

Article

Not peer-reviewed version

Precision and Accuracy Assessment of Cephalometric Analyses Performed by Deep Learning Artificial Intelligence with and without Human Augmentation

[Sumer Panesar](#) , Alyssa Zhao , Eric Hollensbe , Ariel Wong , Surya Sruthi Bhamidipalli , [George Eckert](#) , [Vinicius Dutra](#) , [Hakan Turkkahraman](#) *

Posted Date: 6 May 2023

doi: 10.20944/preprints202305.0391.v1

Keywords: artificial intelligence; machine learning; orthodontics; radiology; cephalometry



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Precision and Accuracy Assessment of Cephalometric Analyses Performed by Deep Learning Artificial Intelligence with and without Human Augmentation

Sumer Panesar ¹, Alyssa Zhao ¹, Eric Hollensbe ², Ariel Wong ¹, Surya Sruthi Bhamidipalli ³, George Eckert ³, Vinicius Dutra ⁴ and Hakan Turkkahraman ^{1,*}

¹ Department of Orthodontics and Oral Facial Genetics, Indiana University School of Dentistry, Indiana, University Purdue University at Indianapolis, Indianapolis, IN 46202, USA; spansar@iu.edu (S.P.)

² Private practice, 3415 S Lafountain St #1, Kokomo, IN 46902

³ Department of Biostatistics and Health Data Science, Indiana University School of Medicine, Indianapolis, IN 46202, USA

⁴ Department of Oral Pathology, Medicine, and Radiology, Indiana University School of Dentistry, Indiana, University Purdue University at Indianapolis, Indianapolis, IN 46202, USA

* Correspondence: haturk@iu.edu

Abstract: The aim was to assess the precision and accuracy of cephalometric analyses performed by artificial intelligence (AI) with and without human augmentation. Four dental professionals with varying experience levels and AI identified 31 landmarks on 30 cephalometric radiographs twice. These landmarks were re-identified by all examiners with the aid of AI. Precision and accuracy were assessed by using intraclass correlation coefficients (ICCs) and mean absolute errors (MAEs). AI revealed the highest precision with a mean ICC of 0.97, while the dental student had the lowest (mean ICC: 0.77). AI/human augmentation method significantly improved precision of the orthodontist, resident, dentist, and dental student by 3.26%, 2.17%, 19.75%, and 23.38%, respectively. The orthodontist demonstrated the highest accuracy with a MAE of 1.57 mm/°. AI/human augmentation method improved the accuracy of the orthodontist, resident, dentist, and dental student by 12.74%, 19.10%, 35.69%, and 33.96%, respectively. AI demonstrated excellent precision and good accuracy in automated cephalometric analysis. The precision and accuracy of the examiners with the aid of AI improved by 10.47% and 27.27%, respectively. The AI/human augmentation method significantly improved the precision and accuracy of less experienced dental professionals to that of an experienced orthodontist.

Keywords: artificial intelligence; machine learning; orthodontics; radiology; cephalometry

1. Introduction

Since the development of cephalometric radiography by Broadbent and Hofrath in 1931 [1], a multitude of lateral cephalometric analyses have been developed for the purpose of orthodontic diagnosis and treatment planning. These analyses allow orthodontists to quantify and analyze relationships in the craniofacial region between the skeleton, dentition, and soft tissue. However, an accurate orthodontic diagnosis derived from a cephalometric radiograph requires accurate identification of key landmarks. Current methods of landmark identification include hand tracing with acetate paper or the use of computer software to manually mark digitized cephalometric radiographs, both of which have been shown to be equally reliable [2].

Regardless of the method of cephalometric tracing, manual identification of anatomical structures presents inherent challenges. Factors such as differences in training and radiograph quality can cause variations in intra-examiner and inter-examiner reliability [3]. Additionally, reliability varies depending on the landmark being identified. The landmarks condylion, gnathion, orbitale,

and anterior nasal spine tended to have the lowest inter-examiner reliability among experienced dental radiologists and orthodontists [4]. Computing technology demonstrates the potential to automate this process through artificial intelligence (AI).

AI is defined as the ability of computers or machines to mimic the human decision-making process through a series of algorithms [5]. One area in which AI has been used in orthodontics is automatic lateral cephalometric landmark identification. Pioneered by Cohen in 1984, preliminary attempts to locate landmarks using AI were limited to easily identifiable landmarks such as sella and menton [6]. Since Cohen's attempt to automate landmark identification, several AI methods have been developed to increase accuracy and precision. Success rates of these methods, defined as a deviation of less than 2 mm between the distance measured from the AI-marked landmark and the expert-marked landmark, have varied from 35 to 84.7% depending on the methodology [5]. Convolutional neural networks (CNNs), a type of deep-learning algorithm, have demonstrated some of the highest accuracy in automatic lateral cephalometric landmark identification [7]. The CNN You-Only-Look-Once version 3 (YOLOv3) is a frontrunner in this aspect [8,9]. Although cephalometric landmark identification with AI has improved in the last few decades, a few limitations still exist. Most studies have used far below the recommended 2300 images to train the AI in order to achieve the same level of accuracy as human examiners [10]. Additionally, few studies have used AI trained from multiple types of x-ray machines, which limits real-world clinical application of the software [11]. A study by Silva et al. addressed these limitations by training a CNN named CEFBOT (RadioMemory Ltd, Belo Horizonte, Brazil) with over 250,000 images from various x-ray sources to automatically perform Arnett's analysis measurements [12].

