EFFECT OF ARTIFiCIAL INTELLIGENCE VERSUS GUIDED LANDMARKS IDENTIFICATION ON THE ACCURACY OF THE LATERAL CEPHALOMETRIC ANALYSIS

Mohamed G. Hemeda[a] , Wael M. Refai [b] , Mohamed G. Elshal [c], Mostafa A. Mohammed [d] ,Ahmed N. Abdelhameed[e].

Abstract

Objective: *This study aimed to compare the effect of Artificial Intelligence versus guided Landmarks identification on the accuracy of the Lateral Cephalometric Analysis.*

Methodology: Three orthodontic specialists identified 17 radiographic landmarks manually for 22 different types of angular and linear measurements of 50 lateral cephalometric radiographs then tracing and analysis were done by Artificial Intelligence based software (Webceph) and Automated cephalometric analysis software (Romexis software). The measurements of the two softwares compared to humans' gold standard (Mean values of the three examiners).

Results: *comparison between humans' gold standard and (Wepceph) the AI's predictions showed no proportional bias in 12 parameters, The mean differences range from 0.2° to 2.9° for angular measurements except Gonial angle 4.55°, Upper 1 to NA angle 3.78° and IMPA 3.72° and from 0.25 to 1.67 mm for linear measurements. Comparison betwe en humans' gold standard and Automated cephalometric analysis software (Romexis software) showed proportional bias in 19 parameters, The mean differences range from 0.16° to 12.67° for angular measurments and from 1.01 to 13.38 mm for linear measurements.*

Conclusions: AI based software is able to identify landmarks of cephalometric X-rays at almost the same quality level as experienced human examiners (current gold standard).comparison between the two types of softwares showed that the accuracy of AI based (Webceph) software is better than the automated cephalometric analysis (Romexis) software.

Keywords: *Cephalometric, manual tracing, Artificial Intelligence, Automated cephalometric analysis.*

Introduction:

Cephalometric analysis has long been, and still is one of the most important tools in evaluating craniomaxillofacial skeletal profile. To perform this, manual tracing of x-ray film and plotting landmarks have been required. This procedure is time-consuming and demands experience. Accordingly, computerized cephalometric systems have been introduced nowadays. However, tracing and plotting still have to be done on the monitor display(11) .

a- Candidate of orthodontics, Department of orthodontics, Faculty of Dentistry, Minia University, Minia, Egypt.
b- Professor of orthodontics, Faculty of Dentistry, Minia University, Minia, Egypt.
c- Lecturer of orthodontics, Faculty of Dentistry, Minia University, Minia, Egypt.
d- Lecturer of orthodontics, Faculty of Dentistry, Minia University, Minia, Egypt.
e- Lecturer of orthodontics, Faculty of Dentistry, Minia University, Minia, Egypt.
Corresponding author
Mohamed G. Hemeda, Department of orthodontics, Faculty of Dentistry, Minia University,minia, Egypt.:email dr.m.hemeda@gmail.com

Artificial intelligence (AI) is developing rapidly. Artificial intelligence is defined as the capability of a machine to imitate intelligent human behavior to perform complex tasks, such as problem solving, object and word recognition, and decision-making(6,15).

Artificial intelligence has recently made substantial strides in perception (the interpretation of sensory information), allowing machines to better represent and interpret complex data. This has led to major advances in applications ranging from web search and self-driving vehicles to natural language processing and computer vision — tasks that until a few years ago could be done only by humans(7).

Deep learning is one of the most evolving areas in artificial intelligence. An automated landmark predicting system, based on a deep learning neural network was invented(11).

Deep learning is a subset of machine learning that is based on a neural network structure loosely inspired by the human brain. Such structures learn discriminative features from data automatically, giving them the ability to approximate very complex nonlinear relationships. While most earlier AI methods have led to applications with subhuman performance, recent deep learning algorithms are able to match and even surpass humans in task-specific applications (9,12).

Due to recent advancements in computing, such AI algorithms can now be used for abstract and complex tasks. Accordingly, these are promising fields in health care. AI algorithms are suitable to assist clinicians in analyzing medical imaging, diagnosis of certain diseases and therefore to support therapeutic decisions (1) .

