Title

Comparison of Conventional and Automated Cephalometric Analysis using Cone-Beam Computed Tomography

Author

Paige, Andrew

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Comparison of Conventional and Automated Cephalometric Analysis using Cone-Beam Computed Tomography

by

Andrew Paige

2023

ABSTRACT OF THE THESIS

Comparison of Conventional and Automated Cephalometric Analysis using Cone-Beam Computed Tomography

by

Andrew Paige

Master of Science in Oral Biology

University of California, Los Angeles, 2023

Professor Sanjay M. Mallya, Chair

Since its advent in 1931, analysis of lateral cephalometric radiographs has been an important aspect of orthodontic diagnosis and treatment planning. This type of orthodontic analysis provides a quantitative evaluation of the positions of various anatomical structures of the face, including skeletal, dental, and soft tissue features. Cephalometric analysis, however, is a time-consuming process and is not always performed in clinical practice. Cephalometric analysis from Cone-Beam Computed Tomography (CBCT) imaging is particularly cumbersome, requiring both reconstruction of a two-dimensional lateral cephalogram as well as subsequent computer-assisted tracing analysis using a program such as Dolphin. As the use of CBCT continues to become more prominent in orthodontics, methodologies to expedite cephalometric analysis may be valuable in facilitating their use by clinicians. Several different commercially available (e.g. CephX) and proprietary products have been created to analyze cephalometric radiographs using automated artificial intelligence (AI), but the accuracy of these products remains incompletely established. Our study examined the accuracy and workflow impacts of

CephX's ABO analysis directly from CBCT compared to manual reconstruction and tracing by human examiners on 40 CBCT volumes. We hypothesized that automatically generated cephalometric analyses from CBCT volumes would differ significantly from those generated by human examiners, and that workflow time would be statistically significantly decreased in the automated analysis. In contrast, overall workflow time was found to be significantly greater for the CephX software than for the human examiners. Furthermore, our study showed that seven of the eleven measurements in the ABO analysis differed statistically significantly from the human examiners: SNB, ANB, SN-MP, L1-NB, L1-MP, LL to E-Plane, and UL to E-Plane. Nine of the eleven measurements showed average measurement error within clinically acceptable limits, while two showed greater average error than is clinically acceptable – U1-SN and L1-MP were the least accurate measurement in our study. Bland-Altman plots were constructed showing that FMA may show a slight tendency towards greater accuracy at high values, while LL to E-Plane may display a slight tendency to underestimate more frequently at low values and overestimate more frequently at high values. Only one measurement (ANB) showed limits of agreement within the maximum allowed difference; ten out of the eleven variables exceeded the maximum allowed difference. Overall, CephX may offer clinically acceptable performance for most variables in the ABO analysis, but needs further improvement overall, particularly with the inclination of the upper and lower incisors (U1-SN and L1-MP).

The thesis of Andrew Paige is approved.

Jimmy Kuanghsian Hu

Yong Kim

Sanjay M. Mallya, Committee Chair

University of California, Los Angeles

2023

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Definition
ABO	American Board of Orthodontics
AI	Artificial intelligence
CBCT	Cone-beam computed tomography
DICOM	Digital imaging and communications in medicine
ICC	Intraclass correlation coefficient
LOA	Limits of agreement
MAR	Measurement agreement rate
MARPE	Miniscrew-assisted rapid palatal expansion
TMJ	Temporomandibular joint

ACKNOWLEDGEMENTS

I would like to acknowledge my research mentor, Dr. Sanjay M. Mallya, whose mentorship, guidance, and support has been instrumental in being able to complete this study at UCLA. It would not have been possible to complete this project without the knowledge, experience, and expertise that he brings in the field of radiology as well as in other specialties of dentistry.

Additionally, I would like to express my gratitude to my committee members Dr. Yong Kim and Dr. Jimmy Kuanghsian Hu. Their feedback, along with that of Dr. Mallya, has been invaluable in this endeavor. I am greatly appreciative of their unwavering support for all of us as orthodontic residents as we have managed our clinical training alongside our research and didactic efforts in this program.

I would also like to thank my co-resident, Dr. Marissa Fabros, whose assistance through this project has been vital to its success. I am thankful to her for all of her assistance on this project from its inception to its conclusion. My other co-residents, my family, and my friends have also all played irreplaceable roles in this journey at UCLA.

INTRODUCTION

a. Cephalometric analysis in orthodontics

Since its advent in orthodontics in 1931, cephalometry has played a significant role in both diagnosis and treatment planning in the field of orthodontics. By locating and marking various anatomical landmarks on the skull, multiple linear and angular measurements can be obtained and utilized in a number of different cephalometric analyses[1, 2]. These analyses can be used to measure different maxillofacial structures absolutely as well as to describe the positions of these structures relative to one another. This form of radiography has played a significant role in allowing orthodontic clinicians to better measure tooth and jaw positions, elucidating the jaw discrepancies which underlie many Class II and Class III malocclusions as well as the possibility of altering skeletal growth through treatment[2].

While tracing of a lateral cephalogram may be considered to be part of a standard orthodontic work-up, it is not done universally in clinical practice. In a survey of practitioners in the American Association of Orthodontists, it was found that only 60% of clinicians "always" take pre-treatment lateral cephalograms, and that less than 40% "always" performed cephalometric analysis on pre-treatment radiographs[3]. Cephalometric tracing is a detail-oriented and time-consuming practice, which may help to explain these findings[3,4].

b. Cone-beam computed tomography in orthodontics

As general imaging technology has advanced, so too have the resources available to orthodontists in diagnosing and evaluating patients. Since the introduction of cone-beam computed tomography (CBCT) into dentistry in Europe in 1998, it has emerged as an important imaging modality in the field of orthodontics[5,6].

As a three-dimensional imaging technology, CBCT has a number of benefits over conventional two-dimensional orthodontic imaging modalities, which suffer from magnification, distortion, structural superimposition, and reduced spatial resolution[7,8]. Furthermore, CBCT has been shown to better visualize impacted teeth, root positioning, dentofacial anomalies such as facial asymmetries or cleft palate, and the temporomandibular joints (TMJs)[9,10]. Despite these advantages, however, there are also limitations to CBCT. Increased exposure to ionizing radiation, higher cost, and limited accessibility are some factors which must be considered. As continued technological advancements diminish these limitations, CBCT could replace traditional radiographs in the treatment of all orthodontic patients[2,10].

CBCT also has the advantage of being able to recreate other images such as a panoramic view of the teeth and cephalometric views of the skull[7,10]. Numerous studies have indicated that these cephalometric views reconstructed from CBCT offer clinically acceptable results compared to traditional two-dimensional cephalograms. Landmark identification and cephalometric analysis can then be performed on these reconstructions with similar or even better precision and accuracy than conventional radiographs[11-14]. When a CBCT scan is obtained for a patient, cephalometric analysis can generally be performed without any additional two-dimensional imaging necessary[12].

c. Automated cephalometric analyses in orthodontics

Currently, there are two widely accepted methodologies for cephalometric analysis: manually, tracing and measuring structures using a sheet of acetate over a physical radiograph, and digitally, using manually identified landmarks with automated measurements on a digitized image[15]. While initial tracing modalities were strictly manual, multiple studies have shown computer-assisted tracing applications to be equivalent to manual workflow in terms of landmark and measurement accuracy while reducing time spent by the clinician. These computer-assisted applications for cephalometric analysis such as Dolphin Imaging, CephNinja, and Quick Ceph now make up the majority of cephalometric tracing efforts[3,15-17].

As the advantages of these computer-assisted modalities have become clear, many researchers have shifted towards attempting to create a fully automated system for cephalometric analysis. There are a number of benefits that have been proposed for fully automated cephalometric analyses: improved workflow, reduced tracing time, and reduced intra- and inter-operator inconsistency are some of these anticipated advantages[15,18-24].