AI-based fully automated decision-making processes put machines in the driver's seat and aim to replace human decisions. However, can machines and humans work together in making better decisions? This new concept is defined as AI/human augmentation, where the goal is to improve human decision-making by machine insights and recommendations. Although previous studies show a promising future for AI-based automatic cephalometric landmark identification and analysis, machines have not yet shown the ability to outperform humans. Therefore, one may theorize that in the realm of cephalometric landmark identification and analysis, the best decisions can be made by the collaboration of AI and humans. However, the literature lacks studies that evaluate how much clinicians with different experience levels benefit from this collaboration. Therefore, the purpose of this study was to assess the precision and accuracy of cephalometric analyses performed by AI alone and AI/human augmentation. We hypothesized that the least experienced clinicians would benefit the most from the AI/human augmentation in cephalometric analyses.

2. Materials and Methods

This study was reviewed and marked as IRB exempt by University Institutional Review Board (Protocol:14334) on March 14, 2022.

2.1. Sample Size Justification

Prior to the start of the study, a power analysis was conducted to determine the sample size. With a sample size of 30 images, the width of the 95% confidence interval for the intraclass correlation coefficients (ICCs) ranged from 0.08 and 0.28 from the estimated ICC, assuming the ICC is between 0.8 and 0.95. In addition, the width of the 95% confidence interval for the percentage of measurements within 1.5 mm of the ideal will range between 20% and 30% from the estimated percentage, assuming the percentage is between 75% and 95%.

2.2. Data Collection

Thirty lateral cephalograms were randomly selected from a pool of adult (age 18 or older) orthodontic patients treated in the School of Dentistry (). Inclusion criteria included patients with class I skeletal relationship (defined as ANB between 0.5 to 5.0°) and Angle class I dental relationship. Patients were excluded if they had craniofacial anomalies, missing or unerupted permanent incisors,

missing or unerupted permanent molars, or poor quality initial cephalometric radiographs. The cephalometric radiographs were saved as JPEG files at a resolution of 200 dpi.

A calibration session led by an orthodontist with 25 years of experience was conducted to clarify the cephalometric landmark identification process and to define each of the landmarks. Four examiners, including an orthodontist with two years of experience, a second-year orthodontics resident, a fourth-year dental student, and a general dentist with one year of experience, each separately and manually identified 31 cephalometric landmarks on the 30 radiographs (Figure 1). The same examiners retraced all radiographs after a period of one week in a randomized order. After completion of the manual tracings, the examiners traced the same cephalometric radiographs with the aid of a deep-learning AI (AI/human augmentation method). For these tracings, AI first located all the landmarks, and the examiners later adjusted points as needed. The AI/human augmentation method tracings were completed again after one week in a randomized order. Additionally, AI alone identified each of the landmarks without any adjustments and repeated the process one week later.

The gold standard for landmark positions were identified with the collaboration of an orthodontist and dental radiologist, each with 25 years of experience within their respective fields. Any disagreements resulted in the point being located at the midpoint both vertically and horizontally of the two points.

