

Accuracy and Reliability of the Artificial Intelligence-Assisted WebCeph Application for Lateral Cephalometric Analysis in Comparison with the Conventional Method

Madiha Abdul Waheed and Amra Minhas Abid

Department of Orthodontics, KRL Hospital, Islamabad, Pakistan

ABSTRACT

Objective: To evaluate the accuracy and reliability of the Artificial Intelligence (AI)-assisted WebCeph application for lateral cephalometric analysis, compared with the manual tracing technique, based on 12 parameters of Steiner's cephalometric analysis.

Study Design: Descriptive, cross-sectional study.

Place and Duration of the Study: Department of Orthodontics, KRL Hospital, Islamabad, Pakistan, between June and November 2024.

Methodology: The study was performed on 30 pre-treatment lateral cephalometric radiographs. Each radiograph was analysed via two techniques: The current gold standard, i.e. conventional manual cephalometric approach and the AI-assisted WebCeph technique. Steiner's linear and angular measurements were obtained. SPSS version 25 was used for data analysis. The interclass correlation coefficient (ICC) was measured between the digital and conventional methods to determine accuracy. An ICC value below 0.75 indicated poor-to-moderate agreement. ICC value within the range of 0.75-0.90 indicated good agreement, while values >0.90 indicated excellent agreement. Intra-operator reliability was determined using a paired t-test. A p-value of $p < 0.05$ was considered as statistically significant. Normality of all the data was assessed using the Shapiro-Wilk test.

Results: All measurements, except SN-OP ($^{\circ}$), showed ICC values above 0.75. An ICC value >0.90 was recorded for five parameters (SNB, ANB, SN-Go-Gn ($^{\circ}$), UL to S-line, and LL to S-line (mm)). Six out of 12 parameters (SNA, U1-NA, L1-NB, Interincisal angle ($^{\circ}$), U1-NA, and L1-NB (mm)) obtained ICC values between 0.75-0.90. On repeated measurements, no statistically significant difference was observed as the p-value was >0.05 for all parameters in both the conventional and WebCeph groups, indicating good reliability.

Conclusion: The WebCeph showed performance at par with the human gold standard, with excellent to good agreement for the majority of the assessed variables in terms of accuracy, as well as acceptable intra-examiner reliability.

Key Words: Artificial Intelligence, Cephalometric analysis, Orthodontics, WebCeph, Human gold standard.

How to cite this article: Waheed MA, Abid AM. Accuracy and Reliability of the Artificial Intelligence-Assisted WebCeph Application for Lateral Cephalometric Analysis in Comparison with the Conventional Method. *J Coll Physicians Surg Pak* 2025; **35(07)**:4-8.

INTRODUCTION

Lateral cephalometric radiographs are a valuable component of the standardised records in orthodontic diagnosis and decision-making.¹ Cephalometric analysis involves tracings and measurements performed on cephalometric radiographs. The current gold standard involving manual tracing of anatomical cephalometric landmarks on acetate sheets is tedious and time-consuming.

Artificial intelligence (AI) refers to the simulation of human intelligence through complex computerised programmes inspired by the biological nervous system.

The introduction and application of AI have provided powerful tools that can aid orthodontists in diagnosis and decision-making.²⁻⁴

Systematic reviews suggest good reliability and accuracy of various AI-based cephalometric applications.^{3,5} One such tool is WebCeph™ (AssembleCircle Corporation, Republic of Korea), a web-based application that involves AI-assisted predictions of cephalometric landmarks and the subsequent automated analysis to provide diagnostic information.⁶ The WebCeph is an AI-supported web-based orthodontic programme, with numerous valuable features including automated cephalometric landmark identification and analysis, surgical simulations, computerised superimposition, case review, case rooms, and digital storage of records, among others. The software also allows for manual revision of the cephalometric landmarks.⁷ The digital platform provides free service for cephalometric landmark detection and analysis, thereby eliminating purchasing cost and expediting the analysis process; hence, the need for performance evaluation is important.

Correspondence to: Dr. Madiha Abdul Waheed,
Department of Orthodontics, KRL Hospital, Islamabad,
Pakistan
E-mail: madihawaheed@hotmail.com

Received: March 15, 2025; Revised: May 15, 2025;
Accepted: May 23, 2025
DOI: <https://doi.org/10.29271/jcpsp.2025.07.4>

Present data show consistent and highly accurate results for automated cephalometric landmark detection, but the evidence is prone to bias.^{8,9} Although the AI-based softwares are gaining rapid popularity over time, evidence regarding their performance in terms of accuracy and reliability is inconclusive. Considering the varying results, the present study was undertaken.

