applied sciences

MDPI

Article

Repeatability and Agreement of CephX AI-Driven Cephalometric Analysis in Patients with Cleft Palate Only: A Comparative Study with Non-Cleft Controls

Alicja Zawiślak [1,2,*], Natalia Kazimierczak [3,4], Wojciech Kazimierczak [3,5], Katarzyna Grocholewicz [1] and Joanna Janiszewska-Olszowska [1,*]

[1] Department of Interdisciplinary Dentistry, Pomeranian Medical University in Szczecin, 70-111 Szczecin, Poland; katarzyna.grocholewicz@pum.edu.pl
[2] Department of Maxillofacial Orthopedics and Orthodontics, Institute of Mother and Child, 01-211 Warsaw, Poland
[3] Kazimierczak Private Medical Practice, 85-009 Bydgoszcz, Poland; natalia.kazimierczak@cm.umk.pl (N.K.); w.kazimierczak@cm.umk.pl (W.K.)
[4] Department of Dentofacial Orthopedics, Collegium Medicum, Nicolaus Copernicus University in Torun, 85-067 Bydgoszcz, Poland
[5] Department of Radiology and Diagnostic Imaging, Collegium Medicum, Nicolaus Copernicus University in Torun, 85-067 Bydgoszcz, Poland
* Correspondence: alicja.zawislak@pum.edu.pl (A.Z.); joanna.janiszewska.olszowska@pum.edu.pl (J.J.-O.)

Abstract

Cephalometric analysis is essential in patients with cleft palate only (CPO) for the assessment of craniofacial morphology and treatment planning. This study evaluated the test–retest repeatability of CephX AI-driven cephalometric analysis and its agreement with expert manual measurements in patients with CPO versus non-cleft controls. Cephalometric radiographs of 50 patients with CPO (25 males and 25 females; aged 14–23 years) and 50 non-cleft controls matched for sex and ethnic origin were traced manually by two experienced orthodontists using CephX. All analyses were repeated after a two-week interval. Repeatability was assessed using Lin's concordance correlation coefficient (CCC), and agreement between AI and manual measurements was evaluated using Bland–Altman analysis. CephX showed high test–retest repeatability for both groups, with CCC values of 0.97 or higher for all cephalometric-variables evaluated. However, the agreement with expert manual measurements varied among cephalometric variables. In the CPO group, 34 of 50 radiographs required manual scale calibration and 31 of 50 generated a landmark-review warning; nevertheless, no skeletal or dental landmarks included in the analysis were manually repositioned by the investigators. CephX provided highly reproducible measurements, but the need for calibration and landmark review indicates that clinician supervision remains important, particularly in patients with altered craniofacial anatomy.

Keywords: cleft palate; cephalometry; artificial intelligence; CephX; craniofacial morphology

Academic Editors: Nir Shpack

Received: 29 July 2026
Revised: 28 September 2026
Accepted: 28 October 2026
Published: 2 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Cleft palate only (CPO) is among the most prevalent craniofacial structural birth defects; it occurs in 0.33 of every 1000 live births. Each cleft type is inherently associated with altered craniofacial morphology and disturbed craniofacial growth, which can be

described using values from cephalometric analysis. Understanding the nature and the severity of skeletal discrepancy resulting both from altered craniofacial morphology and from essential surgical interventions allows reasonable planning of orthodontic treatment and surgical interventions. Thus, in cleft patients, performing cephalometric analysis is essential for preparing a treatment plan [1].

It is known that craniofacial morphology in CPO is significantly different compared to that in unilateral cleft lip and palate (UCLP) or in bilateral cleft lip and palate (BCLP) [2]. It has been found that CPO is associated with a maxillary deficiency and a mandibular distal rotation with a reduced ANB angle and increased intermaxillary angle [3,4].

Currently, an increasing number of AI-supported cephalometric analysis services using automatic landmark identification are available on the market, including the most popular ones—CephX, WeDoCeph, WebCeph, and AudaxCeph—and it can be expected that new software will emerge on the market. Automatic cephalometric analysis based on automatic landmark identification is characterized by excellent accuracy and repeatability [5]. Numerous studies have been published assessing applications of these systems for automatic landmark cephalometric and AI-driven cephalometric analysis [5–13].

In general, it is reported that AI-driven cephalometric landmark identification is more repeatable than that manually performed, even by experienced clinicians [14,15]. However, despite improving performance, the reliability and accuracy of AI algorithms in cephalometric analysis is still questionable [16]. The performance of AI algorithms highly depends on the volume and quality of the training datasets. Furthermore, the accuracy of AI systems in segmenting and interpreting medical images is reliant on comprehensive datasets that encompass diverse patient characteristics, including age, sex, ethnic background, and disease stages [17]. Therefore, it is hypothesized that the identification of cephalometric landmarks in patients with orofacial clefts, including CPO, might be associated with reduced reliability. Despite the increasing use of AI tools and the routine need for cephalometric analysis in cleft patients, there is a notable lack of studies conducted on these specific patients. A single study could be found assessing the reliability of AI-driven cephalometric analysis in cleft lip and palate patients [18]. No studies on CPO patients could be found so far; it can be assumed that these tools have not been tested on CPO patients. That is why the aim of the present study was to assess the reliability of AI-driven cephalometric analysis in CPO patients compared to non-cleft individuals. The aim of the present study was to assess the test–retest repeatability of CephX AI-driven cephalometric analysis in patients with CPO and its agreement with expert manual measurements, compared with non-cleft controls. We hypothesized that CPO-specific craniofacial morphology would be associated with lower AI repeatability and poorer agreement with expert manual measurements than that observed for non-cleft controls.

2. Materials and Methods

2.1. Sample

This study included 50 patients with CPO (25 males and 25 females) aged 14–23 years and 50 non-cleft controls. Controls were matched to the CPO group for sex and ethnic origin. Participants were selected from patients treated at the study outpatient clinic. The sample size was verified using an online calculator (clincalc.com), based on the observed median and interquartile range values of SNA; the estimated standardized effect size was approximately 0.73 for manual tracing by orthodontists and 1.30 for AI tracing. Assuming a two-sided significance level of 0.05, a power of 80%, and equal allocation between groups, a minimum of 31 participants per group were required for the manual tracing and 11 for the AI.