Several proprietary systems for automated cephalometric analysis have been developed by a variety of different research teams. Evaluation of the accuracy of these systems has led to a diversity of results. Some researchers have indicated that their automated systems may be comparable to an experienced orthodontist[22-24], while others have found that their systems are not yet clinically equivalent to an experienced practitioner[15,18].

In addition to these proprietary systems, there are also several commercially available automated cephalometric analysis programs such as CephX and Phimentum. Among these, CephX appears to have the strongest literature base with respect to its accuracy and reproducibility[25-28]. Research has again led to a mixture of results, with some studies reporting clinically acceptable performance[25,27], while others have found that the software is not yet reliable enough[26]. Measurements from CephX have been found to be reproducible relative to other computer-assisted tracing applications[28].

While there remains no clear consensus on the accuracy of fully automated cephalometric analysis, it is clear that machine learning technology is evolving rapidly in healthcare[22]. As such, some programs such as CephX now offer the ability to obtain cephalometric analysis directly from CBCT volumes[29]. While there has been a significant amount of research focused on automated analysis of conventional two-dimensional cephalograms, the accuracy of CephX's automated cephalometric analysis directly from CBCT has not yet been investigated. Some researchers have developed proprietary models for cephalometric analysis directly from CBCT, and these models appears promising yet still not sufficiently equivalent in tracing accuracy to an orthodontic clinician[4,30]. As CephX's commercially available product has not yet been publicly evaluated, research assessing the accuracy of its cephalometric analysis directly from CBCT is warranted.

OBJECTIVE AND SPECIFIC AIMS

OBJECTIVE: The goal of this study is to assess the feasibility of CephX's automatically generated cephalometric analysis from three-dimensional cone-beam computed tomography (CBCT) volumes as a replacement for computer-assisted cephalometric reconstruction and tracing.

SPECIFIC AIMS:

(1) To compare the accuracy of automatically generated cephalometric analyses from CBCT to conventional computer-assisted tracings using lateral cephalograms reconstructed from CBCT volumes.

   a. Evaluate the accuracy of automatic cephalometric analysis compared to conventional computer-assisted tracing and analysis (e.g. Dolphin Imaging).

   b. Statistical analysis of data to detect significant differences in measurements between these two modalities.

(2) To compare the impact on workflow of using automated cephalometric analyses directly from CBCT as a replacement for conventional modalities for cephalometric analysis.

   a. Evaluate the time spent per cephalometric analysis between fully automated and computer-assisted tracing conditions.

   b. Statistical analysis to detect significant difference in workflow time between these two modalities.

DESIGN AND METHODOLOGY

a. Experimental design

Patient selection: Pre-treatment CBCT records from 40 patients (based on a previously calculated minimum of 35 patients) of any age and sex seen by the University of California, Los Angeles (UCLA) Section of Orthodontics were randomly selected for this study[25]. The following inclusion and exclusion criteria were used to define the study population:

> Inclusion criteria: Patient seen by the UCLA Section of Orthodontics, pre-treatment CBCT taken at UCLA, patient in permanent dentition.
>
> Exclusion criteria: Pre-treatment CBCT not taken at UCLA, gross asymmetry present, lack of image resolution or poor image quality, presence of craniofacial deformity, obfuscation of critical dental landmarks, broad prosthetic restorations or multiple missing teeth.

Study protocol: Pre-treatment CBCT volumes from each patient were analyzed using both computer-assisted and automated cephalometric analysis procedures.

> Automated cephalometric analysis: Three-dimensional DICOM files from patient CBCT volumes were deidentified and uploaded directly to CephX. The software automatically converted these files to two-dimensional lateral cephalometric radiographs, which were subsequently traced and analyzed using its artificial intelligence algorithm. A subset of

ten CBCT volumes were uploaded and analyzed twice by the CephX software to determine intra-rater reliability.

Manual computer-assisted cephalometric analysis: Three-dimensional DICOM files from patient CBCT volumes were deidentified and uploaded into Dolphin. Reconstructed two-dimensional lateral cephalograms were produced by the two examiners (trained orthodontic residents at UCLA) using the software's "Build X-Rays" functionality. Each reconstructed lateral cephalogram was manually traced in Dolphin independently by the same examiners. If discrepancies existed between right- and left-sided structures, the midpoint of the two structures was chosen. Furthermore, a subset of ten radiographs was traced twice by each examiner greater than one week apart in order to determine intra-examiner reliability.

b. Data collection

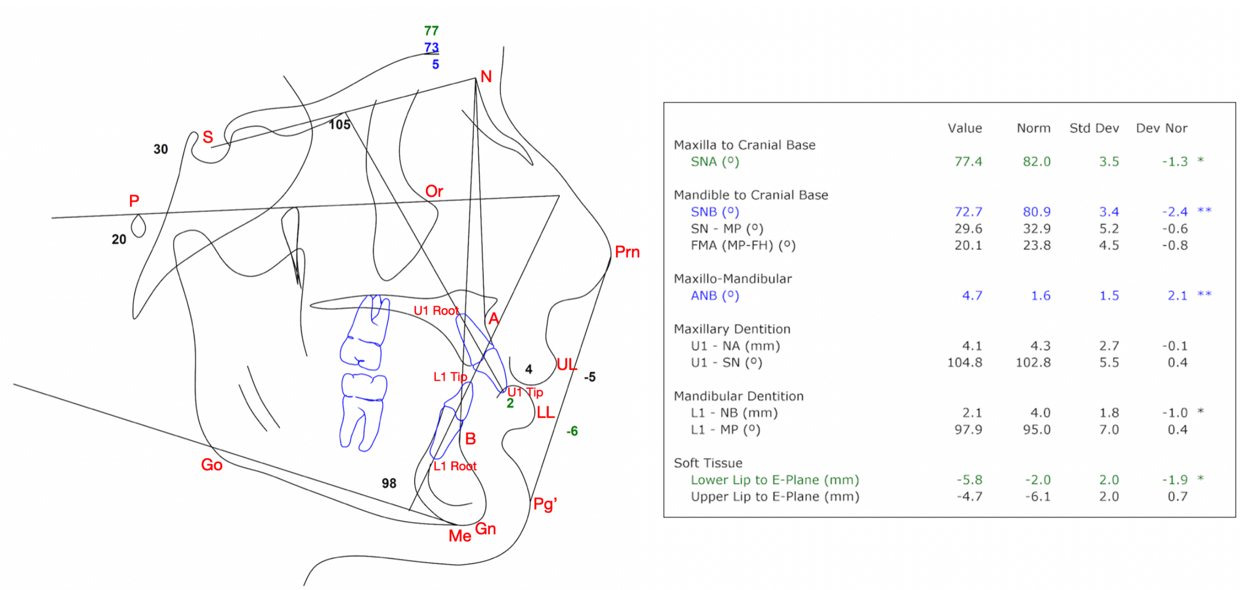
Direct measurements: From each automated and computer-assisted tracing analysis, we collected the following skeletal, dental, and soft tissue measurements used in the ABO cephalometric analysis: SNA (°), SNB (°), SN-MP (°), FMA (°), ANB (°), U1-NA (mm), U1-SN (°), L1-NB (mm), L1-MP (°), Lower Lip to E-Plane (mm), and Upper Lip to E-Plane (mm). The constellation of these eleven values involves identification of seventeen different anatomical landmarks: Sella, Nasion, A Point, B Point, Porion, Orbitale, Gonion, Menton, U1 Tip, U1 Root, L1 Tip, L1 Root, Tip of Nose, Soft Tissue Pogonion, Upper Lip, and Lower Lip. Further information about the ABO analysis, including definitions of its landmarks and measurements, can be found in Tables 1 and 2, respectively. A sample ABO analysis produced from Dolphin and CephX can be seen in Figures 1 and 2. Additionally, we measured the workflow time for each modality of cephalometric analysis. Workflow time was measured using a stopwatch and defined as the time elapsed from initiating upload of the DICOM files into the respective software to the time when cephalometric measurements (e.g. SNA) first became available.