This study is aimed at evaluating the accuracy and reliability of the WebCeph cephalometric analysis. Such AI-based tools can save time and excessive effort, thereby enhancing clinical productivity.

METHODOLOGY

A descriptive, cross-sectional study was conducted at the Department of Orthodontics, KRL Hospital, Islamabad, Pakistan. Non-probability, consecutive sampling technique was employed. Sample size was 30, which was calculated by using the correlation sample size calculator; significance level was 5%, power of test was 80%, and correlation coefficient was $r = 0.5$.¹⁰ Sample size was 30. Both male and female patients, between 12 and 35 years of age, reporting to the orthodontic clinic, were included in the study. Standardised, good-quality radiographs were selected. All the selected radiographs were captured by the same operator, using the same equipment. Patients with gross asymmetry, craniofacial deformity and syndromes, unerupted or missing permanent incisors and molars, impacted teeth, and those who underwent prior orthodontic treatment were excluded.

This study received ethical approval from the Ethical Review Committee of the KRL Hospital, Islamabad, Pakistan. The study was conducted over a duration of six months between June and November 2024. Fifty lateral cephalometric radiographs that matched the criteria were chosen. Numbers were assigned from one to 50. Thirty out of the 50 radiographs were randomly selected using the random.org, a randomisation utility.

Each radiograph was analysed via two techniques: The conventional manual cephalometric technique and the digital AI-assisted WebCeph technique. Hand tracings were carried out on transparent acetate sheets on an illuminated view box using a lead pencil. Cephalometric landmarks and planes were marked. Bilateral structures were averaged and presented as a single landmark (Figure 1).¹¹ Steiner's cephalometric analysis measurements, eight angular and four linear (Table I), were carried out manually and recorded for statistical evaluation. WebCeph analysis was carried out by importing large-resolution JPG versions of all cephalograms, provided by the radiographic imaging services, to the WebCeph web application. Angular and linear measurements of Steiner's analysis were obtained and recorded. To check for intra-operator reliability, 10 out of the 30 radiographs were randomly selected and re-evaluated at a 4-week interval using both the digital and the conventional method.

SPSS Statistics (IBM Corporation, USA) version 25.0 was used for statistical analysis. Descriptive statistics were measured for qualitative and quantitative parameters. Quantitative parameters

i.e., age, angular measurements (SNA, SNB, ANB, SN-Go-Gn, U1-NA, L1-NB, SN-OP, and Interincisal angle), and linear measurements (U1-NA, L1-NB, UL to S line, and LL to S line) were measured in terms of mean and standard deviation (SD).

Interclass correlation coefficient (ICC) was measured between the digital and the conventional method to determine accuracy. ICC value below 0.75 indicated poor or moderate agreement. ICC value within the range of 0.75-0.90 indicated good agreement, while values greater than 0.90 indicated excellent or high-degree of measurement agreement. Normality of all the data was assessed using the Shapiro-Wilk's test, and the parametric test was selected. Intra-operator reliability, at 4-week interval, was determined using the paired t-test. A p-value was generated and compared. Statistical significance was set at $p < 0.05$.¹²

RESULTS

ICC for comparison between the manual and the AI-based WebCeph method exhibited the following results: All measurements, except SN-OP ($^{\circ}$), showed ICC values >0.75 , denoting good agreement in terms of accuracy (Table II). A higher ICC value >0.9 , *i.e.* excellent agreement, was obtained for five parameters, *i.e.* SNB, ANB, SN-Go-Gn ($^{\circ}$), UL to S-line, and LL to S-line (mm), while six of the 12 parameters, *i.e.* SNA, U1-NA, L1-NB, interincisal angle ($^{\circ}$) U1-NA, and L1-NB (mm), obtained ICC values between 0.75 and 0.90.