The comparison between AI and automated cephalometric analysis in Landmarks identification was found to be a point of worthy of investigation in the sake of improving the ease of orthodontic treatment. Accordingly, this study was inducted.

Patients and methods:

1- Ethical regulations:

This study design was approved by the Research Ethics Committee of the Faculty of Dentistry, Minia University, Minia, Egypt.

The investigation of this study was carried out in compliance with the Declaration of Helsinki(2).

All radiographs were coded to mask any patient informations.

2- Sample collection:

The sample of the study included fifty lateral cephalometric radiographs(16) taken from archived lateral cephalometric radiographs and patients seeking orthodontic treatment in the orthodontic clinics in the Faculty of Dentistry, Minia University.

Digital lateral cephalograms of the subjects were taken on a digital cephalometric machine in a standing position with relaxed lips, teeth in centric occlusion and the subject's head in such a position that the Frankfort horizontal plane was parallel to the floor.

All lateral cephalograms were then transferred to a computer loaded with Romexis software and the hard copies were printed with the help of an X-ray printer.

The inclusion criteria for cephalograms were as follows:

1-The X-rays should be of good quality to permit identification of landmarks.

2- All the radiographs should be taken from the same machine.

3- All the radiographs should show the calibration ruler.

3-Tracying and analysis:

The lateral cephalometric radiographs were traced by three methods:

A) Human Examiners analysis (Manual Identification of Cephalometric Landmarks):

Three orthodontic specialists manually identified 17 radiographic landmarks on each cephalometric radiograph (Table 1).

A total of 22 different types of angular and linear measurements (Table 2) were analyzed and compared among groups to evaluate the skeletal relationship between the cranial base and the mandible or maxilla, the relationship between mandible and maxilla, and the dentoalveolar relationship.

After placing registration points on the hard copies of the lateral cephalograms, landmarks were traced manually on tracing paper using 0.5 mm 3H pencil on a view box using transilluminated light in a dark room. Any stray light radiations were eliminated by covering margins of the view box around the radiograph with a black paper.

No more than 10 radiographs were traced in a single session to minimize errors due to examiner fatigue. landmarking was carried out in five sessions for each examiners. A time interval longer than 6 weeks elapsed between first and second analyses

Mean values of the three examiners for each parameter were defined as humans' gold standard and compared to group (B) and group (C)'s predictions.

B) AI based software (Webceph) analysis:

Webceph is an Artificial Intelligence Orthodontic Program which can detect anatomical landmarks automatically in seconds and displays them on the screen. The use of AI in Orthodontics is limited to supervized learning such as objects or point recognition. Webceph is example of cephalometric software programs which are trained to recognized points in cephalometric radiographs to facilitate cephalometric analysis.(5) AI always detects identical position which implies that AI may be the reliable option for repeatedly identifying multiple cephalometric landmarks.(8)

Calculations relevant to AI were performed by laptop (HP Notebook - 15-da0053wm), figure 1 showed AI cephalometric tracing.

(Figure 1: A.I driven cephalometric tracing)

C) Automated cephalometric analysis software (Romexis software):
The Romexis software allows capturing, viewing and processing all types of data in the same system – with all 2D and 3D images and CAD/CAM cases in the same database, figure 2 showed automatic tracing by Romexis software.

(Figure 2: automatic tracing)

Table 1: Cephalometric landmarks

landmark name	Symbol	Definition
Sella	S	Center of sella turcica or The geometric centre of the pituitary fossa.
Nasion	N	Anterior point of nasofrontal suture in the median plane.
Orbitale	Or	The most inferior point of orbit.
Porion	Po	The most superiorly positioned point of the external auditory meatus Pterygoid point (Pt)The intersection of the inferior border of the foramen rotundum with the posterior wall of the pterygomaxillary fissure.
Subspinale	A	The deepest point on contour of alveolar projection between anterior nasal spinal and prosthion or Most concave point on anterior maxillary surface.
Supramentale	B	The deepest point on contour of alveolar projection between infradentale and pogonion or Most concave point on anterior mandibular surface.
Pogonion	Pg	The most anterior point of symphysis.
Gnathion	Gn	A point located by taking the midpoint between the anterior (pogonion) and inferior (menton) points of the bony chin or The most downward and forward point on the symphysis.
Menton	Me	The lowest point on mandibular symphysis.
Gonion	Go	Intersection of the line connecting the most distal aspect of the condyle to the distal border of the ramus (ramus plane) and the line at the base of the mandible (mandibular plane).
Articulare	Ar	The point of intersecting of the posterior margin of the ascending ramus and the outer margin of the cranial base.
Upper incisor edge	UIE	Incisal edge of the most anterior upper incisor.
Lower incisor edge	LIE	Incisal edge of the most anterior lower incisor.
Upper incisor apex	UIA	The root apex of the most prominent upper incisor.
Lower incisor apex	LIA	The root apex of the most prominent lower incisor.
Anterior Point for the occlusal plane	APOcc	The midpoint of the incisor overbite in occlusion.
Posterior point for the occlusal plane	PPOcc	The most distal point of the contact between the most posterior molars in occlusion.

