Imaging Science in Dentistry 2025; 55: 11-21
https://doi.org/10.5624/isd.20240089

# Comparative accuracy of artificial intelligence-based AudaxCeph software, Dolphin software, and the manual technique for orthodontic landmark identification and tracing of lateral cephalograms

Maryam Foroozandeh[1], Fatemeh Salemi[1,*], Abbas Shokri[1], Nasrin Farhadian[2], Nesa Aeini[3,4], Roghayyeh Hassanzadeh[4,5]

[1]*Department of Oral and Maxillofacial Radiology, Dental School, Dental Research Center, Hamadan University of Medical Sciences, Hamadan, Iran*
[2]*Department of Orthodontics, School of Dentistry, Dental Research Center, Hamadan University of Medical Sciences, Hamadan, Iran*
[3]*Farhangian Dental Clinic, Hamadan, Iran*
[4]*Student Research Committee, Hamadan University of Medical Sciences, Hamadan, Iran*
[5]*Department of Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran*

## ABSTRACT

**Purpose**: The aim of this study was to compare the accuracy of AI-based AudaxCeph software, Dolphin software, and the manual technique for identifying orthodontic landmarks and tracing lateral cephalograms.
**Materials and Methods**: In this cross-sectional study, 23 anatomical landmarks were identified on 60 randomly selected lateral cephalograms, and 5 dental indices, 4 skeletal indices, and 1 soft tissue index were measured. Each cephalogram was traced using 4 different methods: manually, with the Dolphin software, with the AudaxCeph software automatically, and with the AudaxCeph software in semi-automatic mode. The intra-class correlation coefficient (ICC) and Bland-Altman plots were used to evaluate the agreement between methods. Inter-observer and intra-observer agreements, calculated using the ICC, confirmed the accuracy, reliability, and reproducibility of the results.
**Results**: There was strong agreement among the AudexCeph (semi-automated or automated) AudaxCeph, Dolphin, and manual methods in measuring orthodontic indices, with ICC values consistently above 0.90. Bland-Altman plots confirmed satisfactory agreement between both versions of AudaxCeph (semi-automated and automated) with the manual method, with mean differences close to 0 and about 95% of data points within the limits of agreement. However, the semi-automated AudaxCeph showed greater agreement and precision than the automated version, as indicated by narrower limits of agreement. The ICC values for inter-observer and intra-observer agreements exceeded 0.98 and 0.99, respectively.
**Conclusion**: The semi-automated AudaxCeph software offers a robust and cost-effective solution for cephalometric analysis. Its high accuracy and affordability make it a compelling alternative to Dolphin software and the manual method. ***(Imaging Sci Dent 2025; 55: 11-21)***

**KEY WORDS**: Artificial Intelligence; Deep Learning; Diagnostic Imaging; Cephalometry; Anatomic Landmarks; Orthodontics

## Introduction

A comprehensive understanding of the growth, development, and analysis of the craniofacial structures, as well as facial and dental morphology, is imperative for dental clinicians and orthodontists in particular, and plays a fundamental role in orthodontic treatment planning and the assessment of treatment outcome.[1,2] Orthodontic science depends on cephalometry for the analysis of craniofacial structures.[3] For over 70 years, orthodontists have utilized cephalometric analysis as a critical technique for diagnos-

Received May 2, 2024; Revised September 28, 2024; Accepted October 15, 2024
Published online December 6, 2024
*Correspondence to : Dr. Fatemeh Salemi
Department of Oral and Maxillofacial Radiology, Dental School, Dental Research Center, Hamadan University of Medical Sciences, Shahid Fahmideh St., Hamadan 654178-38741, Iran
Tel) 98-9183126863, E-mail) Salemifatemeh1991@gmail.com


Imaging Science in Dentistry · pISSN 2233-7822 eISSN 2233-7830

ing and planning treatments for craniofacial and dental discrepancies.[4,5] Therefore, accurate lateral cephalometric analysis is highly important in orthodontics.[6]

Anatomical landmarks are used for cephalometric analyses. Some anatomical lines and angles are formed by connecting these landmarks, enabling linear and angular measurements for orthodontic diagnosis and treatment planning.[7] Lateral cephalometric analysis and tracing are performed by trained orthodontists, but some concerns exist with respect to the time-consuming nature of manual tracing, as well as significant intra-operator and inter-operator differences using the manual technique, highlighting the need for automated landmark identification software programs to maximize the accuracy and reliability of cephalometric analyses.[3]

The Dolphin software is commonly used for cephalometric analysis. The accuracy of cephalometric analyses performed by this software, compared with the manual technique, has been well documented.[4,8] However, high cost is a major drawback of this software, and many dental clinicians in developing countries cannot afford it.

The AudaxCeph software (Audax, Ljubljana, Slovenia) is a digital image information management system that enables the observation of radiographs on a display monitor, and allows linear and angular measurements of orthodontic landmarks with its built-in tools and features. Artificial intelligence (AI) algorithms in cephalometric analysis utilize deep learning and convolutional neural networks (CNNs) to automate the identification of anatomical landmarks on radiographs. These algorithms are trained on large datasets of labeled images to learn the patterns and features associated with specific anatomical landmarks. These programs automate the identification of cephalometric points, evaluate landmarks, calculate angles and distances, and generate automated reports with diagnoses. These software programs have the advantage of automatically performing cephalometric analysis within seconds. AudaxCeph is a company that offers fully automated (AI mode) cephalometric tracing software for clinical use.[9]

The rapidity and reproducibility of these algorithms in automatic cephalometric analyses allow them to be widely used in clinical settings if their accuracy is confirmed for diagnosis and treatment planning.[4,6] Therefore, this study aimed to compare the accuracy of the AudaxCeph software versus the Dolphin software and the manual technique for orthodontic landmark identification and the tracing of lateral cephalograms.

## Materials and Methods

The protocol of this cross-sectional study was approved by the ethics committee of Hamadan University of Medical Sciences (IR.UMSHA.REC.1400.409), and all methods were carried out in accordance with the relevant guidelines and regulations.

