

Original Article

Comparison Between the Accuracy of Different AI-Based Cephalometric Analysis and Conventional Manual Analysis

Mohammed A. Mahmood^{*1}, Adham A. Abdulrahman², Anwar A. Amin², Hadi M. Ismail²

Abstract

Objective: This study compared artificial intelligence-based cephalometric analysis platforms with conventional manual analysis using the Steiner Cephalometric analysis system.

Methods: Three AI online software platforms: WEBCEPH, AssembleCircle Corp, South Korea, CEPHIO, Cephio Sp. Z o.o., Poland, and VOXEL Voxe3Di, United Kingdom, were evaluated using 25 cephalometric radiographs and compared with two manual tracings performed using (Lab Pronto, Blue Sky Plan) for the Steiner Cephalometric analysis system. Independent sample t-test was used to compare the two manual analysis. The Kruskal-Wallis H-test, followed by pairwise comparisons, was used to evaluate differences between AI-based and manual analyses, with significance at $p \leq 0.05$.

Results: No significant discrepancies were found between the two manual analyses. However, the Kruskal-Wallis H-test and pairwise comparisons revealed significant differences between the AI platforms regarding the SNA angle and the U1 to NA angle. Specifically, VOXEL AI analysis differed from WEBCEPH (at $p=0.003$) and CEPHIO (at $p=0.011$) for these measurements.

Conclusions: The findings highlight that while AI platforms such as VOXEL show promising alignment with manual analysis across most measurements, discrepancies, particularly in specific angles, underscore the need for cautious integration and ongoing validation of AI in orthodontic practice. The selection of appropriate AI-based analysis is crucial; however, manual analysis remains the gold standard for cephalometric analysis.

Keywords: Artificial intelligence, Orthodontics, Diagnosis, Automated cephalometric analysis.

Submitted: April 20, 2025, Accepted: July 20, 2025, Published: December 1, 2025.

Cite this article as: Mahmood MA, Abdulrahman AA, Amin AA, Ismail HM. Comparison Between the Accuracy of Different AI-Based Cephalometric Analysis and Conventional Manual Analysis. Sulaimani Dent J. 2025;12(3):1-9.

DOI: <https://doi.org/10.17656/sdj.10211>

1. Department of Basic Science, College of Dentistry, University of Sulaimani, Sulaimani, Iraq.
2. Department of Orthodontics, College of Dentistry, University of Sulaimani, Sulaimani, Iraq.

* Corresponding author: mohammed.mahmood@univsul.edu.iq.



Published by College of Dentistry, University of Sulaimani



Introduction

Artificial intelligence (AI) may contribute to improved healthcare quality due to the increased quality of diagnostic methods and the elimination of diagnostic errors in daily medical practice. For a century, cephalometric analysis was introduced to the field of orthodontics and became one of the important tools in diagnostic procedures. Moreover, it has been used to evaluate the treatment procedure and monitor the treatment outcomes¹⁻⁴. Manual tracing using tracing paper or tracing software is the most conventional method to identify landmarks and analyze lateral cephalometric radiographs⁵. Digital analysis appeared, and several programs were introduced to perform the analysis, which led to the speed and the accuracy of cephalometric analysis improving significantly⁶. Although analysis with AI was a big development in orthodontics, supervision is still mandatory in identifying the landmarks⁷.

The evolution in the field of cephalometric analysis has been marked by the implementation of artificial intelligence (AI) in analysis processes, which is increasing dramatically^{6,7}. Several studies indicated that AI-based analysis is a useful tool for identifying cephalometric landmarks and performing analysis, and its measurement can be trusted for orthodontic diagnosis^{5,6,8}. As such, the number of studies confirming the reliability and validity of AI-based analysis is still increasing. One of the apparent advantages of AI-based analysis is its speed, which saves orthodontists time during the examination phase.

