Zaheer *et al. BMC Medical Informatics and Decision Making* (2024) 24:271
https://doi.org/10.1186/s12911-024-02664-3

BMC Medical Informatics and Decision Making

RESEARCH **Open Access**

Comparison of semi and fully automated artificial intelligence driven softwares and manual system for cephalometric analysis

Rumeesha Zaheer[1], Hafiza Zobia Shafique[1], Zahra Khalid[2*], Rooma Shahid[1], Abdullah Jan[2], Tooba Zahoor[1], Ramsha Nawaz[1] and Mehak ul Hassan[1]

Abstract

Background Cephalometric analysis has been used as one of the main tools for orthodontic diagnosis and treatment planning. The analysis can be performed manually on acetate tracing sheets, digitally by manual selection of landmarks or by recently introduced Artificial Intelligence (AI)-driven tools or softwares that automatically detect landmarks and analyze them. The use of AI-driven tools is expected to avoid errors and make it less time consuming with effective evaluation and high reproducibility.

Objective To conduct intra- and inter-group comparisons of the accuracy and reliability of cephalometric tracing and evaluation done manually and with AI-driven tools that is WebCeph and CephX softwares.

Methods Digital and manual tracing of lateral cephalometric radiographs of 54 patients was done. 18 cephalometric parameters were assessed on each radiograph by 3 methods, manual method and by using semi (WebCeph) and fully automatic softwares (Ceph X). Each parameter was assessed by two investigators using these three methods. SPSS was then used to assess the differences in values of cephalometric variables between investigators, between softwares, between human investigator means and software means. ICC and paired T test were used for intra-group comparisons while ANOVA and post-hoc were used for inter-group comparisons.

Results Twelve out of eighteen variables had high intra-group correlation and significant ICC p-values, 5 variables had relatively lower values and only one variable (SNO) had significantly low ICC value. Fifteen out of eighteen variables had minimal detection error using fully-automatic method of cephalometric analysis. Only three variables had lowest detection error using semi-automatic method of cephalometric analysis. Inter-group comparison revealed significant difference between three methods for eight variables; Witts, NLA, SNGoGn, Y-Axis, Jaraback, SNO, MMA and McNamara to Point A.

Conclusion There is a lack of significant difference among manual, semiautomatic and fully automatic methods of cephalometric tracing and analysis in terms of the variables measured by these methods. The mean detection errors were the highest for manual analysis and lowest for fully automatic method. Hence the fully automatic AI software has the most reproducible and accurate results.

*Correspondence:
Zahra Khalid
zahrakhalid26@gmail.com
Full list of author information is available at the end of the article

Keywords Artificial Intelligence, WebCeph, CephX, Cephalometric analyses, Lateral cephalogram

Introduction

The use of lateral cephalogram and cephalometric analysis has been the gold standard for orthodontic diagnosis and treatment planning for decades. It was introduced by Broadbent in 1931 to study skeletal and dental malocclusions and dentofacial discrepancies [1, 2].

The cephalometric analysis can be conducted manually or by computerized methods. The manual method has been used traditionally and is taught and practiced mostly at undergraduate level in dental schools. It had been the method of choice for cephalometric analysis for almost two decades. Whereas, the digital/computerized methods are the recent innovation in dentistry that utilize AI (Artificial Intelligence) to record craniofacial landmarks and calculate desired measurements for various analyses. Due to this rapid advancement in technology and use of AI in digital radiography and other fields of dentistry, the traditional method of cephalometric analysis is being replaced by these latest computerized methods. The AI-based methods of cephalometric analysis can be semi-automatic or fully automatic. The fully-automatic method uses AI to trace, identify landmarks, and calculate the cephalometric measurements. Whereas, the semi-automatic method involves a combination of both methods i.e., manual selection of landmarks followed by automated calculation of values. Nowadays, smartphone applications are also used to facilitate cephalometric analysis [3, 4].

The traditional manual method of cephalometric analysis is prone to systematic and random errors and is also time consuming [5]. It has often been seen that there's lack of intra-rater and inter-rater consistency and reproducibility. However, the computerized method is easier, quicker, more precise, allows data sharing and is capable of standardized evaluation with higher reproducibility [6, 7]. It is capable of completely eliminating the mechanical errors in drawing lines between landmarks and by-hand measurements with a protractor [8]. Moreover, it avoids the hassle of preparing and storing hardcopies of radiographs and radiographic analysis and provides a digital copy to the dental practice and the patients. However, there may be some disadvantages of these computerized softwares which include errors due to factors such as the quality of the radiograph, magnification errors, and due to adjustment of density, contrast and image quality. Moreover, most of these softwares are relatively expensive [9, 10].

These softwares are becoming increasingly popular due to so many merits which outweigh the minor demerits. Based on literature review, the evidence regarding the accuracy and reliability of one method of cephalometric analysis over others is inconclusive. A number of studies have shown that there's no significant difference in accuracy between manual method and AI-based softwares for cephalometric analysis [2, 6, 11, 12]. Al-Nasseri reported angular measurements in cephalometrics to be comparable for two methods but of low clinical importance [11]. Farooq et al. assessed 30 variables from five analyses (Steiners, Witts, Tweeds, McNamara, Rakosi Jarabacks) in their study on 50 patients. They did not find any statistically significant difference but reported at least five parameters to have varying results between two methods [2].

However, some studies [2, 3, 10, 13] found statistically significant differences between manual and AI-based methods. Hwang et al., (2020) reported AI to be more accurate than manual method for 14 out of 46 landmarks measured in their study, while another 14 variables were found to be more accurately measured by manual method as compared to AI-based method. The results for these 28 (14+14) variables were statistically significant while remaining 18 out of 46 did not show statistically significant differences between AI and manual methods [13]. Another recent study used a fully automated algorithm and concluded 23 cephalometric landmarks to be identified quickly with high accuracy [14].

Considering these varied findings, the primary aim of this study was to compare the accuracy of cephalometric analysis conducted using AI-driven tools and conventional method of manual tracing. The secondary aim was to compare the reproducibility and reliability of the three methods i.e. the manual and the AI-driven softwares (WebCeph [15] and CephX [16]) used for cephalometric analysis.