The required sample size for this study was calculated to be 60 based on the information obtained from a previous study, considering a mean difference of 3.2, a significance level of 5%, and an analysis power of 80%.[6] This study compared the accuracy of cephalometric analyses performed by the AudaxCeph software versus the Dolphin software and the manual technique (as the gold standard) on 60 lateral cephalograms of patients who presented to the Oral Radiology Department of School of Dentistry, Hamadan University of Medical Sciences between 2021 and 2022. All radiographs used in this study were digital radiographs obtained using a ProMax X-ray unit (Planmeca, Helsinki, Finland) with the Frankfurt plane parallel to the horizon and the following exposure settings: 64-70 kVp tube potential, 12 s exposure time, and 5-15 mA tube current. The lateral cephalograms were randomly retrieved from the archives.

The inclusion criteria were as follows: 1) lateral cephalograms of individuals with permanent dentition, 2) lateral cephalograms enabling straightforward identification of anatomical landmarks, and 3) lateral cephalograms with optimal contrast and good quality.

The following lateral cephalograms were excluded: 1) radiographs of patients with a missing upper or lower first molar or incisor(s), 2) distorted radiographs, 3) cephalograms with poor quality complicating landmark identification, or exhibiting artifacts, anomalies, or asymmetry.[10,11]

In total, 23 landmarks were identified, and 10 skeletal and dental indices and 1 soft tissue index were measured (Table 1). Each lateral cephalogram was traced and evaluated 4 times: manually, with the Dolphin software, with the AudaxCeph software automatically, and with the AudaxCeph software in semi-automated mode (with manual landmark identification). Initially, the operator placed a tracing paper on lateral cephalograms and used a ruler and a template for linear and angular measurements (Fig. 1). The Dolphin software (Los Angeles, USA) is a semi-automated program that allows the operator to identify cephalometric landmarks on lateral cephalograms with a mouse on a screen. The software then performs the linear and angular measurements for the desired analysis by connecting the defined landmarks (Fig. 2). The AudaxCeph software can perform all orthodontic analyses requested by dental clinicians. This software identifies all the required landmarks automatically on the image for the type of analysis selected by the operator (Ricketts and Steiner analyses in the present study). The accuracy of the AudaxCeph software was evaluated in 2 modes. In the AI mode, the operator clicked on the automatic tracing option, and AudaxCeph's AI software marked the location for the landmarks. The positions of the landmarks were not changed after the software chose their locations. In the semi-automated mode, the operator identified the landmarks on the images, and then the software made the measurements automatically (Fig. 3).

To minimize errors, only 5 lateral cephalograms were evaluated per session, and 24-hour intervals were considered between the measurement sessions to prevent operator fatigue and eye strain. The template was calibrated using a precise ruler scale before the measurements were made. The observers included a final-year dental student and an oral and maxillofacial radiologist with over 5 years of experience, both of whom were thoroughly calibrated and trained prior to the study. Additionally, an orthodontist with over 15 years of experience provided comprehensive training on the use of the Dolphin software to the observers and supervised the manual analyses. In order to assess the intra-observer agreement, all parameters were

**Table 1.** Measured indices

| Dental indices | Skeletal indices | Soft tissue index |
|---|---|---|
| I-I | SNA (degrees) | LL -E line |
| U1 - NA (degrees) | SNB (degrees) | |
| U1 - NA (mm) | ANB (degrees) | |
| L1 - NB (degrees) | GoGn-SN (degrees) | |
| L1 - NB (mm) | | |

**Fig. 1.** Landmark identification and manual tracing of a lateral cephalogram.

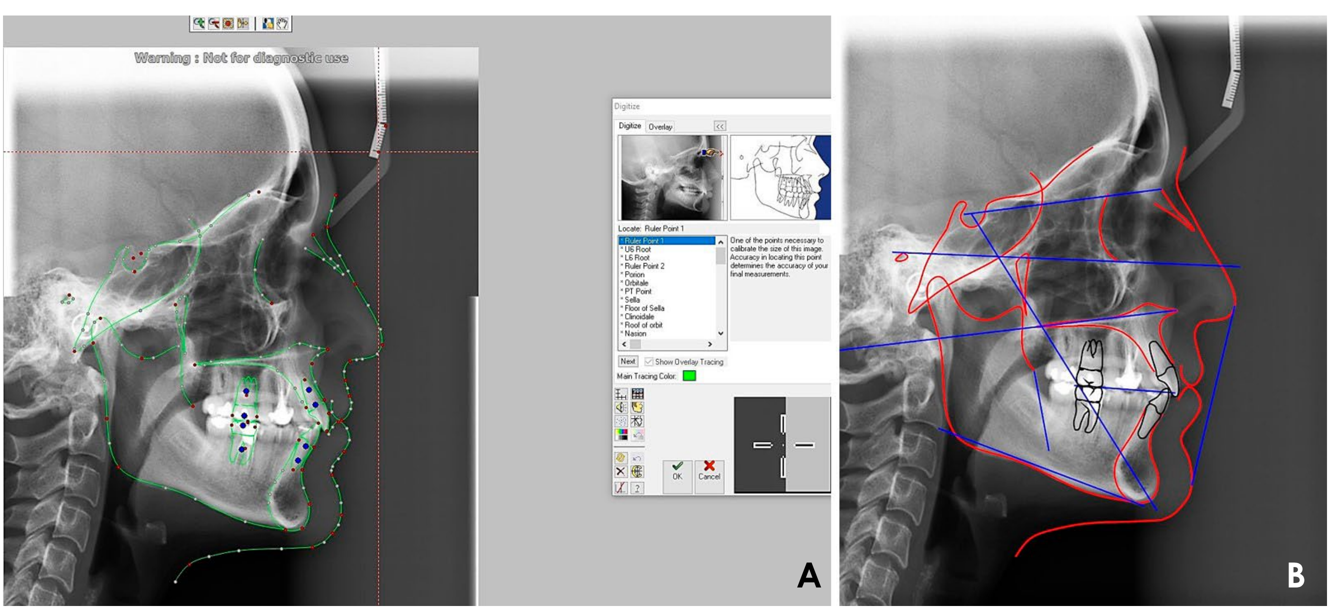


**Fig. 2.** A. User environment of the Dolphin software. B. Final cephalometric analysis performed using the Dolphin software.