The introduction of AI into cephalometric analysis marks a pivotal advancement in the development of diagnostic instruments in orthodontics. Historically, landmark identification was conducted using hand tracing techniques and was regarded as the gold standard for lateral cephalometric research. The emergence of digital analysis has enhanced the efficiency and accuracy of cephalometric evaluations². The automated methods for identifying and analysing cephalometric landmarks represent a significant advancement in cephalometric analysis procedures. They have markedly enhanced efficiency, rendering them progressively appropriate for routine diagnosis in orthodontics⁹. Their improved speed and precision provide significant advantages, indicating a considerable change in the domain. The rising endorsement of daily utilization by orthodontic professionals underscores the escalating importance of these automated techniques in orthodontic practice¹⁰.

The advent of AI-driven cephalometric analysis has resulted in a significant decrease in the time needed for evaluation. Nonetheless, the precision of the research requires further examination, and other studies have emerged in the sector during the past three years. Cephalometry plays a crucial role in orthodontic

diagnosis and treatment planning, making the integration of AI a compelling proposition and stimulating research to assess its reliability and usefulness¹. Furthermore, AI systems can perpetually enhance their capabilities by assimilating new data, rendering them versatile for various craniofacial deformities and heterogeneous patient demographics. This adaptability is essential in orthodontics, where individual anatomical differences greatly influence diagnostic and treatment results. The capacity of AI systems to learn and adapt provides a versatile instrument that can align with the changing clinical requirements and progress in orthodontic research.

Notwithstanding encouraging outcomes, thorough comparative studies are imperative to assess the efficacy of diverse AI-based platforms in relation to conventional manual techniques utilizing standardized systems such as the Steiner Cephalometric analysis system. These studies confirm the validity and dependability of AI systems in various therapeutic contexts and patient populations. By comparing different AI platforms, researchers can identify the most effective systems, guiding orthodontists in selecting the best tools for their practice. Therefore, this study aimed to compare different AI-based cephalometric analysis platforms and compare each of these platforms with conventional manual cephalometric analysis using a definite analysis system (Steiner Cephalometric analysis system).

Materials and methods

The study was approved by the ethical committee of the College of Dentistry at the University of Sulaimani with reference number (CoD-EC-24-004) on August 29, 2024.

Sample Selection

In this observational study, data were retrospectively extracted from lateral cephalometric radiographs of individuals currently undergoing orthodontic treatment, sourced from the patient database of the College of Dentistry, University of Sulaimani. For this study, 25 patients who had undergone lateral cephalometric radiographs between 2021 and 2024 were selected. The G Power software (version 3.1, Heinrich Heine-University Dusseldorf, Germany) was utilized to determine the statistical power, aiming for 80%, at a significance level (alpha) of 0.05 using a two-sided paired t-test. Based on findings from a prior study¹², it was determined that a sample size of 25 images would be sufficient, assuming an effect size of 0.49.

All cephalometric radiographs were obtained using the PaX-I smart (Vatech, Seoul, Korea) with settings of 85 kVp, 10 mA, and an exposure time of 1.2-3.3 seconds.

The same technician took all the radiographs. Patients were in a standing position with the Frankfort plane parallel to the ground, teeth in centric occlusion, and lips relaxed. Inclusion criteria were cephalometric radiographs of patients aged 14-30 years of both sexes, and the radiographs were selected randomly regardless of malocclusion type to maintain generalizability. Exclusion criteria included radiographs with artifacts, prosthetic restorations, and patients with cleft lip, palate, or trauma.

All cephalometric images were exported as JPG files and saved on a computer. Each radiograph was analysed twice using (Lab Pronto and BlueSky Plan) (Figure 1). Tracing was performed directly on the software.

Cephalometric Analysis Systems: Steiner's cephalometric analysis system (Figure 2) serves as the common framework for both AI-based and manual analyses. Three distinct AI-based platforms: WEBCEPH, Assemble Circle Corp, South Korea, CEPHIO, Cephio Sp. z o.o., Poland, and VOXEL Voxel3Di, United Kingdom, were individually employed (Figures 3, 4, and 5). Steiner analysis is available in all three AI-based systems, and ten readings were selected which were present across all platforms. The same quality radiograph was uploaded to all three AI analysis platforms, and the automated digitization was saved without any manual modifications. The next step involved ordering Steiner cephalometric analysis using the predetermined options in the AI analysis systems. Prior to formal data collection, a calibration session was conducted. Ten randomly selected cephalograms were traced by both examiners, and intraclass correlation coefficients (ICC) were calculated to ensure consistency ($ICC > 0.9$ for all parameters). For AI-based methods, all radiographs were scaled using the embedded reference rulers within each platform prior to analysis.

