World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

Development of 3D Artificial Intelligence for maxillofacial morphology identification

I Gusti Aju Wahju Ardani [1,*], Riezki Dwianggraini Wahyudi [1], Olivia Halim [1], Aya Dini Oase Caesar [1], Shirley Gautama [1], Ida Bagus Narmada [1], Ervina Restiwulan Winoto [1] and Aziz Fajar [2]

[1] *Orthodontic Department, Faculty of Dental Medicine, Universitas Airlangga, Surabaya, Indonesia.*
[2] *Data Science Technology, Faculty of Advanced Technology and Multidiscipline, Universitas Airlangga, Surabaya, Indonesia.*

World Journal of Advanced Research and Reviews, 2025, 26(03), 099–109

Publication history: Received on 26 March 2025; revised on 03 May 2025; accepted on 06 May 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.26.3.1647

Abstract

Introduction: Artificial Intelligence (AI) in 3D image analysis using Cone Beam Computed Tomography (CBCT) can be used to provide accurate and reliable orthodontic diagnosis. When deciding on an orthodontic treatment plan and evaluation, it is crucial to identify the maxillary and mandibular morphology. Objectives: To identify the maxillary and mandibular morphology of patients with malocclusion at the Dental Hospital Universitas Airlangga's Orthodontic Specialist Clinic using AI. This project is a preliminary study into the development of 3D AI-based digital tracing software. Material and methods: A total of 17 CBCT x-rays of class I malocclusion patients with Javanese ethnicity were divided into training and validation samples. After being manually annotated, the training samples were loaded into deep learning software. Deep learning using Convolutional Neural Network (CNN) is repeated until the manual annotation points and prediction points reach the most accurate coordinates. The results were validated using the validation samples. Results: The lowest MSE in maxillary morphology is at the as point (808.4) and the highest is at the ANS point (3043.8), while in mandibular morphology, the lowest MSE is at the Pg point (927) and the highest is at the Cd-MR point (8675). Even though there are still a number of anatomical landmark locations with high error rates, the outcomes of deep learning are fairly acceptable. Conclusion: CNN-based AI deep learning models can be used to identify maxillary and mandibular anatomical landmarks on CBCT x-rays, however additional data are still required to maximize the deep learning outcomes.

Keywords: Artificial Intelligence; Cone Beam Computed Tomography; Anatomical Landmark; Maxillary and Mandibular Morphology

1. Introduction

One technology that is currently quite popular and developing quickly is Artificial Intelligence (AI). AI is the ability of a computer system to carry out tasks that often call for human intelligence. This technology can use the data in the system to analyze and make judgments [1]. In dental radiology, AI has been utilized to enhance picture interpretation [2]. Computerized cephalometric analysis, such as the WebCeph program, allows for direct digital measurement of cephalometric angles without the need for traditional methods involving protractors and acetate tracing paper. This is made possible by advancements in software and computer technology [3].

In the field of orthodontics, radiographic examination is a supportive examination that is used to examine growth patterns, malocclusion, soft tissue morphology, jaw connection, incisor inclination, facial development, and treatment limits [4]. Standardization is required for the measurement of landmark points, lines/planes, and angles on the facial skeleton in cephalometric radiography, one of the supporting radiographs in the area of orthodontics [5]. A crucial step

* Corresponding author: I Gusti Aju Wahju Ardani

in creating a treatment plan is labeling anatomical landmarks in both soft and hard tissue on cephalometric radiographs [6].

One of the biggest challenges for inexperienced practitioners when assessing cephalometric radiographs is figuring out which landmark points to use [7]. In cephalometric radiography analysis, "projection errors" and "tracking errors" are two types of errors that might happen. When converting three-dimensional objects into two-dimensional radiographs, projection errors happen. These mistakes vary depending on the relative positions of the film, object, and x-ray tube. According to Gravely (1974), tracking errors can be brought on by a variety of factors, including superimposition of structures that obscures cephalometric landmarks, movement during exposure that blurs the image, a lack of contrast between the film and emulsion grains, measurement errors related to pencil line thickness, and limits on human perception [8].

Three-dimensional, irregularly shaped structures make up the maxilla and mandible. As stated by Dong et al. (2021) three-dimensional structures' real circumstances cannot be fully captured by two-dimensional photographs. As an alternative to medical computed tomography (CT), cone-beam computed tomography (CBCT) provides imaging findings with a lower radiation dosage and a higher resolution (submillimeter) [9]. More precise location of the condylion, gonion, and orbital can be achieved with CBCT since it can produce three-dimensional pictures without the distortion and superimposition of bone and other dental structures observed on traditional radiography [10,11].

Researchers hope to create 3D AI software on maxillofacial morphology in orthodontic patients as AI advances. Maxillofacial morphology is thought to be significant because it can assist orthodontists in making treatment decisions, monitoring, and assessing orthodontic treatment. Since homogeneous samples are necessary for the development of AI software, this study's participants were patients of Javanese ethnicity.