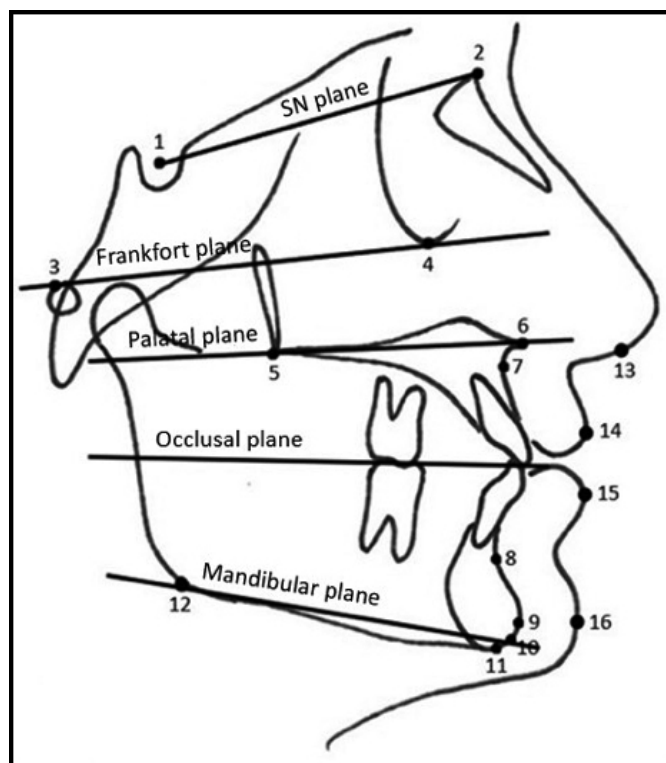


Figure 1: Cephalometric landmarks and planes. (1) Sella (S), (2) Nasion (N), (3) Porion (Po), (4) Orbitale (Or), (5) Posterior nasal spine (PNS), (6) Anterior nasal spine (ANS), (7) A point, (8) B point, (9) Pogonion (Pog), (10) Gnathion (Gn), (11) Menton (Me), (12) Gonion (Go), (13) S point (Steiner analysis), (14) Labial superius (LS), (15) Labial inferius (LI), and (16) Soft tissue pogion (Pog').

Table I: Cephalometric measurements.

Angular parameters (°)	
SNA	Anteroposterior position of the maxilla relative to the anterior cranial base
SNB	Anteroposterior position of the mandible relative to the anterior cranial base
ANB	The difference between SNA and SNB angles defines the mutual relationship in the sagittal plane of the maxillary and mandibular bases
U1-NA	Angle between the nasion-A point (NA) line and the long axis of the upper incisor
L1-NB	Angle between the nasion-B point (NB) line and the long axis of the lower incisor
SN-Go-Gn	Angle between SN plane and the mandibular plane (Go-Gn)
SN-OP	Angle between the SN plane and the occlusal plane
Interincisal angle	The angle between the axis of the upper incisor and the axis of the lower incisor
Linear parameters (mm)	
U1-NA	Linear measurement from the tip of the upper central incisor to the NA line
L1-NB	Linear measurement from the tip of lower central incisor to NB line
UL to S line	Linear measurement from the most prominent point of the upper lip to Steiner's S line
LL to S line	Linear measurement from the most prominent point of the lower lip to Steiner's S line

Table II: Comparison between the conventional and the digital WebCeph methods.

Parameters	Conventional vs. WebCeph	
	ICC ^a	95% CI ^b
Angular parameters (°)		
SNA	0.814	0.614-0.911
SNB	0.900	0.791-0.952
ANB	0.906	0.701-0.963
U1-NA	0.899	0.433-0.967
L1-NB	0.821	-0.044-0.946
SN-Go-Gn	0.940	0.870-0.972
SN-OP	0.672	0.214-0.854
Interincisal angle	0.887	0.124-0.967
Linear parameters (mm)		
U1-NA	0.856	0.605-0.939
L1-NB	0.885	0.703-0.950
UL to S line	0.910	0.790-0.959
LL to S line	0.917	0.825-0.960

^aICC, interclass correlation coefficient (>0.9 excellent; >0.75 - 0.90 good; <0.75 poor to moderate). ^bCI, confidence interval.

Table III: Mean differences, standard deviation, and correlation coefficient (intra-examiner error) for repeated measurements of digital and conventional tracings.

