applied sciences

MDPI

Article

Verification of Convolutional Neural Network Cephalometric Landmark Identification

Moshe Davidovitch [1,*], Tatiana Sella-Tunis [1], Liat Abramovicz [1], Shoshana Reiter [2], Shlomo Matalon [3] and Nir Shpack [1]

[1] Department of Orthodontics, Maurice and Gabriella Goldschlager School of Dental Medicine, Tel Aviv University, Tel Aviv 6997801, Israel
[2] Department of Oral Pathology, Oral Medicine, and Maxillofacial Imaging, Sackler Faculty of Medicine, Maurice and Gabriella Goldschlager School of Dental Medicine, Tel Aviv University, Tel Aviv 6997801, Israel
[3] Department of Oral Rehabilitation, The Goldschleger School of Dental Medicine, Sackler Faculty of Medicine, Tel Aviv 6997801, Israel
* Correspondence: davidom@tauex.tau.ac.il

Abstract: Introduction: The mass-harvesting of digitized medical data has prompted their use as a clinical and research tool. The purpose of this study was to compare the accuracy and reliability of artificial intelligence derived cephalometric landmark identification with that of human observers. **Methods**: Ten pre-treatment digital lateral cephalometric radiographs were randomly selected from a university post-graduate clinic. The x- and y-coordinates of 21 (i.e., 42 points) hard and soft tissue landmarks were identified by 6 specialists, 19 residents, 4 imaging technicians, and a commercially available convolutional neural network artificial intelligence platform (CephX, Orca Dental, Hertzylia, Israel). Wilcoxon, Spearman and Bartlett tests were performed to compare agreement of human and AI landmark identification. **Results**: Six x- or y-coordinates (14.28%) were found to be statistically different, with only one being outside the 2 mm range of acceptable error, and with 97.6% of coordinates found to be within this range. **Conclusions**: The use of convolutional neural network artificial intelligence as a tool for cephalometric landmark identification was found to be highly accurate and can serve as an aid in orthodontic diagnosis.

Keywords: artificial intelligence; convolutional neural networks; lateral cephalometric radiographs; diagnostics

Citation: Davidovitch, M.; Sella-Tunis, T.; Abramovicz, L.; Reiter, S.; Matalon, S.; Shpack, N. Verification of Convolutional Neural Network Cephalometric Landmark Identification. *Appl. Sci.* **2022**, *12*, 12784. https://doi.org/10.3390/app122412784

Academic Editor: Maria Filomena Botelho

Received: 5 November 2022
Accepted: 29 November 2022
Published: 13 December 2022

1. Introduction

The delineation and representation of human facial form has found expression throughout human existence. Ancient Egyptian and Greek cultures developed methods mathematical and otherwise, such as anthropometrics, for this purpose [1]. Until Roentgen's first report of X-rays in 1895 [2], which also were first used in dentistry in the same year [3], these can be summarized as attempts to quantify faces and bodies according to externally visualized cues and proportions. X-rays now allowed for visualization, albeit two-dimensional, of the internal hard tissues which support and provide the form outwardly seen. Broadbent, Wingate and Hofrath codified its place in orthodontics when they independently developed cranium orienting cephalostats and methods to standardize cephalometric radiology, in 1931 [4,5]. Brodie with Downs [6,7] initiated the emergence of a plethora of cephalometric analyses which continue to be used as diagnostic tools to quantify craniofacial characteristics [8–17].

Proficiency in radiographic anatomic landmark identification has facilitated patient diagnosis and treatment evaluation and a significant researcher tool. However, the expanded use of this method also revealed its inherent limitations as initially described by Graber [18], and ensued with descriptions of these according to errors due to superimposition of intervening structures, magnification of structures closer to the X-ray source and

poor observer landmark identification reliability, all of which limit the effectiveness of this tool [19–22].

The advent of computed tomography (CT) in 1973, provided a one-to-one 3D radiographic image resolving the aforementioned shortcomings [23]; however, this method was appropriate for structures larger than teeth or jaws and required much higher doses of radiation than conventional clinical radiology. The introduction of the cone beam CT (CBCT) in 2001, enabled its use in dentistry. This technology now allows orthodontists to receive highly accurate representations of cranial structures, including 3D renderings, within a single diagnostic record while exposing the patient to less radiation than a conventional panoramic radiograph [24,25].

Digitization of cephalometrics has occurred together with other medical diagnostic measures. Initially, traditional radiographs were used directly or scanned into software from which an operator could manually plot the necessary points [26–30]. This approach was only feasible within an institutional setting where staff could be directed and were available to do so, until technological advances bypassed this function [31]. In addition, automatic landmark recognition from the now digitally acquired image has also been evolving [32–37]. The availability of the needed volume of pertinent digital data has catalyzed the development of various modes of machine learning and artificial intelligence (AI), which are being applied to perform autonomous landmark recognition [38–53].

The most commonly used machine learning method of image recognition and classification are neural networks. These are currently applied in identification of objects, faces, traffic signs and to generate vision in self-driving cars, etc. [54]. Convolutional neural networks (CNNs), like neural networks, are made up of "neurons" with learnable weights and biases. Each neuron receives several inputs, takes a weighted sum over them, passes it through an activation function and responds with an output [55]. In this manner, CNNs are an attempt to mimic the decision-making tasks performed by our own central nervous system [56].

The objective of the present study is to evaluate the accuracy and reliability of automatic computer-generated lateral cephalometric landmark recognition by the Algoceph® convolutional neural network (CNN) AI system (Orca Dental, Hertzliya, Israel). Agreement between human operator manual lateral cephalometric landmark identification with those automatically derived will be compared to access this. The working hypothesis of the present study is that there will be no significant difference between trained human and a taught dedicated automated artificial intelligence tool (Algoceph®) in identifying common lateral cephalometric anatomic landmarks.

2. Material and Methods

Ten digital lateral cephalometric radiographs of subjects submitting for treatment in a university post-graduate orthodontic clinic, without congenital craniofacial/dental anomalies, history of facial trauma and a fully erupted permanent dentition, excluding third molars, were randomly selected. The digital radiograph picture size was of a magnification of 1.08–1.13, resolution (digital) max 5.7 lp/mm, image field (digital) 24/27 × 18/30 cm and image pixel size 48 µm.

All digital radiographs were reviewed and data points selected manually by 30 operators (7 experienced orthodontic faculty members, 9 third year and 10 first year orthodontic residents, and 4 imaging center technicians), and automatically using Algoceph® point detection AI (Orca Dental, Hertzliya, Israel). A total of 21 commonly referenced soft and hard tissue lateral cephalometric points were selected (Figure 1), as defined by Jacobson and Jacobson (Table 1) [57].