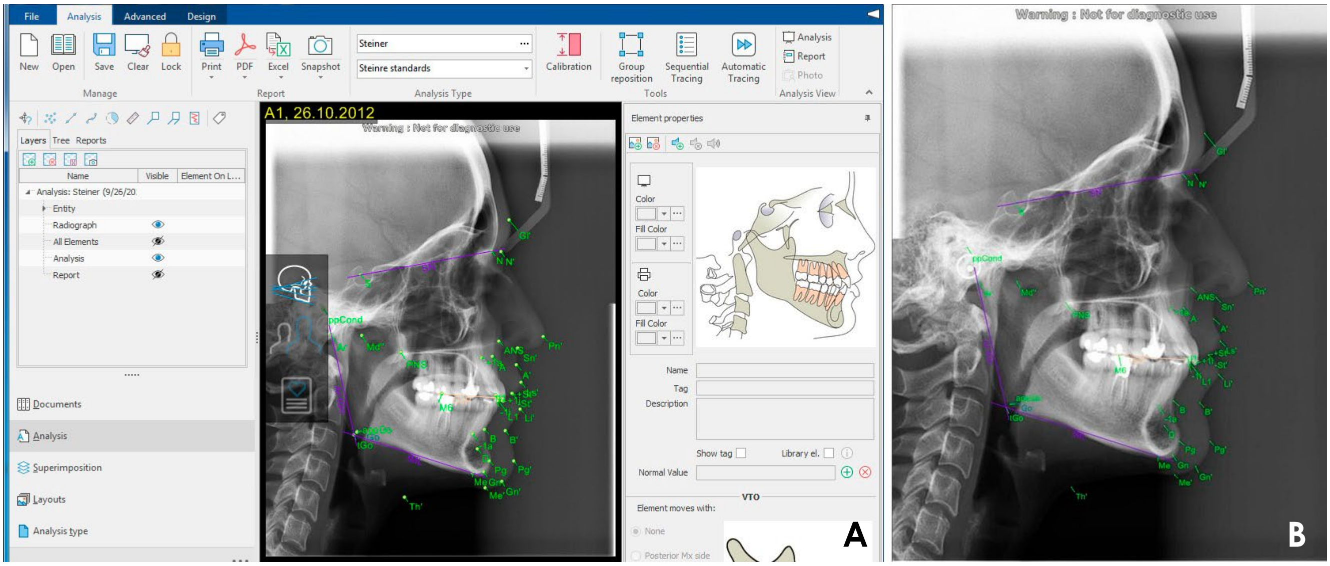


**Fig. 3.** A. User environment of the AudaxCeph software. B. Final cephalometric analysis performed using the AudaxCeph software.

measured again by the same observer after 2 weeks.

The mean and standard deviation (SD) of all indices were reported separately for each tracing method. The Kolmogorov-Smirnov test was used to assess the normality of data distribution.

The intraclass correlation coefficient (ICC) with a 95% confidence interval (CI) was calculated to assess the level of agreement between the measurements obtained with the 3 tracing methods. These calculations were based on the mean of 3 measurements (k = 3), using an absolute agreement, 2-way random-effects model. ICC values were interpreted as follows: values less than 0.5 indicated poor agreement, values between 0.5 and 0.75 reflected moderate agreement, values between 0.75 and 0.90 were interpreted as good agreement, and values greater than 0.90 indicated excellent agreement.

To evaluate the accuracy and reliability of the data, both inter-observer and intra-observer agreement were examined using the ICC. A random sample comprising 25% of the observations was selected for this assessment. Inter-observer agreement was determined by comparing measurements made by different observers, while intra-observer agreement was evaluated by comparing repeated measurements taken by the same observer.

In addition, Bland-Altman (BA) plots were created to graphically illustrate the agreement between the measurements obtained using the 2 different methods. These plots represent agreement by defining the statistical limits of agreement (LOAs) based on the mean and standard deviation of the differences between 2 measurements made by 2 different methods.

The “*irr*” package was used to calculate the ICC and the “*ggplot2*” package was employed to create BA plots in the R statistical software, version 4.2. Descriptive statistical analyses were also performed using SPSS version 23.0 (IBM Corp, Armonk, NY, USA). The significance level was set at 0.05.

## Results

Table 2 presents the ICC values for inter-observer and intra-observer agreement for all the indices. The ICC values for the inter-observer agreement exceeded 0.98 for each index and tracing method, indicating an excellent level of agreement between the observers. Similarly, the ICC values for the intra-observer agreement exceeded 0.99, demonstrating excellent agreement of each observer with himself in the measurements.

These high ICC values demonstrate the strong repeatability and reliability of the results, since the measurements were consistently reliable both when taken by different observers and when repeated by the same observer over time.

The mean and SD of the skeletal, dental, and soft tissue indices, separated by tracing methods, are presented in Table 3. The mean values of all orthodontic indices determined using the 4 tracing methods (manual, Dolphin, automated AudaxCeph, and semi-automated AudaxCeph) differed only minimally from one another., suggesting a high level of consistency among these methods.

Tables 4 and 5 display the ICC values and their 95%

**Table 2.** Intraclass correlation coefficient (ICC) values for inter-observer and intra-observer agreements of the tracing methods

| Measurements | Inter-observer | | | Intra-observer |
|---|---|---|---|---|
| | Semi-automated AudaxCeph | Dolphin | Manual | Manual |
| | ICC (95% CI) | ICC (95% CI) | ICC (95% CI) | ICC (95% CI) |
| Skeletal | | | | |
| GOGn-SN (degrees) | 1.000 (1.000-1.000) | 1.000 (0.999-1.000) | 0.984 (0.947-0.995) | 0.997 (0.995-0.998) |
| SNA (degrees) | 1.000 (0.999-1.000) | 0.997 (0.990-0.999) | 0.996 (0.983-0.999) | 0.999 (0.998-1.000) |
| SNB (degrees) | 1.000 (1.000-1.000) | 0.999 (0.997-1.000) | 0.996 (0.989-0.999) | 0.999 (0.999-1.000) |
| ANB (degrees) | 1.000 (1.000-1.000) | 0.999 (0.996-1.000) | 0.983 (0.951-0.994) | 0.994 (0.990-0.996) |
| Dental | | | | |
| U1-NA (mm) | 1.000 (1.000-1.000) | 1.000 (1.000-1.000) | 0.994 (0.948-0.998) | 0.999 (0.998-0.999) |
| U1-NA (degrees) | 1.000 (1.000-1.000) | 1.000 (1.000-1.000) | 0.997 (0.991-0.999) | 1.000 (0.999-1.000) |
| L1-NB (mm) | 0.999 (0.997-1.000) | 1.000 (0.999-1.000) | 0.987 (0.940-0.996) | 0.997 (0.996-0.998) |
| L1-NB (degrees) | 1.000 (0.999-1.000) | 1.000 (1.000-1.000) | 0.996 (0.978-0.999) | 0.999 (0.998-0.999) |
| I-I (degrees) | 1.000 (0.999-1.000) | 1.000 (1.000-1.000) | 0.999 (0.997-1.000) | 0.998 (0.997-0.999) |
| Soft tissue | | | | |
| Lower lip to E line (cm) | 1.000 (1.000-1.000) | 0.998 (0.995-0.999) | 0.997 (0.986-0.999) | 1.000 (0.999-1.000) |