Table 1. Definition of cephalometric landmarks used in the ABO analysis[31]

Landmark	Definition
Skeletal	
Sella (S)	The midpoint of the cavity of sella turcica
Nasion (N)	The anterior point of the intersection between the nasal and frontal bones
Porion (P)	The midpoint of the upper contour of the external auditory canal
Orbitale (Or)	The lowest point on the inferior margin of the orbit
Point A (A)	The innermost point on the contour of the premaxilla between anterior nasal spine and the incisor tooth
Point B (B)	The innermost point on the contour of the mandible between the incisor tooth and the bony chin
Gonion (Go)	The most posterior and inferior point at the angle of the mandible
Gnathion (Gn)	The most anterior and inferior point on the contour of the chin
Menton (Me)	The most inferior point on the contour of the chin
Dental	
U1 tip	The point of tip of upper central incisor
U1 root	The point of root of upper central incisor
L1 tip	The point of tip of lower central incisor
L1 root	The point of root of lower central incisor
Soft Tissue	
Pronasale (Prn)	The most prominent point of apex nasi
Upper lip (UL)	The most prominent point of the border of the upper lip
Lower lip (LL)	The most prominent point of the border of the lower lip
Soft tissue pogonion (Pg’)	The most anterior soft tissue point of the chin in the midsagittal plane

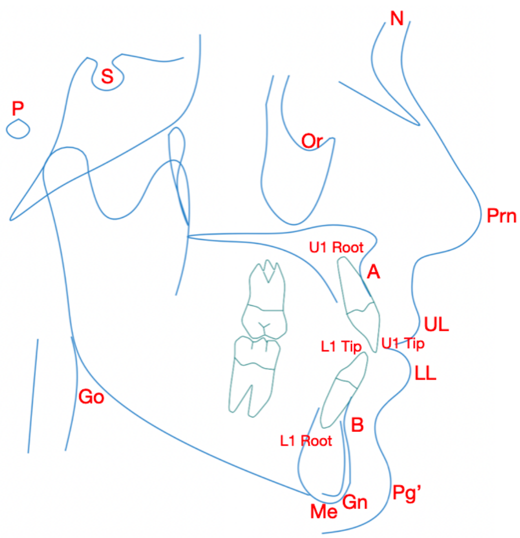
Table 2. Definition of cephalometric measurements used in the ABO analysis[31]

Measurement	Definition
Skeletal	
SNA (°)	The angle formed between points S, N, and A
SNB (°)	The angle formed between points S, N, and B
ANB (°)	The angle formed between points A, N, and B
SN-MP (°)	The angle formed by mandibular plane (Go-Me) and line S-N
FMA (MP-FH) (°)	The angle formed by mandibular plane (Go-Me) and Frankfort horizontal plane (P-Or)
Dental	
U1-NA (mm)	The distance between point U1 tip and line N-A
U1-SN (°)	The angle formed by upper incisor axes and line N-A
L1-NB (mm)	The distance between point L1 tip and line N-B
L1-MP (°)	The angle formed by lower incisor axes and line N-B
Soft Tissue	
LL/E-Plane (mm)	The distance between point LL and E plane (line Prn-Pg')
UL/E-Plane (mm)	The distance between point UL and E plane (line Prn-Pg')



	Value	Norm	Std Dev	Dev Nor	
Maxilla to Cranial Base					
SNA (º)	77.4	82.0	3.5	-1.3	*
Mandible to Cranial Base					
SNB (º)	72.7	80.9	3.4	-2.4	**
SN - MP (º)	29.6	32.9	5.2	-0.6	
FMA (MP-FH) (º)	20.1	23.8	4.5	-0.8	
Maxillo-Mandibular					
ANB (º)	4.7	1.6	1.5	2.1	**
Maxillary Dentition					
U1 - NA (mm)	4.1	4.3	2.7	-0.1	
U1 - SN (º)	104.8	102.8	5.5	0.4	
Mandibular Dentition					
L1 - NB (mm)	2.1	4.0	1.8	-1.0	*
L1 - MP (º)	97.9	95.0	7.0	0.4	
Soft Tissue					
Lower Lip to E-Plane (mm)	-5.8	-2.0	2.0	-1.9	*
Upper Lip to E-Plane (mm)	-4.7	-6.1	2.0	0.7	

Figure 1. Sample ABO cephalometric tracing (left) and analysis (right) generated using Dolphin Imaging software.



Descriptor	Meas.	Type	Mean	Sd	Patient	Graph	Comment
SNA		Deg	82.0	2.0	83.1	-(*)+	
SNB		Deg	80.0	2.0	81.52	-(*)+	
SN-MP		Deg	33.0	4.0	29.9	-(*)+	
FMA		Deg	28.0	3.0	22.11	-(*|)+	Low Mandibular Plane
ANB		Deg	2.0	2.0	1.58	-(*)+	
U/1 -NA	mm	mm	4.0	2.0	5.04	-(*)+	
U/1 to SN		Deg	103.0	4.0	107.09	-(|*)+	
L/1 -NB	mm	mm	4.0	2.0	3.88	-(*)+	
L1-MP		Deg	92.0	3.4	87.53	-(*|)+	
LOWER LIP to E-LINE		mm	-2.0	2.0	-3.95	-(*)+	
UPPER LIP to E-LINE		mm	-4.0	2.0	-5.99	-(*)+	

Figure 2. Sample ABO cephalometric tracing (left) and analysis (right) generated using CephX software.

c. Statistical Analysis

Evaluation of the accuracy of automated cephalometric analysis: We used the measurements obtained from the CephX and Dolphin ABO analyses from each examiner to determine the mean and standard deviation for all examiner and CephX measurements. Intra-class correlation coefficients (ICCs) were calculated to determine the intra-rater reliability for both the human examiners and the CephX software. The measurement error for each variable was calculated between the human examiners to establish reliability between the raters, and mean values were calculated between the examiners to serve as the true value for each patient. After calculating the true values, each measurement from CephX was subjected to paired t-test with equal variance analysis to determine any differences between the software and the human examiners for each variable. The average measurement error was calculated, and Bland-Altman plots were constructed to better characterize the strength and direction of any measurement bias in the CephX analysis.

Assessment of workflow impact: Measurements were obtained from each examiner and from the CephX software, which were used to determine the mean, minimum and maximum, and standard deviation of workflow time. A t-test was used for statistical comparison of workflow times between the

RESULTS

a. Characteristics of the study sample

The study sample included 40 patients seen in the UCLA Section of Orthodontics. With respect to patient sex, 15 (37.5%) were male and 25 (62.5%) were female. The mean age at initial CBCT was 21y10m. Regarding patient ethnicity, 3 (7.5%) were African American, 6 (15.0%) were Asian, 11 (27.5%) were Caucasian, 18 (45.0%) were Hispanic, and 2 (5.0%) were Middle Eastern. The most common indication for CBCT in our study was for surgical planning (55.0%), followed by TMJ concerns (25%), planning of miniscrew-assisted rapid palatal expansion (MARPE) (17.5%), and evaluation of a missing tooth (2.5%).

Table 3. Characteristics of the study sample

Variable	Result
Total Subjects	40
Sex	
Male	15 (37.5%)
Female	25 (62.5%)
Age at Initial CBCT	
Mean	21y10m
Low	12y3m
High	61y5m
Range	49y2m
Ethnicity	
African American	3 (7.5%)
Asian	6 (15.0%)
Caucasian	11 (27.5%)
Hispanic	18 (45.0%)
Middle Eastern	2 (5.0%)
Indication for CBCT	
MARPE	7 (17.5%)
Missing Tooth	1 (2.5%)
Surgical Planning	22 (55.0%)
TMJ Concerns	10 (25.0%)

b. Assessment of workflow impact

The time elapsed from initiating upload of the DICOM files into the respective software to the time when cephalometric measurements first became available was calculated for all cases for both the human examiners and the AI software. This information is detailed in Table 2, which shows that the mean workflow time for the human examiners was 5 minutes and 28 seconds, with a standard deviation of 52 seconds. The mean workflow time for the CephX software was 28 minutes and 53 seconds, with a standard deviation of 4 minutes. The mean workflow time for the CephX software was significantly greater than the human examiners ($p<.001$), and even the minimum AI workflow time took longer than the maximum time spent by the human examiners.

