Iran J Orthod. 2025 December; 20(2):e1265. doi:10.22034/ijo.2025.2064140.1265.
Published online 2025 December 1 Research Article
Original Article **Open Access**

Artificial Intelligence for Automated Cephalometric Analysis in 2D and 3D Craniofacial Imaging

Neeraj Kumar Dudy [1], Shubhnita Verma [2], Prasad Chitra [3*]

[1] Resident, Department of Orthodontics, Army College of Dental Sciences, Secunderabad, India
[2] Associate Professor, Department of Orthodontics, Gitam Dental College and Hospital, Vishakhapatnam, India
[3] Professor, Department of Orthodontics, Army College of Dental Sciences Secunderabad, India

***Corresponding author:** Prasad Chitra

Address: Department of Orthodontics, Army College of Dental Sciences, Gabbilalpet Main Road, Secunderabad, India.

Email: prasadchitra@yahoo.co.uk

Received: 2025 June 25; **Revised:** 2025 September 14; **Accepted:** 2025 October 26

Abstract

Aim: Cephalometry is a fundamental tool for orthodontic diagnosis and treatment planning. Manual landmark annotation is labor-intensive, prone to errors, and exhibits significant intra- and inter-examiner variability. Artificial intelligence (AI) offers a promising solution by providing automated landmark identification at clinically acceptable accuracy levels. This study evaluates the accuracy and reliability of an AI-based system for cephalometric landmark annotation using both lateral cephalograms and cone-beam computed tomography (CBCT) scans.

Methods: A convolutional neural network (CNN) model, based on EfficientNet-B0 pretrained on ImageNet, was developed for end-to-end regression of two-dimensional (2D) coordinates in lateral cephalograms, while a grid-search optimized Automated Landmark Identification system for CBCT (ALICBCT) was employed for three-dimensional (3D) landmark localization. The dataset included 405 lateral cephalograms and 52 CBCT scans from multiple diagnostic centers across India to enhance heterogeneity and generalizability. Twenty-three landmarks were annotated in lateral cephalograms and 18 in CBCTs. Intra- and inter-examiner reliability was assessed using kappa statistics, and comparative evaluation between modalities was performed using paired t-tests. Accuracy was quantified using mean radial error (MRE) and success detection rate (SDR) at clinically relevant thresholds. P-value<0.05 was considered as significant.

Results: For lateral cephalograms, the AI system achieved an average MRE of 1.21 ± 1.00 mm and an SDR of 88.2% at a 2-mm threshold. In CBCT scans, the model achieved an average MRE of 0.95 ± 0.75 mm and an SDR of 94.6%, demonstrating superior precision of 3D imaging. Heatmaps and qualitative assessment confirmed high landmark consistency, particularly for dental and midline skeletal points. Intra- and inter-examiner agreements were substantial ($\kappa = 0.78$ and 0.72, respectively). Comparative analysis indicated that CBCT-based landmark identification outperformed lateral cephalograms for most landmarks, except select midline points.

Conclusion: This study demonstrated that the AI-based system achieved clinically acceptable accuracy and reliability for 2D and 3D cephalometric landmark detection, with CBCT outperforming lateral cephalograms. These findings confirm the feasibility of automated landmark identification in orthodontic diagnostics.

Keywords: Artificial Intelligence, Automated Landmark Identification, Cone-Beam Computed Tomography, Lateral Cephalogram

Copyright © 2025, Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/) which permits copy and redistribute the material just in noncommercial usages, provided the original work is properly cited.

Background

Orthodontic treatment planning is crucial for achieving optimal functional efficiency and aesthetics. Individualized treatment planning is essential for every patient to achieve superior efficiency, predictability, long-term stability, and integration with other dental specialties, such as surgical procedures (orthognathic surgery). Cephalometric analysis is an integral part of diagnosis and treatment planning in orthodontics, emphasizing the morphometric measurement of craniofacial growth and the analysis of spatial relationships between soft and hard tissues (1–3). Conventional cephalometric analysis began in 1948 when William B Downs introduced his linear and angular measurement analysis. This initiated the advent of manual landmark annotation, followed by linear and angular measurements (4).

Other analyses were then introduced in orthodontics for improved diagnosis and treatment planning. The shift from manual to digital cephalometry began in the 1990s. It improved accuracy, reproducibility, segmentation time, and ease of data handling. Conventional cephalometrics is based on two-dimensional (2D) lateral cephalograms, which overlap right and left soft and hard tissues, and are prone to magnification errors and difficulty in marking specific landmarks, such as orbitale, porion, and gonion. Three-dimensional (3D) cephalometrics obtained from the annotation of landmarks on a cone-beam computed tomography (CBCT) scan reduces these errors to an extent. However, annotation of landmarks on both lateral cephalograms and CBCTs is very laborious, time-consuming, and prone to intra- and inter-observer errors (5,6).

In recent years, artificial intelligence (AI), particularly its branches of machine learning (ML) and deep learning (DL), has been increasingly applied to improve the precision, consistency, and dependability of landmark identification in both 2D and 3D cephalometric analyses. Machine learning involves algorithms that learn patterns from structured datasets, enabling decision-making based on underlying statistical relationships. Deep learning, a more advanced subset of ML, employs convolutional neural networks (CNNs), which are multilayered models designed to process data in a way that mimics the neural architecture of the human brain. DL has demonstrated superior performance compared to traditional ML methods in many medical imaging applications, primarily due to its ability to manage high-dimensional radiographic data and automatically extract and learn complex features, such as edges, contours, and shapes, without requiring manual input (7–9).

Since landmark identification is a significant source of error in cephalometric analysis due to observer variability, evaluating whether AI-based approaches can offer a more accurate and time-efficient alternative to conventional manual methods is essential. Many studies have been published, reporting results for the current subject, which achieve accuracy comparable to that of manual processes. These studies used datasets from their native region diagnostic centers or trained AI models on the Institute of Electrical and Electronics Engineers International Symposium on Biomedical Imaging (IEEE) 2015 datasets, which were not heterogeneous. Thus, applicability becomes a concern. No study to date has included a regional dataset for both lateral cephalograms and CBCTs to increase the applicability of the findings.

Thus, this study focused on the development of a CNN-based model built on EfficientNet-B0 to perform end-to-end regression of (x,y) coordinates for lateral cephalograms and the automated landmark identification (ALI) system for CBCTs (ALICBCT) using a grid search approach to optimize various hyperparameters, employing threefold cross-validation for CBCTs for complete automatic cephalometric landmark annotations from an Indian population dataset. The research focused on 23 landmark annotations from lateral cephalograms as the first part of the project, and 18 landmarks from CBCTs. The dataset was collected from multiple regions across the Indian subcontinent to enhance heterogeneity and improve the generalizability of the findings.

We compared the accuracy and precision of annotated landmarks with manual and benchmarked techniques. For future research, additional landmarks, including linear and angular measurements, can be included to facilitate comprehensive diagnosis and treatment planning in orthodontics.

Methods

The Ethics Committee of the institute approved the present study under the code ACDS/IEC/215/5/2023.

Experimental Dataset

This study focused on automatically annotating cephalometric landmarks in both lateral cephalograms and CBCT scans. The dataset was obtained from the Department of Radiology,

Secunderabad, and regional diagnostic centers in three cities in India (Mumbai, Lucknow, and Chandigarh). A statistical power of 80% at a significance level of α=0.05 using a two-sided paired t-test was assumed by the G*power software (Version 3.1, Heinrich Heine University Düsseldorf, Germany). Based on an effect size of 0.46, derived from prior studies, 405 lateral cephalograms were required (9–11). For CBCTs, calculations showed that 52 CBCT scans were required with an effect size of 0.49 according to a previous study (12–14).

All lateral cephalograms were obtained in Digital Imaging and Communications in Medicine (DICOM) format using the NewTom GO 2D/3D Vision machine (NewTom, Cefla Medical, Charlotte, NC, USA). Cone-beam computed tomography (CBCT) scans were also acquired using the same equipment. A total of 410 lateral cephalograms and 52 CBCT scans collected between 2022 and 2025 were included in the study.

Image Orientation and Landmark Annotation

Two open-source software packages, Insight Segmentation and Registration Toolkit (ITK-SNAP), Version 3.8 (http://www.itksnap.org) (15), and 3D Slicer, version 4.11 (http://www.slicer.org) (16), were utilized by experienced clinicians to orient the scans and identify anatomical landmarks.

