

Accuracy of Automatic Cephalometric Analysis Programs on Lateral Cephalograms of Preadolescent Children

Min Sun Song¹, Seong-Oh Kim^{1,2}, Ik-Hwan Kim¹, Chung-min Kang^{1,2}, Je Seon Song^{1,2}

¹Department of Pediatric Dentistry, College of Dentistry, Yonsei University

²Oral Science Research Center, College of Dentistry, Yonsei University

Abstract

The aim of this study was to evaluate the accuracy of 3 different automatic landmark identification programs on lateral cephalograms and the clinical acceptability in pediatric dentistry.

Sixty digital cephalometric radiographs of 7 to 12 years old healthy children were randomly selected. Fourteen landmarks were chosen for assessment and the mean of 3 measurements of each landmark by a single examiner was defined as the baseline landmarks. The mean difference between an automatically identified landmark and the baseline landmark was measured for each landmark on each image.

The total mean difference of 3 automatic programs compared to the baseline landmarks were 2.53 ± 1.63 mm. Errors among 3 programs were not significantly different for 12 of 14 landmarks except Orbitale and Gonion. The automatic landmark identification programs showed significant higher mean detection errors than the manual method. The programs couldn't be used as the 1st tool to replace human examiners. But considering short consuming time, these results indicate that all 3 programs have sufficient validity to be used in pediatric dental clinic.

Key words : Automatic identification, Landmark, Accuracy, Cephalometric analysis, Preadolescent children

I. 서 론

측모 두부 계측법은 측모방사선 사진에서 관찰되는 해부학적 구조물을 계측하는 형태학적 연구로, 계측 결과를 이용하여 개인의 안모 및 악골 관계를 파악하고, 골격적 문제점을 진단할 수 있으며, 앞으로의 성장 방향을 예측할 수 있다[1]. 1931년 Broadbent[2]는 측모두부방사선 계측법을 처음으로 도입하였는데, 측모두부방사선 사진에서 관찰되는 특정 경조직 및 연조직 구조물의 기준점을 계측점으로 정의한 후, 계측점 사이의 거리와 각도를 측정함으로써 부정교합을 수치화 및 분석하였다. 정확한 해

부학적 계측점 탐지와 분석은 성공적인 교정 치료를 위하여 필수적이라고 하였다.

측모 두부 계측법은 주로 수동 측정법과 컴퓨터 보조 방법의 두가지 방법이 이용된다. 수동 측정법은 분석 시 검사자가 직접 계측점 사이의 길이 및 각도를 재기 때문에 많은 시간이 소요되며 측정 오차가 큰 단점이 있다. 컴퓨터 보조 방법은 자동으로 계측되기 때문에 소요 시간은 단축되었으나, 여전히 검사자의 숙련도에 따른 계측점 식별 오차가 존재한다[3]. 따라서, 최근에는 인공지능 프로그램을 학습시켜 자동으로 해부학적 계측점을 식별하는 방법이 연구되고 있다. Cohen과 Linney 등[4]이 1984

Corresponding author : Je Seon Song

Department of Pediatric Dentistry, College of Dentistry, Yonsei University, 50-1 Yonsei-ro, Seodaemun-gu, Seoul, 03722, Korea

Tel: +82-2-2228-3176 / Fax: +82-2-392-7420 / E-mail: songjs@yuhs.ac

Received November 14, 2020 / Revised February 9, 2021 / Accepted January 11, 2021

※This study was supported by a grant from Laonpeople INC., (2-2019-0036).

년 최초로 자동 계측점 식별 방법을 개발한 이후로 자동 계측점 인식 정확성을 높이기 위한 다양한 연구가 꾸준히 보고되고 있으며, 대부분의 계측점에서 검사자의 계측 결과와 높은 상관관계를 나타냈으나, 여전히 모든 계측점을 2.0 mm 오차 범위 내로 식별하는 데에는 한계가 존재하였다[5].

또한, 이전의 연구들은 대부분 영구치열기의 성인 환자들을 대상으로 진행되었으며, 소아 환자의 측모두부방사선 사진을 평가한 사례는 많지 않았다. 그 이유는, 소아의 골이 미성숙하여 연조직과의 낮은 흑화도 차이로 중첩 시 구별이 어려우며, 성장 발달로 인한 해부학적 구조물의 크기 및 형태의 변화와 유치 및 영구치 치배의 혼재로 인하여 성인과 비교하여 소아의 해부학적 구조를 인식하는 것이 더 어렵기 때문이다[6].

이에, 이번 연구에서는 이미 출시된 3종의 측모두부방사선 계측 프로그램을 이용하여 연세대학교 치과대학병원에 내원한 소아 환자의 측모 두부 계측점을 식별한 후, 각 계측점 별 정확도를 비교 분석하고 프로그램들의 임상 적용 가능성을 알아보는 것이 목적이다.

II. 연구 대상 및 방법

이 연구는 연세대학교 치과대학병원 임상 연구 윤리 위원회 (Institutional Review Board, IRB)의 승인을 받아 시행되었다(IRB No: 2-2019-0036).

1. 연구 대상

2019년 1월부터 2020년 8월까지 연세대학교 치과대학병원에 내원하여 측모두부방사선 사진을 촬영한 만 7세에서 12세 사이의 환자 중 상하악 영구 중절치가 맹출한 환자를 대상으로 하였다. 구강 내에 교정장치 또는 공간유지장치 등으로 인하여 영상의 질이 저하되어 교정 분석을 진행하기 어려운 경우, 두개 안면 영역의 비대칭 및 결함이 있는 환자들과 영구 중절치의 선천성 기형 또는 결손을 가진 환자들은 연구 대상자에서 제외하였다. 위 기준을 만족시키는 환자들 중 나이별로 남, 여 각 5명씩 무작위 추출하여 총 60명의 환자를 대상으로 연구를 진행하였으며, 연구 대상자의 성별 평균 연령은 남, 여 각각 9.5세였다.