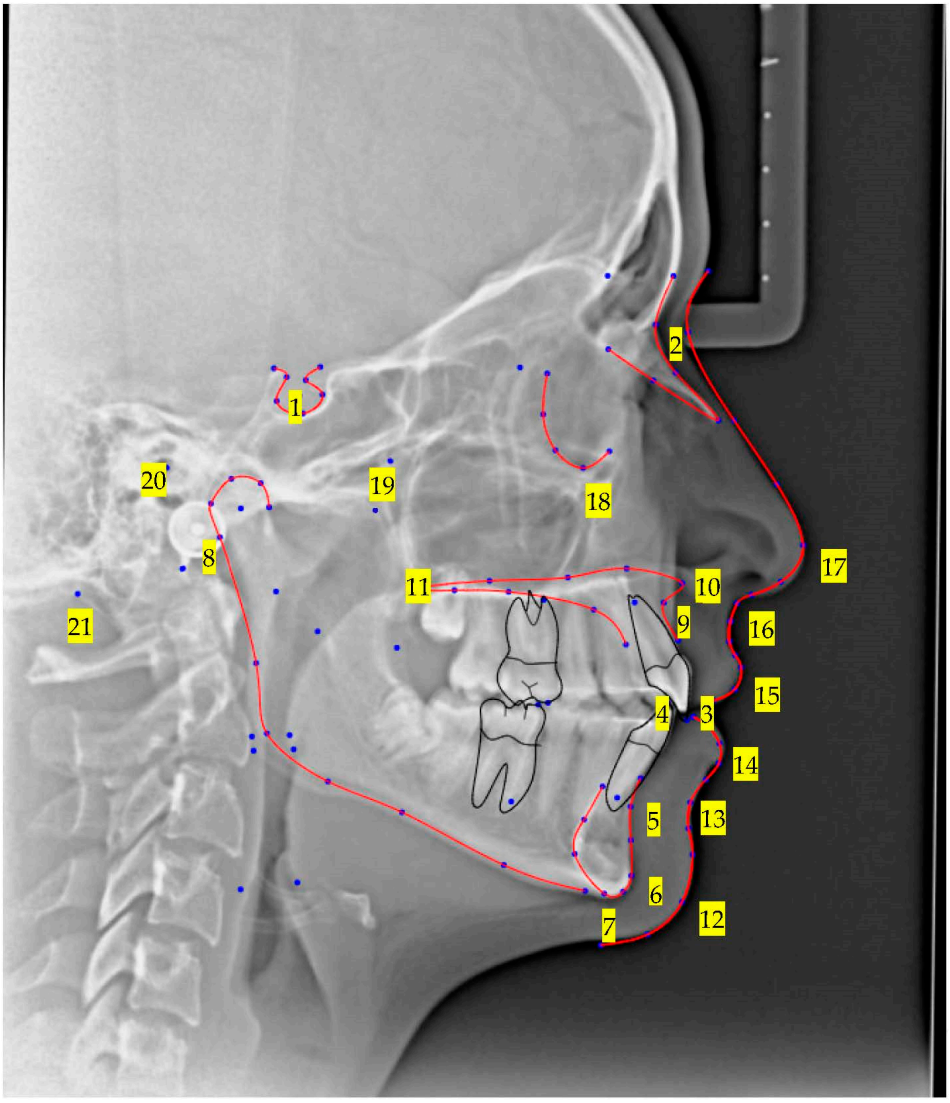

Figure 1. Localization of the 21 lateral cephlometric landmarks used in the present study as defined by Jacobson and Jacobson (see Table 1).

Table 1. Description of hard and soft tissue cranial landmarks used for comparative evaluation of human and AI detection.

	Landmark	Definition
1	Sella	Midpoint of sella turcica
2	Nasion	Most anterior point on frontonasal suture
3	Upper incisor tip (UI)	Tip of most prominent upper central incisor
4	Lower incisor tip (LI)	Tip of most prominent lower central incisor
5	B point	Deepest bony point on mandibular symphysis between pogonion and infradentale
6	Pogonion (Pog)	Most anterior point of mandibular symphysis
7	Menton	Lowest point on mandibular symphysis
8	Articulare	Junction between inferior surface of the cranial base and the posterior border of the ascending ramus of the mandible
9	A point	deepest point of premaxilla concavity bellow ANS
10	ANS	Tip of anterior nasal spine

Table 1. *Cont.*

	Landmark	Definition
11	PNS	Posterior limit of bony palate
12	Soft pogonion (Softpog)	Most anterior soft tissue point of soft chin
13	Soft B	The deepest soft tissue point between chin and subnasale
14	Lower lip	The most anterior point of lower lip
15	Upper lip	The most anterior point of upper lip
16	Subnasale	The junction where base of the columella of the nose meets the upper lip
17	Softnose	Most anterior point of nose tip
18	Orbitale	Most inferior point on the orbital margin
19	PTM	The intersection of the inferior border of the foramen rotundum with the posterior wall of the pterygomaxillary fissure
20	Porion	Most superior point of outline of external auditory meatus
21	Basale	The most inferior point on the anterior border of the foramen magnum in the midsagittal plane

A training algorithm for AI recognition of these points was performed to imprint the CNN according to Anuse and Vyas (Figure 2) [58]. Values underwent forward and backward passes according to Glorot and Bengio until target accuracy is equivalent to or decreasing slowly from ground truth [59]. Subsequent to training, algorithm detection is performed and mapped point coordinates (x, y) are superimposed onto the original image with x and y-coordinates for each point being recorded given a defined 2 mm circumscribed precision range, according to Wang CW [60]. Each point was plotted 5 times using the AI method for which no differences in point detection were found, therefore, each point was noted as a single location. Manual landmark plotting was undertaken once by each observer.

Statistical analyses were carried out using SPSS version 23 (IBM Corp, Armonk, New York, NY, USA). Non-parametric tests were required since descriptive statistics and the Kolmogorov–Smirnov tests for normality revealed that the data was not normally distributed at a level of significance where $p < 0.05$. Associations between AI and manual landmark detection were assessed using Wilcoxon Rank Sum Test (to compare AI to operator group), Spearman's Correlation (to compare findings at each delineated landmark), and Bartlett's Test (to test differences between variances of both AI and operators).

Signed informed consent for use of medical records for teaching and research purposes was obtained for each subject prior to inclusion and as a standard procedure in agreeing to undergo treatment in the university postgraduate orthodontic clinic. Ethics committee approval was sought but deemed unnecessary.

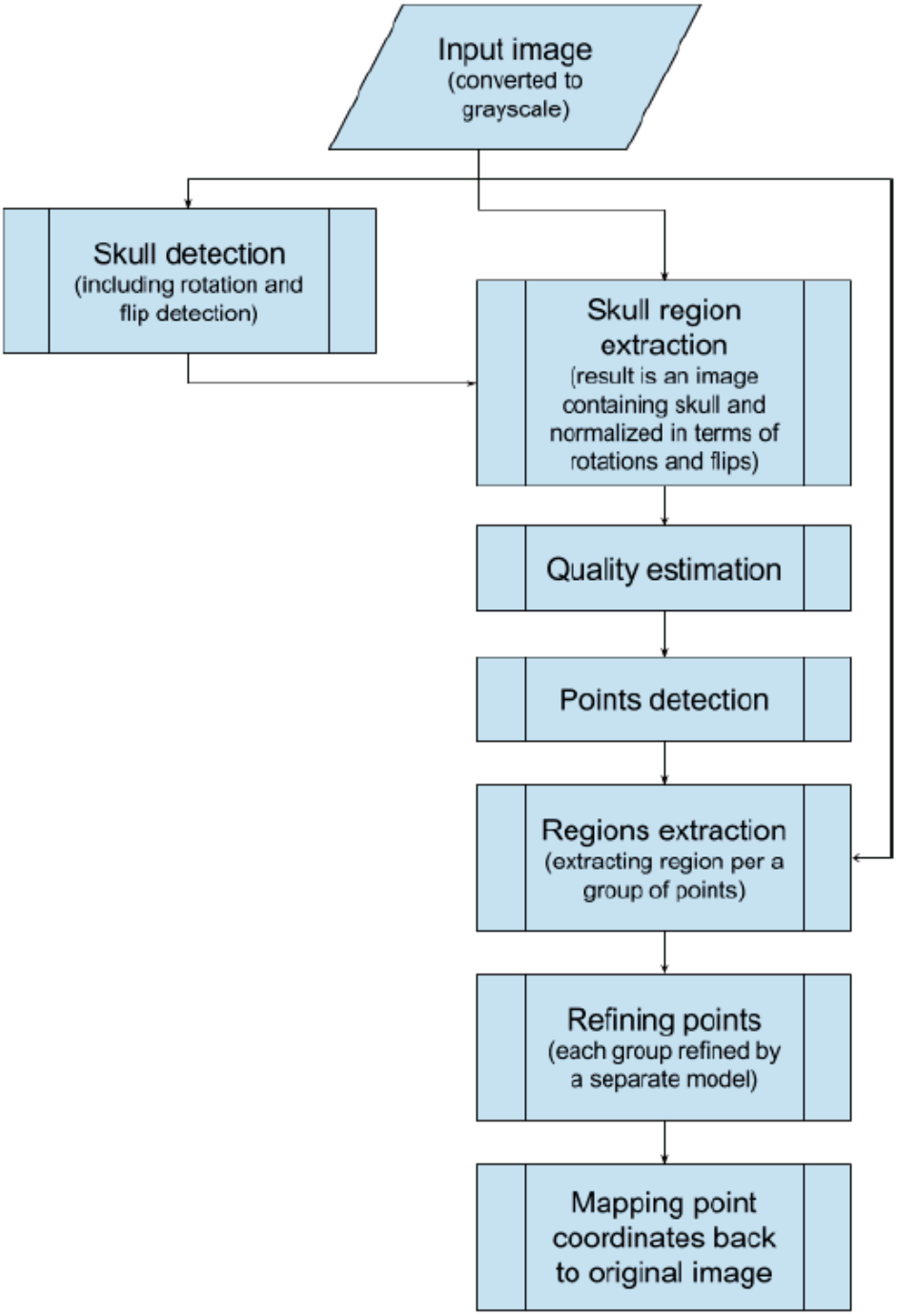

Figure 2. Description of phases of algorithm for detection of defined landmarks.