The inclusion criteria for the CPO group were diagnosis of isolated cleft palate only (CPO), treatment at the study outpatient clinic, and availability of a lateral cephalometric

radiograph of sufficient diagnostic quality to allow reliable cephalometric analysis. The exclusion criteria were the presence of craniofacial anomalies other than CPO, including associated syndromic conditions; Pierre Robin sequence; and insufficient radiographic quality preventing reliable identification of cephalometric landmarks. All lateral cephalometric radiographs were obtained using the same imaging equipment (Cranex Tom digital X-ray system, Soredex, Tuusula, Finland) and according to a standardized acquisition protocol.

All cephalometric radiographs were anonymized and assigned randomly generated identification codes prior to analysis. All information that could reveal patient identity, clinical characteristics, or treatment status was removed from the images available to the examiners. Two orthodontists independently performed cephalometric analysis, blinded to each other's results. Each examiner assessed all radiographs twice, with a two-week interval between assessments. For the second assessment, the radiographs were presented in a newly randomized order, and the examiners had no access to their previous measurements. The manual reference used for comparison with the AI-generated measurements was defined as a mean value from four measurements performed by the expert orthodontists.

2.2. Cephalometric Analysis

Cephalometric landmarks used have been presented in Figure 1.

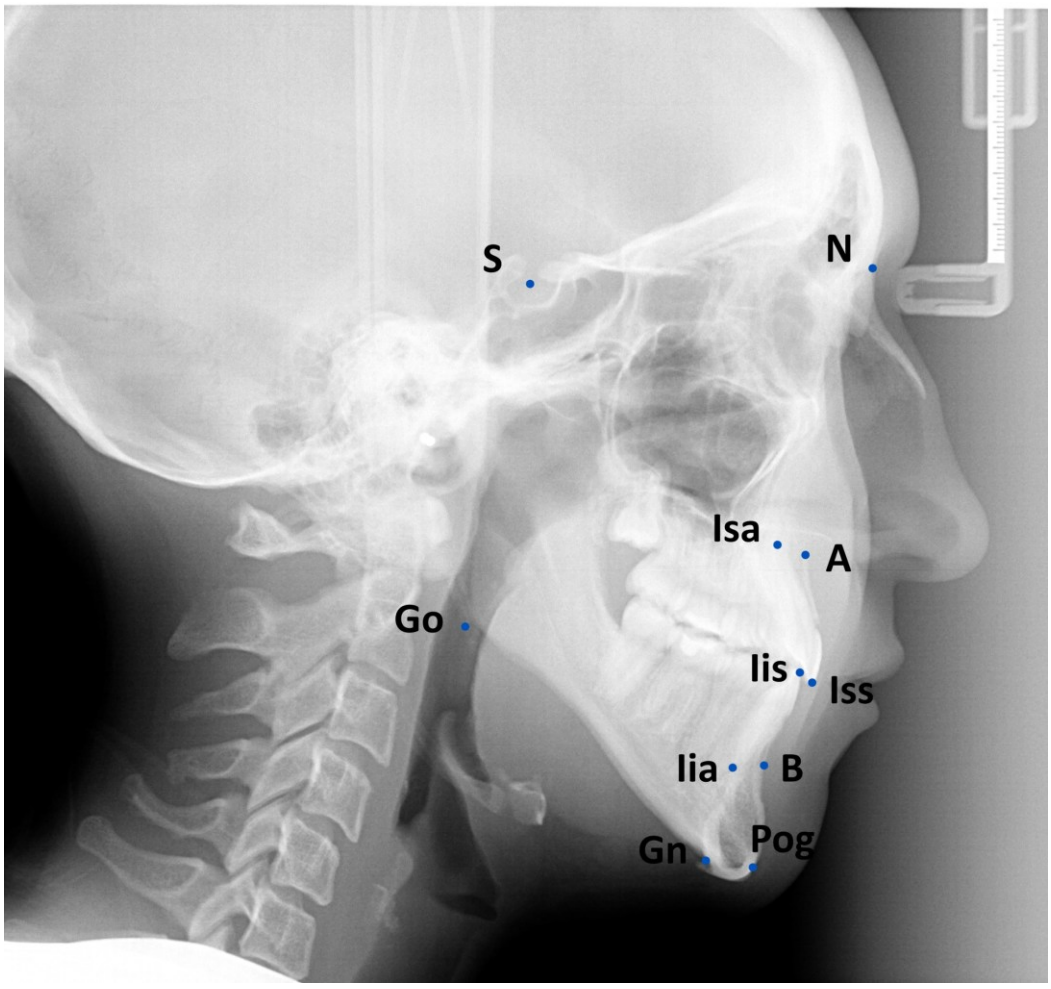


Figure 1. Cephalometric landmarks used in this study for patients with cleft palate only (CPO). S—sella, N—nasion, Isa—incisivus superioris apex, Iis—incisivus inferioris superior, Iia—incisivus

inferioris apex, Iss—incisivus superioris superior, Gn—gnathion, Go—gonion, A—point A (maxillary anterior point), B—point B (mandibular anterior point), Pog—pogonion.

Cephalometric landmarks used have been listed and defined in Table 1.

Table 1. Explanation of cephalometric landmarks used in present study.

Cephalometric Landmark	Definition
S (sella)	The center of sella turcica
N (nasion)	The most anterior point of frontonasal suture
A (Point A)	The most distal point of the anterior surface of the maxilla
B (Point B)	The most distal point of the anterior surface of the mandible
Iss (incisivus superioris superior)	Incisal edge of upper central incisor (the more protrusive one)
Isa (incisivus superioris apex)	Apex of upper central incisor
Iis (incisivus inferioris superior)	Incisal edge of lower central incisor (the more protrusive one)
Iia (incisivus inferioris apex)	Apex of lower central incisor
Go gonion	Constructed apex of gonial angle
Gn gnathion	The lowest point of the frontal part of mandibular body
Pog (pogonion)	The most protruding point on the chin

Cephalometric variables have been listed and defined in Table 2.

Table 2. Cephalometric variables used in present study.

