

Automated landmark identification on cone-beam computed tomography: *Accuracy and reliability*

Ali Ghowsi^a; David Hatcher^b; Heeyeon Suh^c; David Wile^d; Wesley Castro^e; Jan Krueger^e;
Joorok Park^f; Heesoo Oh^g

ABSTRACT

Objectives: To evaluate the accuracy and reliability of a fully automated landmark identification (ALI) system as a tool for automatic landmark location compared with human judges.

Materials and Methods: A total of 100 cone-beam computed tomography (CBCT) images were collected. After the calibration procedure, two human judges identified 53 landmarks in the x, y, and z coordinate planes on CBCTs using Checkpoint Software (Stratovan Corporation, Davis, Calif). The ground truth was created by averaging landmark coordinates identified by two human judges for each landmark. To evaluate the accuracy of ALI, the mean absolute error (mm) at the x, y, and z coordinates and mean error distance (mm) between the human landmark identification and the ALI were determined, and a successful detection rate was calculated.

Results: Overall, the ALI system was as successful at landmarking as the human judges. The ALI's mean absolute error for all coordinates was 1.57 mm on average. Across all three coordinate planes, 94% of the landmarks had a mean absolute error of less than 3 mm. The mean error distance for all 53 landmarks was 3.19 ± 2.6 mm. When applied to 53 landmarks on 100 CBCTs, the ALI system showed a 75% success rate in detecting landmarks within a 4-mm error distance range.

Conclusions: Overall, ALI showed clinically acceptable mean error distances except for a few landmarks. The ALI was more precise than humans when identifying landmarks on the same image at different times. This study demonstrates the promise of ALI in aiding orthodontists with landmark identifications on CBCTs. (*Angle Orthod.* 2022;92:642–654.)

KEY WORDS: Automated; 3D landmark identification; Accuracy; CBCT; Reliability; Landmark error

^a Adjunct Assistant Professor, Department of Orthodontics, Arthur A. Dugoni School of Dentistry, University of the Pacific, San Francisco, Calif, USA.

^b Adjunct Associate Professor, Department of Orthodontics, Arthur A. Dugoni School of Dentistry, University of the Pacific, San Francisco, Calif, USA.

^c Orthodontic Resident, Department of Orthodontics, Arthur A. Dugoni School of Dentistry, University of the Pacific, San Francisco, Calif, USA.

^d Research Scientist, President, and Chief Technology Officer, Stratovan Corporation, Davis, Calif, USA.

^e Staff Scientist, Stratovan Corporation, Davis, Calif, USA.

^f Associate Professor, Department of Orthodontics, Arthur A. Dugoni School of Dentistry, University of the Pacific, San Francisco, Calif, USA.

^g Professor and Chair, Department of Orthodontics, Arthur A. Dugoni School of Dentistry, University of the Pacific, San Francisco, Calif, USA.

Corresponding author: Dr. Heesoo Oh, Professor and Chair, Department of Orthodontics, Arthur A. Dugoni School of Dentistry, University of the Pacific, 155 Fifth Street, San Francisco, CA 94103 (e-mail: hoh@pacific.edu)

Accepted: April 2022. Submitted: December 2021.

Published Online: June 2, 2022

© 2022 by The EH Angle Education and Research Foundation, Inc.

INTRODUCTION

Medical imaging continues to play an increasing role in health care and is an integral part of medicine and dentistry. Since the introduction of cone-beam computed tomography (CBCT) in the late 1990s, orthodontists have rapidly adopted three-dimensional (3D) imaging technology in lieu of taking multiple two-dimensional (2D) radiographs for diagnostic records for orthodontic treatment and maxillofacial surgery. CBCT can help visualize the patient in three dimensions and can provide comprehensive information regarding anatomical spatial relationships.¹ Accurate landmark identification, however, is a prerequisite for accurate and reliable 3D image analysis. So far, craniofacial image analyses and their interpretation have primarily been performed by human experts such as orthodontists, oral and maxillofacial radiologists, and surgeons. However, identifying cephalometric landmarks in 3D images of a highly complex 3D object such as the skull is a challenging and cumbersome task that is subject to

random and systematic errors, which lead to inconsistencies within and across evaluators.² As a result, only a limited number of cephalometric landmarks by human judges have been used for most orthodontic analyses. Thus, an automatic 3D landmark identification system is needed.

With recent advances in computational processing power, significant progress has been made to create automatic landmark identification tools on 2D and 3D images.³⁻¹¹ The application of software-driven automatic landmark location for 3D imaging could greatly benefit clinicians and researchers in objectively and efficiently analyzing various parts of the craniofacial region, including structures that have not been rigorously studied, such as the condyle, glenoid fossa, nasal cavity, airway structures, nerve canals, and foramina in the jaws.

Although there are many landmarks that can be located on CBCTs, previous studies have only tested a limited number of cephalometric landmarks. Only a few studies have attempted to include a large number of noncephalometric landmarks on 2D images.^{9,10} A recent systematic review reported that deep learning produces relatively high accuracy for detecting landmarks in the majority of 2D images; data on 3D imaging are sparse, but promising.¹² To evaluate the accuracy of the automated landmark identification (ALI), a comparison between ALI and the “ground truth” or gold standard is needed.⁹ At the present time, there is no clear threshold regarding what magnitude of errors is considered clinically acceptable.¹¹

The purpose of this study was to evaluate the accuracy of an ALI software (Stratovan Corporation, Davis, Calif) compared with human landmark identification on CBCTs. The accuracy of ALI was evaluated by the mean error distance and the mean absolute error in the x, y, z coordinates of landmarks between the humans and the ALI and a successful detection rate (SDR). In addition, the shape of the envelope of error of landmarks by the ALI for 16 representative landmarks in three dimensions was evaluated.

MATERIALS AND METHODS

This study was approved by the institutional review board of the University of the Pacific (2021-95). The study sample was retrospectively collected from a local imaging center. Initially, 167 CBCT scans were screened. The inclusion criteria were (1) CBCT scans taken with a 0.3-mm³ voxel size level and at least 16 × 13 cm or greater field of view and (2) presence of all permanent teeth erupted in the dental arches. All ages, sexes, and skeletal discrepancies were included. The exclusion criteria were (1) restorative work with significant scatter and (2) presence of craniofacial

deformities, syndromes, or cleft lip and palate. The sample comprised 100 CBCT volumes. These volumes were obtained using various imaging devices because they came from many different offices and imaging centers. However, 95% of the scans were captured using an Imaging Science International CBCT scanner (Hatfield, PA), with only 5% using a NewTom device (Verona, Italy). The volumes were imported as Digital Imaging and Communications in Medicine file formats to the Checkpoint software (Stratovan Corporation, Davis, Calif). All CBCTs were anonymized, and new serial numbers were assigned.