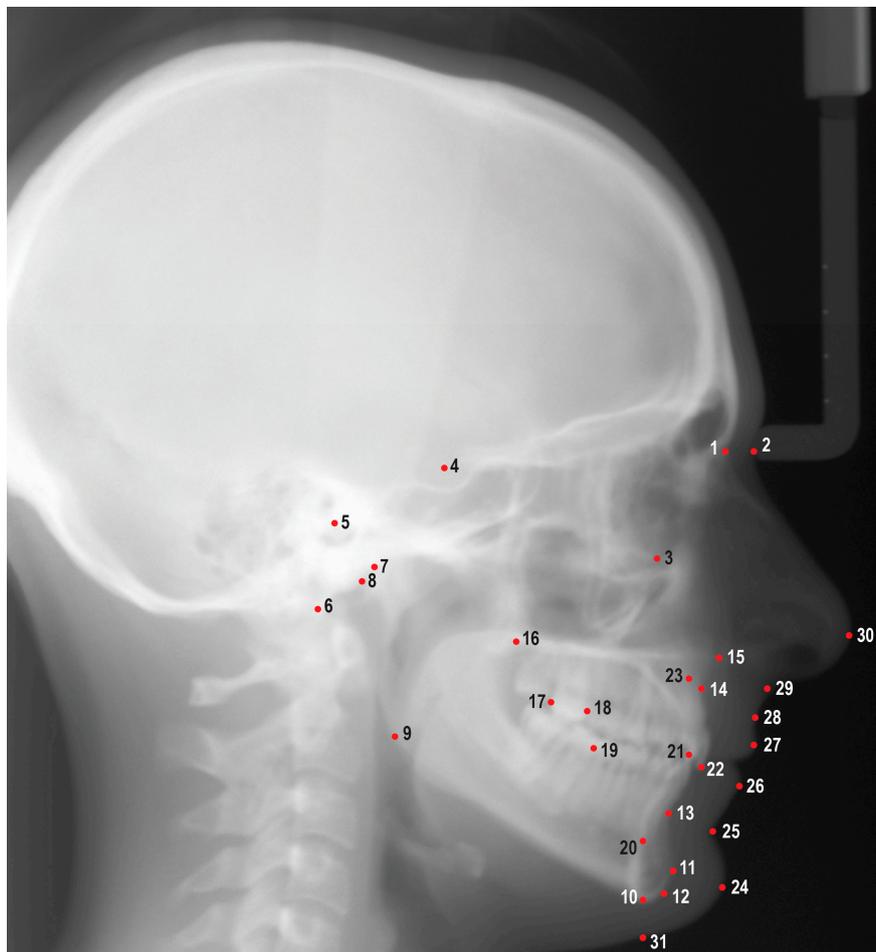


Figure 1. Cephalometric landmarks used in this study. 1: Nasion, 2: Soft tissue nasion, 3: Orbitale, 4: Sella, 5: Porion, 6: Basion, 7: Condylion, 8: Articulare, 9: Gonion, 10: Menton, 11: Pogonion, 12: Gnathion, 13: B point, 14: A point, 15: Anterior nasal spine, 16: Posterior nasal spine, 17: Distal of upper first molar, 18: Mesial of upper first molar, 19: Mesial of lower first molar, 20: Lower incisor root apex, 21: Lower incisor incisal edge, 22: Upper incisor incisal edge, 23: Upper incisor root apex, 24: Soft tissue pogonion, 25: Soft tissue B point, 26: Lower lip, 27: Upper lip, 28: Soft tissue A point, 29: Subnasale, 30: Pronasale, 31: Soft tissue menton.

2.3. Software

RadioCef software (RadioMemory Ltd, Belo Horizonte, Brazil) with CEFBOT module was used for automated cephalometric landmark identification. This software uses four subsystems to locate cephalometric landmarks. Subsystem 1 is a convolutional neural network (CNN) that processes radiographs into probability maps of cephalometric landmark positions. Subsystem 2 is also a CNN that partitions radiographs by vectorizing skeletal and soft tissue borders and estimates the positions of cranial landmarks. Subsystem 3 combines the two CNN subsystems to estimate the position of landmarks, whereas subsystem 4 acts as a quality check to validate the previous results [12]. CEFBOT has been trained with substantial cephalometric data with over 250,000 cephalometric images from multiple x-ray sources and providers [12].

2.4. Data Analyses

Using the traced landmarks, 26 cephalometric measurements were measured for each of the traced radiographs (Supplementary Table S1). Precision was evaluated by comparing the initial tracing to the final tracing completed after a period of one week for each of the examiners, with and without the aid of AI. The accuracy of each of these measurements was evaluated by comparing each examiners' values, with and without the aid of AI, to the gold standard values through mean absolute error (MAE). MAEs were expressed as millimeters or degrees depending on if it was a linear or angular measurement.

2.5. Statistical Analyses

For both precision and accuracy analyses, MAEs with 95% confidence intervals were provided for differences between the gold standard and each initial measurement. For precision analysis, ICCs and Bland Altman plots were used to measure the agreement between initial and final measurements. All the analyses were done using SAS 9.4 (NC, Cary) software.

3. Results

3.1. Precision Analysis

Table 1 shows the ICCs of the cephalometric measurements for each group. Mean ICC values were also calculated for all groups of examiners by averaging the individual cephalometric measurement ICC values. According to the results, AI showed excellent precision (ICCs>0.90). The precision of the examiners without the aid of AI generally followed their level of experience and exposure to the cephalometric analyses. The orthodontist, resident, dentist, and dental student had a mean ICC of 0.92, 0.92, 0.81, and 0.77, respectively. With the aid of AI, the precision of the orthodontist, resident, dentist, and dental student improved by 3.26%, 2.17%, 19.75%, and 23.38%, respectively (Figure 2). Overall, the precision of the examiners improved by 10.47% with the aid of AI. As we expected, the least experienced clinicians benefitted the most from the help of AI.

Table 1. Precision of the examiners with and without the aid of AI given by the intraclass correlation coefficients (ICCs).