Cephalometric Variable	Landmarks Used	Definition and Interpretation
SNA [º]	S, N, A	Angle between the sella–nasion line and point A; indicates maxillary position relative to the cranial base
SNB [º]	S, N, B	Angle between the sella–nasion line and point B; indicates mandibular position relative to the cranial base
ANB [º]	A, N, B	Angle between nasion–point A and nasionpoint B; indicates the anteroposterior relationship between the maxilla and mandible
U1/NA [º]	Iss, Isa, N, A	Angle between the upper incisor axis and the NA line; indicates inclination of upper incisors
U1/NA [mm]	Iss, Isa, N, A	Linear distance from the upper incisor to the NA line; indicates position of upper incisors
L1/NB [º]	Iia, Iis, N, B	Angle between the lower incisor axis and the NB line; indicates inclination of lower incisors
L1/NB [mm]	Iia, Iis, N, B	Linear distance from the lower incisor to the NB line; indicates position of lower incisors
U1-L1 [º]	Iss, Isa, Iia, Iis	Interincisal angle; indicates the angulation between upper and lower incisors
Go-Gn/SN [º]	Go, Gn, S, N	Mandibular plane angle; angle between mandibular plane and the cranial base; indicates vertical growth pattern
Pog-NB [mm]	Pog, N, B	Linear distance from pogonion to the NB line; indicates chin prominence

Cephalometric analysis was performed independently by two specialists in orthodontics experienced in the treatment of patients with orofacial clefts. Cephalometric analysis based on manually identified landmarks was performed in specialized software (Ortodoncja 9, Ortobajt, Wrocław, Poland). Subsequently, all cephalometric radiographs were analyzed using the online AI service CephX (Cephalometric Analysis Software, version 4.02; Orca Dental AI Ltd., Herzliya, Israel; date of manufacture: 21 April 2024). Two weeks later, each orthodontist repeated the complete analysis independently, and the AI

analysis was repeated by uploading the radiographs as new cases. In the CPO group, manual scale calibration was required in 34 of 50 cases. This calibration concerned image scaling and did not involve manual modification of AI-identified skeletal or dental landmarks. In 31 of 50 CPO cases, the software generated a warning recommending review of landmark identification; these cases were visually reviewed, but none of the skeletal or dental landmarks included in the analysis were manually repositioned by the investigators. Therefore, landmark identification and measurement generation remained automated, while the overall workflow required clinician supervision and, in some cases, manual scale calibration. By contrast, none of the control cases required manual scale calibration or generated a landmark-review warning (0/50, 0%). The outline of the facial profile and the points marked by AI are presented in Figure 2.

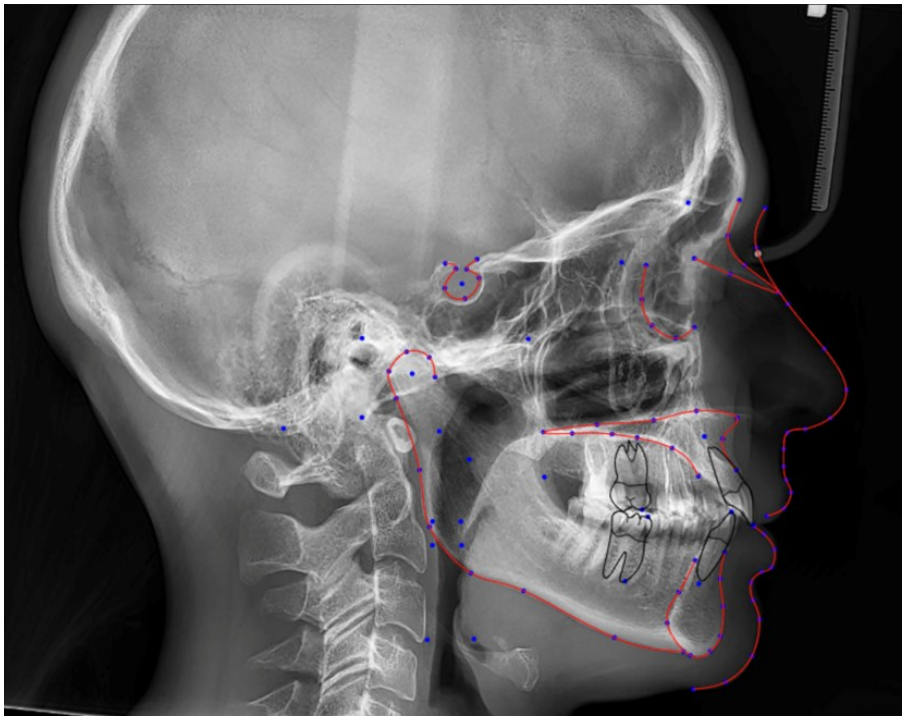


Figure 2. AI-assisted cephalometric analysis of a patient from the control group. Blue dots indicate AI-identified cephalometric landmarks, and the red line indicates the automatically generated facial profile tracing.

2.3. Statistical Analysis

Statistical analysis was performed using R statistical software (version 4.3.3; R Foundation for Statistical Computing, Vienna, Austria) [19] on Windows 11 Pro 64-bit (build 22631; Microsoft Corporation, Redmond, WA, USA), using the packages blandr (version 0.6.0), report (version 0.5.8), patchwork (version 1.2.0), gtsummary (version 1.7.2), epiR (version 2.0.75), ggplot2 (version 3.5.0), and dplyr (version 1.1.4) [20,21]. A two-sided significance level of $\alpha = 0.05$ was used. Numerical and categorical sample characteristics were summarized descriptively. Because sex was deliberately balanced by matching, no inferential test of sex distribution between groups was required. Age was compared between groups using the Wilcoxon rank-sum test.

Repeatability was assessed using test–retest analyses. Manual intra-rater repeatability was evaluated by comparing the first and second measurements within each orthodontist, and AI test–retest repeatability was evaluated by comparing the first and second CephX analyses. Lin's concordance correlation coefficient (CCC) was used as the primary measure of repeatability [22]. Inter-examiner agreement between the two orthodontists was assessed using CCC.

Agreement between AI-generated measurements and the manual reference was evaluated separately using Bland–Altman [23]. The manual reference used for this comparison was a mean of four measurements by the expert orthodontists.

For the Bland–Altman analysis, the manual reference value for each radiograph and variable was calculated as the mean of four measurements obtained by the two expert orthodontists (two assessments per examiner). Pairwise differences between the AI-generated measurements and this manual reference were used to estimate the mean bias and the lower and upper 95% limits of agreement (LoA), calculated as the bias ± 1.96 times the standard deviation of the paired differences. Confidence intervals were calculated for the bias and both limits of agreement to assess the precision of these estimates [24]. A paired t-test was used to assess whether the mean difference differed from zero.

3. Results

Figure 3 illustrates a lateral cephalogram of a patient with CPO, highlighting characteristic craniofacial features of this condition, including a skeletal Class III tendency, retruded maxilla, and a reduced nasolabial angle.

