# Effectiveness of Human-Artificial Intelligence Collaboration in Cephalometric Landmark Detection

Van Nhat Thang Le, Jae-Gon Kim, Yeon-Mi Yang, Dae-Woo Lee

Department of Pediatric Dentistry, School of Dentistry, Jeonbuk National University, Jeonju, Republic of Korea

## INTRODUCTION

- In orthodontics, one of the major goals is to resolve craniofacial discrepancies and properly align teeth for aesthetics and function. To achieve this, a detailed analysis should be performed on a lateral cephalogram. The accurate position of landmarks can give the clinician an exact analysis. Thus, localization of landmark position is important.
- Recently, automated cephalometric detection was developed along with convolutional neural network (CNN) models to shorten consuming time, improve accurate landmark detection, and reduce random errors. Among CNN models, “Deep Anatomical Context Feature Learning” (DACFL) outperformed state-of-the-art methods on the ISBI 2015 with high performance. A research question is raised: does DACFL achieve high performance on private dataset?
- In another aspect, most research has been focused on head-to-head comparisons between artificial intelligence (AI) and human in landmarks detection. AI's support for humans in diagnosis may become more useful and practical. This leads to a question: will human-AI collaboration do better than a single? Therefore, the aim of our study was to evaluate the performance of DACFL model on private dataset and the impact of AI-based support on the beginners in cephalometric landmark detection.

## MATERIALS AND METHODS

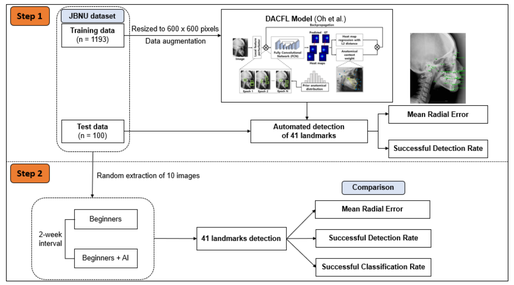


**Figure 1.** Workflow diagram of the study plan.

- 1293 lateral cephalograms of 6- to 18-year-old children and adolescents who visited the Department of Pediatric Dentistry from 2008 to 2018.
- 41 landmarks were manually identified by dental residents.

**Table 1.** List of anatomical landmarks

| No | Landmarks |
|---|---|
| 1 | Sella |
| 2 | Porion |
| 3 | Basion |
| 4 | Hinge Axis |
| 5 | Pterygoid |
| 6 | Nasion |
| 7 | Orbitale |
| 8 | A-Point |
| 9 | PM |
| 10 | Pogonion |
| 11 | B-point |
| 12 | Posterior Nasal Spine |
| 13 | Anterior Nasal Spine |
| 14 | R1 |
| 15 | R3 |
| 16 | Articulare |
| 17 | Menton |
| 18 | Maxilla 1 crown |
| 19 | Maxilla 1 root |
| 20 | Mandible 1 crown |
| 21 | Mandible 1 root |
| 22 | Maxilla 6 distal |
| 23 | Maxilla 6 root |
| 24 | Mandible 6 distal |
| 25 | Mandible 6 root |
| 26 | Glabella |
| 27 | Soft tissue Nasion |
| 28 | Pronasale |
| 29 | Columella |
| 30 | Subnasale |
| 31 | Soft tissue A |
| 32 | Upper Lip |
| 33 | Stms |
| 34 | Stmi |
| 35 | Lower Lip |
| 36 | Soft tissue B |
| 37 | Soft tissue Pogonion |
| 38 | Gnathion |
| 39 | Gonion |
| 40 | APOcc |
| 41 | PPOcc |

- Beginners were 20 final year students. 10 cephalograms that had been analyzed twice (at a 2-week interval) were used to evaluate the support ability of AI If the students change their answers by replacing AI’s answers, the changes were recorded.
- Evaluation metrices include mean radial error (MRE), successful detection rate (SDR) within 2-, 2.5-, 3-, and 4-mm threshold, and successful classification rate (SCR).

**Table 2.** List of measurements

| No | Measurements |
|---|---|
| 1 | ANB |
| 2 | SNA |
| 3 | SNB |
| 4 | ODI |
| 5 | APDI |
| 6 | FHI |
| 7 | FHA |
| 8 | MW |