Table 2: Linear and angular measurements

parameter	Definition
SNA (°)	Angle determined by points S, N, and A.
SNB (°)	Angle determined by points S, N, and B.
ANB (°)	Angle determined by points A, N, and B.
Gonial angle (°)	Angle determined by points Ar, Go and Me.
Upper gonial angle (°)	Angle determined by points Ar, Go and N.
Lower gonial angle (°)	Angle determined by points N, Go and Me.
Saddle angle (°)	Angle determined by points N, S and Ar.
Articular angle (°)	Angle determined by points S, Ar and Go.
Y axis (°)	Angle formed by the intersection of S-Gn line to Frankfort horizontal plane anteriorly.
Wit's appraisal (mm)	Distance between points of A and B to the occlusal plane.
Interincisal angle (°)	Angle formed by the intersection of the mandibular incisor axis to the maxillary incisor axis.
Upper incisor / SN (°)	Angle between upper incisor axis and SN line posteriorly.
Upper incisor / FH (°)	Angle between upper incisor axis and Frankfurt horizontal plane posteriorly.
Upper incisor to NA (°)	Angle formed by the intersection of the maxillary incisor axis to the plane between points N and A .
Upper incisor to NA (mm)	Perpendicular distance from the tip of the maxillary incisor to the plane between points N and A .
Lower incisor to NB (°)	Angle formed by the intersection of the mandibular incisor axis to the plane between points N and B.
Lower incisor to NB (mm)	Perpendicular distance from the tip of the mandibular incisor to the plane between points N and B.
Upper incisor to A vertical (mm)	Perpendicular distance from the tip of the maxillary incisor to a line drawn perpendicular from point A to Frankfort horizontal plane.
Lower incisor to A pog (mm)	Perpendicular distance from the tip of the mandibular incisor to the plane between points A and Pog.
FMA (°)	Angle formed between the Frankfort horizontal plane and the mandibular (Go-Me) plane.
FMIA (°)	Angle formed between the Frankfort horizontal plane and the mandibular incisor axis.
IMPA (°)	Angle formed between the mandibular plane and the mandibular incisor axis.

Statistical analysis :

The collected data were coded, tabulated, and statistically analyzed using SPSS program (Statistical Package for Social Sciences) software version 25 and MedCalc software version 12.

Descriptive statistics were done for parametric quantitative data by mean, Standard deviation (SD) and minimum and maximum of range.

To assess the agreement between manual method and software methods many analyses were carried out:

-Cronbach's alpha coefficient and Interclass correlation coefficients to assess the reliability between different examiners and between different methods.

-Paired Samples T test to assess the significance of mean difference between the manual method and each software method.

-Pearson's correlation coefficient to assess the correlation between each two methods for each parameter.

-Bland–Altman plots were made for all investigated parameters to illustrate differences between the manual and each software method versus the two measurements average. In these plots, The mean differences between the two analyses as well as the 95% limits of agreement (mean difference± 1.96 x standard deviation of the differences) was visualized.

Finally, to determine if there were any proportional biases, Simple linear regression analyses is performed for each parameter with the difference between the manual and each software method as the dependent variable (criterion) and the mean of both analyses as the independent variable (predictor). The resulting regression lines were finally added to the corresponding Bland–Altman plot.

The level of significance was taken at (P value ≤ 0.05).