3. Results

A customized (human) operator in situ scattergram showing the envelope of detection around each evaluated anatomical landmark depicted as yellow ovals surrounding each "averaged" landmark is shown in Figure 3. Since Algoceph® was "taught" until target accuracy matched ground truth, the 5 trials performed for each landmark resulted in repeated single point localization (Figure 4).

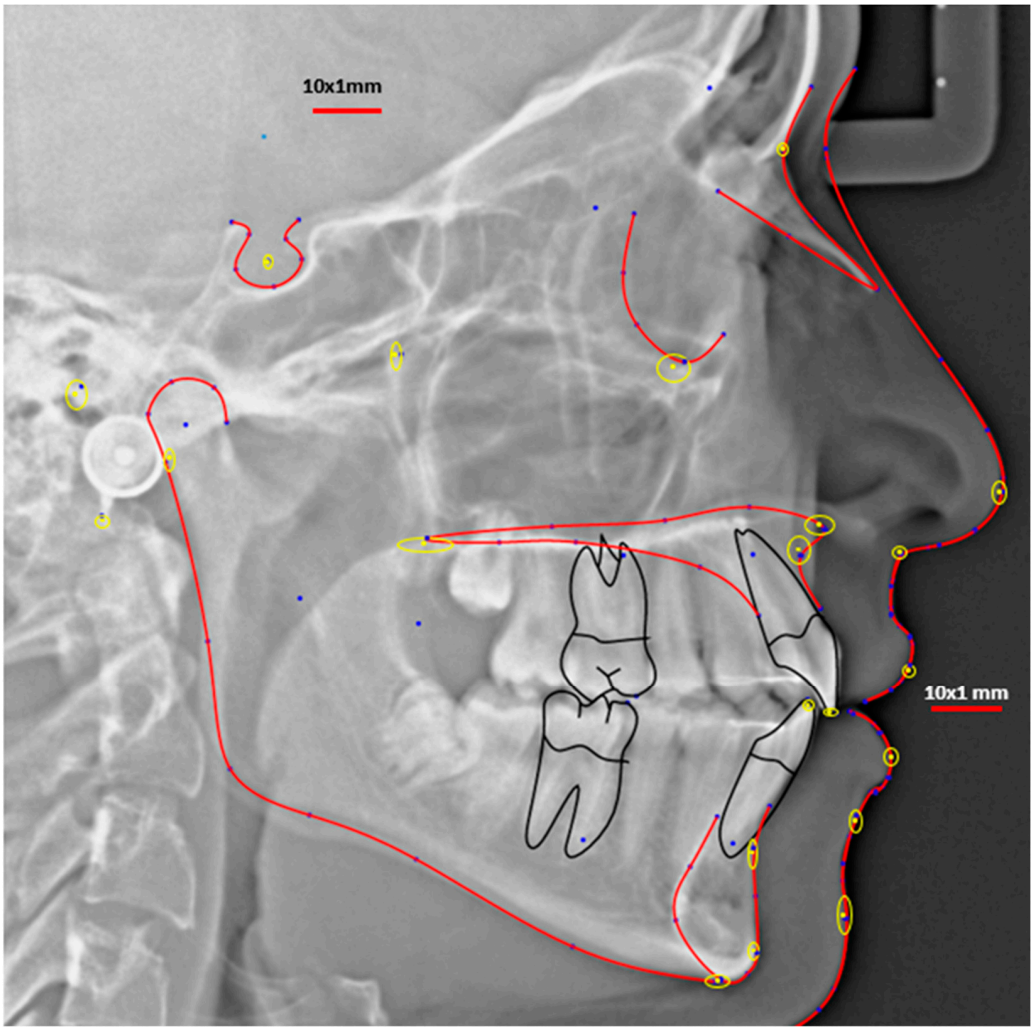

Figure 3. In situ scattergram of operator selected anatomic landmarks. The range of landmark selection is indicated by the yellow ovals calculated as 2 std dev from the mean, the mean being the determined landmark position.

Comparison between operator (avgX, avgY) and automatic (algoX, algoY) landmark detection are shown in Table 2. All landmark identification points were found to be statistically similar except; SoftpogY, UpperlipY, OrbitaleX, PTMX, PorionY and BasaleX. For SoftpogY the difference was found to be 2.67 mm $\pm$ 2.55 mm, whereas for the remaining landmarks the mean recognition error was less than 1.5 mm. Furthermore, comparison of agreement between AI landmark detection and the x and y coordinates of each landmark as selected by the human operators found that 36 out of 42 (85.72%) of these coordinates were found to be highly correlated ($r > 0.90$). The aforementioned outliers were found to be moderately correlated ($r = 0.729–0.891$) (Table 2).