Methods

This was a cross-sectional, comparative, and quantitative study approved by the Ethics Research Committee of Armed Forces Institute of Dentistry. The study started with the recruitment of participants in December 2021 and continued until mid-May, 2022. The study involved 54 participants presenting to the Orthodontics Department of AFID. The details about the study were explained to the patients and informed consent was taken regarding use of their lateral cephalometric radiographs for the study with anonymity ensured.

The sample included both male and female patients seeking orthodontic treatment. The inclusion criteria included patients reporting to the department for treatment of malocclusion with fixed braces (as lateral ceph is advised to all patients undergoing orthodontic treatment for initial diagnostic work up), having no history of cleft lip or palate or any other deformities. The criteria for recruitment excluded those who had any history of craniofacial deformity and/or transverse discrepancy or signs and symptoms of temporomandibular joint (TMJ) disorders.

The lateral cephalograms of 54 participants were selected for the study. All the radiographs were taken by the same machine (Sirona Dental Systems GmbH®). The standard protocol for recording lateral cephalometric radiographs was followed whereby the patients were positioned in the cephalostat with the sagittal plane perpendicular to the path of the X-rays, the Frankfort plane parallel to the floor, natural head position with unstrained lips and teeth in centric occlusion i.e. with teeth in maximum intercuspation by asking the patients to bite on their back teeth. After that, each radiograph was traced and analyzed for various cephalometric parameters by both manual and automatic methods. The operational definitions for these methods are as follows:

- *Manual method*: It is the traditional method of direct measurement of linear and angular cephalometric parameters using a ruler and protractor. The procedure involves using an overlay of acetate tracing paper secured over the radiograph followed by tracing and identification of the landmarks. Finally, the lines are drawn with pencil and ruler and measurements are recorded with the ruler and a protractor.
- *Automatic Method*: It is the most recent technique where AI-driven softwares are used for cephalometric analysis. After importing the digital radiograph(s) into the computer or smartphone application, the AI-driven softwares trace the radiograph, identify and mark the craniofacial landmarks and complete the analysis by linear and angular measurements of cephalometric parameters. Examples of such softwares are QuickCeph and CephX.
- *Semi-Automatic Method*: This is a combination of manual and automatic methods. In this procedure, identification of landmarks and magnification correction can be done manually by the operator after importing the radiograph in the software. Whereas, the tracing and calculation of linear and angular measurements of cephalometric parameters is done by the software/AI. For example, Dolphin imaging, WebCeph, OneCeph, Ceph Ninja application. In current study, we used WebCeph and only performed magnification correction manually while tracing, landmark identification and cephalometric measurements were all calculated by the software. The magnification correction was done using 10-point rule i.e. based on the length of the cephalostat rod, that was 10 mm, hanging from the cepalostat machine and recorded on the cephalogram.

Each lateral cephalometric radiograph was traced and analyzed manually by two investigators. For the AI based analysis the cephalograms were imported to the softwares (WebCeph and CephX) for analysis by two investigators. Both the investigators in each group did the analysis within a period of two weeks. Eighteen commonly used cephalometric parameters which are used in making a diagnosis of skeletal, dental and soft tissue relationships or proportions were used (Figs. 1 and 2).

Statistical analysis

The sample size calculation was done using the PS software version 3.1.6 that assumed that the difference in the response of matched pairs is normally distributed with mean difference and standard deviation of 2 each. The sample size was calculated to be minimum 30 for paired analysis and 54 for independent analysis. The mean values of the 18 variables calculated by the three methods of cephalometric analysis were compared. The data was found to be normal using Shapiro-Wilk tests hence parametric tests were applied. One way ANOVA (Analysis of variance) and post-hoc was used to compare the groups. The degree of agreement between examiners for each method of cephalometric analysis or intra-group comparisons was conducted using intraclass correlation (ICC; two-way mixed effect model, absolute agreement at 95% confidence interval) and paired-sample t-test. An ICC of 0.7 is generally acceptable and above 0.9 is considered excellent. Both tests (ICC and Paired T-Test) were conducted to compare two investigators with each other for all three methods of cephalometric analyses i.e., manual, as well as two different AI-driven softwares (WebCeph and CephX). The type I error of 5% was allowed and the statistical analyses were performed by the computer program SPSS version 26.0 [17] and Microsoft Excel.

Results

The mean age of the sample was 16.7 ± 1.6 years and the median age was 15.5 years. The sample comprised of 39% males(n=21) and 61% females(n=33). The mean and standard deviation of each variable has been presented in Table 1 by each method of analysis.

The intra-examiner reliability of each 'method of analysis' was assessed using intra-class correlation and paired sample T-test by comparing the two examiners (1 and 2) performing these analyses (Table 2). The coefficients for ICC showed high correlation between them for all three (manual, semi-automatic, and fully-automatic) cephalometric analysis for 12 out of 18 variables as indicated in Table 2. These variables included SNA (Sella-Nasion