Inclusion Criteria

For lateral cephalograms, inclusion required images taken in natural head position with clearly visible and complete anatomical structures.

For CBCT scans, the inclusion criteria included:

1. Scans acquired in an upright position using the same machine
2. Presence of all central incisors without defects or missing teeth

Exclusion Criteria

CBCT and lateral cephalogram images were excluded if the patient had:

1. A history of orthodontic treatment or orthognathic surgery
2. Cleft lip and palate or other craniofacial syndromes

Landmarks

Lateral Cephalograms

Twenty-three anatomical landmarks were annotated, comprising 19 hard tissue and 4 soft tissue points. The hard tissue landmarks included sella (S), porion (Po), basion (Ba), articulare (Ar), gonion (Go), anterior nasal spine (ANS), posterior nasal spine (PNS), point A, point B, apex of upper incisor (AUI), tip of upper incisor (TUI), tip of lower incisor (TLI), apex of lower incisor (ALI), point G (center of mandible), pogonion (Pog), gnathion (Gn), menton (Me), nasion (N), and orbitale (Or). The soft tissue landmarks were: soft tissue pogonion (Pog'), labrale superior (Ls), labrale inferior (Li), and soft tissue nose (No) (Figure 1A).

CBCTs

Eighteen anatomical landmarks were identified for analysis. These included: condylion (Co), gonion (Go), lower incisor edge (LIE), lower molar (LMol), orbitale (Or), porion (Po), apex of upper incisor (AUI), tip of upper incisor (TUI), upper molar (UMol), subspinale (A), anterior nasal spine (ANS), supramentale (B), gnathion (Gn), menton (Me), nasion (N), posterior nasal spine (PNS), and pogonion (Pog) (Figure 1B).

Overview of Deep Learning Algorithms

Lateral Cephalograms

A CNN model was developed using EfficientNet-B0 pretrained on ImageNet as the backbone for end-to-end regression of landmark coordinates. The model, referred to as Landmark Net, utilized EfficientNet-B0 as a feature extractor, with the classification head replaced by a custom regression module to predict the (x, y) coordinates of 23 anatomical landmarks (totalling 38 output values).

The modified regression head consisted of the following layers:

- Linear (1280 → 1024) → Rectified linear unit (ReLU) → Dropout (0.5)
- Linear (1024 → 512) → ReLU → Dropout (0.5)
- Linear (512 → 38)

Preprocessing

All input images were resized to 224×224 pixels using the Albumentations library. Image normalization was performed using standard ImageNet mean and standard deviation values.

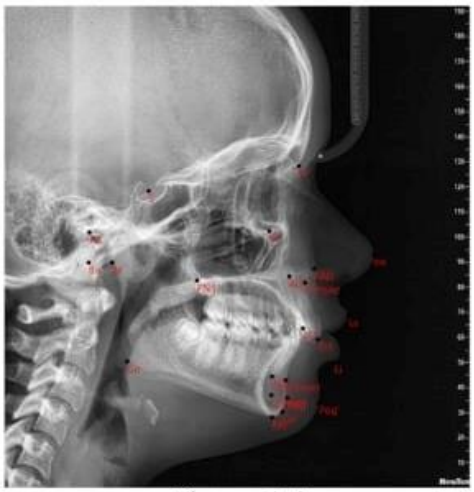


Figure 1-A

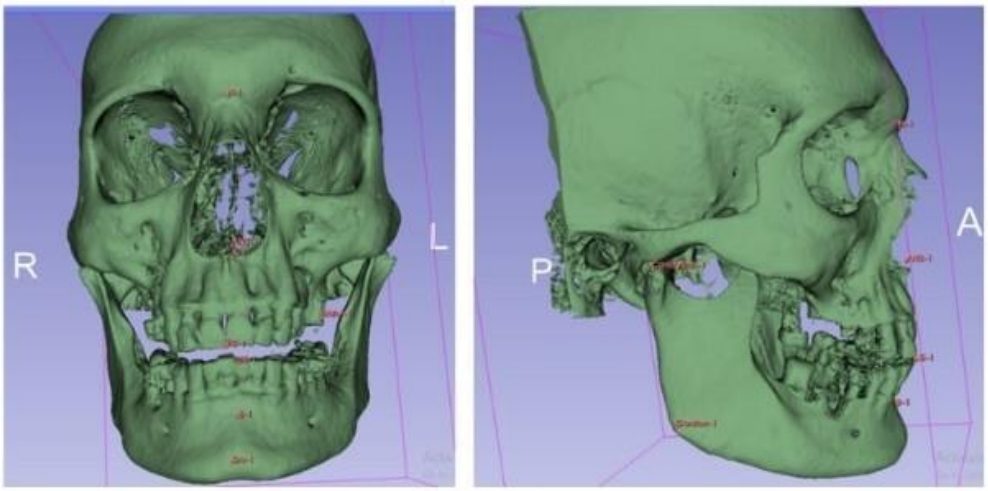


Figure 1-B

Figure 1 : Cephalometric landmarks for lateral cephalograms.
Figure 1A) Frontal view and **Figure 1B)** sagittal view presentation of landmarks on a CBCT reconstructed model.

Training

The model was trained using the following parameters:

- Loss function: mean squared error (MSE)
- Epochs: approximately 100
- Batch size: 16–32
- Regularization: Early stopping and model checkpointing were employed to prevent overfitting and retain the best-performing model.

Inference Pipelin

The inference process involved the following steps:

1. Load and convert the image to red, green, blue (RGB) format
2. Resize the image and normalize using ImageNet statistics
3. Predict the landmark coordinates using the trained model
4. Rescale the predicted coordinates back to the original image dimensions

Two-panel visualizations were generated for each image, displaying the original cephalogram on the left and the predicted landmarks on the right, with red dots and corresponding labels (Figure 2).

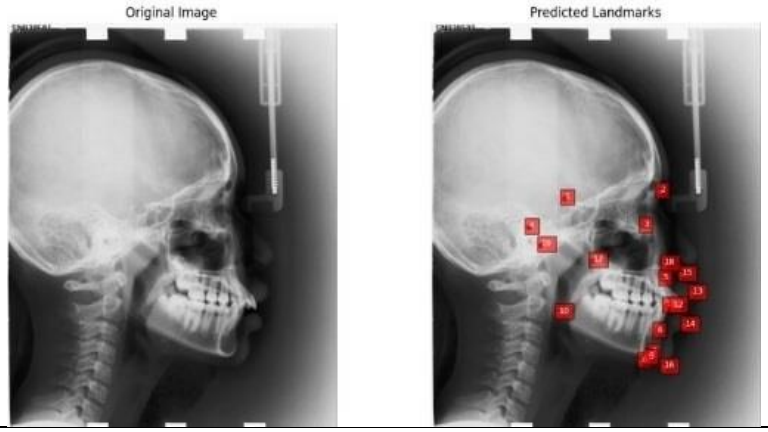


Figure 2: Original cephalogram on the left and the predicted landmarks on the right with red dots and corresponding labels.

Outputs

The output included:

- Portable network graphics (PNG) visualizations of each image with predicted landmarks
- A comma-separated values (CSV) file containing the landmark coordinates, formatted as: image name, x1, y1, ..., x19, y19

CBCTs

The Automatic Landmark Identification (ALI) system is a technology designed to automatically place anatomical landmarks on 3D CBCT images, addressing the challenges of manual landmarking, which is time-consuming, subjective, and prone to error.

ALI for CBCT (ALICBCT): We employed ALICBCT, an open-source, fully automated tool that reformulates landmark detection as a classification problem using a virtual agent inside a volumetric image. This approach leverages deep reinforcement learning (DRL) and a DenseNet-based feature network to guide the agent's navigation. Landmark prediction proceeds in a multi-scale manner (coarse-to-fine), with a verification step for final refinement. Each landmark has its own independently trained agent, enabling reproducibility and customization.

ALICBCT has been validated on up to 119 craniofacial landmarks with a mean localization error below the clinical tolerance of 2 mm and is available as an extension in the 3D Slicer platform (with open-source code and tutorials) (17).

Stratovan Corporation ALI Software: In some studies, Stratovan's Checkpoint ALI software was used. This system was trained using grid-search hyperparameter tuning and threefold cross-validation, and its accuracy was benchmarked against a ground truth established by expert judges (ICC>0.99). It reliably detects 53 landmarks (cephalometric, mandibular, TMJ, and dental) with a mean error of ~1.57 mm and is deterministic, producing identical results for identical inputs, thereby improving reproducibility (18).