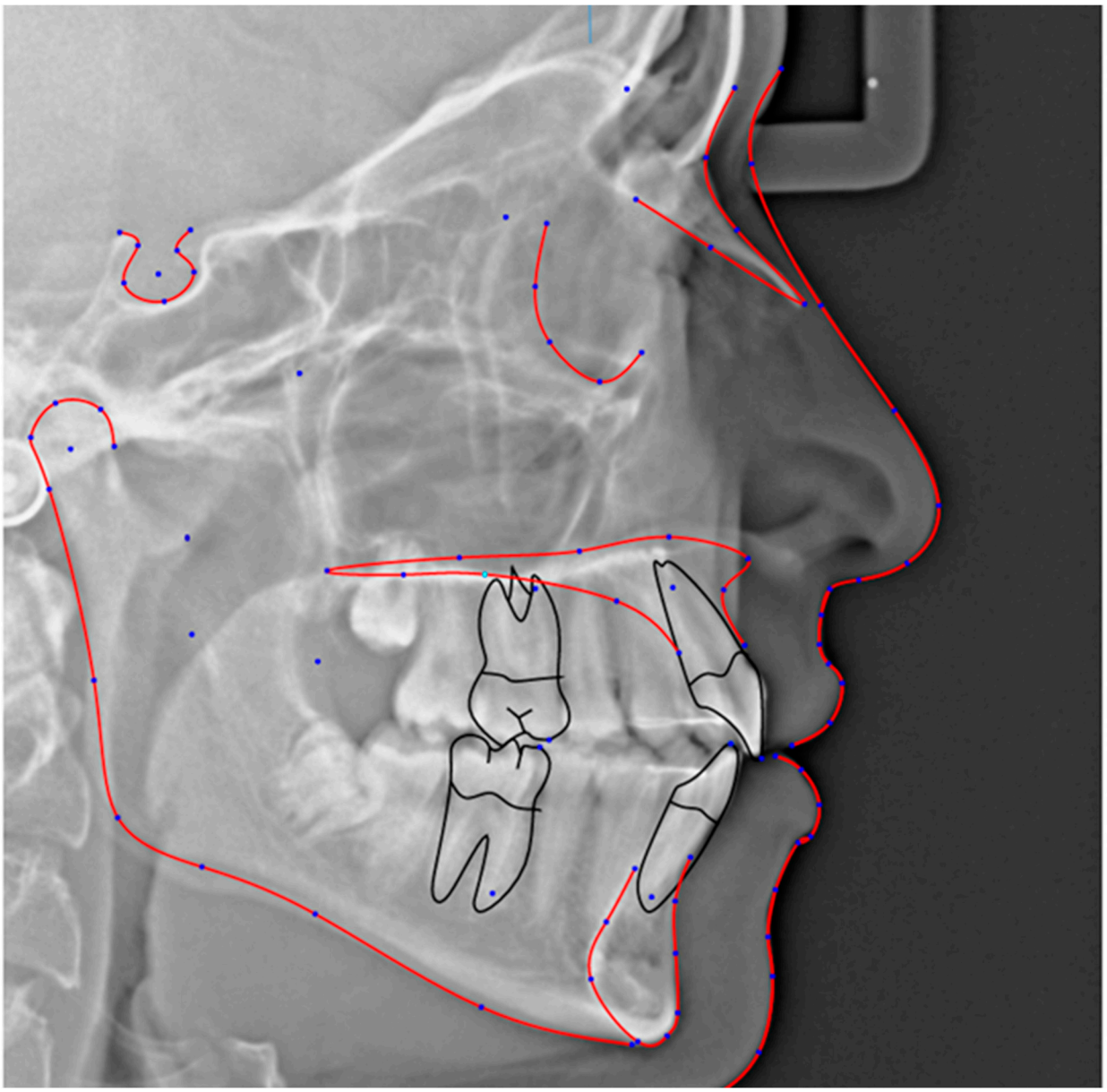

Figure 4. Example of automatic landmark location performed by the AI application of Algoceph. Note that identical outcomes of 5 separate trials provide non-scattered single point outcomes.

Table 2. Differences and correlations between Algoceph (algoX and algoY) and the operators' average (avgX and avgY). Note: significant differences are marked in bold, $p < 0.01$. ** significant Spearmen correlations. r > 0.729, $p < 0.01$.

Landmark X/Y Coordinate		Differences between Measurement Scores			Spearman Correlation	Mean Recognition Error
		Mean	**Std. Deviation**	***p***		**Mean (mm) ± SD**
1	Sella avgX	54.67	3.73	0.114	0.988 **	0.14 ± 0.39
	Sella algoX	54.81	3.74			
2	Sella avgY	139.04	6.61	0.959	0.903 **	0.05 ± 1.28
	Sella algoY	138.99	6.51			

Table 2. *Cont.*

Landmark X/Y Coordinate		Differences between Measurement Scores			Spearman Correlation	Mean Recognition Error
		Mean	**Std. Deviation**	***p***		**Mean (mm) ± SD**
3	Nasion avgX	119.94	5.75	0.285	0.952 **	0.17 ± 1.23
	Nasion algoX	119.78	6.58			
4	Nasion avY	150.26	7.08	0.878	0.976 **	0.21 ± 1.27
	Nasion algoY	150.05	6.30			
5	Ui avgX	125.41	4.39	0.799	0.912 **	0.17 ± 1.04
	Ui algoX	125.58	4.44			
6	Ui avgY	73.52	6.92	0.721	0.988 **	0.25 ± 1.02
	Ui algoY	73.26	6.33			
7	Li avgX	122.10	4.38	0.445	0.927 **	0.19 ± 1.05
	Li algoX	121.91	4.30			
8	Li avgY	75.68	6.34	0.114	0.998 **	0.39 ± 0.77
	Li algoY	76.07	5.81			
9	B point avgX	115.18	6.03	0.878	0.964 **	0.04 ± 1.13
	B point algoX	115.22	5.77			
10	B point avgY	56.65	6.23	0.921	0.988 **	0.01 ± 0.65
	B point algoY	56.65	5.91			
11	Pog avgX	116.02	7.16	0.721	0.915 **	0.11 ± 1.00
	Pog algoX	115.91	7.01			
12	Pog avgY	43.80	7.38	0.657	0.988 **	1.18 ± 0.90
	Pog algoY	42.61	7.32			
13	Menton avgX	109.56	7.03	0.891	0.976 **	0.07 ± 0.86
	Menton algoX	109.49	6.78			
14	Menton avgY	37.95	7.75	0.721	0.998 **	0.12 ± 0.71
	Menton algoY	37.83	7.42			
15	Articulare avgX	42.62	2.62	0.959	0.879 **	0.08 ± 1.11
	Articulare algoX	42.54	2.84			
16	Articulare avgY	108.06	5.97	0.799	0.915 **	0.08 ± 2.29
	Articulare algoY	108.14	4.50			
17	A point avgX	121.51	4.55	0.444	0.903 **	0.15 ± 1.03
	A point algoX	121.35	4.58			
18	A point avgY	95.44	6.24	0.721	0.964 **	0.18 ± 1.16
	A point algoY	95.26	5.22			
19	ANS avgX	125.92	4.27	0.872	0.915 **	0.88 ± 1.25
	ANS algoX	125.03	4.12			
20	ANS avgY	100.68	6.58	0.884	0.988 **	0.43 ± 1.29
	ANS algoY	100.25	5.61			
21	PNS avgX	75.83	4.27	0.782	0.867 **	0.13 ± 1.40
	PNS agoX	75.70	3.82			

Table 2. *Cont.*

Landmark X/Y Coordinate		Differences between Measurement Scores			Spearman Correlation	Mean Recognition Error
		Mean	**Std. Deviation**	***p***		**Mean (mm) ± SD**
22	PNS avgY	98.51	5.29	0.918	0.976 **	0.23 ± 1.46
	PNS algoY	98.75	4.15			
23	Soft pog avgX	126.92	6.66	0.086	0.964 **	0.48 ± 1.67
	Soft pog algoX	127.40	5.88			
24	**Soft pog avgy**	**44.06**	**6.83**	**0.022**	**0.842 ** **	2.67 ± 2.55
	Soft pog algoy	**46.74**	**5.83**			
25	Soft b avgX	126.20	5.16	0.878	0.988 **	0.05 ± 1.25
	Soft b algoX	126.15	4.68			
26	Soft b avgY	57.78	6.76	0.203	0.988 **	0.45 ± 0.98
	Soft b algoY	58.24	5.99			
27	Lower lip avgX	134.90	4.40	0.959	0.891 **	0.04 ± 1.05
	Lower lip algoX	134.95	4.42			
28	Lower lip avgY	68.75	7.41	0.721	0.998 **	0.03 ± 0.88
	Lower lip algoY	68.79	6.64			
29	Upper lip avgX	137.73	4.58	0.169	0.939 **	0.31 ± 0.96
	Upper lip algoX	137.41	4.83			
30	**Upper lip avgY**	**82.05**	**7.28**	**0.017**	**0.964 ** **	1.11 ± 1.16
	Upper lip algoY	**83.17**	**6.53**			
31	Subnasale avgX	136.32	4.44	0.541	0.915 **	0.10 ± 1.33
	Subnasale algoX	136.43	4.75			
32	Subnasale avgY	96.84	7.21	0.386	0.964 **	0.35 ± 1.42
	Subnasale algoY	96.48	6.05			
33	Soft nose avgX	150.25	5.35	0.381	0.976 **	0.30 ± 1.27
	Soft nose algoX	150.55	5.96			
34	Soft nose avgY	108.63	8.31	0.918	0.975 **	0.01 ± 0.75
	Soft nose algoY	108.63	7.65			
35	**Orbitale avgX**	**105.67**	**3.90**	**0.037**	**0.976 ** **	1.07 ± 1.29
	Orbitale algoX	**106.74**	**4.56**			
36	Orbitale avgY	123.12	6.63	0.878	0.915 **	0.16 ± 1.09
	Orbitale algoY	122.96	6.10			
37	**PTM avgX**	**70.19**	**4.03**	**0.028**	**0.939 ** **	0.99 ± 0.98
	PTM algoX	**71.19**	**4.38**			
38	PTM avgY	123.11	6.27	0.241	0.927 **	0.98 ± 1.95
	PTM algoY	124.10	5.00			
39	Porion avgX	32.75	2.57	0.285	0.729 **	0.64 ± 1.49
	Porion algoX	32.11	3.25			
40	**Porion avgY**	**120.08**	**4.38**	**0.036**	**0.830 ** **	1.14 ± 1.41
	Porion algoY	**121.23**	**4.27**			

Table 2. *Cont.*

Landmark X/Y Coordinate		Differences between Measurement Scores			Spearman Correlation	Mean Recognition Error
		Mean	**Std. Deviation**	***p***		**Mean (mm) ± SD**
41	**Basale avgX**	**35.89**	**3.15**	**0.005**	**0.903 ****	1.03 ± 0.90
	Basale algoX	**34.86**	**3.36**			
42	Basale avgY	100.71	5.18	0.959	0.976 **	0.02 ± 1.20
	Basale algoY	100.74	4.83			

Computer aggregated correlation between AI and human observations are shown in Figure 5. It can be seen that absolute overlap between scores was found (r = 0.99, $p < 0.001$) (Figure 5). Bartlett's Test to show differences in variances between AI and observers showed this to be small for both AI ($\chi^2 = 2.98$, $p = 0.98$), and operators ($\chi^2 = 2.72$, $p = 0.96$). These results indicate that disparity between scores for each point is similar without regard to landmark chosen or method of location (AI or operator) (Figure 6a,b).