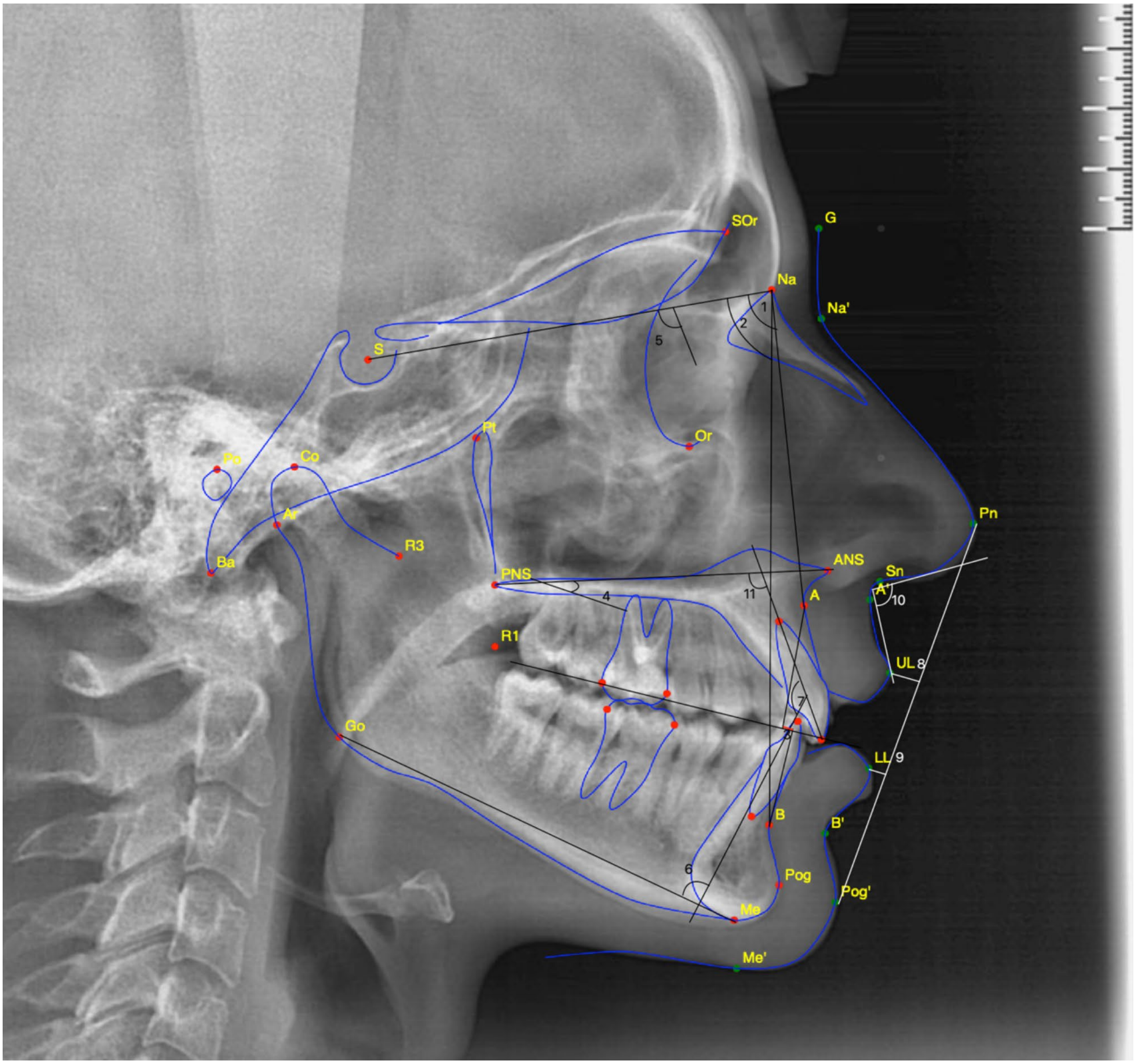

Fig. 1 Landmarks and points
1. SNA **2**. SNB **3**. Wits (Distance between A perpendicular and B perpendicular on occlusal plane) **4**. MMA **5**. ISN **6**. IMPA **7**. II **8**. E line upper lip **9**. E line lower lip **10**. NLA **11**. IPP

plane to point A angle), SNB (Sella-Nasion plane to point B angle), IMPA(Lower incisor to mandibular plane angle), Inter-Incisal angle(II), E-line to upper and lower lip, SN-Go-Gn (Sella Nasion-Gonion-Gnathion angle), Y-Axis, Jaraback, SNP (Sella-Nasion-Palatal plane angle), MMA (Maxillo-Mandibular plane angle), and IPP (Upper Incisor to Palatal plane angle) However, the ICC coefficient for five variables had relatively lower values, including manual analysis of Witts (0.597), NLA (Naso-labial angle)(0.635) and McNamara to Point A (0.578), semi-automatic analysis of Upper Incisal Inclination (0.629), and both manual and semi-automatic analyses of McNamara to Pogonion (0.503 and 0.621, respectively). The ICC coefficient values for other methods of analyses for these 5 variables were high. The lowest ICC correlation was recorded for semi-automatic analysis of SNO (0.293). Similarly, the p-values for the ICC analyses were highly statistically significant ($p \leq 0.001$) for all the variables except semi-automatic cephalometric analysis of SNO ($p = 0.108$). All these aforementioned results depict high correlation of the two examiners performing all the analyses.

The paired sample t-test compared the difference in mean values (of the 18 variables) to test the variability