CI: confidence interval

**Table 3.** Orthodontic indices for each tracing method

| Measurements | Automated Audax | Dolphin | Manual | Semi-automated Audax |
|---|---|---|---|---|
| Skeletal | | | | |
| GOGn-SN (degrees) | 33.75 ± 5.09 | 33.97 ± 4.67 | 33.88 ± 4.78 | 33.90 ± 4.70 |
| SNA (degrees) | 82.00 ± 3.66 | 81.59 ± 3.99 | 81.48 ± 3.99 | 81.51 ± 3.98 |
| SNB (degrees) | 78.00 ± 4.14 | 78.23 ± 4.32 | 78.15 ± 4.33 | 78.28 ± 4.34 |
| ANB (degrees) | 3.97 ± 2.27 | 3.38 ± 2.61 | 3.35 ± 2.61 | 3.38 ± 2.61 |
| Dental | | | | |
| U1-NA (mm) | 3.49 ± 3.29 | 4.52 ± 3.68 | 4.49 ± 3.62 | 4.51 ± 3.68 |
| U1-NA (degrees) | 21.07 ± 6.53 | 21.14 ± 7.46 | 21.07 ± 7.43 | 21.17 ± 7.41 |
| L1-NB (mm) | 5.21 ± 1.84 | 5.48 ± 2.14 | 5.50 ± 2.17 | 5.51 ± 2.17 |
| L1-NB (degrees) | 28.54 ± 5.57 | 29.64 ± 6.18 | 29.57 ± 6.14 | 29.66 ± 6.18 |
| I-I (degrees) | 126.12 ± 8.41 | 125.55 ± 9.06 | 125.55 ± 9.05 | 125.58 ± 9.09 |
| Soft tissue | | | | |
| Lower lip to E line (cm) | −0.19 ± 2.90 | −0.74 ± 2.69 | −0.67 ± 2.71 | −0.71 ± 2.70 |

CIs for the agreement among orthodontic index measurements obtained using different methods. Table 4 presents the ICC values and CIs for the manual, Dolphin, and automated AudaxCeph methods. Table 5 details the ICC values and CIs for the manual, Dolphin, and semi-automated AudaxCeph methods. All estimated ICC values for all indices were above 0.90, indicating a very high level of agreement among the methods. The 95% CIs for these ICC values were consistently above 0.90 and had lower and upper limits of more than 90%, confirming excellent agreement. The F-values calculated for the ICC were also very high, with $P$-values less than 0.001. This indicates that the ICC values are significantly different from 0, demonstrating that the methods closely agreed with each other and exhibited no substantial differences between them. Overall, these results confirm strong agreement between the measurement methods and confirm high reliability and precision across all measurement methods.

**Table 4.** Intra-class correlation coefficient (ICC) values for assessing the agreement among the manual, Dolphin, and automated AudaxCeph tracing methods in measuring orthodontic indices

| Measurements | ICC | 95% confidence interval | | F-test with true value 0 | |
|---|---|---|---|---|---|
| | | Lower bound | Upper bound | Value | *P*-value |
| Skeletal | | | | | |
| GOGn-SN (degrees) | 0.990 | 0.985 | 0.994 | 99.8 | <0.05 |
| SNA (degrees) | 0.980 | 0.969 | 0.988 | 54.8 | <0.05 |
| SNB (degrees) | 0.995 | 0.992 | 0.997 | 194.0 | <0.05 |
| ANB (degrees) | 0.963 | 0.935 | 0.978 | 32.2 | <0.05 |
| Dental | | | | | |
| U1-NA (mm) | 0.957 | 0.920 | 0.976 | 29.4 | <0.05 |
| U1-NA (degrees) | 0.983 | 0.973 | 0.989 | 56.8 | <0.05 |
| L1-NB (mm) | 0.983 | 0.973 | 0.990 | 67.8 | <0.05 |
| L1-NB (degrees) | 0.976 | 0.960 | 0.986 | 48.6 | <0.05 |
| I-I (degrees) | 0.983 | 0.974 | 0.989 | 60.4 | <0.05 |
| Soft tissue | | | | | |
| Lower lip to E line (cm) | 0.980 | 0.964 | 0.988 | 60.6 | <0.05 |

**Table 5.** Intra-class correlation coefficient (ICC) values for assessing the agreement among the manual, Dolphin, and semi-automated AudaxCeph tracing methods in measuring orthodontic indices

| Measurements | ICC (95% CI) | F-test with true value 0 | |
|---|---|---|---|
| | | Value | *P*-value |
| Skeletal | | | |
| GOGn-SN (degrees) | 0.999 (0.998-0.999) | 972 | <0.05 |
| SNA (degrees) | 0.999 (0.998-0.999) | 1072 | <0.05 |
| SNB (degrees) | 0.999 (0.999-1.000) | 1215 | <0.05 |
| ANB (degrees) | 0.998 (0.997-0.999) | 472 | <0.05 |
| Dental | | | |
| U1-NA (mm) | 0.999 (0.998-0.999) | 988 | <0.05 |
| U1-NA (degrees) | 1.000 (0.999-1.000) | 2570 | <0.05 |
| L1-NB (mm) | 0.998 (0.997-0.999) | 602 | <0.05 |
| L1-NB (degrees) | 0.999 (0.999-1.000) | 1786 | <0.05 |
| I-I (degrees) | 1.000 (0.999-1.000) | 2025 | <0.05 |
| Soft tissue | | | |
| Lower lip to E line (cm) | 0.999 (0.999-1.000) | 2175 | <0.05 |

CI: confidence interval

BA plots were created to provide a comprehensive view of the agreement and discrepancies between the tracing methods for all orthodontic indices (Figs. 4-6). Each BA plot depicts a pairwise comparisons among the 4 tracing methods for an orthodontic index, including 5 comparisons: automated AudaxCeph versus Dolphin, automated AudaxCeph versus manual, Dolphin versus manual, semi-automated AudaxCeph versus Dolphin, and semi-automated AudaxCeph versus manual, all within a single plot. Thus, Figure 4 presents BA plots comparing the tracing methods for skeletal indices. Figure 5 displays BA plots for dental indices, and Figure 6 illustrates a BA plot for the soft tissue index.

