치의과학박사 학위논문

Pattern of Mean Error Value Change
in Artificial Intelligence–assisted Hard Tissue Landmark
Identification of Lateral Cephalogram Images of Class III
Patients Treated with Two–jaw Orthognathic Surgery and
Surgical Orthodontic Treatment

양악 악교정수술과 수술 교정치료를 받은 골격성 III급
부정교합 환자의 측모 두부계측방사선사진 영상에서
인공지능을 이용한 경조직 계측점 식별 시
평균 오차의 변화 양상

2022 년 8 월

서울대학교 대학원

치의과학과 치과교정학 전공

홍 미 희

Pattern of Mean Error Value Change
in Artificial Intelligence–assisted Hard Tissue Landmark
Identification of Lateral Cephalogram Images of Class III
Patients Treated with Two–jaw Orthognathic Surgery and
Surgical Orthodontic Treatment

지도교수 백 승 학

이 논문을 치의과학박사 학위논문으로 제출함

2022 년 6 월

서울대학교 대학원

치의과학과 치과교정학 전공

홍 미 희

홍미희의 치의과학박사 학위논문을 인준함

2022 년 7 월

위 원 장   양 일 형   (인)

부위원장   백 승 학   (인)

위   원   최 진 영   (인)

위   원   허 경 회   (인)

위   원   박 기 호   (인)

−ABSTRACT−

Pattern of Mean Error Value Change
in Artificial Intelligence−assisted Hard Tissue Landmark
Identification of Lateral Cephalogram Images of Class III
Patients Treated with Two−jaw Orthognathic Surgery and
Surgical Orthodontic Treatment

Mihee Hong, DDS, MSD

Department of Orthodontics, Graduate School,

Seoul National University

(Directed by Professor Seung−Hak Baek, DDS, MSD, PhD)

Objective: Recently, auto digitization of hard tissue landmarks on lateral cephalograms (Lat−cephs) has reported, with regard to artificial intelligence models using cascade convolutional neural network (CNN). The aim of this study was to investigate the pattern of accuracy change in artificial intelligence (AI)−assisted hard tissue landmark identification in serial Lat−cephs of Class III patients who underwent two−jaw orthognathic surgery and orthodontic treatment using a cascade CNN algorithm.

Materials and Methods: A total of 3,188 Lat−cephs of 797 Class III patients were allocated into the training and validation sets (3,004 Lat−cephs of 751 patients) and test set (184 Lat−cephs of 46 patients; subdivided into the genioplasty and non−genioplasty groups, n=23 per group) for landmark identification using a cascade CNN model. Each Class III patient in the test set had four Lat−

cephs: initial (T0), pre-surgery [T1, presence of orthodontic brackets (OBs)], post-surgery [T2, presence of OBs and surgical plates and screws (SPS)], and debonding [T3, presence of SPS and fixed retainers (FR)]. After mean errors of 20 hard tissue landmarks between human gold standard and the cascade CNN model were calculated, statistical analysis was performed.

Results: Results are as follows. (1) The total mean error was 1.17 mm without significant difference among the four time-points (T0, 1.20 mm; T1, 1.14 mm; T2, 1.18 mm; T3, 1.15 mm). (2) In comparison of two time-points [(T0, T1) vs. (T2, T3)], ANS, A point, and B point showed an increase in error ($P<0.01$; $P<0.05$; $P<0.01$), while distal contact point of the maxillary first molar (Mx6D) and distal contact point of the mandibular first molar (Md6D) showed a decrease in error ($P<0.01$; $P<0.01$). (3) No difference in errors existed at B point, Pogonion, Menton, crown tip of the mandibular central incisor (Md1C), and root apex of the mandibular central incisor (Md1R) between the genioplasty and non-genioplasty groups.

Conclusion: The cascade CNN model can be used for auto-digitization of hard tissue landmark in serial Lat-cephs including initial, pre-and post-surgery, and debonding time points despite presence of OB, SPS, FR, genioplasty, and bone remodeling.

Keywords: Cascade convolutional neural network, Artificial intelligence, hard tissue landmark identification, serial cephalograms, orthognathic surgery, genioplasty

Student number: 2020-34778

Table of Contents

I. INTRODUCTION

Major dental hospitals in Korea have reported their high prevalence of Class III malocclusion and negative social recognition of the prognathic appearance.[1,2] Numerous patients with skeletal Class III malocclusion in Korea have undergone two-jaw orthognathic surgery (TJ-OGS) along with technical achievement in this area.[3]

The four stages should be performed precisely and routinely for preparing successful TJ-OGS treatment outcome: initial cephalograms were used for diagnosis and gross treatment planning for pre-surgical orthodontic treatment and orthognathic surgery; pre-surgical cephalograms were analysed for planning of the direction and amount at surgical movement; post-surgical cephalograms were evaluated for surgical outcome assessment and post-surgical orthodontic treatment planning; and debonding cephalograms were necessary for comprehensive assessment of orthodontic treatment and orthognathic surgery.[4,5] Furthermore, outcome assessment of pre- and post- surgical orthodontic procedures and TJ-OGS is supported by superimposition of serial cephalograms obtained at each time-point for the same patient. Accurate and reliable identification of cephalometric landmarks is an essential prerequisite to perform these procedures.

An artificial intelligence (AI) algorithm including convolutional neural network (CNN) can help clinicians detect cephalometric landmarks with an accuracy that is close to that of human experts.[6-13] Previous AI studies have regarded the accuracy within a range of 2 mm as a clinically acceptable performance in landmark

1

identification.[9,13−16] However, it appears to be a lenient standard for appropriate clinical use. Therefore, use of stricter criteria (i.e., range within at least 1.5 mm) is necessary in determining the accuracy of landmark identification for clinical relevance.

In addition, previous studies on the accuracy of automated landmark identification[9,14−16] reported their models trained and validated using initial lateral cephalograms only. However, pre− and post−surgical lateral cephalograms, and debonding cephalogram images contain additional metal images including orthodontic brackets (OB), surgical plates and screws (SPS), and fixed retainer (FR). Although initial cephalograms can produce baseline algorithms for AI− supported landmark detection, the existence of OB, SPS, and FR adjacent to landmarks might affect the accuracy of AI detection in serial lateral cephalograms along the four stages.

To the best of our knowledge, no study has compared the accuracy of automated landmark identification in serial cephalograms at the four time−points covering from the initial, pre−surgery, post− surgery, to debonding stages in orthognathic surgery cases. Therefore, the purpose of the study was to investigate the pattern of accuracy change in AI−assisted hard−tissue landmark identification in serial lateral cephalograms of Class III patients who underwent TJ−OGS and pre− and post−surgical orthodontic treatment using a cascade CNN algorithm and strict criteria for determining the degree of accuracy.

II. REVIEW OF LITERATURE

1. High prevalence of Class III skeletal malocclusion and orthognathic surgery in Korea

According to socioeconomic development and interest in appearance, epidemiological studies on malocclusion have been conducted. Im et al.[1] investigated 676 patients who had visited at the department of orthodontics, Seoul National Dental Hospital at 1992 and 2002. The percentage of Class III malocclusion were 54.4% and 48.1%, respectively. Mandibular prognathism was one of the major chief complaints and its percentage was 24.4% and 17.5 %, respectively. The portion of patients treated with orthognathic surgery were increased from 14.8% at 1992 and 25.0% at 2002.

Yeongdong severance dental hospital analyzed their data from 2008 to 2012. Piao et al.[2] reported that skeletal Class III malocclusion based on ANB angle was 31.4% among 7,476 patients and the orthognathic surgery rate was 18.5% among 4,861 actual orthodontic cases. Furthermore, 70% of surgical patients had skeletal Class III malocclusion. Along with the advancement of surgical technique, severe skeletal Class III patients have been treated with TJ−OGS in Korea.[3]

In summary, the half of patients in orthodontic department of the major university dental hospitals in Seoul, Korea was Class III malocclusion. Furthermore, more than 2/3 of orthognathic surgery patients belonged to Class III malocclusion.

3

2. The role of lateral cephalograms in diagnosis, planning, treatment, and assessment for orthognathic surgery with orthodontic treatment

Cephalometric measurement and analysis are clinically crucial in diagnosing and planning, especially for surgical treatment. Nielsen[4] compared three methods of cephalometric evaluation of growth and treatment change by superimposing serial headfilms, especially for the maxilla. This study revealed that the "best fit" method, which superimposed at ANS, underestimates the vertical eruption of the teeth by 30% to 50%. However, the structural method proved no significant differences in vertical displacement of the selected landmarks when compared with the implant method.

Johnston et al.[5] investigated the treatment outcome in patients with Class III malocclusion, who underwent surgical orthodontic treatment, using consecutive patients' cephalometric images. They mentioned that the dentoskeletal parameters influencing the clinical decision of orthognathic surgery are as follows: ANB, mandibular incisor inclinations, Wits measurement, gonial angle, Sella-Nasion distance, and maxillary-mandibular length ratio. In addition, they evaluated achievement of ideal inclination of the maxillary and mandibular incisors in their pre-surgical orthodontic treatment phases. Half of the mandibular incisors was retroclined (< 87°); maxillary incisors showed significant compensatory proclination (> 115°).

Since overjet and maxillary incisor inclination could be significantly different depending on the recording timing after surgery (i.e. mild

relapse), it is necessary to use standard interval for taking surgical cephalometric films as record-taking protocols.

3. Artificial intelligence in cephalometric landmark identification

Treatment outcome can be presented as cephalometric values relative to ideal or acceptable ranges at standardized specific timepoints. However, the cephalometric values rely on the accuracy of landmark identification. Drawbacks of manual analysis of cephalograms are time-consuming and inter-observer variability.