2. Material and methods

The study included patients, either male or female, between the ages of 18 and 25, who were of Javanese ethnicity, had class I skeletal malocclusion with ANB 0–4°, had a complete soft file CBCT, and had not received orthodontic treatment. The patient does not have any systemic medical history, including diabetes mellitus, heart disease, stroke, HIV, osteoporosis, mandibular displacement, facial defects, or abnormalities in the quantity, size, or shape of teeth. Nor does she have a history of maxillofacial trauma.

The study used primary data from August to December 2022 from the CBCT scans of patients at the Orthodontic Specialist Clinic, University of Airlangga Dental Hospital Surabaya. The device used is an Intel i9 CPU GPU Nvidia Quadro 8000 RAM 64 GB computer equipped with AI software development for the Sepuluh November Institute of Technology laboratory and annotation software for identifying anatomical landmarks.

A CBCT examination was performed on a sample of 17 patients. The CBCT samples were split into two sets: three more samples were utilized to confirm the results of identifying anatomical landmarks, and the remaining fourteen samples were used to train AI software. Each patient has 640 slices of CBCT data, which are stored as DICOM data with an image size of 704x704 per slice. After extracting each slice in .jpeg format, the clinical expert chose the right slice for each anatomical landmark.

A technician developed manual annotation software, which the researcher used to manually identify anatomical landmarks. Three-dimensional point coordinates (x, y, z) are the annotation's output. The deep learning technique will use the annotation results as ground truth (GT). The data and ground truth pairs used in this study will be split into two categories: training data and validation data. Validation data is used to determine whether the deep learning model developed can generalize data that was not utilized in training. Training data is used to conduct the deep learning process for patient samples. The Python programming language is then used to input the training sample data into AI software for deep learning. Deep learning is performed repeatedly until the prediction points and manual annotation points (ground truth) are at the most precise coordinates. Validation was then performed on the fourteen samples. Next, the results of learning to identify the anatomical landmarks of maxillofacial morphology were tested using three test data sets.

Deep Learning uses a Convolutional Neural Network (CNN) consisting of convolution layers that will perform feature extraction. These features will be detected during the training process. In deep learning, it is also necessary to select an error function. The error function contains a function to calculate the error between the deep learning prediction results and ground truth. The deep learning architecture used is a deep learning architecture for regression, where the deep learning model used will estimate the coordinate values for each 3-dimensional axis.

Error measurements were carried out using the TensorFlow library and Keras (a method for calculating the distance between two points from AI). The error function used in this deep learning model is Mean Squared Error (MSE). MSE is the average squared error between the actual value and the predicted value. An MSE value that is low or close to zero indicates that the prediction results are in accordance with ground truth data. The Mean Squared Error (MSE) formula is as follows:

$$MSE = \frac{\sum_{t=1}^{n}(\mathrm{At} - \mathrm{Ft})^2}{\mathrm{n}}$$

At: Ground truth value

Ft: Prediction/validation value

n: The amount of data

3. Results

The research findings are represented by coordinate numbers (x, y, z), which are values derived from the axial, coronal, and sagittal dimensions. The coordinate numbers are separated into two categories: Ground Truth (GT) data and Prediction data. The outcomes of AI deep learning, which is done consistently to produce low Mean Square Error (MSE) values, are used to generate prediction data. Ground Truth data, on the other hand, comes from manual annotation.

Table 1 Recapitulation of Prediction Data and Ground Truth Deep Learning for Point as, ANS, M1-R, Cp-L, B, and Pg

Sample	Data	as	ANS	M1-R	Cp-L	B	Pg
A	Prediction	[218 501 356]	[167 513 361]	[310 391 211]	[124 295 659]	[515 575 309]	[615 561 351]
	GT	[220 601 316]	[193 675 349]	[329 505 212]	[161 409 675]	[532 605 303]	[607 633 307]
B	Prediction	[211 519 346]	[150 499 355]	[303 382 214]	[135 286 657]	[521 573 335]	[623 556 355]
	GT	[233 539 330]	[206 567 332]	[329 428 207]	[158 324 665]	[496 565 347]	[560 589 376]
C	Prediction	[217 509 347]	[159 492 357]	[307 379 219]	[141 288 652]	[515 568 331]	[616 548 354]
	GT	[228 542 336]	[202 576 351]	[327 431 216]	[153 321 654]	[513 564 299]	[577 589 383]
D	Prediction	[218 506 350]	[161 498 357]	[306 383 216]	[137 290 654]	[514 569 326]	[616 553 354]
	GT	[242 522 345]	[200 548 362]	[347 419 227]	[172 291 676]	[509 544 331]	[594 562 354]
E	Prediction	[209 529 339]	[137 482 349]	[296 370 216]	[149 278 653]	[523 568 361]	[631 549 359]
	GT	[234 528 319]	[200 550 334]	[326 419 206]	[99 305 633]	[512 514 347]	[579 522 353]
F	Prediction	[221 498 364]	[179 544 368]	[320 406 203]	[105 305 668]	[518 583 285]	[608 575 345]
	GT	[283 505 327]	[217 528 365]	[390 385 209]	[241 259 636]	[521 526 350]	[591 534 347]
G	Prediction	[212 525 337]	[142 473 348]	[297 368 222]	[157 280 647]	[518 562 362]	[626 540 359]

