RELIABILITY AND ACCURACY OF ARTIFICIAL INTELLIGENCE–BASED SOFTWARE FOR CEPHALOMETRIC DIAGNOSIS: A DIAGNOSTIC STUDY

Murtazaev Saidmurodkhon Saidaloyevich
Doctor of Medical Sciences, Professor

Rakhmonov Ulugbek
Master's student Department of Pediatric Therapeutic Dentistry
Tashkent State Medical University, Tashkent, Republic of Uzbekistan
https://doi.org/10.5281/zenodo.20116221

Abstract:
Artificial intelligence (AI) is transforming cephalometric diagnostics in orthodontics through platforms such as CephX and WebCeph. These AI-based tools utilize machine learning methods and neural networks for automatic identification of cephalometric landmarks and for performing angular and linear measurements, thereby improving diagnostic accuracy, consistency, and efficiency. CephX demonstrated higher reliability and accuracy compared to WebCeph, particularly for FMA and L1–MP measurements, while WebCeph showed greater discrepancies in SNA, ANB, U1–NA, and L1–MP measurements. Both platforms allow manual adjustment of landmarks, which is essential for ensuring clinical accuracy. Key challenges in AI implementation include variability in results across skeletal classes, data privacy concerns, and the need for clinician supervision to verify AI-generated outputs. The expanding use of AI in orthodontics offers opportunities to optimize cephalometric analysis, improve treatment planning, and standardize diagnostic procedures.

Keywords:
Orthodontics; cephalometric analysis; artificial intelligence; CephX; WebCeph; lateral cephalogram; automatic landmark detection; angular measurements; linear measurements; skeletal Class I; skeletal Class II; diagnostic accuracy; digital orthodontics.

Introduction

Orthodontic practice requires comprehensive diagnostics to develop a precise treatment plan. This process typically involves collecting various types of data, including photographs, radiographs, and diagnostic models. Among these, 2D lateral cephalometric radiography plays a key role in cephalometric analysis, which is fundamental for orthodontists [1, 2]. Cephalometric analysis helps to classify dentoalveolar and skeletal malocclusions, determine the severity of orthodontic problems, evaluate craniofacial structures before orthodontic, implant, and surgical interventions, and track both growth and treatment progress [3]. Historically, cephalometric analysis was performed manually using acetate paper and a protractor—a labor-intensive and error-prone process. The advent of digital software simplified this procedure by facilitating the manual placement of reference points and thus increasing efficiency. However, it was found that both manual and digital tracing exhibit similar error levels in landmark placement [8]. With technological advancements, new AI-based software has been developed to further accelerate cephalometric analysis. These programs offer a fully automated system that identifies cephalometric landmarks without operator input, locating all points and performing measurements in just a few seconds [9, 10]. An additional advantage of AI-based systems is their consistency and reliability. While different specialists may interpret cephalometric analysis results differently, AI consistently provides uniform results, thereby enhancing diagnostic reliability. Furthermore, AI systems are trained on large datasets of annotated cephalometric radiographs, which allows the AI

to identify patterns from numerous cases [11-13]. The application of AI has a number of unstudied aspects, particularly concerning its generalizability across different populations, as it may not perform equally well in all age and ethnic groups, given that investment in AI research comes primarily from Asia [12]. Subtle clinical judgment may still be required when managing and interpreting complex cases. Moreover, ethical and legal issues, such as data privacy and the interaction between clinicians, remain areas of active research [12, 14]. Despite these challenges, AI shows great promise for the future, especially if its accuracy continues to meet acceptable standards. Accelerating cephalometric analysis can significantly streamline the treatment planning process for orthodontists [10]. Nevertheless, research on this topic is currently limited, and existing studies are often based on small sample sizes. As AI-based software evolves, further research will be necessary to validate its efficacy and expand its clinical application.

Materials and Methods:

Cephalometric radiographs taken before the start of treatment were selected from the archives of the TSMU orthodontic clinic for the period of 2022 to 2025. All radiographs were obtained using the same apparatus. Inclusion and exclusion criteria were applied to create a more homogeneous sample and minimize errors.