Results:

Reliability Coefficient and interclass correlation between examiners (Inter-observer):

Inter-observer reliability was very high through all parameters analyzed in this study (all ICC > 0.900 with p< 0.001).

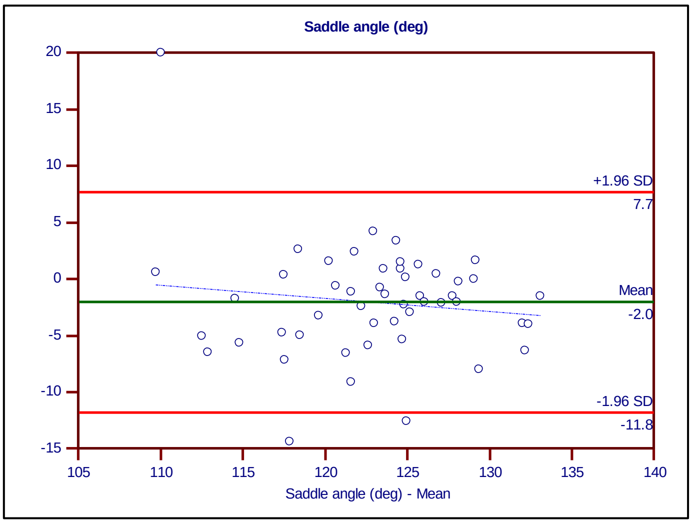
Comparison of AI (Webceph) predictions to humans' gold standard:

Comparison of Webceph (AI) software predictions to the humans' gold standard (Table 3). As regarding Pearson's correlation coefficient (r) for each parameters between the two methods of analysis, there was positive significant correlation for each parameter, mean difference between the two analysis for each parameter using Paired Samples T test showed significant difference between the two means for all parameters except SNB, articular angle, Y axis, interincisal angle, lower 1 to NB angle and distance and FMIA

Fig (3) is an example of Bland Altman plots which done for all parameters. The differences between the humans' gold standard and the predictions of Webceph software were plotted against the averages of the two measurements (X-axis).

The green lines illustrate the mean difference of both measurements (when these lines were extremely close to the zero line, there was no clinically relevant consistent bias). The red lines illustrate the 95% limits of agreement The dashed blue lines represent the linear regression line with the difference as the dependent variable (criterion) and the mean as the independent variable (predictor).

Simple linear regression analyses for SNA, SNB, ANB, Upper gonial angle, Saddle angle, Articular angle, Y axis, Wits appraisal, Upper 1/ SN, Lower 1 to NB angle, Lower 1 to A Pog and FMIA showed no statistically relevant p-values (all P> 0.05) no proportional bias (agreement) but Gonial angle, Lower gonial angle, Interincisal angle, Upper 1 / FH angle, Upper 1 to NA angle and distance, Lower 1 to NB distance, Lower 1 to A vertical , FMA and IMPA which show (P < 0.05). Therefore, these parameters exhibited proportional bias (no agreement)(Table3).



(Figure 3: Bland–Altman plots of saddle angle of Webceph software)

Comparison of Automated cephalometric analysis (Romexis) software to the human's gold standard:

Comparison of Automated cephalometric analysis (Romexis) software to the human's gold standard (Table 4). As regarding Pearson's correlation coefficient for each parameters between the two methods of analysis, there was insignificant correlation for all parameters with exception of FMA which showed positive significant correlation.

As regarding mean difference between the two methods for each parameter using Paired Samples T test showed significant difference between the two means for all parameters except SNB, upper gonial angle and lower 1 to A pog.

Fig (4) is example of Bland Altman plots which done for all parameters. The differences between the human's gold standard and the predictions of Romexis software method were plotted against the averages of the two measurements(X-axis).