A total of 53 landmarks were used, which included 21 orthodontic cephalometric landmarks, 12 mandibular skeletal landmarks, 6 temporomandibular joint (TMJ) landmarks, and 14 dental and dentoalveolar landmarks (Table 1).

Ground Truth

To establish ground truth (true position) for the selected landmarks, two judges (Drs. Ghowsi and Hatcher) independently performed manual landmark identification on rendered volumes in the Checkpoint software. In addition to the 3D surface-rendered model, the software had multiplanar reconstruction images in the axial, sagittal, and coronal views (Figure 1). Judge 1 was an experienced second-year orthodontic resident with at least 2 years of experience in landmark identification on 3D imaging. Judge 2 was an oral and maxillofacial radiologist who has been considered an expert in 3D imaging for more than 20 years. For the first round of calibration, the two judges used six randomly selected CBCT images based on operational definitions of the 53 selected landmarks. After the first calibration, the data were analyzed by comparing the x, y, and z coordinates for the landmarks between the two judges. An additional four CBCT images were used for the second round of calibration. After calibration, the two judges identified 53 landmarks on 100 CBCT images. The interjudge reliability between the two judges was found to be greater than 0.99. The ground truth was created by calculating the mean value of the x, y, and z coordinates for each landmark across both judges' landmark identification.

Automated Landmark Identification

The ALI was trained using a grid-search on various hyperparameters using threefold cross-validation. For each set of parameters, an instance of the model with those specific parameters was trained on two-thirds of the data and evaluated using the remaining one-third as a testing set. Once a set of desirable hyperparameters was identified and the results were

Table 1. Definition of 3D Landmarks

Number	Landmark	Definition
Orthodontic cephalometric landmarks		
1	Sella	Midpoint of the fossa for the pituitary gland
2	Nasion	Midline structure that is the most anterior point of the frontonasal suture
3	Basion	Midline structure found on the anterior region of the foramen magnum
4	ANS	Midline structure that is the most anterior point of the anterior nasal spine
5	PNS	Midline structure that is the most posterior point of the posterior nasal spine
6	A-Point	Midline point that is the deepest concavity on the maxilla between the ANS and prosthion
7	B-Point	Midline structure that is the deepest concavity on the mandibular symphysis between the infradentale and pogonion
8	Pogonion	Midline structure that is the most anterior part of the bony chin
9	Menton	Midline structure that is the most inferior part of the bony chin
10, 11	Orbitale ^a	Lowest point in the inferior region of the orbit
12, 13	Porion ^a	Most superior point on the external auditory meatus
14, 15	Condyle Post ^a	Most posterior point on the condyle
16, 17	Gonion ^a	Point where the ramus and mandibular planes intersect and the most posterior, inferior, and lateral point at the angle of the mandible
18	UR1 Incisal Edge	Upper right central incisor's most inferior point on the incisal edge
19	UR1 Root Apex	Upper right central incisor's root apex
20	LR1 Incisal Edge	Lower right central incisor's most superior point on the incisal edge
21	LR1 Root Apex	Lower right central incisor's root apex
22, 23	Antego Notch ^a	Concavity that is the most anterior region of the gonion
24, 25	Post Gonion ^a	Most posterior point of the gonion
26, 27	Coronoid ^a	Most superior and center point of the coronoid
28, 29	Lingula ^a	Ridge on the medial surface of the ramus on the mandible
Mandibular and TMJ landmarks		
30, 31	Mental for Inferior ^a	Most inferior portion directly under the landmarked mental foramen
32, 33	Mental for Ant ^a	Foramen in the mandible that is around the lower premolars identified at the anterior most point of the foramen
34, 35	Condyle Med Pole ^a	Most medial point on the condyle
36, 37	Condyle Lat Pole ^a	Most lateral point on the condyle
38, 39	Malleus ^a	Small bone in the middle ear that can be located in the axial view
Dental and dentoalveolar landmarks		
40	Max Midline	Midline point on the maxilla between the upper 1's
41	Mand Midline	Midline point on the mandible between the lower 1's
42	LL3-Alv bone	ABP ^b adjacent to the center of the lower left canine from the axial view
43	LL5-Alv bone	ABP adjacent to the center of the lower left second premolar from the axial view
44	LL7-Alv bone	ABP adjacent to the center of the lower left second molar from the axial view
45	LR3-Alv bone	ABP adjacent to the center of the lower right canine from the axial view
46	LR5-Alv bone	ABP adjacent to the center of the lower right second premolar from the axial view
47	LR7-Alv bone	ABP adjacent to the center of the lower right second molar from the axial view
48	UL3-Alv bone	ABP adjacent to the center of the upper left canine from the axial view
49	UL5-Alv bone	ABP adjacent to the center of the upper left second premolar from the axial view
50	UL7-Alv bone	ABP adjacent to the center of the upper left second molar from the axial view
51	UR3-Alv bone	ABP adjacent to the center of the upper right canine from the axial view
52	UR5-Alv bone	ABP adjacent to the center of the upper right second premolar from the axial view
53	UR7-Alv bone	ABP adjacent to the center of the upper right second molar from the axial view

^a Bilateral landmarks (right and left).

^b ABP indicates alveolar bone point.

consistent among each of the three testing folds, a final model was trained using these hyperparameters with all of the available data. A total of 53 landmarks were identified for 100 CBCTs by the ALI, and x, y, z coordinates were exported to an Excel (Microsoft Corp, Redmond, Wash) file.

Evaluation of Accuracy of ALI

To evaluate the accuracy of the ALI, the mean error distance (mm) and the mean absolute error (mm) in the

x, y, and z coordinates of the landmarks between the ground truth and the ALI were calculated. An SDR was also calculated.

The error distance of a landmark on each CBCT image between ALI and the ground truth by human judges was calculated with a 3D Euclidian distance formula, where x_1 , y_1 , and z_1 are the mean coordinate values of the two human judges (ground truth) and x_2 , y_2 , and z_2 are the coordinate values of ALI.⁶