Cephalometric measurements	Conventional method		Digital WebCeph method	
	Difference (Mean ± SD ^a)	Paired t-test p-values	Difference (Mean ± SD ^a)	Paired t-test p-values
Angular parameters (°)				
SNA	-0.10 ± 1.45	0.832	-0.05 ± 0.67	0.819
SNB	0.10 ± 0.99	0.758	0.06 ± 0.51	0.718
ANB	-0.20 ± 1.23	0.619	-0.15 ± 0.41	0.279
U1-NA	0.40 ± 2.07	0.555	-0.30 ± 1.06	0.394
L1-NB	0.70 ± 3.97	0.591	-0.15 ± 0.94	0.627
SN-Go-Gn	-0.30 ± 1.34	0.496	-0.01 ± 0.03	0.343
SN-OP	-0.40 ± 1.08	0.269	-0.01 ± 1.04	0.976
Interincisal angle	-0.90 ± 4.53	0.546	0.41 ± 0.58	0.053
Linear parameters (mm)				
U1-NA	0.10 ± 1.17	0.794	0.16 ± 0.45	0.293
L1-NB	-0.20 ± 0.54	0.269	-0.11 ± 0.23	0.170
UL to S line	0.25 ± 0.63	0.244	0.09 ± 0.33	0.436
LL to S line	0.35 ± 0.71	0.153	0.08 ± 0.51	0.662

^aSD, standard deviation; (p >0.05, not significant).

Paired t-test for intra-examiner error exhibited no statistically significant difference (p >0.05, Table III) in both the conventional and the WebCeph groups, indicating good reliability. The largest differences noted in consecutive tracing trials were 0.30° and 0.41° for the digital WebCeph technique and 0.90° and 0.70° for the conventional approach.

DISCUSSION

With current advancements in AI technology, great achievements in the orthodontic domain are anticipated. While tracing accuracy and reliability can be a limiting factor in conventional cephalometry,¹³ studies indicate that AI-based applications

show landmark detection at par with human experts,¹⁴ and greater reliability than conventional, *i.e.* always detected identical landmark positions upon repeated trials.^{15,16} Recent studies on WebCeph also show acceptable intra-observer reliability.^{12,17,18} Results from the present study exhibited no statistically significant difference (p >0.05) between the digital and the conventional groups, indicating good reliability.

While some studies including the present study, evaluating accuracy of the WebCeph in comparison with the traditional tracing method show acceptable results,^{12,18} suggesting that the WebCeph can be an aid to the orthodontists, literature showing contradictory conclusions exists. Comparing the

findings of the present study to similar studies aimed at assessing the accuracy of the fully-automated WebCeph software, some differences were observed. A recent study by Baig *et al.* showed significant inaccuracies and a lack of reliability in AI-based fully-automated lateral cephalometric analysis using the WebCeph software, in comparison with the gold-standard hand-tracing approach. Statistically significant differences were obtained for 10 out of the 11 measurements.¹⁹ Similar results were noted by other studies, although the results are promising for the identification of certain points.²⁰

Kunz *et al.* in their study comparing the WebCeph with the human gold standard, showed no significant mean difference in any of the nine examined measurements. However, WebCeph exhibited a high possibility of proportional bias. Accuracy was not clinically acceptable for the WebCeph dental analysis.²¹

Comparing the WebCeph with the semi-automated AutoCAD software, *i.e.* manual landmark identification, followed by automated angular and linear calculations, Yassir *et al.* in their study showed similar findings, with poor landmark detection and inconsistent results with the automated WebCeph. Authors, therefore, suggest caution when using the software for cephalometric analysis, with supervision by an experienced clinician.²²

Similarly, in another study, WebCeph showed significant differences ($p < 0.05$) in landmarks recognised by the digital application. Human experts showed excellent reproducibility ($ICC \geq 0.9943$), whereas the WebCeph showed good reproducibility with $ICC \geq 0.7868$.²³ The authors concluded that the WebCeph produced significant errors, with inconsistent and incorrect landmark identification.

Another recent study evaluating the accuracy of the fully-automated WebCeph and OrthoDx softwares vs. non-automated manual landmark marking *via* the Dolphin software showed statistically significant favourable results for the angular parameters. Linear parameters and soft tissue measurements showed weak correlation. Therefore, manual intervention is required in order to minimise errors when using AI-assisted fully-automated software for cephalometric evaluation.²⁴ The present study showed excellent-to-good agreement for all angular and linear measurements, except the SN-OP ($^\circ$), which produced an ICC value of 0.672, indicating poor-to-moderate agreement.