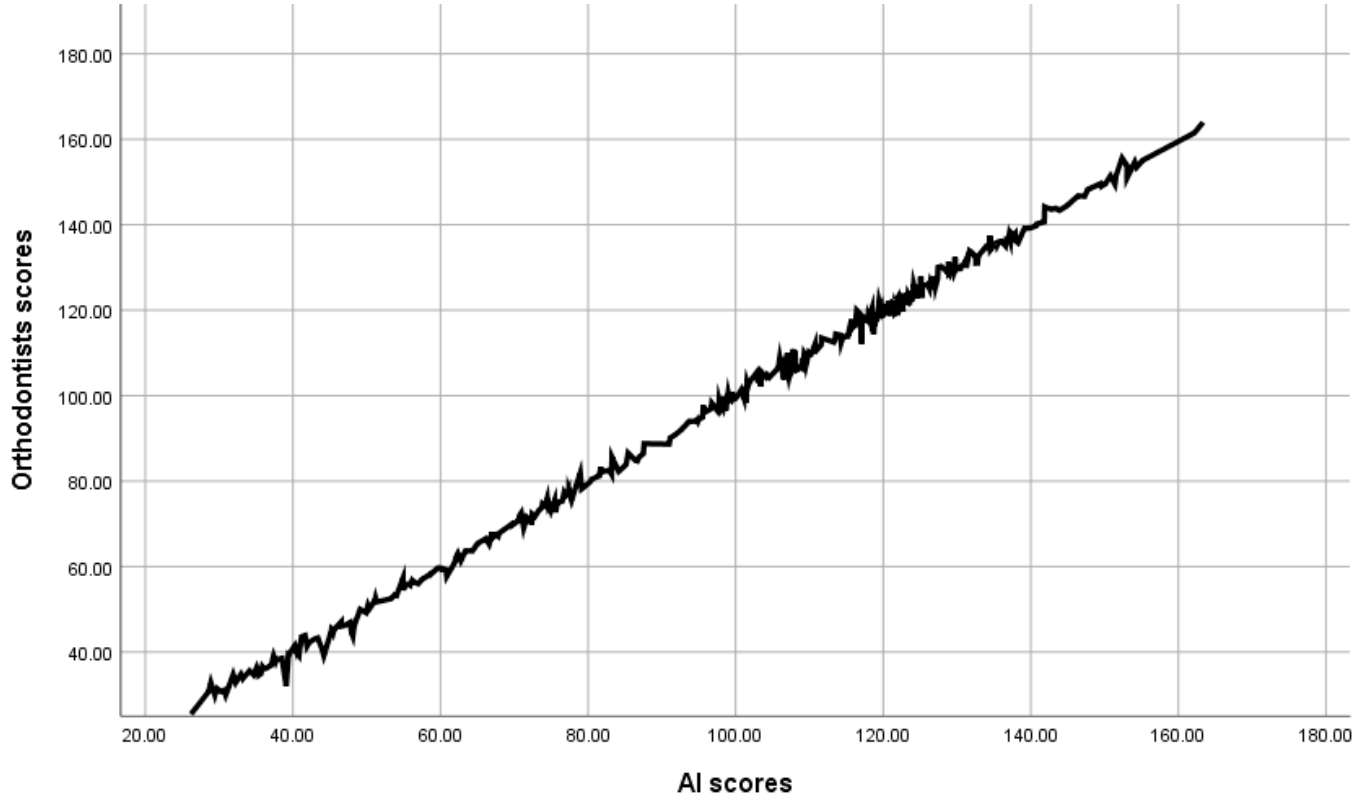

Figure 5. Correlation between general AI and observer scores.

Mean-Algo Y Differences	Sella	Nasion	UCI	LCI	B Point	Pogonion	Menton	Articulare	A Point	ANS	PNS
Mean-Algo 1	0.04	0.07	-0.13	0.19	-0.38	0.01	-0.14	-2.67	-0.52	-0.46	-0.63
Mean-Algo 2	-0.08	0.66	-0.83	0.23	0.55	-0.24	-0.05	3.65	1.66	0.88	0.46
Mean-Algo 3	0.37	0.95	0.39	0.61	0.46	0.34	0.62	1.11	-0.43	0.77	0.16
Mean-Algo 4	0.03	0.23	0.44	0.16	1.03	-0.60	3.40	2.93	2.02	0.03	-0.41
Mean-Algo 5	0.17	0.56	0.36	-0.92	-0.70	0.55	0.08	0.63	1.30	1.65	-0.66
Mean-Algo 6	-0.80	-0.28	-0.44	-0.52	-0.86	0.18	-0.77	0.27	-1.13	-1.11	0.40
Mean-Algo 7	0.19	-0.67	-0.41	-0.86	0.67	0.23	0.76	0.75	0.31	1.28	-0.46
Mean-Algo 8	0.84	0.16	0.05	-0.26	-0.03	0.78	0.44	-0.55	0.32	-0.29	1.33
Mean-Algo 9	-0.64	0.81	-0.01	-0.29	0.03	0.86	-0.51	-0.39	-0.15	-0.29	0.00
Mean-Algo 10	-0.95	-0.48	-0.18	-0.09	-0.18	2.39	0.02	0.41	-0.89	-0.50	-1.86
Absolutes Sum	4.11	4.87	3.22	4.14	4.88	6.19	6.78	13.33	8.73	7.27	6.38
Avr Absolute dif	0.41	0.49	0.32	0.41	0.49	0.62	0.68	1.33	0.87	0.73	0.64

Mean-Algo Y Differences	Soft Pogonion	Soft B	Lower Lip*	Upper Lip*	Subnasale	Soft Nose*	Orbitale	PTM	Porion	Basale
Mean-Algo 1	0.27	-0.60	-0.02	-0.45	-0.59	0.05	-0.62	0.86	-0.72	-0.28
Mean-Algo 2	-0.55	0.20	-0.17	-0.27	0.60	0.87	-0.22	0.38	-0.72	1.11
Mean-Algo 3	0.04	-0.49	0.33	0.42	1.03	0.09	0.19	-1.14	0.13	-0.53
Mean-Algo 4	-0.10	0.50	1.60	-0.33	-0.07	1.38	2.40	1.54	1.16	1.81
Mean-Algo 5	0.48	-1.49	-0.13	-1.54	0.37	-0.26	-0.77	0.01	-1.86	-1.17
Mean-Algo 6	0.09	-0.91	-0.43	0.15	0.14	-1.00	-1.06	-1.57	0.19	1.45
Mean-Algo 7	-4.35	-0.44	-0.65	0.15	0.09	0.29	0.93	-1.51	-0.75	0.37
Mean-Algo 8	-0.31	-0.49	0.13	-1.29	0.25	-0.42	0.93	0.02	1.28	0.04
Mean-Algo 9	0.89	0.74	0.93	1.17	0.01	-0.04	0.49	0.55	-1.60	-0.39
Mean-Algo 10	0.64	-1.58	-0.45	-1.49	-0.40	-0.43	-0.34	0.55	-1.60	-0.39
Absolutes Sum	7.71	7.43	4.83	7.25	3.53	4.81	7.96	8.13	10.01	7.53
Avr Absolute dif	0.77	0.74	0.48	0.72	0.35	0.48	0.80	0.81	1.00	0.75