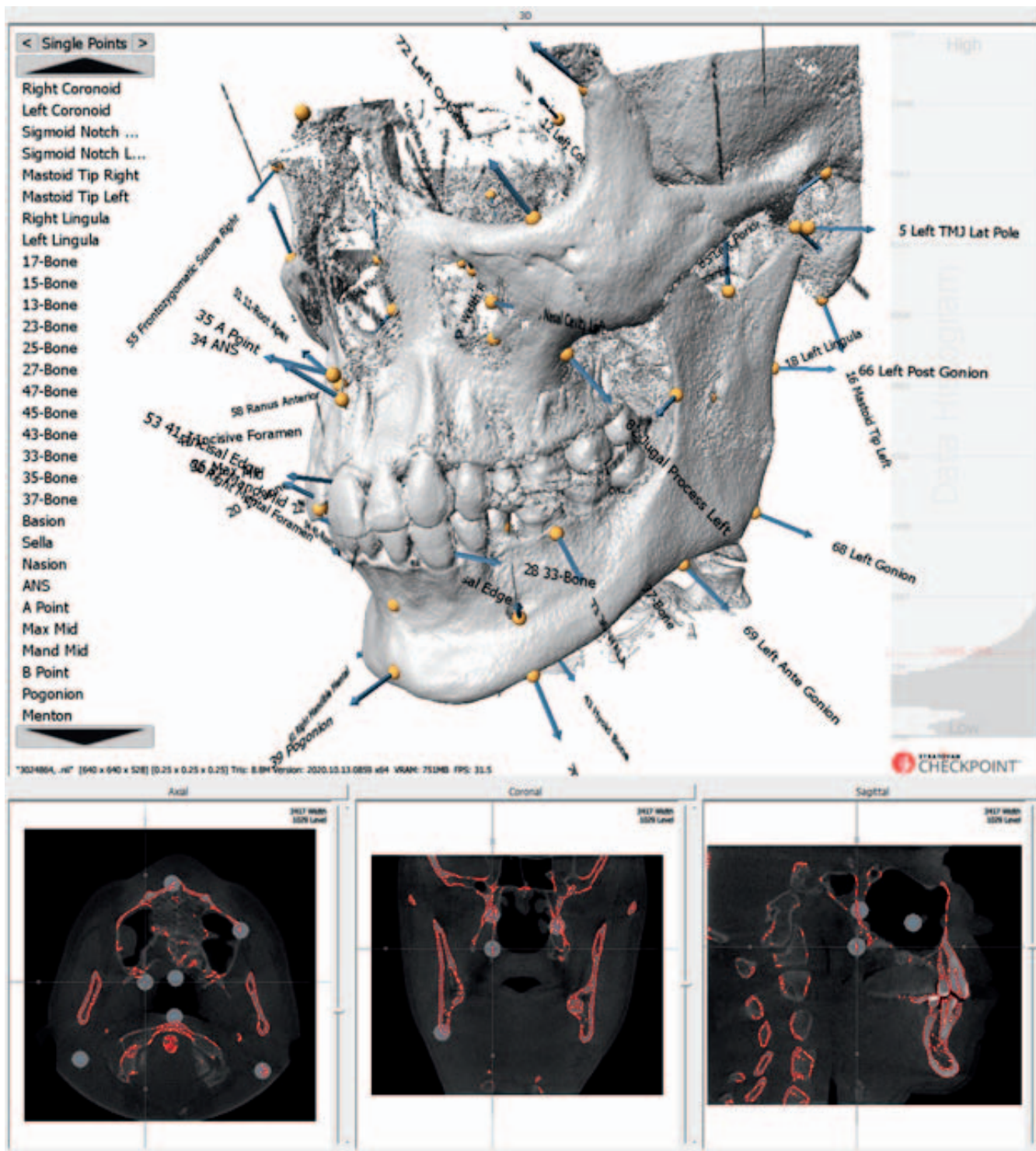


Figure 1. Axial, sagittal, and coronal slices and 3D rendering views that were used when landmarking.

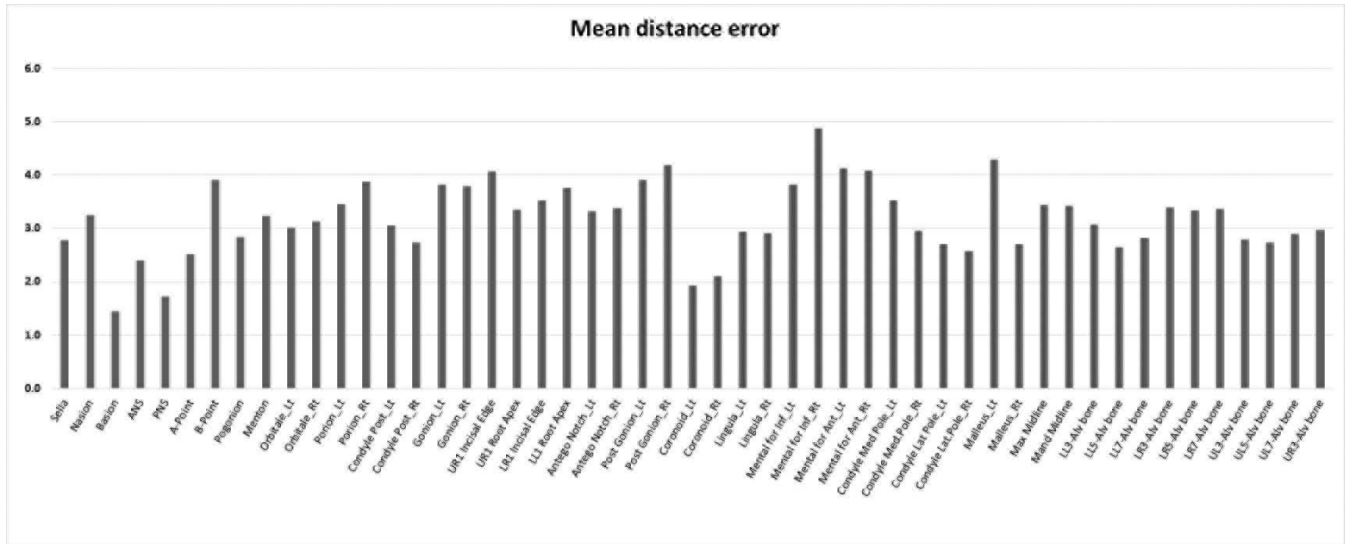


Figure 2. Mean error distance (mm) of all landmarks.

Error Distance (ED)

$$= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

Mean error distance was calculated as the following, where *i* indicates each image and *n* is the total number of images:

$$\text{Mean Error Distance} = \frac{1}{n} \sum_{i=1}^n \text{ED}_i$$

The SDR percentage represents the percentage of images in which each landmark was located within a precision range.² Common error distance ranges of ≤2 mm, 2.5 mm, 3 mm, and 4 mm were used to divide the groups according to the number of accurate identifications in SDR.²

Envelope of Errors of ALI

Scatterplots were generated for 16 of 21 cephalometric landmarks to compare the shape of the envelope of error between ALI and human judges. Average of ALI and human judge identification coordinates for each landmark were plotted on the origin with the individual estimates arrayed around them. The 95% confidence ellipses for Judge 1, Judge 2, and ALI were constructed.¹³

Statistical Analysis

Intraclass correlation coefficient (ICC) was used to evaluate the interjudge reliability for all measurements, and 10 images were tested twice through the ALI to assess its internal reliability. Simple descriptive statis-

tics, such as mean, standard deviation (SD), and percentage, were performed. Data were managed with Excel 2013 (Microsoft Corp) and were then analyzed using Statistical Package for the Social Sciences version 27 (IBM Corp, Armonk, N.Y.).