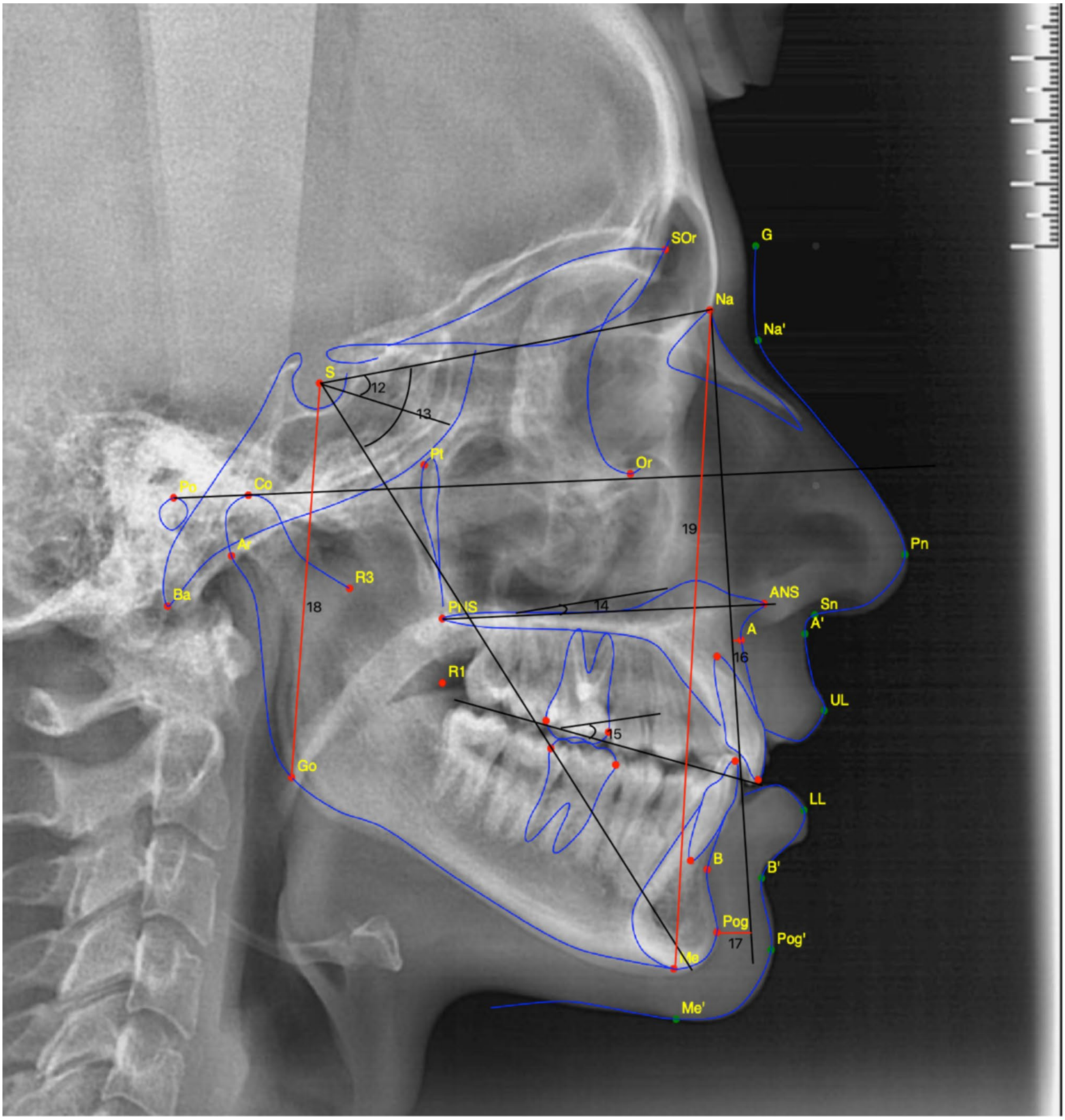

Fig. 2 Landmarks and points
12. SN-Mand **13**. Y-axis **14**. SNP **15**. SNO **16**. McNamara to point A **17**. McNamara to Pogonion
18. PFH (Posterior facial height) **19**. AFH (Anterior facial height) [Jaraback ratio is PFH/AFH]

between the two examiners performing all analyses (Table 2). An insignificant p-value ($p>0.05$) in this case supports our results of high correlation and minimal variability between the two examiners. The results revealed statistically insignificant p-value for all three analyses of ten variables including SNB, Witts, IMPA, IPP, II angle, SNGoGn, Y-Axis, Jaraback, SNP and MMA. However, a significant difference in mean values of following eight analyses were recorded; manual analyses of E-line to upper lip distance, NLA, SNO, IPP, McNamara to Point A, and McNamara to Pogonion and semi-automatic analysis of SNA. The remaining analyses for these 8 variables also showed statistically insignificant results. These results also support our following results of highest accuracy of fully-automatic without a single discrepancy recorded in ICC and paired-sample T-test.

Table 1 Descriptives (mean, standard deviations and standard error of mean)

Variable	Method	Mean	Std Dev	SE Mean	Variable	Method	Mean	Std Dev	SE Mean
SNA (degrees)	Manual 1	80.61	4.26	0.58	**SNB (degrees)**	Manual 1	75.74	3.99	0.54
	Manual 2	80.23	3.82	0.52		Manual 2	75.89	3.91	0.52
	Semi-Auto 1	82.12	3.53	0.48		Semi-Auto 1	77.09	3.45	0.47
	Semi-Auto 2	81.55	3.73	0.51		Semi-Auto 2	76.80	3.73	0.51
	Fully Auto 1	80.87	3.55	0.48		Fully Auto 1	76.25	3.67	0.50
	Fully Auto 2	80.93	3.52	0.48		Fully Auto 2	76.16	3.71	0.51
Wits (mm)	Manual 1	1.21	3.66	0.497	**MMA (degrees)**	Manual 1	25.43	5.89	0.80
	Manual 2	1.65	2.83	0.38		Manual 2	25.91	5.85	0.79
	Semi-Auto 1	3.93	3.23	0.44		Semi-Auto 1	24.37	4.48	0.61
	Semi-Auto 2	4.20	3.20	0.44		Semi-Auto 2	24.58	4.35	0.59
	Fully Auto 1	4.61	2.99	0.41		Fully Auto 1	28.61	4.75	0.65
	Fully Auto 2	4.70	2.99	0.41		Fully Auto 2	28.34	4.70	0.64
Upper Incisor Inclination (ISN) (degrees)	Manual 1	106.81	8.75	1.19	**IMPA (degrees)**	Manual 1	95.39	9.21	1.25
	Manual 2	106.23	9.33	1.27		Manual 2	95.00	9.05	1.23
	Semi-Auto 1	105.80	15.86	2.16		Semi-Auto 1	94.05	6.56	0.89
	Semi-Auto 2	107.52	9.02	1.23		Semi-Auto 2	94.89	7.10	0.96
	Fully Auto 1	107.46	7.73	1.05		Fully Auto 1	92.47	7.14	0.97
	Fully Auto 2	107.37	7.70	1.05		Fully Auto 2	92.85	7.12	0.97
Inter-Incisal Angle (II) (degrees)	Manual 1	123.43	11.68	1.59	**Upper Incisor to palatal plane (IPP) (degrees)**	Manual 1	115.37	8.76	1.19
	Manual 2	124.22	11.91	1.62		Manual 2	113.91	8.89	1.21
	Semi-Auto 1	123.65	18.85	2.57		Semi-Auto 1	116.53	8.52	1.16
	Semi-Auto 2	124.47	11.49	1.56		Semi-Auto 2	116.03	8.29	1.13
	Fully Auto 1	122.30	11.23	1.53		Fully Auto 1	116.78	7.46	1.02
	Fully Auto 2	122.01	11.15	1.52		Fully Auto 2	116.88	7.30	0.99
E-Line Upper lip (mm)	Manual 1	-2.16	4.15	0.56	**E-Line Lower lip(mm)**	Manual 1	-1.04	4.09	0.56
	Manual 2	-2.91	3.86	0.53		Manual 2	-1.09	3.55	0.48
	Semi-Auto 1	-1.24	3.53	0.48		Semi-Auto 1	-0.07	3.48	0.47
	Semi-Auto 2	-1.63	3.23	0.44		Semi-Auto 2	-0.42	3.44	0.47
	Fully Auto 1	-1.97	3.06	0.42		Fully Auto 1	-1.0	3.53	0.48
	Fully Auto 2	-1.94	3.11	0.42		Fully Auto 2	-0.99	3.54	0.48
NLA (degrees)	Manual 1	108.22	13.29	1.81	**SNGoGn (degrees)**	Manual 1	34.02	6.89	0.94
	Manual 2	99.58	14.35	1.95		Manual 2	34.65	7.71	1.05
	Semi-Auto 1	105.34	20.39	2.77		Semi-Auto 1	32.84	6.06	0.83
	Semi-Auto 2	102.71	18.17	2.47		Semi-Auto 2	33.12	6.34	0.86
	Fully Auto 1	111.83	10.92	1.49		Fully Auto 1	37.74	5.66	0.77
	Fully Auto 2	112.15	10.89	1.48		Fully Auto 2	37.75	5.66	0.77
Y-Axis (degrees)	Manual 1	69.19	4.84	0.66	**Jaraback ratio (degrees)**	Manual 1	65.02	5.24	0.71
	Manual 2	69.81	5.06	0.69		Manual 2	64.27	6.04	0.82
	Semi-Auto 1	68.39	4.16	0.57		Semi-Auto 1	66.52	4.91	0.67
	Semi-Auto 2	68.57	4.35	0.59		Semi-Auto 2	66.31	4.99	0.68
	Fully Auto 1	60.37	4.24	0.58		Fully Auto 1	61.50	3.99	0.54
	Fully Auto 2	59.99	3.25	0.44		Fully Auto 2	61.50	4.00	0.54
SNO (degrees)	Manual 1	19.57	5.79	0.79	**SNP (degrees)**	Manual 1	9.83	4.03	0.55
	Manual 2	20.81	5.37	0.73		Manual 2	9.21	2.68	0.36
	Semi-Auto 1	15.04	6.86	0.93		Semi-Auto 1	8.58	2.96	0.40
	Semi-Auto 2	16.09	14.89	2.03		Semi-Auto 2	8.51	2.92	0.39
	Fully Auto 1	13.34	4.79	0.65		Fully Auto 1	9.31	2.98	0.41
	Fully Auto 2	13.51	4.83	0.66		Fully Auto 2	9.42	2.96	0.40