Table 4. Comparison of workflow time between human examiners and CephX

	Human Examiners			**CephX**			**Significance**
	Min	Max	Mean ± SD	Min	Max	Mean ± SD	p-value
Duration (seconds)	252	464	328 ± 52	1378	2260	1733 ± 240	<.001
Duration (minutes)	4 min 12 sec	7 min 44 sec	5 min 28 sec ± 52 sec	22 min 58 sec	37 min 40 sec	28 min 53 sec ± 4 min	

b. Intra- and inter-rater reliability for human examiners and CephX

Intra-rater reliability was determined using the intra-class correlation coefficient (ICC) for each of the human examiners and for the CephX software. The ICC values and 95% confidence intervals are detailed in Table 3. Both human examiners showed excellent intra-rater reliability, with ICC values greater than 0.9 for all measurements. The lowest ICC values for the human examiners were found with U1-NA measurements, showing ICC values of .983 and .982 for the examiners. The CephX software showed perfect ICC values, with 100% measurement agreement between repeated analyses of the same input data (i.e. the same CBCT volumes).

Table 5. Intra-class correlation coefficients for human examiners and CephX

	Examiner 1		Examiner 2		CephX	
	ICC	95%	ICC	95%	ICC	95%
Skeletal						
SNA (°)	.993	.972-.998	.978	.872-.995	1.000	-
SNB (°)	.998	.991-.999	.995	.942-.999	1.000	-
SN-MP (°)	.998	.992-.1.000	.996	.918-.999	1.000	-
FMA (MP-FH) (°)	.998	.914-1.000	.998	.991-.999	1.000	-
ANB (°)	.998	.993-1.000	.996	.981-.999	1.000	-
Dental						
U1-NA (mm)	.983	.936-.996	.982	.933-.996	1.000	-
U1-SN (°)	.989	.957-.997	.992	.966-.998	1.000	-
L1-NB (mm)	.998	.990-1.000	.997	.989-.999	1.000	-
L1-MP (°)	.997	.989-.999	.998	.991-.999	1.000	-
Soft Tissue						
LL/E-Plane (mm)	.998	.993-1.000	.997	.986-.999	1.000	-
UL/E-Plane (mm)	.998	.993-1.000	.998	.990-.999	1.000	-

Inter-rater reliability was established by calculating the mean absolute differences between measurements made by each examiner for each variable, and these findings are shown in Table 4. Measurement differences of up to 2 measurement units (mm or degree) are generally considered to be clinically acceptable – the average error values for each of the eleven measurements made by the human examiners fell below 1 measurement unit[28]. The lowest variability between examiners was seen with soft tissue measurements (LL to E-Plane and UL to E-Plane), L1-NB, and ANB. The greatest variability between examiners was seen with SNA, SN-MP, U1-SN, and L1-MP, though all remained below a difference of 1 measurement unit.

For each variable, the measurement agreement rate (MAR) was defined and calculated as the percentage of time in which the error was less than a specified number of measurement units. All linear measurements in the study showed 100% MAR within 2 mm between the examiners, and all angular measurements showed 90% or greater MAR within 2 degrees. Altogether, the examiners displayed high inter-rater reliability, with the mean error for all eleven variables falling below 1 unit, and all showing 90% or greater MAR within 2 units.

Table 6. Mean measurement error between human examiners

	Absolute Difference	Measurement Agreement Rate (%)			
	Mean Error ± SD	<1 unit	<2 units	<3 units	<4 units
Skeletal					
SNA (°)	.84 ± .54	60.0	92.5	100.0	100.0
SNB (°)	.54 ± .38	87.5	100.0	100.0	100.0
ANB (°)	.42 ± .31	95.0	100.0	100.0	100.0
SN-MP (°)	.81 ± .76	70.0	90.0	97.5	100.0
FMA (MP-FH) (°)	.75 ± .75	70.0	97.5	97.5	97.5
Dental					
U1-NA (mm)	.59 ± .43	82.5	100.0	100.0	100.0
U1-SN (°)	.93 ± .66	62.5	95.0	100.0	100.0
L1-NB (mm)	.39 ± .32	95.0	100.0	100.0	100.0
L1-MP (°)	.85 ± .68	60.0	92.5	100.0	100.0
Soft Tissue					
LL/E-Plane (mm)	.36 ± .21	100.0	100.0	100.0	100.0
UL/E-Plane (mm)	.34 ± .30	97.5	100.0	100.0	100.0

c. Accuracy of CephX compared to human examiners

In order to evaluate the accuracy of CephX, mean measurement values and standard deviations were calculated for both the software and the human examiners. A paired t-test was used to evaluate these values for any statistically significant differences between the AI and the human examiners. There was a statistically significant difference between CephX and the human examiners in seven of the eleven measurements: SNB, ANB, SN-MP, L1-NB, L1-MP, LL to E-Plane, and UL to E-Plane ($p<.05$). There was no statistically significant difference found between CephX and the human examiners in values for SNA, FMA, U1-NA, and U1-SN. The mean values for each measurement and the differences between the tracing modalities are shown in further details in Table 5.

Using the mean measurement values between the human examiners as the true value for each patient, the absolute error between CephX and true value was calculated for each measurement for each patient. The results of this analysis are shown in Table 6. When comparing CephX and human examiners, the average error was below 1 measurement unit for only two variables: ANB and UL to E-Plane. The average error was within the clinically acceptable 2 unit range for nine of the eleven variables -- SNA, SNB, SN-MP, FMA, U1-NA, L1-NB, and LL to E-Plane. U1-SN and L1-MP showed a mean error greater than 2 degrees, with L1-MP showing a difference of 4.05 ± 2.89 degrees. This was the lowest agreement seen between the CephX software and the true value from the human examiners.

As was done between the human examiners, the MAR was calculated for each of the measurements made by CephX. The most frequent agreement within the clinically accepted 2 units was seen with ANB and UL to E-Plane, the two measurements that also held the lowest mean error. U1-SN and L1-MP, the two measurements with the highest mean error, showed the least frequent agreement within 2 units at 42.5% and 30% of the time, respectively. SN-MP, FMA, U1-SN, and L1-MP all showed agreement within 2 measurement units less than 70% of the time. L1-MP showed the most measurement error overall, with greater than 4 mm error 50% of the time, and clinically unacceptable error (greater than 2 degrees) 70% of the time. These values are also shown in more details in Table 6.

