Comparative Validation of the Mixed and Permanent Dentition at Web-Based Artificial Intelligence Cephalometric Analysis

Sunhahn Shin, Donghyun Kim

Division of Pediatric Dentistry, Department of Dentistry, Ewha Womans University Mokdong Hospital

Abstract

This retrospective study aimed to evaluate the difference in measurement between conventional orthodontic analysis and artificial intelligence orthodontic analysis in pediatric and adolescent patients aged 7 - 15 with the mixed and permanent dentition.

A total of 60 pediatric and adolescent patients (30 mixed dentition, 30 permanent dentition) who underwent lateral cephalometric radiograph for orthodontic diagnosis were randomly selected. Seventeen cephalometric landmarks were identified, and 22 measurements were calculated by 1 examiner, using both conventional analysis method and deep learning-based analysis method. Errors due to repeated measurements were assessed by Pearson's correlation coefficient. For the mixed dentition group and the permanent dentition group, respectively, a paired t-test was used to evaluate the difference between the 2 methods.

The difference between the 2 methods for 8 measurements were statistically significant in mixed dentition group: APDI, SNA, SNB, Mandibular plane angle, LAFH ($p < 0.001$), Facial ratio ($p = 0.001$), U1 to SN ($p = 0.012$), and U1 to A-Pg ($p = 0.021$). In the permanent dentition group, 4 measurements showed a statistically significant difference between the 2 methods: ODI ($p = 0.020$), Wits appraisal ($p = 0.025$), Facial ratio ($p = 0.026$), and U1 to A-Pg ($p = 0.001$).

Compared with the time-consuming conventional orthodontic analysis, the deep learning-based cephalometric system can be clinically acceptable in terms of reliability and validity. However, it is essential to understand the limitations of the deep learning-based programs for orthodontic analysis of pediatric and adolescent patients and use these programs with the proper assessment.

Key words : Cephalometric radiography, Deep learning, Artificial intelligence, Orthodontic diagnosis

Ⅰ. Introduction

Early diagnosis and treatment of malocclusion in mixed dentition is one of the important topics in pediatric dentistry. Lateral cephalometric radiography is an essential element in evaluating the cause of malocclusion by determining the relationship between the teeth, skeleton and soft tissue[1]. The pediatric dentist's judgment is important in establishing a treatment plan and recognizing patients' needs. These orthodontic treatment decisions may vary between pediatric dentists due to differences of opinion. Clinicians with less experience have trouble making the right decision[2].

Manual tracing is time-consuming and comes with a high risk of misreading values. Manual tracing is susceptible to error in landmark identification due to radiographic film magnification[3,4]. Manual tracing has been replaced by digital tracing gradually. Computer-assisted orthodontic analysis, such as V-Ceph (Osstem, Seoul, Korea), FACAD (Ilexis AB, Linköping, Sweden), and Dolphin (Dolphin Imaging, Chatsworth, CA, USA), has many advantages: the ease of use, unnecessariness of hard copies, and modification of digital images[5]. In spite of these advantages, several factors influence the selection of software, such as the complexity of installation procedures, high subscription and update cost, and the significant effort required to master the training software[6].

According to recent studies, the clinical adaptation of deep learning models has shown outstanding performance in dentistry[7,8]. Many algorithms have been developed to automatically detect cephalometric landmarks for orthodontic analysis through various artificial intelligence models[9-11]. Previous studies showed the accuracy and reperoducibility of landmark identification using artificial intelligence[6,10,11]. Although few studies have evaluated angular and linear measurements due to the difficulty of analyzing the reproducibility of lines and angles, evaluation of a new diagnostic system is important to assess whether it provides immediately useful results for treatment planning[12].

WebCeph (Assemblecircle, Seoul, Korea) is an orthodontic analysis platform and web-based cloud system that clinicians are available free of charge. It can save patient records, predict outcomes after orthodontic treatment, and perform cephalometric analysis using deep learning-based algorithms. WebCeph is a system developed by image training a deep learning model for adult patients with permanent dentition, and its effectiveness is reported in orthognathic surgery and adult orthodontic diagnosis[2]. However, there are insufficient studies to evaluate the accuracy of orthodontic analysis programs based on deep learning models in children and adolescents.

This study aims to validate the reliability of the application of WebCeph, a clinically acceptable deep learning-based system in adult patients, to children and adolescents without fine tuning. This study focused on whether these data were clinically useful for adolescent patients by evaluating 22 cephalometric measurements obtained through both conventional and deep learning-based methods.