Manual Analysis: Tracing software was used and the points placed by operators to perform the Steiner system. This conventional method was carried out by two experienced orthodontists, each with a minimum of 15 years of clinical proficiency in cephalometric analysis. Both linear and angular measurements derived from manual and AI-based analyses were accurately considered during the comparative assessment (Table 1).

Landmark Identification and Measurement Process

1. Identification of Landmarks: The following landmarks were identified on each radiograph:

1. Sella (S): The midpoint of the pituitary fossa.
2. Nasion (N): The most anterior point on the frontonasal suture.
3. A Point (A): The deepest point on the contour of the premaxilla.
4. B Point (B): The deepest point on the contour of the mandible.
5. Pogonion (Pog): The most anterior point on the chin.
6. Menton (Me): The lowest point on the mandible.
7. Gonion (Go): The most posterior-inferior point on the angle of the mandible.
8. Upper Incisor Tip (U1): The tip of the upper central most protruded incisor.
9. Lower Incisor Tip (L1): The tip of the lower central incisor.
2. Tracing Landmarks: The identified landmarks were traced on digital images using tracing software. The orthodontists ensured precision by consistently using the same set of landmarks for every analysis.

Statistical Analysis:

Microsoft Excel, IBM SPSS statistical package version 29, and DATAtab Team (2024). DATAtab: Online Statistics Calculator. DATAtab e.U. Graz, Austria, were used to tabulate and analyze the data; the Shapiro-Wilk test was used to test the data's normality. Independent sample t-test was used to show the inter-examiner differences between the two manual tracings. Kruskal-Wallis H-test was used to represent the differences between all groups.

Results

Twenty-five cephalometric radiographs were used in this study, the distribution of the data summarized in Table 1. The radiographs were analyzed by three AI online software programs and manual tracing by two experienced orthodontists. The resulting data was tabulated in Microsoft Excel and analyzed by IBM SPSS Statistics version 29.0.1.0 and DATAtab Team (2024). DATAtab: Online Statistics Calculator. DATAtab e.U. Graz, Austria.

Shapiro-Wilk test showed that the distribution of data in the manual analysis groups was normal. In contrast, for most of the AI analysis, it deviated from normality, and accordingly, the Kruskal-Wallis H-test (non-parametric analysis) was used to show differences between the groups. The SNA angle measurement shows significant difference among groups (p-value= 0.007), also U1 to NA(deg) shows significant difference between groups (p-value= 0.002).

The t-test indicates no significant differences between the two manual analyses (Table 2) (t = -0.25, p-value = 1).

Median and interquartile range were calculated for all variables, then the results of the Kruskal Wallis H-test with pairwise comparison between groups were shown. VOXEL and WEBCEPH results were different in the analysis of SNA (p-value= 0.003); also, VOXEL and CEPHIO, and VOXEL and WEBCEPH were significantly different regarding the measurement of U1 to NA angle (p-value= 0.031).

Table 1: Steiner Cephalometric Analysis System.²⁴

Measurement	Description
SNA Angle	Angle formed by the SN line and the NA line.
SNB Angle	Angle formed by the SN line and the NB line.
ANB Angle	Angle formed by the NA line and the NB line.
Occlusal Plane to SN Angle	Measures the angle between the occlusal plane and the Sella-Nasion line.
Mandibular plane angle (Go-Gn to SN)	Indicates the vertical relationship between the cranial base and the mandibular plane.
U1-NA Angle	Measures the inclination of the upper incisors relative to the NA line.
U1-NA Linear	Measures the linear distance of the upper incisors to the NA line.
L1-NB Angle	Measures the inclination of the lower incisors relative to the NB line.
L1-NB Linear	Measures the linear distance of the lower incisors to the NB line.
Interincisal Angle	Measures the angle formed by the intersection of the long axes of the upper and lower incisors.