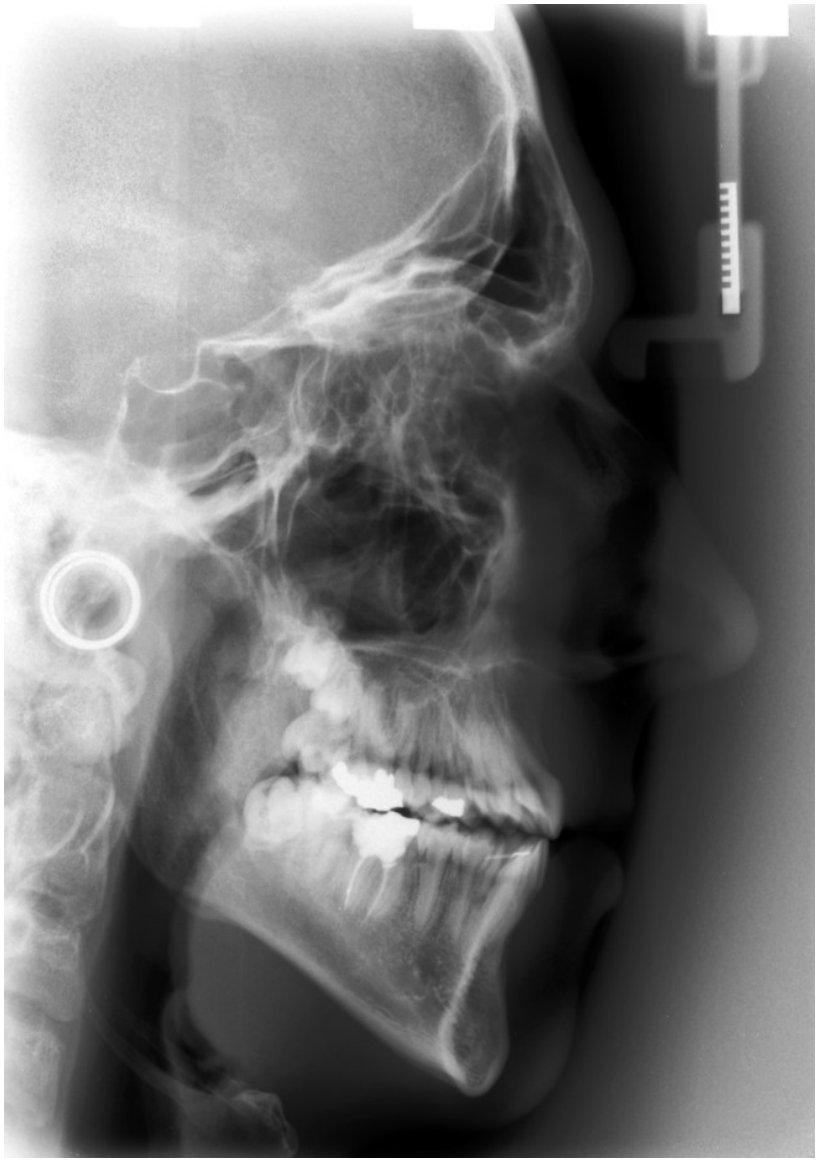


Figure 3. Lateral cephalogram of a patient with cleft palate only (CPO), showing characteristic craniofacial features.

The distribution of the study and control groups according to age and sex has been presented in Table 3.

Table 3. Demographic characteristics of the overall sample and group-specific comparisons.

Characteristic	*N*	Overall Sample [a]	Group		p
			CPO, $n_1 = 50$ [a]	Control, $n_2 = 50$ [a]	
Sex:	100				—
Female		50 (50.00%)	25 (50.00%)	25 (50.00%)	
Male		50 (50.00%)	25 (50.00%)	25 (50.00%)	
Age, years[b]	100	17.00 (15.72, 18.00)	15.81 (15.26, 16.96)	17.00 (17.00, 18.00)	<0.001[d]
Age group:	100				<0.001[c]
Pediatric		73 (73.00%)	46 (92.00%)	27 (54.00%)	
Adult		27 (27.00%)	4 (8.00%)	23 (46.00%)	

[a] *n (%), unless otherwise indicated;* [b] *median (interquartile range);* [c] *Pearson's chi-squared test (age group);* [d] *Wilcoxon rank-sum test (age). The dash indicates that no statistical test was performed for sex distribution because sex was matched by design.*

Sex distribution was deliberately balanced by matching, with 25 males and 25 females in each group. Age was significantly lower in the CPO group (Mdn = 15.81 years, IQR: 15.26–16.96) than in the control group (Mdn = 17.00 years, IQR: 17.00–18.00; $p < 0.001$).

In the CPO group, 34 of 50 radiographs generated the message: "Scale calibration is estimated. Consider recalibration for maximum accuracy. Click 'manual calibration' to set the scale or continue anyway." In each of these cases, manual scale calibration was performed. This step concerned image scaling only; the positions of the AI-identified skeletal and dental landmarks were not manually modified.

In two CPO cases, the automatically generated soft-tissue outline appeared incorrect in the lower-lip and chin region. These cases were submitted to the manufacturer for manual review. However, soft-tissue landmarks were not included among the cephalometric variables evaluated in the present study; therefore, these soft-tissue inaccuracies did not affect the measurements included in the analysis.

In 31 of 50 CPO cases, the software generated the message: "Landmarks require overview. AI-generated landmarks for this case have not passed the requirements. When the analysis is completed, review landmark positioning before using the results." These cases were visually reviewed, but none of the skeletal or dental landmarks included in the analysis were manually repositioned by the investigators. By contrast, no landmark-review warnings were generated in the control group (0/50, 0%).

Detailed results are presented in Table 4, highlighting variations in assessment outcomes between manual and AI-driven cephalometric analyses.

Table 4. Comparison of manual and AI-driven cephalometric variables across repeated measurements in CPO and control groups.