Hutton et al.[6] evaluated the application of active shape models for cephalometric landmark identification. A training set of hand-annotated images with the resulting models was applied to 63 test images. 13% of the 16 landmarks were within 1 mm, 35% within 2 mm, and 74% within 5 mm. Although there was not enough accuracy for automated landmark identification, there was possibility of a time-saving tool as a first-estimation of the landmarks.

Leonardi et al.[7] described several techniques for automatic landmark identification of cephalograms including image filtering plus knowledge-based landmark search, model-based approaches, and soft-computing approaches, and hybrid approaches.

They mentioned that automatic cephalometric analysis from 1966 to 2006 was methodologically unsound in terms of inclusion criteria, error level, and standard deviation of the mean error. From their literature review, they suggested that less than 2 mm difference

between automatic landmark location and human operator was considered successful and less than 4 mm distance as an acceptable distance. Furthermore, they implied that there would be higher level of errors in reference plane/line constructed by two landmarks. In addition, they pointed out that the landmark identification errors should be understood in aspect of following a pattern envelop. Overall, errors of the automated system were greater than those of the manual approaches, which means that it is not useful for clinical purposes.

Due to possible inaccuracies by converting procedures from head films into digital images, it is necessary to perform landmarks identification from direct digital cephalograms. Therefore, Leonardi et al.[8] evaluated the accuracy of automatic landmarks identification for 41 direct digital cephalograms using Cellular Neural Networks algorithms, which comprised of neurons, the set of processing elements. 10 landmarks including Nasion, A Point, Basion, Porion, Pterygoid, B Point, Pogonion, Protuberance Menti, upper incisor edge, and lower Incisor edge were used for automatic identification. They reported very small difference at most within 0.59 mm. Nasion, Pogonion, Porion, and Protuberance Menti were more reliable in the horizontal dimension and A point and Pterygoid point in the vertical dimension.

In summary, the errors of automated landmarks identification were reduced using soft copies of digital cephalograms and improved algorithms. Although various automated landmark detection methods were introduced, it is still questionable how to achieve the clinically acceptable 'within 2 mm' range.

Lee et al.[10] reported a new framework for identifying cephalometric landmarks using the Bayesian Convolutional Neural Networks, which comprised extraction of 'region of interest' landmarks and estimation of landmark position. A mean landmark error was 1.53 ± 1.74 mm. A successful detection rate (SDR) was 82.11% in 2 mm range, 92.28% in 3 mm rage, and 95.95% in 4 mm. This framework presented the possibility of automatic cephalometric landmarks as a diagnosis tool and for education.

Vandaele et al.[11] proposed a multi-resolution tree-based method for accelerating landmark detection. The key parameters were identified and were compared to expert ground truths. The algorithms are integrated in the open source Cytomine software (Cytomine Co. SA, Liege, Belgium) with parameter configuration guidelines for end-users.

Khanagar et al.[12], in their systematic review, reported that AI technology has been used for cephalometric landmarks identification, orthodontic extractions decision, cervical vertebra maturation determination, orthognathic surgery simulation for facial attractiveness, prediction of orthodontic treatment need, and planning orthodontic treatment. Artificial neural networks or convolutional neural networks are major algorithms. Simplifying the tasks in quick time can increase efficiency.

Kim et al.[13] investigated the accuracy of landmarks identification algorithm on 100 cephalograms from 10 university hospitals in Korea using the cascade CNN model. A total of 3,150 lateral cephalograms were collected from multi-centers for training of an automated landmark digitization model. For 100 test set, two

orthodontists digitized the anatomic landmarks using the V-ceph software (version 8.0, Osstem, Seoul, Korea). The mean position between two orthodontists was used for the gold standard of landmark position (1.31 ± 1.13 mm). Absolute distance between the mean position between two orthodontists and the automatic digitized position by the algorithm was 1.36 ± 0.98 mm. The authors considered the cascade CNN model as a preliminary screening tool for diagnosis and assessment of cephalograms.

4. Acceptable range and success criteria

Arık et al.[9] defined the SDR as "ratio of the corresponding landmarks within the proximity of the precision range from the ground truth location". Further, they mentioned clinically acceptable SDR of 2 mm. They reported that the 67% to 75% of accuracy was achieved for their framework.

Wang et al.[14] investigated the accuracy of automatic landmark detection approaches of cephalograms of 300 patients. Ground truth data was prepared manually by two experts. When 4 mm precision range is regarded acceptable in clinical practice, only three approaches presented that 80% of landmarks remained within 4 mm precision range. However, when 2 mm precision range is regarded acceptable in clinical practice, only one approach presented that 70% of landmarks remained within 2 mm precision range.

Wang et al.[15] reported that manual tracing of landmarks and anatomic structures during orthodontic treatment planning is subjective as well as time consuming. As a solution, they compared seven automatic methods for analyzing cephalometric X-ray image.

8

Main criteria for evaluation included "mean radial error", and "successful detection rate" for automatic cephalogram landmark identification.

The formulation of mean radial error R is $R = \sqrt{\Delta x^2 + \Delta y^2}$. Δx indicates absolute distance of the horizontal direction from the obtained landmark to the referenced landmarks. Δy means the absolute distance of the vertical direction between the two landmarks. For "successful detection rate", if the absolute difference between the two landmarks is no greater then z mm, the detection of landmark is regarded as being successful. $p_z = \frac{\#\{j: \|L_d(j) - L_r(j)\| < z\}}{\#\Omega} \times 100\%$; L_d: the location of the detected landmark; L_r: the location of the referenced landmark; z: precision measurements for evaluation, such as 2 mm, 3 mm and 4 mm; $j \in \Omega$; $\# \Omega$: the number of detections

Hwang et al.[16] tested their automatic identification whether the new AI algorithm was superior to clinically trained human experts. They compared the mean detection errors between AI and human to those between human examiners (1.46 ± 2.97 mm vs. 1.50 ± 1.48 mm). As a result, AI showed the same degree of accuracy in identifying cephalometric landmarks compared to human experts.

III. MATERIALS AND METHODS

1.Data set

A total of 3,188 lateral cephalograms of 797 patients with Class III malocclusion were used for the training and validation sets and the test set for automated landmark identification using the cascade CNN model. The inclusion criteria were as follows: (1) Class III patient who underwent pre- and post-surgical orthodontic treatment and TJ-OGS with/without genioplasty and (2) Class III patient whose serial lateral cephalograms were available (3) Class III patient whose digital cephalometric radiograms can be collected using digital imaging and communication in medicine (DICOM) format. The exclusion criterion was Class III patient who had craniofacial deformities.

The training and validation sets for automated landmark identification by the cascade CNN model included 3,004 lateral cephalograms of 751 Class III patients from 10 institutions (Table 1). Some of the patients who belonged to the training or validation set had more than four lateral cephalograms because additional progress lateral cephalograms were taken between time-points, while some of them had missing lateral cephalograms at specific timepoints.

For the test set, Class III patients with cephalograms obtained at the following four timepoints were selected: initial (T0), pre-surgery (T1, taken at least 1 month before TJ-OGS; presence of

10

OBs), post-surgery (T2, taken at least 2 months after TJ-OGS; presence of OBs and S-PS), and debonding (T3, presence of S-PS, FR, and bone remodeling change). As a result, the test set consisted of 184 cephalograms of 46 Class III patients from eight institutions (Table 1). It was subdivided into the genioplasty and non-genioplasty groups (n = 23 patients per group). Their characteristics are enumerated in Figure 1.

2. Ethical approval

This nationwide multicenter study was reviewed and approved by the Institutional Review Board (IRB) Committee of 10 institutions: Seoul National University Dental Hospital (ERI18002), Kyung Hee University Dental Hospital (KH-DT19006), Kyungpook National University Dental Hospital (KNUDH-2019-03-02-00), Asan Medical Center (2019-0408), Ewha University Medical Center (EUMC 2019-04-017-009), Wonkwang University Dental Hospital (WKDIRB201903-01), Ajou University Dental Hospital (AJIRB-MED-MDB-19-039), Korea University Anam Hospital (K2019-0543-010), Chonnam National University Dental Hospital (CNUDH-EXP-2021-001), and Chosun University Dental Hospital (CUDHIRB 1901 005 R01).

3. Cascade CNN

Data sets were obtained from 10 centers using anonymized DICOM file format. Since finding the exact location of hard tissue landmarks in a large lateral cephalogram image is relatively difficult, a fully

11

automated landmark prediction algorithm with the cascade network was developed. [13] Two steps were followed: 1) detection of the region of interest (256 × 256 and 512 × 512 pixels depending on the landmark) using the RetinaNet[17] and 2) prediction of the landmark using the U−Net[18] (Figure 2).

4. Cephalometric landmarks

Definitions of 12 skeletal and eight dental landmarks are presented in Figure 3 and Table 2. The landmarks were digitized by a single orthodontist who had 20 years of experience (human gold standard) and by the cascade CNN model.

5. Intra−examiner reliability

Twenty randomly selected lateral cephalogram images were re− digitized with an interval of 2 weeks by the same operator. The measurement error for the whole landmarks were assessed using Dahlberg's formula. The overall Dahlberg's error was 0.38 mm, ranged from 0.07 mm to 0.61 mm by each landmark. (Table 3)

6. Definition of errors (Table 4)

The mean values of absolute errors for each landmark were calculated using the absolute distance between the human gold standard and AI−assisted detection. The degree of error (Err) was allocated into excellent (Err < 1.0 mm), good (1.0 ≤ Err <1.5 mm), fair (1.5 ≤ Err < 2.0 mm), acceptable (2.0 ≤ Err < 2.5 mm), and

unacceptable (2.5 mm ≤ Err) groups.