Table 2: Showing the distribution of the study sample (median and interquartile range), and the differences between and within the study groups.

Measurements	VOXEL	WEBCEPH	CEPHIO	Manual 1	Manual 2	Kruskal Wallis Test (p-value)
SNA	79.12 (4.81) ^A	82.69(2.67) ^B	80.9(3.5) ^{AB}	80(5) ^{AB}	80.9(4.2) ^{AB}	0.007
SNB	74.81(0)	76.99(0.29)	76.8(0.02)	76(0.33)	77.7(0.31)	0.193
ANB	3.94(3.66)	5.97(4.73)	4.3(5.9)	5(5)	4.2(4.5)	0.695
Occlusal plane to SN angle	18.75(5.02)	16.86(7.5)	16(5.6)	14(5)	15.8(5.5)	0.089
Mandibular plane angle (Go-Gn to SN)	33.77(9.44)	32.44(7.31)	34.9(7)	32(7)	34.3(6.9)	0.374
U1 to NA (mm)	4.42(3.62)	3.42(2.23)	4.6(3)	4(3)	4.1(2.6)	0.188
U1 to NA (deg)	27.92(8.31) ^D	21.29(9.35) ^C	20.7(9.9) ^C	28(11) ^{CD}	25.8(8.8) ^{CD}	0.002
L1 to NB (mm)	4.64(3.49)	5.12(4.09)	3.8(5.5)	5(5)	3.9(3)	0.751
L1 to NB (deg)	23.57(10.3)	27.41(7.87)	28.6(11.2)	30(10)	27.9(12)	0.211
Interincisal angle	121.76(21.18)	127.45(19.45)	127.5(17.6)	120(19)	123.7(23.6)	0.152



Figure 1: Screenshot of the Steiner analysis by manual analysis using (BlueSky Plan).

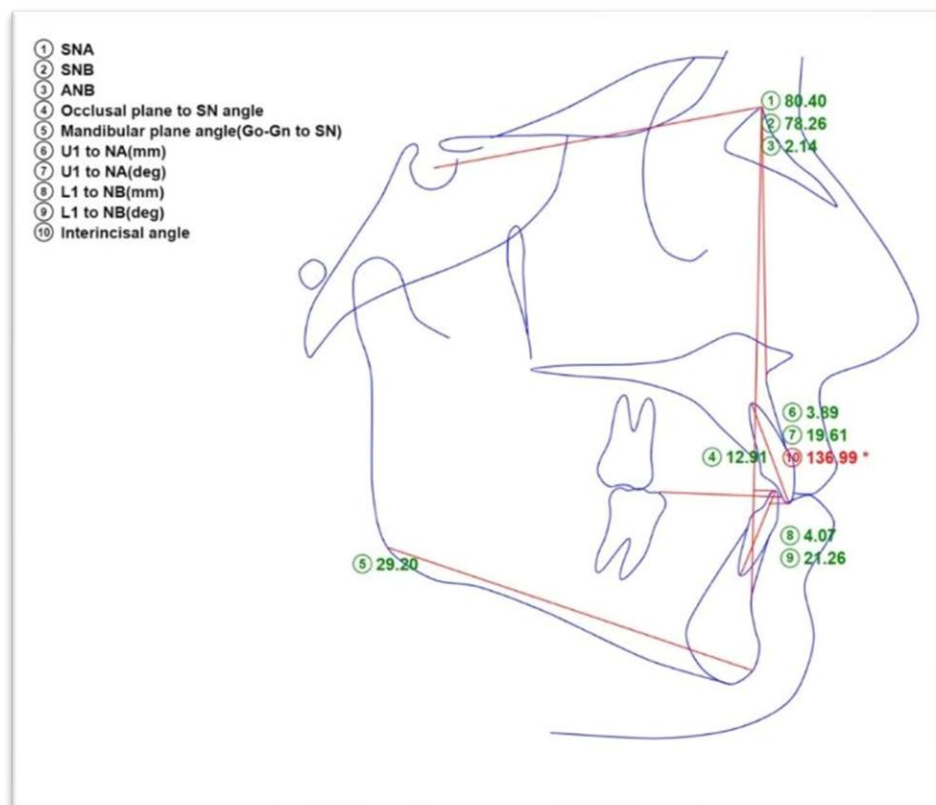


Figure 2: Steiner Analysis Measurements.