Cephalometric Variables	Group	Manual Assessment		AI Assessment	
		First Measurement ($n_1 = 50$) [a]	Second Measurement ($n_2 = 50$) [a]	First Measurement ($n_1 = 50$) [a]	Second Measurement ($n_2 = 50$) [a]
SNA angle	CPO	79.30 (75.60, 81.75)	79.35 (74.78, 81.43)	78.79 (76.38, 80.84)	78.70 (76.25, 80.84)
	Control	82.25 (79.68, 84.30)	82.30 (79.85, 84.65)	83.26 (80.72, 85.56)	83.18 (80.72, 85.56)
SNB angle	CPO	77.85 (75.83, 82.40)	78.10 (75.63, 82.78)	76.92 (75.22, 80.76)	77.09 (75.22, 80.76)
	Control	79.75 (77.33, 81.93)	80.60 (77.98, 82.58)	79.61 (77.46, 81.53)	79.61 (77.46, 81.53)
ANB angle	CPO	−0.10 (−3.00, 2.25)	−0.30 (−3.10, 2.28)	1.28 (−1.41, 2.48)	1.28 (−1.38, 2.48)
	Control	2.85 (1.00, 4.68)	2.35 (0.45, 4.45)	4.10 (2.47, 5.56)	4.10 (2.47, 5.56)

U1/NA angle	CPO	26.55 (22.30, 33.28)	25.50 (20.93, 33.23)	26.00 (20.39, 32.44)	25.09 (20.88, 33.71)
	Control	23.75 (18.23, 29.05)	23.50 (16.50, 28.95)	21.96 (18.54, 25.88)	21.96 (18.54, 25.88)
U1/NA, mm	CPO	3.50 (2.20, 5.58)	3.40 (1.95, 5.68)	4.69 (2.99, 6.61)	4.59 (2.99, 6.28)
	Control	3.25 (2.13, 4.95)	5.10 (3.45, 6.75)	3.50 (1.65, 5.34)	3.50 (1.65, 5.34)
L1/NB angle	CPO	22.40 (17.43, 26.40)	22.20 (17.20, 25.83)	20.41 (15.49, 22.32)	20.41 (15.49, 22.50)
	Control	27.25 (22.03, 31.58)	25.40 (20.13, 30.63)	23.42 (17.84, 27.34)	23.42 (17.80, 27.34)
L1/NB, mm	CPO	2.40 (1.23, 3.20)	2.45 (1.00, 3.38)	3.40 (1.33, 4.20)	3.11 (1.57, 4.20)
	Control	3.65 (1.73, 4.88)	4.10 (2.43, 5.80)	3.88 (2.11, 5.34)	3.88 (2.11, 5.34)
Interincisal angle	CPO	132.20 (121.80, 140.73)	132.40 (123.80, 145.35)	132.39 (126.20, 143.20)	132.39 (125.48, 142.33)
	Control	127.45 (121.93, 136.43)	130.75 (120.33, 139.65)	130.71 (122.89, 139.11)	130.42 (122.89, 139.11)
GoGn/SN angle	CPO	35.95 (30.53, 39.88)	35.45 (30.70, 39.85)	40.26 (36.35, 43.76)	40.26 (36.54, 43.86)
	Control	29.00 (25.95, 37.68)	30.65 (25.38, 36.40)	33.92 (30.55, 41.68)	33.92 (30.55, 41.68)
Pog/NB mm	CPO	1.50 (0.38, 2.65)	1.50 (0.70, 2.80)	2.66 (1.04, 3.99)	2.31 (1.05, 3.74)
	Control	1.15 (0.30, 2.38)	1.25 (0.40, 2.40)	1.41 (0.60, 2.78)	1.41 (0.60, 2.78)

[a] Values are presented as median (interquartile range).

Measurements differed to varying degrees between AI and manual assessment. Agreement between the AI-generated measurements and the manual reference is presented in Table 5 and should be interpreted using the observed bias and limits of agreement rather than repeatability alone.

Table 5. Results of Bland–Altman analysis for assessing agreement between orthodontist and AI-based program in measuring clinical variables across the groups.

Variable	Group	Bias	95% CI (Bias)	ULoA	95% CI (ULoA)	LLoA	95% CI (LLoA)	Spread of LoA	*p*
SNA angle	CPO	−0.17	−0.73–0.38	3.67	2.70–4.62	−4.01	−4.97–−3.05	7.68	0.535
	Control	−0.92	−1.32–−0.52	1.86	1.17–2.56	−3.71	−4.40–−3.01	5.57	**<0.001**
SNB angle	CPO	1.05	0.60–1.51	4.20	3.41–4.98	−2.09	−2.87–−1.30	6.28	**<0.001**
	Control	0.47	0.25–0.68	1.94	1.58–2.31	−1.01	−1.38–−0.64	2.96	**<0.001**
ANB angle	CPO	−1.22	−1.87–−0.57	3.24	2.12–4.35	−5.69	−6.80–−4.58	8.93	**<0.001**
	Control	−1.40	−1.73–−1.08	0.83	0.27–1.39	−3.64	−4.19–−3.08	4.47	**<0.001**
U1/NA angle	CPO	0.86	−0.68–2.40	11.46	8.81–14.11	−9.74	−12.38–−7.10	21.21	0.266
	Control	0.80	−0.18–1.78	7.55	5.86–9.24	−5.95	−7.64–−4.26	13.51	0.107
U1/NA, mm	CPO	−0.69	−1.68–0.30	6.15	4.44–7.85	−7.53	−9.24–−5.82	13.68	0.167
	Control	0.63	0.22–1.04	3.45	2.74–4.15	−2.19	−2.90–−1.49	5.64	**0.003**
L1/NB angle	CPO	1.96	1.12–2.81	7.76	6.32–9.22	−3.84	−5.29–−2.39	11.61	**<0.001**
	Control	3.02	2.28–3.75	8.10	6.83–9.36	−2.06	−3.32–−0.79	10.16	**<0.001**
L1/NB, mm	CPO	−0.53	−0.89–−0.17	1.97	1.34–2.60	−3.03	−3.66–−2.41	5.01	**0.005**
	Control	0.07	−0.29–0.43	2.55	1.93–3.16	−2.40	−3.01–−1.78	4.95	0.687
Interincisal angle	CPO	−1.31	−2.71–0.11	8.42	5.99–10.84	−11.02	−13.44–−8.60	19.44	0.069
	Control	−2.24	−3.33–−1.16	5.26	3.38–7.13	−9.75	−11.62–−7.88	15.01	**<0.001**
GOGN/SN angle	CPO	−4.75	−5.41–−4.09	−0.16	−1.30–0.99	−9.34	−10.49–−8.20	9.19	**<0.001**
	Control	−4.52	−4.91–−4.13	−1.82	−2.49–1.15	−7.21	−7.88–−6.45	5.40	**<0.001**
POG/NB mm	CPO	−1.01	−1.31–−0.71	1.06	0.55–1.58	−3.08	−3.61–−2.56	4.14	**<0.001**
	Control	−0.28	−0.49–−0.07	1.16	0.80–1.53	−1.72	−2.08–−1.36	2.88	**0.009**

Bold values indicate statistically significant results ($p < 0.05$).