*= most protruding
UCI= upper centreal incisor edge
LCI= lower centrel incisor edge

(**a**)

Mean-Algo X Differences	Sella	Nasion	UCI	LCI	B Point	Pogonion	Menton	Articulare	A Point	ANS	PNS
Mean-Algo 1	0.04	0.34	-0.23	-0.24	0.25	0.60	0.38	-0.46	0.50	0.99	0.28
Mean-Algo 2	-0.20	0.16	0.64	0.04	0.69	0.60	-0.14	0.27	0.78	0.28	-0.16
Mean-Algo 3	-0.44	0.09	-0.36	0.10	-0.43	-0.52	-0.25	-0.53	-0.19	-0.70	2.58
Mean-Algo 4	0.10	0.55	0.69	-0.03	0.14	-0.69	-0.12	2.43	1.86	2.68	1.56
Mean-Algo 5	-0.15	-0.27	-0.07	0.10	-0.39	-0.53	-0.80	1.25	-0.55	-1.02	-0.31
Mean-Algo 6	-0.52	0.15	-0.86	0.45	-0.46	-0.30	0.78	-0.11	-0.58	0.49	-0.33
Mean-Algo 7	-0.25	0.77	0.98	0.90	0.38	0.34	0.26	0.11	0.80	0.00	0.50
Mean-Algo 8	-0.01	0.43	0.58	0.16	0.03	0.27	0.46	-0.82	0.23	-0.69	0.00
Mean-Algo 9	-0.67	0.43	0.04	0.23	-0.22	-0.18	0.21	0.10	0.21	-0.54	-0.11
Mean-Algo 10	-0.54	-0.27	-0.24	-1.21	-1.56	-0.60	-0.01	-0.54	-0.13	-0.08	-0.81
Absolutes Sum	2.92	3.47	4.70	3.46	4.56	4.63	3.42	6.63	5.82	7.46	6.64
Avr Absolute dif	0.29	0.35	0.47	0.35	0.46	0.46	0.34	0.66	0.58	0.75	0.66

Mean-Algo X Differences	Soft Pogonion	Soft B	Lower Lip*	Upper Lip*	Subnasale	Soft Nose*	Orbitale	PTM	Porion	Basale
Mean-Algo 1	-0.54	0.74	-0.45	0.52	0.37	-0.67	-0.42	0.11	0.08	1.73
Mean-Algo 2	-0.12	-0.22	0.74	0.62	0.36	0.61	0.81	-1.28	2.20	0.51
Mean-Algo 3	-0.90	-0.14	-0.35	-0.88	-0.78	0.07	-1.38	0.35	-0.50	-0.26
Mean-Algo 4	-0.38	0.79	1.75	0.19	-0.35	-0.67	0.89	0.39	0.24	0.64
Mean-Algo 5	-0.70	-0.17	0.34	-0.18	-1.12	-0.53	-2.30	-1.31	-0.96	0.56
Mean-Algo 6	-1.34	-1.09	0.30	-0.89	-0.53	-0.07	0.44	-4.43	-0.82	-0.02
Mean-Algo 7	-0.03	0.18	-0.12	0.70	0.41	0.49	0.73	-1.06	0.40	0.84
Mean-Algo 8	0.01	0.40	-0.17	0.39	0.49	-0.40	-1.05	-0.39	-0.04	0.63
Mean-Algo 9	-0.03	-0.22	0.70	0.21	-0.73	0.01	-0.40	0.30	1.89	0.56
Mean-Algo 10	-0.59	-1.77	-1.55	-0.42	0.43	0.84	-0.48	0.11	0.52	-0.16
Absolutes Sum	4.64	5.71	6.47	4.99	5.57	4.37	8.92	9.74	7.64	5.89
Avr Absolute dif	0.46	0.57	0.65	0.50	0.56	0.44	0.89	0.97	0.76	0.59

*= most protruding
UCI= upper centreal incisor edge
LCI= lower centrel incisor edge

(**b**)

Figure 6. (**a**) Differences between the Algoceph and observer mean for the 21 landmarks of 10 different cephalograms in the vertical y plane. (**b**) Differences between the Algoceph and the and observer mean for the 21 landmarks of 10 different cephalograms in the horizontal x plane.

When the 4 sub-categories of observers were compared using a Repeated Measures Analysis of Variance no significant differences in agreement in landmark localization existed for nearly all points (p ranged between 0.063 and 0.913) (Table 2). The only exceptions were found to be PNSX and SoftnoseY. When the location of these points was examined with regard to their x,y-coordinates, it was found that differences for PNSX were $F(1,9) = 10.44$, $p = 0.01$, and that first-year residents (M = 76.33, SD = 3.63) and third-year residents (M = 76.89, SD = 4.19) differed more (higher) in the vertical axes compared to specialists (M = 74.93, SD = 4.19) and technicians (M = 75.00, SD = 3.80). For Softnose Y, $F(1,9) = 9.80$, $p = 0.01$, it was found that third-year residents (M = 44.32, SD = 7.03), specialists (M = 44.52, SD = 6.56) and imaging technicians (M = 44.71, SD = 6.12) differed on the y axes (lower) compared to first-year residents (M = 42.92, SD = 6.88).

4. Discussion

The use of AI as a management tool has already found an application in orthodontic patient selection/referral within public health care systems [61]. The use of deep learning in AI as a diagnostic tool to perform cephalometric landmark identification can potentially eliminate intra/inter-observer variation as well as vastly reduce the time invested and decreased efficiency in performing this task manually [62]. In order to do so, there have been several methods of application of deep learning in landmark detection [63]. The method used in the present study is regression-based deep learning as described by Noothout et al. [64].

Yue et.al. proposed as correct/acceptable a ±2 mm differential in human versus computerized lateral cephalometric landmark identification, where if more than 20% of the total localizations were unacceptable equated as a failed comparison [43]. Based on this definition of "correct", the purpose of the present study was to compare the identification of lateral cephalometric landmark accuracy by the latest generation of artificial intelligence to that of human observers with varied amounts of aggregated clinical experience. The amount of error in landmark identification was the difference between that produced by experienced observers and by AI.

Twenty-one cranial landmarks (Table 1) were registered according to automatic point detection (Algoceph®). Human observer identification was plotted as an envelope surrounding the former which was taken as an origin (0,0), with each such point delineated into its x,y coordinates. It was found that in 36 out of 42 (85.7%) coordinates no statistically significant differences were found between AI and all human observers (Table 2 and Figure 6a,b).

It was found that coordinates of six anatomic landmarks showed statistically significant differences between human and automatic identification; SoftpogY, whose measurement error was clinically insignificant (2.67 mm ± 2.55), UpperlipY which was clinically acceptable (1.11 mm ± 1.16), as were OrbitaleX (1.07 mm ± 1.29), PTMX (0.99 mm ± 0.98), PorionY (1.14 mm ± 1.41), and BasaleX (1.03 mm ± 0.90). These landmarks were also previously described as highly prone to erroneous identification by Baumrind [19,20], however, the differences found in the present study were small enough to deem them diagnostically relevant (Table 2).

Adoption of digital technologies in orthodontics were initially applied to data storage and automated clear aligner production. Current efforts are to "teach" these tools to perform diagnostics and treatment planning [65]. Doing so stipulates the understanding that the reliability of any measurement derived from radiographic analysis depends on the reproducibility in the identification of defined landmarks. Factors such as the quality of the radiographs (contrast, scaling, etc.), and operator reliability have been shown to influence the magnitude of identification error. Earlier studies of the performance of AI to identify (fewer) cephalometric landmarks reported far lower levels of accuracy than those found in the present study using the CephX Algo method (Table 3) [36,37,44–46]. A more recent report by Kunz et al. described findings in agreement with those of the present study [66]. Taken together; these suggest that the CephX Algo method has reduced

the above sources of measurement error to the extent that its output can be accepted as diagnostically accurate.

Table 3. Mean measurement error (mm) of human vs. AI from early studies.