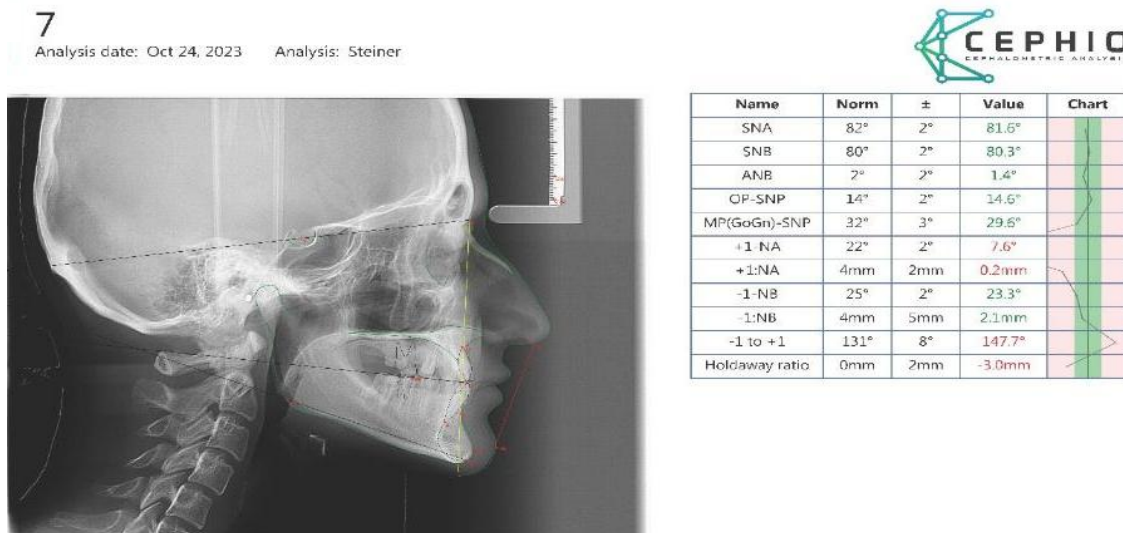


Figure 3: Screenshot of the CEPHIO analysis system and the output report.

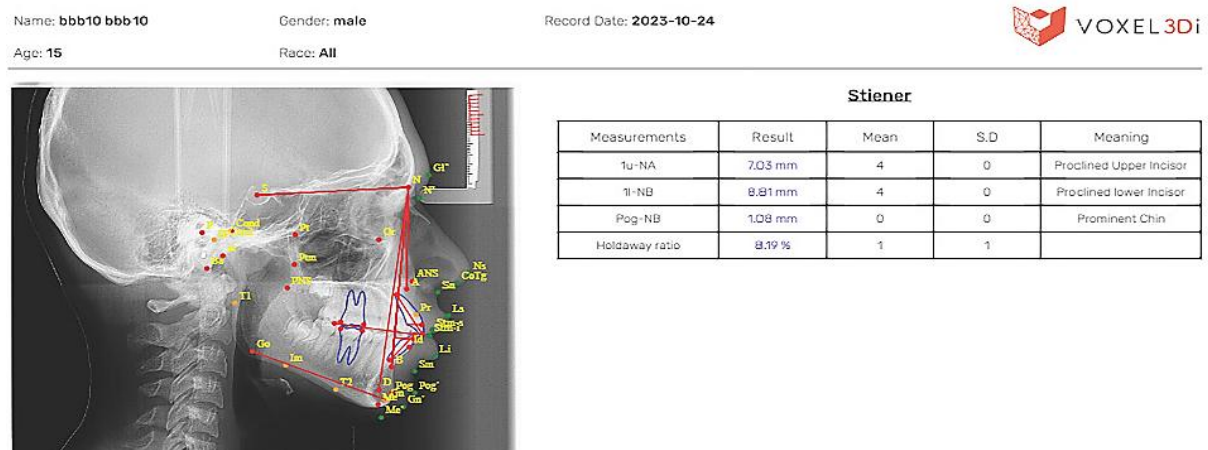


Figure 4: Screenshot of the VOXEL analysis system and the output report.