By employing these validated systems, we ensured that 3D landmark detection was precise, efficient, and reproducible, while avoiding the need for developing a new 3D CNN model from scratch, an approach that would have required large proprietary datasets and significant computational resources.

The ALICBCT system successfully identified 18 anatomical landmarks across 52 CBCT scans for our study. The resulting x, y, and z coordinates of each landmark were exported to a Microsoft Excel file (Microsoft Corp., Redmond, WA) for further analysis.

Evaluation Metrics

The performance of the proposed system was assessed both qualitatively and quantitatively. Two experienced orthodontists qualitatively reviewed the landmark predictions for accuracy and clinical relevance.

Quantitative evaluation was conducted using two widely adopted metrics: success detection rate (SDR) and mean radial error (MRE). Data was collected and entered in MS Excel, and was analyzed using SPSS version 26. Descriptive and inferential statistical tests were applied as seen fit. Intergroup comparison of the variables was done using independent sample T- test and inter and intra examiner reliability was calculated using Kappa statistic.

1. Success Detection Rate (SDR)

SDR quantifies the proportion of predicted landmarks that fall within a predefined Euclidean distance threshold (e.g., 2, 3, or 4 mm) from the ground truth landmarks. It is defined as:

$$SDR=\mathbb{Z}\frac{\{i_i\|f_d(j)-f_r(j)\|<11\}}{\mathbb{Z}\Psi}x100\%$$

where:

- fd(j) and fr(j) are the predicted and actual coordinates of landmark jjj, respectively.
- H is the predefined error threshold (2, 3, or 4 mm).
- #Ψ is the total number of landmarks across all test images.

2. Mean Radial Error (MRE)

MRE represents the average Euclidean distance between predicted and ground truth landmark coordinates. The standard deviation (SD) of these distances reflects the variability of prediction errors. The equations are as follows:

$$\mathrm{MRE}=\frac{\sum_{i=1}^{n}R_i}{N_1}$$

$$\mathrm{R}=\sqrt{(\mathrm{x},-\mathrm{x}_0)^2+(\mathrm{y}_1-\mathrm{y}_0)^2}$$

$$\mathrm{SD}=\sqrt{\frac{1}{\mathrm{N}-1}\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{R_i}-\mathrm{MRE})^2}$$

where:

- Ri is the Euclidean distance between each predicted and actual landmark pair.
- (x1,y1) and (x0,y0) are the coordinates of the predicted and actual landmarks.
- N is the number of test images.

Results

Subjective Evaluation

Figure 3 illustrates the Euclidean distances (Ri) between the predicted and actual landmarks, with predicted landmarks marked in red and reference (manually annotated) landmarks in black. Visual assessment by two experienced orthodontists confirmed that the proposed algorithm achieved high accuracy in landmark localization. In many instances, the predicted landmarks closely matched or even surpassed the precision of manual annotations, the current gold standard, across both lateral cephalograms and CBCT scans.

Kappa Statistics

For intra-examiner = 0.78

For inter-examiner = 0.72

The intra- (κ = 0.78) and inter-examiner (κ = 0.72) agreements fall within the 'substantial' category, indicating that the reliability of landmark identification is adequate for clinical and research applications.

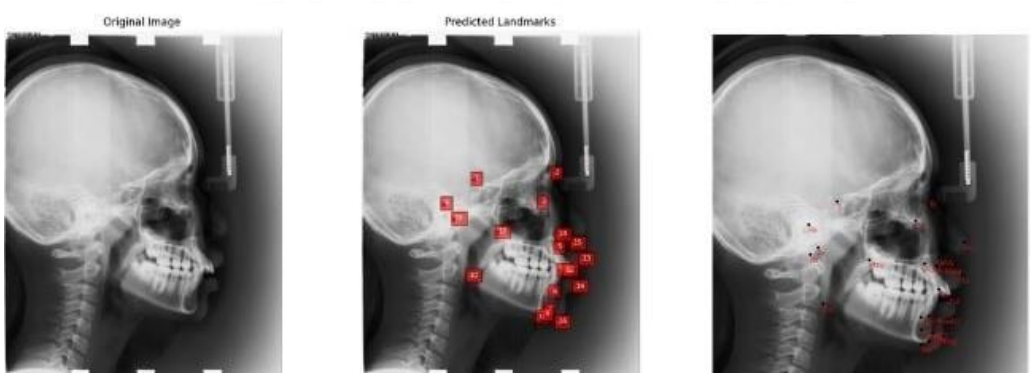


Figure 3: Original cephalogram on the left, Euclidean distances (Ri) between the predicted landmarks marked in red (middle) and reference (manually annotated) landmarks in black (right).

Quantitative Evaluation of SDR

For Lateral Cephalograms

Table 1 summarizes lateral cephalogram landmark annotations at the 2-mm threshold. The AI model demonstrated high accuracy (≥90% SDR) for several landmarks, particularly those located along the midline or in regions with high radiographic contrast, including sella (95.33%), menton (96.9%), pogonion (93.25%), and the apices of the upper and lower incisors (~93–94%). In contrast, lower accuracy was observed for landmarks such as soft tissue pogonion (75.33%), porion (75.34%), gonion (65.33%), and articulare (59.33%), likely due to anatomical superimposition, reduced clarity, and positional variability.

Accuracy improved significantly at the 3-mm threshold, with most landmarks achieving SDRs of 90% or higher. By the 4-mm threshold, nearly all landmarks exceeded 95% SDR, with several reaching 100%, including the upper and lower lips, menton, and gnathion. Notably, even complex landmarks such as porion (94.7%) and articulare (93.3%) showed strong performance, highlighting the model's robustness under practical clinical tolerance levels.

Table 1: Success detection rate (SDR) % for 23 automatically annotated landmarks for lateral cephalogram

Landmarks/ SDR (%)	2mm	3mm	4mm
Upper lip (Ls)	93	99.33	100
Lower lip (Li)	90.67	100	100
Soft tissue nose (No)	90.1	97.45	96
Soft tissue Pogonion (Pog')	75.33	93.33	98
Sella (S)	95.33	98.7	98.7
Nasion (Na)	91.26	98	96.7
Orbitale (Or)	85.67	98.78	98
Porion (Po)	75.34	84	94.7
Subspinale (A-point)	78.23	86	96.7
Supramentale (B-point)	90.1	96.7	98.7
Pogonion (Pog)	93.25	97.3	99.3
Menton (Me)	96.9	99.3	99.3
Gnathion (Gn)	95	99.3	100
Gonion (Go)	65.33	85.3	93.3
Apex of upper incisor (AUI)	90.34	92.3	98.34
Tip of upper incisor (TUI)	91.15	91.1	99
Apex of lower incisor (ALI)	93.45	93.3	99.65
Tip of lower incisor (TLI)	91.78	92.55	98.86
Posterior nasal spine (PNS)	97.67	97.33	98.7
Anterior nasal spine (ANS)	92.45	94.7	98
Articulare (Ar)	59.33	84	90
Basion (Ba)	78.67	88.1	94.67
Centre of mandible (G point)	96.45	99	100

For CBCT

Table 2 summarizes AI-based landmark detection on CBCT images, demonstrating high accuracy levels, particularly at stringent thresholds. At the 2-mm SDR level, several critical craniofacial landmarks, including nasion, anterior nasal spine (ANS), posterior nasal spine (PNS), and the tips and apices of upper and lower incisors, achieved 100% detection accuracy, highlighting the algorithm's strength in identifying well-defined skeletal structures.

Despite these successes, a few landmarks posed challenges. Orbitale (56.92%), gonion (64.5%), and supramentale/point B (68.5%) had relatively lower accuracy at the 2-mm threshold, possibly due to their anatomical complexity, curvature, and reduced radiographic contrast in CBCT scans.

At the 3-mm threshold, accuracy significantly improved: the accuracy of orbitale, gonion, and point B rose to 64.2%, 77.3%, and 83%, respectively. All remaining landmarks exceeded 94% SDR, with several maintaining 100%. By the 4-mm threshold, nearly all landmarks achieved an SDR of ≥88%, including gonion (88.4%) and orbitale (70.3%), reflecting the algorithm's robust clinical performance under realistic tolerance margins.

