**	.205	**.001**	<**.001**
Go-Me, mm	**.009**	.388	.164	**.006**
U1 to SN, degree	.220	.999	.283	.273
U1 to NA, degree	**.001**	.147	<**.001**	.084
U1 to NA, mm	<**.001**	.317	<**.001**	**.021**
L1 to MP, degree	.072	.943	.079	.158
L1 to NB, degree	**.001**	.858	**.001**	**.007**
L1 to NB, mm	**.002**	.725	**.002**	**.017**
UL to E-line, mm	**.001**	.827	**.006**	**.001**
LL to E-line, mm	**.002**	.917	**.003**	**.009**
IHP Group (n = 20)				
SNA, degree	**.001**	.530	**.029**	**.001**
SNB, degree	**.026**	.907	**.028**	.078
ANB, degree	**.001**	.531	**.015**	<**.001**
FMA, degree	<**.001**	.093	**.005**	<**.001**
Go-Me, mm	.335	<**.001**	.406	.406
U1 to SN, degree	.167	.968	.187	.284
U1 to NA, degree	<**.001**	.498	**.004**	<**.001**
U1 to NA, mm	**.004**	.997	**.012**	**.009**
L1 to MP, degree	.089	.977	.111	.168
L1 to NB, degree	<**.001**	.337	**.010**	<**.001**
L1 to NB, mm	<**.001**	.220	**.009**	<**.001**
UL to E-line, mm	**.001**	.913	**.005**	**.001**
LL to E-line, mm	**.003**	.614	**.041**	**.003**

excels in identifying internal landmarks with distinct radiographic boundaries. However, its reliability diminished when applied to soft tissue measurements, such as UL to E-line (mm) and LL to E-line (mm), likely due to the challenges of locating landmarks along the external contours of soft tissues. WebCeph and Ceph Assistant exhibited more variable performance. WebCeph achieved good agreement for some parameters like SNB (^), and U1 to SN (^), but struggled with linear measurements, particularly Go-Me (mm), U1 to NA (mm), and L1 to NB (mm) measurements. Ceph Assistant exhibited similar variability, with notable weaknesses in linear measurements like Go-Me (mm). Interestingly, Ceph Assistant performed better for soft tissue measurements, suggesting that its algorithms may prioritize external landmark identification differently. These findings highlight the importance of selecting AI cephalometric software based on the clinician's specific needs. For applications in which skeletal landmarks are the primary focus, AudaxCeph may provide more consistent results, whereas Ceph Assistant appeared to be more reliable for soft tissue measurements. The variability observed in WebCeph and Ceph Assistant for linear parameters further reinforces the need for manual verification, particularly when assessing critical dimensions such as Go-Me (mm). These insights may help orthodontists integrate AI-assisted cephalometry more effectively into clinical workflows, balancing efficiency with accuracy.

A novel contribution of this study was the evaluation of AI performance in cephalograms with incorrect head posture. Positional errors reduced the landmarking accuracy across most parameters, in particular for linear parameters. The impact of positional errors was most pronounced for Go-Me (mm), which relies on the precise identification of mandibular landmarks. All three systems exhibited poor agreement for this parameter in the IHP group. These results were in agreement with prior literature, which consistently highlighted the challenges associated with the localization of the Go landmark and ruler calibration.[19] In contrast, angular measurements such as SNB (^) and SN-MP (^) obtained good-to-excellent agreement in the IHP, suggesting that angular measurements are more robust to positional variability. These findings are of clinical relevance, as they highlight the limitations of actual, fully automated systems in analyzing cephalograms that may not meet ideal imaging conditions.

Beyond the accuracy factor, one other primary concern related to AI applications in orthodontics is data protection.[27] Although online AI-based cephalometric tools offer the possibility of faster and less time-consuming cephalometry, they raise significant concerns regarding patient data privacy and security.[28] To comply with regulations such as the GDPR (General Data Protection Regulation)[27] in the European Union and HIPAA (Health Insurance Portability and Accountability Act),[27] patient data should be anonymized before uploading, by removing both direct and indirect identifiers.[28] Additionally, encryption protocols should secure data both in transit and at rest, using advanced standards like the Advanced Encryption Standard with a key length of at least 256 bits (AES-256).[29] While waiting for a centralized protocol for data encryption and sharing for AI-based technology, informed consent is crucial to notify patients of risk of data breaches and loss of privacy and their rights regarding data processing and storage.[27] Implementing these safeguards can help balance AI's clinical utility with ethical and legal responsibilities.

Limitations

The data provided by this study must be interpreted considering limitations related to the retrospective design and the impossibility of defining the exact degree of rotation of the patient's head in the IHP group.

CONCLUSIONS

- Fully automated AI-based cephalometric software demonstrated variable accuracy depending on the parameters concerned, with angular measurements showing higher reliability compared to linear parameters.
- This pattern was exacerbated under conditions involving head positional errors in cephalograms.
- Therefore, the observed discrepancies for key measurements indicate that clinician oversight remains essential and that semi-automated systems that allow manual corrections, rather than fully automated systems, should be preferred.

ACKNOWLEDGMENT

This work was supported by the European Union's NextGenerationEU initiative under the Italian Ministry of University and Research as part of the PNRR - M4C2-I1.3 Project PE00000019 'HEAL ITALIA', CUP I53C22001440006, awarded to Dr. Alessandro Polizzi as PhD Student.