Ⅱ. Materials and Methods

This retrospective study was conducted at Ewha Womans University Mokdong Hospital. The study was approved by the Institutional Review Board (IRB) of the hospital (IRB No. 2021-03-029-003) prior to the study.

1. Study population

The subjects were 60 pediatric and adolescent patients aged 7 - 15 years for orthodontic diagnosis between 2013 and 2019. Patients were randomly selected for this retrospective study and consisted of 30 patients with mixed dentition and 30 patients with permanent dentition. The average age of the mixed dentition group was 8.33 years (range, 7 - 11 years) and the permanent dentition group was 12.67 years (range, 11 - 15 years). The percentage of men and women was equally distributed in both groups. The exclusion criteria were specified as follows: unerupted or missing incisors and first permanent molars, unerupted teeth overlying the incisors' apices, malformed teeth, maxillofacial deformities, and severe skeletal deformations.

All cephalometric radiographs used in this study were obtained from subjects for orthodontic diagnosis prior to orthodontic treatment. The cephalometric radiographs were obtained with the following specifications: dose, 15 mAs; tube voltage, 80 - 90 kV; scanning time, 0.6 - 1.0 sec; and field of view, 20.1 × 25.2 cm. Cephalometric radiographs were downloaded from an imaging system (INFINITT Dental PACS; Infinitt, Seoul, Korea) and stored as DICOM files for V-Ceph and JPEG images for WebCeph. The original image was 2,510 × 2,000 pixels with 300 dpi resolution.

2. Study design

1) Evaluation of the Cephalometric radiograph using conventional tracing

All 60 cephalometric radiographs were traced by 1 pediatric dentist using V-Ceph version 7.0 (Osstem, Seoul, Korea) and retraced by the same examiner 2 weeks later for the intra-examiner reproducibility. Landmark identification for cephalometric analysis was performed manually on the digital image with a mouse-controlled cursor by the examiner. A total of 17 landmarks were defined on each cephalometric radiograph and 22 selected skeletal and dental parameters were measured

(Table 1).

2) Evaluation of the Cephalometric radiograph using deep learning-based programs

An individual account was created on the WebCeph website for this study. The cephalometric radiographs were uploaded to the WebCeph website and traced by the website within a few seconds. The actual size of each image was calibrated by measuring the known distance (20.0 mm) between 2 points in the digital cassette using the WebCeph ruler on the uploaded image screen. This calibration was standardized all cephalometric images. Images were traced automatically by the deep learning-based program, and the examiner did not modify the orthodontic analysis results. All measurements were obtained from the WebCeph database for each image.

3. Statistical analysis

The mean differences and standard deviation of the differences between the repeated measurements for each patient group were calculated. The intra-examiner reproducibility was evaluated by the Pearson's correlation coefficient. Systematic errors were calculated by paired measurement comparisons of conventional and deep learning-based tracings by paired t-test, respectively, for the mixed and permanent dentition groups. The data were analyzed by IBM SPSS Statistics software (ver. 22.0, SPSS Inc., Chicago, IL, USA).

Ⅲ. Results

1. Reliability of repeated conventional tracing in the mixed and permanent dentition groups

Table 2 gives the reliability of repeated measurements by a single examiner for the conventional tracing. In general, the correlation coefficients of all measurements in both mixed and permanent dentition groups were above 0.9 (strong correlation), except LAFH, which had a correlation of 0.82 for the

Table 1. Cephalometric measurements used in this study

Measurements	Definition
Skeletal Measurement	
Facial angle	Angle between Po-Or and N-Pg
Convexity angle	Angle between N-A and A-Pg
FMA	Angle between Frankfort horizontal plane and Mandibular plane
APDI	Summation of measurements of Palatal plane angle, Facial angle, and A-B plane angle
ODI	Summation of measurements of Palatal plane angle and A-B to Mandibular plane angle
SNA	Angle between S-N and N-A
SNB	Angle between S-N and N-B
ANB	Angle between A-N and N-B
Wits appraisal (mm)	Distance between points A and B, both projected perpendicular on the functional occlusal plane
Mandibular plane angle	Angle between S-N and Go-Gn
LAFH (mm)	Distance between ANS and Me (lower anterior facial height)
Facial ratio (PFH/AFH)	Ratio percent of PFH and AFH measurements
SN to occlusal plane	Angle between S-N and the Functional occlusal plane
Dental Measurement	
U1 to SN	Angle between long axis of upper incisor and S-N
U1 to NA angle	Angle between long axis of upper incisor and N-A
U1 to NA (mm)	Distance from Upper incisor edge to N-A
U1 to A-Pg (mm)	Distance from Upper incisor edge to A-Pg
Inter incisal angle	Angle between long axis of upper and lower incisors
IMPA	Angle between long axis of lower incisor and the mandibular plane
L1 to NB angle	Angle between long axis of lower incisor and N-B
L1 to NB (mm)	Distance from long axis of lower incisor and N-B
L1 to A-Pg (mm)	Distance form long axis of lower incisor and A-Pg