2. 연구 방법

1) 측모 두부 계측점 측정

모든 측모두부방사선 사진은 1명의 숙련된 조사자와 자동 측모두부방사선 분석 프로그램들에 의해서 계측되었다. 1명의 숙

련된 조사자를 대조군으로 설계하였고, 조사자는 계측 프로그램 (V-ceph, version 5.5, Osstem, Seoul, Korea)을 사용하여 수동으로 해부학적 계측점을 식별하였다. 자동 분석 프로그램은 각각 I군은 V-ceph version 8.2 (Osstem, Seoul, Korea), II군은 Ceppro version 2.01 (DDHaim, Seoul, Korea), III군은 LaonCeph version 2.1 (Laonpeople, Seoul, Korea)로 나뉘었으며, 3개의 군 모두 조사자가 계측한 동일한 60명의 측모방사선 사진을 대상으로 계측점을 자동 식별하였다(Table 1, Fig. 1). 조사자의 관찰자 내 신뢰도를 평가하기 위해 모든 측모두부방사선 사진에 대한 관찰을 3주 간격으로 3번에 걸쳐 시행하였다. 급내 상관계수(Intra-class correlation coefficient)를 이용하여 평가한 결과, 관찰자 내 신뢰도는 0.90로 높은 신뢰도를 보였다.

측모두부계측을 위하여 자주 사용되는 14개의 계측 기준점을 식별하였다. 각 계측점은 Ear-rod를 원점으로 하는 2차원 좌표로 표현하였으며, y축은 Ear-rod를 수직 이등분하는 선으로, x축은 y축과 직각을 이루면서 Ear-rod를 이등분하는 선으로 정의하였다. 14개 계측점의 식별 정확도를 비교하기 위하여 검사자가 3회 반복 측정한 결과의 평균을 기준 좌표로 정의하였으며, 기준이 되는 좌표와 자동 계측점 식별 프로그램이 탐지한 좌표 사이의 거리 차(mm)를 자동 프로그램의 계측점 식별 오차로 평가하였다. 오차는 이미지 처리 프로그램(Image J, version 1.41, National Institutes of Health, Bethesda, MD)을 이용하여 계측하여 평가하였다(Fig. 2, Table 2). 검사자의 평균 식별 오차는 3회 반복 측정된 좌표 값과 기준 좌표 사이의 거리 차의 평균으로 설정하였다.

2) 통계 분석

각 14개 계측점에 대해 3개 군 및 검사자가 표기한 좌표와 기준 좌표 간의 직선 거리의 평균 및 표준편차를 구하여 검사자와 3개 군의 계측점 별 식별 정확성을 평가하였다. 3개 군 사이의 정확도 차이를 비교하기 위하여 일원배치 분산분석(One way analysis of variance, ANOVA)으로 통계적 유의차 검정 및 Tukey 사후 검정을 시행하였다. 검사자와 I군, II군 및 III군 사이의 식별 능력 비교는 독립변수 t검정(Student's t-test)으로 통계적 유의성

Table 1. Automatic landmark identification programs used in this study and sample grouping

Group	Programs	Manufacturer
I	V-ceph, version 8.2	Osstem, Seoul, Korea
II	Ceppro, version 2.1	DDHaim, Seoul, Korea
III	Laonceph, version 2.1	Laonpeople, Seoul, Korea