The repeatability and agreement of the manual cephalometric measurements are presented in Table 6.

Table 6. Repeatability and agreement of cephalometric measurements performed by the orthodontists.

Variable	Group	CCC	CI 95%	Scale Shift	Location Shift	Bias Correction
SNA angle	CPO	0.92	0.86–0.95	1.08	−0.07	1.00
	Control	0.89	0.82–0.93	1.02	0.04	1.00
SNB angle	CPO	0.97	0.95–0.98	0.98	0.00	1.00
	Control	0.94	0.91–0.97	1.07	0.15	0.99
ANB angle	CPO	0.97	0.95–0.98	1.07	−0.07	1.00
	Control	0.91	0.85–0.95	0.95	−0.12	0.99
U1/NA angle	CPO	0.94	0.90–0.96	1.14	−0.15	0.98
	Control	0.96	0.94–0.98	1.07	−0.05	1.00
U1/NA, mm	CPO	0.94	0.90–0.96	1.16	0.07	0.99
	Control	0.71	0.56–0.82	1.07	0.38	0.93
L1/NB angle	CPO	0.88	0.79–0.93	0.99	−0.11	0.99
	Control	0.91	0.85–0.94	1.08	−0.08	0.99
L1/NB, mm	CPO	0.98	0.97–0.99	1.05	0.03	1.00
	Control	0.84	0.74–0.90	1.11	0.25	0.97
Interincisal angle	CPO	0.92	0.88–0.96	1.06	0.19	0.98
	Control	0.91	0.86–0.95	1.08	0.09	0.99
GOGN/SN angle	CPO	0.97	0.95–0.98	0.98	−0.01	1.00
	Control	0.95	0.92–0.97	0.94	−0.04	1.00
POG/NB mm	CPO	0.96	0.93–0.98	0.95	0.15	0.99
	Control	0.96	0.94–0.98	0.93	0.04	1.00

It is visible that the agreement differs for individual variables. The SNA angle has a high concordance in the CPO group (CCC = 0.92) but a slightly lower concordance in the control group (CCC = 0.89). U1/NA (mm) and L1/NB (mm) in the control group had lower CCC values (0.71 and 0.84, respectively), which indicates a higher variability in repeated measurements.

AI test–retest repeatability is presented in Table 7.

Table 7. Test–retest repeatability of AI-driven cephalometric measurements.

Variable	Group	CCC	CI 95%	Scale Shift	Location Shift	Bias Correction
SNA angle	CPO	0.98	0.96–0.99	1.00	−0.01	1.00
	Control	1.00	1.00–1.00	1.00	−0.01	1.00
SNB angle	CPO	0.99	0.99–1.00	0.99	−0.01	1.00
	Control	1.00	1.00–1.00	0.98	−0.01	1.00
ANB angle	CPO	1.00	1.00–1.00	1.01	0.00	1.00
	Control	1.00	1.00–1.00	0.99	0.00	1.00
U1/NA angle	CPO	0.96	0.94–0.98	1.04	0.01	1.00
	Control	0.99	0.99–1.00	1.03	0.02	1.00
U1/NA, mm	CPO	0.97	0.94–0.98	0.99	−0.05	1.00
	Control	0.98	0.97–0.99	1.04	0.03	1.00
L1/NB angle	CPO	0.99	0.98–0.99	1.04	−0.03	1.00
	Control	1.00	1.00–1.00	1.00	0.00	1.00
L1/NB, mm	CPO	1.00	1.00–1.00	1.00	0.00	1.00
	Control	1.00	1.00–1.00	1.00	0.00	1.00
Interincisal angle	CPO	1.00	1.00–1.00	1.00	−0.01	1.00
	Control	1.00	1.00–1.00	1.00	−0.01	1.00
GOGN/SN angle	CPO	0.97	0.94–0.98	0.93	0.05	1.00

	Control	1.00	1.00–1.00	1.00	0.01	1.00
POG/NB mm	CPO	0.98	0.97–0.99	1.02	−0.02	1.00
	Control	1.00	1.00–1.00	1.00	0.01	1.00

The AI software showed high test–retest repeatability across all evaluated variables. This finding reflects the reproducibility of repeated AI measurements and should not be interpreted as evidence of accuracy or agreement with the manual reference.

CephX achieved CCC values of 0.97 or higher for every evaluated variable in both the CPO and control groups. For several measurements, including ANB angle, L1/NB (mm), and the interincisal angle, the CCC was 1.00. These results indicate the high test–retest reproducibility of the AI-generated measurements.

4. Discussion

AI-assisted cephalometric analysis has the potential to reduce the variability associated with repeated landmark identification, but its clinical usefulness depends not only on reproducibility but also on agreement with expert measurements and the need for clinician supervision [5,8,14].

In patients with CPO, altered craniofacial morphology may make landmark identification particularly challenging. Manual landmarking is also subject to inter- and intra-observer variability [3], which makes separate evaluation of repeatability and agreement important.

The present study evaluated CephX in patients with CPO and non-cleft controls. Previous studies have assessed AI-assisted cephalometry in non-cleft populations, while Tageldin et al. [18] evaluated AI landmark identification in patients with unilateral cleft palate. The most relevant previous studies are summarized in the shortened Table 8 and discussed in relation to the present findings.

Table 8 summarizes the selected studies most directly relevant to the agreement, repeatability, and reliability of AI-assisted cephalometric analysis.

Table 8. Selected studies evaluating AI-assisted cephalometric analysis in relation to manual or alternative cephalometric methods.

Study	AI System and Comparison	Main Finding Relevant to the Present Study
Koz et al., 2025 [25]	WebCeph and OneCeph vs. manual analysis	High overall compatibility with manual analysis, although discrepancies occurred for selected angular and linear measurements.
Tageldin et al., 2025 [18]	WebCeph vs. expert manual landmark identification in patients with unilateral cleft palate	Significant differences were found for selected landmarks, including A-point, ANS, and Or; manual verification of AI-identified landmarks was recommended.
Bor et al., 2024 [26]	CephX, WeDoCeph and WebCeph vs. orthodontist-performed digital tracing	Differences occurred in angular and linear measurements; orthodontist supervision was recommended, particularly in atypical skeletal relationships.
Kazimierczak et al., 2024 [8]	CephX, WebCeph and AudaxCeph; repeated AI analyses	Differences were observed among AI platforms for several cephalometric variables; repeated analyses were also used to evaluate reproducibility.
Guinot-Barona et al., 2024 [9]	CephX vs. measurements performed by experienced orthodontists	Differences between AI and orthodontist measurements were identified in 7 of 24 cephalometric parameters.
Meriç et al., 2020 [14]	CephX, CephNinja and Dolphin vs. manual tracing	CephX showed variability for several measurements; manual correction of CephX landmarks improved agreement with the other methods.