## OVERVIEW OF RESULTS

**Table 3.** Results of landmark detection in terms of mean radial error

| No. | Landmarks | AI | |
|---|---|---|---|
| | | MRE (mm) | SD |
| 1 | Sella | 0.76 | 0.44 |
| 2 | Porion | 1.40 | 1.20 |
| 3 | Basion | 2.09 | 1.98 |
| 4 | Hinge axis | 1.70 | 1.06 |
| 5 | Pterygoid | 2.38 | 1.46 |
| 6 | Nasion | 1.33 | 0.87 |
| 7 | Orbitale | 2.23 | 1.81 |
| 8 | A-point | 1.43 | 1.11 |
| 9 | PM | 1.32 | 0.88 |
| 10 | Pogonion | 1.19 | 0.96 |
| 11 | B-point | 1.69 | 1.22 |
| 12 | Posterior nasal spine | 1.63 | 1.37 |
| 13 | Anterior nasal spine | 1.27 | 0.91 |
| 14 | R1 | 2.32 | 1.62 |
| 15 | R3 | 1.84 | 1.21 |
| 16 | Articulare | 1.03 | 0.74 |
| 17 | Menton | 1.22 | 0.91 |
| 18 | Maxilla 1 crown | 1.03 | 0.92 |
| 19 | Maxilla 1 root | 3.31 | 2.78 |
| 20 | Mandible 1 crown | 0.90 | 0.52 |
| 21 | Mandible 1 root | 2.89 | 4.04 |
| 22 | Maxilla 6 distal | 1.41 | 1.86 |
| 23 | Maxilla 6 root | 2.20 | 1.53 |
| 24 | Mandible 6 distal | 1.64 | 2.32 |
| 25 | Mandible 6 root | 2.97 | 2.50 |
| 26 | Glabella | 5.18 | 5.13 |
| 27 | Soft tissue nasion | 3.10 | 2.44 |
| 28 | Pronasale | 2.06 | 8.16 |
| 29 | Columella | 1.05 | 0.87 |
| 30 | Subnasale | 1.06 | 0.99 |
| 31 | Soft tissue A | 1.21 | 1.34 |
| 32 | Upper lip | 1.48 | 4.10 |
| 33 | Stms | 1.82 | 1.46 |
| 34 | Stmi | 1.03 | 0.79 |
| 35 | Lower lip | 1.16 | 0.90 |
| 36 | Soft tissue B | 2.08 | 2.70 |
| 37 | Soft tissue pogonion | 4.70 | 10.39 |
| 38 | Gnathion | 1.34 | 2.08 |
| 39 | Gonion | 2.70 | 2.14 |
| 40 | APOcc | 1.02 | 1.19 |
| 41 | PPOcc | 2.33 | 2.86 |
| Average | | 1.87 | 2.04 |

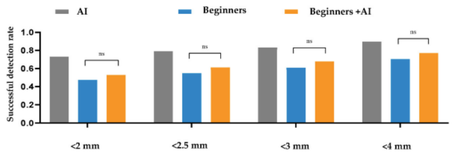


**Figure 2.** Comparison between the beginner-only and beginner-AI groups in terms of the successful detection rate. A t-test was applied to compare the average successful detection rates between the beginner-only and beginner-AI groups within 2, 2.5, 3, and 4-mm thresholds. The beginner-AI collaboration improved the successful detection rates within 2, 2.5, 3, and 4-mm thresholds.
**Abbreviations:** AI, artificial intelligence; ns, not significant

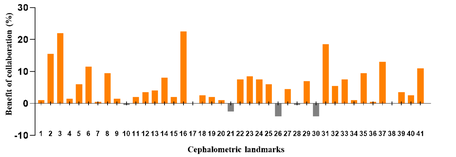


**Figure 3.** Benefit of beginner-AI collaboration in the detection of cephalometric landmarks. Based on successful detection rate for each landmark within a 2-mm threshold, the benefits of be-ginner-AI collaboration were analyzed. In general, this collaboration has shown positive impact on majority of cephalometric landmarks.

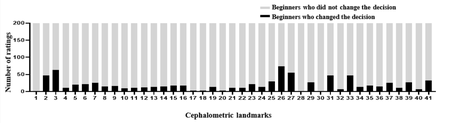


**Figure 4.** Number of decision changes among beginners across 41 landmarks. In the second experiment, the beginners traced the anatomical landmarks on 10 images with the AI’s answer view. The recorded changes are represented as number of ratings. In general, the number of decision changes was small despite being shown at most anatomical landmarks.

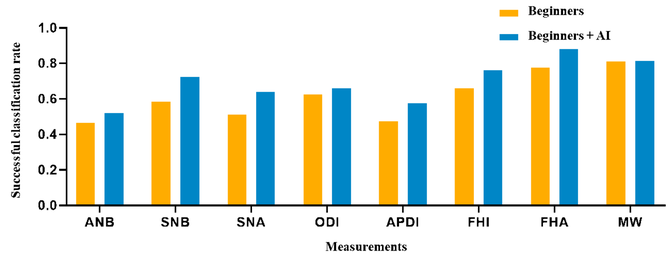


**Figure 5.** Comparison of eight clinical measurements between the beginner-only and beginner-AI groups. From the SCRs of two groups, a figure was presented to demonstrate the AI’s sup-port. As a result, the beginner-AI collaboration improved the SCRs of eight clinical measurements.
**Abbreviation**: SCR, successful classification rate

## RESULTS

- In the table 3, DACFL showed the average of MRE was 1.87 mm. Among 41 landmarks, sella had the smallest MRE (0.76 mm) while glabella had the largest MRE (5.18 mm).
- As presented in Figure 2, DACFL achieved the averages of SDR of 73.32%, 80.39%, 85.61%, and 91.68% within 2-, 2.5-, 3-, and 4-mm, respectively. Across all ranges, sella had the largest SDR while glabella had the smallest SDR. SDRs of maxilla 1 root (38%), mandible 6 root (36%), glabella (32%), and soft tissue nasion (38%) were low within 2-mm threshold. Across all ranges, SDRs were improved with AI support. However, the improvements were insignificant. Within 2-mm threshold, beginners-AI collaboration improves 5.33% of SDR.
- AI showed positive impact on almost landmarks (Figure 3) and there were small changes in the positions of the landmarks (Figure 4).
- With AI support, SCRs of all measurements were increased up. The average of SCR increased 8.38% (Figure 5).

## CONCLUSION

- Our study showed that the DACFL model achieved an SDR of 73.17% within a 2-mm threshold on the private dataset.
- The beginner-AI collaboration improved the SDR by 5.33% within a 2-mm threshold and improved the SCR by 8.38% when com-pared with beginners.
- These results suggest that the DACFL model is applicable to clinical orthodontic diagnosis.