Then, the accuracy percentage (AP) was calculated using a formula (percentage of the excellent and good groups among the total degree of error groups), which means that the error range within 1.5 mm was considered accurate. The degree of accuracy was defined as "very high" ($90\% \leq$ AP), "high" ($70 \leq$ AP $< 90\%$), "medium" ($50 \leq$ AP $< 70\%$), and "low" (AP $< 50\%$).

7. Statistics

Repeated measures ANOVA, and post-hoc test for within-subject by Tukey's adjustment for multiple comparisons were performed to find out the difference between T0, T1, T2, and T3 stages. Repeated measures Multivariate Analysis of Variance (MANOVA) was performed to find out the difference between before-surgery stage (T0 and T1) and after surgery stage (T2 and T3). Statistical analysis was done using SPSS ver. 23.0 (IBM Corp., Armonk, NY, USA) and SAS 9.4 (SAS Institute Inc., Cary, NC, USA.). P-values of < 0.05 were considered statistically significant.

IV. RESULTS

1. Evaluation of the total landmarks (Table 4)

The total landmarks showed a good mean error value (1.17 mm), and the total AP had a high degree of accuracy (74.2%).

2. Evaluation of the skeletal landmarks (Table 4, Figure 4)

Nasion and Sella showed an excellent mean error value and a very high degree of accuracy (0.59 mm and 95.1%; 0.46 mm and 100%, respectively), while Porion and Orbitale showed a good mean error value and a high degree of accuracy (1.07 mm and 76.1%; 1.21 mm and 73.9%, respectively). On the other hand, Basion showed a fair mean error value (1.64 mm) and a medium degree of accuracy (63.1%).

ANS and A point showed a good mean error value and a medium degree of accuracy (1.39 mm and 65.2%; 1.41 mm and 63.0%, respectively). PNS had a good mean error value (1.19 mm) and a high degree of accuracy (72.7%). Pogonion, Menton and Articulare showed an excellent mean error value and a very high degree of accuracy (0.79 mm and 91.3% and 0.77 mm and 93.5%, 0.77 mm and 93.5%, respectively). B point showed a good mean error value (1.15 mm) and a high degree of accuracy (77.2 %).

3. Evaluation of the dental landmarks (Table 4, Figure 4)

The crown tip of the maxillary central incisor (Mx1C) showed an excellent mean error value (0.44 mm) and a very high degree of accuracy (97.8%) while the distal contact point of the maxillary first molar (Mx6D) had a good mean error value (1.43 mm) and a medium degree of accuracy (64.1%). On the other hand, the root apex of the maxillary central incisor (Mx1R) and the distobuccal root apex of the maxillary first molar (Mx6R) had a fair mean error value and a medium degree of accuracy (1.55 mm and 57.6%; 1.68 mm and 51.6%, respectively).

The crown tip of the mandibular central incisor (Md1C) demonstrated an excellent mean error value (0.49 mm) and a very high degree of accuracy (97.3%), while the root apex of the mandibular central incisor (Md1R) had a fair mean error value (1.57 mm) and a medium degree of accuracy (58.2%). The distal contact point of the mandibular first molar (Md6D) had a fair mean error value (1.67 mm) and medium accuracy (51.6%), and distal root apex of the mandibular first molar Md6R exhibited an acceptable mean error value (2.03 mm) and a low degree of accuracy (41.3%).

4. Comparison of the mean errors among the four timepoints (T0, T1, T2, and T3) (Table 5)

No significant difference was found in the overall mean errors (P > 0.05). Only three landmarks, namely ANS, Mx6D, and Md6D

15

showed a significant difference in the mean errors among the four timepoints (ANS, increase in the mean error from T0 and T1 to T2, $P < 0.01$; Mx6D, decrease in the mean error from T0 to T2, $P < 0.05$; Md6D, decrease in the mean error from T0 to T2 and T3, $P < 0.01$).

5.Comparison of the mean errors between the two timepoints [(T0, T1) vs. (T2, T3)] (Table 5, Figures 5 and 6)

ANS, A point, and B point showed an increase in the mean error after TJ−OGS than before TJ−OGS (ANS, $P < 0.01$; A point, $P < 0.05$; B point, $P < 0.01$), while Mx6D and Md6D showed a decrease in the mean error after TJ−OGS than before TJ−OGS (all $P < 0.01$).

6.Comparison of the mean errors between the genioplasty and non−genioplasty groups (Table 6)

No significant difference in the mean errors in the landmarks located adjacent to the genioplasty area (B point, Pogonion, Menton, Md1C, and Md1R) existed in each timepoint between the two groups, except Md1R at T1 ($P<0.05$).

V. DISCUSSION

Since TJ—OGS induces the position change and bone remodeling in the skeletal structures and produces the metallic images of the OB, SPS, and FR, the accuracy and reliability of cephalometric hard-tissue landmark identification in serial lateral cephalograms are important for assessment of treatment outcomes.[19]

As total landmarks exhibited a good mean error value and a high degree of accuracy (1.17 mm and 74.2%, respectively, Table 4) without significant difference among the four time-points (T0, 1.20 mm; T1, 1.14 mm; T2, 1.18 mm; T3, 1.15 mm; P > 0.05, Table 5), accuracy of the AI—assisted digitization was not significantly affected by the presence of OB, SPS, FR, and bone remodeling change during orthodontic treatment and TJ—OGS.

Regardless of the degree of accuracy of each landmark (Table 4, Figure 4), none of the five cranial base landmarks exhibited a significant difference in the mean errors among the four time-points (T0, T1, T2 and T3) and between the two time-points [(T0, T1) vs. (T2, T3)] (Table 5). In this study, the mean error of Porion was 1.07 mm, while previous studies reported their errors with a range of 0.53 to 1.89mm.[8,10,14,16] Accuracy of the cranial base landmarks can be regarded as baseline for comparison of serial lateral cephalograms because the positions of these cranial base landmarks are not affected by TJ—OGS.

Three error patterns were found in the maxillary skeletal landmarks.

First, the mean errors of ANS were different among the four time-points (T0, 1.07 mm; T1, 1.22 mm; T2, 1.78 mm; T3, 1.49 mm, $P <$ 0.01; Table 5, Figure 5) and presented an increased error value after TJ-OGS than before it [(T0, T1) vs. (T2, T3), $P < 0.01$; Table 5, Figure 6]. This suggested that the metal image of the SPS adjacent to ANS as well as surgical shape modification of ANS [20, 21] (Figure 1) could affect the accuracy of AI-assisted landmark detection.

Second, although the error of A point was not significantly different among the four time-points (T0, 1.27 mm; T1, 1.28 mm, T2, 1.50 mm, T3, 1.59 mm, Table 5), it presented an increase in the mean error value after TJ-OGS than before it [(T0, T1) vs. (T2, T3), $P < 0.05$; Table 5, Figure 6]. This occurred because A point might be less affected by the metal image of the SPS installed at the maxilla and have a lower chance for surgical shape modification, compared to ANS (Figure 1). Furthermore, A point might be less affected by the metal image of SPS installed lateral to the pyriform aperture in the maxilla and have a lower chance for surgical shape modification relative to ANS.

Third, in case of posterior impaction and/or anteroposterior movement of the maxilla, the position of PNS had to be changed. However, for PNS, no significant difference was found either among the four time-points (T0, 1.16 mm; T1, 1.14 mm, T2, 1.29 mm, T3, 1.17 mm; $P > 0.05$, Table 5) or between the two time-points [(T0, T1) vs. (T2, T3), $P > 0.05$; Table 5]. No significant difference in accuracy between time points means that the amount error of landmark at four or two timepoints was neither significantly

18

increased nor decreased. This might be due to (1) an absence of the metal image of the SPS within the region of interest of PNS and (2) an easily defined the end point of the hard palate.

There are three patterns for the errors in the mandibular skeletal landmarks. First, since there were no metal images within the region of interest of Articulare and Menton, their mean errors were not significantly different among the four time-points and between the two time-points (all P > 0.05, Table 5).

Second, the mean error of Pogonion was not significantly different among the four time-points and between the two time-points (P > 0.05; Table 5), which suggests that the metal image of the SPS adjacent to Pognion (Figure 1) might not affect the accuracy of AI-assisted landmark detection.

Third, although the mean errors of B point did not significantly differ among the four time-points (T0, 1.00 mm; T1, 1.01 mm; T2, 1.29 mm; T3, 1.31 mm, P > 0.05; Table 5), comparison of the two time-points revealed an increase in error after TJ-OGS than before TJ-OGS [(T0, T1) vs. (T2, T3), P < 0.01; Table 5, Figure 6]. These findings suggest that the metal image of the SPS adjacent to the B point (Figure 1) might affect the accuracy of AI-assisted landmark detection.

There are two sources of errors in the dental landmarks. First, regardless of the degree of accuracy in the dental landmarks (Table 4), Mx1C, Md1C, Mx1R, Md1R, Mx6R, and Md6R did not exhibit significant difference in the mean errors among the four time-

points and between the two time-points (all P > 0.05; Table 5). Second, the mean errors of Mx6D and Md6D were significantly different among the four time-points (Mx6D: T0, 1.66 mm; T1, 1.63 mm, T2, 1.20 mm, T3, 1.23 mm; Md6D, T0, 2.15 mm; T1, 1.71 mm, T2, 1.51 mm, T3, 1.33 mm; all P < 0.01, Table 5, Figure 5) and presented decreased mean error values after TJ-OGS than before TJ-OGS [(T0, T1) vs. (T2, T3), all P < 0.01; Table 5, Figure 6].

The possible reasons for these might be the following: (1) Horizontal and vertical overlapping of the right and left maxillary and mandibular first molars made it difficult to accurately locate the Mx6D and Md6D at T0 lateral cephalogram; and (2) Orthodontic treatment and TJ-OGS improved the alignment of the maxillary and mandibular dentition and corrected the cant, shift and yaw of the maxilla and mandible, reducing the discrepancy between double images of the maxillary and mandibular first molars.