RESULTS

The interjudge reliability between the two human judges was excellent with an average ICC >0.99, except for four coordinates; only two were lower than 0.75 (Menton_y, 0.62; Left Orbitale_y, 0.72). The specific ICC values for all orthodontic cephalometric landmarks (range of 0.62 to 0.99), TMJ landmarks (range of 0.98 to 0.99), mandibular skeletal (range of 0.89 to 0.99), and dental and dentoalveolar landmarks (range of 0.98 to 0.99) are shown in the Appendix. The ICC for the ALI was 1 (identical) when the same CBCT volumes were traced twice. This indicated that the ALI algorithm is deterministic and produces precisely the same results given the same input, which is not the case with human judges.

Overall performance of the ALI system compared with the human judges in identifying landmarks on CBCT is shown in Table 2 and Figure 2. The mean absolute error for all 53 landmarks was 1.28 mm in the x-axis, 1.72 mm in the y-axis, and 1.72 mm in the z-axis. The ALI mean absolute error for all coordinate planes was 1.57 mm. The mean error distance between ALI and the ground truth was 3.19 ± 2.6 mm. For orthodontic cephalometric landmarks, ALI had a mean absolute error of less than 2 mm for the x, y, and z coordinates of sella, basion, ANS, PNS, A-Point, Menton, left porion, right condylian, and left condylian,

Table 2. Overall Performance of the ALI System: Mean Absolute Error at Each Coordinate, the Mean Error Distance With SD of Each Landmark, and SDRs for 2-, 2.5-, 3-, and 4-mm Range Criteria

Number	Landmark	Mean Absolute Error (mm)			Mean Error Distance (mm)		SDR (%)			
		x	y	z	Mean	SD	2 mm	2.5 mm	3 mm	4 mm
Orthodontic cephalometric landmarks										
1	Sella	1.82	1.28	1.08	2.79	3.61	45	66	75	87
2	Nasion	1.47	0.85	2.16	3.25	3.92	50	64	68	80
3	Basion	1.03	0.48	0.52	1.45	1.10	79	88	93	98
4	ANS	0.55	1.86	1.02	2.40	1.71	51	62	78	92
5	PNS	0.48	1.12	0.87	1.72	0.97	76	84	92	95
6	A-Point	0.47	1.68	1.35	2.51	1.61	45	57	69	86
7	B-Point	1.47	1.57	2.77	3.91	4.61	18	32	53	73
8	Pogonion	1.13	0.78	2.65	2.83	4.43	57	70	79	82
9	Menton	1.01	1.76	1.99	3.24	5.78	51	66	75	84
10	Orbitale_Lt	2.49	1.33	0.59	3.01	2.00	30	45	58	77
11	Orbitale_Rt	2.67	1.11	0.63	3.14	2.06	36	49	59	71
12	Porion_Lt	1.49	1.82	1.84	3.45	2.05	21	38	48	72
13	Porion_Rt	1.70	2.23	1.85	3.88	2.81	18	24	36	66
14	Condyle Post_Lt	1.61	1.04	1.90	3.06	2.14	39	48	60	78
15	Condyle Post_Rt	1.49	0.96	1.62	2.73	1.76	37	56	71	84
16	Gonion_Lt	0.63	2.15	2.91	3.82	3.04	31	38	47	67
17	Gonion_Rt	0.54	2.38	2.68	3.79	2.55	30	40	46	59
18	UR1 Incisal Edge	1.20	2.66	2.12	4.06	3.73	19	34	45	63
19	UR1 Root Apex	0.72	2.13	1.94	3.35	1.87	25	39	55	69
20	LR1 Incisal Edge	0.94	2.37	1.74	3.52	3.53	29	43	60	78
21	LR1 Root Apex	1.07	2.03	2.40	3.76	4.94	26	36	53	72
Mandibular and TMJ skeletal landmarks										
22	Antego Notch_Lt	1.39	2.62	1.07	3.33	2.58	34	47	53	69
23	Antego Notch_Rt	1.31	2.55	1.31	3.38	3.33	45	56	61	70
24	Post Gonion_Lt	0.67	0.74	3.66	3.91	2.61	27	37	44	60
25	Post Gonion_Rt	0.68	1.06	3.82	4.19	2.78	25	34	38	49
26	Coronoid_Lt	0.66	0.72	1.46	1.92	1.93	70	75	78	85
27	Coronoid_Rt	0.67	0.79	1.62	2.10	2.61	72	76	78	84
28	Lingula_Lt	0.95	1.50	1.91	2.94	1.76	32	52	62	79
29	Lingula_Rt	0.96	1.53	1.80	2.91	1.52	39	48	56	77
30	Mental for Inf_Lt	1.90	2.94	0.86	3.82	2.20	22	28	46	63
31	Mental for Inf_Rt	2.44	3.27	1.79	4.87	6.81	26	35	42	62
32	Mental for Ant_Lt	2.27	2.25	1.88	4.12	5.84	24	38	53	74
33	Mental for Ant_Rt	2.67	2.33	1.38	4.08	2.17	21	29	37	53
34	Condyle Med Pole_Lt	1.83	1.69	1.79	3.52	1.54	14	25	36	66
35	Condyle Med Pole_Rt	1.61	1.34	1.57	2.95	1.46	32	46	57	78
36	Condyle Lat Pole_Lt	1.28	1.31	1.36	2.71	2.11	45	57	70	79
37	Condyle Lat Pole_Rt	1.17	1.17	1.49	2.57	2.31	54	64	73	86
38	Malleus_Lt	2.37	2.47	1.71	4.29	1.62	7	16	21	42
39	Malleus_Rt	1.37	1.33	1.34	2.71	1.10	23	47	65	93
Dental and dentoalveolar landmarks										
40	Max Midline	1.30	2.17	1.70	3.44	2.37	22	40	52	68
41	Mand Midline	0.83	2.09	2.00	3.43	3.94	30	46	57	80
42	LL3-Alv bone	1.02	1.75	1.83	3.07	2.61	31	48	59	81
43	LL5-Alv bone	1.14	1.48	1.33	2.65	1.51	39	59	70	84
44	LL7-Alv bone	1.06	1.78	1.34	2.82	1.56	39	53	66	80
45	LR3-Alv bone	1.15	1.96	2.04	3.39	4.18	33	43	57	78
46	LR5-Alv bone	1.37	1.99	1.69	3.34	3.50	24	47	56	82
47	LR7-Alv bone	1.40	1.95	1.72	3.37	2.12	24	35	50	73
48	UL3-Alv bone	0.81	1.75	1.46	2.80	1.39	31	51	65	83
49	UL5-Alv bone	1.09	1.66	1.24	2.73	1.30	36	49	61	83
50	UL7-Alv bone	1.10	1.70	1.49	2.89	1.40	30	42	56	78
51	UR3-Alv bone	1.08	1.90	1.36	2.97	2.31	31	47	59	81
52	UR5-Alv bone	1.17	1.85	1.39	2.99	1.71	28	49	61	81
53	UR7-Alv bone	1.30	1.70	1.97	3.35	1.57	20	31	49	67
Mean of all landmarks		1.28	1.72	1.72	3.19	2.60	35	48	59	75