Quantitative Evaluation of MRE

For Lateral Cephalograms

Table 3 presents the mean radial errors (MREs) for each landmark on lateral cephalograms, comparing the proposed AI algorithm to manual annotations. The overall average MRE was 1.21±1.00 mm. Midline dental landmarks showed the greatest precision due to their well-defined anatomical boundaries. Midline skeletal points such as menton and nasion displayed moderate error, while landmarks like gonion, point A, point B, and porion exhibited higher variability. These outliers likely reflect the influence of overlapping structures and complex craniofacial anatomy, suggesting areas where future algorithmic refinement may further enhance performance.

Table 2: Success detection rate (SDR) % for 18 automatically annotated landmarks of CBCTs

Landmarks	2mm	3mm	4mm
Condylion (Co)	98	98.6	99.1
Lower incisal edge (LIE)	95.6	100	100
Lower molar (LMol)	94.4	96.6	98
Nasion (Na)	100	100	100
Orbitale (Or)	56.92	64.2	70.3
Porion (Po)	88.3	94.4	100
Subspinale (A-point)	90	92.4	93.4
Supramentale (B-point)	68.5	83	93.4
Pogonion (Pog)	90.46	98	100
Menton (Me)	95.56	98.1	100
Gnathion (Gn)	93.4	97.7	100
Gonion (Go)	64.5	77.3	88.4
Apex of upper incisor (AUI)	100	100	100
Tip of upper incisor (TUI)	100	100	100
Apex of lower incisor (ALI)	99.3	100	100
Tip of lower incisor (TLI)	99.1	100	100
Posterior nasal spine (PNS)	100	100	100
Anterior nasal spine (ANS)	95.3	96.2	98.4

Table 3 : MRE for 23 automatically annotated landmarks of lateral cephalograms

Landmarks	MRE±SD (mm)
Upper lip (Ls)	1.34±1.02
Lower lip (Li)	1.54±1.10
Soft tissue nose (No)	1.52±1.08
Soft tissue Pogonion (Pog')	1.00±0.50
Sella (S)	1.70±1.10
Nasion (Na)	1.25±1.60
Orbitale (Or)	1.30±1.60
Porion (Po)	1.50±1.80
Subspinale (A-point)	1.60±1.50
Supramentale (B-point)	1.80±1.60
Pogonion (Pog)	1.55±1.05
Menton (Me)	1.10±0.60
Gnathion (Gn)	1.30±1.20
Gonion (Go)	2.10±1.40
Apex of upper incisor (AUI)	0.50±0.30
Tip of upper incisor (TUI)	0.42±0.28
Apex of lower incisor (ALI)	0.34±0.30
Tip of lower incisor (TLI)	0.25±0.20
Posterior nasal spine (PNS)	1.40±0.92
Anterior nasal spine (ANS)	1.25±1.02
Articulare (Ar)	1.30±1.5
Basion (Ba)	1.5±1.10
Centre of mandible (G point)	0.34±0.16
Average	1.21±1.00

For CBCTs

Table 4 summarizes the MREs for CBCT-based automatic landmark identification. The average MRE across all landmarks was 1.06±0.75 mm, indicating generally high accuracy.

Dentoalveolar landmarks, including incisor tips and apices, and nasal spines, showed exceptionally low MREs (0.25–0.7 mm). Moderately higher errors (1.0–1.5 mm) were observed for menton, pogonion, point A, and point B, though still within clinically acceptable limits. Orbitale, due to its complex morphology and overlapping structures, recorded the highest MRE (2.9±2.5 mm), underscoring the ongoing challenge in the automated localization of this landmark.

Table 4 : MRE for 18 automatically annotated landmarks of CBCTs

Landmarks	MRE±SD (mm)
Condylion (Co)	1.3±1.0
Lower incisal edge (LIE)	0.6±0.9
Lower molar (LMol)	1.3±1.5
Orbitale (Or)	2.9±2.5
Nasion (Na)	0.7±0.4
Porion (Po)	1.2±0.4
Subspinale (A-point)	1.1±0.8
Supramentale (B-point)	1.4±1.1
Pogonion (Pog)	1.3±0.7
Menton (Me)	1.4±0.6
Gnathion (Gn)	1.3±0.6
Gonion (Go)	1.4±1.2
Apex of upper incisor (AUI)	0.6±0.4
Tip of upper incisor (TUI)	0.5±0.4
Apex of lower incisor (ALI)	0.6±0.3
Tip of lower incisor (TLI)	0.7±0.7
Posterior nasal spine (PNS)	0.4±0.3
Anterior nasal spine (ANS)	0.5±0.4
Average	1.06±0.75

Comparison between 2D and 3D Modalities (Paired T-test)

Table 5 presents a comparative evaluation between modalities, revealing that CBCTs consistently achieved higher SDRs at 2 mm than lateral cephalograms except at the gonion, likely due to anatomical superimposition, reduced clarity, and positional variability.

Table 6 presents a comparative evaluation between modalities, revealing that CBCTs consistently achieved higher SDRs at 3 mm than lateral cephalograms except at the porion and menton, as lateral cephalograms also achieved higher SDR values for these landmarks.

Table 7 presents a comparative evaluation between modalities, revealing that CBCT consistently achieved higher SDRs at 4 mm than lateral cephalograms except at the AUI, TUI, TLI, and PNS, as lateral cephalograms also annotate these landmarks with higher accuracy.

Table 8 presents a comparative evaluation of modalities, revealing that CBCTs consistently achieved higher MREs than lateral cephalograms except at AUI, where lateral cephalogram values showed greater precision and accuracy.

Heatmaps for Annotated Landmarks

Green: lower error, higher consistency accuracy/reliability

Yellow/Orange: moderate deviation

Red: higher error, poorer reliability

Table 5: Comparison between lateral cephalogram and CBCT for SDR for 2mm

Landmark	Modality	Sample Size	Mean (mm)	Standard Deviation	Standard Error Mean	*P- value
Orbitale	Lateral Cephalogram	52	85.7887	1.11298	0.15434	0.001
	CBCT	52	56.2638	1.21255	0.16815	
Nasion	Lateral Cephalogram	52	91.5206	0.28969	0.04017	0.001
	CBCT	52	98.7885	1.48636	0.20612	
Porion	Lateral Cephalogram	52	75.4271	0.54083	0.07500	0.001
	CBCT	52	88.1500	0.80647	0.11184	
Subspinale	Lateral Cephalogram	52	78.6773	1.39852	0.19394	0.001
	CBCT	52	89.9615	1.42748	0.19796	
Supramentale	Lateral Cephalogram	52	90.8787	1.13135	0.15689	0.001
	CBCT	52	68.3115	0.56174	0.07790	
Pogonion	Lateral Cephalogram	52	92.5035	1.42323	0.19737	0.001
	CBCT	52	90.3460	1.20781	0.16749	
Menton	Lateral Cephalogram	52	96.4960	0.52724	0.07311	0.001
	CBCT	52	95.5962	1.31762	0.18272	
Gnathion	Lateral Cephalogram	52	95.1467	0.47645	0.06607	0.001
	CBCT	52	93.6865	0.80635	0.11182	
Gonion	Lateral Cephalogram	51	64.8635	1.31889	0.18468	0.148
	CBCT	52	64.5365	0.92758	0.12863	
TUI	Lateral Cephalogram	51	90.7233	1.59600	0.22349	0.001
	CBCT	52	98.7885	1.48636	0.20612	
AUI	Lateral Cephalogram	52	90.8787	1.13135	0.15689	0.001
	CBCT	52	98.7885	1.48636	0.20612	
ALI	Lateral Cephalogram	52	92.5035	1.42323	0.19737	0.001
	CBCT	52	100.0135	1.35502	0.18791	
TLI	Lateral Cephalogram	52	91.5306	0.28949	0.04015	0.001
	CBCT	52	99.9942	1.36661	0.18951	
PNS	Lateral Cephalogram	52	96.6037	0.61583	0.08540	0.001
	CBCT	52	98.7885	1.48636	0.20612	
ANS	Lateral Cephalogram	52	92.4765	1.42977	0.19827	0.001
	CBCT	52	95.7462	0.98289	0.13630	