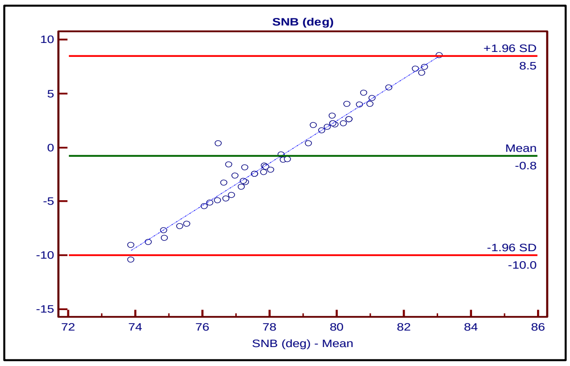
Table 3: Mean Difference, Bland-Altman Upper and Lower Limits of Agreements, and 95% CI of Differences Between Humans' gold standard and Webceph software
*: Significant level at P value < 0.05 (using Paired Samples T test) CI: Confidence Interval

	Bland-Altman Plot			**Pearson's correlation**		**Paired Samples T test**			
Humans' gold standard – Webceph Software	**Limits of agreement**		**P value (linear regression)**	**r**	**P value**	**Mean difference**	**Mean difference 95% CI**		**P value**
	Lower	**Upper**					**Lower**	**Upper**	
SNA (deg)	-6.23	4.24	0.889	0.748	<0.001*	-0.99	-1.75	-0.23	*0.011**
SNB (deg)	-3.50	3.93	0.539	0.916	<0.001*	0.22	-0.32	0.76	*0.422*
ANB (deg)	-3.72	1.40	0.545	0.915	<0.001*	-1.16	-1.54	-0.78	*<0.001**
Gonial angle (deg)	-2.96	12.05	**0.005***	0.866	<0.001*	4.55	3.46	5.64	*<0.001**
Upper gonial angle (deg)	-2.56	7.55	0.360	0.821	<0.001*	2.49	1.76	3.23	*<0.001**
Lower gonial angle (deg)	-1.58	5.79	**0.005***	0.957	<0.001*	2.11	1.57	2.64	*<0.001**
Saddle angle (deg)	-11.77	7.70	0.369	0.670	<0.001*	-2.03	-3.45	-0.62	*0.006**
Articular angle (deg)	-9.14	10.98	0.068	0.740	<0.001*	0.92	-0.54	2.38	*0.209*
Y axis (deg)	-4.26	5.20	0.363	0.867	<0.001*	0.47	-0.22	1.16	*0.175*
Wits appraisal (mm)	-2.42	3.50	0.852	0.954	<0.001*	0.54	0.11	0.97	*0.015**
Interincisal angle (deg)	-13.15	10.57	**0.010***	0.876	<0.001*	-1.29	-3.01	0.43	*0.138*
Upper 1 / SN (deg)	-6.09	11.89	0.319	0.834	<0.001*	2.90	1.60	4.21	*<0.001**
Upper 1/ FH (deg)	-6.68	12.23	**0.046***	0.805	<0.001*	2.77	1.40	4.15	*<0.001**
Upper 1 to NA angle (deg)	-6.33	13.88	**0.004***	0.755	<0.001*	3.78	2.31	5.24	*<0.001**
Upper 1 to NA distance (mm)	-2.82	6.16	**<0.001***	0.731	<0.001*	1.67	1.02	2.32	*<0.001**
Lower 1 to NB angle (deg)	-9.19	10.10	0.989	0.805	<0.001*	0.45	-0.94	1.85	*0.517*
Lower 1 to NB distance (mm)	-7.09	5.50	**0.021***	0.574	<0.001*	-0.79	-1.71	0.12	*0.087*
Upper 1 to A vertical distance (mm)	-2.39	4.03	**0.002***	0.873	<0.001*	0.82	0.36	1.29	*0.001**
Lower 1 to A pog distance (mm)	-2.06	2.57	0.532	0.914	<0.001*	0.25	-0.08	0.59	*0.135*
FMA (deg)	-2.80	8.47	**0.001***	0.931	<0.001*	2.84	2.02	3.65	*<0.001**
FMIA (deg)	-7.26	9.34	0.273	0.882	<0.001*	1.04	-0.17	2.24	*0.090*
IMPA (deg)	-10.40	2.97	**0.002***	0.934	<0.001*	-3.72	-4.68	-2.75	*<0.001**

Table 4: Mean Difference, Bland-Altman Upper and Lower Limits of Agreements, and 95% CI of Differences Between humans' gold standard and Romexis software
*: Significant level at P value < 0.05 (using Paired Samples T test)
CI: Confidence Interval