Measurements	Orthodontist	Orthodontist	Resident	Resident	Dentist	Dentist	Student	Student	Student	Mean	Mean
		AI		AI		AI		AI	AI	without AI	with AI
SNA (°)	0.92	0.93	0.91	0.93	0.71	0.98	0.39	0.94	0.97	0.73	0.95
SN-Palatal Plane (°)	0.93	0.93	0.89	0.98	0.62	0.96	0.74	0.96	0.92	0.80	0.96
SNB (°)	0.96	0.94	0.92	0.98	0.77	0.98	0.92	0.96	0.99	0.89	0.97
FMA (°)	0.95	0.93	0.94	0.98	0.88	0.97	0.88	0.98	0.99	0.91	0.97
SN-MP (°)	0.99	0.95	0.95	0.99	0.88	0.98	0.95	0.99	1.00	0.94	0.98
Y-Axis (°)	0.97	0.96	0.92	0.98	0.79	0.98	0.93	0.97	0.99	0.90	0.97
ANB (°)	0.81	0.91	0.84	0.73	0.60	1.00	0.13	0.89	0.91	0.60	0.88
ANS-PNS (mm)	0.78	0.92	0.86	0.91	0.58	0.98	0.65	0.93	0.99	0.72	0.94
Co-Gn (mm)	0.95	0.96	0.96	0.99	0.82	0.98	0.88	0.97	1.00	0.90	0.98
Ba-S-N (°)	0.97	0.96	0.94	0.98	0.87	0.94	0.87	0.96	1.00	0.91	0.96
U1-SN (°)	0.95	0.96	0.93	0.95	0.91	0.93	0.90	0.95	0.98	0.92	0.95
U1-NA (°)	0.93	0.96	0.96	0.94	0.92	0.93	0.90	0.95	0.97	0.93	0.95
U1-NA (mm)	0.88	0.92	0.90	0.81	0.71	0.96	0.22	0.91	0.91	0.68	0.90
L1-MP (°)	0.94	0.96	0.95	0.96	0.96	0.98	0.90	0.93	0.99	0.94	0.96
L1-NB (°)	0.92	0.98	0.94	0.94	0.93	0.97	0.84	0.88	0.99	0.91	0.94
L1-NB (mm)	0.94	0.96	0.96	0.98	0.94	1.00	0.92	0.97	0.99	0.94	0.98
Interincisal angle (°)	0.96	0.98	0.96	0.96	0.96	0.96	0.90	0.94	0.99	0.95	0.96
Upper Lip to E-Plane (mm)	0.99	0.99	0.99	0.99	0.99	0.98	0.96	0.98	1.00	0.98	0.99
Lower Lip to E-Plane (mm)	1.00	0.99	0.99	0.99	0.99	0.99	0.97	0.99	1.00	0.99	0.99
N'-Sn' (mm)	0.75	0.93	0.85	0.97	0.67	0.98	0.85	0.97	0.99	0.78	0.96
Sn'-Me' (mm)	0.95	0.97	0.97	0.96	0.95	0.82	0.85	0.94	1.00	0.93	0.92
N-ANS (mm)	0.94	0.92	0.75	0.98	0.56	0.99	0.81	0.94	0.89	0.77	0.96
ANS-Me (mm)	0.98	0.99	0.99	0.99	0.99	1.00	0.98	0.99	0.98	0.99	0.99
Ar-Go (mm)	0.96	0.92	0.93	0.99	0.81	0.97	0.91	0.99	0.98	0.90	0.97

NA-APo (°)	0.83	0.87	0.84	0.70	0.59	1.00	0.13	0.90	0.91	0.60	0.87
FH-NPo (°)	0.81	0.95	0.88	0.95	0.71	0.94	0.65	0.94	0.96	0.76	0.95
Mean	0.92	0.95	0.92	0.94	0.81	0.97	0.77	0.95	0.97	0.86	0.95

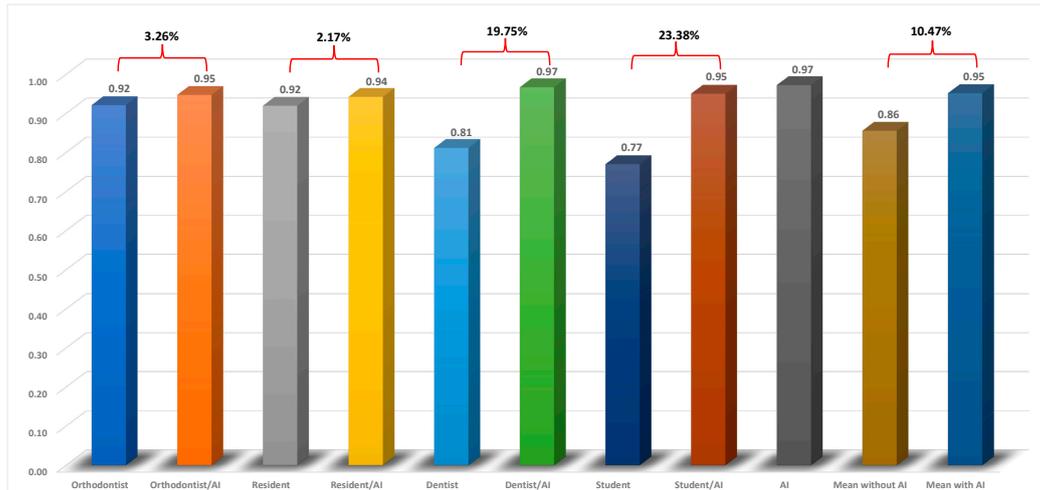


Figure 2. Mean intraclass correlation coefficient for each examiner group and percentages of improvement in precision.