Advances in AI technology are rapid, but AI models and algorithms require further refinement and testing. Although findings from the present study indicate good agreement between the WebCeph technique and the manual cephalometric tracing method, at present, digital technology cannot completely overtake or replace the orthodontist's role in cephalometric diagnosis and clinical decision-making. Systematic reviews and meta-analyses on AI-assisted cephalometric landmark detection propose further research due to high risk of bias in the existing literature.^{9,25} A recent umbrella review

illustrated erroneous automated cephalometric landmark detection with limited accuracy, suggesting verification from a trained orthodontist.²⁵

A key limitation of this study is that the landmark detection and evaluation by human expert was done by one examiner only. Despite sufficient clinical experience, assessment by human experts can be susceptible to errors. Therefore, for a more accurate gold standard assessment, a mean value for each parameter examined by more than two orthodontists could be obtained.

CONCLUSION

Accuracy of the AI-assisted WebCeph cephalometric analysis is at par with the human gold standard. Excellent agreement was obtained for five of the 12 cephalometric parameters. Six of the 12 parameters indicated good agreement. In terms of intra-examiner reliability, both the WebCeph and the human gold standard showed acceptable results at detecting identical landmark positions upon repeated trials.

ETHICAL APPROVAL:

This study received ethical approval from the Ethical Review Committee of the KRL Hospital, Islamabad, Pakistan (Ref. No: KRL-HI-ERC-May21/25).

PATIENTS' CONSENT:

Informed consent was obtained from all participants included in the study.

COMPETING INTEREST:

The authors declared no conflict of interest.

AUTHORS' CONTRIBUTION:

MAW: Study design, data collection, analysis, and manuscript writing.

AMA: Study design and critical review.

Both authors approved the final version of the manuscript to be published.

REFERENCES

1. Baeshen HA, Helal NM, Basri OA. Significance of cephalometric radiograph in orthodontic treatment plan decision. *J Contemp Dent Pract* 2019; **20(7)**:789-93. doi: 10.5005/jp-journals-10024-2598.
2. Pethani F. Promises and perils of artificial intelligence in dentistry. *Aust Dent J* 2021; **66(2)**:124-35. doi: 10.1111/ADJ.12812.
3. Khanagar SB, Al-Ehaideb A, Vishwanathaiah S, Maganur PC, Patil S, Naik S, *et al.* Scope and performance of artificial intelligence technology in orthodontic diagnosis, treatment planning, and clinical decision-making - A systematic review. *J Dent Sci* 2021; **16(1)**:482-92. doi: 10.1016/J.JDS.2020.05.022.
4. Siddiqui TA, Sukhia RH, Ghandhi D. Artificial intelligence in dentistry, orthodontics and Orthognathic surgery: A literature review. *J Pak Med Assoc* 2022; **72(Suppl 1)**:S91-6. doi: 10.47391/JPMA.AKU-18.