Table 7. Sample means and paired t-tests for comparison between conventional and AI methods

	Human Examiners	**CephX**	**Paired Differences**	**Significance**
	Mean ± SD	Mean ± SD	Mean ± SD	p-value
Skeletal				
SNA (°)	81.9 ± 3.9	82.2 ± 3.7	.31 ± 1.58	.219
SNB (°)	79.7 ± 5.9	80.3 ± 5.6	.58 ± 1.16	.003
ANB (°)	2.2 ± 4.1	1.9 ± 4.2	-0.27 ± .78	.033
SN-MP (°)	32.3 ± 6.5	34.1 ± 7.0	1.70 ± 1.90	<.001
FMA (MP-FH) (°)	25.3 ± 6.0	25.0 ± 5.6	-0.33 ± 2.65	.436
Dental				
U1-NA (mm)	5.2 ± 2.6	5.5 ± 2.9	.24 ± 1.55	.326
U1-SN (°)	106.7 ± 8.6	106.6 ± 7.7	-0.11 ± 2.88	.814
L1-NB (mm)	6.3 ± 3.5	5.3 ± 3.5	-1.02 ± 1.19	<.001
L1-MP (°)	90.1 ± 9.4	86.7 ± 9.2	-3.39 ± 3.66	<.001
Soft Tissue				
LL/E-Plane (mm)	-0.4 ± 3.2	-1.4 ± 3.6	-.92 ± 1.53	<.001
UL/E-Plane (mm)	-3.8 ± 3.5	-3.4 ± 3.4	.35 ± .90	.018

Table 8. Mean measurement error for CephX

	Absolute Difference	**Measurement Agreement Rate (%)**			
	Mean Error ± SD	<1 unit	<2 units	<3 units	<4 units
Skeletal					
SNA (°)	1.34 ± .86	35.0	75.0	92.5	100.0
SNB (°)	1.05 ± .75	47.5	87.5	97.5	100.0
ANB (°)	.64 ± .51	80.0	95.0	100.0	100.0
SN-MP (°)	1.95 ± 1.63	32.5	62.5	80.0	90.0
FMA (MP-FH) (°)	1.86 ± 1.90	47.5	67.5	80.0	85.0
Dental					
U1-NA (mm)	1.22 ± .96	50.0	80.0	95.0	97.5
U1-SN (°)	2.35 ± 1.64	27.5	42.5	57.5	80.0
L1-NB (mm)	1.37 ± .74	32.5	82.5	97.5	100.0
L1-MP (°)	4.05 ± 2.89	15.0	30.0	45.0	50.0
Soft Tissue					
LL/E-Plane (mm)	1.51 ± .93	30.0	77.5	92.5	100.0
UL/E-Plane (mm)	.75 ± .60	67.5	95.0	100.0	100.0

In order to better characterize the strength and direction of the measurement bias shown by CephX, Bland-Altman plots were created for all skeletal, dental and soft tissue measurements. Figure 3 shows the Bland-Altman plots for the five skeletal measurements and contains the average bias for each. On average, CephX overestimated the values for SNA, SNB, and SN-MP for 0.31°, 0.58°, and 1.70°, respectively. CephX underestimated the values for ANB and FMA by 0.27° and 0.33°, respectively. For SNA, SNB, and ANB, the strength and direction of the measurement bias appears to be consistent across both high and low mean measurement values. FMA may show a slight tendency towards greater accuracy at higher measurement values. The limits of agreement for skeletal measurements made by CephX were within the maximum allowed difference of +/- 2 measurement units for only one variable: ANB.

Figure 4 shows the Bland-Altman plots for the four dental measurements and contains the average bias for each value. On average, CephX overestimated only the measurements for U1-NA, by a mean amount of 0.24 mm. U1-SN (0.11°), L1-NB (1.02 mm), and L1-MP (3.39°) were all underestimated on average by the software. All four of these dental measurements appeared to show consistent strength and direction of measurement bias across both high and low mean values, and none showed limits of agreement within the maximum allowed difference.

Figure 5 depicts the Bland-Altman plots for the two soft tissue measurements in our study. On average, CephX tended to underestimate the values for LL to E-Plane (0.92°), while overestimating the values for UL to E-Plane (0.35°). UL to E-Plane appears to show consistent strength and direction of bias across high and low mean values, while LL to E-Plane may show a slight tendency to underestimate more frequently at low values and overestimate more frequently at high values. No soft tissue measurements showed displayed limits of agreement within the maximum allowed difference. In aggregate, CephX showed no overall predilection for under- or overestimation, as it overestimated five out of eleven and underestimated six out of eleven measurements.

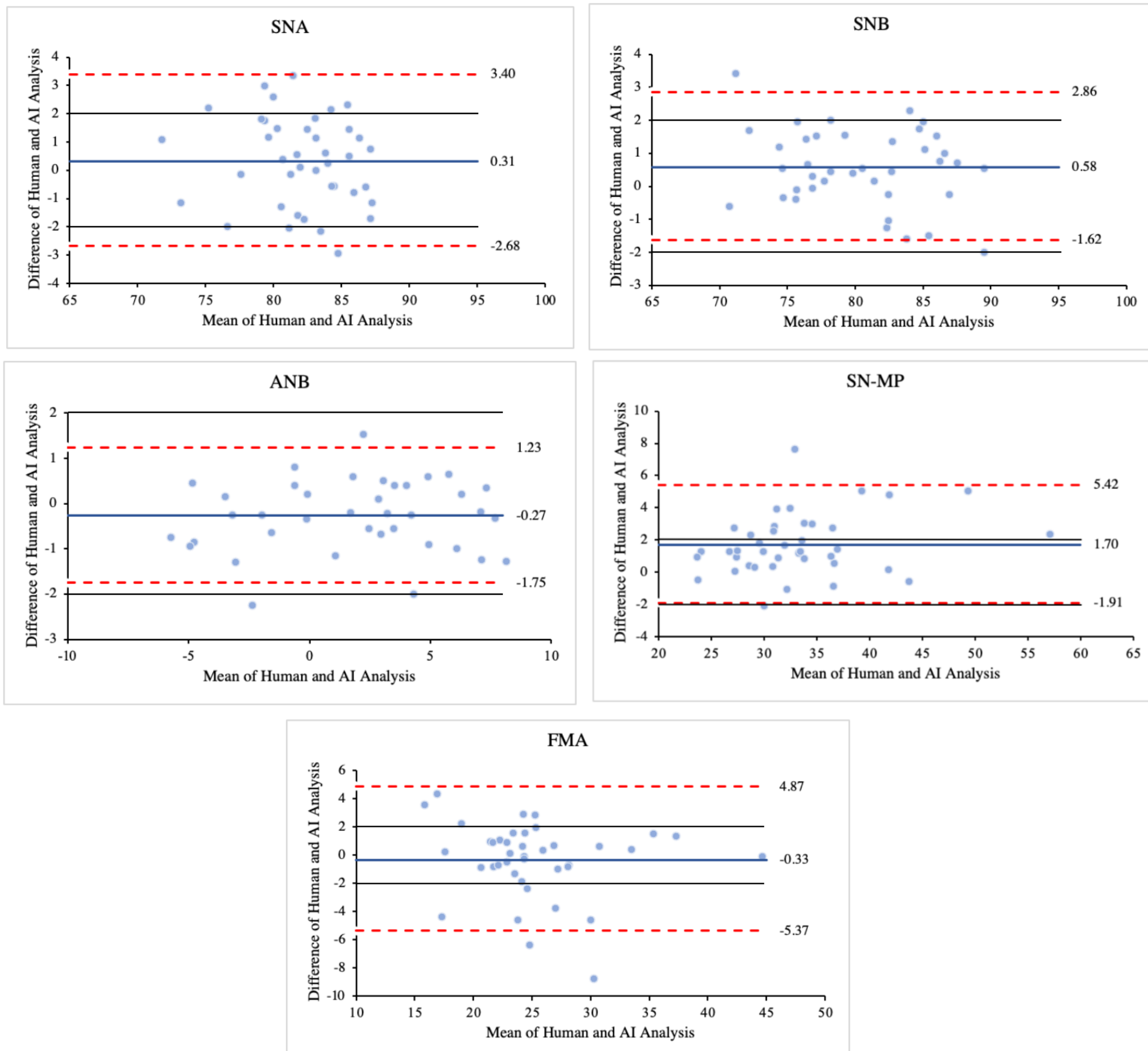


Figure 3. Bland-Altman plots for skeletal measurements between conventional and AI methods. Each plot depicts the difference between each measurement made by CephX and human examiners (y-axis) over the mean of each measurement made by CephX and the human examiners (x-axis). The blue line represents the mean bias between measurements, and the red-hashed lines represent the upper and lower limits of agreement. The black lines represent the maximum allowed difference between measurements.