which is considered highly accurate. All 21 orthodontic cephalometric landmarks had a mean absolute error of less than 3 mm (Table 2). The mean SDR within a 4-mm error distance range was 78% for all orthodontic cephalometric landmarks. Basion and PNS showed the highest accuracy at a mean error distance of 1.45 ± 1.1 mm and 1.72 ± 0.97 mm, respectively, and a 79% and 76% SDR in a 2-mm error distance range, respectively. A slightly more than 90% SDR within a 3-mm range resulted for both landmarks. On the other hand, UR1 Incisal Edge showed the lowest accuracy at a mean error distance of 4.06 ± 3.73 mm and both the right and left gonion points at a mean error distance of 3.79 ± 2.55 mm and 3.82 ± 3.04 mm, respectively.

ALI in the mandibular and TMJ regions showed good accuracy, with a mean absolute error of less than 3 mm for all landmarks, except the right mental foramen inferior and posterior gonion; the mean SDR within a 4-mm range was 71% for the mandibular and TMJ regions. For the dental and dentoalveolar landmarks, the ALI had a mean absolute error of less than 3 mm for all landmark coordinates. The mean SDR within a 4-mm range for all dental and dentoalveolar landmarks, including the incisor edge and root apex, was 77%.

Figure 3 shows scatterplots depicting the envelope of errors in three different planes. Overall, the shape of the envelope of error and 95 % confidence ellipse by ALI were similar to that of human judges, with a few significant outliers noted for both.

DISCUSSION

The present study was conducted to test the accuracy and reliability of an ALI developed by the Stratovan Corporation in comparison with the ground truth performed by human judges. Across all three coordinate planes, 98% of landmarks had a mean absolute error of less than 3 mm compared with human judges. The present study included many bilateral anatomical landmarks and showed higher accuracy than a study by Shahidi et al., which reported a 63.57% SDR with <3 mm error distance when 14 cephalometric landmarks on 28 CBCTs were identified.⁶ Multiple factors might affect the accuracy of ALI systems: sample size, computation method, cross-validation folds, image resolution, and so on. Most of the previous studies focused on standard cephalometric landmarks on the midsagittal plane, and only a few bilateral landmarks were tested. The present study included multiple noncephalometric bilateral landmarks, such as mental and mandibular foramen, coronoid process, and condylar points, which are useful in evaluating transverse dimension, asymmetry, and shape analyses.

According to the clinical requirement, if the difference between the ALI and the ground truth was less than 2 mm, the ALI was considered to be correct; if less than 4 mm, it was acceptable.⁴ The current study showed a 75% SDR for all 53 landmarks (78% for the cephalometric landmarks) within a 4-mm error distance range, which is considered a clinically acceptable range for cephalometric measurements. However, the mean error distance might not be the most relevant result to assess the accuracy of the ALI when error distributions are not spherical, as shown in the envelope of error scattergrams in Figure 3. Therefore, the mean absolute error for each x, y, and z coordinate was reported to evaluate the directional error distribution on each plane, which can provide different clinical significance. In addition, the clinically acceptable amount of error in landmark placement depends entirely on clinical application. Further studies are needed to assess the effects of landmark location errors in each direction (x, y, and z coordinates) on the values of the linear and angular measures of some widely used standard cephalometric and noncephalometric head-film analyses.

In this study, two expert judges manually located 53 standard cephalometric and noncephalometric landmarks on 100 CBCT images. During the calibration session, the six images used were reevaluated by Judge 1 to confirm precise landmarking. Judge 1 had a mean error of 0.7 mm from one session to another. The ALI outperformed with an error of 0 mm from one session to another, which is a significant advantage over human landmark identification. It is well documented that locating landmarks on a 3D surface using volume rendering introduces significantly more errors.¹⁴ For instance, when the landmarks were placed on the 3D surface alone, they could jump to the opposite side of the CBCT volume if they did not interface with the surface exactly as intended. Therefore, multiplanar reconstruction images were employed to increase accuracy; however, this approach requires more time to complete the landmark identification task.^{3,4}

Another aim of this study was to determine whether the envelope of error by ALI was similar to that of human judges. Park et al. demonstrated that the envelope of error on CBCT was not spherical, and the midline structures were typically identified more accurately.¹⁵ They also found that, among human judges, the most reliable landmarks were the incisal edges of the incisors, whereas the least reliable was gonion.¹⁵ In the current study, the most reliable landmarks between the human judges were sella, left malleus, right coronoid, and nasion. More than 30 new anatomical landmarks were included in the current study, such as landmarks for the mental foramen,