Table 2. Reproducibility of repeated measurements using a conventional analysis method in the mixed and permanent dentition groups (inter-examiner error)

Measurements	Mixed dentition			Permanent dentition		
	Difference (Mean ± SD)	r	p	Difference (Mean ± SD)	r	p
Facial angle	-0.30 ± 1.15	0.952	0.000	0.07 ± 0.24	0.997	0.000
Convexity angle	0.20 ± 0.92	0.990	0.000	0.04 ± 0.21	0.989	0.000
FMA	-0.35 ± 1.30	0.951	0.000	0.02 ± 0.29	0.969	0.000
APDI	-0.72 ± 1.85	0.968	0.000	-0.02 ± 0.27	0.957	0.000
ODI	0.66 ± 3.00	0.941	0.000	-0.06 ± 0.26	0.978	0.000
SNA	-1.25 ± 1.35	0.906	0.000	-0.04 ± 0.33	0.995	0.000
SNB	-1.96 ± 1.69	0.900	0.000	0.11 ± 0.29	0.977	0.000
ANB	0.70 ± 1.28	0.903	0.000	-0.12 ± 0.48	0.983	0.000
Wits appraisal (mm)	-0.40 ± 1.73	0.937	0.000	-0.08 ± 0.27	0.999	0.000
Mandibular plane angle	1.44 ± 1.89	0.911	0.000	-0.05 ± 0.39	0.968	0.000
LAFH (mm)	1.74 ± 2.53	0.829	0.000	-0.04 ± 0.34	0.929	0.000
Facial ratio (PFH/AFH)	-0.49 ± 1.02	0.951	0.000	0.04 ± 0.26	0.959	0.000
SN to occlusal plane	0.24 ± 1.05	0.997	0.000	-0.15 ± 0.30	0.998	0.000
U1 to SN	-0.39 ± 1.62	0.982	0.000	-0.04 ± 0.22	0.958	0.000
U1 to NA angle	-0.50 ± 1.96	0.971	0.000	0.05 ± 0.23	0.972	0.000
U1 to NA (mm)	-0.08 ± 0.59	0.959	0.000	0.03 ± 0.17	0.997	0.000
U1 to A-Pg (mm)	-0.22 ± 0.54	0.983	0.000	0.03 ± 0.18	0.998	0.000
Inter incisal angle	-0.58 ± 2.84	0.973	0.000	0.00 ± 0.29	0.975	0.000
IMPA	0.36 ± 1.50	0.978	0.000	0.03 ± 0.16	0.974	0.000
L1 to NB angle	0.22 ± 0.86	0.991	0.000	0.08 ± 0.23	0.979	0.000
L1 to NB (mm)	0.03 ± 0.68	0.958	0.000	0.05 ± 0.21	0.967	0.000
L1 to A-Pg (mm)	-0.16 ± 0.57	0.977	0.000	-0.04 ± 0.21	0.998	0.000

p values from Pearson's correlation test.
r = Pearson's correlation coefficient.

mixed dentition group. As indicated by the correlation coefficients in both groups, intra-examiner reproducibility was high, and the reliability of repeated measurements was good.

2. Measurement differences between conventional analysis and deep learning-based cephalometric analysis in the mixed and permanent dentition groups

Table 3 shows the measurement differences between the conventional and deep learning-based methods in the mixed dentition group. Statistically significant differences were detected for APDI, SNA, SNB, Mandibular plane angle, LAFH, Facial ratio, U1 to SN, and U1 to A-Pg. The measurement differences between the 2 methods in the permanent dentition group are reported in Table 4. In general, the magnitude of the differ-

ence between sample means was small. Only 4 measurements (ODI, Wits appraisal, Facial ratio, and U1 to A-Pg) have statistically significant differences.

IV. Discussion

Lateral cephalometric image analysis was first introduced by Broadbent[13] in 1931. It has been used as a tool for diagnosing malocclusion, treatment planning, a growth pattern, and treatment outcome analysis. In the past, anatomical landmarks were manually traced on cephalometric radiographs. Despite being the most important procedure in cephalometric analysis, landmark identification is error-prone and time-consuming[14]. After the introduction of computer-aided analysis, cephalometric analysis using deep learning models was developed.