Previous studies have shown that AI-assisted cephalometric systems may provide reproducible measurements but can differ from expert or manual analyses on selected variables [8,9,14,25,26]. Tageldin et al. [18], in patients with unilateral cleft palate, reported significant differences in selected landmarks and recommended manual verification of AI-generated landmark positions. These findings are consistent with the present observation

that high AI test–retest repeatability does not by itself establish agreement with expert manual measurements and that clinician supervision remains important.

As in previous reports, the control group displayed more balanced craniofacial relationships, with measurements generally falling within or close to normal ranges. Previous findings suggest that CPO patients generally exhibit patterns associated with craniofacial anomalies, such as retrusive maxilla and mandible, incisor protrusion, and steep mandibular planes [3].

In the present study, the magnitude of the AI–manual differences varied among cephalometric variables. For clinical interpretation, differences of 2 mm or 2° have been considered clinically significant in conventional cephalometric assessment [27]. Nevertheless, agreement should be interpreted using the observed bias and limits of agreement for each variable rather than from test–retest CCC values alone.

The high test–retest CCC values obtained for CephX indicate high computational reproducibility rather than accuracy. An algorithm may reproduce a measurement consistently while maintaining a systematic difference from expert-defined measurements. The frequent need for manual scale calibration and landmark review in the CPO group further supports the need for clinician supervision, particularly in patients with altered craniofacial anatomy.

Limitations of this Study

This study has several limitations. First, the single-center sample limits generalizability to broader populations. The CPO and control groups were matched for sex and ethnic origin but were not age-matched, and this age difference should be considered when interpreting between-group comparisons. Only skeletal and dental landmarks were analyzed. This study evaluated a single AI platform (CephX), so the findings should not be generalized to other systems. Manual landmark identification is subject to observer variability; intra-rater and inter-examiner results should therefore be considered alongside AI test–retest repeatability. Radiographs of insufficient diagnostic quality preventing reliable identification of cephalometric landmarks were excluded. Finally, the cross-sectional design limits conclusions beyond the measurement performance evaluated here.

5. Conclusions

1. CephX demonstrated high test–retest repeatability in cephalometric measurements in both patients with CPO and non-cleft controls; however, repeatability was not equivalent to agreement with expert manual measurements, which varied among cephalometric variables.
2. In the CPO group, the frequent need for manual scale calibration and landmark review indicates that clinician supervision remains important when CephX is used in patients with altered craniofacial anatomy.

The findings support the use of CephX as a clinician-supervised tool for reproducible cephalometric measurement in the population studied, rather than as an autonomous replacement for expert assessment.

Author Contributions: Conceptualization, A.Z. and J.J.-O.; methodology, A.Z. and J.J.-O.; software, A.Z. and J.J.-O.; validation, A.Z. and J.J.-O.; formal analysis, A.Z., J.J.-O., N.K. and W.K.; investigation, A.Z., J.J.-O., N.K. and W.K.; resources, A.Z., J.J.-O., N.K. and W.K.; data curation, A.Z.; writing—original draft preparation, A.Z. and J.J.-O.; writing—review and editing, A.Z., J.J.-O. and K.G.; visualization, A.Z.; supervision, J.J.-O.; project administration, A.Z. and J.J.-O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki. According to the decision KB-006/04/2022/Z approved on 21 January 2022 of the Bioethical Committee Pomeranian Medical University in Szczecin, this study is exempt from ethical approval.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used and analyzed during the current study are available from the corresponding authors on reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

CPO	cleft palate only
AI	artificial intelligence
ANB	angle between points A, Nasion, and B
A (Point A)	the most distal point on the anterior surface of the maxilla; represents maxillary position in cephalometric analysis.
B (Point B)	the most distal point on the anterior surface of the mandible; represents mandibular position in cephalometric analysis.
BCLP	bilateral cleft lip and palate
CCC	Lin's concordance correlation coefficient
CI	confidence interval
Go	gonion
Gn	gnathion
LoA	limits of agreement
N	nasion
Pog	pogonion
R	statistical programming language used for data analysis and visualization
S	sella
SNA	sella–nasion–point A angle
SNB	sella–nasion–point B angle
UCLP	Unilateral Cleft Lip and Palate
U1/NA [°]	angle between the upper incisor axis and the NA line
U1/NA [mm]	linear distance between the upper incisor and the NA line
L1/NB [°]	angle between the lower incisor axis and the NB line
L1/NB [mm]	linear distance between the lower incisor and the NB line
U1–L1 [°]	interincisal angle

References

1. Salari, N.; Darvishi, N.; Heydari, M.; Bokaee, S.; Darvishi, F.; Mohammadi, M. Global prevalence of cleft palate, cleft lip and cleft palate and lip: A comprehensive systematic review and meta-analysis. *J. Stomatol. Oral Maxillofac. Surg.* **2022**, *123*, 110–120. https://doi.org/10.1016/j.jormas.2021.05.008.
2. Azouz, V.; Ng, M.; Patel, N.; Murthy, A.S. Low incidence of maxillary hypoplasia in isolated cleft palate. *Maxillofac. Plast. Reconstr. Surg.* **2020**, *42*, 8. https://doi.org/10.1186/s40902-020-00252-9.
3. Zawiślak, A.; Wędrychowska-Szulc, B.; Grocholewicz, K.; Janiszewska-Olszowska, J. Craniofacial cephalometric morphology in Caucasian adult patients with cleft palate only (CPO). *Diagnostics* **2023**, *13*, 2058. https://doi.org/10.3390/diagnostics13122058.
4. Zawiślak, A.; Wędrychowska-Szulc, B.; Grocholewicz, K.; Janiszewska-Olszowska, J. Craniofacial cephalometric morphology in Polish adolescents with cleft palate only. *J. Clin. Med.* **2024**, *13*, 4507. https://doi.org/10.3390/jcm13154507.
5. Bao, H.; Zhang, K.; Yu, C.; Li, H.; Cao, D.; Shu, H.; Liu, L.; Yan, B. Evaluating the accuracy of automated cephalometric analysis based on artificial intelligence. *BMC Oral Health* **2023**, *23*, 191. https://doi.org/10.1186/s12903-023-02881-8.
6. Gupta, S.; Shetty, S.; Natarajan, S.; Nambiar, S.; Mv, A.; Agarwal, S. A comparative evaluation of concordance and speed between smartphone app-based and artificial intelligence web-based cephalometric tracing software with the manual tracing method: A cross-sectional study. *J. Clin. Exp. Dent.* **2024**, *16*, e11–e17. https://doi.org/10.4317/jced.60899.