	GT	[231 537 336]	[194 570 365]	[328 428 212]	[108 327 644]	[512 536 352]	[596 545 353]
H	Prediction	[213 520 344]	[149 493 353]	[302 378 215]	[142 286 654]	[519 568 344]	[624 551 357]
	GT	[216 533 371]	[160 564 357]	[324 418 202]	[109 286 637]	[498 539 371]	[568 543 371]
I	Prediction	[212 516 351]	[154 510 358]	[306 387 210]	[127 290 661]	[522 576 324]	[622 562 353]
	GT	[223 523 385]	[164 546 362]	[339 409 210]	[161 291 640]	[494 544 346]	[576 555 347]
J	Prediction	[216 507 355]	[162 517 360]	[309 391 208]	[122 294 662]	[519 577 313]	[618 565 351]
	GT	[231 526 373]	[191 578 333]	[364 408 204]	[168 283 669]	[510 550 415]	[604 583 382]
K	Prediction	[222 494 360]	[178 521 366]	[318 396 212]	[118 299 660]	[514 576 296]	[605 563 346]
	GT	[233 547 381]	[202 575 378]	[351 422 207]	[177 335 648]	[529 558 369]	[633 564 398]
L	Prediction	[220 503 354]	[166 505 361]	[311 386 214]	[130 292 656]	[515 572 317]	[613 556 350]
	GT	[234 534 335]	[196 578 341]	[343 438 216]	[158 307 660]	[510 562 331]	[577 583 366]
M	Prediction	[219 502 356]	[169 514 362]	[312 391 212]	[124 296 659]	[516 574 308]	[612 560 349]
	GT	[255 523 336]	[199 556 363]	[353 427 199]	[116 263 648]	[540 544 306]	[613 567 340]
N	Prediction	[220 500 353]	[169 507 361]	[312 388 216]	[130 294 656]	[513 571 314]	[611 556 352]
	GT	[228 537 328]	[216 575 350]	[346 410 198]	[74 330 702]	[541 536 286]	[577 559 319]

The results above are then processed to determine the error function for each data using Mean Squared Error (MSE). This research is a 3-dimensional research where the average MSE error function is better seen in all dimensions (axial, coronal and sagittal). To get the average MSE error function, it can be seen in Table 2 and Table 3 which was obtained by calculating using the MSE average formula as follows:

$$\text{Mean MSE} = \frac{[(\text{MSE Results X}) + (\text{MSE Results Y}) + (\text{MSE Results Z})]}{3}$$

Table 2 Recapitulation of Mean Error Function Data Mean Squared Error (MSE) of Maxillary Morphology

Sample	A	as	is	ANS	PNS	P1-L	P1-R	M1-L	M1-R
A	5909	3868	3985	9021	651	1633	979	154	4453
B	3156	380	2216	2763	1140	940	292	2322	947
C	3019	444	1753	2980	1426	655	303	2627	1038
D	1444	286	1670	1349	3213	873	578	1645	1033
E	3211	342	1095	2939	1586	1124	1611	3884	1134
F	622	1754	1000	570	4980	2997	1444	5135	1792
G	4166	169	3312	12402	2019	1267	935	2305	1554
H	3338	302	1854	1726	2788	475	1212	3258	751
I	1212	442	1163	471	1663	956	1136	4419	524
J	2221	303	1500	1764	1446	794	800	2350	1110
K	2178	1124	1315	1212	2659	615	471	1761	597
L	2051	506	1342	2210	940	987	412	1382	1244
M	1288	712	1058	888	2814	1186	670	2588	1049
N	2936	686	1360	2318	1710	627	769	2022	655
Mean	2625,1	808,4	1758,7	3043,8	2073,9	1080,5	829,48	2560,7	1277,1

The annotation points and deep learning validation result points with the biggest (least accurate) and smallest (most accurate) MSE, as determined by calculations made in Table 2 and Table 3, are shown in Figure 1-6. The image below shows the display of deep learning validation results along with captions: Yellow dots represent CBCT deep learning prediction points, and red dots represent Ground Truth (GT) points.

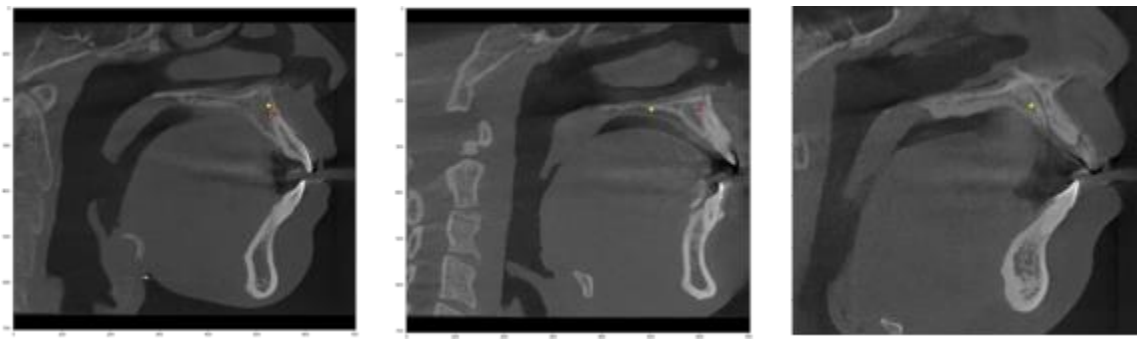

Figure 1 (a) Validation results of as points with the smallest MSE (b) Validation results of as points with the largest MSE (c) as test results with deep learning test data