The BA plots revealed overall agreement between the automated AudaxCeph and Dolphin methods, as well as between the automated AudaxCeph and manual methods, for measurements of all skeletal, dental, and soft tissue indices. Although the LOAs were slightly wide, the mean difference was close to 0. Importantly, 95% of the data points fell within ±2 SDs of the mean difference, or within the LOAs, indicating that despite the slightly wide LOAs, both pairs of tracing methods show a satisfactory level of agreement. The BA plots comparing the Dolphin and manual methods show that for both skeletal and dental indices, the LOAs were narrow, with mean differences close to 0. Approximately 95% of data points fell within these limits, indicating strong agreement between the methods. Some indices, such as the dental index L1-NB (mm), had points on the LOA boundaries, but these differences were still within an acceptable range, reflecting generally good agreement. For soft tissue indices, over 93% of observations remained within the LOAs, demonstrating that despite some variability, the methods still exhibited good agreement. Finally, the BA plots comparing the semi-automated AudaxCeph and Dolphin methods reveal a satisfactory level of agreement across all skeletal, dental, and soft tissue indices. When comparing the semi-automated AudaxCeph with the manual method, the results show that for most indices, 95%

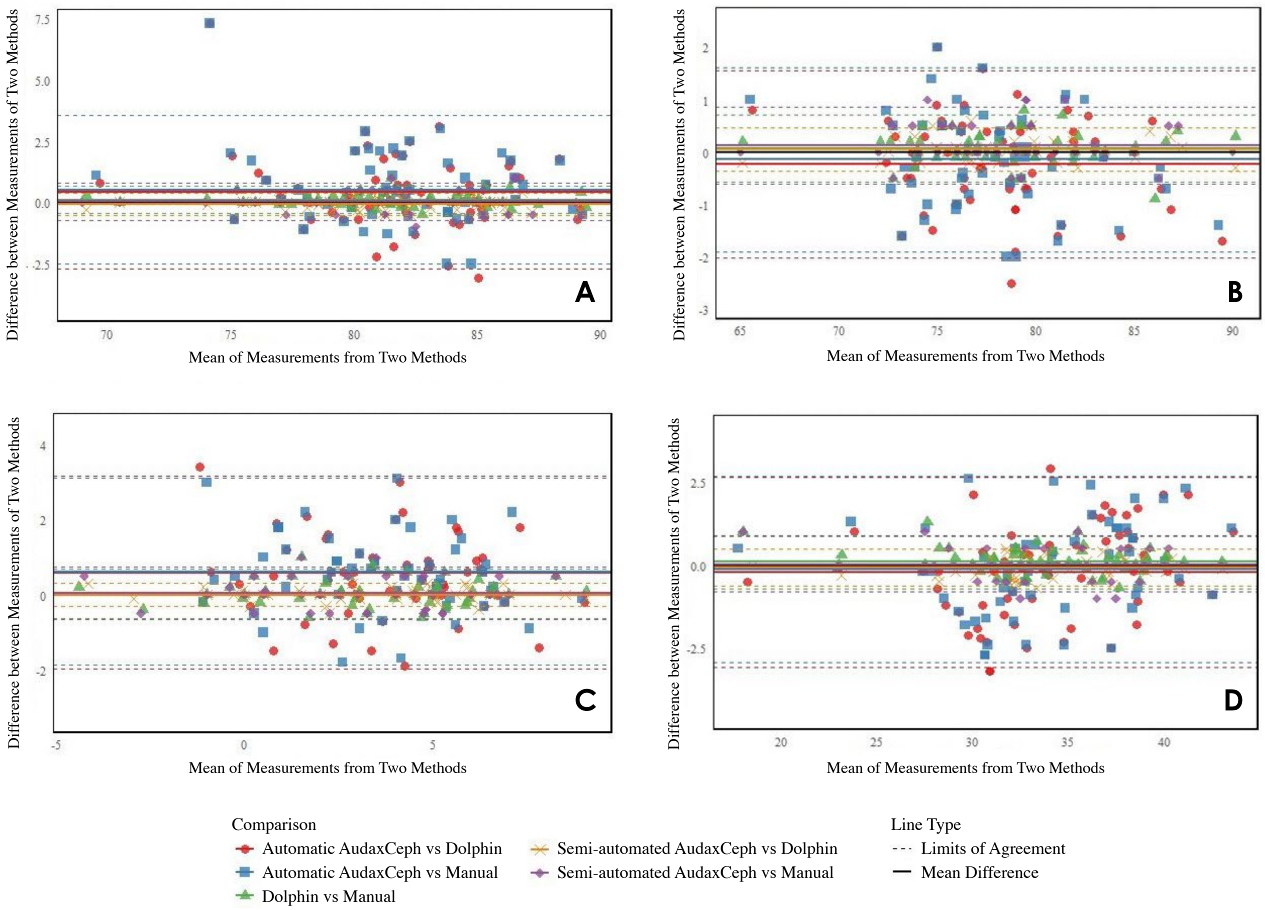


**Fig. 4.** Bland-Altman plot comparing the measurement methods for skeletal indices: A. SNA, B. SNB, C. ANB and D. GoGn-SN.

of the points fell within the LOA. However, for 3 specific indices - GoGn-SN, U1-NA (mm), and lower lip to E-line - 90% of the points were within the LOAs, indicating slightly lower but still acceptable, agreement for these measurements.

## Discussion

Similar to other professional fields, advances in science and technology revolutionized dentistry and caused a shift toward digitization and application of AI. Digital software programs for cephalometric analysis offer several advantages, including time savings, enhanced intra-examiner and inter-examiner agreement, high reproducibility, and the elimination of operator-related errors.[3]

AudaxCeph software has attracted interest from dentists and radiologists due to its low or nonexistent installation cost, which helps reduce overall patient expenses, optimize treatment outcomes, and save time. However, clinicians must always prioritize patient health and well-being, customizing the diagnostic setup and treatment to align with each patient's medical history.