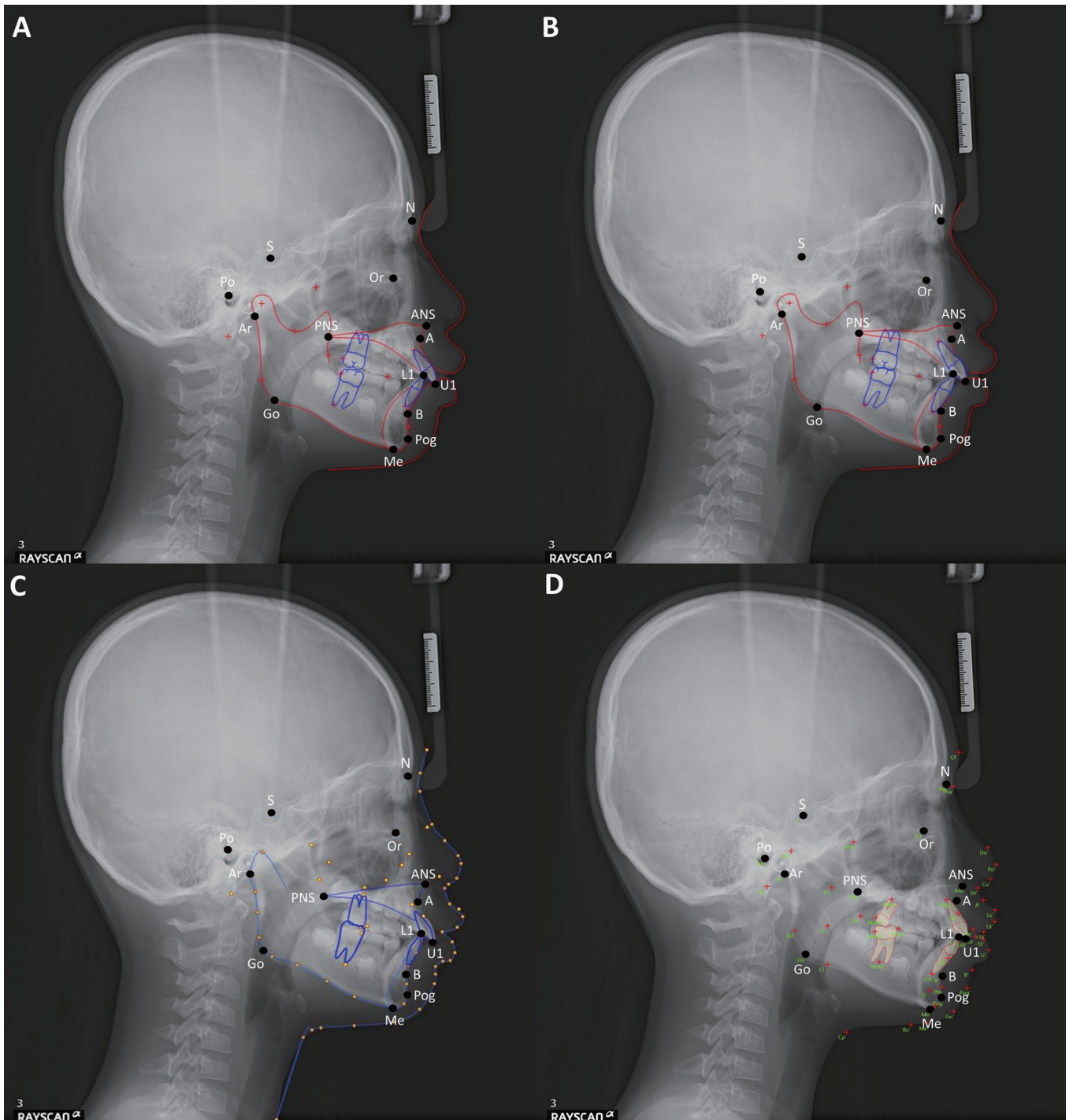


Fig. 1. Lateral cephalograms of 4 different landmark identification methods. (A) The 14 landmarks used in this study on the lateral cephalogram, detected by the manual method. The same 14 landmarks identified by the 3 automated landmark identification programs of (B) V-ceph, ver.8.2, (C) Ceppo, ver.2.1, and (D) Laonceph, ver.2.1.

Table 2. The 14 cephalometric landmarks employed in the present study

Landmarks	Definition
Sella (S)	Midpoint of the hypothysial fossa
Nasion (N)	Most concave point of naso-frontal suture
Porion (Po)	Uppermost point of the external ear meatus
Articulare (Ar)	Intersection of the posterior border of the ramus with the base of the occipital bone
Orbitale (Or)	Lowermost point of the orbit
A point (A)	Deepest point of anterior maxilla
B point (B)	Deepest point of anterior mandible
Anterior nasal spine (ANS)	The Tip of anterior nasal spine
Posterior nasal spine (PNS)	Posterior spine of the palatine bone
Pogonion (Pog)	Most anterior point of the chin
Menton (Me)	Most inferior point of the chin
Gonion (Go)	Intersection of the lines tangent to the posterior and inferior border of mandible
Upper 1st incisor tip (U1)	Tip of the upper first incisor
Lower 1st incisor tip (L1)	Tip of the lower first incisor

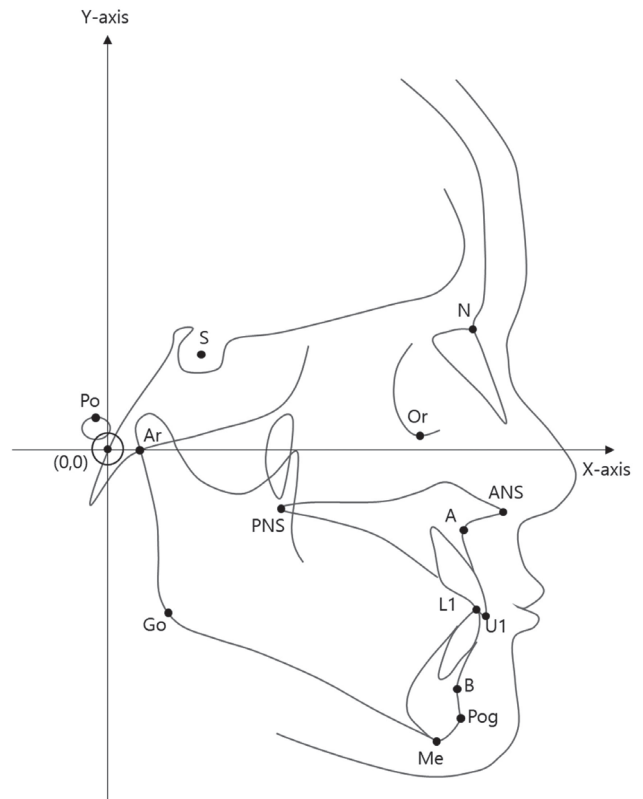


Fig. 2. The vertical and horizontal reference lines and coordinate system defined for this study.

을 살펴보았다. 모든 통계 분석은 통계 프로그램(SPSS for Windows, version 25.0; IBM Inc., Chicago, USA)을 이용하였다.

III. 연구 결과

1. 검사자의 계측점 식별 정확도

검사자의 각 계측점 별 식별 오차의 평균과 표준편차를 표로 나타내었다(Table 3). Sella (S), Porion (Po), Articulare (Ar), Posterior Nasal Spine (PNS), Point B (B)과 Menton (Me)이 평균 1.0 mm 이내의 오차 범위를 나타냈다. 가장 식별 오차가 낮은 계측점은 S로 평균 0.74 ± 0.43 mm이었고, Gonion (Go)이 1.46 ± 1.37 mm로 가장 높은 오차를 나타냈다. 검사자의 전체 평균 식별 오차는 1.30 ± 0.72 mm이었다.