Deep learning algorithms use convolution filters and pool-

Table 3. Measurement differences between conventional and deep learning-based cephalometric analysis in the mixed dentition group

Measurements	V-Ceph (Mean ± SD)	WebCeph (Mean ± SD)	Difference Between Two Methods (Mean ± SD)	p
Facial angle	85.47 ± 3.20	85.28 ± 3.57	0.19 ± 1.81	0.561
Convexity angle	7.73 ± 5.88	8.01 ± 6.43	-0.29 ± 1.60	0.335
FMA	28.38 ± 3.88	28.50 ± 4.06	-0.12 ± 1.94	0.737
APDI	81.92 ± 5.94	80.00 ± 6.74	1.91 ± 2.10	0.000
ODI	72.48 ± 6.81	73.10 ± 7.96	-0.62 ± 3.56	0.350
SNA	81.44 ± 3.13	80.02 ± 3.00	1.42 ± 1.69	0.000
SNB	78.01 ± 3.98	76.51 ± 3.62	1.50 ± 1.47	0.000
ANB	3.43 ± 2.65	3.55 ± 2.71	-0.12 ± 0.81	0.434
Wits appraisal (mm)	-2.21 ± 3.40	-2.11 ± 4.40	-0.10 ± 1.95	0.792
Mandibular plane angle	34.25 ± 3.90	37.14 ± 4.29	-2.89 ± 1.63	0.000
LAFH (mm)	58.06 ± 3.46	63.06 ± 4.13	-5.00 ± 3.04	0.000
Facial ratio (PFH/AFH)	63.70 ± 3.04	62.58 ± 3.21	1.12 ± 1.69	0.001
SN to occlusal plane	20.79 ± 5.14	23.42 ± 11.63	-1.63 ± 10.90	0.197
U1 to SN	102.32 ± 7.08	100.56 ± 8.46	1.76 ± 3.60	0.012
U1 to NA angle	20.88 ± 6.67	21.38 ± 7.74	-0.50 ± 3.01	0.369
U1 to NA (mm)	3.23 ± 1.65	3.53 ± 1.93	-0.31 ± 1.28	0.200
U1 to A-Pg (mm)	5.70 ± 2.52	5.75 ± 2.60	-1.20 ± 2.70	0.021
Inter incisal angle	130.57 ± 9.15	129.19 ± 10.71	1.37 ± 4.70	0.120
IMPA	91.11 ± 5.33	92.21 ± 6.59	-1.10 ± 3.88	0.130
L1 to NB angle	25.13 ± 4.94	26.33 ± 5.93	-0.06 ± 1.12	0.788
L1 to NB (mm)	5.16 ± 1.90	5.40 ± 2.25	-0.23 ± 0.92	0.174
L1 to A-Pg (mm)	3.33 ± 2.25	3.15 ± 2.54	0.18 ± 1.17	0.410

p values from paired t-test.

ing layers to extract characteristics from images and analyze patterns. Many deep learning models have been enhanced by improving filter sizes, various ideas, and locations. The convolutional neural network, a deep neural network with multiple hidden layers, is a suitable structure for learning 2-dimensional images, such as cephalometric radiographs. These algorithms are useful for less experienced clinicians as they provide consistent results of landmark detection and measurement analysis[2]. The previous studies only investigated the accuracy of orthodontic diagnosis in patients with permanent teeth using artificial intelligence-based algorithm[6]. It is meaningful to conduct a study on the reliability of deep learning-based orthodontic analysis systems in young patients with the mixed dentition and permanent dentition.

WebCeph is a program recently developed in Korea through deep learning training of Korean adult orthodontic patients. After receiving a lateral cephalometric radiograph as input, in step 1, WebCeph detects the lateral facial region using the Support Vector Machine for the Histogram Oriented Gradients processed image. In step 2, landmarks are detected using the Ensemble of Regression Tree algorithm and the Gradient Boosting Algorithm for the detected lateral facial region. In step 3, using Faster R-CNN (ResNet), it is designed to detect a more precise location around the detected landmark location[15].

In this study, this program was chosen from among several deep learning-based programs to evaluate the accuracy of cephalometric measurements in children and adolescents. WebCeph is now free of charge, and clinicians can easily access it from mobile and computers in 22 languages. If the accuracy of the orthodontic analysis in both pediatric and adolescent patients is proven, it will help clinicians in actual clinical practice.