The manual method was used as the gold standard reference to compare with the AudaxCeph in assessing its measurement accuracy in this study.[12,13] The choice to use manual analysis in this study is due to the high cost of the Dolphin software license, which prevents many centers from accessing this software, leading them to continue with manual methods. Chen et al. noted that the traditional hand-tracing process can be time-consuming, and the linear and angular cephalometric measurements obtained manually with a ruler and protractor may be prone to errors. Time pressures in the clinical environment can also contribute to decreased reliability. Measurement errors associated with the thickness of the pencil line and the perceptive limits of the human eye further contribute to tracing errors. Inconsistency in landmark identification is a significant source of error in manual cephalometry analysis. This error varies with each landmark and is influenced by the experience and training of the observers.[14] Computer-aided cephalo-

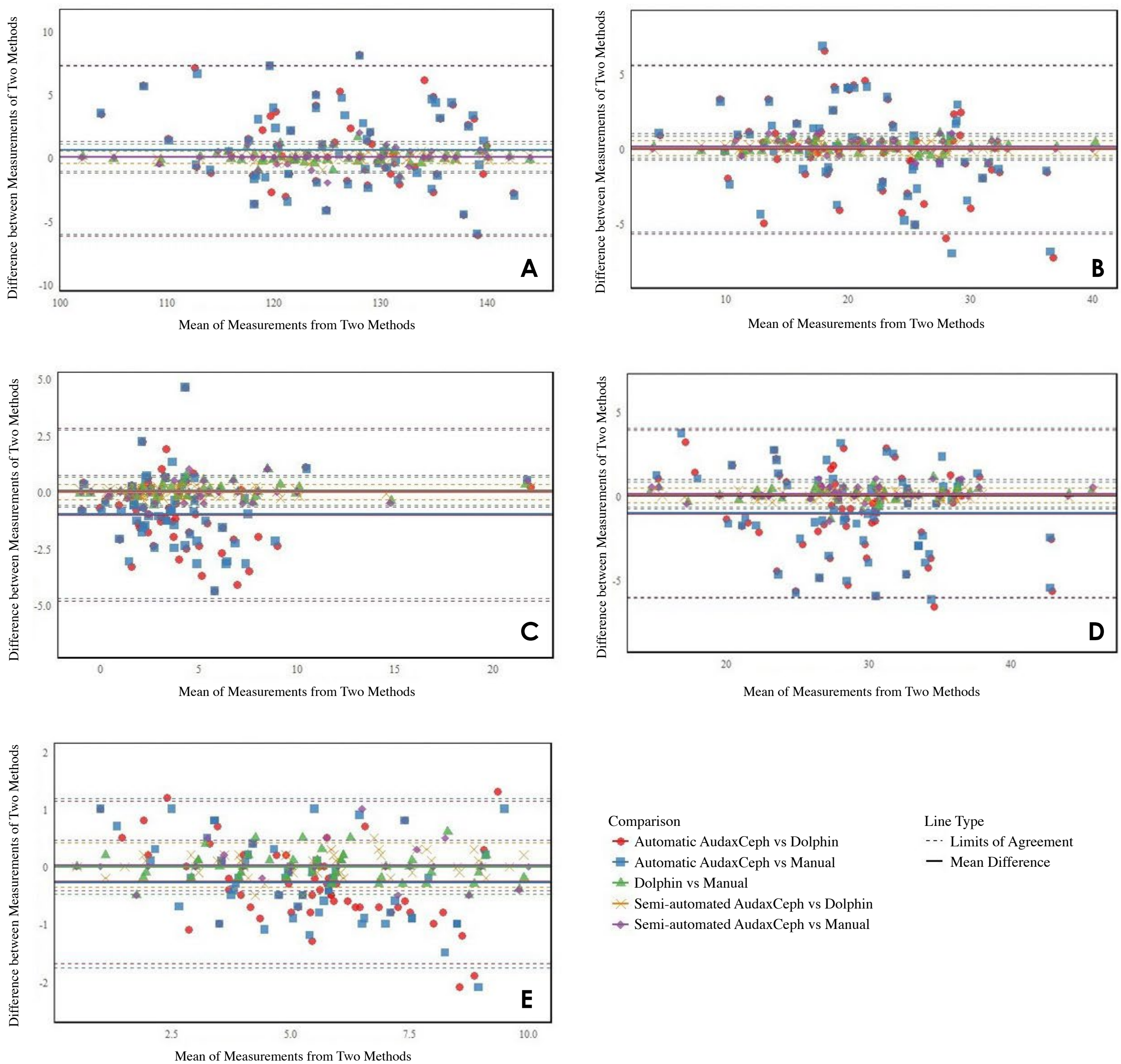


**Fig. 5.** Bland-Altman plot comparing the measurement methods for dental indices: A. I-I (degrees), B. U1-NA (degrees), C. U1-NA (mm), D. L1-NB (degrees) and E. L1-NB (mm).

metric analysis offers faster data acquisition and analysis compared to manual methods. Additionally, the image can be manipulated to enhance its visual appearance by adjusting the brightness and contrast and zooming in on the image, which can facilitate easier landmark identification. The accuracy of the Dolphin software for measuring orthodontic indices has been previously confirmed.[4,8,15] Therefore, in this study, the accuracy of the AudaxCeph software was compared with that of the manual analysis method, which served as the gold standard. In light of the disadvantages of the manual method and the advantages of digital methods, the accuracy of the AudaxCeph software was also compared with one of the most commonly used digital software, the Dolphin software.

This study employed Bland-Altman plots to evaluate the agreement between different methods for measuring orthodontic indices. Both the semi-automated and automated versions of AudaxCeph showed satisfactory agreement with manual tracing across all indices, with mean differences close to 0 and 95% of the points falling within the LOA.