The positions of the cranial base landmarks were not changed at all by orthognathic surgery. In addition, the positions of Articulare is not affected by orthognathic surgery due to its radiographic definition.[3,13]

The positions of the skeletal landmarks in the maxilla and mandible are directly affected by orthognathic surgery, while the locations of the dental landmarks are indirectly or relatively changed from the surgical movement of the maxilla and mandible.

Since the dental landmarks including Mx6D, Mx6R, Md6D, and Md6R have double images especially in facial asymmetry cases,

they were digitized on the average point of the right and left teeth. However, the adverse effect of facial asymmetry on the identification accuracy of the positions of the maxillary and mandibular molars at T0 and T1 were decreased at T2 and T3 along with restoring facial asymmetry by orthognathic surgery (Table 5).

No significant difference was found in the mean errors in the landmarks adjacent to the genioplasty area including B point, Pogonion, Menton, Md1C, and Md1R (all $P > 0.05$, Table 6). The possible reasons for this are as follows: (1) Menton and Md1C were located relatively far from the SPS installed at the symphysis and their shapes were not affected by genioplasty; (2) Since Pogonion and B point are the most forward and deepest points on the anterior surface of the symphysis, respectively, they can be easily identified despite the presence of the metal image of the SPS; and (3) Although Md1R had a fair mean error value and a medium degree of accuracy (1.57 mm and 58.2%, respectively), these patterns were not aggravated at T2 and T3 despite the presence of the metal image of the SPS.

The cascade CNN algorithm proposed in this study showed a possibility of hard tissue landmark identification from bony anatomies in serial lateral cephalograms despite the presence of OB, SPS, FR, genioplasty, and bone remodeling. However, since Mx1R, Mx6R, Md1R, Md6D, and Md 6R showed more than 1.5 mm of error and less than 60% of accuracy percentage, it is necessary to increase the accuracy and reliability of landmark identification of the dental landmarks, especially the distal root apex of the

21

mandibular first molar in the further study. When the AI-assisted hard tissue landmark identification is used, clinicians should consider these characteristics.

VI. CONCLUSION

The cascade CNN model can be used for auto-digitization of hard tissue landmark in serial Lat-cephs including initial, pre-and post-surgery, and debonding time points despite presence of OB, SPS, FR, genioplasty, and bone remodeling.

REFERENCES

1. Im DH, Kim TW, Nahm DS, Chang YI. Current trends in orthodontic patients in Seoul National University Dental Hospital. Korean J Orthod. 2003;33:63-72.

2. Piao Y, Kim SJ, Yu HS, Cha JY, Baik HS. Five-year investigation of a large orthodontic patient population at a dental hospital in South Korea. Korean J Orthod. 2016;46:137-45.

3. Lim SW, Kim M, Hong M, Kang KH, Kim M, Kim SJ et al. Comparison of one-jaw and two-jaw orthognathic surgery in patients with skeletal Class III malocclusion using data from 10 multi-centers in Korea: Part 1. Demographic and skeletodental characteristics. Korean J Orthod. 2022;52:66-74.

4. Nielsen IL. Maxillary superimposition: a comparison of three methods for cephalometric evaluation of growth and treatment change. Am J Orthod Dentofacial Orthop. 1989;95:422-31.

5. Johnston C, Burden D, Kennedy D, Harradine N, Stevenson M. Class III surgical-orthodontic treatment: a cephalometric study. Am J Orthod Dentofacial Orthop. 2006;130:300-9.

6. Hutton TJ, Cunningham S, Hammond P. An evaluation of active shape models for the automatic identification of cephalometric landmarks. Eur J Orthod. 2000;22:499-508.

7. Leonardi R, Giordano D, Maiorana F, Spampinato C. Automatic cephalometric analysis. Angle Orthod. 2008;78:145-51.

8. 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; 717102:1-12.

9. Arık SÖ, Ibragimov B, Xing L. Fully automated quantitative cephalometry using convolutional neural networks. J Med Imaging (Bellingham). 2017;4:014501-11.

10. Lee JH, Yu HJ, Kim MJ, Kim JW, Choi J. Automated cephalometric landmark detection with confidence regions using Bayesian convolutional neural networks. BMC Oral Health. 2020;20:270-9.

11. Vandaele R, Aceto J, Muller M, Péronnet F, Debat V, Wang CW, et al. Landmark detection in 2D bioimages for geometric morphometrics: a multi−resolution tree−based approach. Sci Rep. 2018;8:538-51.

12. Khanagar SB, Al−Ehaideb A, Vishwanathaiah S, Maganur PC, Patil S, Naik S, et al. Scope and performance of artificial intelligence technology in orthodontic diagnosis, treatment planning, and clinical decision−making − A systematic review. J Dent Sci. 2021;16:482-92.

13. Kim J, Kim I, Cho JH, Hong MH, Kang KH, Kim MJ, et al. Accuracy of automated identification of lateral cephalometric landmarks using cascade convolutional neural networks on lateral cephalograms from nationwide multi−centers. Orthod Craniofac Res. 2021;24 Suppl 2:59-67.

14. Wang CW, Huang CT, Hsieh MC, Li CH, Chang SW, Li WC, et al. Evaluation and Comparison of Anatomical Landmark Detection Methods for Cephalometric X−Ray Images: A Grand Challenge. IEEE Trans Med Imaging. 2015;34:1890-900.

15. Wang CW, Huang CT, Lee JH, Li CH, Chang SW, Siao MJ, et al. A benchmark for comparison of dental radiography analysis algorithms. Med Image Anal. 2016;31:63-76.

16. Hwang HW, Park JH, Moon JH, Yu Y, Kim H, Her SB, et al.

9. Arık SÖ, Ibragimov B, Xing L. Fully automated quantitative cephalometry using convolutional neural networks. J Med Imaging (Bellingham). 2017;4:014501-11.

10. Lee JH, Yu HJ, Kim MJ, Kim JW, Choi J. Automated cephalometric landmark detection with confidence regions using Bayesian convolutional neural networks. BMC Oral Health. 2020;20:270-9.

11. Vandaele R, Aceto J, Muller M, Péronnet F, Debat V, Wang CW, et al. Landmark detection in 2D bioimages for geometric morphometrics: a multi−resolution tree−based approach. Sci Rep. 2018;8:538-51.

12. Khanagar SB, Al−Ehaideb A, Vishwanathaiah S, Maganur PC, Patil S, Naik S, et al. Scope and performance of artificial intelligence technology in orthodontic diagnosis, treatment planning, and clinical decision−making − A systematic review. J Dent Sci. 2021;16:482-92.

13. Kim J, Kim I, Cho JH, Hong MH, Kang KH, Kim MJ, et al. Accuracy of automated identification of lateral cephalometric landmarks using cascade convolutional neural networks on lateral cephalograms from nationwide multi−centers. Orthod Craniofac Res. 2021;24 Suppl 2:59-67.

14. Wang CW, Huang CT, Hsieh MC, Li CH, Chang SW, Li WC, et al. Evaluation and Comparison of Anatomical Landmark Detection Methods for Cephalometric X−Ray Images: A Grand Challenge. IEEE Trans Med Imaging. 2015;34:1890-900.

15. Wang CW, Huang CT, Lee JH, Li CH, Chang SW, Siao MJ, et al. A benchmark for comparison of dental radiography analysis algorithms. Med Image Anal. 2016;31:63-76.

16. Hwang HW, Park JH, Moon JH, Yu Y, Kim H, Her SB, et al.

Automated identification of cephalometric landmarks: Part 2-Might it be better than human?. Angle Orthod. 2020;90:69-76.

17. Lin TY, Goyal P, Girshick R, He K, Dollar P. Focal Loss for Dense Object Detection. IEEE Trans Pattern Anal Mach Intell. 2020;42:318-27.

18. Ronneberger O, Fisher P, U-Net BT.: Convolutional Networks for Biomedical Image Segmentation. International Conference on Medical Image computing and Computer-Assisted Intervention. 2015;234-41.

19. Roden-Johnson D, English J, Gallerano R. Comparison of hand-traced and computerized cephalograms: landmark identification, measurement, and superimposition accuracy. Am J Orthod Dentofacial Orthop. 2008;133:556-64.

20. Seigo O, Noriko N, Yuya N, Hitoshi Y, Tokutano M, Takako K, et al. Effects of vertical movement of the Anterior Nasal Spine on the Maxillary stability after LeFort 1 Osteotomy for Pitch correction. J Craniofacial Surg. 2015;26:481-85.

21. Venkategowda PR, Prakash AT, Roy ET, Shetty KS, Thakkar S, Maurya R. Stability of vertical, horizontal and angular parameters following superior repositioning of Maxilla by Le Fort I osteotomy: A cephalometric study. J Clin Diagn Res. 2017;11:ZC10-ZC14.