According to previous studies, the differences between manual and digital tracing methods were not clinically significant, and both methods were reliable[6,16]. This study evalu-

Table 4. Measurement differences between conventional and deep learning-based cephalometric analysis in the permanent dentition group

Measurements	V-Ceph (Mean ± SD)	WebCeph (Mean ± SD)	Difference Between Two Methods (Mean ± SD)	p
Facial angle	87.22 ± 2.87	87.21 ± 2.86	0.41 ± 0.62	0.935
Convexity angle	8.62 ± 5.50	8.68 ± 5.59	-0.36 ± 0.45	0.478
FMA	26.28 ± 6.41	26.36 ± 6.42	-0.08 ± 0.53	0.392
APDI	81.90 ± 5.85	81.89 ± 5.82	-0.20 ± 0.45	0.976
ODI	74.09 ± 7.69	74.29 ± 7.67	-0.79 ± 0.50	0.020
SNA	84.02 ± 3.20	84.19 ± 3.19	-0.17 ± 0.62	0.146
SNB	79.81 ± 3.62	79.75 ± 3.60	0.25 ± 0.52	0.573
ANB	4.21 ± 2.40	4.45 ± 2.55	-0.24 ± 0.64	0.054
Wits appraisal (mm)	-0.19 ± 4.43	-0.01 ± 4.65	-1.18 ± 0.42	0.025
Mandibular plane angle	31.85 ± 5.71	32.02 ± 5.70	-0.17 ± 0.49	0.068
LAFH (mm)	65.48 ± 6.66	65.62 ± 6.58	-0.14 ± 0.49	0.138
Facial ratio (PFH/AFH)	67.21 ± 5.24	67.41 ± 5.10	-0.19 ± 0.45	0.026
SN to occlusal plane	16.51 ± 4.47	16.70 ± 4.56	-0.19 ± 0.55	0.066
U1 to SN	106.44 ± 8.02	106.37 ± 7.97	0.06 ± 0.50	0.496
U1 to NA angle	22.89 ± 5.94	23.08 ± 5.85	-0.20 ± 0.49	0.139
U1 to NA (mm)	4.82 ± 2.20	4.98 ± 2.18	-1.16 ± 0.44	0.054
U1 to A-Pg (mm)	7.92 ± 3.67	8.15 ± 3.59	-1.23 ± 0.35	0.001
Inter incisal angle	126.03 ± 10.67	126.11 ± 10.49	-0.48 ± 0.67	0.517
IMPA	94.01 ± 6.01	94.06 ± 6.08	-0.05 ± 0.45	0.578
L1 to NB angle	27.34 ± 5.22	27.41 ± 5.38	-0.17 ± 0.48	0.461
L1 to NB (mm)	6.86 ± 2.94	6.79 ± 2.92	0.07 ± 0.31	0.210
L1 to A-Pg (mm)	3.99 ± 3.19	3.99 ± 3.07	0.53 ± 0.44	0.998

p values from paired t-test.

ated the reliability by comparing deep learning-based tracing and digital tracing as a conventional method. The measurements repeated with the conventional digital cephalometric method showed high reproducibility in both groups (Table 2). This data showed the examiner could accurately reproduce measurements, and there was no difficulty in identifying the landmarks equally. These results correspond with previous studies that showed high reliability of the repeated measurements[12,17,18]. In this study, 2 repeated measurements were obtained to evaluate the reliability of the examiner based on the previous research method[12]. However, a more accurate assessment of intra-observer reliability can be obtained if repeated measurements were performed 3 times.

The influence of examiner's clinical experience on landmark identification is significant along with the tracing method itself. In general, it is known that inter-examiner errors more frequently occur than intra-examiner errors[19]. Therefore, all measurements in this study were obtained by 1 examiner to minimize these errors. However, limitations to the reliability of the results may exist depending on a single examiner's experience and proficiency. Thus, more reliable study results can be obtained by comparing 3 repeated measurements by 2 or more proficient examiners.

Table 3 shows the comparison results between the measurements of 2 methods by paired t-test in the mixed dentition group. The statistically significant differences between the 2 methods were detected in 8 of the 22 cephalometric variables (APDI, SNA, SNB, Mandibular plane angle, LAFH, Facial ratio, U1 to SN, and U1 to A-Pg). The comparison between the measurements of 2 methods in the permanent dentition group is shown in Table 4. In the permanent dentition group, there were no statistically significant difference between the 2 methods in most measurements, except for APDI, Wits appraisal, Facial ratio, and U1 to A-Pg.