REFERENCES

1. Leonardi R, Giordano D, Maiorana F, Spampinato C. Automatic cephalometric analysis: a systematic review. *Angle Orthod.* 2008;78(1):145–151.
2. Celik E, Polat-Ozsoy O, Toygar Memikoglu TU. Comparison of cephalometric measurements with digital versus conventional cephalometric analysis. *Eur J Orthod.* 2009;31(3):241–246.
3. Chen S-K, Chen Y-J, Yao C-CJ, Chang H-F. Enhanced speed and precision of measurement in a computer-assisted digital cephalometric analysis system. *Angle Orthod.* 2004;74(4):501–507.
4. Houston W. The analysis of errors in orthodontic measurements. *Am J Orthod.* 1983;83(5):382–390.
5. Houston W, Maher R, McElroy D, Sherriff M. Sources of error in measurements from cephalometric radiographs. *Eur J Orthod.* 1986;8(3):149–151.
6. Kamoen A, Dermaut L, Verbeeck R. The clinical significance of error measurement in the interpretation of treatment results. *Eur J Orthod.* 2001;23(5):569–578.
7. Hwang H-W, Park J-H, Moon J-H, et al. Automated identification of cephalometric landmarks: Part 2-Might it be better than human? *Angle Orthod.* 2020;90(1):69–76.
8. Moon J-H, Kim M-G, Cho SJ, et al. Evaluation of automated photograph-cephalogram image integration using artificial intelligence models. *Angle Orthod.* 2024;94(6):595–601.
9. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. *Nat Med.* 2019;25(1):30–36.
10. Lindner C, Wang C-W, Huang C-T, Li C-H, Chang S-W, Cootes TF. Fully automatic system for accurate localisation and analysis of cephalometric landmarks in lateral cephalograms. *Sci Rep.* 2016;6(1):33581.
11. Polizzi A, Leonardi R. Automatic cephalometric landmark identification with artificial intelligence: an umbrella review of systematic reviews. *J Dent.* 2024:105056.
12. Prince STT, Srinivasan D, Duraisamy S, Kannan R, Rajaram K. Reproducibility of linear and angular cephalometric measurements obtained by an artificial-intelligence assisted software (WebCeph) in comparison with digital software (AutoCEPH) and manual tracing method. *Dental Press J Orthod.* 2023;28(01):e2321214.
13. Jeon S, Lee KC. Comparison of cephalometric measurements between conventional and automatic cephalometric analysis using convolutional neural network. *Prog Orthod.* 2021;22:1–8.
14. Zaheer R, Shafique HZ, Khalid Z, 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.
15. Mahto RK, Kafle D, Giri A, Luintel S, Karki A. Evaluation of fully automated cephalometric measurements obtained from web-based artificial intelligence driven platform. *BMC Oral Health.* 2022;22(1):132.
16. Yassir, Yassir A., Aya R., Salman, and Sarah A. Nabbat. "The accuracy and reliability of WebCeph for cephalometric analysis." Journal of Taibah University Medical Sciences. 17.1(2022): 57–66.
17. Meriç P, Naoumova J. Web-based fully automated cephalometric analysis: comparisons between app-aided, computerized, and manual tracings. *Turk J Orthod.* 2020;33(3):142.
18. 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):481–490.
19. Hendrickx J, Gracea RS, Vanheers M, et al. Can artificial intelligence-driven cephalometric analysis replace manual tracing? A systematic review and meta-analysis. *Eur J Orthod.* 2024;46(4).

20. Brodie AG. *On the Growth Pattern of the Human Head*. University of Illinois at Urbana-Champaign; 1940.
21. Yoon Y-J, Kim K-S, Hwang M-S, Kim H-J, Choi E-H, Kim K-W. Effect of head rotation on lateral cephalometric radiographs. *Angle Orthod.* 2001;71(5):396–403.
22. American Board of Orthodontics. Cephalometric Review. Highlights of Tracing, ABO Ceph Analysis, Regional Anatomy, Superimposition Techniques & Interpretation.
23. Hwang H-W, Moon J-H, Kim M-G, Donatelli RE, Lee S-J. Evaluation of automated cephalometric analysis based on the latest deep learning method. *Angle Orthod.* 2021;91(3): 329–335.
24. de Queiroz Tavares Borges Mesquita G, Vieira WA, Vidigal MTC, et al. Artificial intelligence for detecting cephalometric landmarks: a systematic review and meta-analysis. *J Digit Imaging.* 2023;36(3):1158–1179.
25. Dot G, Rafflenbeul F, Arbotto M, Gajny L, Rouch P, Schouman T. Accuracy and reliability of automatic three-dimensional cephalometric landmarking. *Int J Oral Maxillofacial Surg.* 2020;49(10):1367–1378.
26. Sommer T, Ciesielski R, Erbersdobler J, Orthuber W, Fischer-Brandies H. Precision of cephalometric analysis via fully and semiautomatic evaluation of digital lateral cephalographs. *Dentomaxillofac Radiol.* 2009;38(6):401–406.
27. Leonardi R, Vaiid N. *Artificial Intelligence in Orthodontics: Concerns, Conjectures, and Ethical Dilemmas*. Elsevier; 2024.
28. Murdoch B. Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Medical Ethics.* 2021;22:1–5.
29. Liu G. The application of data encryption technology in computer network communication security. *Mob Inf Syst.* 2022; 2022(1):3632298.