Figure 1(a) shows the validation results of the axle point with the smallest MSE in sample G, namely 169 and Figure 1(b) shows the validation results of as point with the largest MSE in sample A, namely 3868. In figure 1(a), it can be seen that the prediction points are close to the GT point, while in Figure 1(b), it can be seen that the distance between the prediction points is very far from the GT point. In Figure 1(c) it can be seen that the prediction results for the as point using deep learning are more posterior than the anatomy should be. Therefore, it can be said that the as point prediction results are still not accurate.

Table 3 Recapitulation of Mean Error Function Data Mean Squared Error (MSE) of Mandibular Morphology

SAMPLE	Cp-L	Cd-AL	Cd-PL	Co-L	Go-L	Sg-L	B	ai	ii	Pg	Gn	Me	Cp-R	Cd-AR	Cd-PR	Co-R	Go-R	Sg-R	Cd-ML	Cd-MR
A	4.874	1.332	3.449	3.019	3.539	5.151	408	3.228	1.845	2.395	5.367	1.367	2.141	3.993	1.438	811	2.490	9.895	1.866	2.322
B	679	805	175	1.988	874	1.558	278	2.255	1.441	1.833	2.737	299	961	1.114	171	274	170	6.799	651	691
C	412	2.904	2.191	3.463	658	1.990	348	4.521	1.532	1.348	4.797	291	97	5.144	2.726	2.050	1.926	7.957	76.632	99.425
D	570	1.927	2.358	5.871	4.135	1.383	225	3.021	1.127	188	3.291	262	1.564	2.668	2.541	3.320	808	5.606	3.782	1.782
E	1.210	1.996	1.443	342	190	1.334	1.078	4.833	1.930	1.156	539	3.345	1.086	1.973	1.433	1.058	2.908	4.854	2.327	2.441
F	7.212	5.916	4.782	9.419	5.257	1.234	2.494	1.333	1.197	658	150	1.260	4.361	2.710	1.789	7.501	1.942	1.011	3.676	4.129
G	1.540	1.729	526	2.142	452	1.454	271	4.637	1.487	320	7.281	1.894	1.602	2.998	775	1.883	2.143	1.511	3.205	1.306
H	459	3.670	2.466	4.311	454	1.464	670	2.291	1.057	1.132	1.706	822	942	2.710	2.124	1.565	942	3.459	2.143	2.037
I	533	1.428	1.325	3.505	3.873	269	764	2.369	1.397	734	5.053	3.755	917	1.246	662	1.250	446	2.013	1.210	688
J	762	1.345	1.479	3.102	3.171	175	3.738	1.502	1.258	494	4.365	1.495	2.033	909	2.175	2.915	1.341	2.812	1.542	969
K	1.640	2.614	2.519	5.302	3.183	93	1.959	866	695	1.163	3.567	2.367	525	505	1.553	3.193	1.913	1.455	1.545	2.211
L	342	3.027	1.961	2.823	518	1.700	107	2.138	942	760	4.073	522	723	310	442	780	510	1.829	639	926
M	425	2.547	983	4.967	2.564	2.762	493	2.226	684	44	2.575	1.098	2.129	443	1.748	2.983	2.683	750	3.748	1.620
N	2.183	1.503	855	2.260	1.514	3.341	931	1.545	902	751	1.334	248	1.766	907	1.589	1.020	281	345	1.367	908
Mean	1.632	2.339	1.894	3.751	2.170	1.708	983	2.626	1.250	927	3.345	1.359	1.489	1.974	1.512	2186	1.465	3.593	7.452	8.675

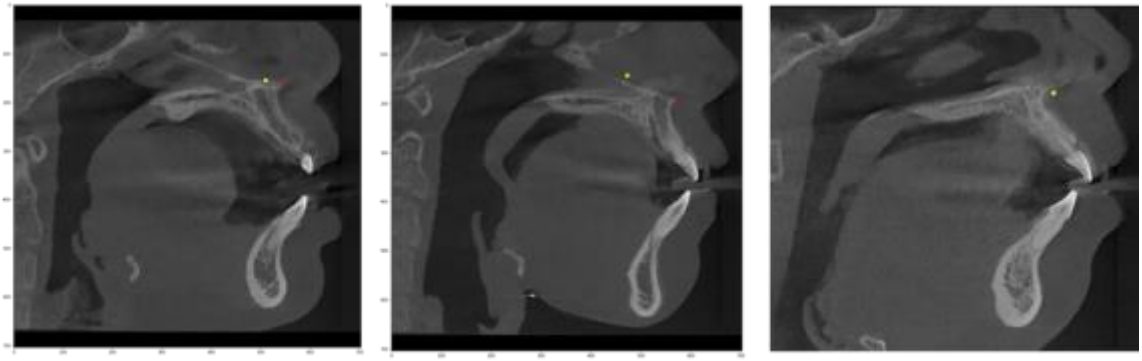

Figure 2 (a) Validation results of ANS points with the smallest MSE (b) Validation results of ANS points with the largest MSE (c) ANS test results with deep learning test data