Table 1 (continued)

Variable	Method	Mean	Std Dev	SE Mean	Variable	Method	Mean	Std Dev	SE Mean
McNamara to Point A (mm)	Manual 1	-1.89	3.86	0.53	**McNamara to Pogonion (mm)**	Manual 1	-9.90	6.67	0.91
	Manual 2	0.07	3.53	0.48		Manual 2	-5.12	6.77	0.92
	Semi-Auto 1	1.01	3.32	0.45		Semi-Auto 1	-5.07	7.13	0.97
	Semi-Auto 2	0.61	3.39	0.46		Semi-Auto 2	-5.92	6.16	0.84
	Fully Auto 1	0.66	2.58	0.35		Fully Auto 1	-5.06	6.08	0.83
	Fully Auto 2	0.69	2.56	0.35		Fully Auto 2	-5.41	5.95	0.81

*Semi-Auto refers to Semi-Automatic WebCeph, Fully Auto refers to Fully Automatic CephX; 1 refers to 1st examiner and 2 refers to 2nd examiner

Table 2 Intra-group comparison using ICC and paired T-test

Variable	Method (by examiners 1 & 2)	Intra-Class Corr	ICC *p*-value	Pair t-test *p*-value	Variable	Method (by examiners 1 & 2)	Intra-Class Corr	ICC *p*-value	Pair t-test *p*-value
SNA	Manual	0.884	0.000	0.289	**SNB**	Manual	0.940	0.000	0.566
	Semi-Auto	0.907	0.000	0.045		Semi-Auto	0.942	0.000	0.212
	Fully Auto	0.996	0.000	0.398		Fully Auto	0.993	0.000	0.306
Wits	Manual	0.597	0.001	0.373	**MMA**	Manual	0.874	0.000	0.373
	Semi-Auto	0.946	0.000	0.167		Semi-Auto	0.941	0.000	0.463
	Fully Auto	0.986	0.000	0.376		Fully Auto	0.975	0.000	0.179
ISN	Manual	0.906	0.000	0.423	**IMPA**	Manual	0.746	0.000	0.218
	Semi-Auto	0.629	0.000	0.352		Semi-Auto	0.931	0.000	0.077
	Fully Auto	0.997	0.000	0.383		Fully Auto	0.964	0.000	0.301
II	Manual	0.949	0.000	0.262	**IP**	Manual	0.940	0.000	0.010
	Semi-Auto	0.722	0.000	0.683		Semi-Auto	0.985	0.000	0.068
	Fully Auto	0.991	0.000	0.297		Fully Auto	0.997	0.000	0.370
E-Line Upper lip	Manual	0.920	0.000	0.011	**E-Line Lower lip**	Manual	0.953	0.000	0.804
	Semi-Auto	0.911	0.000	0.145		Semi-Auto	0.962	0.000	0.052
	Fully Auto	0.998	0.000	0.547		Fully Auto	0.998	0.000	0.529
NLA	Manual	0.635	0.000	0.000	**SNGoGn**	Manual	0.828	0.000	0.414
	Semi-Auto	0.916	0.000	0.073		Semi-Auto	0.972	0.000	0.334
	Fully Auto	0.990	0.000	0.294		Fully Auto	1.000	0.000	0.576
Y-Axis	Manual	0.830	0.000	0.224	**Jaraback**	Manual	0.819	0.000	0.218
	Semi-Auto	0.983	0.000	0.239		Semi-Auto	0.896	0.000	0.627
	Fully Auto	0.919	0.000	0.175		Fully Auto	0.999	0.000	1.000
SNO	Manual	0.887	0.000	0.010	**SNP**	Manual	0.732	0.000	0.150
	Semi-Auto	0.293	0.108	0.607		Semi-Auto	0.925	0.000	0.749
	Fully Auto	0.987	0.000	0.279		Fully Auto	0.985	0.000	0.285
McNamara to Point A	Manual	0.578	0.000	0.000	**McNamara to Pogonion**	Manual	0.503	0.001	0.000
	Semi-Auto	0.821	0.000	0.266		Semi-Auto	0.621	0.000	0.375
	Fully Auto	0.996	0.000	0.488		Fully Auto	0.972	0.000	1.88