5. Khanagar SB, Al-ehaideb A, Maganur PC, Vishwanathaiah S, Patil S, Baeshen HA, *et al.* Developments, application, and performance of artificial intelligence in dentistry - A systematic review. *J Dent Sci* 2021; **16**(1):508-22. doi: 10.1016/J.JDS.2020.06.019.
6. WEBCEPH. (Accessed 8 May 2025). Available from: <https://webceph.com/en/guide/>.
7. Kilinc DD, Kircelli BH, Sadry S, Karaman A. Evaluation and comparison of smartphone application tracing, web based artificial intelligence tracing and conventional hand tracing methods. *J Stomatol Oral Maxillofac Surg* 2022; **123**(6):e906-15. doi: 10.1016/J.JORMAS.2022.07.017.
8. Schwendicke F, Chaurasia A, Arsiwala L, Lee JH, Elhennawy K, Jost-Brinkmann PG, *et al.* Deep learning for cephalometric landmark detection: Systematic review and meta-analysis. *Clin Oral Investig* 2021; **25**(7):4299. doi: 10.1007/S00784-021-03990-W.
9. de Queiroz Tavares Borges Mesquita G, Vieira WA, Vidigal MTC, Travencolo BAN, Beaini TL, Spin-Neto R, *et al.* Artificial intelligence for detecting cephalometric landmarks: A systematic review and meta-analysis. *J Digit Imaging* 2023; **36**(3):1158. doi: 10.1007/S10278-022-00766-W.
10. Zamrik OM, Iseri H. The reliability and reproducibility of an Android cephalometric smartphone application in comparison with the conventional method. *Angle Orthod* 2021; **91**(2):236-42. doi: 10.2319/042320-345.1.
11. Alexander J. Radiographic Cephalometry: From Basics to Videoimaging. Chicago: Quintessence Pub. Co., 1995. Available from: https://catalog.nlm.nih.gov/permalink/01NLM_INST/m5fc0v/alma997924983406676.
12. Mahto RK, Kafle D, Giri A, Luintel S, Karki A. Evaluation of fully automated cephalometric measurements obtained from web-based artificial intelligence driven platform. *BMC Oral Health* 2022; **22**(1):132. doi: 10.1186/S12903-022-02170-W.
13. Kamoen A, Dermaut L, Verbeeck R. The clinical significance of error measurement in the interpretation of treatment results. *Eur J Orthod* 2001; **23**(5):569-78. doi: 10.1093/EJO/23.5.569.
14. Hwang HW, Moon JH, Kim MG, Donatelli RE, Lee SJ. Evaluation of automated cephalometric analysis based on the latest deep learning method. *Angle Orthod* 2021; **91**(3):329-35. doi: 10.2319/021220-100.1.
15. 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. doi: 10.2319/022019-129.1.
16. Kazimierczak W, Gawin G, Janiszewska-Olszowska J, Dyszkiewicz-Konwinska M, Nowicki P, Kazimierczak N, *et al.* Comparison of three commercially available, AI-driven cephalometric analysis tools in orthodontics. *J Clin Med* 2024; **13**(13):3733. doi: 10.3390/JCM13133733.
17. Zaheer R, Shafique HZ, Khalid Z, Shahid R, Jan A, Zahoor T, *et al.* Comparison of semi and fully automated artificial intelligence driven softwares and manual system for cephalometric analysis. *BMC Med Inform Decis Mak* 2024; **24**(1):271. doi: 10.1186/S12911-024-02664-3.
18. Prince STT, Srinivasan D, Duraisamy S, Kannan R, Rajaram K. Reproducibility of linear and angular cephalometric measurements obtained by an artificial-intelligence assisted software (WebCeph) in comparison with digital software (AutoCEPH) and manual tracing method. *Dental Press J Orthod* 2023; **28**(1):e2321214. doi: 10.1590/2177-6709.28.1.E2321214.OAR.
19. Baig N, Gyasudeen KS, Bhattacharjee T, Chaudhry J, Prasad S. Comparative evaluation of commercially available AI-based cephalometric tracing programs. *BMC Oral Health* 2024; **24**(1):1241. doi: 10.1186/s12903-024-05032-9.
20. Moreno M, Gebeile-Chauty S. Comparative study of two software for the detection of cephalometric landmarks by artificial intelligence. *Orthod Fr* 2022; **93**(1):41-61. doi: 10.1684/ORTHODFR.2022.73.
21. Kunz F, Stelzig-Eisenhauer A, Widmaier LM, Zeman F, Boldt J. Assessment of the quality of different commercial providers using artificial intelligence for automated cephalometric analysis compared to human orthodontic experts. *J Orofac Orthop* 2023; **86**(3):145-60. doi: 10.1007/S00056-023-00491-1.
22. Yassir YA, Salman AR, Nabbat SA. The accuracy and reliability of WebCeph for cephalometric analysis. *J Taibah Univ Med Sci* 2021; **17**(1):57-66. doi: 10.1016/J.JTUMED.2021.08.010.
23. Silva TP, Pinheiro MCR, Freitas DQ, Gaeta-Araujo H, Oliveira-Santos C. Assessment of accuracy and reproducibility of cephalometric identification performed by 2 artificial intelligence-driven tracing applications and human examiners. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2024; **137**(4):431-40. doi: 10.1016/J.OOOO.2024.01.011.
24. Duran GS, Gokmen S, Topsakal KG, Gorgulu S. Evaluation of the accuracy of fully automatic cephalometric analysis software with artificial intelligence algorithm. *Orthod Craniofac Res* 2023; **26**(3):481-90. doi: 10.1111/OCR.12633.
25. Polizzi A, Leonardi R. Automatic cephalometric landmark identification with artificial intelligence: An umbrella review of systematic reviews. *J Dent* 2024; **146**:105056. doi: 10.1016/j.jdent.2024.105056.

• • • • •