Figure 2(a) shows the validation results of ANS points with the smallest MSE in sample I, namely 471 and Figure 2(b) shows the validation results of ANS point with the largest MSE in sample G, namely 12402. In figure 2(a), it can be seen that the distance between the prediction points is quite close to the GT point, while in Figure 2(b), it can be seen that the distance between the prediction points is very far from the GT point. In Figure 2(c) it can be seen that the ANS point prediction results using deep learning are more posterior than the anatomy should be. Therefore, it can be said that the ANS point prediction results are still not accurate.

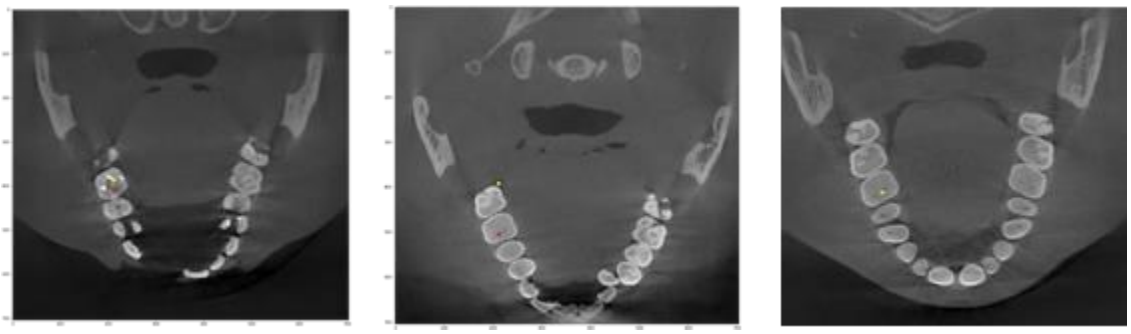

Figure 3 (a) Validation results of M1-R points with the smallest MSE (b) Validation results of M1-R points with the largest MSE (c) M1-R test results with deep learning test data

Figure 3(a) shows the validation results for point M1-R with the smallest MSE in sample I, namely 524, and Figure 3(b) shows the validation results for point M1-R with the largest MSE in sample A, namely 4453. In figure 3(a), it can be seen that the distance between the prediction points is quite close to the GT point, while in Figure 3(b), it can be seen that the distance between the prediction points is far from the GT point. In Figure 3(c) it can be seen that the M1-R point prediction results using deep learning are very slightly deviated from the anatomy they should be with a good level of accuracy. Therefore, it can be said that the prediction results for point M1-R are quite accurate.

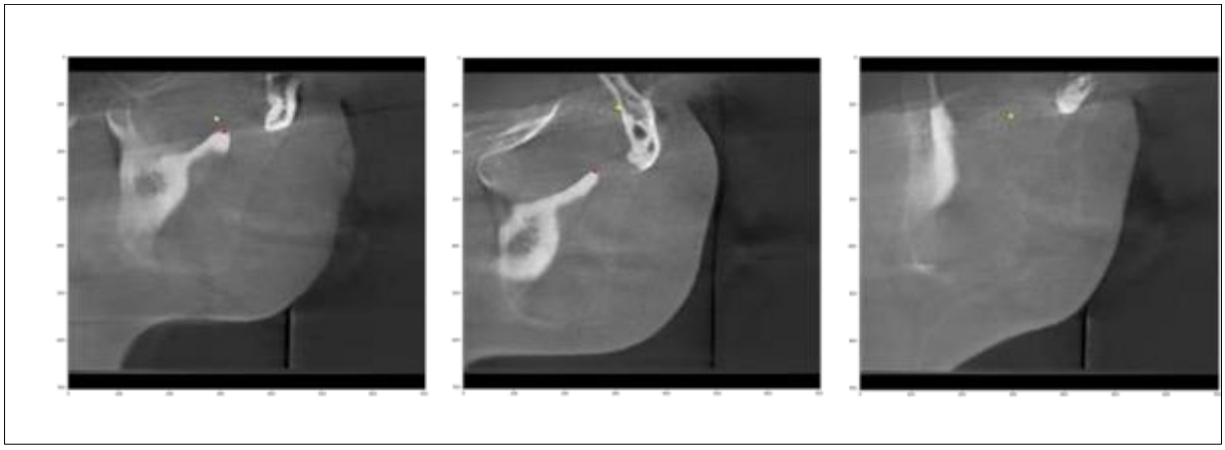

Figure 4 (a) Validation results of Cp-L points with the smallest MSE (b) Validation results of Cp-L points with the largest MSE (c) Cp-L test results with deep learning test data

Figure 4(a) shows the validation results of the Cp-L point with the smallest MSE on sample L, namely 342 and Figure 4(b) shows the validation results of the Cp-L point with the largest MSE on sample F, namely 7212. In figure 4(a), it can be seen that the distance between the prediction points is quite close to the GT point, while in Figure 4(b), it can be seen that the distance between the prediction points is far from the GT point. In Figure 4(c) it can be seen that the prediction results for the Cp-L point using deep learning are not in accordance with the anatomy it should be. Therefore, it can be said that the prediction results for the Cp-L point are still not accurate.