Humans' gold standard – Romexis software	Bland-Altman Plot			Pearson's correlation		Paired Samples T test			
	Limits of agreement		P value (linear regression)	r	P value	Mean difference	Mean difference 95% CI		P value
	Lower	Upper					Lower	Upper	
SNA (deg)	-5.55	9.42	**<0.001***	-0.086	0.554	1.94	0.85	3.02	*0.001**
SNB (deg)	-10.01	8.47	**<0.001***	0.009	0.948	-0.77	-2.11	0.57	*0.252*
ANB (deg)	-3.67	9.19	**<0.001***	-0.041	0.781	2.76	1.82	3.70	*<0.001**
Gonial angle (deg)	-14.52	10.01	**<0.001***	-0.033	0.817	-2.25	-4.03	-0.48	*0.014**
Upper gonial angle (deg)	-8.32	8.63	**<0.001***	0.230	0.108	0.16	-1.07	1.38	*0.800*
Lower gonial angle (deg)	-13.07	8.33	**<0.001***	0.186	0.196	-2.37	-3.92	-0.81	*0.004**
Saddle angle (deg)	-20.60	11.78	0.259	0.146	0.313	-4.41	-6.75	-2.06	*<0.001**
Articular angle (deg)	-8.22	22.19	0.693	0.202	0.160	6.99	4.78	9.19	*<0.001**
Y axis (deg)	-11.44	5.50	**<0.001***	0.272	0.056	-2.97	-4.20	-1.74	*<0.001**
Wits appraisal (mm)	0.81	22.96	**0.016***	0.156	0.280	11.89	10.28	13.49	*<0.001**
Interincisal angle (deg)	-37.54	12.19	**<0.001***	0.124	0.392	-12.67	-16.28	-9.07	*<0.001**
Upper 1 / SN (deg)	-11.58	20.44	**<0.001***	0.154	0.285	4.43	2.11	6.75	*<0.001**
Upper 1/ FH (deg)	-9.43	22.25	**<0.001***	0.171	0.234	6.41	4.11	8.71	*<0.001**
Upper 1 to NA angle (deg)	-12.82	17.78	**0.003***	0.145	0.314	2.48	0.26	4.70	*0.029**
Upper 1 to NA distance (mm)	-21.52	-5.24	**<0.001***	-0.130	0.366	-13.38	-14.56	-12.20	*<0.001**
Lower 1 to NB angle (deg)	-10.07	25.51	**<0.001***	-0.167	0.247	7.72	5.14	10.30	*<0.001**
Lower 1 to NB distance (mm)	-11.89	0.62	**<0.001***	0.065	0.655	-5.64	-6.54	-4.73	*<0.001**
Upper 1 to A vertical distance (mm)	-14.24	-1.39	**<0.001***	0.194	0.178	-7.82	-8.75	-6.89	*<0.001**
Lower 1 to A pog distance (mm)	-8.40	6.38	0.447	0.054	0.707	-1.01	-2.08	0.06	*0.064*
FMA (deg)	-16.21	7.49	**<0.001***	**0.293**	**0.039***	-4.36	-6.08	-2.64	*<0.001**
FMIA (deg)	-24.56	11.38	**<0.001***	0.033	0.822	-6.59	-9.19	-3.98	*<0.001**
IMPA (deg)	-7.62	29.53	**<0.001***	0.068	0.639	10.96	8.26	13.65	*<0.001**

Simple linear regression analyses for each parameter showed statistically relevant p-values (all p < 0.05), therefore, these parameters exhibited proportional bias, with exception of saddle angle, articular angle and lower 1 to A pog. Which showed no statistically relevant p-values (P > 0.05).



(Figure 4: Bland-Altman plots of ANB of Romexis software)

Discussion:

In this study, fifty lateral cephalometric radiographs were taken from the same machine with good quality to permit identification of landmarks. 17 radiographic landmarks were identified and 22 different types of angular and linear were measured on each cephalometric x-ray of the fifty lateral cephalometric radiographs manually by three orthodontic specialists to decrease human errors.

No more than 10 radiographs were traced in a single session to minimize errors due to examiner fatigue.

Tracing by the two softwares was done without editing any point position manually after landmark identification automatically by the softwares to detect the accuracy of the softwares without any human help, which software more accurate and which one was near to humans' gold standard the AI software or the automated software in landmark identification and in angular and linear measurements.