Table 1. Composition of the training, validation, and test datasets

		Training set	Validation set	Test set	sum
A	Seoul National University Dental Hospital	1,292	100	52	1,444
B	Kyung Hee University Dental Hospital	607	100	48	755
C	Kyungpook National University Dental Hospital	133	30	20	183
D	Asan Medical center	144	32	24	200
E	Ewha University Medical center	116	20	12	148
F	Wonkwang University Dental Hospital	95	26	8	129
G	Ajou University Dental Hospital	84	20	12	116
H	Korea University Anam Hospital	62	25	0	87
I	Chonnam National University Dental Hospital	48	16	8	72
J	Chosun University Dental Hospital	45	9	0	54
Total	lateral cephalograms	2,626	378	184	3,188
	Class III patients	751		46	797

Table 2. The definition of hard–tissue cephalometric landmarks

Compartment			Landmark	Description
Skeletal landmark	Cranial Base		Nasion (N)	The most anterior point on the frontonasal suture in the midsagittal plane
			Sella (S)	Center of the Sella Turcica
			Porion (Por)	The most superior point of the external auditory meatus
			Orbitale (Or)	The most inferior point of the orbital cavity contour
			Basion (Ba)	The most posterior and inferior point of the occipital bone
	Maxilla	anterior	ANS	The tip of anterior nasal spine
			A point	The deepest point between ANS and the upper incisal alveolus
		posterior	PNS	The most posterior point of the hard palate
	Mandible	anterior	B point	The deepest point between Pogonion and the lower incisal alveolus
			Pogonion (Pog)	The most anterior point on the symphysis
		posterior	Articulare (Ar)	Intersection between the inferior cranial base surface and the posterior surface of condyle
		bottom	Menton (Me)	The most inferior point on the symphysis
Dental landmark	Maxillary dentition	anterior	Mx1C	Crown tip of the maxillary central incisor
			Mx1R	Root apex of the maxillary central incisor
		posterior	Mx6D	Distal contact point of the maxillary first molar
			Mx6R	Distobuccal root apex of the maxillary first molar
	Mandibular dentition	anterior	Md1C	Crown tip of the mandibular central incisor
			Md1R	Root apex of the mandibular central incisor
		posterior	Md6D	Distal contact point of the mandibular first molar
			Md6R	Distal root apex of the mandibular first molar

Table 3. Intra-examiner reliability using Dahlberg's error

Compartment			Landmark	Dahlberg's error (mm)
Skeletal landmark	Cranial Base		Nasion	0.39
			Sella	0.07
			Porion	0.48
			Orbitale	0.25
			Basion	0.35
	Maxilla	anterior	ANS	0.48
			A point	0.48
		posterior	PNS	0.52
	Mandible	anterior	B point	0.45
			Pogonion	0.28
		bottom	Menton	0.61
		posterior	Articulare	0.28
Dental landmark	Maxillary dentition	anterior	Mx1C	0.14
			Mx1R	0.32
		posterior	Mx6D	0.18
			Mx6R	0.16
	Mandibular dentition	anterior	Md1C	0.15
			Md1R	0.43
		posterior	Md6D	0.46
			Md6R	0.56
Total				0.38

Table 4. The absolute values of error, distribution of error, accuracy percentage, and degree of accuracy for each landmark

Compartment			Landmark	Absolute value of error (Err)		Distribution (number/percentage)					Accuracy	
				Mean (mm) [Q1, Q3]	SD (mm)	Excellent (Err<1.0 mm)	Good (1.0≤Err<−1.5 mm)	Fair (1.5 ≤Err<2.0 mm)	Acceptable (2.0 ≤Err<2.5 mm)	Unacceptable (2.5 mm ≤Err)	Accuracy Percentage (number/percentage)	Degree of Accuracy
Skeletal landmark	Cranial Base		Nasion	0.59 [0.27, 0.76]	0.48	157 (85.3%)	18 (9.8%)	4 (2.2%)	3 (1.6%)	2 (1.1%)	175 (95.1%)	very high
			Sella	0.46 [0.27, 0.60]	0.23	180 (97.8%)	4 (2.2%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	184 (100%)	very high
			Porion	1.07 [0.60, 1.46]	0.69	103 (56.0%)	37 (20.1%)	24 (13.0%)	14 (7.6%)	6 (3.3%)	140 (76.1%)	high
			Orbitale	1.21 [0.49, 1.54]	1.01	92 (50.0%)	44 (23.9%)	21 (11.4%)	12 (6.5%)	15 (8.2%)	136 (73.9%)	high
			Basion	1.64 [0.60, 2.06]	1.61	82 (44.6%)	34 (18.5%)	21 (11.4%)	13 (7.1%)	34 (18.5%)	116 (63.1%)	medium
	Maxilla	anterior	ANS	1.39 [0.60, 1.83]	1.01	72 (39.1%)	48 (26.1%)	23 (12.5%)	14 (7.6%)	27 (14.7%)	120 (65.2%)	medium
			A point	1.41 [0.70, 2.03]	0.99	86 (46.7%)	30 (16.3%)	22 (12.0%)	18 (9.8%)	28 (15.2%)	116 (63.0%)	medium
		posterior	PNS	1.19 [0.54, 1.54]	0.89	97 (52.7%)	37 (20.1%)	24 (13.0%)	13 (7.1%)	13 (7.1%)	134 (72.7%)	high
	Mandible	anterior	B point	1.15 [0.52, 1.44]	0.96	106 (57.6%)	36 (19.6%)	21 (11.4%)	8 (4.3%)	13 (7.1%)	142 (77.2%)	high
			Pogonion	0.79 [0.35, 0.99]	0.68	140 (76.1%)	28 (15.2%)	7 (3.8%)	1 (0.5%)	8 (4.3%)	168 (91.3%)	very high

				Err	SD	[Q1, Q3]						AP	
		bottom	Menton	0.77	0.44	[0.49, 0.97]	143 (77.7%)	29 (15.8%)	8 (4.3%)	3 (1.6%)	1 (0.5%)	172 (93.5%)	very high
		posterior	Articulare	0.77	0.45	[0.43, 1.02]	138 (75.0%)	34 (18.5%)	10 (5.4%)	1 (0.5%)	1 (0.5%)	172 (93.5%)	very high
Dental landmark	Maxillary dentition	anterior	Mx1C	0.44	0.37	[0.27, 0.54]	178 (96.7%)	2 (1.1%)	2 (1.1%)	1 (0.5%)	1 (0.5%)	180 (97.8%)	very high
		anterior	Mx1R	1.55	1.09	[0.80, 1.95]	63 (34.2%)	43 (23.4%)	36 (19.6%)	12 (6.5%)	30 (16.3%)	106 (57.6%)	medium
		posterior	Mx6D	1.43	1.08	[0.69, 1.90]	76 (41.3%)	42 (22.8%)	23 (12.5%)	17 (9.2%)	26 (14.1%)	118 (64.1%)	medium
		posterior	Mx6R	1.68	1.08	[0.88, 2.10]	51 (27.7%)	44 (23.9%)	35 (19.0%)	19 (10.3%)	35 (19.0%)	95 (51.6%)	medium
	Mandibular dentition	anterior	Md1C	0.49	0.64	[0.27, 0.53]	172 (93.5%)	7 (3.8%)	0 (0.0%)	2 (1.1%)	3 (1.6%)	179 (97.3%)	very high
		anterior	Md1R	1.57	1.04	[0.87, 2.05]	64 (34.8%)	43 (23.4%)	29 (15.8%)	18 (9.8%)	30 (16.3%)	107 (58.2%)	medium
		posterior	Md6D	1.67	1.24	[0.87, 2.10]	54 (29.3%)	41 (22.3%)	33 (17.9%)	30 (16.3%)	26 (14.1%)	95 (51.6%)	medium
		posterior	Md6R	2.03	1.35	[0.98, 2.66]	46 (25.0%)	30 (16.3%)	25 (13.6%)	28 (15.2%)	55 (29.9%)	76 (41.3%)	low
Total				1.17	1.04		2100 (57.1%)	631 (17.1%)	368 (10.0%)	227 (6.2%)	354 (9.6%)	2731 (74.2%)	high

Err, absolute value of error; SD, standard deviation; Accuracy Percentage (AP); error range within 1.5 mm was considered accurate.

The degree of accuracy was defined as very high (90% ≤ AP), high (70% ≤ AP < 90%), medium (50% ≤ AP < 70%), and low (AP < 50%).

Q1, the first quartile; Q3, the third quartile.

32

Table 5. Comparison of mean errors among the four time-points (T0, T1, T2 and T3) and between two time-points [(T0, T1) vs. (T2, T3)]

Compartment			Landmark	Initial stage (T0) Mean (mm) [Q1, Q3]	SD (mm)	Pre-surgery stage (T1) Mean (mm) [Q1, Q3]	SD (mm)	Post-surgery stage (T2) Mean (mm) [Q1, Q3]	SD (mm)	Debonding stage (T3) Mean (mm) [Q1, Q3]	SD (mm)	among T0, T1, T2 and T3 stages [§] P-value	Tukey HSD test	(T0, T1) vs. (T2,T3) [§] P-value	Contrast matrix with the MANOVA
Skeletal landmark	Cranial Base		Nasion	0.58 [0.32, 0.76]	0.42	0.59 [0.27, 0.69]	0.48	0.55 [0.25, 0.81]	0.39	0.64 [0.27, 0.85]	0.60	0.698		0.852	
			Sella	0.48 [0.32, 0.62]	0.23	0.43 [0.27, 0.51]	0.19	0.41 [0.27, 0.54]	0.20	0.50 [0.27, 0.71]	0.27	0.155		0.986	
			Porion	1.04 [0.69, 1.45]	0.55	1.07 [0.48, 1.57]	0.76	1.17 [0.60, 1.49]	0.85	1.01 [0.60, 1.41]	0.57	0.493		0.566	
			Orbitale	1.19 [0.48, 1.68]	0.88	1.15 [0.46, 1.59]	0.88	1.39 [0.54, 1.59]	1.40	1.10 [0.48, 1.45]	0.76	0.454		0.618	
			Basion	1.41 [0.43, 2.06]	1.32	1.59 [0.70, 1.73]	1.60	1.95 [0.64, 2.52]	1.94	1.61 [0.67, 1.93]	1.52	0.148		0.092	
	Maxilla	anterior	ANS	1.07 [0.49, 1.56]	0.78	1.22 [0.41, 1.78]	0.97	1.78 [0.96, 2.28]	1.22	1.49 [0.99, 1.93]	0.87	0.003 **	T0[a], T1[a], T2[b], and T3[ab]	0.003 **	(T0, T1) < (T2, T3)
			A point	1.27 [0.67, 1.64]	0.89	1.28 [0.61, 1.94]	0.78	1.50 [0.73, 2.00]	1.07	1.59 [0.76, 2.44]	1.16	0.151		0.040*	(T0, T1) < (T2, T3)
		posterior	PNS	1.16 [0.53, 1.71]	0.79	1.14 [0.54, 1.41]	0.87	1.29 [0.61, 1.55]	1.09	1.17 [0.49, 1.57]	0.82	0.823		0.587	
	Mandible	anterior	B point	1.00 [0.43, 1.22]	0.97	1.01 [0.49, 1.42]	0.61	1.29 [0.54, 1.42]	1.24	1.31 [0.64, 1.81]	0.91	0.142		0.008**	(T0, T1) < (T2, T3)
			Pogonion	0.66 [0.32, 0.98]	0.48	0.80 [0.34, 1.03]	0.72	0.82 [0.43, 1.08]	0.69	0.86 [0.34, 1.05]	0.77	0.277		0.260	
		bottom	Menton	0.83	0.52	0.70	0.39	0.74	0.38	0.82	0.45	0.298		0.786	