*** independent sample T- test**

Table 6: Comparison between lateral cephalogram and CBCT for SDR at 3mm

Landmark	Modality	Sample Size	Mean (mm)	Standard Deviation	Standard Error Mean	*P value
Orbitale	Lateral Cephalogram	52	98.2475	0.62755	0.08703	0.001
	CBCT	52	64.5365	0.92758	0.12863	
Nasion	Lateral Cephalogram	52	96.6037	0.61583	0.08540	0.001
	CBCT	52	98.7885	1.48636	0.20612	
Porion	Lateral Cephalogram	52	85.7887	1.11298	0.15497	0.349
	CBCT	52	95.7077	1.02686	0.14240	
Subspinale	Lateral Cephalogram	52	86.3592	0.52709	0.07309	0.001
	CBCT	52	93.0077	1.15959	0.16081	
Supramentale	Lateral Cephalogram	52	96.4938	0.52452	0.07274	0.001
	CBCT	52	83.4423	1.58935	0.22040	
Pogonion	Lateral Cephalogram	52	96.6037	0.61583	0.08540	0.001
	CBCT	52	98.5577	2.23531	0.30998	
Menton	Lateral Cephalogram	52	99.3338	0.61575	0.08539	0.549
	CBCT	52	99.1558	2.04723	0.28390	
Gnathion	Lateral Cephalogram	52	99.3338	0.61575	0.08539	0.001
	CBCT	52	97.5981	1.44269	0.20007	
Gonion	Lateral Cephalogram	52	85.7887	1.11298	0.15437	0.001
	CBCT	52	77.5769	1.48570	0.20603	
AUI	Lateral Cephalogram	52	92.4765	1.42977	0.19827	0.001
	CBCT	52	98.7885	1.48636	0.20612	
TUI	Lateral Cephalogram	52	92.4765	1.42977	0.19827	0.001
	CBCT	52	98.7885	1.48636	0.20612	
ALI	Lateral Cephalogram	52	92.5035	1.42323	0.19737	0.001
	CBCT	52	98.7885	1.48636	0.20612	
TLI	Lateral Cephalogram	52	92.6177	0.56248	0.07800	0.001
	CBCT	52	98.7885	1.48636	0.20612	
PNS	Lateral Cephalogram	52	96.5742	0.70080	0.09718	0.001
	CBCT	52	98.7885	1.48636	0.20612	
ANS	Lateral Cephalogram	52	94.1969	0.73799	0.10234	0.001
	CBCT	52	95.7077	1.02686	0.14240	

*** Independent sample T- test**

Table 7: Comparison between lateral cephalogram and CBCT for SDR at 4mm

Landmark	Modality	Sample Size	Mean (mm)	Standard Deviation	Standard Error Mean	*P- value
Orbitale	Lateral Cephalogram	52	97.6902	1.67587	0.23240	0.001
	CBCT	52	71.0231	1.66364	0.23071	
Nasion	Lateral Cephalogram	52	96.7490	0.44164	0.06124	0.001
	CBCT	52	98.7885	1.48636	0.20612	
Porion	Lateral Cephalogram	52	94.7650	0.45464	0.06305	0.001
	CBCT	52	98.7885	1.48636	0.20612	
Subspinale	Lateral Cephalogram	52	96.7650	0.45464	0.06305	0.001
	CBCT	52	93.5500	2.20810	0.28236	
Supramentale	Lateral Cephalogram	52	98.7419	0.42224	0.05855	0.001
	CBCT	52	93.5500	2.03610	0.28236	
Pogonion	Lateral Cephalogram	52	99.3650	0.48923	0.06305	0.009
	CBCT	52	98.7885	1.48636	0.20612	
Menton	Lateral Cephalogram	52	99.3650	0.45464	0.74738	0.001
	CBCT	52	98.7885	1.49563	0.21723	
Gnathion	Lateral Cephalogram	52	100.0650	0.43232	0.06305	0.001
	CBCT	52	98.7885	1.42929	0.23573	
Gonion	Lateral Cephalogram	52	93.3650	0.45464	0.08999	0.001
	CBCT	52	88.1692	0.81278	0.11271	
AUI	Lateral Cephalogram	52	98.4050	0.45464	0.06463	0.078
	CBCT	52	98.7885	1.48763	0.22263	
TUI	Lateral Cephalogram	52	99.0650	0.47453	0.09737	0.202
	CBCT	52	98.7885	1.41273	0.27848	
ALI	Lateral Cephalogram	52	99.7550	0.48787	0.02737	0.001
	CBCT	52	98.7885	1.42363	0.25636	
TLI	Lateral Cephalogram	52	98.9250	0.45464	0.08383	0.528
	CBCT	52	98.7885	1.45728	0.24536	
PNS	Lateral Cephalogram	52	98.7650	0.45272	0.06363	0.917
	CBCT	52	98.7885	1.47838	0.27474	
ANS	Lateral Cephalogram	52	98.0650	0.43223	0.05252	0.001
	CBCT	52	99.1846	2.03396	0.28206	

*** Independent sample T- test**

Table 8: Comparison between lateral cephalogram and CBCT for MRE

Landmark	Modality	Sample Size	Mean (mm)	Standard Deviation	Standard Error Mean	*P- value
Orbitale	Lateral Cephalogram	52	1.385	0.0998	0.0138	0.001
	CBCT	52	2.762	0.1992	0.0276	
Nasion	Lateral Cephalogram	52	1.256	0.0343	0.0048	0.001
	CBCT	52	0.694	0.1434	0.0199	
Porion	Lateral Cephalogram	52	1.472	0.0344	0.0048	0.001
	CBCT	52	1.237	0.1030	0.0143	
Subspinale	Lateral Cephalogram	52	1.200	0.1990	0.0276	0.001
	CBCT	52	1.079	0.0959	0.0133	
Supramentale	Lateral Cephalogram	52	1.754	0.0528	0.0073	0.001
	CBCT	52	1.327	0.0580	0.0080	
Pogonion	Lateral Cephalogram	52	1.462	0.0667	0.0093	0.001
	CBCT	52	1.285	0.0847	0.0117	
Menton	Lateral Cephalogram	52	1.198	0.0646	0.0090	0.001
	CBCT	52	1.377	0.0607	0.0084	
Gnathion	Lateral Cephalogram	52	1.337	0.1067	0.0148	0.001
	CBCT	52	1.300	0.0943	0.0131	
Gonion	Lateral Cephalogram	52	2.091	0.0441	0.0061	0.001
	CBCT	52	1.329	0.0940	0.0130	
Apex of upper incisor	Lateral Cephalogram	52	0.592	0.0686	0.0095	0.001
	CBCT	52	0.552	0.0548	0.0076	
Tip of upper incisor	Lateral Cephalogram	52	0.547	0.0896	0.0124	0.765
	CBCT	52	0.543	0.0570	0.0079	
Apex of lower incisor	Lateral Cephalogram	52	0.546	0.0904	0.0125	0.001
	CBCT	52	0.599	0.0733	0.0102	
Tip of lower incisor	Lateral Cephalogram	52	0.507	0.0743	0.0103	0.001
	CBCT	52	0.670	0.0898	0.0125	
PNS	Lateral Cephalogram	52	1.466	0.0761	0.0106	0.001
	CBCT	52	0.446	0.0478	0.0066	
ANS	Lateral Cephalogram	52	1.297	0.0917	0.0127	0.001
	CBCT	52	0.487	0.0943	0.0131	