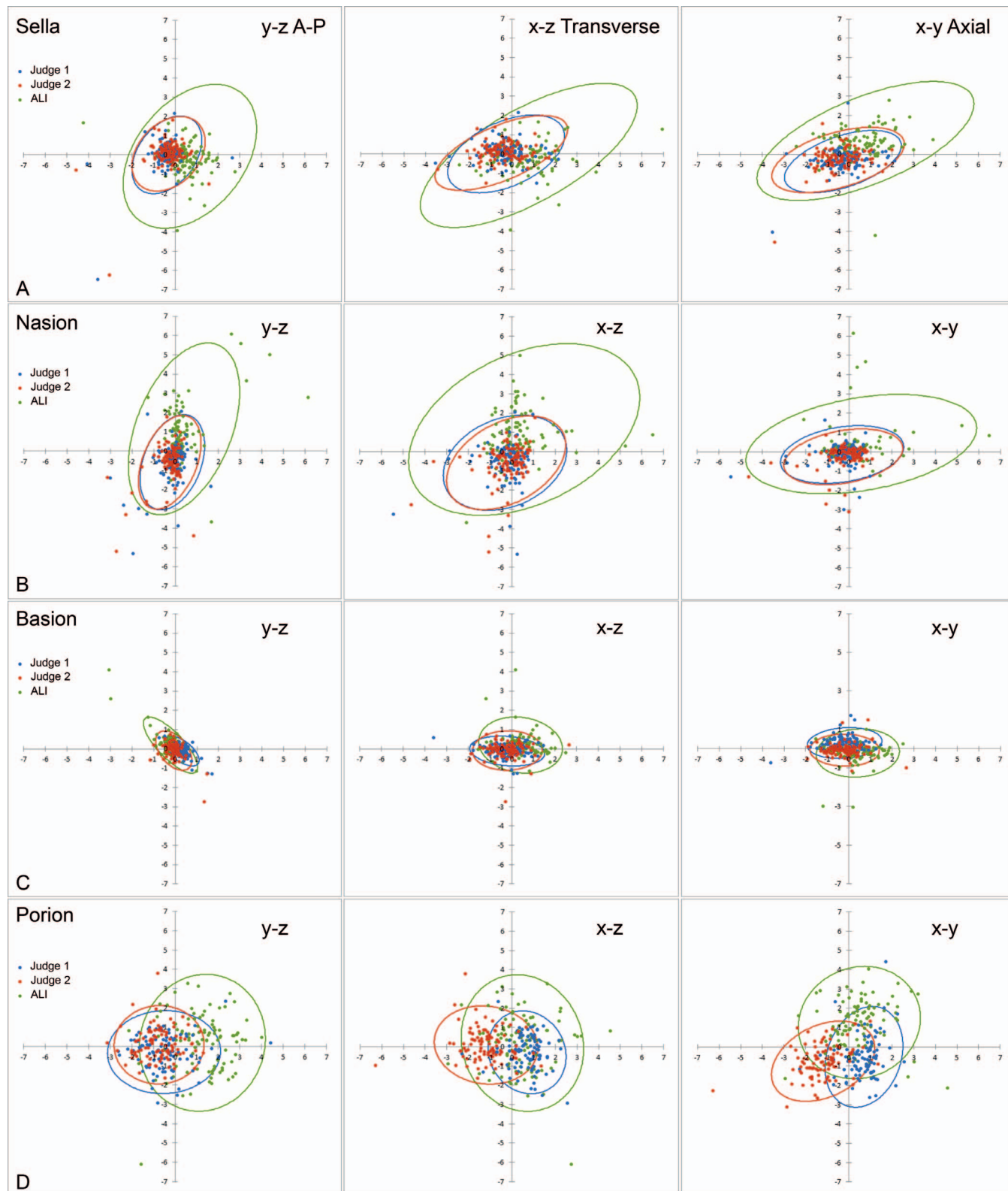


Figure 3. Scatterplots with 95% confidence ellipses depicting the envelope of error in different planes of view. Judge 1 is depicted in blue, Judge 2 is depicted in red, and ALI is depicted in green. For bilateral landmarks, the right-side landmark is presented. (A) Sella. (B) Nasion. (C) Basion. (D) Porion. (E) Orbitale. (F) ANS. (G) PNS. (H) A-Point. (I) B-Point. (J) Pogonion. (K) Menton. (L) Gonion. (M) UR1 Incisor Edge. (N) UR1 Root Apex. (O) LR1 Incisal Edge. (P) LR1 Root Apex.

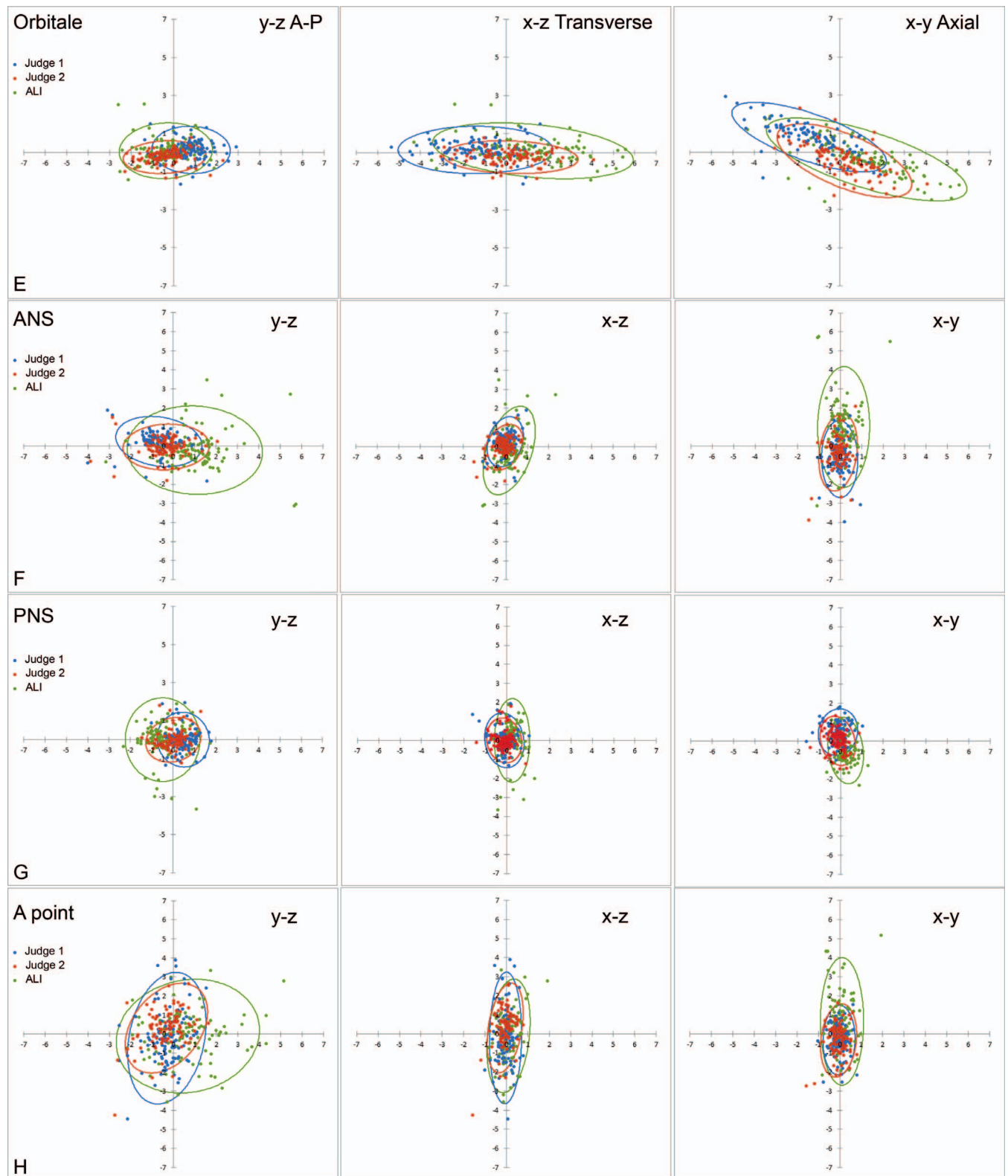


Figure 3. Continued.