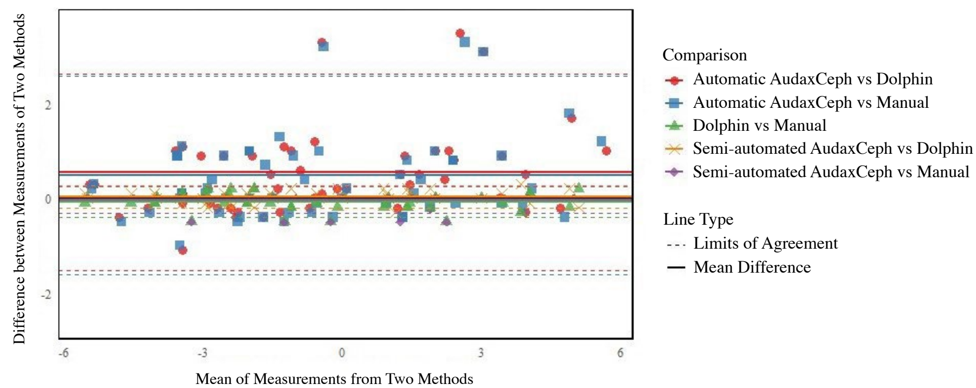


**Fig. 6.** Bland-Altman plot comparing the measurement methods for the lower lip to E line (cm) as a soft tissue index.

However, the semi-automated AudaxCeph demonstrated greater precision, as evidenced by narrower limits of agreement than the automated version, which had wider limits, indicating less consistency. This variability is likely due to inherent differences in the measurement processes and user monitoring. The semi-automated AudaxCeph, with only minor adjustments, can attain accuracy levels comparable to manual tracing, making it a practical and cost-effective alternative. It is likely that with further user revisions, the automated version of AudaxCeph could potentially achieve the desired level of accuracy.

Ristau et al.[16] compared the fully-automated and semi-automated AudaxCeph in the identification of 30 landmarks on 60 lateral cephalograms, and found no significant differences between the manual tracing by 2 orthodontists and the fully-automated software for landmark analysis except for the X and Y dimensions of the porion and the Y dimension of the L1 apex. They concluded that fully-automated AudaxCeph was suitable for tracing, and added that the difference in the identification of the porion between the tracing methods in their study was due to the fact that the porion is a double-sided structure that is often visualized as double and blurred, making its identification and tracing difficult. In the X dimension, the most common landmarks with >2 mm differences between the observers' findings with AI and software were the porion (35.8%), U1 apex (21.7%), and orbit (15.8%). These landmarks in the Y dimension were 40% for the L1 apex, 35.8% for the U1 apex, 27.5% for the porion, and 24.2% for the B point. In the clinical setting, the porion and orbit in the X dimension had no significant effect on linear and angular measurements, and therefore, they did not cause any significant clinical problems. In the Y dimension, only the porion affected the horizontal Frankfurt plane, and the points of incisor apices impacted the angular measurements of the incisors. The researchers attributed these discrepancies primarily to the placement of the X-Y axis origin and secondarily to operator eye fatigue, as each operator evaluated 30 lateral cephalograms within 1 hour and 45 minutes - a duration likely to cause eye fatigue. In the current study, each observer evaluated only 5 lateral cephalograms per session, with a 24-hour interval between sessions to prevent eye fatigue and minimize errors.

A search of the literature conducted by the authors revealed no studies similar to the current investigation, which compares the accuracy of AudaxCeph software with that of Dolphin and manual techniques. Therefore, the novelty of this study is its primary strength. However, some studies have compared the accuracy of Dolphin with other software programs.

Kazimierczak et al.[17] compared the accuracy and repeatability of cephalometric analysis results using 3 commercial AI-driven programs: CephX, WebCeph, and AudaxCeph. A total of 124 cephalograms were analyzed. The AI systems demonstrated high agreement for most parameters (ICC3 > 0.9). However, notable differences were observed in the measurements of angular convexity and the occlusal plane, suggesting that the programs employ different methodologies. Some analyses exhibited high variability in results, indicative of errors. Repeatability analysis within each program showed perfect agreement. The researchers concluded that AI-driven cephalometric analysis tools offer high potential for reliable and efficient orthodontic assessments and demonstrate high agreement in repeated analyses. Nevertheless, the observed discrepancies and high variability in some analyses highlight the need for standardization across AI platforms and a critical evaluation of automated results by clinicians, particularly for parameters that significantly impact treatment. Unlike the present study, the results were not compared with a gold standard method.

Kasinathan et al.[18] evaluated the errors and reliability of landmark identification and soft tissue cephalometric measurements using Dolphin orthodontic software compared to the manual technique. They found the results of the 2 methods to be almost comparable; however, digital tracing and identification of landmarks offer advantages such as easy archiving, retrieval, and transfer. Generally, the digital technique is superior to the manual technique for both daily application and research purposes, due to these advantages. Although the landmarks evaluated in their study differed from those in the present study, their findings on the comparable accuracy of the manual and Dolphin tracing techniques align with these results, as most data obtained from both techniques were consistent. Kumar et al.[10] conducted a comparison of cephalometric analysis between the CephNinja Android application and the NemoCeph computer software for orthodontic landmark analysis. They assessed 100 lateral cephalograms from orthodontic patients and found that the differences between the 2 tracing methods were not significant for 70% of the indices. They concluded that CephNinja provides satisfactory results compared to NemoCeph and can be reliably used as an alternative.

In AI mode, AudaxCeph exhibited some differences compared to other tracing methods. This variation was anticipated, as landmark identification is more of a cognitive process than a precise objective. Validating landmark identification proves challenging because the landmarks identified on tracing paper or by Dolphin software represent the cognitive perspective of the operator. Furthermore, each dental clinician has their own method of identifying landmarks, and there is no standardized protocol for this process.

Kim et al.[19] evaluated a fully automatic landmark identification model based on deep learning, using real clinical data, and verified its accuracy while considering inter-examiner variability. A total of 950 lateral cephalometric images from Yonsei Dental Hospital were utilized in the study. Two calibrated examiners manually identified the 13 most crucial landmarks to serve as references. The proposed deep learning model features a 2-step structure: a region of interest machine and a detection machine, each comprising 8 convolution layers, 5 pooling layers, and 2 fully connected layers. They concluded that the deep learning model is capable of achieving fully automatic identification of cephalometric landmarks and can outperform human examiners for some landmarks. Considering inter-examiner variability is beneficial for assessing the clinical applicability of deep learning methods in cephalometric landmark identification.

Kim et al.[6] analyzed 2,075 lateral cephalograms using a fully automated cephalometric analysis method that employs deep learning and a corresponding web application. They found that 23 cephalometric landmarks were automatically identified with high accuracy immediately, and the proposed algorithm yielded results highly similar to the ground truth for classifying anatomical types. Although this method outperformed the most accurate existing deep learning-based method, some landmarks were not sufficiently accurate for clinical use. In the present study, BA plots confirmed that both versions of AudaxCeph demonstrated satisfactory agreement with the Dolphin method in measuring all orthodontic landmarks. However, while the automatic version of AudaxCeph requires additional user revisions to reach clinically acceptable accuracy, the semi-automated version can achieve comparable accuracy to the Dolphin software with only minor user adjustments.