Landmark	Liu et al. [19]	Hutton et al. [8]	Saad et al. [20]	Tanikawa et al. [21]	Rudolph et al. [7]	CephX Algo
Sella	0.94	5.5	3.24	2.1	5.06	0.148
Nasion	2.32	5.6	2.95	1.7	2.57	0.27
Orbitale	5.28	5.5	3.4	2.24	2.46	1.08
Porion	2.43	7.3	3.48	3.63	5.67	1.3
ANS	2.9	3.8	2.7	2.32	2.64	0.97
Point A	4.29	3.3	2.54	2.13	2.33	0.23
Point B	3.96	2.6	2.22	3.12	1.85	0.04
Pogonion	2.53	2.7	3.65	1.91	1.85	1.18
Menton	1.9	2.7	4.4	1.59	3.09	0.12
UI	2.36	2.9	3.65	1.78	NAD	0.3
LI	2.86	NAD	3.14	1.81	NAD	0.35

5. Conclusions

The convolutional neural network artificial intelligence method for determining lateral cephalometric landmark identification was found to be significantly correlated to human identification of 21 lateral cephalometric radiographic anatomic landmarks. This implies that this application of AI can be used to reduce the time expenditure and human error involved in performing this task manually.

Author Contributions: Conceptualization, writing—original draft preparation, M.D.; Methodology, formal analysis, T.S.-T.; Data curation, L.A.; Data curation, S.R.; Investigation, supervision, S.M.; Project administration, writing—review and editing, N.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review and approval were waived for this study due to all medical records included in the data were derived from subjects that provided signed informed consent permitting these records to be used for teaching and research purposes without any additional radiographic exposure being required to conduct this study, and proper measures taken to protect the personal information associated with each subject. A statement to this regard is present in the "Methods" section of the manuscript.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: All data is available upon request from the correspondence author.

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

References

1. Cuff, T. Biometric method, past, present, and future. In *Historical Anthropometrics*; Irwin, J.O., Ed.; Ashgate: Aldershot, UK, 1998; pp. 363–375.
2. Tubiana, M. Wilhelm Conrad Röntgen and the discovery of X-rays. *Bull. Acad. Natl. Med.* **1996**, *180*, 97–108. [PubMed]
3. Forrai, J. History of X-ray in dentistry. *Rev. Clín. Pesq. Odontol.* **2007**, *3*, 205–211.
4. Broadbent, B. A new X-ray technique and its application to orthodontia. *Angle Orthod.* **1931**, *1*, 45–66.
5. Trenouth, M.J.; Gelbier, S. Development of the norm concept in orthodontics. *Dent. Hist.* **2012**, *56*, 39–52.
6. Brodie, A.G.; Downs, W.B.; Goldstein, A.; Myer, E. Cephalometric appraisal of orthodontic results. *Angle Orthod.* **1938**, *8*, 261–265.
7. Downs, W.B. Variations in facial relationship. *Am. J. Orthod.* **1948**, *34*, 813–840. [CrossRef] [PubMed]
8. Steiner, C.C. Cephalometrics for you and me. *Am. J. Orthod.* **1953**, *39*, 729–755. [CrossRef]
9. Tweed, C.H. The Frankfort-Mandibular Incisor Angle (FMIA) In Orthodontic Diagnosis, Treatment Planning and Prognosis. *Angle Orthod.* **1954**, *24*, 121–169.
10. Sassouni, V. A roentgenographic cephalometric analysis of cephalo-facio-dental relationships. *Am. J. Orthod.* **1955**, *41*, 735–764. [CrossRef]

11. Broadbent, B.H., Sr.; Broadbent, B.H., Jr.; Golden, W.H. *Bolton Standards of Dentofacial Developmental Growth*; C. V. Mosby: St. Louis, MO, USA, 1975.
12. Moorrees, C.F.; Lebret, L. The mesh diagram and cephalometrics. *Angle Orthod.* **1962**, *32*, 214–231.
13. Björk, A. Prediction of mandibular growth rotation. *Am. J. Orthod.* **1969**, *55*, 585–599. [CrossRef] [PubMed]
14. Jarabak, J.R.; Fizzell, J.A. *Technique and Treatment with Light-Wire Edgewise Appliance*, 2nd ed.; The C. V. Mosby Company: St. Louis, MO, USA, 1972.
15. Wits, A.J. The Wits appraisal of jaw disharmony. *Am. J. Orthod.* **1975**, *67*, 125–138.
16. Ricketts, R.M.; Bench, R.; Gugino, C.; Hilgers, J.; Schulhof, R. Visual treatment objective or V.T.O. In *Bioprogressive Therapy*; Rocky Mountain Orthodontics: Denver, CO, USA, 1979; pp. 35–54.
17. McNamara, J.A. A method of cephalometric evaluation. *Am. J. Orthod.* **1984**, *86*, 449–469. [CrossRef] [PubMed]
18. Graber, T.M. Problems and limitations of cephalometric analysis in orthodontics. *J. Am. Dent. Assoc.* **1956**, *53*, 439–454. [CrossRef] [PubMed]
19. Baumrind, S.; Frantz, R.C. The reliability of head film measurements. 1. Landmark identification. *Am. J. Orthod.* **1971**, *60*, 111–127. [CrossRef]
20. Baumrind, S. Toward a general model for clinical craniofacial research. In *Essays in Honor of Robert E*; Hunter, W.S., Carlson, D.S., Eds.; Moyers. Monograph 24 Craniofacial Growth Series; Center for Human Growth and Development, University of Michigan: Ann Arbor, MI, USA, 1991; pp. 37–71.
21. Trpkova, B.; Major, P.; Prasad, N.; Nebbe, B. Cephalometric landmarks identification and reproducibility: A meta analysis. *Am. J. Orthod. Dentofac. Orthop.* **1997**, *112*, 165–170. [CrossRef]
22. Hans, M.G.; Palomo, J.M.; Valiathan, M. History of imaging in orthodontics from Broadbent to cone-beam computed tomography. *Am. J. Orthod. Dentofac. Orthop.* **2015**, *148*, 914–921. [CrossRef]
23. Ambrose, J.; Hounsfield, G. Computerized transverse axial tomography. *Br. J. Radiol.* **1973**, *46*, 148–149. [CrossRef]
24. van Vlijmen, O.J.C.; Kuijpers, M.A.R.; Berge, S.J.; Schols, J.G.J.H.; Maal, T.J.J.; Breuning, H.; Kuijpers-Jagtman, A.M. Evidence supporting the use of cone-beam computed tomography in orthodontics. *J. Am. Dent. Assoc.* **2012**, *143*, 241–252. [CrossRef]
25. Ludlow, J.B.; Walker, C. Assessment of phantom dosimetry and image quality of i-CAT FLX cone-beam computed tomography. *Am. J. Orthod. Dentofacial. Orthop.* **2013**, *144*, 802–817. [CrossRef]
26. Sloan, R.F. Computer applications in orthodontics. *Int. Dent. J.* **1980**, *30*, 189–200.
27. Baumrind, S.; Miller, D.M. Computer-aided headfilm analysis: The University of California San Francisco method. *Am. J. Orthod.* **1980**, *78*, 41–65. [CrossRef]
28. Richardson, A. A comparison of traditional and computerized methods of cephalometric analysis. *Eur. J. Orthod.* **1981**, *3*, 15–20. [CrossRef] [PubMed]
29. BeGole, E.A. Verification and standardization of cephalometric coordinate data. *Comput. Programs Biomed.* **1981**, *12*, 212–216. [CrossRef] [PubMed]
30. BeGole, E.A. Software development for the management of cephalometric radiographic data. *Comput. Programs Biomed.* **1981**, *11*, 175–182. [CrossRef] [PubMed]
31. Konchak, P.A.; Koehler, J.A. A Pascal computer program for digitizing lateral cephalometric radiographs. *Am. J. Orthod.* **1985**, *87*, 197–200. [CrossRef] [PubMed]
32. Cohen, A.M.; 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. [CrossRef]
33. Lévy-Mandel, A.D.; Venetsanopoulos, A.N.; Tsotsos, J.K. Knowledge-based landmarking of cephalograms. *Comput. Biomed. Res.* **1986**, *19*, 282–309. [CrossRef]
34. Parthasarathy, S.; Nugent, S.T.; Gregson, P.G.; Fay, D.F. Automatic landmarking of cephalograms. *Comput. Biomed. Res.* **1989**, *22*, 248–269. [CrossRef]
35. Cardillo, J.; Sid-Ahmed, M.A. An image processing system for locating craniofacial landmarks. *IEEE Trans. Med. Imaging* **1994**, *13*, 275–289. [CrossRef]
36. Rudolph, D.J.; Sinclair, P.M.; Coggins, J.M. Automatic computerized radiographic identification of cephalometric landmarks. *Am. J. Orthod. Dentofac. Orthop.* **1998**, *113*, 173–179. [CrossRef] [PubMed]
37. Hutton, T.J.; Cunningham, S.; Hammond, P. An evaluation of active shape models for the automatic identification of cephalometric landmarks. *Eur. J. Orthod.* **2000**, *22*, 499–508. [CrossRef] [PubMed]
38. Desvignes, M.; Romaniuk, B.; Clouard, R.; Demoment, R.; Revenu, M.; Deshayes, M.J. First steps toward automatic location of landmarks on X-ray images. In Proceedings of the 15th International Conference on Pattern Recognition ICPR-2000, Barcelona, Spain, 3–7 September 2000; pp. 275–278.
39. Innes, A.; Ciesielski, V.; Mamutil, J.; John, S. Landmark Detection for Cephalometric Radiology Images using Pulse Coupled Neural Networks. In Proceedings of the International Conference on Artificial Intelligence (IC-AI'02), Las Vegas, NV, USA, 24–27 June 2002; pp. 511–517.
40. Giordano, D.; Leonardi, R.; Maiorana, F.; Spampinato, C. Cellular neural networks and dynamic enhancement for cephalometric landmarks detection. In *Artificial Intelligence and Soft Computing*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2006; Volume 4029, pp. 768–777.