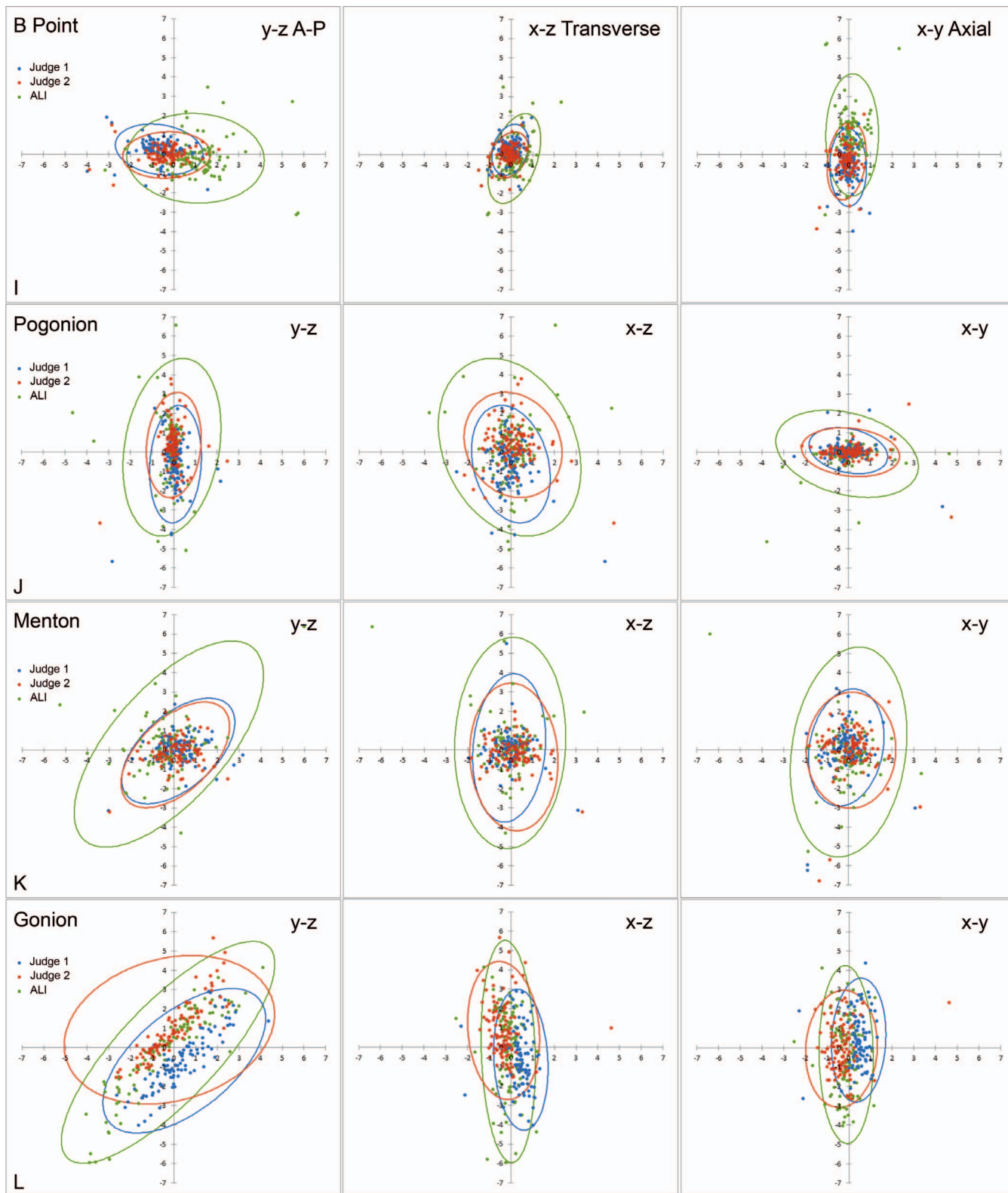


Figure 3. Continued.

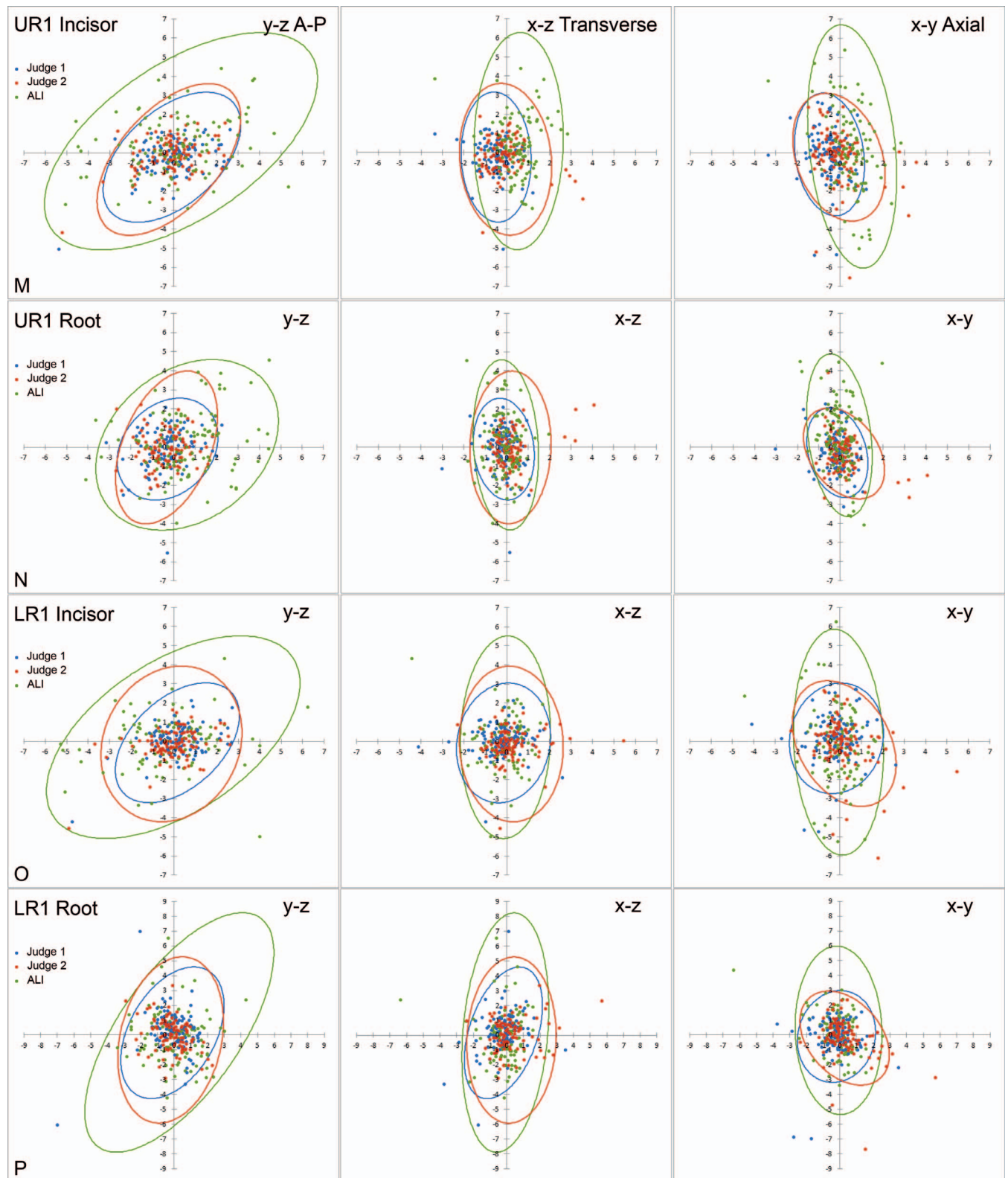


Figure 3. Continued.