*** Independent sample T- test**

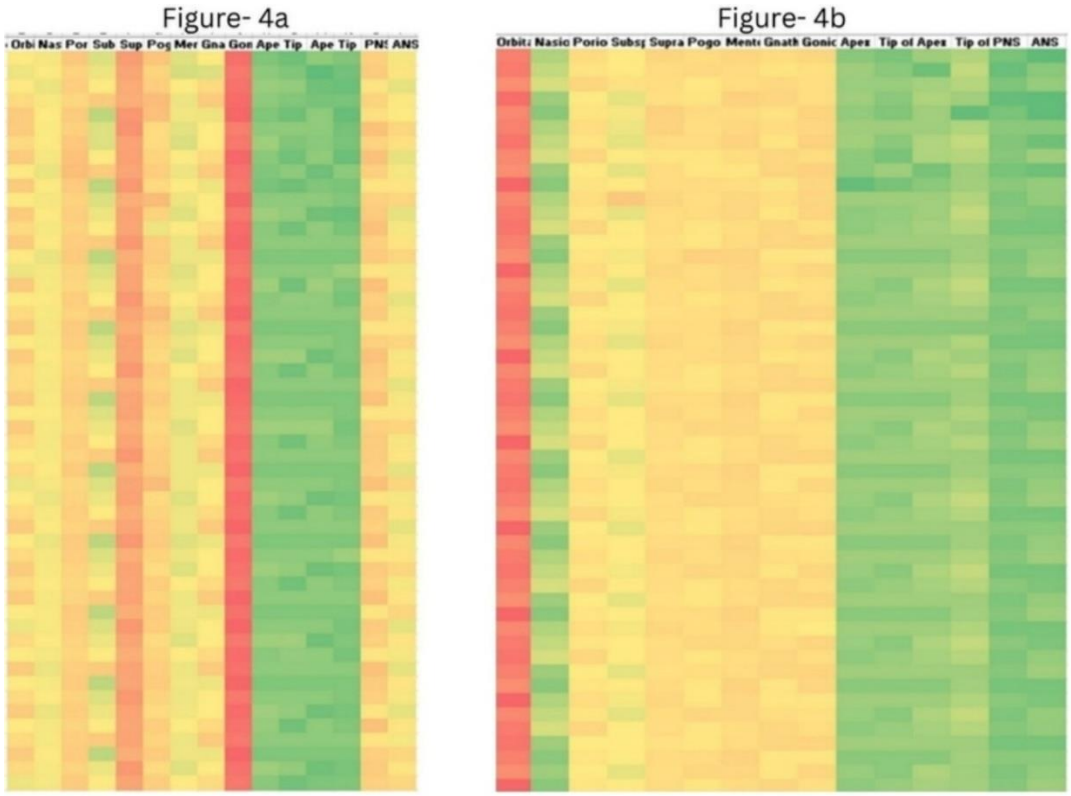


Figure 4: Figure 4a) Heatmaps for annotated landmarks for the lateral cephalogram. The heatmap demonstrates variability in landmark annotation accuracy across different cephalometric points in lateral cephalograms. ANS, PNS, and incisor tips showed higher reliability (predominantly green), whereas landmarks like orbitale, nasion, and pogonion exhibited moderate variability (yellow–orange). Gonion consistently displayed the lowest reliability (red), as it is a constructed landmark and is prone to examiner variation and AI detection errors. **Figure 4b)** Heatmaps for annotated landmarks for CBCTs. The heatmap demonstrates differential reproducibility across cephalometric landmarks in CBCTs. ANS, PNS, AUI, TUI, TLI, and ALI consistently showed high reliability (green), while midfacial skeletal points such as subspinale, supramentale, pogonion, and menton showed moderate reproducibility (yellow). In contrast, orbitale demonstrated the least consistency (red), because it is difficult to locate orbitale due to overlapping structures and anatomical variability in 3D.

Discussion

Recent advancements in artificial intelligence (AI) have significantly enhanced medical imaging applications, particularly in orthodontics and craniofacial diagnostics. This study evaluated the accuracy and precision of an AI-based automatic landmark annotation system by comparing its outputs with manually annotated reference landmarks across both lateral cephalogram and CBCT datasets. Using key performance metrics, namely the mean radial error (MRE) and success detection rate (SDR), the study provides valuable insights into the reliability and clinical viability of automated landmark detection.

For CBCT images, performance was superior to that of lateral cephalograms for most of the annotated landmarks. Certain dental and mid-facial skeletal points like AUI and menton show equal accuracy in both lateral cephalograms and CBCTs (Figure 5a). Furthermore, CBCT-based annotations demonstrated lower average MREs (0.95 mm) compared to lateral cephalograms (1.21 mm), highlighting the advantage of three-dimensional imaging in reducing landmark ambiguity and anatomical superimposition (Figure 5b).

Landmarks such as the incisor tips and apices (AUI, TUI, ALI, and TLI) demonstrated excellent precision across both modalities, which is attributable to their distinct anatomical visibility. This finding validates the AI algorithm's strong reliability in dental landmark localization regardless of imaging type.

Moreover, the findings align with the broader trend of ongoing improvements in deep learning-based image analysis, as reflected by the progressively lower error margins in recent

literature. These results support the clinical applicability of AI for landmark identification, demonstrating accuracy levels consistent with or exceeding those reported in earlier studies, and are reliable for use in clinical workflows.

While many previous studies have employed the publicly available IEEE cephalometric dataset for algorithm training and benchmarking (11,19–25), these datasets have limitations in scope and diversity. Their constrained variability may hinder generalizability across broader clinical populations (26). To address this, the present study utilized a newly compiled, diverse dataset derived from the affiliated institution and multiple diagnostic centers. This expanded dataset enhanced model robustness and relevance to the Indian population, improving its applicability across heterogeneous real-world clinical scenarios.

In the realm of AI-driven medical imaging, dataset complexity is influenced not only by technical factors like image resolution but also by biological variability, including differences in craniofacial morphology, age, sex, ethnicity, malocclusion type, and bone density (27–30). These factors significantly challenge AI performance and necessitate the development of algorithms trained on large, demographically diverse datasets. By incorporating a wide range of anatomical and demographic variability, the present study makes a meaningful contribution to the advancement of AI systems that are both accurate and generalizable across complex clinical landscapes.

Artificial intelligence (AI) applications in cephalometric analyses have seen substantial progress, with increasing adoption in orthodontic diagnostics and treatment planning. This study evaluated the performance of an AI-based automatic landmark annotation system on both lateral cephalograms and CBCT images, benchmarked against manual annotations using standard metrics including success detection rate (SDR) and mean radial error (MRE). The results underscore the clinical feasibility and technical reliability of the proposed model, especially in CBCT datasets.

These findings have direct implications for orthodontics, craniofacial diagnostics, and treatment planning. The AI model demonstrated the ability to accurately and consistently identify anatomical landmarks, especially in CBCT scans, which could substantially reduce the time and variability associated with manual landmarking. The high SDR and low MRE values suggest that the AI system performs at or above clinically acceptable standards, enabling its integration into routine diagnostic workflows.

In particular, landmarks with distinct anatomical boundaries, such as the tips and apices of incisors, showed exceptional performance across both imaging modalities, reinforcing the reliability of the model for dental landmark detection. These efficiencies can potentially enhance diagnostic consistency, streamline clinical workflows, and support evidence-based treatment planning.