7. Lee, J.; Bae, S.R.; Noh, H.K. Commercial artificial intelligence lateral cephalometric analysis: Part 1—The possibility of replacing manual landmarking with artificial intelligence service. *J. Clin. Pediatr. Dent.* **2023**, *47*, 106–118. https://doi.org/10.22514/jocpd.2023.085.
8. Kazimierczak, W.; Gawin, G.; Janiszewska-Olszowska, J.; Dyszkiewicz-Konwińska, M.; Nowicki, P.; Kazimierczak, N.; Serafin, Z.; Orhan, K. Comparison of three commercially available, AI-driven cephalometric analysis tools in orthodontics. *J. Clin. Med.* **2024**, *13*, 3733. https://doi.org/10.3390/jcm13133733.
9. Guinot-Barona, C.; Alonso Pérez-Barquero, J.; Galán López, L.; Barmak, A.B.; Att, W.; Kois, J.C.; Revilla-León, M. Cephalometric analysis performance discrepancy between orthodontists and an artificial intelligence model using lateral cephalometric radiographs. *J. Esthet. Restor. Dent.* **2024**, *36*, 555–565. https://doi.org/10.1111/jerd.13156.
10. Kang, S.; Kim, I.; Kim, Y.J.; Kim, N.; Baek, S.H.; Sung, S.J. Accuracy and clinical validity of automated cephalometric analysis using convolutional neural networks. *Orthod. Craniofacial Res.* **2024**, *27*, 64–77. https://doi.org/10.1111/ocr.12683.
11. El-Dawlatly, M.; Attia, K.H.; Abdelghaffar, A.Y.; Mostafa, Y.A.; Abd El-Ghafour, M. Preciseness of artificial intelligence for lateral cephalometric measurements. *J. Orofac. Orthop.* **2024**, *85*, 27–33. https://doi.org/10.1007/s00056-023-00459-1.
12. Duran, G.S.; Gökmen, Ş.; Topsakal, K.G.; Görgülü, S. Evaluation of the accuracy of fully automatic cephalometric analysis software with artificial intelligence algorithm. *Orthod. Craniofacial Res.* **2023**, *26*, 481–490. https://doi.org/10.1111/ocr.12633.
13. Junaid, N.; Khan, N.; Ahmed, N.; Abbasi, M.S.; Das, G.; Maqsood, A.; Ahmed, A.R.; Marya, A.; Alam, M.K.; Heboyan, A. Development, application, and performance of artificial intelligence in cephalometric landmark identification and diagnosis: A systematic review. *Healthcare* **2022**, *10*, 2454. https://doi.org/10.3390/healthcare10122454.
14. Meriç, P.; Naoumova, J. Web-based fully automated cephalometric analysis: Comparisons between app-aided, computerized, and manual tracings. *Turk. J. Orthod.* **2020**, *33*, 142–149. https://doi.org/10.5152/TurkJOrthod.2020.20062.
15. Mosleh, M.A.; Baba, M.S.; Malek, S.; Almaktari, R.A. Ceph-X: Development and evaluation of 2D cephalometric system. *BMC Bioinform.* **2016**, *17*, 499. https://doi.org/10.1186/s12859-016-1370-5.
16. Hendrickx, J.; Gracea, R.S.; Vanheers, M.; Winderickx, N.; Preda, F.; Shujaat, S.; Jacobs, R. Can artificial intelligence-driven cephalometric analysis replace manual tracing? A systematic review and meta-analysis. *Eur. J. Orthod.* **2024**, *46*, cjae029. https://doi.org/10.1093/ejo/cjae029.
17. Jayakumar, S.; Sounderajah, V.; Normahani, P.; Harling, L.; Markar, S.R.; Ashrafian, H.; Darzi, A. Quality assessment standards in artificial intelligence diagnostic accuracy systematic reviews: A meta-research study. *NPJ Digit. Med.* **2022**, *5*, 11. https://doi.org/10.1038/s41746-021-00544-y.
18. Tageldin, M.A.; Yacout, Y.M.; Eid, F.Y.; Abdelhafiz, S.H. Accuracy of cephalometric landmark identification by artificial intelligence platform versus expert orthodontist in unilateral cleft palate patients: A retrospective study. *Int. Orthod.* **2025**, *23*, 100990. https://doi.org/10.1016/j.ortho.2025.100990.
19. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2024. Available online: https://www.R-project.org/ (accessed on 2 November 2026).
20. Datta, D. *blandr: A Bland-Altman Method Comparison Package for R*, version 0.6.0; Zenodo: Geneva, Switzerland, 2017. https://doi.org/10.5281/zenodo.824514.
21. Makowski, D.; Lüdecke, D.; Patil, I.; Thériault, R.; Ben-Shachar, M.; Wiernik, B. Automated Results Reporting as a Practical Tool to Improve Reproducibility and Methodological Best Practices Adoption. CRAN. 2023. Available online: https://easystats.github.io/report/ (accessed on 2 November 2026).
22. Lin, L. A note on the concordance correlation coefficient. *Biometrics* **2000**, *56*, 324–325.
23. Bland, J.M.; Altman, D.G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1986**, *327*, 307–310.
24. Altman, D.G.; Bland, J.M. Commentary on quantifying agreement between two methods of measurement. *Clin. Chem.* **2002**, *48*, 801–802.
25. Koz, S.; Uslu-Akcam, O. Artificial Intelligence-Supported and App-Aided Cephalometric Analysis: Which One Can We Trust? *Diagnostics* **2025**, *15*, 559. https://doi.org/10.3390/diagnostics15050559.
26. Bor, S.; Ciğerim, S.Ç.; Kotan, S. Comparison of AI-Assisted Cephalometric Analysis and Orthodontist-Performed Digital Tracing Analysis. *Prog. Orthod.* **2024**, *25*, 41. https://doi.org/10.1186/s40510-024-00539-x.
27. Proffit, W.R.; Fields, H.W.; Sarver, D.M. *Contemporary Orthodontics*, 4th ed.; Mosby Elsevier: St. Louis, MO, USA, 2006.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.