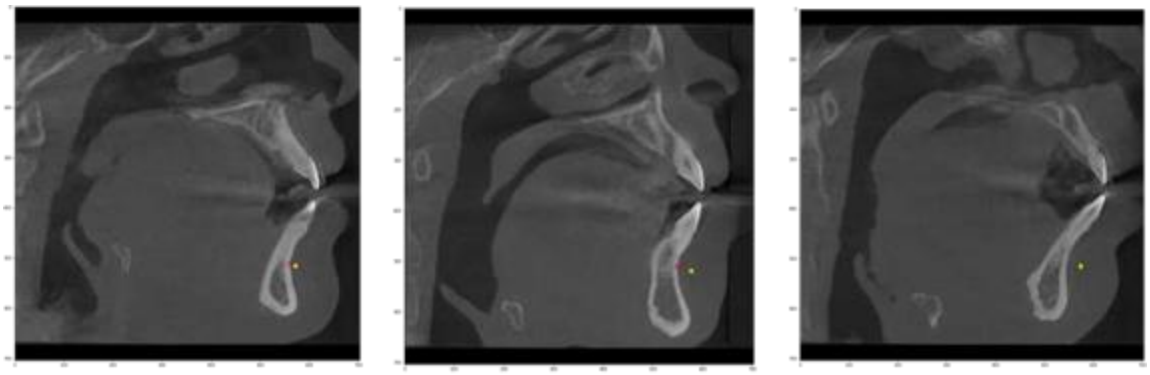

Figure 5 (a) Validation results of B points with the smallest MSE (b) Validation results of B points with the largest MSE (c) B test results with deep learning test data

Figure 5(a) shows the validation results for point B with the smallest MSE on sample L, namely 107 and Figure 5(b) shows the validation results for point B with the largest MSE on sample J, namely 3738. In figure 5(a), it can be seen that the distance between the prediction points is very close to the GT point, while in Figure 5(b), it can be seen that the distance between the prediction points is quite far from the GT point. In Figure 5(c) it can be seen that the prediction results for point B using deep learning are slightly more anterior than the anatomy should be with a fairly good level of accuracy. Therefore, it can be said that the prediction results for point B are quite accurate.

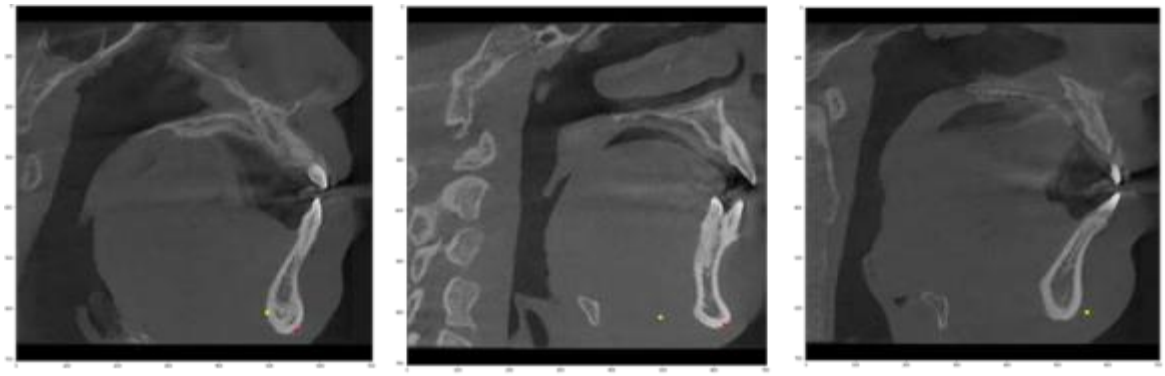

Figure 6 (a) Validation results of Pg points with the smallest MSE (b) Validation results of Pg points with the largest MSE (c) Pg test results with deep learning test data

Figure 6(a) shows the validation results for point Pg with the smallest MSE in sample M, namely 44, and Figure 6(b) shows the validation results for point Pg with the largest MSE in sample A, namely 2395. In figure 6(a), it can be seen that the distance between the prediction points is quite close to the GT point, while in Figure 6(b), it can be seen that the distance between the prediction points is quite far from the GT point. In Figure 6(c) it can be seen that the prediction results for point Pg using deep learning are slightly deviated from the anatomy they should be with a fairly good level of accuracy. Therefore, it can be said that the prediction results for point Pg are quite accurate.