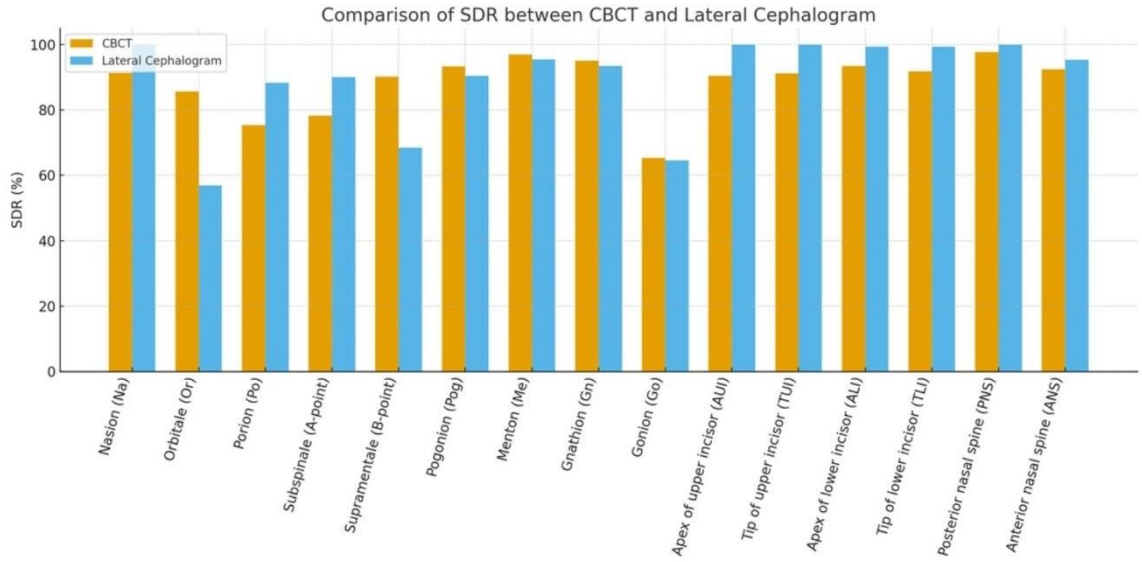


Figure 5A

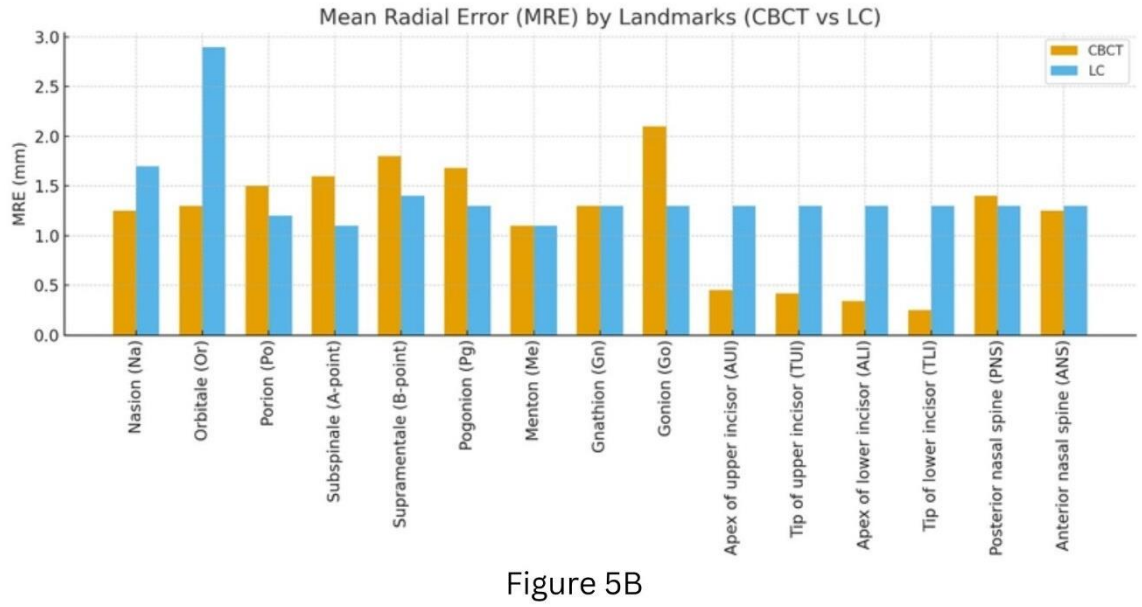


Figure 5B

Figure 5: Figure 5A) Bar chart comparing the success detection rate (SDR) at 2 mm between lateral cephalograms (LC) and CBCT for each landmark. **Figure 5B)** Bar chart comparing the MRE for AI-based landmark detection between lateral cephalograms (LC) and CBCT.

A key observation from earlier research is the inverse relationship between dataset size and model accuracy, where larger datasets do not necessarily translate to better performance unless quality and diversity are rigorously maintained. For instance, Hwang et al. trained on 1,983 annotated lateral cephalograms but reported an SDR of only 73.2% (31). In contrast, Lee et al. (9) with a smaller dataset of 150 images, achieved an SDR of 86%). This highlights the significance of data quality over mere quantity.

Our study strategically combined a high-quality dataset from the affiliated institution's radiology department and multiple diagnostic centers across the Indian subcontinent, focusing on 405 lateral cephalograms and 52 CBCT scans. This approach ensured anatomical diversity while preserving consistency in annotation. The model achieved high SDR values (88.2% for cephalograms and 94.6% for CBCT at the 2 mm threshold) and low MREs (1.21 mm and 0.95 mm, respectively), outperforming many previous studies in both 2D and 3D domains.

CBCT imaging consistently demonstrated superior accuracy and lower error across landmarks due to its inherent 3D structure, which mitigates

the common challenges of 2D imaging, such as superimposition and distortion. These findings corroborate those of Hendrickx et al., who reported a mean error range of 1.0–5.8 mm in 3D imaging and >80% accuracy in 2D methods at a 2-mm threshold (26). Mesquita et al.(32) similarly found agreement rates of 79% and 90% at 2- and 3-mm thresholds, respectively, with a mean divergence of 2.05 mm. In contrast, Serafin et al.(33) noted a pooled error of 2.44 mm (95% CI: 1.83–3.05) across studies . The present study's results, MREs of 0.95 mm (CBCT) and 1.21 mm (LC), demonstrate a clear improvement over these reported averages.

Recent advances in deep learning architectures, especially convolutional neural networks (CNNs) and spatial configuration networks (SCNs), have significantly improved landmark detection performance. State-of-the-art methods now report MREs as low as 1.038 mm and SDRs exceeding 97% at 2-mm thresholds in 2D imaging (28,34–36). In the CBCT domain, hybrid models such as HRNetV2 combined with MLPs have achieved MREs of 0.61 mm and SDRs of up to 99.93% at 4-mm thresholds (37). This outpaces older methods such as the ALI platform (MRE ~1.57 mm, SDR 45–79%) and some commercial solutions like Imagoworks Inc., which recorded MREs of 1.7±0.1 mm and SDRs of 88.16% and 94.35% at 3 mm and 4 mm, respectively (38-39).

The current model's performance with SDRs of 94.6% (CBCT) and 88.2% (LC) at 2-mm thresholds and competitive MRE values aligns with or surpasses these benchmarks, emphasizing the efficacy of using diverse, well-annotated datasets and optimized CNN architectures.

Limitations

This study has several limitations that should be acknowledged. Firstly, the dataset scope was constrained by ethical considerations and data-sharing limitations. While heterogeneous data is important for ensuring the generalizability of the model across populations, the current dataset may not fully capture this variability. To address this, multi-institutional collaborations could be pursued to enhance dataset heterogeneity and evaluate inter-center performance variations. Secondly, annotation bias represents a potential source of error. Variability within and between observers in annotating data could introduce subtle biases that affect model performance. Such variability could be minimized by supervision from a third author, helping to resolve conflicts between observers and improve annotation consistency. Thirdly, population representation in the dataset may be limited. The current sample may not fully reflect the spectrum of anatomical variation across different ages, ethnicities, and malocclusion types. Expanding the dataset to include more ethnically diverse and age-varied populations, as well as a wider range of malocclusion types, would help improve the model's applicability. Finally, the model was primarily developed and tested in a controlled environment, and its robustness in real-world clinical settings has not yet been fully evaluated. To address this, clinical validation studies are needed to measure the impact of AI-assisted landmarking on diagnostic accuracy, treatment planning, and patient outcomes.

Conclusion

This study evaluated the accuracy and precision of AI-based automatic landmark annotation in lateral cephalograms and CBCT images by comparing the outcomes with manual landmarking standards. The AI model demonstrated high reliability across both imaging modalities, achieving an average mean radial error (MRE) of 1.21 mm and a successful detection rate (SDR) of 85.8% within a 2-mm threshold for lateral cephalograms. In CBCT images, performance was notably higher, with an average MRE of 0.95 mm and an SDR of 94.4%. These results are consistent with, and in some cases surpass, those reported in earlier studies, reflecting ongoing progress in deep learning applications for medical imaging.

The clinical implications of these findings are significant. AI-assisted landmark annotation can streamline diagnostic workflows, reduce clinician workload, and improve the consistency and objectivity of cephalometric assessments. Nonetheless, for successful clinical integration, the identified limitations, particularly the need for validation in larger and more heterogeneous populations, as well as across a broader range of anatomical complexities, must be addressed.

Acknowledgements

We thank Dr. Ipshita Potlia for assistance with the statistical analysis.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available upon request.