2. 검사자와 자동 측모두부방사선 계측 프로그램의 비교

검사자와 3개의 프로그램 군 사이의 비교에서는 모든 계측점

Table 3. Mean detection errors for manual landmark identification

Landmark	Mean Difference \pm SD (mm)
S	0.74 ± 0.43
N	1.09 ± 0.63
Po	0.91 ± 0.56
Or	1.15 ± 0.65
Ar	0.84 ± 0.49
ANS	1.06 ± 0.64
PNS	0.90 ± 0.58
Point A	1.11 ± 0.69
Point B	0.94 ± 0.59
Pog	1.11 ± 0.69
Go	1.46 ± 1.37
Me	0.98 ± 0.68
U1	1.12 ± 0.75
L1	1.01 ± 0.69
Mean	1.30 ± 0.72

S = Sella, N = Nasion, Po = Porion, Or = Orbitale, Ar = Articulare, ANS = Anterior Nasal Spine, PNS = Posterior Nasal Spine, Pog = Pogonion, Go = Gonion, Me = Menton, U1 = Upper 1st incisor tip, L1 = Lower 1st incisor tip

들에서 검사자의 평균 식별 오차가 3개의 실험군보다 유의미하게 낮았다(Table 4).

3. 자동 측모두부방사선 분석 프로그램의 계측점 식별 정확도와 세 프로그램 군 사이의 비교

검사자가 직접 식별하였을 때와 마찬가지로 S가 모든 실험군들에서 가장 낮은 측정 오차를 나타냈는데, I군에서는 평균 1.32 ± 0.81 mm, II군은 1.68 ± 0.92 mm, III군은 1.45 ± 0.87 mm로 측정되었다. 14개 계측점 중 S만 유일하게 평균 2.0 mm 이내의 오차 범위 내에 탐지되었다. Po, Orbitale (Or), Go과 A가 평균 3.0 mm 이상의 오차를 나타냈으며, 특히 Go은 II군에서 평균 6.42 ± 2.94 mm의 오차가 측정되었다.

3종의 자동 측모 두부 분석 프로그램 군들 사이의 계측점 식별 오차를 비교하였을 때, 12개의 계측점인 S, Nasion (Na), Po, Ar, Anterior Nasal Spine (ANS), PNS, Point A (A), B와 Pog에서는 군 간 통계적으로 유의미한 차이는 관찰되지 않았다. 반면, Or의 경우 식별 오차에 관하여, I군이 II군보다 평균 1.2 mm, III

군보다는 평균 0.8 mm 낮은 통계적으로 유의미한 차이를 보였다(Table 5, $p = 0.000, 0.003$). II군과 III군 사이에는 유의한 차이를 보이지 않았다(Table 5, $p = 0.545$). Go의 경우에는 II군이 I군, III군과 비교하여 각각 평균 3.29 mm, 2.49 mm로 높은 오차를 가졌으며 이는 통계학적으로 유의하였다(Table 5, $p = 0.000, 0.000$).

4. 계측점 식별 오차 한계 범위에 따른 계측점의 분포

서로 다른 식별 오차 범위 한계를 설정한 후, 3개 군이 식별한 각 오차 한계 범위 내에 포함되는 계측점 분포를 확인한 결과는 다음과 같다. 1.0 mm 이내의 오차 범위 안에 포함되는 계측점은 3개 군 모두에서 관찰되지 않았다. 계측점 식별 오차 범위 한계를 2.0 mm 이내로 설정하였을 때는, I군과 III군에서는 14개의 계측점 중 1개, II군에서는 2개의 계측점이 포함되었다. I군과 III군으로 계측하였을 때, 모든 계측점이 4.0 mm 이내의 오차 범위 내에서 식별되었으며, II군은 14개 중 13개의 계측점이 4.0 mm 범위 이내의 오차를 나타내어 그 뒤를 이었다(Table 6).

Table 4. Comparing difference of mean detection errors between automated identification programs and human examiner

Landmark	Mean Difference (mm) between Examiner and Groups					
	Group I - Examiner		Group II - Examiner		Group III - Examiner	
	Net Difference	p	Net Difference	p	Net Difference	p
Sella	-0.59		-0.75		-0.72	
Nasion	-1.13		-1.06		-0.98	
Porion	-1.60		-1.80		-2.21	
Orbitale	-1.17		-2.26		-1.96	
Articulare	-1.17		-1.03		-1.19	
ANS	-1.09		-1.51		-1.38	
PNS	-1.35	0.000	-1.59	0.000	-1.79	0.000
Point A	-1.85		-1.50		-1.60	
Point B	-1.25		-1.52		-1.75	
Pogonion	-0.90		-1.43		-1.28	
Gonion	-1.92		-5.21		-2.72	
Menton	-0.99		-1.32		-1.3	
U1	-1.25		-1.39		-1.28	
L1	-1.05		-1.29		-1.39	

p value from Student's t-test

ANS = Anterior Nasal spine, PNS = Posterior Nasal Spine, U1 = Upper 1st incisor tip, L1 = Lower 1st incisor tip

Table 5. Comparisons of landmark detection errors among three automated identification programs