After the intra-group comparisons, the three methods of analyses were compared for inter-group differences for each variable. Table 3 shows the comparison of mean detection errors in each group and the one with lowest error was considered most reliable, consistent, and accurate method for analyses. More than 80% variables i.e., 15 out of 18 variables had minimal detection error using fully-automatic method of cephalometric analysis. Only three variables had lowest detection error using semi-automatic method of cephalometric analysis, i.e., Y-Axis, SNP, and MMA. The p-values for Y-Axis and MMA were statistically significant. (Table 3)

The difference in mean values of variables assessed manually, by semi-automatic and fully-automatic methods were analyzed using ANOVA and tested against $p<0.05$. The results for ANOVA showed significant difference between three methods for eight variables; Witts, NLA, SNGoGn, Y-Axis, Jaraback, SNO, MMA and McNamara to Point A (Table 3). Therefore, post-hoc analyses were performed to study these differences. The results for post-hoc analyses revealed following:

Table 3 Inter-group comparison of accuracy of manual vs. semi-automatic vs. fully-automatic using ANOVA

S.no	Variable	Detection Error Manual		Detection Error Semi-Auto		Detection Error Fully Auto		More Accurate Results from	ANOVA*
		Mean (degrees)	SD	Mean (degrees)	SD	Mean (degrees)	SD		*P*-value
1	SNA	0.379	2.60	0.573	2.06	-0.05	0.46	Fully-Auto	0.049
2	SNB	-0.29	1.80	0.31	1.74	0.09	0.63	Fully-Auto	0.268
3	Wits (mm)	-4.35	3.56	-0.27	1.44	-0.09	0.71	Fully-Auto	0.000***
4	ISN	0.58	5.31	-1.71	13.42	0.09	0.79	Fully-Auto	0.864
5	IMPA	1.39	8.18	-0.84	3.42	0.08	2.67	Fully-Auto	0.097
6	II	-0.79	5.17	-0.82	14.63	0.29	2.07	Fully-Auto	0.679
7	E-Line Upper lip (mm)	0.75	2.08	0.38	1.91	-0.02	0.31	Fully-Auto	0.251
8	E-Line Lower lip (mm)	0.05	1.63	0.35	1.28	-0.02	0.28	Fully-Auto	0.413
9	NLA	8.64	13.08	2.63	10.57	-0.31	2.17	Fully-Auto	0.004***
10	SNGoGn	-0.63	5.61	-0.27	2.06	-0.01	0.19	Fully-Auto	0.000***
11	Y-Axis	-0.63	3.76	-0.18	1.11	0.38	2.05	Semi-Auto	0.000***
12	Jaraback	0.75	4.41	0.20	3.05	0.00	0.21	Fully-Auto	0.000***
13	SNO	-1.24	3.40	-1.05	14.94	-0.16	1.09	Fully-Auto	0.000***
14	SNP	0.62	3.12	0.07	1.57	-0.10	0.72	Semi-Auto	0.183
15	MMA	-0.48	3.93	-0.21	2.09	0.27	1.45	Semi-Auto	0.000***
16	IPP	1.46	4.00	0.50	1.97	-0.09	0.76	Fully-Auto	0.348
17	McNamara A (mm)	-1.96	3.84	0.39	2.61	-0.03	0.31	Fully-Auto	0.004***
18	McNamra Pog (mm)	-4.78	7.26	0.85	6.99	0.36	1.97	Fully-Auto	0.083

*One-way ANOVA was performed to compare mean values of each variable across three methods of analysis

*** Highly statistically significant p-value

- Mean value of Witts analysis to be significantly higher by Semi-automatic method of analysis than by Manual method.
- Mean values of NLA, SNGoGn, and MMA to be significantly greater by Fully-automatic method than both Manual and Semi-automatic methods.
- Mean values of Y-Axis and Jaraback to be significantly higher using Manual and Semi-automatic methods of analyses as compared to their values by Fully-automatic method.
- Mean value of SNO to be significantly greater by Manual method than by using Semi- and Fully-automatic methods.
- Mean value of McNamara to Point A to be significantly greater by Semi- and Fully-automatic methods than by manual method.

Discussion

The study provided a detailed comparative analysis of eighteen cephalometric parameters obtained by manual and AI based methods. Each method was performed by two investigators for the assessment of intra-group reliability i.e., author 1's manual analysis vs. author 2's manual analysis, author 1's semi-automatic analysis vs. author 2's semi-automatic analysis and author 1's fully-automatic analysis vs. author 2's fully-automatic analysis.

The intra-group comparisons were conducted using ICC and paired T-test and the goal was to obtain high ICC values with significant p-values and insignificant p-values for paired T-test to reconfirm lack of significant difference between examiners. Both the desired results were proof of high intra-group reliability. Our results (Table 2) showed that both examiners had high correlation for majority of variables, irrespective of method of analysis. 12 out of 18 variables had high intra-group correlation and significant ICC p-values, 5 variables had relatively lower ICC values and only one variable (SNO) had significantly low ICC value. The ICC revealed that the inter-examiner reliability was excellent for fully-automatic method of analysis i.e., 0.919 to 1.000. For semi-automatic method, it ranged from 0.621 to 0.985 with majority values above 0.8. Only one ICC value for semi-automatic method was very low (i.e., 0.293) for SNO, which could be due to an error. Finally, for manual method, the values ranged from 0.503 to 0.953. Therefore, we failed to reject the null hypothesis since there was no significant difference in the cephalometric analysis performed by 2 examiners using manual, semi-automatic WebCeph and fully-automatic CephX softwares. A similar study compared WebCeph and AutoCAD at different time intervals by a single examiner and reported both programs to have adequate reliability [4].