Inclusion criteria:

• Patients with digital images of assured quality that can be integrated into an artificial intelligence system (.DICOM, .JPEG).

• Lateral cephalometric radiograph with a calibration ruler.

• Skeletal Class I or II.

Exclusion criteria:

• Lateral cephalometric radiographs with artifacts, low resolution, motion blur, or areas obscured by foreign objects, etc.

• Patients with unerupted, impacted, or missing incisors.

• Patients with craniofacial anomalies or pronounced skeletal asymmetry.

Out of 60 samples, 30 cephalograms were assigned to skeletal class I, and 30 to skeletal class II. The class I group included 12 males and 18 females, while the class II group included 15 males and 15 females. The age of the study participants ranged from 14 to 26 years. The skeletal class was determined by the value of the patient's ANB angle. The ANB angle is a widely used cephalometric measurement that assesses the anteroposterior (sagittal) relationship between the maxilla and mandible. An ANB value from 0 to 3.9 was classified as skeletal class I, and an ANB value of 4 and above was classified as skeletal class II.

Image acquisition procedure:

Lateral cephalometric radiographs at the TSMU orthodontic clinic are taken with the patient's teeth in maximum intercuspation, with the patient's head aligned to the Frankfort horizontal plane.

Image calibration:

After the images were loaded into each program, they were ready for calibration and tracing. The CephX program allows for image calibration in increments of 5 mm, 10 mm, 15 mm, and 20 mm, while WebCeph allows for calibration in increments of 10 mm, 20 mm, and 30 mm. A distance of 10 mm was selected as the calibration value for images traced in both programs.

Results

Table 1

Lateral Cephalogram Tracing Landmarks
Ruler Point 1 – A point for calibrating the image size. The accuracy of its placement affects the final measurements.
Ruler Point 2 – A point for calibrating the image size. The accuracy of its placement affects the final measurements.

S (Sella) – The center of the sella turcica (pituitary fossa).
N (Nasion) – The most anterior point of the frontonasal suture in the median plane.
Po (Porion) – The most superior and external point of the external auditory meatus.
Or (Orbitale) – The lowest and most anterior point of the orbital margin.
Point A (Subspinale) – The deepest point on the concave profile of the anterior maxilla between the anterior nasal spine and the alveolar crest.
Pg (Pogonion) – The most anterior point of the bony chin in the median plane.
Me (Menton) – The lowest point of the mandibular symphysis in the median plane.
Gn (Gnathion) – The most anterior and inferior point of the chin (midpoint between Pg and Me).
Point B (Supramentale) – The deepest point on the concave profile of the mandible between the chin and the alveolar crest.
Go (Gonion) – The most convex point along the inferior border of the mandibular ramus.
Go' (Constructed Gonion) – The intersection of the mandibular plane and a line bisecting the angle between the Me–Go line and the Ar–mandibular ramus line.
Posterior Ramus Point – The most posterior point on the superior border of the mandibular ramus.
Ar (Articulare) – The point of intersection of the posterior border of the ascending ramus of the mandible and the external border of the posterior cranial base.
Cd (Condylion) – The most posterior and superior point of the condylar head.
U1 Apex (Upper Incisor Apex) – The root apex of the most anterior maxillary central incisor.
U1 Tip (Upper Incisor Tip) – The incisal edge of the most anterior maxillary central incisor.
L1 Tip (Lower Incisor Tip) – The incisal edge of the most anterior central incisor of the mandible.
L1 Apex (Lower Incisor Apex) – The root apex of the most anterior central incisor of the mandible.
ANS (Anterior Nasal Spine) – The anterior nasal spine, its tip on the midline.
PNS (Posterior Nasal Spine) – The posterior nasal spine, its tip on the midline.
Prn (Pronasale) – A point on the anterior contour of the nose.
Ls (Labialis Superior) – The most anterior point on the contour of the upper lip.
Li (Labialis Inferior) – The most anterior point on the contour of the lower lip.
Pog' (Soft Tissue Pogonion) – A point on the anterior contour of the soft tissue chin.