41. Cootes, T.; Taylor, C.; Cooper, D.; Graham, J. Active Shape Models-Their training and application. *Comput. Vis. Image Und.* **1995**, *61*, 38–59. [CrossRef]
42. Leonardi, R.; Giordano, D.; Maiorana, F.; Spampinato, C. Automatic Cephalometric Analysis. *Angle Orthod.* **2008**, *78*, 145–151. [CrossRef]
43. Yue, W.; Yin, D.; Li, C.; Wang, G.; Xu, T. Automated 2-D cephalometric analysis on X-ray images by a model-based approach. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 1615–1623.
44. Liu, J.; Chen, Y.; Cheng, K. Accuracy of computerized automatic identification of cephalometric landmarks. *Am. J. Orthod. Dentofac. Orthop.* **2000**, *118*, 535–540. [CrossRef]
45. Saad, A.A.; El-Bialy, A.; Kandil, A.H.; Sayed, A.A. Automatic cephalometric analysis using active appearance model and simulated annealing. In Proceedings of the International Conference on Graphics, Vision and Image Processing, Cairo, Egypt, 19–21 December 2005.
46. Tanikawa, C.; Masakazu, Y.; Kenji, T. Automated Cephalometry: System Performance Reliability Using Landmark-Dependent Criteria. *Angle Orthod.* **2009**, *6*, 1037–1046. [CrossRef]
47. Leonardi, R.; Giordano, D.; Maiorana, F. An evaluation of cellular neural networks for the automatic identification of cephalometric landmarks on digital images. *J. Biomed. Biotechnol.* **2009**, *2009*, 717102. [CrossRef]
48. Vucinic, P.; Trpovski, Z.; Scepan, I. Automatic landmarking of cephalograms using active appearance models. *Eur. J. Orthod.* **2010**, *32*, 233–241. [CrossRef]
49. Ciresan, D.C.; Meier, U.; Masci, J.; Gambardella, L.M.; Schmidhuber, J. Flexible, high performance convolutional neural networks for image classification. In Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, Barcelona, Spain, 16–22 July 2011; Volume 2.
50. Shahidi, S.; Oshagh, M.; Gozin, F.; Salehi, P.; Danaei, S. Accuracy of computerized automatic identification of cephalometric landmarks by a designed software. *Dentomaxillofac Radiol.* **2013**, *42*, 20110187. [CrossRef] [PubMed]
51. Kaur, A.; Singh, C. Automatic cephalometric landmark detection using Zernike moments and template matching. *Signal Image Video Process.* **2013**, *9*, 117–132. [CrossRef]
52. Lindner, C.; Cootes, T.F. Fully automatic cephalometric evaluation using Random Forest regression-voting. In Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI), New York, NY, USA, 16–19 April 2015.
53. Arik, S.O.; Ibragimov, B.; Xing, L. Fully automated quantitative cephalometry using convolutional neural networks. *J. Med. Imaging* **2017**, *4*, 14501. [CrossRef] [PubMed]
54. Hassoun, M.H. *Fundamentals of Artificial Neural Networks*; The MIT Press: Cambridge, MA, USA, 1995.
55. Collobert, R.; Weston, J. A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning. In Proceedings of the 25th International Conference on Machine Learning, ICML '08, Helsinki, Finland, 1 January 2008; ACM: New York, NY, USA, 2008; pp. 160–167.
56. Craig, M.; Adapa, R.; Pappas, I.; Menon, D.; Stamatakis, E. Deep graph convolutional neural networks identify frontoparietal control and default mode network contributions to mental imagery manuscript. In Proceedings of the 2018 Conference on Cognitive Computational Neuroscience, Philadelphia, PA, USA, 5–8 September 2018.
57. Jacobson AJacobson, R.L. (Ed.) *Radiographic Cephalometry: From Basics to 3D Imaging*, 2nd ed.; Quintessence Publishing Co. Limited: New Malden, UK, 2006.
58. Anuse, A.; Vyas, V. A novel training algorithm for convolutional neural network. *Complex Intell. Syst.* **2016**, *2*, 221–234. [CrossRef]
59. Glorot, X.; Yoshua, B. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS), Sardinia, Italy, 13 May 2010; Volume 9.
60. Wang, C.W.; Huang, C.T.; Hsieh, M.C.; Li, C.H.; Chang, S.W.; Li, W.C.; Vandaele, R.; Maree, R.; Jodogne, S.; Geurts, P.; et al. Evaluation and comparison of anatomical landmark detection methods for cephalometric X-Ray images: A grand challenge. *IEEE Trans. Med. Imaging* **2015**, *34*, 1890–1900. [CrossRef]
61. Mohamed, M.; Ferguson, D.J.; Venugopal, A.; Alam, M.K.; Makki, L.; Vaid, N.R. An artificial intelligence based application to optimize orthodontic referrals in a public oral healthcare system. *Semin. Orthod.* **2021**, *27*, 157–163. [CrossRef]
62. Oh, K.; Oh, l.-S.; Van Nhat Le, T.; Lee, D.-W. Deep anatomical context feature learningfor cephalometric landmark detection. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 806–817. [CrossRef]
63. Mohammad-Rahimi, H.; Nadimi, M.; Rohban, M.H.; Shamsoddin, E.; Lee, V.Y.; Motamedian, S.R. Machine learning and orthodontics, current trends and the future opportunities: A scoping review. *Am. J. Orthod. Dentfac. Orthop.* **2021**, *160*, 17–92. [CrossRef]
64. Noothout, J.M.H.; De Vos, B.D.; Wolterink, J.M.; Postma, E.M.; Smeets, P.A.M.; Takx, R.A.P. Deep learning-based regression and classification for automatic landmark localizationin medical images. *IEEE Trans. Med. Imaging* **2020**, *39*, 4011–4022. [CrossRef]
65. Retrovey, J.M. The role of AI and machine learning in contemporary orthodontics. *APOS Trends Orthod.* **2021**, *11*, 74–80. [CrossRef]
66. Kunz, F.; Stellzig-Eisenhauer, A.; Zeman, F.; Boldt, J. Artificial intelligence in orthodontics: Evaluation of a fully automated cephalometric analysis using a customized convolutional neural network. *J. Orofac. Orthop./Fortschr. der Kieferorthopädie* **2020**, *81*, 52–68. [CrossRef] [PubMed]