The measurements that showed a significant difference in both groups were Facial ratio and U1 to A-Pg. Facial ratio, which is defined as the ratio percent of PFH (S - Go) and AFH (N - Me), is considered to be associated with difficulties in landmark identification itself. Differences in landmark identification used in combination of gonion and menton, which are difficult to detect, might increase the magnitude of the discrepancy. According to previous studies, the uncertainty in detecting the menton point may be caused by the difficulty of locating a landmark on curved anatomical boundaries[17]. The gonion point is defined by bisecting the angle to the tangent to the mandibular plane and the posterior border of the ramus. The detection of this landmark is difficult because it corresponds to an incorrectly defined outline related with bilateral anatomic structures and often projecting as overlapping image[12,17].

U1 to A-Pg, a variable associated with the maxillary incisors, A point and pogonion, is another measurement that showed a significant difference between 2 methods in both groups. In previous studies, it was reported that the tracing of incisor location was difficult and the incisor angular measurement was likely to change depending on the tracing methods[14,20,21]. As the A point is known to have poor reproducibility, in children with mixed dentition, locating A point may be more dif-

ficult because the image is blurred and overlaps due to the presence of tooth germs in the maxillary anterior region[12].

The number of measurements showing a statistically significant difference between the 2 methods was observed more in the mixed dentition group than in the permanent teeth group (8 measurements in the mixed dentition group, 4 measurements in the permanent dentition group). Fig. 1 is an example of the differences between WebCeph and V-Ceph program in landmark identification of a patient with mixed dentition. The radiographs traced by two methods were superimposed. Although the detection of landmarks was within the error range of 2 mm, reported as clinically acceptable in previous studies[10,11], the angular or linear measurements composed of a combination of landmarks may have a larger error due to the summation of the errors. Comparing the measurements showing a significant difference in the 2 groups, the measurements related to the identification error of nasion (N) are observed in the mixed dentition group. Previous research has indicated that it is difficult to identify nasion (N) when the nasofrontal suture is not accurately visualized, suggesting the possibility of such errors[22].

Since WebCeph is a deep learning model trained only on adult with permanent teeth, errors may occur in cephalometric measurements for pediatric and adolescent patients. It would
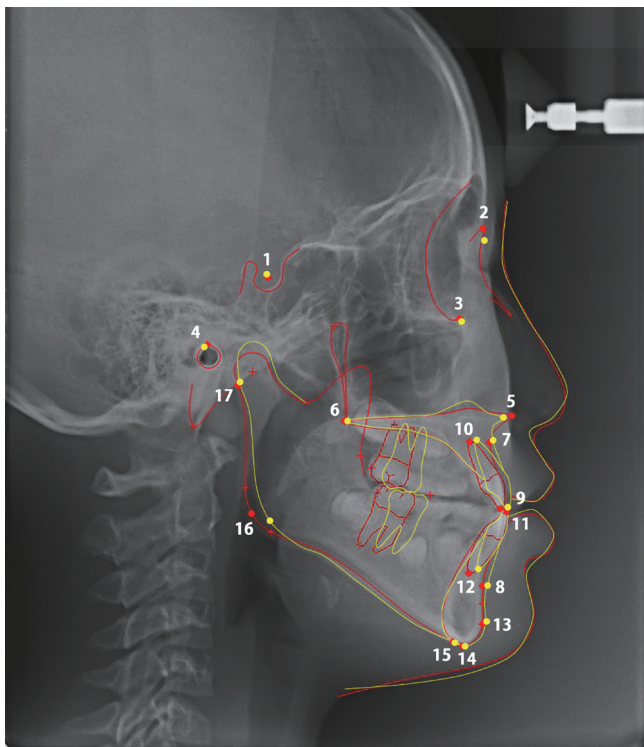

Fig. 1. Differences between the 2 methods in landmark identification of a patient with mixed dentition. While the yellow dots are landmarks detected by the deep learning-based WebCeph program, the red dots are landmarks detected by the conventional V-Ceph program.
1 = Sella Turcica, 2 = Nasion, 3 = Orbitale, 4 = Porion, 5 = Anterior nasal spine, 6 = Posterior nasal spine, 7 = A point, 8 = B point, 9 = Upper incisor edge, 10 = Upper incisor apex, 11 = Lower incisor edge, 12 = Lower incisor apex, 13 = Pogonion, 14 = Gnathion, 15 = Menton, 16 = Gonion, 17 = Articulare

have been difficult to accurately detect skeletal and dental measurements on cephalometric radiographs of patients with mixed dentition. The performance of a diagnostic system could be improved by fine-tuning existing deep learning models[23]. Developing a fine-tuned deep learning-based cephalometric system for patients with mixed dentition could provide more accurate analysis results.