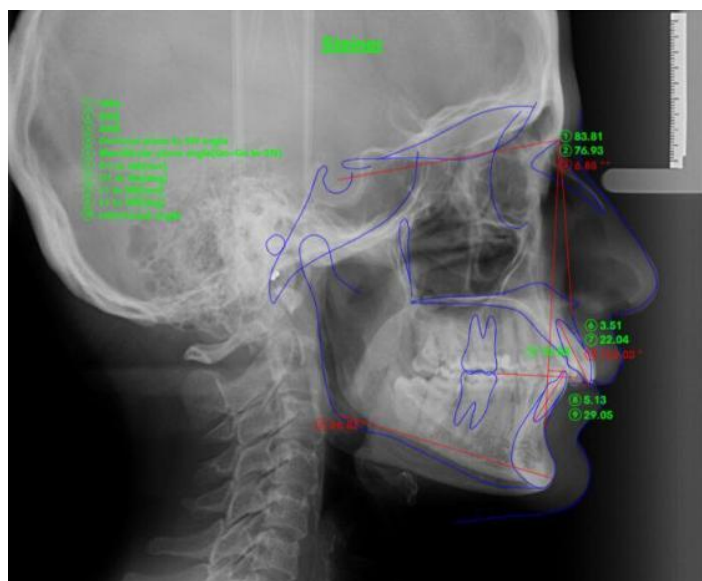


Figure 5: Screenshot of the WEBCEPH analysis system and the readings.

Discussion

This study investigated the accuracy of AI-based Steiner cephalometric analysis by comparing three commercial online systems to manual analysis performed by two experienced orthodontists with extensive expertise in orthodontics and cephalogram tracing. The comparison revealed discrepancies in landmark identification and subsequent angular measurements between the manual tracings and the AI software outputs. These findings highlight inaccuracies within the AI programs, indicating their current limitations in providing reliable cephalometric tracings.

Most of the studies in this regard compared AI analysis with manual or human surveillance analysis to find out the accuracy of such analysis^{11,13,14}. These studies generally employed varying quantities of cephalograms for testing and validating the database, ranging from a dozen to a thousand. Additionally, regarding the clinicians involved in manually identifying landmarks, variability was exhibited in their number and levels of clinical experience in cephalometric tracing. However, this study and most other studies comparing manual analysis with AI digitization analysis shared a common finding, accentuating this innovation's advantages of time saving and accuracy.

Although computer science advancements have facilitated the extensive use of computers in cephalometry, resulting in improved accuracy, manual analysis skills are still necessary^{15,16}. In addition, Cephalometric imaging software, including popular options like Dolphin Imaging, Blue Sky Plan (this study), and QuickCeph, often involves a manual landmark placement process. In the hands of an experienced clinician, this task typically consumes an average of 10 to 15 minutes per case. Despite the clinician's expertise, the manual placement of landmarks is time-consuming and susceptible to errors^{17,18}. This study has found that there are no significant differences between the AI analysis system and the manual system (Table 2), while significant differences between AI systems exist. For example, VOXEL and WEBCEPH analysis showed significant differences in SNA, and with CEPHIO in U1 to NA. This may be due to the differences in landmark identification and reading for the same radiographs. These differences in AI analysis systems for cephalometric analysis come from variations in algorithm types, training data quality, data annotation practices, image preprocessing techniques, training and optimization methods, computational resources, and clinical validation processes. These factors collectively influence the accuracy, reliability,

and generalizability of AI systems in orthodontic diagnostics¹⁹.

VOXEL AI analysis showed statistical insignificance compared with other AI analysis. This may partly be due to the Convolutional Neural Network (CNN). A CNN filter is designed to detect specific features within the input data, producing a corresponding feature map that highlights those characteristics. The parallel processing nature of CNNs enables them to efficiently recognize patterns within an input, whether an image or a data sequence. The initial layers typically learn low-level features such as edges, textures, or colours, while deeper layers progressively learn more complex and abstract features.