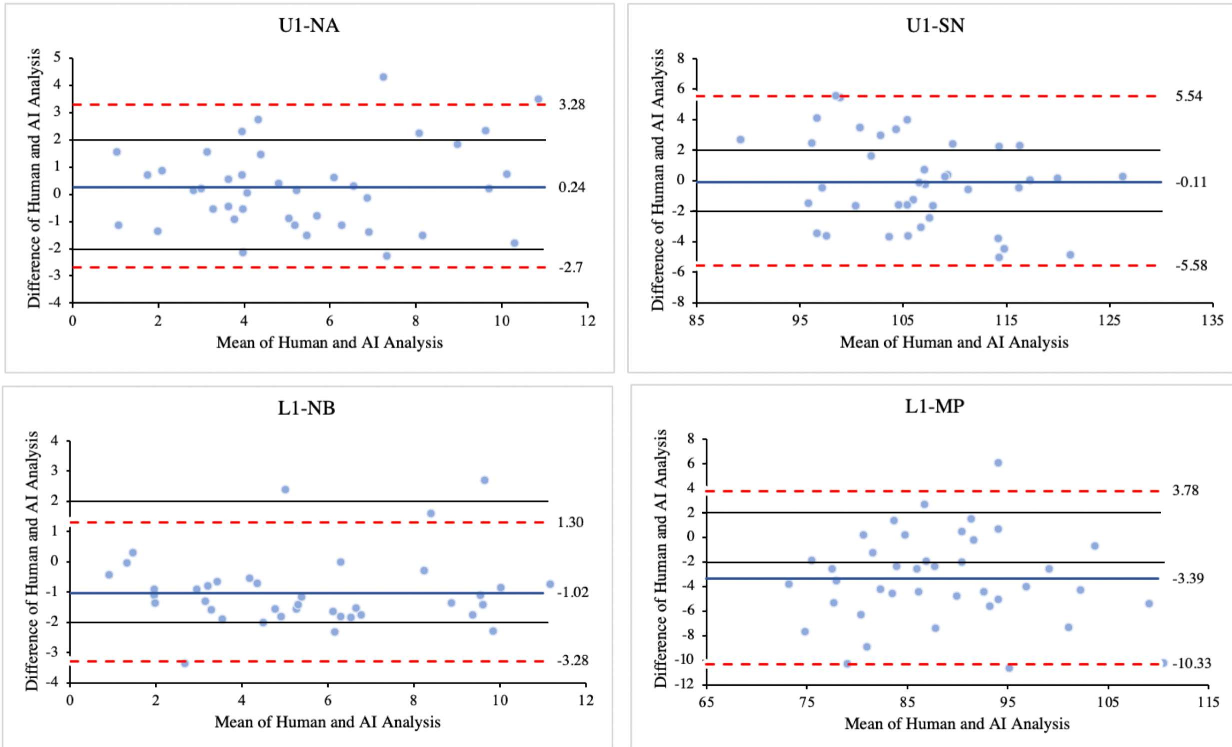


Figure 4. Bland-Altman plots for dental measurements between conventional and AI methods. Each plot depicts the difference between each measurement made by CephX and human examiners (y-axis) over the mean of each measurement made by CephX and the human examiners (x-axis). The blue line represents the mean bias between measurements, and the red-hashed lines represent the upper and lower limits of agreement. The black lines represent the maximum allowed difference between measurements.

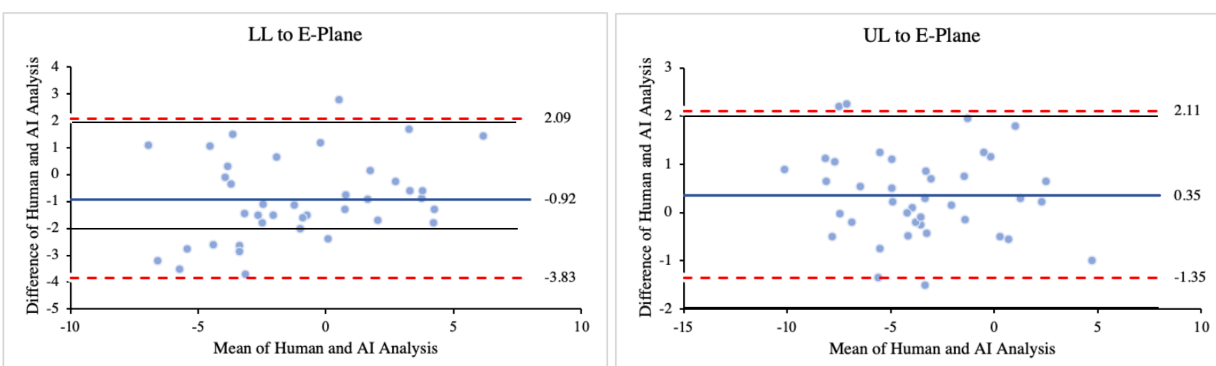


Figure 5. Bland-Altman plots for soft tissue measurements between conventional and AI methods. Each plot depicts the difference between each measurement made by CephX and human examiners (y-axis) over the mean of each measurement made by CephX and the human examiners (x-axis). The blue line represents the mean bias between measurements, and the red-hashed lines represent the upper and lower limits of agreement. The black lines represent the maximum allowed difference between measurements.

DISCUSSION

While tracing of lateral cephalometric radiographs is a component of a standard orthodontic work-up, studies show that this is frequently not the case in clinical practice, with only 60% of orthodontists “always” taking pre-treatment lateral cephalograms and less than 40% “always” performing cephalometric analysis on these images[3]. Studies suggest that the time-consuming and detail-oriented nature of this analysis may be in part to blame for its lack of routine use in clinical practice[3,4].

As imaging techniques in orthodontics and in dentistry continue to evolve, so too will methods of obtaining and tracing lateral cephalometric radiographs. Three-dimensional CBCT imaging offers better visualization of a variety of craniofacial structures, and does not suffer from the magnification or distortion that can be seen in two-dimensional imaging[7,8,9,10]. CBCT also has the benefit of being able to reconstruct traditional views of the teeth and skull, such as panoramic and lateral cephalometric radiographs, which can be analyzed with similar or even better accuracy than their traditional two-dimensional counterparts[11-14]. While CBCT imaging does currently have limitations in terms of cost, accessibility, and increased radiation dose, technological advancements continue to diminish these limitations, and CBCT imaging may eventually replace two-dimensional imaging techniques for all orthodontic patients[2,10].

Presently, there are two widely accepted methodologies for cephalometric analysis, both using manual landmark identification. In a completely manual analysis, landmarks are identified using a sheet of acetate paper overlying a physical radiograph, while in digital analysis, landmarks are identified virtually on a digitized image[15]. Both methodologies are well validated in the literature, with computer-assisted workflows resulting in more efficient tracing times compared to manual analysis[3,15-17]. Nevertheless, researchers and clinicians continue to try to find new ways to make cephalometric analysis more accurate and more time-efficient, and there are now AI algorithms that can be used to trace these images. These algorithms have largely focused on two-dimensional imaging, and while some studies have demonstrated acceptable accuracy, others have disputed these findings[15,18, 22-28]. Despite the fact that two-dimensional AI analyses have yet to be fully validated in the literature, programs such as CephX now offer the ability to obtain cephalometric analysis directly from CBCT volumes[29]. The present study is the first that seeks to better evaluate the accuracy and workflow of this direct-from-CBCT automated cephalometric analysis.

Our study first examined the impacts of the fully automated cephalometric analysis on workflow time compared to the computer-assisted analysis using human examiners. While proponents of automated cephalometric analysis cite reduced tracing times as one of its benefits, our study found a significantly greater workflow time from CBCT upload to the time when analysis became available for the automated analysis compared to the human examiners ($p<.001$) [15-17]. Even the minimum AI workflow time of 22 minutes and 58 seconds took longer than the maximum workflow time of 7 minutes and 44 seconds spent by the human examiners. While our findings shows that CephX takes significantly longer from start to finish, it is important to note that the automated analysis does not require doctor time, and can therefore be delegated to assistants or other staff members in the clinic. This is not the case with the computer-assisted workflow, which requires image reconstruction and landmark identification by a trained examiner. Therefore, while manual reconstruction and analysis can be accomplished more quickly in absolute terms, it remains less efficient overall with respect to doctor time.