Landmark	Landmark Detection Errors (Mean ± SD)			Statistic Comparisons between Groups		
	Group I	Group II	Group III	I-II	I-III	II-III
S	1.32 ± 0.81	1.68 ± 0.92	1.45 ± 0.87	0.059	0.798	0.374
N	2.32 ± 1.29	2.25 ± 1.13	2.17 ± 1.43	0.986	0.890	0.982
Po	2.53 ± 1.21	2.72 ± 1.25	3.13 ± 2.00	0.854	0.068	0.346
Or	2.24 ± 1.31	3.34 ± 1.47	3.04 ± 1.43	0.000	0.003	0.545
Ar	2.03 ± 1.33	1.88 ± 1.16	2.05 ± 1.55	0.883	1.000	0.841
ANS	2.15 ± 1.28	2.56 ± 1.39	2.44 ± 1.43	0.252	0.566	0.945
PNS	2.26 ± 1.29	2.49 ± 1.26	2.69 ± 1.36	0.669	0.167	0.788
A	3.02 ± 1.60	2.68 ± 1.46	2.77 ± 1.39	0.495	0.740	0.980
B	2.23 ± 1.30	2.50 ± 1.67	2.73 ± 1.48	0.669	0.162	0.779
Pog	2.09 ± 1.28	2.63 ± 1.24	2.47 ± 1.39	0.061	0.288	0.882
Go	3.13 ± 2.00	6.42 ± 2.94	3.93 ± 2.30	0.000	0.186	0.000
Me	2.03 ± 1.27	2.35 ± 1.29	2.34 ± 1.30	0.415	0.455	1.000
U1	2.40 ± 1.68	2.53 ± 1.55	2.42 ± 1.39	0.950	1.000	0.969
L1	2.09 ± 1.17	2.34 ± 1.32	2.44 ± 1.34	0.653	0.699	0.967

p value from One-way ANOVA, Tukey's Post Hoc test. The bonferroni-corrected *p* values were reported to correct significance tests to adjust for multiple testing.

S = Sella, N = Nasion, Po = Porion, Or = Orbitale, Ar = Articulare, ANS = Anterior Nasal Spine, PNS = Posterior Nasal Spine, A = Point A, B = Point B, Pog = Pogonion, Go = Gonion, Me = Menton, U1 = Upper 1st incisor tip, L1 = Lower 1st incisor tip

Table 6. The percentage of the landmarks identified in the range of different thresholds of errors in different automatic cephalometric analysis programs

Mean Error Range (mm)	Landmarks, n (%)		
	Group I	Group II	Group III
< 1.0	0 (0.00)	0 (0.00)	0 (0.00)
< 2.0	1 (7.14)	2 (14.29)	1 (7.14)
< 3.0	12 (85.71)	12 (85.71)	11 (78.57)
< 4.0	14 (100.00)	13 (92.86)	14 (100.00)

Group I : V-ceph version 8.2, Group II : Ceppro version 2.1, Group III : Laonceph version 2.1

IV. 총괄 및 고찰

부정 교합은 크게 골격적 문제와 치아 교합의 문제로 나뉘며, 개인의 골격적 문제를 평가하기 위해서는 측모두부방사선 계측이 필수적이다. 측모두부방사선 분석의 진단학적 가치는 해부학적 계측점 식별의 정확성과 정밀도에 크게 의존하는데[7], 계측점 식별의 정확성은 측정 오류가 없는 것을 의미하며, 정밀도는 같은 대상에 대한 반복 측정 시, 측정값의 편차가 크지 않고 서로 근접함을 뜻한다[8]. 대부분의 측모방사선 분석법의 결과 오류는 체계적 오차와 무작위 오차로 이루어지는데[3], 체계적 오

차는 측정 대상이 동일하지 않거나 두 명 이상의 검사자, 혹은 검사자의 분석 시기가 다를 때 발생하는 문제점으로 연구 진행 시 필연적으로 발생하는 한계적인 반면, 무작위 오차는 계측점 탐지에서의 오차와 측정 단계에서의 오차를 의미한다. 그 중, 측모 두부 계측 분석법의 무작위 오차는 특정 계측점 식별의 어려움과 계측점 자체의 부정확한 정의에서 기인한다[8]. Savage 등 [9]은 계측점 식별과 측정 시 발생하는 오차 모두 측모방사선 분석 결과에 영향을 주지만, 계측점 식별에서의 다양성이 측정에서의 다양성보다 약 5배 크다고 발표하였다. 즉, 정확한 계측점 식별이 측모방사선 분석 결과의 신뢰성에 상당한 영향을 미친

다. 따라서 임상가의 주관성에 기인하는 검사자 사이의 계측점 식별 오차를 줄임과 동시에 높은 정확성을 갖는 자동 측모두부 방사선 분석 방법을 개발하기 위해서 많은 연구들이 진행되었으며, 실제로 임상에서 사용할 수 있도록 자동 측모두부 분석 프로그램이 출시되었다. 하지만 상품화된 자동 측모방사선 분석 프로그램들의 신뢰도 및 정확성을 평가한 연구는 많지 않았다.

이 연구는 시판된 자동 측모방사선 분석 프로그램 3가지 종류를 사용하여 연세대학교 치과대학병원에 내원한 소아 환자들의 측모 두부 계측점을 식별한 후 정확성을 평가하였고, 각 프로그램에 따른 차이를 통계적으로 분석하였다. Rakosi[10]과 Tong 등[11]은 측모 두부 계측 분석 시 임상적으로 허용 가능한 계측점 식별 오차 범위는 2.0 mm 내외라고 보고하였으며, Yue 등[12]은 평균 2.0 mm의 오차 범위는 정확하다고 평가할 수 있고, 4.0 mm 이내의 범위는 임상적으로 허용 가능한 오차라고 발표하였다. 이번 연구 결과 실험에 포함된 V-ceph, Ceppro, Laonceph 프로그램은 각각 평균 2.54 ± 1.63 mm, 2.53 ± 1.63 mm, 2.54 ± 1.63 mm의 오차를 나타내어 3종의 프로그램 모두 평균 2.0 mm 대의 오차 범위 내에서 계측점을 식별함을 알 수 있었다. 따라서 소아 환자들의 교정 분석 시, 자동 계측점 식별 프로그램을 보조적으로 사용할 수 있을 것으로 사료된다.