Dental landmark				T0		T1		T2		T3		p-value§	post-hoc§	p-value‡	post-hoc‡
				[0.51, 1.05]		[0.43, 0.88]		[0.48, 0.87]		[0.47, 1.04]					
		posterior	Articulare	0.76	0.39	0.75	0.42	0.73	0.40	0.84	0.58	0.540		0.616	
				[0.48, 0.97]		[0.42, 0.95]		[0.42, 0.99]		[0.42, 1.10]					
	Maxillary dentition	anterior	Mx1C	0.48	0.37	0.49	0.55	0.41	0.26	0.38	0.18	0.355		0.096	
				[0.27, 0.54]		[0.24, 0.54]		[0.24, 0.54]		[0.26, 0.49]					
			Mx1R	1.83	1.24	1.37	1.14	1.56	1.02	1.54	1.17	0.166		0.714	
				[0.92, 2.38]		[0.64, 1.59]		[0.84, 2.11]		[0.72, 1.87]					
		posterior	Mx6D	1.66	1.18	1.63	1.31	1.20	0.80	1.23	0.88	0.032 *	T0b, T1ab, T2a, and T3ab	0.008 **	(T2, T3) < (T0, T1)
				[0.75, 2.24]		[0.70, 2.32]		[0.58, 1.59]		[0.64, 1.60]					
Dental landmark			Mx6R	1.89	1.15	1.65	1.08	1.57	1.03	1.60	1.05	0.349		0.194	
				[0.97, 2.71]		[0.99, 2.15]		[1.08, 1.88]		[0.72, 2.04]					
	Mandibular dentition	anterior	Md1C	0.48	0.45	0.62	1.10	0.47	0.39	0.37	0.31	0.215		0.096	
				[0.22, 0.56]		[0.27, 0.54]		[0.27, 0.54]		[0.24, 0.39]					
			Md1R	1.77	1.22	1.52	1.00	1.37	0.96	1.63	0.95	0.303		0.400	
				[0.88, 2.53]		[0.84, 1.95]		[0.72, 1.62]		[0.90, 2.18]					
		posterior	Md6D	2.15	1.79	1.71	0.91	1.51	1.14	1.33	0.75	0.008 **	T0b, T1ab, T2a, and T3a	0.003 **	(T2, T3) < (T0, T1)
				[1.21, 2.52]		[0.96, 2.39]		[0.67, 2.01]		[0.80, 1.88]					
			Md6R	2.21	1.64	2.08	1.36	1.94	1.25	1.89	1.12	0.579		0.242	
				[0.97, 2.70]		[0.94, 2.80]		[0.83, 2.67]		[1.16, 2.70]					
Total				1.20	1.10	1.14	1.02	1.18	1.09	1.15	0.95	0.376		0.895	

§ Repeated measures analysis of variance (ANOVA) test and post-hoc test for 'within-subject' by Tukey's adjustment for multiple comparisons were performed.

a, b, ab: 'a' and 'b' indicate statically significant difference between a and b; while, 'ab' indicates that there was no significant difference between a and ab and between ab and b.

‡ Repeated measures multivariate analysis of variance (MANOVA) test was performed.

T0, initial stage; T1, pre-surgery stage [presence of orthodontic brackets (OBs)]; T2, post-surgery stage [presence of OBs and surgical plates and screws (SPS)], and T3, debonding stage [presence of SPS and fixed retainers (FR)].

Q1, the first quartile; Q3, the third quartile.

Table 6. Comparison of mean errors in each time-point (from T0 to T3) between the genioplasty and non-genioplasty groups

Land-mark	Group 1 (genioplasty group)										Comparison of mean error among T0, T1, T2 and T3 stages [§]	Group 2 (non-genioplasty group)										Comparison of mean error among T0, T1, T2 and T3 stages[§]	Comparison of the genioplasty and non-genioplasty groups[‡]				
	Total		Initial stage (T0)		Pre-surgery stage (T1)		Post-surgery stage (T2)		Debonding stage (T3)			Total		Initial stage (T0)		Pre-surgery stage (T1)		Post-surgery stage (T2)		Debonding stage (T3)			Total	(T0)	(T1)	(T2)	(T3)
	Mean (mm) [Q1, Q3]	SD (mm)	Mean (mm) [Q1, Q3]	SD (mm)	Mean (mm) [Q1, Q3]	SD (mm)	Mean (mm) [Q1, Q3]	SD (mm)	Mean (mm) [Q1, Q3]	SD (mm)	P-value	Mean (mm) [Q1, Q3]	SD (mm)	Mean (mm) [Q1, Q3]	SD (mm)	Mean (mm) [Q1, Q3]	SD (mm)	Mean (mm) [Q1, Q3]	SD (mm)	Mean (mm) [Q1, Q3]	SD (mm)	p-value	p-value	p-value	p-value	p-value	p-value
B point	1.08 [0.52, 1.47]	0.67	0.87 [0.43, 1.20]	0.46	0.99 [0.48, 1.56]	0.60	1.21 [0.70, 1.67]	0.70	1.25 [0.65, 1.83]	0.82	0.184	1.22 [0.52, 1.37]	1.19	1.13 [0.27, 1.68]	1.30	1.03 [0.54, 1.29]	0.63	1.37 [0.51, 1.30]	1.63	1.37 [0.60, 1.80]	1.01	0.543	0.453	0.386	0.855	0.670	0.682
Pog	0.76 [0.37, 0.97]	0.66	0.61 [0.34, 0.87]	0.38	0.66 [0.27, 1.03]	0.39	0.81 [0.43, 0.73]	0.82	0.95 [0.49, 0.97]	0.87	0.109	0.82 [0.34, 1.10]	0.69	0.71 [0.27, 1.03]	0.57	0.95 [0.36,1.03]	0.93	0.84 [0.43, 1.44]	0.56	0.77 [0.34, 1.22]	0.66	0.463	0.677	0.468	0.171	0.898	0.436
Menton	0.73 [0.48, 0.96]	0.42	0.71 [0.49, 0.99]	0.35	0.71 [0.49, 1.02]	0.47	0.68 [0.43, 0.91]	0.38	0.82 [0.48, 0.73]	0.48	0.578	0.81 [0.49, 1.02]	0.45	0.95 [0.54, 1.14]	0.63	0.68 [0.36, 0.87]	0.28	0.79 [0.58, 0.35]	0.38	0.81 [0.43, 1.07]	0.43	0.186	0.344	0.117	0.813	0.334	0.926
Md1C	0.41 [0.27, 0.51]	0.33	0.58 [0.27, 0.61]	0.53	0.38 [0.27, 0.51]	0.18	0.39 [0.27, 0.48]	0.25	0.30 [0.17, 0.36]	0.14	0.062	0.56 [0.25, 0.54]	0.85	0.39 [0.17, 0.48]	0.33	0.87 [0.27, 0.61]	1.52	0.56 [0.27, 0.61]	0.48	0.44 [0.24, 0.49]	0.41	0.156	0.220	0.149	0.138	0.155	0.137
Md1R	1.50 [0.82, 2.01]	0.94	1.64 [0.65, 2.21]	1.13	1.22 [0.77, 1.63]	0.68	1.61 [0.97, 2.10]	1.03	1.53 [0.87, 2.23]	0.88	0.380	1.64 [0.88, 2.06]	1.13	1.90 [1.02, 2.53]	1.32	1.83 [1.08, 2.43]	1.18	1.12 [0.54, 1.45]	0.84	1.73 [0.91, 2.16]	1.03	0.091	0.358	0.484	0.04*	0.082	0.484

[§] Repeated measures analysis of variance (ANOVA)

[‡] Independent t-test was performed.

*, P<0.05

Q1, the first quartile; Q3, the third quartile.