Before starting to calculate the humans' gold standard and to compare between the measurements of manual and the two softwares analysis reliability coefficient, interclass correlation (ICC) and Cronbach's alpha tests were done between the three examiner to make sure that the values of the measurements of the manual method of the three examiner within small range and near to each other, Interclass correlation showed ICC > 0.9 with p value < 0.001* and Cronbach's alpha > 0.7 , So there was excellent inter-rater reliability between the three examiners for each parameter.

Comparison of Webceph (AI) software predictions to the humans' gold standard by Pearson's correlation coefficient (r) which was done to assess the correlation between each two methods for each parameter between the two methods of analysis, there was positive significant correlation for each parameter. Bland–Altman plots were made for all investigated parameters to illustrate differences between the manual and software method, Simple linear regression analyses no statistically relevant p-values (P> 0.05) no proportional bias (agreement) for 12 parameters (SNA, SNB, ANB, Upper gonial angle, Saddle angle, Articular angle, Y axis, Wits appraisal, Upper 1/ SN, Lower 1 to NB angle, Lower 1 to A Pog and FMIA) and there was no agreement for 10 parameters (Gonial angle, Lower gonial angle, Interincisal angle, Upper 1 / FH angle, Upper 1 to NA angle and distance, Lower 1 to NB distance, Lower 1 to A vertical , FMA and IMPA) which show (P < 0.05) but the mean difference of this parameters was (4.55°, 2.11°, -1.29°, 2.7°, 3.78°, 1.6 mm, -0.79 mm, 0.82 mm, 2.84° and -3.72°) respectively. The mean differences range from 0.2° to 2.9° for angular measurements except Gonial angle 4.55°, Upper 1 to NA angle 3.78° and IMPA 3.72° and from 0.25 to 1.67 mm for linear measurements would probably not make a difference in the treatment and is insignificant for a clinical decision (9,3) .This result was similar to that of Felix Kunz et al(8) which has a very high correlation between the predictions of the AI and the humans' gold standard. Silva et al(14) showed CEFBOT (an artificial intelligence (AI)-based cephalometry software) was able to perform all but one of the ten measurements. ICC values > 0.94 were found for the remaining eight measurements, while the Frankfurt horizontal plane - true horizontal line (THL) angular measurement showed the lowest reproducibility (human, ICC = 0.876; CEFBOT, ICC = 0.768). Measurements performed by the human examiner and by CEFBOT were not statistically different. Nishimoto et al (11) trained the regression network with 7803 images. Angles and lengths in cephalometric analysis, predicted by the neural network, were not significantly different from those, calculated by the coordinate values, and plotted manually.

Comparison of Automated cephalometric analysis (Romexis) software to the human's gold standard by Pearson's correlation coefficient for each parameters between the two methods of analysis, there was insignificant correlation for all parameters with exception of FMA which showed positive significant correlation. Mean difference between the two methods for each parameter using Paired Samples T test showed significant difference between the two means for all parameters except SNB, upper gonial angle and lower 1 to A pog. Bland Altman plots of all parameters, Simple linear regression analyses for each parameter showed statistically relevant p-values (all $p < 0.05$), therefore, these parameters exhibited proportional bias, with exception of saddle angle, articular angle and lower 1 to A pog. which showed no statistically relevant p-values ($P > 0.05$). The mean differences range from 0.16° to 12.67° for angular measurments and from 1.01 to 13.38 mm for linear measurements. The result were similar to those of Chen et al(3) which showed the differences of landmark location between original cephalometric radiographs and their digital counterparts were statistically significant, the reliability of landmark identification in digital images was comparable to that in original radiographs except for four points.The results were not in accordance with that of Roden-johnson et al(13) in which no difference in the identification of cephalometric landmarks made manually vs digitally with Quick Ceph 2000. There was no difference in acquiring consistent cephalometric values for the measurements required by the American Board of Orthodontics for the Phase III clinical examination manually vs digitally by using Quick Ceph 2000. The differences in the results could be explained that Roden-johnson et al used another program of automated cephalometric analysis software.