자동 계측점 식별 프로그램 간 정확도 비교에서는, Or과 Go를 식별할 때 실험군 간 유의미한 차이가 관찰되었다. Or은 V-ceph 프로그램으로 측정하였을 때 Ceppro와 Laonceph 프로그램보다 약 1.0 mm가량 낮은 측정 오차를 나타냈으며 Go를 측정하였을 때는 V-ceph과 Laonceph이 Ceppro 프로그램보다 두드러지게 낮은 식별 오차를 가졌다. 그러나 위 두가지 계측점을 제외한 나머지 12개의 계측점에서는 3군 사이의 유의미한 차이는 관찰되지 않았다. 이는 3종의 프로그램이 서로 유사한 정확도를 가진다는 결과를 보여주었다.

이번 연구에서의 자동 계측점 식별 오차는 이전의 비슷한 주제의 선행 연구들과 유사하거나 다소 낮다. 1998년 Rudolph 등[13]은 14명의 측모방사선 사진을 대상으로 15개의 계측점을 식별하였는데 총 평균 3.07 mm의 오차를 보고하였고, 2000년 Hutton 등[14]의 연구에서는 63명의 측모방사선을 대상으로 16개의 계측점을 식별한 결과 평균 4.08 mm의 오차를 발표하였다. 2000년 Liu 등[15]은 38명을 대상으로 13개의 계측점을 식별하였는데 총 2.86 mm의 평균 식별 오차를 나타내어 이번 연구 결과와 유사함을 알 수 있다. 한편 최근에 발표된 몇몇 연구들은 이번 연구보다 더 정확성이 높은 결과를 보고했다. Giordano 등[16]은 계측점 평균 식별 오차를 1.07 mm로 발표하였고, Yue 등[12]의 연구에서는 12개의 계측점 중 71%이 2.0 mm 이내의 평균 오차를 나타냈으며, Leonardi 등[17]은 10개의 계측점을 자동

으로 식별한 결과 최대 0.59 mm의 낮은 오차를 보였다고 발표하였다. 이는 S만이 유일하게 14개의 계측점 중 2.0 mm 이내의 식별 오차를 나타내며, 전체적으로 평균 2.53 mm의 오차를 보인 이번 연구보다 더 높은 정확도를 보임을 알 수 있다. 하지만 Giordano 등[16]은 8개의 계측점만을 연구에 포함하였으며, Yue 등[12]의 연구는 A, B, Lower 1st incisor tip (L1) 계측점을, Leonardi 등[17]은 ANS, PNS, Or 계측점을 식별 대상에서 제외하였다. 계측점의 계측 신뢰도에 대하여 보고한 이전의 연구들에 의하면, Po과 Or은 계측점 주위의 다양한 해부학적 구조의 중첩으로 명확하게 식별하기 어렵다. 또한 A, B, Me과 Go는 점진적으로 구부러지는 곡선 위에 존재하는 기준점이기 때문에, 명확한 직선 변연 위에 위치하는 계측점인 Upper 1st incisor tip이나 L1과 비교하여 정확한 위치를 파악하는 것에 한계가 있다[18,19]. 따라서 최근에 발표된 연구들의 경우 상대적으로 식별하기 어려운 계측점을 제외하고 소수의 계측점을 가지고 측정하였기 때문에 측정된 오차가 과소평가 되었을 가능성을 생각해볼 수 있으며, 이번 연구에는 Downs 분석법이나 McNamara 분석법에서 공통적으로 사용되는 14개 계측점을 포함함으로써 3종의 자동 분석 프로그램들의 효율성을 더 정확하게 확인할 수 있었다.

그러나 이 연구에는 몇 가지 한계점이 있다. 우선 상품화된 3개의 자동 측모방사선 분석 프로그램들의 학습 데이터의 대상이 성인 또는 영구치열기의 청소년에 국한되어 있는지, 혼합치열기 어린이도 포함되어 있는지 파악하지 못했다는 점이다. 2010년 Tanikawa 등[6]은 기존의 성인 및 청소년의 측모 방사선에 적합한 자동 계측점 인식 시스템을 사춘기 이전의 어린이를 대상으로 추가적으로 학습시킨 후, 사춘기 전 아동의 측모 두부 계측점 식별의 정확도 차이를 비교 평가하였다. 그 결과, Ptm, Pog, B, Me과 Gnathion에서는 두 시스템 모두 80% 이상의 유사한 성공률을 보였으나, 9개의 계측점(PNS, A, Condylion, S, Or, ANS, L1, Ar, N)은 소아를 대상으로 학습시킨 시스템이 기존의 방법과 비교하여 유의미하게 높은 성공률의 개선을 보였음을 보고하였다. 이는 소아 환자들의 특성에 대한 학습의 유무에 따라서 계측점 식별 능력에 유의미한 차이가 나타남을 의미한다. 그러나 이번 연구에서는 연구자가 직접 프로그램을 기계 학습시키지 않고 이미 출시된 프로그램들의 정확도를 평가하였기 때문에, 추후 연구에서는 프로그램의 학습 대상을 분석하고, 더 많은 소아 환자들의 측모방사선 사진을 학습 자료에 포함시켜서 정확도를 높이는 것이 필요하다.