Without the aid of AI, the cephalometric measurements which yielded the moderate precision were maxillary length, U1-NA (mm), SNA, ANB, and convexity angles. With AI/human augmentation method, 23 out of 26 measurements showed excellent precision, while U1-NA (mm), ANB, and convexity angles improved to good precision (Table 1).

3.2. Accuracy Analysis

To compare each examiner's overall accuracy with and without AI's aid, an average of the MAEs was calculated for each group (Table 2). The average MAE of CEFBOT was 1.83 (in mm/°) which was within the clinically acceptable range (MAEs < 2 mm/°). The examiners' accuracy without AI generally followed their level of experience and exposure to the cephalometric analyses. The orthodontist and resident both had MAEs within the clinically acceptable range (1.57 and 1.99 mm/°, respectively) while the accuracy of the dentist and dental student were out of the range with MAEs of 2.55 and 2.68 mm/°, respectively. With the aid of AI, the accuracy of the orthodontist, resident, dentist, and dental student improved by 12.74%, 19.10%, 35.69%, and 33.96%, respectively (Figure 3). AI/human augmentation method revealed clinically acceptable accuracies for all the examiner groups. Although the examiners' accuracy improved by 27.27% with the aid of AI, the least experienced clinicians benefitted the most from the AI/human augmentation method of cephalometric landmark identification.

Table 2. Accuracy of the examiners with and without the aid of AI given by the mean absolute errors (MAEs) from the gold standard.

Measurements	Orthodontist	Orthodontist	Resident	Resident	Dentist	Dentist	Student	Student	AI	Mean	Mean
		AI		AI		AI		AI		without AI	with AI
SNA (°)	1.67	1.20	1.69	1.66	3.24	1.61	3.57	1.77	1.86	2.54	1.56
SN-Palatal Plane (°)	1.50	1.18	1.66	1.10	2.87	1.51	1.76	1.47	1.58	1.95	1.32
SNB (°)	1.17	0.86	1.38	1.22	2.93	1.31	1.42	1.24	1.49	1.73	1.16
FMA (°)	1.41	1.65	3.00	2.91	1.81	1.52	3.05	2.21	3.18	2.32	2.07
SN-MP (°)	1.20	1.36	1.44	2.51	2.98	1.33	2.00	1.62	2.99	1.91	1.71
Y-Axis (°)	0.83	0.77	1.26	0.92	3.29	1.12	1.36	0.94	1.09	1.69	0.94
ANB (°)	0.97	0.71	1.07	0.66	1.21	0.78	2.90	0.91	0.70	1.54	0.77
ANS-PNS (mm)	2.21	1.78	3.44	2.36	3.35	1.88	2.56	1.58	2.06	2.89	1.90
Co-Gn (mm)	2.49	1.66	1.68	1.30	2.99	1.38	2.44	1.48	1.37	2.40	1.46
Ba-S-N (°)	3.03	2.76	3.36	2.38	5.74	3.39	3.89	3.17	1.99	4.01	2.93
U1-SN (°)	2.19	1.84	3.25	1.94	5.78	2.99	4.69	3.48	2.39	3.98	2.56
U1-NA (°)	2.62	1.75	3.24	2.42	4.24	3.17	5.36	3.28	2.52	3.87	2.66
U1-NA (mm)	1.05	0.84	1.35	0.85	1.53	1.04	3.78	1.04	0.80	1.93	0.94
L1-MP (°)	1.96	2.19	2.88	2.60	2.38	2.04	3.18	2.61	3.08	2.60	2.36
L1-NB (°)	1.87	1.86	2.70	1.98	2.28	2.21	4.37	2.56	2.16	2.81	2.15
L1-NB (mm)	0.55	0.46	0.56	0.49	0.52	0.56	0.84	0.58	0.58	0.62	0.52
Interincisal angle (°)	2.60	1.96	4.16	2.86	4.21	3.82	7.12	4.47	2.73	4.52	3.28
Upper Lip to E-Plane (mm)	0.33	0.40	0.43	0.35	0.37	0.34	0.30	0.57	0.85	0.36	0.42
Lower Lip to E-Plane (mm)	0.47	0.47	0.44	0.39	0.45	0.47	0.38	0.49	0.83	0.44	0.46
N'-Sn' (mm)	1.92	1.38	1.48	1.38	2.33	1.75	1.39	1.44	2.60	1.78	1.49
Sn'-Me' (mm)	1.72	1.74	1.81	1.59	1.15	1.64	1.84	1.63	1.81	1.63	1.65
N-ANS (mm)	0.94	1.21	1.73	1.08	2.80	1.45	1.11	1.10	1.49	1.65	1.21
ANS-Me (mm)	1.16	1.32	1.38	1.07	1.60	1.13	1.25	0.82	1.49	1.35	1.09
Ar-Go (mm)	1.76	1.63	1.91	3.10	1.97	1.65	2.77	2.60	3.11	2.10	2.25