Figure 1. Composition of the test set. T0, initial; T1, pre−surgery; T2, post−surgery; T3, debonding
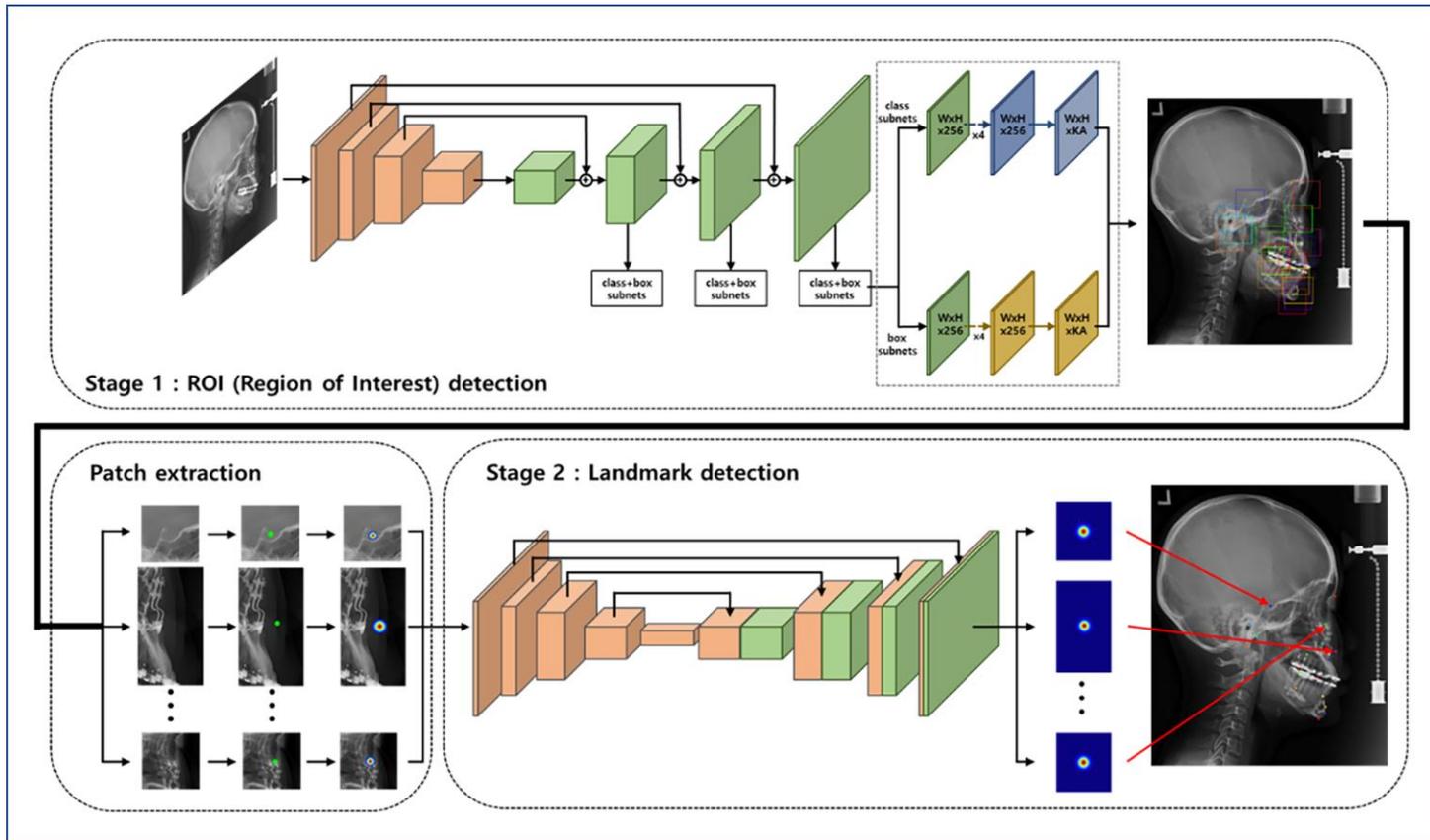
Figure 2. General schematic of the cascade convolutional neural network algorithm for artificial intelligence–assisted hard tissue landmark identification
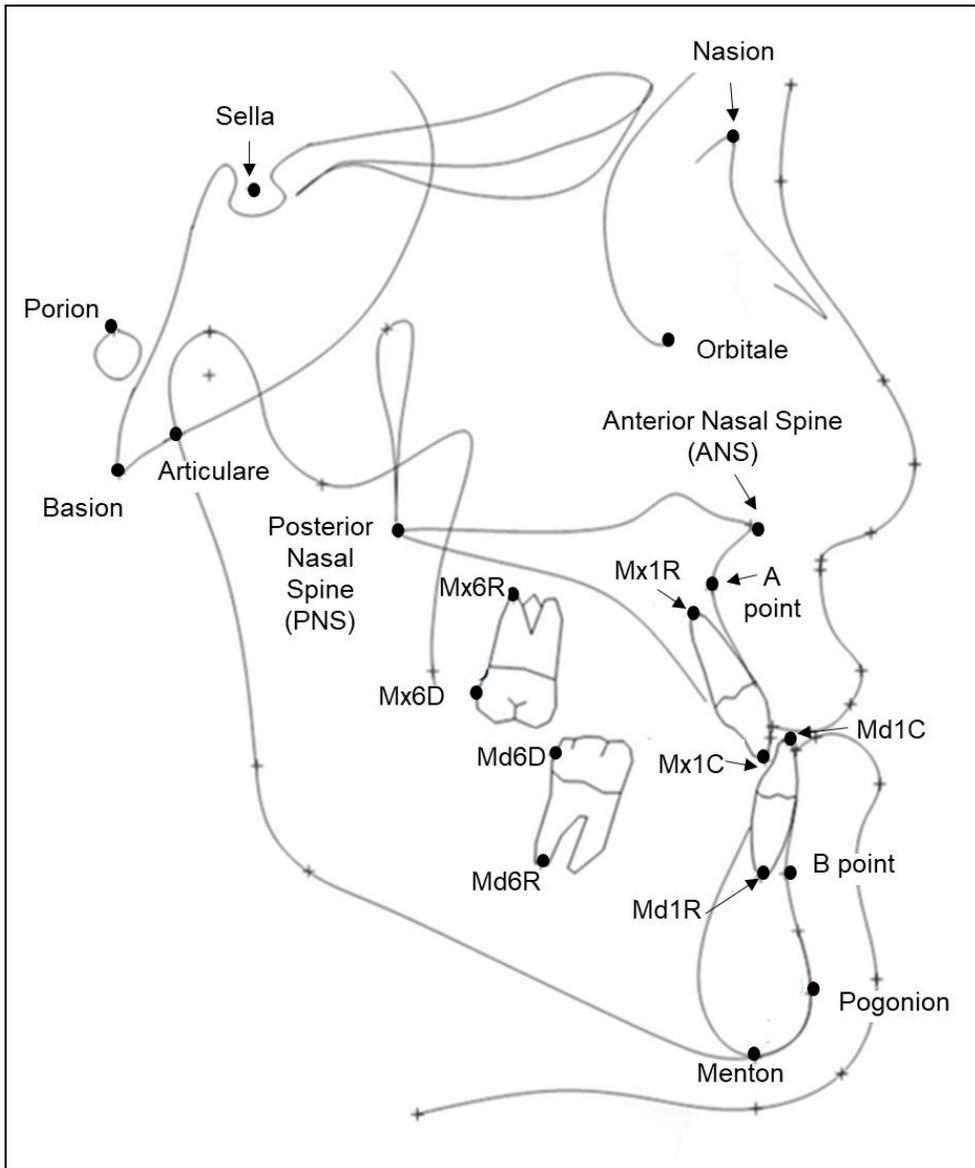
Figure 3. The skeletal and dental landmarks
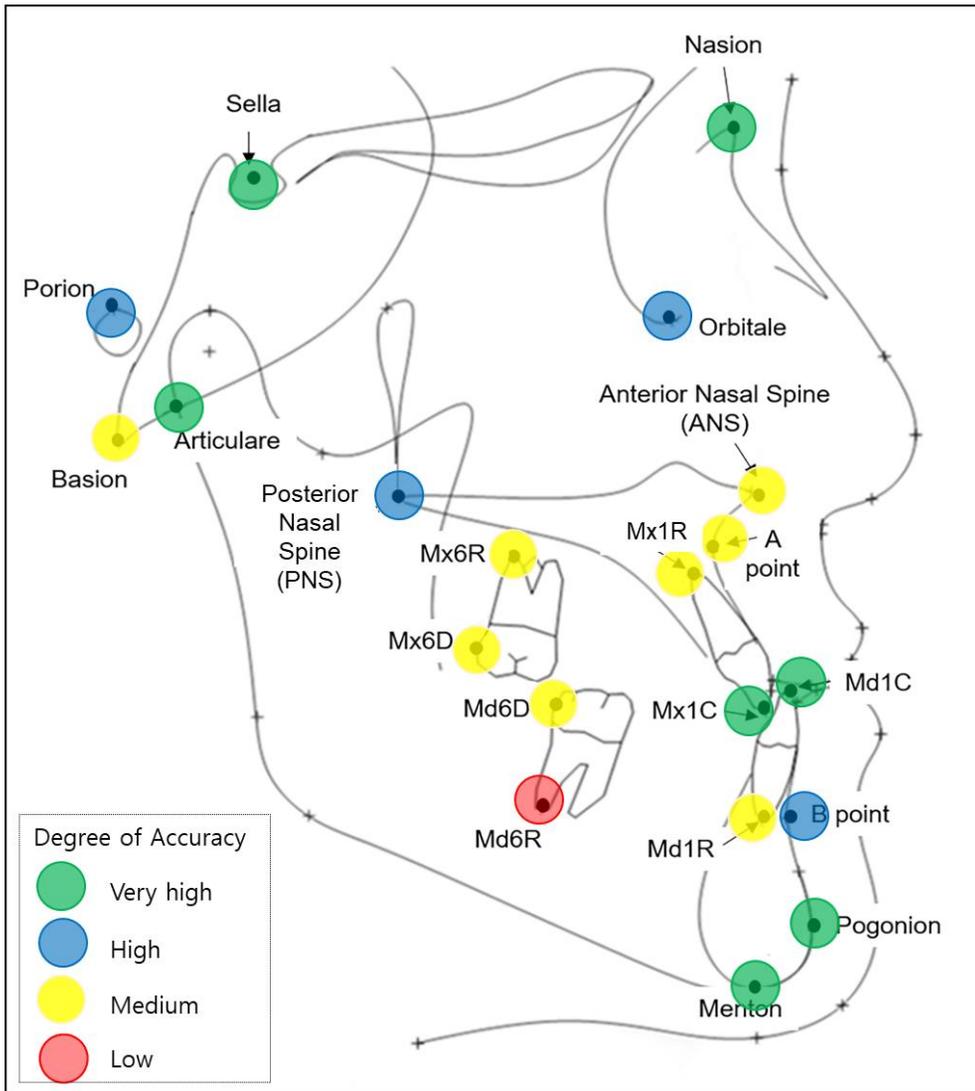
Figure 4. Degree of accuracy for 20 hard tissue landmarks. Green circle: degree of accuracy is very high, which means accuracy percentage is ≥ 90%. Blue circle: degree of accuracy is high, which means accuracy percentage is 70% ≤ AP < 90%. Yellow circle: degree of accuracy is medium, which means accuracy percentage is 50% ≤ AP < 70%. Red circle: degree of accuracy is low, which means accuracy percentage is < 50%.
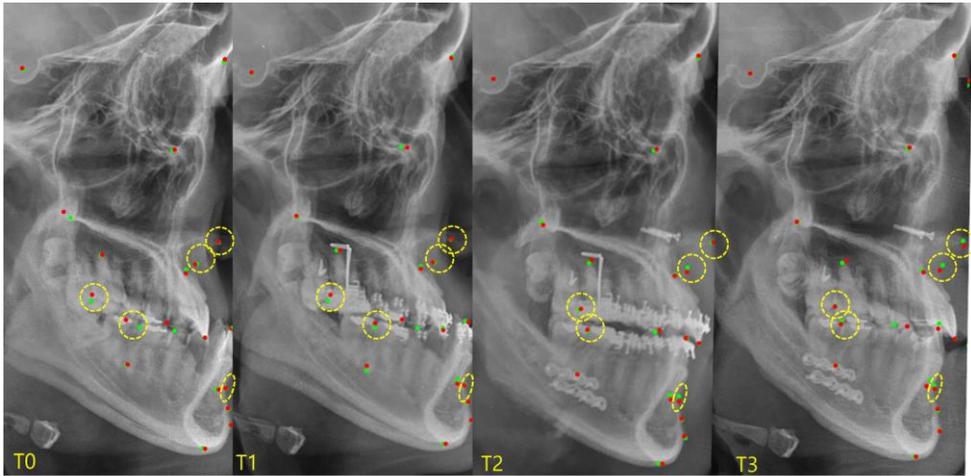
Figure 5. Examples of increase in the errors in ANS, A−point, B point and decrease in the errors in Mx6D and Md6D; Green dots by human and red dots by artificial intelligence. T0, initial; T1, pre−surgery; T2, post−surgery; T3, debonding
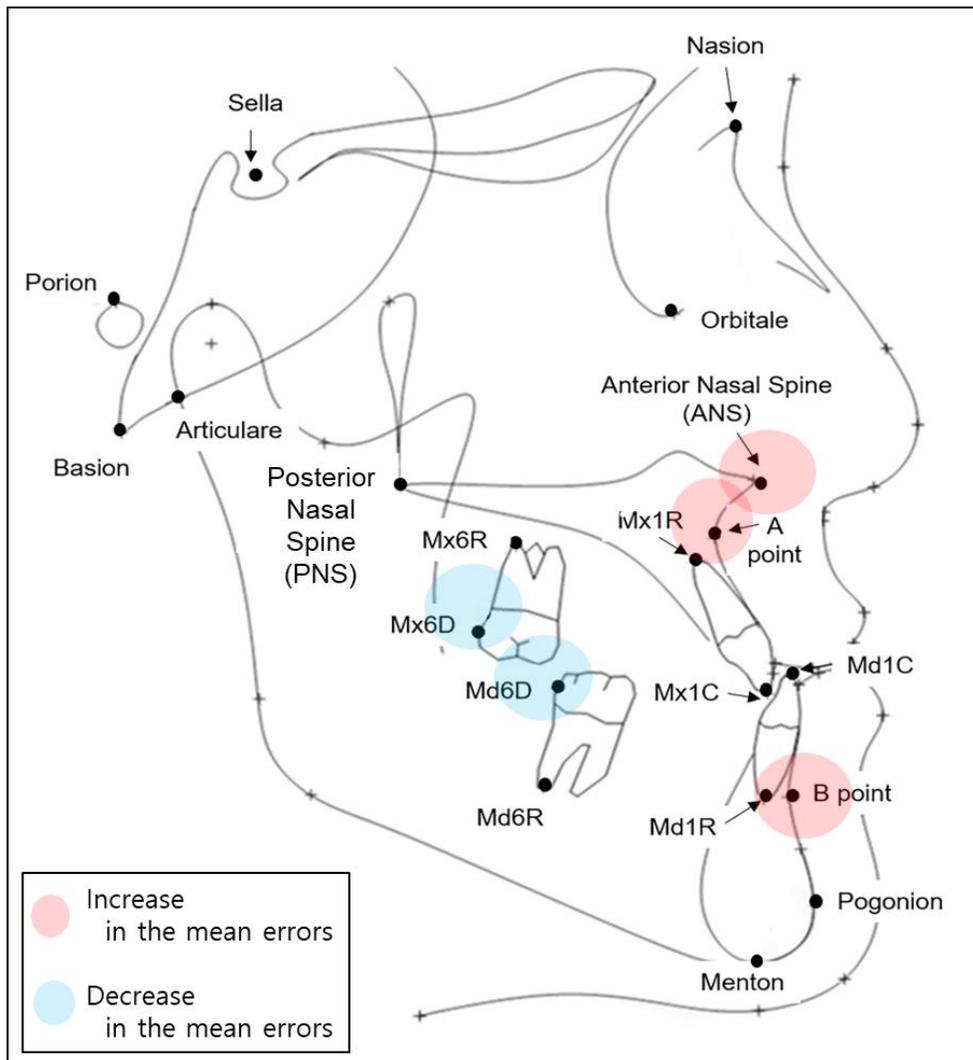
Figure 6. In comparison between pre−and post− surgical time points, the mean errors of ANS, A point, and B point were statistically increased (P<0.01; P<0.05; P<0.01). On the other hand, the mean errors of Mx6D and Md6D were decreased (P<0.01; P<0.01). However, these values are clinically insignificant.