두 번째는 평가에 포함된 계측점 안에 연조직 계측점이 포함되지 않았다는 점이다. 두부 계측은 측모 방사선 촬영을 통해 치아 교정 치료 및 악교정 수술의 관점에서 치아-골격 및 연조직 구조를 검사하는 것이 목적이기 때문에 환자의 연조직 분석도

중요하다. Arnett과 Bergaman[20]이 연조직 분석을 교정 치료의 진단과 치료계획 수립의 중요한 열쇠로 묘사하였을 만큼, 연조직 측모 두부 계측 분석은 수직 및 수평적 차원에서 안모 형태를 평가하며, 치료 전과 후의 심미성을 비교 분석하는 기준이 된다[21]. 연조직은 방사선 사진에서 경조직보다 흑화도 대비가 크지 않아, 계측점 식별 판단이 어려울 수 있으나, 이후 연구에서는 Hwang 등[22]의 논문과 같이 대표적인 연조직 계측점을 포함 기준에 추가하여 치료 전과 후의 환자의 안모 변화를 평가하고 치료 결과를 다각적으로 판단하는 것이 필요할 것으로 보인다.

세 번째는 대조군이 단일 검사자라는 점이다. Anuwong-nukroh 등[23]은 수동 측정법과 자동 계측점 분석 프로그램(OrthoTrac, version 6.14, Carestream Dental, Ontario, United States) 사이의 계측점 식별 정확성을 비교한 연구에서 13개의 계측점 중 5개의 계측점(Sella, Nasion, Gnathion, Pogonion, Gonion)에서 유의미한 차이가 없으며, 평균적으로 x축 0.89 mm, y축 1.32 mm의 식별 오차를 가짐을 발표하였다. 이 때 Anuwongnukroh 등[23]은 2명의 검사자가 식별한 좌표의 평균을 기준 좌표로 하여 식별 오차를 계측하였다. 반면 이번 연구에서는 3개의 프로그램의 정확성을 평가하는 기준을 단일 검사자가 3회 측정된 결과값의 평균으로 설정하였으며 그 결과 3개의 프로그램들 평균 2.53 mm의 식별 오차를 나타냈고, 검사자가 3개의 실험군과 비교하여 14개의 계측점에서 모두 유의미하게 낮은 오차를 보였다. 비교 연구에 있어서는 표준화가 매우 중요한데[8], 2명 이상의 검사자가 측정할 경우 필연적으로 발생할 수 있는 오차를 고려해볼 때, 1명의 검사자가 모든 측정을 시행할 경우 오차를 최소화할 수 있는 장점이 있으며[24], 이번 연구에서 관찰자 내 일치도는 0.90로 높았다. 그러나 측모두부방사선 계측의 경우 관찰자 내 다양성보다 관찰자 간 다양성이 더 크며, 때때로 검사자 사이의 차이가 방사선 이미지의 질보다 측정 결과에 큰 영향을 미치기 때문에 단일 검사자의 측정 값을 기준으로 할 경우 검사자의 경험 및 능력에 따라서 결과에 큰 차이가 존재할 수 있다[25]. 즉, 계측점 식별 오차가 1.0 mm 이내일 경우에 측모 두부 분석 결과를 정확하다고 판단할 수 있음을 고려하면[26,27] 이번 연구에 사용된 3개의 자동 측모 분석 프로그램들은 숙련된 교정 전문가를 대체하기 어려우며, 유사한 주제의 이전 연구[23]와 비교하여 낮은 정확도를 나타낸다고 분석할 수 있으나, 정확도 평가에 사용된 기준이 단일 검사자의 평균 측정값이므로 결과의 신뢰도에 한계가 존재할 가능성을 배제할 수 없다. 따라서 향후 연구에서는 2명 이상의 숙련된 검사자들의 반복 측정 결과의 평균을 기준값으로 설정한다면 보다 신뢰도 높은 연구가 될 것이다.

V. 결 론

이 연구는 연세대학교 치과대학병원에 내원한 만 7 - 12세 환자 60명을 대상으로 자동 측모두부방사선 분석 프로그램의 계측점 식별 정확도와 프로그램에 따른 정확도 차이 및 임상 적용 가능성을 알아보려고 하였다. 자동 측모두부방사선 분석 프로그램은 평균 2.53 mm의 식별 오차를 나타내었고, 3개의 프로그램 사이의 성능에는 유의미한 차이는 없었다. 검사자와의 비교에서는 14개의 계측점에서 모두 검사자가 식별하였을 때가 자동 측모두부방사선 분석 프로그램에 비해서 유의미하게 낮은 식별 오차를 보였다. 자동 측모두부방사선 분석 프로그램은 사람을 대체할 정도의 정확도를 갖지는 못하나 임상적으로 허용 가능한 범위 이내의 측정 오차를 나타내며, 분석할 때 소요되는 시간이 매우 짧으므로 소아 환자들을 대상으로 교정 분석을 진행할 때, 효율적인 보조 진단 도구로 사용할 수 있을 것으로 사료된다.

Authors' Information

Min Sun Song <https://orcid.org/0000-0001-6532-8080>
 Seong-Oh Kim <https://orcid.org/0000-0002-8620-1377>
 Ik-Hwan Kim <https://orcid.org/0000-0003-4444-532X>
 Chung-Min Kang <https://orcid.org/0000-0001-7813-3741>
 Je Seon Song <https://orcid.org/0000-0001-8620-5629>

References

1. Shahidi S, Oshagh M, Danaei SM, *et al.* : Accuracy of computerized automatic identification of cephalometric landmarks by a designed software. *Dentomaxillofac Radiol*, 42: 20110187, 2013.
2. Broadbent BH : A new x-ray technique and its application to orthodontia. *Angle Orthod*, 1:45-66, 1931.
3. Baumrind S, Miller DM : Computer-aided head film analysis: the University of California San Francisco method. *Am J Orthod*, 78:41-65, 1980.
4. Cohen AM, Ip HH, Linney AD : A preliminary study of computer recognition and identification of skeletal landmarks as a new method of cephalometric analysis. *Br J Orthod*, 11:143-154, 1984.
5. Arik S, Ibragimov B, Xing L : Fully automated quantitative cephalometry using convolutional neural networks. *J Med Imaging*, 4:014501, 2017.
6. Tanikawa C, Yamamoto T, Yagi M, Takada K : Automatic rec-