NA-APo (°)	1.63	1.31	1.82	1.17	2.45	1.38	4.27	1.42	1.26	2.54	1.32
FH-NPo (°)	1.60	1.31	2.51	1.60	1.83	1.20	2.20	1.47	1.66	2.04	1.40
Mean	1.57	1.37	1.99	1.61	2.55	1.64	2.68	1.77	1.83	2.20	1.60

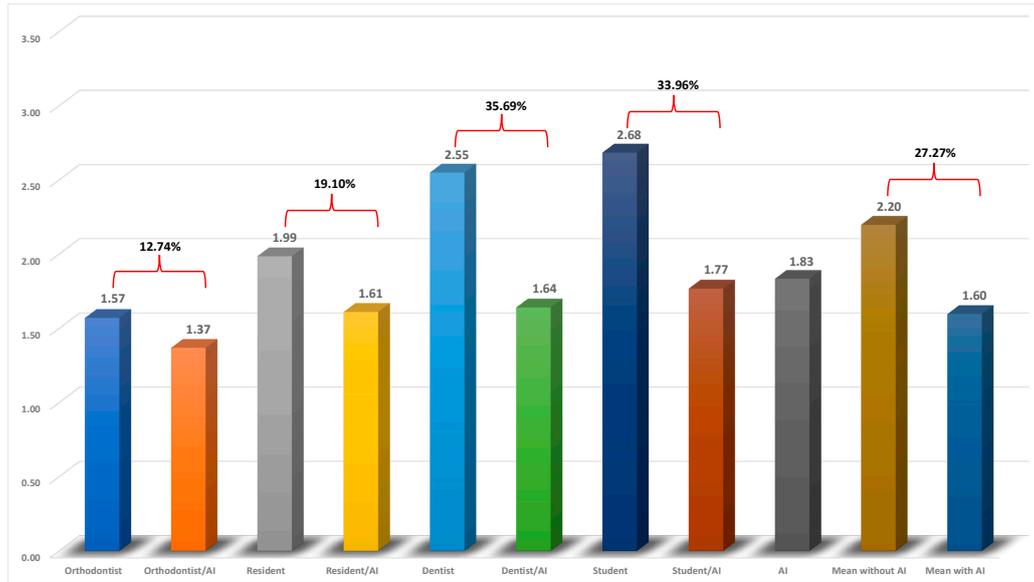


Figure 3. Average mean absolute errors for each examiner group and the percentages of improvement in accuracy.

Without the aid of AI, 13 out of 26 cephalometric measurements (maxillary and mandibular lengths, posterior face height, SNA, FMA, cranial base, U1-SN, L1-MP, L1-NB, interincisal, convexity and facial angles) yielded MAEs out of the clinically acceptable range (Table 2). With AI/human augmentation method, 5 of these measurements (SNA, maxillary and mandibular lengths, convexity and facial angles) improved to be in the clinically acceptable range, while others remained out of the range although they showed some improvement in their accuracies (Table 2).

4. Discussion

The precision of AI alone obtained in this study was similar to the levels of precision reported in other studies [9,12]. Silva and his colleagues found that the ICC of AI ranged from 0.768 to 0.997 for the 10 Arnett's measurements. Excluding the lowest ICC measurement of Frankfort plane to true horizontal line (ICC = 0.768), the remaining ICC values were greater than 0.94, indicating excellent precision [12]. This is similar to our study in that the ICC values of AI traced cephalometric measurements ranged from 0.890 to 1.00 across the 26 cephalometric measurements and demonstrated the greatest precision of all examiners. Hwang had also found that the YOLOv3 deep-learning software had excellent precision and marked each landmark at the exact same position compared to the human inter-examiner reliability of 0.97 mm [9]. The accuracy of AI derived cephalometric measurements differed from Silva's study which found that there were no statistically significant differences in accuracy of 9 of the 10 Arnett's measurements used, and AI was unable to measure the 10th measurement [9]. In this study, 10 out of 26 AI derived measurements revealed accuracies out of the clinically acceptable range. An interesting note was that of the 4 of these measurements involved the gonion landmark (posterior face height, FMA, SN-MP, and L1-MP). Posterior face height and L1-MP were significantly underestimated, whereas SN-MP and FMA were overestimated. These errors were consistent with AI software identifying gonion too superiorly relative to the gold standard, which is a likely source contributing to the mean error. This error was not limited to our study, as Hwang found that the mean error of YOLOv3 was 2.9 mm in the identification of the gonion landmark [9].