References

1. Graber LW, Vanarsdall RL, Vig KWL, Huang GJ. Orthodontics: Current Principles and Techniques. 6th ed. Philadelphia: Elsevier; 2017. p: 257-430.
2. Proffit WR, Larson B, Fields H, Sarver DM. Contemporary Orthodontics. 6th ed. Philadelphia: Elsevier. 2018. P:279-403.
3. Nijkamp PG, Habets LL, Aartman IH, Zentner A. The influence of cephalometrics on orthodontic treatment planning. Eur J Orthod. 2008;30(6):630-5. doi: 10.1093/ejo/cjn059.
4. Hans MG, Palomo JM, Valiathan M. History of imaging in orthodontics from Broadbent to cone-beam computed tomography. Am J Orthod Dentofacial Orthop. 2015;148(4):578-86. doi:10.1016/j.ajodo.2015.06.014.
5. Kunz F, Stellzig-Eisenhauer A, Zeman F, Boldt J. Artificial Intelligence in orthodontics : evaluation of a fully automated cephalometric analysis using a customized convolutional neural network. J Orofac Orthop. 2020;81(1):52-68. doi: 10.1007/s00056-019-00203-8.
6. Arık SÖ, Ibragimov B, Xing L. Fully automated quantitative cephalometry using convolutional neural networks. J Med Imaging (Bellingham). 2017;4(1):014501. doi: 10.1117/1.JMI.4.1.014501.
7. Schwendicke F, Golla T, Dreher M, Krois J. Convolutional neural networks for dental image diagnostics: a scoping review. J Dent. 2019;91:103226. doi: 10.1016/j.jdent.2019.103226.
8. Schwendicke F, Chaurasia A, Arsiwala L, Lee JH, Elhennawy K, Jost-Brinkmann PG, et al. Deep learning for cephalometric landmark detection: systematic review and meta-analysis. Clin Oral Investig. 2021;25(7):4299-309. doi: 10.1007/s00784-021-03990-w.
9. 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(1):270. doi: 10.1186/s12903-020-01256-7.
10. Oh K, Oh IS, Le VNT, Lee DW. Deep anatomical context feature learning for cephalometric landmark Detection. IEEE J Biomed Health Inform. 2021;25(3):806-17. doi: 10.1109/JBHI.2020.3002582.
11. Zhao X, Xu T, Peng L, Li S, Zhao Y, Liu H, et al. Recognition and segmentation of teeth and mandibular nerve canals in panoramic dental X-rays by Mask RCNN. Displays. 2023;78(1):102447. doi: 10.1016/j.displa.2023.102447.
12. Kang SH, Jeon K, Kang SH, Lee SH. 3D cephalometric landmark detection by multiple stage deep reinforcement learning. Sci Rep. 2021;11(1):17509. doi: 10.1038/s41598-021-97116-7.
13. Yun HS, Hyun CM, Baek SH, Lee SH, Seo JK. A semi-supervised learning approach for automated 3D cephalometric landmark identification using computed tomography. PLoS One. 2022;17(5):e0267573. doi:10.1371/journal.pone.0267573.
14. Yun HS, Hyun CM, Baek SH, Lee SH, Seo JK. A semi-supervised learning approach for automated 3D cephalometric landmark identification using computed tomography. PLoS One. 2022;17(9):e0275114. doi: 10.1371/journal.pone.0275114.
15. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage. 2006;31(3):1116-28. doi: 10.1016/j.neuroimage.2006.01.015.
16. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magn Reson Imaging. 2012;30(9):1323-41. doi: 10.1016/j.mri.2012.05.001.
17. Gillot M, Miranda F, Baquero B, Ruellas A, Gurgel M, Al Turkestani N, et al. Automatic landmark identification in cone-beam computed tomography. Orthod Craniofac Res. 2023;26(4):560-567. doi: 10.1111/ocr.12642.
18. Ghowsi A, Hatcher D, Suh H, Wile D, Castro W, Krueger J, et al. Automated landmark identification on cone-beam computed tomography: Accuracy and reliability. Angle Orthod. 2022;92(5):642-54. doi: 10.2319/122121-928.1.
19. Zhong Z, Li J, Zhang Z, Jiao Z, Gao X. An attention-guided deep regression model for landmark detection in cephalograms. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, Yap P-T, Khan A, editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI. Cham: Springer International Publishing; 2019. p. 540-548. doi:10.1007/978-3-030-32226-7_60.
20. Song Y, Qiao X, Iwamoto Y, Chen YW, Chen Y. An efficient deep learning based coarse-to-fine cephalometric landmark detection method. IEICE Trans Inf Syst. 2021;E104.D(8):1359–66. doi:10.1587/transinf.2021EDP7001.
21. Zeng M, Yan Z, Liu S, Zhou Y, Qiu L. Cascaded convolutional networks for automatic cephalometric landmark detection. Med Image Anal. 2021;68:101904. doi: 10.1016/j.media.2020.101904.
22. Chen CC, Wu YF, Aung LM, Lin JC, Ngo ST, Su JN, et al. Automatic recognition of teeth and periodontal bone loss measurement in digital radiographs using deep-learning Artificial Intelligence. J Dent Sci. 2023;18(3):1301-09. doi: 10.1016/j.jds.2023.03.020.
23. Qian J, Cheng M, Tao Y, Lin J, Lin H. CephaNet: An improved faster R-CNN for cephalometric landmark detection. In: Proceedings of the 16th International

Symposium on Biomedical Imaging (ISBI); 2019 Apr 8-11; Venice, Italy. IEEE; 2019. p. 868-871. doi:10.1109/ISBI.2019.8707446.
24. Noothout JMH, De Vos BD, Wolterink JM, Postma EM, Smeets PAM, Takx RAP, et al. Deep learning-based regression and classification for automatic landmark localization in medical images. IEEE Trans Med Imaging. 2020;39(12):4011-22. doi: 10.1109/TMI.2020.3009002.
25. Kwon HJ, Koo HI, Park J, Cho NI. Multistage probabilistic approach for the localization of cephalometric landmarks. IEEE Access. 2021;9:21306–14. doi: 10.1109/ACCESS.2021.3052460.
26. Hendrickx J, Gracea RS, Vanheers M, Winderickx N, Preda F, Shujaat S, et al. Can Artificial Intelligence-driven cephalometric analysis replace manual tracing? A systematic review and meta-analysis. Eur J Orthod. 2024;46(4):cjae029. doi: 10.1093/ejo/cjae029.
27. Kim MJ, Liu Y, Oh SH, Ahn HW, Kim SH, Nelson G. Automatic cephalometric landmark identification system based on the multi-stage convolutional neural networks with CBCT combination images. Sensors (Basel). 2021;21(2):505. doi: 10.3390/s21020505.
28. Wang S, Li H, Li J, Zhang Y, Zou B. Automatic analysis of lateral cephalograms based on multiresolution decision tree regression voting. J Healthc Eng. 2018;2018:1797502. doi: 10.1155/2018/1797502.
29. Darkwah WK, Kadri A, Adormaa BB, Aidoo G. Cephalometric study of the relationship between facial morphology and ethnicity: review article. Transl Res. 2018;12:20–4. doi: 10.1016/j.tria.2018.07.001.
30. Tanikawa C, Lee C, Lim J, Oka A, Yamashiro T. Clinical applicability of automated cephalometric landmark identification: part I-patient-related identification errors. Orthod Craniofac Res. 2021;24 Suppl 2:43-52. doi: 10.1111/ocr.12501.
31. 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(1):69-76. doi: 10.2319/022019-129.1.
32. de Queiroz Tavares Borges Mesquita G, Vieira WA, Vidigal MTC, Travençolo BAN, Beaini TL, Spin-Neto R, et al. Artificial Intelligence for detecting cephalometric landmarks: a systematic review and meta-analysis. J Digit Imaging. 2023;36(3):1158-79. doi: 10.1007/s10278-022-00766-w.
33. Serafin M, Baldini B, Cabitza F, Carrafiello G, Baselli G, Del Fabbro M, et al. Accuracy of automated 3D cephalometric landmarks by deep learning algorithms: systematic review and meta-analysis. Radiol Med. 2023;128(5):544-55. doi: 10.1007/s11547-023-01629-2.
34. Šavc M, Sedej G, Potočnik B. Cephalometric landmark detection in lateral skull X-ray images by using improved spatial configuration-net. Appl Sci. 2022;12(9):4644. doi:10.3390/app12094644.
35. Le VNT, Kang J, Oh IS, Kim JG, Yang YM, Lee DW. Effectiveness of Human-Artificial Intelligence Collaboration in Cephalometric Landmark Detection. J Pers Med. 2022;12(3):387. doi: 10.3390/jpm12030387.
36. Yao J, Zeng W, He T, Zhou S, Zhang Y, Guo J, et al. Automatic localization of cephalometric landmarks based on convolutional neural network. Am J Orthod Dentofacial Orthop. 2022;161(3):e250-e259. doi: 10.1016/j.ajodo.2021.09.012.
37. Takahashi K, Shimamura Y, Tachiki C, Nishii Y, Hagiwara M. Cephalometric landmark detection without X-rays combining coordinate regression and heatmap regression. Sci Rep. 2023;13(1):20011. doi: 10.1038/s41598-023-46919-x.
38. Ghowsi A, Hatcher D, Suh H, Wile D, Castro W, Krueger J, et al. Automated landmark identification on cone-beam computed tomography: accuracy and reliability. Angle Orthod. 2022;92(5):642-54. doi: 10.2319/122121-928.1.
39. Park J, Yoon S, Kim H, Kim Y, Lee U, Yu H. Clinical validity and precision of deep learning-based cone-beam computed tomography automatic landmarking algorithm. Imaging Sci Dent. 2024;54(3):240-50. doi: 10.5624/isd.20240009.