- ognition of anatomic features on cephalograms of preadolescent children. *Angle Orthod*, 80:812-820, 2010.
7. Stabrun AE, Danielsen K : Precision in cephalometric landmark identification. *Eur J Orthod*, 4:185-196, 1982.
 8. Houston WJ : The analysis of errors in orthodontic measurements. *Am J Orthod*, 83:382-390, 1983.
 9. Savage AW, Showfety KJ, Yancey J : Repeated measures analysis of geometrically constructed and directly determined cephalometric points. *Am J Orthod Dentofacial Orthop*, 91:295-299, 1987.
 10. Rakosi T : An atlas of manual cephalometric radiography. Wolfe Medical Publications, 7-19, 1982.
 11. Tong W, Nugent ST, Fay DF, *et al.* : Landmarking of cephalograms using a microcomputer system. *Comput Biomed Res*, 23:358-379, 1990.
 12. Yue W, Yin D, Xu T, *et al.* : Automated 2-D cephalometric analysis on X-ray images by a model-based approach. *IEEE Trans Biomed Eng*, 53:1615-1623, 2006.
 13. Rudolph DJ, Sinclair PM, Coggins JM : Automatic computerized radiographic identification of cephalometric landmarks. *Am J Orthod Dentofacial Orthop*, 113:173-179, 1998.
 14. Hutton TJ, Cunningham S, Hammond P : An evaluation of active shape models for the automatic identification of cephalometric landmarks. *Eur J Orthod*, 22:499-508, 2000.
 15. Liu JK, Chen YT, Cheng KS : Accuracy of computerized automatic identification of cephalometric landmarks. *Am J Orthod Dentofacial Orthop*, 118:535-540, 2000.
 16. Giordano D, Leonardi R, Distefano ML, *et al.* : Automatic Landmarking of Cephalograms by Cellular Neural Networks. *Artificial Intelligence in Medicine*, 3581:333-342, 2005.
 17. 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, 2009.
 18. Baumrind S, Frantz RC : The reliability of head film measurements. 1. Landmark identification. *Am J Orthod*, 60:111-127, 1971.
 19. Kazandjian S, Kiliaridis S, Mavropoulos A : Validity and reliability of a new edge-based computerized method for identification of cephalometric landmarks. *Angle Orthod*, 76:619-624, 2006.
 20. Arnett GW, Bergman RT : Facial keys to orthodontic diagnosis and treatment planning. Part I. *Am J Orthod Dentofacial Orthop*, 103:299-312, 1993.
 21. Arnett GW, Jelic JS, Bergman R, *et al.* : Soft tissue cephalometric analysis: diagnosis and treatment planning of dentofacial deformity. *Am J Orthod Dentofacial Orthop*, 116:239-253, 1999.
 22. Hwang HW, Park JH, Lee SJ, *et al.* : Automated identification of cephalometric landmarks: Part 2- Might it be better than human? *Angle Orthod*, 90:69-76, 2020.
 23. Anuwongnukroh N, Dechkunakorn S, Kangern S, *et al.* : Accuracy of automatic cephalometric software on landmark identification. *Mater Sci Eng*, 265:1-6, 2017.
 24. Erkan M, Gurel HG, Nur M, Demirel B : Reliability of four different computerized cephalometric analysis programs. *Eur J Orthod*, 34:318-321, 2012.
 25. McWilliam JS, Welander U : The effect of image quality on the identification of cephalometric landmarks. *Angle Orthod*, 48:49-56, 1978.
 26. Trpkova B, Major P, Prasad N, Nebbe B : Cephalometric landmarks identification and reproducibility: a meta analysis. *Am J Orthod Dentofacial Orthop*, 112:165-170, 1997.
 27. Richardson A : A comparison of traditional and computerized methods of cephalometric analysis. *Eur J Orthod*, 3:15-20, 1981.

국문초록

소아 환자 대상의 자동 계측점 식별 프로그램의 정확성 평가

송민선¹ · 김성오^{1,2} · 김익환¹ · 강정민^{1,2} · 송제선^{1,2}

¹연세대학교 치과대학 소아치과학교실

²연세대학교 치과대학 구강과학연구소

이 연구의 목적은 소아 환자들의 측모방사선 사진을 대상으로 시판되는 3종의 자동 계측점 식별 프로그램의 정확성을 평가하고 소아치과 임상에서의 사용 가능성을 예측하는 것이다.

영구 중절치가 맹출한 만 7 - 12세 건강한 어린이 60명의 측모방사선 사진에 14개의 계측점을 표시하였다. 1명의 검사자가 3회 반복 측정된 결과의 평균을 기준점으로 정의하여 자동으로 식별된 계측점과의 거리 차이를 계측하였다.

3종의 자동 계측점 식별 프로그램은 평균 2.53 mm의 오차를 나타냈다. Orbitale과 Gonion을 제외한 12개의 계측점에서 3종의 프로그램 사이에 유의미한 차이는 없었으나, 검사자가 모든 계측점에서 3종의 프로그램보다 유의미하게 높은 정확도를 보였다. 이 연구를 통하여 사춘기 전 소아의 측모방사선 분석 시 자동 계측점 식별 프로그램이 검사자를 대체할 정도는 아니나 짧은 소요시간과 임상 허용 가능한 범위 이내의 정확도를 갖는 효과적인 진단 보조기구임을 알 수 있다.