After the intra-group comparisons to assess the reliability of each method, the next aim of the study was to find the most accurate/valid method with minimal detection error by comparing the three methods of analyses. The results showed that 15 out of 18 variables had minimal detection error using fully-automatic method of cephalometric analysis. Remaining three variables had lowest detection error using semi-automatic method of cephalometric analysis. Hence, fully-automatic AI software CephX had the most reliable and accurate results, followed by semi-automatic WebCeph software.

Yassir et al [4] compared WebCeph (both semi and fully automated options) and AutoCAD (semi-automated) and reported adequate reproducibility with both. Their results were in agreement with our study as they found no significant difference between two programs using Paired T-Test or ICC. Another study [18] comparing the reliability of the android smartphone-based app One-Ceph (semi-automated) with computer cephalometric tracing program Dolphin Imaging software (semi-automated) reported only four out of 15 parameters showing significant differences. One more study [19] comparing diagnostic accuracy of two smartphone applications (CephNinja and OneCeph, both semi automated) with Viewbox (semi automated) software for cephalometric analysis reported OneCeph to have high validity as compared Viewbox while the reliability of CephNinja was comparable to Viewbox. Hence, they also concluded the performance of these digital softwares to be satisfactory with high potential to replace manual methods in future.

While comparing our results of manual versus computerized, our study found significant differences for several variables (Witts, NLA, SNGoGn, Y-Axis, Jaraback, SNO, MMA and McNamara to Point A). The study of Farooq et al., [2] reported consistency between the two methods for majority of the variables except 1-NA, Y-axis and interincisal angle. Hence, the two studies had agreement regarding difference in manual vs. computerized for one variable i.e., Y-Axis only. Both studies had higher mean value of Y-Axis recorded by manual method as compared to computerized methods. Nasseri [11] found no significant difference in any measurements while comparing manual with digital, however he concluded angular measurements to be more comparable between manual and computerized methods of analyses. Albarakati et al. [3] reported that both conventional and digital cephalometric analysis, although showed some statistically significant differences, most differences were not clinically significant and therefore both are highly reliable. Erkan et al., however, reported no statistically significant difference was found between manual and digital programs [6].

Finally, one study similar to ours by Sommer et al. [20] compared the three methods as well but their findings were different from ours. The fully-automatic method for cephalometric analysis was found not to be reliable. In summary, reviewing results of majority of the studies, we conclude that even when the differences were statistically significant, they were clinically acceptable. Thus, this study indicates that all three methods provide quality results and hence the computerized (AI-based) methods are quite reliable, validating most studies that have compared different cephalometric tracing methods.

Our study also compared the inter-examiner reliability of these methods which has not been discussed in the previous studies. This shows that since the AI based methods of cephalometric analysis are quite accurate therefore they are helpful in saving the clinicians time and effort in analyzing the patients lateral cephalograms and making a quick diagnosis, even by the chairside. In modern world, time is a significant factor and tracing the radiographs, marking landmarks and doing all the calculations by hand could be very tiring and time consuming. Moreover the extra materials used for tracing like acetate tracing sheets and for their storage are also not required. These softwares are now becoming indispensable in the digital orthodontic workflow in orthodontic clinic setups.

The study strengths included adequate sample size, multiple appropriate statistical tests and examiners/operators with experience in orthodontics. Moreover, unlike most studies which compared only two i.e., either manual vs. digital or two digital applications or softwares, our study provided comparison of all three methods of analyses. In addition, the study had two examiners for each method of analysis with high ICC. Even though majority of similar studies had one examiner to avoid bias, in our study we included two examiners to assess difference in fully-automatic and semi-automatic methods when performed by two different examiners. The study wanted to assess the likelihood of differences in values when two different examiners conducted them.

Subdivision of the sample according to vertical and sagittal skeletal patterns could have further verified accuracy detection under different skeletal patterns as the morphology of the condyle also varies in varying skeletal patterns as described by Antonino Lo Giudice et al. [21]. This may be done in future studies. Cephalometric accuracy of 3D structures using AI segmentation tools may also be performed in the future as shown in recent studies on fully automatic segmentation of the mandible and upper airway using convoluted neural networks [22, 23].

Conclusion

There is a lack of significant difference in the majority of variables among the manual, semi automatic and fully automatic methods of cephalometric tracing and analysis. The mean detection errors were the highest for manual analysis and lowest for fully automatic method. Hence the fully automatic AI software has the most reproducible and accurate results. Moreover, these computerized cephalometric softwares are freely available, easy to use, efficient, and can help reduce human error.

Abbreviations

AI	Artificial Intelligence
ICC	Intra Class Correlation
SNA	Sella-Nasion-Point A angle
SNB	Sella-Nasion-Point B angle
MMA	Maxillo-mandibular plane angle
IMPA	Incisor to Mandibular plane angle
ISN	Upper incisor to SN Plane angle
IPP	Upper incisor to Palatal plane angle
II	Interincisal angle
E-line	Esthetic line
NLA	Nasolabial angle
SNGoGn	Sella Nasion-Gonion-Gnathion angle
SNO	Sella Nasion to Occlusal plane angle
SNP	Sella Nasion to Palatal plane angle

Acknowledgements
Not Applicable.

Author contributions
RZ contributed to the design, data collection, data entry and writing of the manuscript. HZS helped in data collection, analysis and manuscript writing. ZK contributed to study design, data collection, manuscript editing and correction and is also the corresponding author. RS contributed to data collection and analysis. AJ supervised the research. TZ helped in data collection. RN and MH also contributed in data collection.All authors have read and reviewed the final manuscript.

Funding
No funding was received for this research.

Data availability
The data used and analyzed in this study can be available on reasonable request to the corresponding author.