4. Discussion

Technology known as artificial intelligence (AI) can facilitate human labor and boost output.[1] AI has witnessed the advancement and use of machine learning in a number of medical domains over the past few decades. The advancement of medical imaging technology has made it possible to automatically identify anatomical landmarks in cephalometric radiographs [12].

Orthodontic diagnosis is a time-consuming process that involves patient assessment, study models analysis, radiography and photography analysis [13]. Because orthodontists analyze patients in a complicated manner, there is a vast range of treatment approaches available. For the purpose of increasing speed, accuracy, and consistency, these diagnostic techniques ought to be automated.

Because cephalometric analysis based on 2D pictures is frequently imprecise, morphological analysis is improved by 3D imaging, such as CBCT [14]. The development of CBCT provides a more accurate and thorough comprehension of diagnostic pictures, enabling the development of treatment strategies that are more likely to be effective. Anatomical cephalometric landmarks must be automatically detected because manual detection takes a lot of time, practice, and error-prone handling, which reduces accuracy. 2D cephalometric analysis has various unique constraints, so a deep learning method approach on 3D CBCT was used to solve this issue.

Convolutional neural networks (CNNs), a type of neural network that performs well on visual tasks like image classification, picture segmentation, object detection, face recognition, and others, are used in the Deep Learning process. CNN extracts information from pictures using convolutional filters; first layers identify edges, second layers identify objects' pieces, and third layers are able to identify entire things, like faces or other intricate geometric structures [15]. Consequently, a CNN-based anatomical landmark detection model in CBCT x-rays is used in this study.

According to the research findings, three additional samples were used as test data for the anatomical landmark determination process, and fourteen samples were used as training and validation samples for the AI software. These results produced fairly good deep learning outcomes, despite the fact that there were still a number of high error points in anatomical landmarks.

In maxillary morphology, if further MSE calculations are carried out, the average obtained for the entire sample at each point which has the lowest error rate is at the as point, namely 808.4 and the highest is at the ANS point, namely 3043.8. In mandibular morphology, if the MSE calculation is carried out further, the average obtained for the entire sample at each point which has the lowest error rate is at the Pg point, namely 927 and the highest is at the Cd-MR point, namely 8675. This shows that, the results of deep learning validation using AI are the best at the as and Pg points and the worst at the ANS and Cd-MR points.

Test results with new samples (test data) on maxillary morphology, at points A, as, is, ANS, PNS, P1-L, P1-R and M1-L have inaccurate results, while at point M1-R the results are quite accurate. The prediction points at points A, PNS, P1-L, P1-R and M1-L are still far from the anatomical landmarks they should be. Meanwhile, even though the as, is and ANS points are still not accurate, the prediction point results are close to the anatomical landmarks they should be. Test results with new samples (test data) on mandibular morphology, at points Co-R, Co-L, Cd-MR, Cd-ML, Cd-AR, Cd-AL, Cd-PR, Cd-PL, Cp-R , Cp-L, Go-R, Go-L, Gn, Me, Sg-R, Sg-L, ii, ai have inaccurate results, while at points B and Pg the results are quite accurate. Prediction points at points Co-R, Co-L, Cd-MR, Cd-ML, Cd-AR, Cd-AL, Cd-PR, Cd-PL, Cp-R, Go-R, Go-L, Sg- R, Sg-L is still far from the anatomical landmark it should be. Meanwhile, at points Cp-L, ai, ii, Me, Gn, although still not accurate, the results of the predicted points are close to the anatomical landmarks they should be. It is still necessary to test more new samples to train machine learning to form memories and record precise anatomical landmark points.

Even if the research mentioned above yielded an error rate that is still somewhat high, it is nevertheless sufficient as a starting point for the development of AI in the field of orthodontics. The obtained disparity is still very large, but with the ongoing development of deep learning techniques, it may be less. The following factors have led to differences in research findings with varying degrees of precision for every patient and the complete anatomical landmark point of the maxillofacial:

1. Deep learning has time constraints AI program that takes a while to learn. It will take roughly six months to receive the prediction results for 14 research samples with total of 29 anatomical landmarks. More accurate and meaningful findings can only be obtained through repeated learning.
2. Small sample size. At the Orthodontic Specialist Clinic, University of Airlangga Dental Hospital Surabaya, CBCT is still infrequently used to plan orthodontic therapy for patients. Following numerous deep learning iterations, training and validation samples were acquired. More samples are required to boost the variety of learning data because the number of collected samples is still extremely little.
3. Despite the fact that Javanese ethnicity was the primary factor in sample selection, there were differences in the research sample, including differences in head sizes, weights, and heights. These size variances also affect how different CBCT radiography results are. There are variations in the field of vision for each slice, and the mandible's anatomical features—particularly at the condyle—are shortened.
4. Presentation of anatomical landmark annotation software is still displayed on each side of the axis, axial / coronal / sagittal. So there are difficulties in the accuracy of 3D annotation. Manual annotation was only carried out from one axis/dimension and annotations from the other two sides were missed.

Compliance with ethical standards

Acknowledgments

Orthodontic Department of Faculty of Dental Medicine, Universitas Airlangga helped make this research possible, and for that the authors are grateful.