Table 2

Angular Measurements (in degrees)

∠SNA – The angle between the S–N and N–A lines.
∠SNB – The angle between the S–N and N–B lines.
∠ANB – The angle between the N–A and N–B lines.
∠SNMP – The angle between the S–N and Go'–Me lines.
∠FMA – The angle between the Po–Or and Go'–Me lines.
∠U1 to SN – The angle between the axis of the most protrusive maxillary central incisor (the line passing through the incisal edge and the root apex) and the S–N line.
∠L1 to MP – The angle between the axis of the most protrusive mandibular central incisor (the line passing through the incisal edge and the root apex) and the Go'–Me line.

Linear Measurements (in mm)
U1 to NA – The distance between the incisal edge of the most protrusive maxillary central incisor and the N–A line.
L1 to NB – The distance between the incisal edge of the most protrusive mandibular central incisor and the N–B line.
U lip to E-Plane – The distance from the Labialis superior point to the esthetic line (Pronasale – soft tissue Pogonion).
L lip to E-Plane – The distance from the Labialis inferior point to the esthetic line (Pronasale – soft tissue Pogonion).

Summary results for the Class I sample:

The summary results for Class I patients are presented in Table 3. For the SNA measurement, a statistically significant difference was observed between the traditional method and WebCeph (81.75 vs. 83.73), and between CephX and WebCeph (81.51 vs. 83.73), but there was no significant difference between the traditional method and CephX. For the ANB measurement, a statistically significant difference was observed between the traditional method and WebCeph (2.3 vs. 3.65), and between CephX and WebCeph (2.28 vs. 3.65), but there was no significant difference between the traditional method and CephX. For the FMA measurement, a statistically significant difference was observed between the traditional method and CephX (21.34 vs. 19.12), but no significant difference was found between the traditional method and WebCeph or between CephX and WebCeph. For the L1MP measurement, a statistically significant difference was observed between the traditional method and WebCeph (90.93 vs. 95.20), but no significant difference was found between the traditional method and CephX or between CephX and WebCeph. For the U1NA measurement, a statistically significant difference was observed between the traditional method and WebCeph (4.5 vs. 2.76) and between CephX and WebCeph (4.13 vs. 2.76), but there was no significant difference between the traditional method and CephX. For the SNB, SNMP, U1SN, L1NB, ULIPE, and LLIPE measurements, no statistically significant differences were found among the three measurement methods.

Table 3

Variable	Traditional Mean±SD)	CephX Mean±SD)	WebCeph Mean±SD)	P-value
SNA	81.75±3.60a	81.51±3.34ab	83.73 ±3.59c	.003**
SNB	79.45±3.59	79.22±3.46	80.09±3.65	.46
ANB	2.30±0.93a	2.28±1.14ab	3.65±1.15c	<.0001***
SNMP	27.83±5.26	29.41±5.18	28.18±5.44	.30
FMA	21.34±4.31a	19.12±4.39b	19.40±4.92ab	.03*
U1SN	103.08±7.11	103.39±7.18	104.13±6.91	.75
U1NA	4.50±2.42a	4.13±2.31a	2.76±1.97b	.0004*
L1MP	89.47±6.09a	92.14±6.05ab	93.90±6.41b	.002**
L1NB	3.48±1.93	2.94±1.87	3.37±1.98	.34
ULIPE	-3.23±2.46	-3.39±2.46	-3.06±2.30	.79
LLIPE	-2.11±2.64	-2.41±2.60	-2.44±2.44	.78

Summary of results for the Class II sample:

The summary results for Class II patients are presented in Table 4. For the SNA measurement, a statistically significant difference is observed between CephX and WebCeph (82.19 vs. 83.8), but no significant difference exists between the Traditional method and WebCeph, nor between the Traditional method and CephX. For the ANB measurement, a statistically significant difference was found between the Traditional method and WebCeph (5.62 vs. 6.51) and between CephX and WebCeph (5.41 vs. 6.51); however, there was no significant difference between the Traditional method and CephX. For the L1MP measurement, a statistically significant difference exists

between the Traditional method and CephX (92.39 vs. 95.46) and between the Traditional method and WebCeph (92.39 vs. 96.51), but no significant difference was found between CephX and WebCeph. For the SNB, SNMP, FMA, U1SN, U1NA, L1NB, ULIPE, and LLIPE measurements, no statistically significant differences were detected among the three measurement methods.