Meriç et al.[11] compared the accuracy of cephalometric analyses using fully-automated tracing, computerized tracing, and semi-automated tracing with manual adjustments, as well as the duration of each method. They assessed these parameters using Dolphin Imaging 13.01, CephNinja 3.51, and fully-automated tracing via the web-based CephX software. They noted a generally higher variability with the CephX compared to the other methods; however, the CephNinja and Dolphin showed results comparable to manual tracing. When manual corrections were applied to landmark identification by the CephX, the results were similar to those obtained with the CephNinja and Dolphin, with the CephX also demonstrating the shortest tracing time. They concluded that CephX analysis, with manual correction, holds promise for clinical applications due to its comparability with the CephNinja and Dolphin and its significantly shorter tracing time.

However, since this study, along with similar previous investigations, excluded lateral cephalograms with artifacts, anomalies, and asymmetry, the reliability of this software for tracing such images remains unknown.[16] Therefore, future studies should focus on more heterogeneous study populations. Additionally, increasing the sample size and evaluating a greater number of landmarks would enhance the reliability of the results. Therefore, it is essential for future studies to thoroughly compare various cephalometric analysis software programs, considering factors such as accuracy and cost-effectiveness, to identify the most effective options available.

In conclusion, Bland-Altman plots confirmed satisfactory agreement between both the semi-automated and automated versions of AudaxCeph and the manual method, with mean differences close to 0 and approximately 95% of data points falling within the LOA in this study. However, the semi-automated AudaxCeph demonstrated greater agreement and precision compared to its automated counterpart, as evidenced by narrower limits of agreement. While the automated version of AudaxCeph requires additional user revisions to reach clinically acceptable accuracy, the semi-automated version can achieve comparable accuracy to manual tracing with only minor user adjustments. Given its advantages-including time savings, reliable accuracy, optimal reproducibility, and lower cost - the semi-automated AudaxCeph represents a viable and cost-effective alternative for widespread use among orthodontists.

**Conflicts of Interest:** None

## Acknowledgements

This study was derived from a DDS thesis (140011129 431) submitted to the Vice Chancellor of Research of Hamadan University of Medical Sciences. The authors wish to express their gratitude to the Vice Chancellor of Research of Hamadan University of Medical Sciences.

## References

1. Chalipa J, Akhoundi MS, Shoshtarimoghaddam E, Nik TH, Imani M. Designing orthodontic craniofacial templates for 8-14 year-old Iranian girls based on cephalometric norms. J Dent (Tehran) 2013; 10: 64-73.
2. Singh Rathore A, Dhar V, Arora R, Diwanji A. Cephalometric norms for Mewari children using steiner's analysis. Int J Clin Pediatr Dent 2012; 5: 173-7.
3. Lee JH, Yu HJ, Kim MJ, Kim JW, Choi J. Automated cephalometric landmark detection with confidence regions using Bayesian convolutional neural networks. BMC Oral Health 2020; 20: 270.
4. 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: ZC20-3.
5. Sayar G, Kilinc DD. Manual tracing versus smartphone application (app) tracing: a comparative study. Acta Odontol Scand 2017; 75: 588-94.
6. 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: 105513.
7. Ludwig B, Bister D, Schott TC, Lisson JA, Hourfar J. Assessment of two e-learning methods teaching undergraduate students cephalometry in orthodontics. Eur J Dent Educ 2016; 20: 20-5.
8. Santoro M, Jarjoura K, Cangialosi TJ. Accuracy of digital and analogue cephalometric measurements assessed with the sandwich technique. Am J Orthod Dentofacial Orthop 2006; 129: 345-51.
9. Nishimoto S, Sotsuka Y, Kawai K, Ishise H, Kakibuchi M. Personal computer-based cephalometric landmark detection with deep learning, using cephalograms on the internet. J Craniofacial Surg 2019; 30: 91-5.
10. Kumar M, Kumari S, Chandna A, Konark, Singh A, Kumar H, et al. Comparative evaluation of CephNinja for Android and NemoCeph for computer for cephalometric analysis: a study to evaluate the diagnostic performance of CephNinja for cephalometric analysis. J Int Soc Prev Community Dent 2020; 10: 286-91.
11. Meriç P, Naoumova J. Web-based fully automated cephalometric analysis: comparisons between app-aided, computerized, and manual tracings. Turk J Orthod 2020; 33: 142-9.
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: 52-68.
13. Serafin M, Baldini B, Cabitza F, Carrafiello G, Baselli G, Del Fabbro M, et al. Accuracy of automated 3D cephalometric landmarks by deep learning algorithms: systematic review and meta-analysis. Radiol Med 2023; 128: 544-55.
14. Chen SK, Chen YJ, Yao CC, Chang HF. Enhanced speed and precision of measurement in a computer-assisted digital cephalometric analysis system. Angle Orthod 2004; 74: 501-7.
15. Thet PH, Kaboosaya B. Reproducibility of computerized cephalometric analysis software compared with conventional manual tracing for analyzing skeletal stability after orthognathic surgery. J Maxillofac Oral Surg 2023; 22: 833-40.
16. Ristau B, Coreil M, Chapple A, Armbruster P, Ballard R. Comparison of AudaxCeph2®'s fully automated cephalometric tracing technology to a semi-automated approach by human examiners. Int Orthod 2022; 20: 100691.
17. Kazimierczak W, Gawin G, Janiszewska-Olszowska J, Dyszkiewicz-Konwińska M, Nowicki P, Kazimierczak N, et al. Comparison of three commercially available, AI-driven cephalometric analysis tools in orthodontics. J Clin Med 2024; 13: 3733.
18. Kasinathan G, Kommi PB, Kumar SM, Yashwant A, Arani N, Sabapathy S. Evaluation of soft tissue landmark reliability between manual and computerized plotting methods. J Contemp Dent Pract 2017; 18: 317-21.
19. Kim YH, Lee C, Ha EG, Choi YJ, Han SS. A fully deep learning model for the automatic identification of cephalometric landmarks. Imaging Sci Dent 2021; 51: 299-306.