Of the five cephalometric measurements that demonstrated moderate precision without the aid of AI, 4 of them (U1-NA distance, SNA, ANB, and convexity angles) involved "Point A" landmark. This suggests "Point A" may pose significant challenges in terms of reproducibility. Possibilities for this include the proximity of this point to the maxillary incisor root, superimposition of nearby structures, blurriness of the curvature in the anterior maxilla, and proximity of the canine fossa.

Subspinale, also known as Point A, has shown good reliability in the x-axis but significant variation in the y-axis [13].

In terms of accuracy averaged among all the examiners, 13 of the 26 measurements demonstrated MAEs greater than 2 mm or degrees without the use of AI. The increased MAEs of FMA and posterior face height may be due to the difficulty of identification of the gonion landmark. Gonion is a constructed point that relies on accurate identification of other landmarks, such as articulare and menton, to create tangent lines to the inferior body of the mandible and posterior border of the ramus. Slight inaccuracies in the tangent lines can dramatically affect the location of gonion. Among the 16 landmarks evaluated by Baumrind and Frantz, Gonion was found to be the least reliable landmark in both the x and y-axes in lateral cephalometric films [14]. Additionally, in the case of FMA, it is possible the orbitale landmark was not accurately identified by all examiners. Orbitale has been shown to have significant variability in lateral cephalometric identification, especially in the y-axis [15]. Inaccuracy of the maxillary length measurement may be due to difficulty of identification of the posterior nasal spine landmark, as it is often superimposed with the 3rd molar germ. Condylion and basion landmarks are also challenging to identify due to the superimposition of other craniofacial structures and thus may explain the higher MAEs of mandibular length and cranial base angle. Basion has shown significant variation in identification in both the x and y-coordinates [15]. The incisor angulation measurements (U1-SN, U1-NA, L1-MP, L1-NB, and interincisal angle) also demonstrated increased MAEs. This could be due to difficulty in the identification of upper and lower incisor apices. Stabrun and Danielson found that upper and lower incisor apices were not reproducibly identified in 75% of cases [16]. Misidentification of the incisor root apices results in alterations of the long axis of the upper and lower incisors, which affects the accuracy of U1-SN, U1-NA, L1-MP, L1-NB, and interincisal angles.

The AI/human augmentation method was a novel aspect of this study. This method demonstrated an increase in precision and accuracy of all examiner groups compared to the manual landmark identification technique, especially with the less experienced student and dentist examiners. This is likely due to the inherent ability of the AI to predictably locate each landmark in a nearly identical position [9]. The less experienced examiners may not have changed the position of landmarks unless there was a significant discrepancy due to a lack of skill in tracing lateral cephalometric radiographs. Most measurements that showed clinical significance overlapped between AI alone and the AI/human augmentation method (posterior face height, FMA, L1-MP, L1-NB, and interincisal angles). This reinforces the idea that the location of difficult to locate landmarks, such as gonion and incisor apices, may not have been modified by examiners.

One limitation of this study is that a single examiner represented an entire dental professional group. Ideally, each experience level would comprise of multiple examiners. Additionally, we were unable to compare the accuracy or precision of individual landmarks. Future studies with multiple examiners and evaluation of the precision and accuracy of individual landmarks are recommended. Despite significant improvements in computing technology and AI in automatic lateral cephalometric landmark identification, AI alone cannot fully replace experienced clinicians. However, AI can serve as an educational tool and aid in landmark identification through the human/AI augmentation method, especially for less experienced clinicians.

5. Conclusions

Deep learning AI software demonstrated excellent precision and good accuracy in automated cephalometric analysis. Overall, precision and accuracy of the examiners with the aid of AI improved by 10.47% and 27.27%, respectively. The AI/human augmentation method significantly improved the precision and accuracy of less experienced dental professionals and students to that of an experienced orthodontist. These groups benefited the most from the approach.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Table S1: Cephalometric measurements and their definitions.

Author Contributions: Conceptualization, Sumer Panesar and Hakan Turkkahraman; Data curation, Sumer Panesar, Alyssa Zhao, Eric Hollensbe, Ariel Wong, Vinicius Dutra and Hakan Turkkahraman; Formal analysis, Sumer Panesar, Surya Bhamidipalli, George Eckert, Vinicius Dutra and Hakan Turkkahraman; Investigation, Sumer Panesar, Vinicius Dutra and Hakan Turkkahraman; Methodology, Sumer Panesar, Alyssa Zhao, Eric Hollensbe, Ariel Wong and Vinicius Dutra; Project administration, Hakan Turkkahraman; Software, Hakan Turkkahraman; Supervision, Hakan Turkkahraman; Validation, Sumer Panesar; Writing – original draft, Sumer Panesar and Hakan Turkkahraman; Writing – review & editing, Sumer Panesar, Ariel Wong, Vinicius Dutra and Hakan Turkkahraman.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was reviewed and marked as IRB exempt by University Institutional Review Board (Protocol:14334) on March 14, 2022.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data underlying this article are available in the article.

Acknowledgments: We would like to thank Craig Eberhardt, EHR systems engineer, for help in filtering patient data.