This parallel feature learning is a fundamental aspect of CNNs. It contributes to their effectiveness in tasks like image recognition, object detection, and various other applications where extracting hierarchical features from input data is essential. Therefore, an average convolutional layer learns from 32 to 512 filters^{20,21}.

Numerous studies have examined the variability in accuracy among several types of AI algorithms^{19,22-25}. This variability has also been reflected in our research, as there were significant differences between the AI systems for some parameters (SNA, and U1 to NA line). At the same time, there were no significant differences between the two manual analyses and the three AI analysis systems.

The principal finding of this study indicates that the AI analysis method is not inherently unreliable but requires refinement across various stages of machine development and training. The outcomes suggest that areas within the AI-based cephalometric analysis process merit further enhancement to boost its reliability and effectiveness.

In parallel, the study's secondary outcome underscores that manual analysis remains the standard method for cephalometric analysis despite being time and resource consuming. This implies that, currently, the traditional manual approach continues to offer a level of precision and dependability that surpasses the capabilities of the AI-based method. The acknowledgment of manual analysis as the benchmark underscores its continued relevance and effectiveness in the field despite the advancements made in AI technologies.

While the study underscored the superior precision of manual analysis compared to AI-based methods, certain limitations persist. For instance, the study's sample size for manual tracing and AI-based analysis was limited.

Furthermore, only two orthodontists conducted the manual analyses, suggesting a potential for bias or limited generalizability of results. It is recommended that future studies address these limitations by increasing the sample size for both manual and AI analyses and involving a larger number of orthodontists to perform manual analyses, enhancing the robustness and reliability of findings.

Conclusion

Orthodontists should be mindful of the limitations and consider AI as a complementary tool rather than solely relying on automated analyses. AI-based analysis undoubtedly saves time and expedites the process. Nevertheless, orthodontists must review the analysis and occasionally adjust landmarks, lines, and planes based on their expertise and contextual information. AI analysis developers must prioritize this flexibility, allowing orthodontists the option to modify and refine analyses within their platforms. This collaboration between AI technology and orthodontic expertise will ensure accurate and tailored treatment plans for patients.