Declarations

Ethics approval and consent to participate
The study was approved by Ethics Research Committee of Armed Forces Institute of Dentistry (AFID), Rawalpindi. Methods of data collection were in accordance with the Helsinki guidelines-Ethical Principles of Medical Research involving patients data (radiographs in this research). Informed consent was taken from the patients and their parents (if less than 16 years)to include their radiographs in this research project.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Armed Forces Institute of Dentistry Rawalpindi, Rawalpindi, Pakistan
[2]Armed Forces Institute of Dentistry, National University of Medical Sciences, Rawalpindi, Pakistan

Received: 13 February 2023 / Accepted: 2 September 2024
Published online: 27 September 2024

References

1. Broadbent B. A new X-ray technique and its application to Orthodontia. The Angle Orthodontist. 1931;1(2):45–66. https://meridian.allenpress.com/angle-orthodontist/article-abstract/1/2/45/55162
2. Farooq MU, Khan MA, Imran S, Sameera A, Qureshi A, Ahmed SA, et al. Assessing the Reliability of Digitalized Cephalometric Analysis in Comparison with Manual Cephalometric Analysis. J Clin Diagn Res. 2016;10(10):20–3.
3. AlBarakati SF, Kula KS, Ghoneima AA. The reliability and reproducibility of cephalometric measurements: a comparison of conventional and digital methods. Dentomaxillofacial Radiology. 2014;41(1):11–7. https://www.birpublications.org/doi/https://doi.org/10.1259/dmfr/37010910
4. Yassir YA, Salman AR, Nabbat SA. The accuracy and reliability of WebCeph for cephalometric analysis. J Taibah Univ Med Sci. 2021;8:10.
5. Celik E, Polat-Ozsoy O, Toygar Memikoglu TU. Comparison of cephalometric measurements with digital versus conventional cephalometric analysis. Eur J Orthod. 2009;31(3):241–6.
6. Erkan M, Gurel HG, Nur M, Demirel B. Reliability of four different computerized cephalometric analysis programs. European Journal of Orthodontics. 2012;34(3):318–21. https://academic.oup.com/ejo/article/34/3/318/425059
7. Aksakalli S, Yilanci H, Gorukmez E, Ramoglu SI. Reliability Assessment of Orthodontic Apps for Cephalometrics. Turkish J Orthod. 2017;29(4):98–102.
8. Chen YJ, Chen SK, Yao JCC, Chang HF. The Effects of Differences in Landmark Identification on the Cephalometric Measurements in Traditional Versus Digitized Cephalometry. Angle Orthodontist. 2004;74(2):155–61. http://meridian.allenpress.com/angle-orthodontist/article-pdf/74/2/155/1380244/0003-3219
9. Nouri M, Hamidiaval S, Baghban AA, Basafa M, Fahim M. Efficacy of a newly designed Cephalometric Analysis Software for McNamara Analysis in comparison with Dolphin Software. J Dent. 2015;12(1):60–9. Available from:. www.jdt.tums.ac.ir.
10. Chen YJ, Chen KS, Chang HF, Chen KC. Comparison of Landmark Identification in Traditional Versus Computer-Aided Digital Cephalometry. Angle Orthodontist. 2000;70(5):387–92. http://meridian.allenpress.com/angle-orthodontist/article-pdf/70/5/387/1377382/0003-3219
11. Al-Nasseri NA. The accuracy of computerised cephalometric analysis compared to conventional manual method. Mustansiria DJ. 2005;2(1):7–14.
12. Kunz F, Stellzig-Eisenhauer A, Zeman F, Boldt J. Artificial intelligence in orthodontics: evaluation of a fully automated cephalometric analysis using a customized convolutional neural network. J Orofac Orthop. 2020;81(1):52–68.
13. 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. 2020;90(1):69–76.
14. Kim H, Shim E, Park J, Kim YJ, Lee U, Kim Y. Web-based fully automated cephalometric analysis by deep learning. Comput Methods Programs Biomed. 2020;194.
15. Korean Intellectual Property Office. WebCeph. WebCephTM patented by the Korean Intellectual Property Office. 2019 [cited 2022 Aug 7]. https://webceph.com/en/about/
16. ORCA Dental AI. CephX. 2021 [cited 2022 Aug 7]. https://cephx.com/about/
17. SPSS Statistics Version 26.0 IBM. Statistical Package for the Social Sciences. SPSS Statistics. 2021 [cited 2022 Aug 7]. https://www.ibm.com/products/spss-statistics
18. Shettigar P, Shetty S, Naik RD, Basavaraddi SM, Patil AK. A comparative evaluation of reliability of an android-based app and Computerized Cephalometric Tracing Program for Orthodontic Cephalometric Analysis. Biomedical Pharmacol J. 2019;12(1):341–6.
19. Livas C, Delli K, Spijkervet FKL, Vissink A, Dijkstara PU. Concurrent validity and reliability of cephalometric analysis using smartphone apps and computer software. Angle Orthod. 2019;89(6):889–96.
20. Sommer T, Ciesielski R, Erbersdobler J, Orthuber W, Fischer-Brandies H. Precision of cephalometric analysis via fully and semiautomatic evaluation of digital lateral cephalographs. Dentomaxillofacial Radiol. 2009;38(6):401–6.
21. Lo Giudice A, Rustico L, Caprioglio A, Migliorati M, Nucera R. Evaluation of condylar cortical bone thickness in patient groups with different vertical facial dimensions using cone-beam computed tomography. Odontology. 2020;108(4):669–75. https://doi.org/10.1007/s10266-020-00510-2. Epub 2020 Mar 31. PMID: 32236830.
22. Lo Giudice A, Ronsivalle V, Spampinato C, Leonardi R. Fully automatic segmentation of the mandible based on convolutional neural networks (CNNs). Orthod Craniofac Res. 2021;24(Suppl 2):100–7. https://doi.org/10.1111/ocr.12536. Epub 2021 Dec 1. PMID: 34553817.

23. Leonardi R, Lo Giudice A, Farronato M, Ronsivalle V, Allegrini S, Musumeci G, Spampinato C. Fully automatic segmentation of sinonasal cavity and pharyngeal airway based on convolutional neural networks. Am J Orthod Dentofacial Orthop. 2021;159(6):824–835.e1. https://doi.org/10.1016/j.ajodo.2020.05.017. PMID: 34059213.

Publisher's note