In a previous study, Alqahtani[6] compared measurements obtained from the analysis of the computer program FACAD® (Ilexis AB, Linköping, Sweden) and that of the deep learning-based algorithm CephX® (CephX Inc., Las Vegas, USA). Alqahtani found statistically meaningful differences in 3 of 16 measurements between the 2 methods. Alqahtani reported that analysis using CephX® was fast and efficient, and the measurements obtained from both methods are correlated and highly reproducible. Similarly, this study showed statistically significant differences in 4 of 22 measurements in the permanent dentition group and 8 of 22 measurements in the mixed dentition group between the 2 methods (Table 3 and 4).

This study had the limitation of being conducted in a single center. If the study entailed collecting data from multiple institutions, more general conclusions could have been drawn. This study was conducted using a single program for cephalometric analysis in children and adolescents. Subsequent studies on the comparative evaluation of various deep learning programs for cephalometric analysis will be more meaningful. Another limitation was that the subjects of this study were selected without considering the skeletal class. Additional studies are needed for a more accurate evaluation by designing patient-classified studies according to skeletal class I, II, and III. Also, further assessment studies using larger samples will benefit the clinical applications of deep learning techniques in dentistry.

The significance of this study lies in the fact that it is the first study to study the reliability of a deep learning-based orthodontic analysis method in patients with mixed and permanent dentitions. The evaluation of deep learning-based orthodontic analysis, which provides immediately useful data for treatment planning, is meaningful for clinical applications. In the cephalometric radiographs of mixed dentition, there may be errors in tooth-related landmarks and measurements due to the overlapping of deciduous teeth and the permanent tooth germ. When performing the orthodontic analysis of children and adolescents using the deep learning-based program, it is important for dentists to be aware of the above limita-

tions and do not rely entirely on this program. In clinical practice, dentists make individual corrections by re-positioning the landmarks after obtaining the results of orthodontic analysis from WebCeph. In using a deep learning-based analysis program as an auxiliary tool for orthodontic diagnosis of pediatric and adolescent patients, the dentist's judgment and abundant clinical experience remain important. Understanding the above limitations, using this program appropriately can help pediatric dentists decide whether or not to begin orthodontic treatment.

Ⅴ. Conclusions

Most of the measurements analyzed by the conventional and deep learning-based cephalometric system did not show a significant difference in the permanent dentition group, with the exception of some variables. However, some measurements showing a statistically significant difference between the 2 methods were more in the mixed dentition group. When using deep learning-based programs for orthodontic analysis of pediatric and adolescent patients, it is recommended to recognize the limitations of this program and use it with the proper judgment.

Authors' Information

Sunhahn Shin *https://orcid.org/0000-0002-4029-8046*
Donghyun Kim *https://orcid.org/0000-0003-2442-3254*

References

1. Kim JY, Kwon JY, Kim KH, Park KT : Study on lateral cphalogram of children with normal occlusion in the primary dentition. *J Korean Acad Pediatr Dent,* 32:649-656, 2005.
2. Kim YH, Park JB, Jung SK, *et al*. : Influence of the Depth of the Convolutional Neural Networks on an Artificial Intelligence Model for Diagnosis of Orthognathic Surgery. *J Pers Med*, 11:356-366, 2021.
3. Baumrind S, Frantz RC : The reliability of head film measurements: 1. Landmark identification. *Am J Orthod*, 60:111-127, 1971.
4. Cohen A, Linney A : A Low Cost System for Computer Based Cephalometric Analysis. *Br J Orthod*, 13:105-108, 1986.
5. Seki K, Okano T : Exposure reduction in cephalography

with a digital photostimulable phosphor imaging system. *Dentomaxillofac Radiol*, 22:127-130, 1993.

6. Alqahtani H : Evaluation of an online website-based platform for cephalometric analysis. *J Stomatol Oral Maxillofac Surg*, 121:53-57, 2020.

7. Jung SK, Kim TW : New approach for the diagnosis of extractions with neural network machine learning. *Am J Orthod Dentofacial Orthop*, 149:127-133, 2016.

8. Lee KS, Jung SK, Choi J, *et al*. : Evaluation of transfer learning with deep convolutional neural networks for screening osteoporosis in dental panoramic radiographs. *J Clin Med*, 9:392-404, 2020.

9. Nishimoto S, Sotsuka Y, Kakibuchi M, *et al*. : Personal computer-based cephalometric landmark detection with deep learning, using cephalograms on the internet. *J Craniofac Surg*, 30:91-95, 2019.

10. Arik SÖ, Ibragimov B, Xing L : Fully automated quantitative cephalometry using convolutional neural networks. *J Med Imaging*, 4:014501, 2017.