In order to develop the true measurement values for each patient with which to compare the AI algorithm, it was first necessary to evaluate and establish intra-rater reliability between the examiners. In this regard, both examiners showed excellent intra-rater reliability across all measurements. The ICC values for both examiners were above 0.9 for all measures, with .982 being the lowest ICC for either examiner on any measure. CephX displayed perfect ICC values across all measurements, showing that the software has perfect agreement when provided with the same input data. Overall, intra-rater reliability was excellent for each of the human examiners as well as for the CephX software.

Next, we evaluated the inter-rater reliability between the human examiners. Our examiners showed a high level of inter-rater reliability across all measurements. This was observed by averaging the absolute differences between measurements taken by each examiner for each patient. The average difference between examiners was below 1 measurement unit for all values, which was well within the 2-unit limit for clinical acceptability. Soft tissue measurements such as LL and UL to E-Plane, as well as L1-NB and ANB, showed the greatest agreement between examiners. SNA, SN-MP, U1-SN, and L1-MP showed the greatest variability between examiners. While all measurements were in close agreement, this may suggest that measurement values which include the landmarks Sella and Nasion are among the most difficult to consistently identify between examiners. The MAR was consistent at 100% agreement within 2 mm for all linear measurements, and at 90% or greater agreement within 2 degrees for all angular measurements. This indicates a slightly greater amount of variability and disagreement in the angular measurements compared to the linear ones. Nevertheless, the overall level of agreement was shown to be very high between the examiners, and their values were averaged to develop the “true” value for each patient’s measurement to be compared against the AI software.

To evaluate the accuracy of the CephX software, means and standard deviations were calculated for each of the measurements taken using each tracing modality. Paired t-tests were used to evaluate the significance of any differences between the two modalities. There was a statistically significant difference between CephX and the human examiners in seven of the eleven cephalometric values. However, this lack of significant difference in these values must be interpreted with caution – as AI error can occur either above or below the true measurement value, measurement error which occurs with equal frequency and magnitude above and below the true value will erroneously show no difference on average between the two groups. For this reason, it is necessary to also examine absolute magnitude of difference between measurements made by CephX and the human examiners, as was done between the examiners previously.

The mean absolute error between CephX and the true value measurements was calculated for each of the eleven variables.. Unlike the human examiners, average error was below 1 measurement unit for only two variables. However, average error was within the clinically accepted 2 measurement units for nine of the eleven variables. Measures of incisor inclination (U1-SN and L1-MP) showed the greatest average measurement error, with L1-MP showing the lowest overall agreement with the human examiners. As with the human examiners, we also examined the MAR for the CephX software. The most frequent agreement within 2 units was seen with the variables ANB and UL to E-Plane, while U1-SN and L1-MP showed the least frequent agreement. While no human examiner showed worse than 90% MAR within 2 measurement units, CephX recorded MAR values of 90% or higher at this level for only two of the eleven variables. The average measurement error by CephX against the true values was higher than that of the human examiners to each other for every variable that was measured in our study.

Finally, Bland-Altman plots were developed to examine the strength and direction of the measurement bias found for each variable. There was no general trend for CephX to over- or underestimate values, as it overestimated five and underestimated six out of the eleven total variables in the ABO analysis. Of these eleven variables, nine appeared to show consistent strength and direction of measurement bias, with two appearing to show slight trends towards variations in measurement bias across high and low mean values. Ten out of eleven measurements displayed limits of agreement which acceptable the maximum allowed difference of +/- 2 measurement units.

With respect to workflow time, our data reject the hypothesis that workflow time for CephX would be lower than for the human examiners. The significantly greater workflow time seen with the AI software was statistically significant at $p<.001$, taking an average of greater than five times as long as the human examiners. As stated above, however, utilization of the AI software requires zero doctor time; therefore, this is a task that can be easily delegated to staff or other team members, which can increase overall efficiency even if the raw time taken for each analysis is indeed higher. Taking these findings into account, and assuming perfect accuracy of the AI software, it may be prudent to manually reconstruct and trace any images for which analysis is time-sensitive, while utilizing the AI for routine cephalometric tracing across the majority of orthodontic patients.

Regarding the accuracy of CephX, our results appear to be fairly consistent with those of the existing literature – that is, while promising, automated cephalometric analysis algorithms do not yet fully approximate the accuracy of measurements made by trained human examiners. While statistically significant differences do not inherently imply clinical significance, seven of the eleven variables showed a statistically significant difference from the human examiners. Two of the eleven variables (U1-SN, 2.35°, and L1-MP, 4.05°) showed measurement error outside of the clinically acceptable 2 unit range, and two others (SN-MP, 1.95°, and FMA, 1.86°) closely approximated this limit. The standard deviation of all measurement differences was also much greater than for the human examiners. Angular measurements compromised the four least accurate variables in the analysis, but also two of three most accurate, which seems to indicate that the software does not have a strong predilection for greater accuracy for either angular or linear measurements. Overall, there is significant promise with software such as CephX – nine of the eleven variables did ultimately show an average error within clinically acceptable limits, and two of the eleven (ANB, 0.64°, and UL to E-Plane, 0.75 mm) were below 1 unit difference. The Bland-Altman plots show that ANB was the only variable for which the limits of agreement of CephX's measurements fell within the maximum allowed difference, with the other variables falling outside of this acceptable range.

As technological advances occur and AI algorithms continue to improve, future studies will be needed in order to better characterize the accuracy of new automated cephalometric analyses which are developed. Our study excluded many patients with factors which would complicate the tracing process -- craniofacial patients, grossly asymmetric patients, and patients with significant deviations from a normal dentition were excluded in the present study. These additional factors can obscure key landmarks and make more the tracing process more challenging. Successful automated cephalometric analysis algorithms will need to be validate not only on study samples such as our own, but also on samples which include a more diverse range of patient types and malocclusions.

While the minimum sample size based on our power analysis was reached, the limitations of our study should also be discussed. Patients vary widely in craniofacial morphology, and as

AI algorithms continue to improve, larger sample sizes with more diverse patient populations will be required to fully validate the accuracy of automated cephalometric analysis software such as CephX. Furthermore, despite our high intra-examiner reliability, inclusion of additional human examiners may be beneficial in further bolstering true patient measurement values against which the AI software can be validated. Finally, our study examined only the accuracy of the ABO analysis. Future studies will be required to look at the accuracy of other analyses utilizing different cephalometric landmarks, as well as the impact of including these additional analyses on workflow time. As more and more cephalometric landmarks are included for the human examiners to analyze, the gap in workflow time will undoubtedly narrow. Therefore, the proposed benefits of AI in terms of workflow time may become clearer as more analyses are included for the human examiners to trace.

CONCLUSIONS

- In our study sample, the three most common CBCT indications were found to be surgical planning, TMJ concerns, and MARPE planning.
- The workflow time for analysis of each CBCT volume was found to be significantly greater for CephX than for the human examiners.
- Both CephX and the human examiners showed excellent intra-rater reliability, with ICC values greater than 0.9 for all measurements.
- A statistically significant difference was found between CephX and the human examiners in values for SNB, ANB, SN-MP, L1-NB, L1-MP, LL to E-Plane, and UL to E-Plane. There was no statistically significant difference for SNA, FMA, U1-NA, and U1-SN.
- Nine of the eleven ABO measurements performed by CephX showed average measurement error within the clinically acceptable two-unit difference: SNA, SNB, ANB, SN-MP, FMA, U1-NA, L1-NB, LL to E-Plane, and UL to E-Plane. U1-SN and L1-MP showed greater average measurement error than is clinically acceptable.
- CephX, on average, showed no overall predilection for over- or underestimation of cephalometric values.
- Nine of the eleven ABO measurements performed by CephX showed consistent strength and direction of measurement bias across high and low values.
- Only one measurement (ANB) showed limits of agreement within the maximum allowed difference; ten out of the eleven variables exceeded the maximum allowed difference.
- FMA may show a slight tendency towards greater accuracy at high values, while LL to E-Plane may display a slight tendency to underestimate more frequently at low values and overestimate more frequently at high values.