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

References

- Leonardi, R.; Giordano, D.; Maiorana, F.; Spampinato, C. Automatic cephalometric analysis. *Angle Orthod* **2008**, *78*, 145-151, doi:10.2319/120506-491.1.
- Albarakati, S.F.; Kula, K.S.; Ghoneima, A.A. The reliability and reproducibility of cephalometric measurements: a comparison of conventional and digital methods. *Dentomaxillofac Radiol* **2012**, *41*, 11-17, doi:10.1259/dmfr/37010910.
- Liu, J.K.; Chen, Y.T.; Cheng, K.S. Accuracy of computerized automatic identification of cephalometric landmarks. *Am J Orthod Dentofacial Orthop* **2000**, *118*, 535-540, doi:10.1067/mod.2000.110168.
- Durão, A.P.; Morosolli, A.; Pittayapat, P.; Bolstad, N.; Ferreira, A.P.; Jacobs, R. Cephalometric landmark variability among orthodontists and dentomaxillofacial radiologists: a comparative study. *Imaging Sci Dent* **2015**, *45*, 213-220, doi:10.5624/isd.2015.45.4.213.
- Hung, K.; Montalvao, C.; Tanaka, R.; Kawai, T.; Bornstein, M.M. The use and performance of artificial intelligence applications in dental and maxillofacial radiology: A systematic review. *Dentomaxillofac Radiol* **2020**, *49*, 20190107, doi:10.1259/dmfr.20190107.
- Cohen, A.M.; Ip, H.H.; Linney, A.D. A preliminary study of computer recognition and identification of skeletal landmarks as a new method of cephalometric analysis. *Br J Orthod* **1984**, *11*, 143-154, doi:10.1179/bjo.11.3.143.
- Ren, R.; Luo, H.; Su, C.; Yao, Y.; Liao, W. Machine learning in dental, oral and craniofacial imaging: a review of recent progress. *PeerJ* **2021**, *9*, e11451, doi:10.7717/peerj.11451.
- Park, J.H.; Hwang, H.W.; Moon, J.H.; Yu, Y.; Kim, H.; Her, S.B.; Srinivasan, G.; Aljanabi, M.N.A.; Donatelli, R.E.; Lee, S.J. Automated identification of cephalometric landmarks: Part 1-Comparisons between the latest deep-learning methods YOLOV3 and SSD. *Angle Orthod* **2019**, *89*, 903-909, doi:10.2319/022019-127.1.
- Hwang, H.-W.; Park, J.-H.; Moon, J.-H.; Yu, Y.; Kim, H.; Her, S.-B.; Srinivasan, G.; Aljanabi, M.N.A.; Donatelli, R.E.; Lee, S.-J. Automated Identification of Cephalometric Landmarks: Part 2-Might It Be Better Than human? *The Angle Orthodontist* **2019**, *90*, 69-76, doi:10.2319/022019-129.1.
- Moon, J.H.; Hwang, H.W.; Yu, Y.; Kim, M.G.; Donatelli, R.E.; Lee, S.J. How much deep learning is enough for automatic identification to be reliable? *Angle Orthod* **2020**, *90*, 823-830, doi:10.2319/021920-116.1.
- Kim, J.; Kim, I.; Kim, Y.-J.; Kim, M.; Cho, J.-H.; Hong, M.; Kang, K.-H.; Lim, S.-H.; Kim, S.-J.; Kim, Y.H.; et al. Accuracy of automated identification of lateral cephalometric landmarks using cascade convolutional neural networks on lateral cephalograms from nationwide multi-centres. *Orthodontics & Craniofacial Research* **2021**, *24*, 59-67, doi:https://doi.org/10.1111/ocr.12493.
- Silva, T.P.; Hughes, M.M.; Menezes, L.D.S.; de Melo, M.F.B.; Takeshita, W.M.; Freitas, P.H.L. Artificial intelligence-based cephalometric landmark annotation and measurements according to Arnett's analysis: can we trust a bot to do that? *Dentomaxillofac Radiol* **2022**, 20200548, doi:10.1259/dmfr.20200548.
- Broch, J.; Slagsvold, O.; Røsler, M. Error in landmark identification in lateral radiographic headplates. *European Journal of Orthodontics* **1981**, *3*, 9-13, doi:10.1093/ejo/3.1.9.
- Baumrind, S.; Frantz, R.C. The reliability of head film measurements. 1. Landmark identification. *Am J Orthod* **1971**, *60*, 111-127, doi:10.1016/0002-9416(71)90028-5.

15. Trpkova, B.; Major, P.; Prasad, N.; Nebbe, B. Cephalometric landmarks identification and reproducibility: a meta analysis. *Am J Orthod Dentofacial Orthop* **1997**, *112*, 165-170, doi:10.1016/s0889-5406(97)70242-7.
16. Stabrun, A.E.; Danielsen, K. Precision in cephalometric landmark identification. *Eur J Orthod* **1982**, *4*, 185-196, doi:10.1093/ejo/4.3.185.

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