Conclusions:

-The comparison between the two types of softwares to humans' gold standard showed that the accuracy of AI based (Webceph) software is better than the automated cephalometric analysis (Romexis) software.

-In most parameters AI showed as accurate an identification of cephalometric landmarks as did human examiners.

References:

1-A Nasef, WM Refai. Application of a New Three Dimensional Method of Analysis for Comparison between the Effects of Two Different Methods of Distalization of the Maxillary First Molar. Egypt Dent J 61, 4195-4201, (2015).

2-Blaikie N, Priest J. Designing social research: The logic of anticipation. John Wiley & Sons; (2019 Jan 15).

3-Chen, Yi Jane, Ssu Kuang Chen, Hsin Fu Chang, and Kun Chee Chen. "Comparison of landmark identification in traditional versus computer-aided digital cephalometry." The Angle Orthodontist 70 (5), 387-392.(2000).

4-Davoudian, Masoud. "The validity of computer assisted system for lateral cephalometric analysis in comparison with conventional method. II. Onyxceph Imaging Software." Iran. J Orthod 4, 67-77.(2009).

5-Finlay, Laetitia M. Craniometry and cephalometry: a history prior to the advent of radiography. The Angle Orthodontist, 50(4), 312-321. (1980).

6-Hashimoto, Daniel A., Guy Rosman, Daniela Rus, and Ozanan R. Meireles. Artificial intelligence in surgery: promises and perils. Annals of surgery, 268(1), 70-76 (2018).

7- Hosny Ahmed, Chintan Parmar, John Quackenbush, Lawrence H. Schwartz, and Hugo JWL Aerts. Artificial intelligence in radiology. Nature Reviews Cancer, 18(8), 500-510.(2018).

8-Kunz, F., A. Stellzig-Eisenhauer, F. Zeman, and J. Boldt. "Evaluation of a fully automated cephalometric analysis using a customized convolutional neural network." J. Orofac. Orthop 81, 52-68(2019).

9-Mnih, Volodymyr, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare and Alex Graves. Human-level control through deep reinforcement learning. Nature, 518(7540), 529-533. (2015).

10-Navarro, Ricardo de Lima, Paula Vanessa Pedron Oltramari-Navarro, Thais Maria Freire Fernandes, Giovani Fidelis de Oliveira, Ana Claudia de Castro Ferreira Conti, Marcio Rodrigues de Almeida, and Renato Rodrigues de Almeida. "Comparison of manual, digital and lateral CBCT cephalometric analyses." Journal of Applied Oral Science 21, 167-176.(2013).

11-Nishimoto, Soh, Yohei Sotsuka, Kenichiro Kawai, Hisako Ishise, and Masao Kakibuchi. Personal computer-based cephalometric landmark detection with deep learning, using cephalograms on the internet. Journal of Craniofacial Surgery, 30(1), 91-95 (2019).

12-Pendleton, Scott Drew, Hans Andersen, Xinxin Du, Xiaotong Shen, Malika Meghjani, You Hong Eng, Daniela Rus, and Marcelo H. Ang. Perception, planning, control, and coordination for autonomous vehicles. Machines, 5(1), 6 (2017).

13-Roden-Johnson, Dustin, Jeryl English, and Ronald Gallerano. Comparison of hand-traced and computerized cephalograms: landmark identification, measurement, and superimposition accuracy. American journal of orthodontics and dentofacial orthopedics, 133(4), 556-564. (2008).

14-Silva, Thaísa Pinheiro, Mariana Mendonça Hughes, Liciane dos Santos Menezes, Maria de Fátima Batista de Melo, Wilton Mitsunari Takeshita, and Paulo Henrique Luiz de Freitas."Artificial intelligence-based cephalometric landmark annotation and measurements according to Arnett's analysis: can we trust a bot to do that?." Dentomaxillofacial Radiology 20200548.(2021).

15-Wong SH, Al-Hasani H, Alam Z and Alam A. Artificial intelligence in radiology: how will we be affected?. European radiology, 29(1), 141-143 (2019).

16-Yu, Sang-Hun, Dong-Seok Nahm, and Seung-Hak Baek. Reliability of landmark identification on monitor-displayed lateral cephalometric images. American Journal of Orthodontics and Dentofacial Orthopedics, 133(6), 790-e1. (2008).