References

1. Tsolakis IA, Tsolakis AI, Elshebiny T, Matthaios S, Palomo JM. Comparing a Fully Automated Cephalometric Tracing Method to a Manual Tracing Method for Orthodontic Diagnosis. *J Clin Med*. 2022;1:6854-2.
2. Bulatova G, Kusnoto B, Grace V, Tsay TP, Avenetti DM, Sanchez FJC. Assessment of automatic cephalometric landmark identification using artificial intelligence. *Orthod Craniofac Res*. 2021;24 (Suppl 2):37-42.
3. Mohammed MH, Omer ZQ, Aziz BB, Abdulkareem JF, Mahmood TMA, Kareem FA, Mohammad DN. Convolutional Neural Network-Based Deep Learning Methods for Skeletal Growth Prediction in Dental Patients. *J Imaging*. 2024;10(11):278.
4. Rauf AM, Mahmood TMA, Mohammed MH, Omer ZQ, Kareem FA. Orthodontic Implementation of Machine Learning Algorithms for Predicting Some Linear Dental Arch Measurements and Preventing Anterior Segment Malocclusion: A Prospective Study. *Medicina (Kaunas)*. 2023;59(11):1973.
5. Meriç P, Naoumova J. Web-based Fully Automated Cephalometric Analysis: Comparisons between App-aided, Computerized, and Manual Tracings. *Turk J Orthod*. 2020;33(3):142-9.
6. Narkhede S, Rao P, Sawant V, Sachdev SS, Arora S, Pawar AM, Reda R, Testarelli L. Digital versus Manual Tracing in Cephalometric Analysis: A Systematic Review and Meta-Analysis. *J Pers Med*. 2024;14(6):566.
7. Polizzi A, Leonardi R. Automatic cephalometric landmark identification with artificial intelligence: An umbrella review of systematic reviews. *J Dent*. 2024;146:105056.
8. Guinot-Barona C, Alonso Pérez-Barquero J, Galán López L, Barmak AB, Att W, Kois JC, Revilla-León M. Cephalometric analysis performance discrepancy between orthodontists and an artificial intelligence model using lateral cephalometric radiographs. *J Esthet Restor Dent*. 2024;36(4):555-65.
9. Jiang F, Guo Y, Yang C, Zhou Y, Lin Y, Cheng F, et al. Artificial intelligence system for automated landmark localization and analysis of cephalometry. *Dentomaxillofac Radiol*. 2023;52(1):20220081.
10. Wang CW, Huang CT, Lee JH, Li CH, Chang SW, Siao MJ, et al. A benchmark for comparison of dental radiography analysis algorithms. *Med Image Anal*. 2016;31:63-76.
11. Çoban G, Öztürk T, Hashimli N, Yağci A. Comparison between cephalometric measurements using digital manual and web-based artificial intelligence cephalometric tracing software. *Dental Press J Orthod*. 2022;27(4):222112.
12. Abdullah RT, Kuijpers MA, Bergé SJ, Katsaros C. Steiner cephalometric analysis: predicted and actual treatment outcome compared. *Orthod Craniofac Res*. 2006;9(2):77-83.
13. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 2007;39(2):175-91.
14. Jeon S, Lee KC. Comparison of cephalometric measurements between conventional and automatic cephalometric analysis using convolutional neural network. *Prog Orthod*. 2021;22(1):14.
15. Mahto RK, Kafle D, Giri A, Luintel S, Karki A. Evaluation of fully automated cephalometric measurements obtained from web-based artificial intelligence driven platform. *BMC Oral Health*. 2022;22(1):132.
16. Hlongwa P. Cephalometric analysis: manual tracing of a lateral cephalogram. *South African Dent J*. 2019;74(6):318-22.
17. Bao H, Zhang K, Yu C, Li H, Cao D, Shu H, et al. Evaluating the accuracy of automated cephalometric analysis based on artificial intelligence. *BMC Oral Health*. 2023;23(1):191.
18. Wang CW, Huang CT, Hsieh MC, Li CH, Chang SW, Li WC, et al. Evaluation and Comparison of Anatomical Landmark Detection Methods for Cephalometric X-Ray Images: A Grand Challenge. *IEEE Trans Med Imaging*. 2015;34(9):1890-900.
19. El-Fegh I, Galhood M, Sid-Ahmed M, Ahmadi M. Automated 2-D cephalometric analysis of X-ray by

image registration approach based on least square approximator. *Annu Int Conf IEEE Eng Med Biol Soc.* 2008;2008:3949-52.

20. Kiełczykowski M, Kamiński K, Perkowski K, Zadurska M, Czochrowska E. Application of Artificial Intelligence (AI) in a Cephalometric Analysis: A Narrative Review. *Diagnostics* (Basel). 2023;13(16):2640.

21. François-Lavet V, Henderson P, Islam R, Bellemare MG, Pineau J. An introduction to deep reinforcement learning. *Foundations and Trends® in Machine Learning.* 2018;11(3-4):219-354.

22. Brownlee J. Deep learning for computer vision: image classification, object detection, and face recognition in python. *Machine Learning Mastery;* 2019.

23. Leonardi R, Giordano D, Maiorana F. An evaluation of cellular neural networks for the automatic identification of cephalometric landmarks on digital images. *J Biomed Biotechnol.* 2009;2009:717102.

24. Tanikawa C, Yamamoto T, Yagi M, Takada K. Automatic recognition of anatomic features on cephalograms of preadolescent children. *Angle Orthod.* 2010;80(5):812-20.

25. Lindner C, Wang CW, Huang CT, Li CH, Chang SW, Cootes TF. Fully Automatic System for Accurate Localisation and Analysis of Cephalometric Landmarks in Lateral Cephalograms. *Sci Rep.* 2016;6:33581.

26. Park JH, Hwang HW, Moon JH, Yu Y, Kim H, Her SB, Srinivasan G, Aljanabi MNA, Donatelli RE, Lee SJ. Automated identification of cephalometric landmarks: Part 1-Comparisons between the latest deep-learning methods YOLOV3 and SSD. *Angle Orthod.* 2019;89(6):903-9.