11. Hwang HW, Moon JH, Lee SJ, *et al*. : Evaluation of automated cephalometric analysis based on the latest deep learning method. *Angle Orthod*, 91:329-335, 2021.

12. Santoro M, Jarjoura K, Cangialosi TJ : Accuracy of digital and analogue cephalometric measurements assessed with the sandwich technique. *Am J Orthod Dentofacial Orthop*, 129:345-351, 2006.

13. Broadbent BH : A new X-ray technique and its application to orthodontia: the introduction of cephalometric radiography. *Angle Orthod*, 51:93-114, 1981.

14. Gravely J, Benzies PM : The clinical significance of tracing error in cephalometry. *Br J Orthod*, 1:95-101, 1974.

15. Kim YH : Web based and artificial intelligence driven orthodontic analysis system. *J Clin Digit Dent*, 1:24-28, 2019.

16. Albarakati S, Kula K, Ghoneima A : The reliability and reproducibility of cephalometric measurements: a comparison of conventional and digital methods. *Dentomaxillofac Radiol*, 41:11-17, 2012.

17. Chen YJ, Chen SK, Chang HF, Chen KC : Comparison of landmark identification in traditional versus computer-aided digital cephalometry. *Angle Orthod*, 70:387-392, 2000.

18. Naoumova J, Lindman R : A comparison of manual traced images and corresponding scanned radiographs digitally traced. *Eur J Orthod*, 31:247-253, 2009.

19. Sayinsu K, Isik F, Trakyali G, Arun T : An evaluation of the errors in cephalometric measurements on scanned cephalometric images and conventional tracings. *Eur J Orthod*, 29:105-108, 2007.

20. Baumrind S, Frantz RC : The reliability of head film measurements: 2. Conventional angular and linear measures. *Am J Orthod*, 60:505-517, 1971.

21. Lim KF, Foong K : Phosphor-stimulated computed cephalometry: reliability of landmark identification. *Br J Orthod*, 24:301-308, 1997.

22. Sekiguchi T, Savara BS : Variability of cephalometric landmarks used for face growth studies. *Am J Orthod*, 61:603-618, 1972.

23. Chi J, Walia E, Eramian M, *et al*. : Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network. *J Digit Imaging*, 30:477-486, 2017.

국문초록

혼합치열과 영구치열 환자를 대상으로 한 웹 기반 인공지능 두부 계측 분석에서의 비교 검증

신선한・김동현

이화여자대학교 의과대학 목동병원 소아치과학교실

이 후향적 연구의 목적은 7 - 15세 사이의 혼합치열기와 영구치열기의 소아 및 청소년 환자에서 기존 교정 분석 방법과 인공 지능을 활용한 교정 분석 방법을 이용한 변수의 차이를 비교하여 평가하는 것이다.

교정 진단을 위해 측면 두부계측 방사선 사진을 촬영한 소아 환자 60명(혼합 치열기 30명, 영구치열기 30명)을 무작위로 선정하였다. V-ceph을 사용한 기존 분석 방법과 WebCeph를 사용한 딥 러닝 기반 분석 방법으로 1명의 검사자가 17개의 두부 측정 계측점을 식별하고, 22개의 측정 항목을 평가했다. 기존 분석 방법의 반복 측정으로 인한 오차는 Pearson의 상관 분석을 사용하여 평가하였다. 혼합치열군과 영구치열군에 대한 각각 두 방법의 차이는 paired t-test를 사용하여 평가하였다.

혼합치열군에서 두 분석 방법의 차이는 8개의 계측항목에서 통계적으로 유의하였다: APDI, SNA, SNB, Mandibular plane angle, LAFH (p < 0.001), Facial ratio (p = 0.001), U1 to SN (p = 0.012), and U1 to A-Pg (p = 0.021). 영구치열군에서는 두 분석 방법 간에 4개의 계측항목이 통계적으로 유의한 차이를 보였다: ODI (p = 0.020), Wits appraisal (p = 0.025), Facial ratio (p = 0.026), and U1 to A-Pg (p = 0.001).

많은 시간이 소요되는 기존의 교정 분석 방법과 비교하였을 때, 딥 러닝 기반 교정 분석 시스템은 측정의 신뢰성과 유효성 측면에서 임상적으로 허용될 수 있다. 하지만 소아 환자의 교정 분석을 위해 딥 러닝 기반 프로그램을 사용할 때에는 이러한 프로그램의 한계점을 인지하고 올바른 판단으로 사용하는 것이 중요하다.