국문 초록

양악 악교정수술과 수술 교정치료를 받은 골격성 III급 부정교합 환자의 측모 두부계측방사선사진 영상에서 인공지능을 이용한 경조직 계측점 식별 시 평균 오차의 변화 양상

홍 미 희

서울대학교 대학원 치의학과 치과교정학 전공

(지도교수: 백 승 학)

연구목적: 최근 직렬 합성곱 신경망 (cascade TJneural network) 인공지능 모델을 사용하여 측모두부계측방사선사진 영상에서 경조직 계측점을 자동 식별(auto-digitization)하는 연구들이 발표되고 있다. 본 연구의 목적은 양악 악교정수술과 술전 및 술후 교정치료를 받은 환자들의 연속적인 측모두부계측방사선사진 영상에서 경조직 계측점을 자동 식별하는 정확도가 촬영 시점에 따라 어떠한 양상으로 변화하는지 평가하는 것이었다.

연구재료 및 방법: 연구 대상은 양악 악교정 수술을 받은 골격성 III급 부정교합 환자 797명의 측모두부계측방사선사진 영상 3,188장이었다. 751명으로부터 확보한 3,004장의 영상을 training set과 validation set으로 활용하였고, 46명 환자 [이부성형술(genioplasty) 시행 군 (n=23), 이부성형술(genioplasty) 비시행군 (n=23)]로부터 확보한 184장의 영상을 test set으로 하였다. Test set은 초진(T0), 교정용 브라켓 영상이 포함된 술전(T1), 수술용 고정판과 나사 (surgical plate and screw) 영상이 나타난 술후(T2), 수술용 고정판과 나사 및

고정식 유지장치가 영상으로 보이는 종료(T3)의 4가지 시점별로 촬영된 측모두부계측방사선사진 영상으로 구성되었다. 인공지능과 치과교정과 전문의1인(human gold standard)이 각각 20개의 경조직 계측점을 digitization 한 후, 이에 따른 오차를 통계적으로 검증하였다.

연구 결과: 그 결과는 다음과 같았다. (1) 전체 평균 오차는 1.17 mm, T0의 평균 오차는 1.20 mm, T1의 평균 오차는 1.14 mm, T2의 평균 오차는 1.18 mm, T3의 평균 오차는 1.15 mm로 나타났다. (2) 술전 (T0, T1)과 술후(T2, T3)의 비교에서는, ANS, A point, B point 의 오차가 통계학적으로는 증가하였으나 임상적으로는 유의미하지 않았고 ($P<0.01$; $P<0.05$; $P<0.01$), Mx6D와 Md6D는 오차가 감소하였다 ($P<0.01$; $P<0.01$). (3) genioplasty 실행군과 genioplasty 비실행군 간의 비교에서 B point, Pogonion, Menton, Md1C, Md1R의 오차는 통계적인 차이가 나타나지 않았다.

결론: 직렬 인공지능 합성곱 신경망 모델은 교정 브라켓, 수술용 고정판과 나사, 고정식 유지장치, 이부성형술 그리고 수술 후 골 개조에도 불구하고, 초진, 술전 교정 치료, 술후 교정치료, 종료 시 촬영된 측모두부계측방사선사진 영상에서 경조직 계측점들을 자동식별 하는데 사용될 수 있다는 결론을 도출하였다.

————————————————————————————————————