Table 4

Variable	Traditional (Mean±SD)	CephX Mean±SD)	WebCeph (Mean±SD)	P-alue
SNA	82.60±3.28ab	82.19±3.06a	83.80±3.08b	.03*
SNB	76.99±3.66	76.98±3.92	77.30±3.33	.88
ANB	5.62±1.48a	5.41±1.51ab	6.51±1.44c	.0006**
SNMP	29.99±6.59	32.21±6.67	30.82±7.37	.27
FMA	23.11±5.72	21.33±5.67	21.59±6.66	.29
U1SN	100.00±10.66	100.32±9.00	100.58±9.56	.96
U1NA	3.12±3.41	2.99±3.02	2.80±2.02	.86
L1MP	92.39±5.92a	95.46±5.49b	96.51±6.18b	.002**
L1NB	5.12±2.81	4.91±3.18	4.93±2.52	.92
ULIPE	-1.57±2.61	-1.60±2.63	-1.30±2.51	.81
LLIPE	-0.31±2.84	-0.62±2.95	-0.59±2.67	.83

Results for the entire sample:

Summary results for the entire sample are presented in Table 5. For the SNA parameter, a statistically significant difference was observed between the Traditional method and WebCeph (82.18 vs. 83.77) and between CephX and WebCeph (81.85 vs. 83.77); however, there was no significant difference between the Traditional method and CephX. For the ANB parameter, a statistically significant difference was identified between the Traditional method and WebCeph (3.96 vs. 5.08) and between CephX and WebCeph (3.85 vs. 5.08), but no significant difference was found between the Traditional method and CephX. For the FMA parameter, a statistically significant difference was established between the Traditional method and CephX (22.23 vs. 20.23), but no significant differences were found between the Traditional method and WebCeph or between CephX and WebCeph. For the L1MP parameter, a statistically significant difference was detected between the Traditional method and CephX (90.93 vs. 93.80) and between the Traditional method and WebCeph (90.93 vs. 95.20), but there was no significant difference between CephX and WebCeph. For the U1NA parameter, a statistically significant difference was noted between the Traditional method and WebCeph (3.81 vs. 2.78), but no significant differences were present between the Traditional method and CephX or between CephX and WebCeph. For the SNB, SNMP, U1SN, L1NB, ULIPE, and LLIPE parameters, no statistically significant differences were found among the three measurement methods.

Table 5

Variable	Traditional Mean±SD)	CephX Mean±SD)	WebCeph Mean±SD)	P-value
SNA	82.18±3.45a	81.85±3.20ab	83.77±3.33c	.0001*
SNB	78.22±3.81	78.10±3.85	78.69±3.75	.51
ANB	3.96±2.07a	3.85±2.06ab	5.08±1.93c	<.0001*
SNMP	28.91±6.03	30.81±6.11	29.50±6.58	.09
FMA	22.23±5.12a	20.23±5.17b	20.50±5.93ab	.02*
U1SN	101.54±9.14	101.86±8.24	102.35±8.49	.80
U1NA	3.81±3.02a	3.56±2.74ab	2.78±1.98b	0.02*
L1MP	90.93±6.16a	93.80±5.98bc	95.20±6.40c	<.0001*

L1NB	4.30±2.54	3.93±2.78	4.15±2.39	.59
ULIPE	-2.40±2.66	-2.50±2.69	-2.18±2.56	.68
LLIPE	-1.21±2.87	-1.51±2.91	-1.51±2.70	.68

Conclusion:

Overall, when compared to cephalograms traced by a clinician, the WebCeph software demonstrated the greatest discrepancies for the SNA, ANB, U1 to NA, and L1 to MP points and measurements. For the CephX software, the largest discrepancies were observed for the FMA and L1 to MP indicators. Compared to the Traditional Method, both AI programs showed fewer statistically significant differences when analyzing the Class II sample. Although this aspect has not been fully explored, there may be differences in the reliability and accuracy of anatomical landmark identification between different skeletal structure types when using AI software for tracing. Further research should be conducted to evaluate this finding. Overall, with the exception of FMA and L1 to MP, the CephX software produced results most similar to the cephalograms traced by a clinician using the Traditional Method. Conversely, the WebCeph software demonstrated more issues with accuracy and reliability compared to the Traditional Method. It is important to note that both CephX and WebCeph feature pop-up windows that notify the clinician that the obtained measurements and analyses are for informational purposes only, and that the position of any landmark should be manually checked or adjusted if necessary. Ultimately, this study showed that the CephX software is more reliable and accurate than WebCeph when compared to cephalograms traced by a clinician. It is recommended to use both AI programs with caution, remembering to utilize the semi-automatic correction tool to manually adjust the tracing as needed to obtain the most accurate analysis results.

References

1. Keim, R. G., Gottlieb, E. L., Nelson, A. H., & Vogels, D. S., 3rd (2002). 2002 JCO study of orthodontic diagnosis and treatment procedures. Part 1. Results and trends. *Journal of clinical orthodontics : JCO*, *36*(10), 553–568.
2. Broadbent, B. H., Broadbent, B. H., & Golden, W. H. (1975). *Bolton standards of dentofacial developmental growth*. Mosby.
3. Graber L. *Orthodontics: Current Principles and Techniques*. Philadelphia, PA: Mosby Elsevier; 2012.
4. Kravitz, N. D., Vogels, D., & Vogels, P. (2023). 2023 JCO orthodontic practice study part 3: Practice growth and staff data. *Journal of clinical orthodontics : JCO*, *57*(12), 709–723.
5. Keim, R. G., Gottlieb, E. L., Nelson, A. H., & Vogels, D. S., 3rd (2008). 2008 JCO study of orthodontic diagnosis and treatment procedures, part 1: results and trends. *Journal of clinical orthodontics : JCO*, *42*(11), 625–640.
6. Dreyer, K. J., & Geis, J. R. (2017). When Machines Think: Radiology's Next Frontier. *Radiology*, *285*(3), 713–718. https://doi.org/10.1148/radiol.2017171183
7. OrthodonticPracticeUS. (2024). *A survey of orthodontic practitioners regarding the routine use of lateral cephalometric radiographs in orthodontic treatment*. Orthodontic Practice US.
8. Baig, N., Gyasudeen, K.S., Bhattacharjee, T. *et al.* Comparative evaluation of commercially available AI-based cephalometric tracing programs. *BMC Oral Health* 24, 1241 (2024). https://doi.org/10.1186/s12903-024-05032-9
9. Power, G., Breckon, J., Sherriff, M., & McDonald, F. (2005). Dolphin Imaging Software: an analysis of the accuracy of cephalometric digitization and orthognathic prediction. *International journal of oral and maxillofacial surgery*, *34*(6), 619–626. https://doi.org/10.1016/j.ijom.2005.04.003

10. A. Cohen, H.-S. Ip, and A. Linney, “A preliminary study of computer recognition and identification of skeletal landmarks as a new method of cephalometric analysis,” *British Journal of Orthodontics*, vol. 11, no. 3, 143-154, 1984.
11. R. Lenoardi, D. Giordano, F. Maiorana, and C. Spampinato, “Automatic cephalometric analysis,” *The Angle Orthodontist*, vol. 78, no. 1, pp. 145-151. 2008.
12. A. Richardson, “A comparison of traditional and computerized methods of cephalometric analysis,” *The European Journal of Orthodontics*, vol. 3, no. 1, pp. 15-20, 1981.
13. Selvarajan, Karthikraja & Natrajan, Karthick & Geevee, Ajith & Elango, Elankalai & Sankar, Hanumanth & Krishnan, Prabhakar. (2024). Accuracy and