REFERENCES

1. Broadbent, B. H. (1931). A new x-ray technique and its application to orthodontia. The Angle Orthodontist, 1(2), 45-66.
2. Proffit WR, Fields Jr HW, Sarver DM. Contemporary orthodontics. United Kingdom: Elsevier Health Sciences; 2018. pp. 2-5, 24-25, 186-191.
3. McCabe, M., & Rinchuse, D. J. (2014). A survey of orthodontic practitioners regarding the routine use of lateral cephalometric radiographs in orthodontic treatment. Orthodontic practice US, 2.
4. Montúfar, J., Romero, M., & Scougall-Vilchis, R. J. (2018). Hybrid approach for automatic cephalometric landmark annotation on cone-beam computed tomography volumes. American Journal of Orthodontics and Dentofacial Orthopedics, 154(1), 140-150.
5. Kapila, S. D., & Nervina, J. M. (2015). CBCT in orthodontics: assessment of treatment outcomes and indications for its use. Dentomaxillofacial radiology, 44(1), 20140282.
6. Scarfe, W. C., Azevedo, B., Toghyani, S., & Farman, A. G. (2017). Cone beam computed tomographic imaging in orthodontics. Australian dental journal, 62, 33-50.
7. Agrawal, J. M., Agrawal, M. S., Nanjannawar, L. G., & Parushetti, A. D. (2013). CBCT in orthodontics: the wave of future. The journal of contemporary dental practice, 14(1), 153.
8. Mah, J. K., Huang, J. C., & Choo, H. (2010). Practical applications of cone-beam computed tomography in orthodontics. The Journal of the American Dental Association, 141, 7S-13S.

9. De Grauwe, A., Ayaz, I., Shujaat, S., Dimitrov, S., Gbadegbegnon, L., Vande Vannet, B., & Jacobs, R. (2019). CBCT in orthodontics: a systematic review on justification of CBCT in a paediatric population prior to orthodontic treatment. European journal of orthodontics, 41(4), 381-389.
10. Abdelkarim, A. (2019). Cone-beam computed tomography in orthodontics. Dentistry journal, 7(3), 89.
11. Kumar, V., Ludlow, J. B., Mol, A., & Cevidanes, L. (2007). Comparison of conventional and cone beam CT synthesized cephalograms. Dentomaxillofacial Radiology, 36(5), 263-269.
12. Kumar, V., Ludlow, J., Soares Cevidanes, L. H., & Mol, A. (2008). In vivo comparison of conventional and cone beam CT synthesized cephalograms. The Angle Orthodontist, 78(5), 873-879.
13. Montúfar, J., Romero, M., Muñoz-Jiménez, V., Scougall-Vilchis, R. J., & Jiménez, B. (2018). Perspective and orthogonal CBCT/CT digitally reconstructed radiographs compared to conventional cephalograms. In Conf. Biomedical Engineering and Sciences| (Vol. 16, pp. 41-45).
14. Moshiri, M., Scarfe, W. C., Hilgers, M. L., Scheetz, J. P., Silveira, A. M., & Farman, A. G. (2007). Accuracy of linear measurements from imaging plate and lateral cephalometric images derived from cone-beam computed tomography. American Journal of Orthodontics and Dentofacial Orthopedics, 132(4), 550-560.
15. Leonardi, R., Giordano, D., Maiorana, F., & Spampinato, C. (2008). Automatic cephalometric analysis: a systematic review. The Angle Orthodontist, 78(1), 145-151.

16. Chen, S. K., Chen, Y. J., Yao, C. C. J., & Chang, H. F. (2004). Enhanced speed and precision of measurement in a computer-assisted digital cephalometric analysis system. The Angle Orthodontist, 74(4), 501-507.
17. Roden-Johnson, D., English, J., & Gallerano, R. (2008). Comparison of hand-traced and computerized cephalograms: landmark identification, measurement, and superimposition accuracy. American journal of orthodontics and dentofacial orthopedics, 133(4), 556-564.
18. Anuwongnukroh, N., Dechkunakorn, S., Damrongsri, S., Nilwarat, C., Pudpong, N., Radomsutthisarn, W., & Kangern, S. (2018). Assessment of the Reliability of Automatic Cephalometric Analysis Software. International Journal of Mechanical Engineering and Robotics Research, 7(1).
19. Kim, H., Shim, E., Park, J., Kim, Y. J., Lee, U., & Kim, Y. (2020). Web-based fully automated cephalometric analysis by deep learning. Computer methods and programs in biomedicine, 194, 105513.
20. Kim, M. J., Liu, Y., Oh, S. H., Ahn, H. W., Kim, S. H., & Nelson, G. (2021). Automatic Cephalometric Landmark Identification System Based on the Multi-Stage Convolutional Neural Networks with CBCT Combination Images. Sensors, 21(2), 505.
21. Kunz, F., Stellzig-Eisenhauer, A., Zeman, F., & Boldt, J. (2020). Artificial intelligence in orthodontics. Journal of Orofacial Orthopedics/Fortschritte der Kieferorthopädie, 81(1), 52-68.
22. Park, J. H., Hwang, H. W., Moon, J. H., Yu, Y., Kim, H., Her, S. B., ... & Lee, S. J. (2019). Automated identification of cephalometric landmarks: Part 1—Comparisons between the latest deep-learning methods YOLOV3 and SSD. The Angle Orthodontist, 89(6), 903-909.

23. Hwang, H. W., Park, J. H., Moon, J. H., Yu, Y., Kim, H., Her, S. B., ... & Lee, S. J. (2020). Automated identification of cephalometric landmarks: Part 2-Might it be better than human? The Angle Orthodontist, 90(1), 69-76.
24. Lindner, C., Wang, C. W., Huang, C. T., Li, C. H., Chang, S. W., & Cootes, T. F. (2016). Fully automatic system for accurate localisation and analysis of cephalometric landmarks in lateral cephalograms. Scientific reports, 6(1), 1-10.
25. Jeon, S., & Lee, K. C. (2021). Comparison of cephalometric measurements between conventional and automatic cephalometric analysis using convolutional neural network. Progress in Orthodontics, 22(1), 1-8.
26. Meriç, P., & Naoumova, J. (2020). Web-based Fully Automated Cephalometric Analysis: Comparisons between App-aided, Computerized, and Manual Tracings. Turkish Journal of Orthodontics, 33(3), 142.
27. Mosleh, M. A. A., Baba, M. S., Malek, S., & Almaktari, R. A. (2016). Ceph-X: development and evaluation of 2D cephalometric system. BMC bioinformatics, 17(19), 193-201.
28. Alqahtani, H. (2020). Evaluation of an online website-based platform for cephalometric analysis. Journal of stomatology, oral and maxillofacial surgery, 121(1), 53-57.
29. Orca Dental AI. Artificial Intelligence Empowered Dentistry. CephX. Retrieved December 2, 2021, from https://cephx.com/.
30. Montúfar, J., Romero, M., & Scougall-Vilchis, R. J. (2018). Automatic 3-dimensional cephalometric landmarking based on active shape models in related projections. American Journal of Orthodontics and Dentofacial Orthopedics, 153(3), 449-458.

31. Bao, H., Zhang, K., Yu, C., Li, H., Cao, D., Shu, H., ... & Yan, B. (2023). Evaluating the accuracy of automated cephalometric analysis based on artificial intelligence. BMC Oral Health, 23(1), 1-10.