antegonial notch point, lingula point, and various condylar points; some of these new landmarks showed promising results. The mental foramen inferior point had low reliability between the two human judges, which could have been attributed to the points jumping from one area to another when identifying landmarks on the 3D surface; the point could have occasionally traveled deep into the foramen. It is worth mentioning that similar observations were found with the human judge's landmark identification. The ALI was less accurate in locating gonion and posterior gonion, which could be due to the same reasons that the human judges encountered as previously reported in 2D radiographs¹⁶ and 3D CBCT images.¹⁵ The scatterplots reveal that the shape of the envelopes of error of gonion points generated by ALI and the human judges were almost identical (Figure 3L).

Interestingly, the ALI was less reliable in locating the incisal edge and root apex of the upper and lower incisors, which was an unexpected finding (Figure 3M-P). The 95% confidence ellipses in both the y-axis and the z-axis were greater with ALI than with Judge 1 and Judge 2. The application of ALI might give rise to errors ranging from -5 mm to 7 mm in the z-axis for the upper incisor edge and -8 mm to 8 mm in the z-axis for the lower incisor root with 95% probability (Figure 3M,P). This could have been attributed to difficulty in distinguishing between teeth in occlusion with one another and between bone and a small area of the root apex. It is worth noting that the incisor edge point is the most easily identifiable and most reliable landmark among human judges.^{15,16} This finding indicates that an outlier detection and feedback system is needed to improve the accuracy of the ALI. Therefore, a hybrid approach that incorporates human supervision along with further development of ALI could help achieve higher accuracy for automatic landmark identification and automatic 3D craniofacial cephalometric and geometric morphometric analyses.

Fully automated landmark identification of 3D craniofacial images is a rapidly advancing field. The most recent systematic review revealed¹¹ that ALI has the potential to be a promising tool for aiding orthodontists with landmark identification on CBCTs. Future studies include further improvement of the accuracy of ALI; the development of new landmarks that have not been previously studied, such as landmarks for airway and the nasal cavity; and the creation of various automated linear and angular measurements.

CONCLUSIONS

- A fully automated landmark identification was used to detect 53 landmarks on 100 CBCTs. The ALI's mean absolute error for all coordinate planes was 1.57 mm

on average. The mean error distance between ALI and the ground truth was 3.19 ± 2.6 mm. ALI resulted in a 75% SDR for all 53 landmarks (78% SDR for the 21 cephalometric landmarks) within a 4-mm error range, which is considered a clinically acceptable range for generating cephalometric measurements.

- The ALI was more precise than humans when identifying landmarks on the same image at different times.
- This study demonstrates the promise of ALI in aiding orthodontists and researchers with CBCT landmark identification in routine clinical practice and the analysis of big data for research in the future.

SUPPLEMENTAL DATA

The Appendix with supplemental data is available online.

REFERENCES

1. Mah JK, Huang JC, Choo H. Practical applications of cone-beam computed tomography in orthodontics. *J Am Dent Assoc.* 2010;141:7S-13S.
2. Lindner C, Wang CW, Huang CT, et al. Fully automatic system for accurate localisation and analysis of cephalometric landmarks in lateral cephalograms. *Sci Rep.* 2016;6:33581.
3. Hassan B, Nijkamp P, Verheij H, et al. Precision of identifying cephalometric landmarks with cone beam computed tomography in vivo. *Eur J Orthod.* 2013;35: 38-44.
4. 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.
5. Montúfar J, Romero M, Scougall-Vilchis RJ. Hybrid approach for automatic cephalometric landmark annotation on cone-beam computed tomography volumes. *Am J Orthod Dentofacial Orthop.* 2018;154:140-150.
6. Shahidi S, Bahrapour E, Soltanimehr E, et al. The accuracy of a designed software for automated localization of craniofacial landmarks on CBCT images. *BMC Med Imaging.* 2014;14:32.
7. Gupta A, Kharbanda OP, Sardana V, et al. A knowledge-based algorithm for automatic detection of cephalometric landmarks on CBCT images. *Int J Comput Assist Radiol Surg.* 2015;10:1737-1752.
8. Mohammad-Rahimi H, Nadimi M, Rohban MH, Shamsoddin E, Lee VY, Motamedian SR. Machine learning and orthodontics, current trends and the future opportunities: a scoping review. *Am J Orthod Dentofacial Orthop.* 2021; 160:170-192.
9. Hwang HW, Park JH, Moon JH, et al. Automated identification of cephalometric landmarks: part 2—might it be better than human? *Angle Orthod.* 2020;90:69-76.
10. Moon JH, Hwang HW, Yu YS, Kim MG, Donatelli RE, Lee SJ. How much deep learning is enough for automatic identification to be reliable? A cephalometric example. *Angle Orthod.* 2020;90:823-830.

11. Dot G, Rafflenbeul F, Arbotto M, Gajny L, Rouch P, Schouman T. Accuracy and reliability of automatic three-dimensional cephalometric landmarking. *Int J Oral Maxillofac Surg.* 2020;49:1367–1378.
12. Schwendicke F, Chaurasia A, Arsiwala L, et al. Deep learning for cephalometric landmark detection: systematic review and meta-analysis. *Clin Oral Investig.* 2021;25:4299–4309.
13. Donatelli RE, Lee SJ. How to report reliability in orthodontic research: part 2. *Am J Orthod Dentofacial Orthop.* 2013;144:315–318.
14. Fernandes TM, Adamczyk J, Poleti ML, Henriques JF, Friedland B, Garib DG. Comparison between 3D volumetric rendering and multiplanar slices on the reliability of linear measurements on CBCT images: an in vitro study. *J Appl Oral Sci.* 2015;23:56–63.
15. Park J, Baumrind S, Curry S, Carlson SK, Boyd RL, Oh H. Reliability of 3D dental and skeletal landmarks on CBCT images. *Angle Orthod.* 2008;89:758–767.
16. Baumrind S, Frantz RC. The reliability of head film measurements: 1. Landmark identification. *Am J Orthod.* 1971;60:111–127.