Disclosure of conflict of interest

The authors report no conflict of interest.

References

[1] Xu Y, Liu X, Cao X, Huang C, Liu E, Qian S, Liu X, Wu Y, Dong F, Qiu CW, Qiu J, Hua K, Su W, Wu J, Xu H, Han Y, Fu C, Yin Z, Liu M, Roepman R, Dietmann S, Virta M, Kengara F, Zhang Z, Zhang L, Zhao T, Dai J, Yang J, Lan L, Luo M, Liu Z, An T, Zhang B, He X, Cong S, Liu X, Zhang W, Lewis JP, Tiedje JM, Wang Q, An Z, Wang F, Zhang L, Huang T, Lu C, Cai Z, Wang F, Zhang J. Artificial intelligence: A powerful paradigm for scientific research. Innovation (Camb). 2021 Oct 28;2(4):100179. doi: 10.1016/j.xinn.2021.100179. PMID: 34877560; PMCID: PMC8633405.

[2] Chen YW, Stanley K, Att W. Artificial intelligence in dentistry: current applications and future perspectives. Quintessence Int. 2020;51(3):248-257. doi: 10.3290/j.qi.a43952. Erratum in: Quintessence Int. 2020;51(5):430. doi: 10.3290/j.qi.a44465. PMID: 32020135.

[3] Yassir YA, Salman AR, Nabbat SA. The accuracy and reliability of WebCeph for cephalometric analysis. J Taibah Univ Med Sci. 2021 Sep 22;17(1):57-66. doi: 10.1016/j.jtumed.2021.08.010. PMID: 35140566; PMCID: PMC8801471.

[4] Wang X, Zhang T, Yang E, Gong Z, Shen H, Wu H, Zhang D. Biomechanical Analysis of Grafted and Nongrafted Maxillary Sinus Augmentation in the Atrophic Posterior Maxilla with Three-Dimensional Finite Element Method. Scanning. 2020 Oct 2;2020:8419319. doi: 10.1155/2020/8419319. PMID: 33093935; PMCID: PMC7556061.

[5] White SC, Pharaoh MJ. Oral Radiology: Principles and Interpretation. 5th ed. New Delhi: Mosby; 2004.

[6] Lindner C, Wang CW, Huang CT, Li CH, Chang SW, Cootes TF. Fully Automatic System for Accurate Localisation and Analysis of Cephalometric Landmarks in Lateral Cephalograms. Sci Rep. 2016 May; 6:1–10.

[7] Jacobson A. Radiographic cephalometry: From Basics to 3-D Imaging. London: Quintessence Publishing Co. Inc; 2006.

[8] Gravely JF, Benzies PM. The clinical significance of tracing error in cephalometry. Br J Orthod. 1974 Apr;1(3):95-101. doi: 10.1179/bjo.1.3.95. PMID: 4525738.

[9] Dong Q, Shi H, Jia Q, Tian Y, Zhi K, Zhang L. Analysis of Three-Dimensional Morphological Differences in the Mandible between Skeletal Class I and Class II with CBCT Fixed-Point Measurement Method. Scanning. 2021 May 7;2021:9996857. doi: 10.1155/2021/9996857. PMID: 34040691; PMCID: PMC8121591.

[10] Orhan K, Bayrakdar IS, Ezhov M, Kravtsov A, Özyürek T. Evaluation of artificial intelligence for detecting periapical pathosis on cone-beam computed tomography scans. Int Endod J. 2020 May;53(5):680-689. doi: 10.1111/iej.13265. Epub 2020 Feb 3. PMID: 31922612.

[11] Ludlow JB, Gubler M, Cevidanes L, Mol A. Precision of cephalometric landmark identification: cone-beam computed tomography vs conventional cephalometric views. Am J Orthod Dentofacial Orthop. 2009 Sep;136(3):312.e1-10; discussion 312-3. doi: 10.1016/j.ajodo.2008.12.018. PMID: 19732656; PMCID: PMC2753840.

[12] Alsubai S. A Critical Review on the 3D Cephalometric Analysis Using Machine Learning. Computers 2022; 11: 154.

[13] Silva TP, Hughes MM, Menezes LDS, de Melo MFB, Freitas PHL, Takeshita WM. Artificial intelligence-based cephalometric landmark annotation and measurements according to Arnett's analysis: can we trust a bot to do that? Dentomaxillofac Radiol. 2022 Sep 1;51(6):20200548. doi: 10.1259/dmfr.20200548. Epub 2022 Aug 5. PMID: 33882247; PMCID: PMC10043619.

[14] Bulatova G, Kusnoto B, Grace V, Tsay TP, Avenetti DM, Sanchez FJC. Assessment of automatic cephalometric landmark identification using artificial intelligence. Orthod Craniofac Res. 2021 Dec;24 Suppl 2:37-42. doi: 10.1111/ocr.12542. Epub 2021 Nov 29. PMID: 34842346.

[15] Bezdan T, Bačanin N. Convolutional Neural Network Layers and Architectures. 2019; 445–451. https://doi.org/10.15308/sinteza-2019-445-451.