Original Article

Application of artificial intelligence in lateral cephalometric analysis

Huang-Ting Lee [a], Po-Yuan Chiu [a,b], Chen-Wen Yen [c], Szu-Ting Chou [a,b], Yu-Chuan Tseng [a,b*]

[a] School of Dentistry, College of Dental Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan
[b] Department of Orthodontics, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan
[c] Department of Mechanical and Electromechanical Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan

Introduction

Convolutional neural network (CNN), a deep learning model, has various applications, including image classification,[1—3] image segmentation,[4,5] natural language processing,[6] facial landmark detection,[7,8] and lane detection.[9] In medicine, CNNs have been used for diagnosing various cancers[10] and detection of early-stage Alzheimer's disease.[11] However, unlike the widespread application of artificial intelligence (AI) in medicine,[12—14] its application in dentistry remains limited.

Since its introduction in 1931 by Broadbent, lateral cephalometric analysis has remained the main diagnostic procedure in orthodontics; this procedure plays a pivotal role in treatment planning.[15] It involves the identification of anatomical anchor points on X-ray images, followed by the measurement of various distances, angles, and ratios for the clarification of the craniofacial structures. As the skull is a highly complicated three-dimensional structure projected onto a single two-dimensional plane in a cephalogram, the images always exhibit overlapping features. Moreover, facial asymmetry, anatomical variations across individuals, and variations in head positioning during image acquisition make it challenging to identify landmarks on lateral cephalograms. This task is performed manually, is time consuming, and causes considerable interclinician[16] and intraclinician variations.[17,18] Given that inaccurate cephalometric analysis may lead to critical consequences, a self-adapting algorithm that can accurately detect cephalometric landmarks is desired.

In 2017, a US study was conducted using a CNN model.[19] In this study, two clinicians annotated 19 landmarks on approximately 400 image samples.[20] The present study was also conducted using a CNN model but improved it with U-net and MobileNetV2 to reduce huge portion of calculations. Compare to the previous study,[19,20] we trained our model with much more sample and evaluated our model with additional dental analysis.

The aim of this study was to develop a unique and excellent artificial intelligence for use in orthodontic lateral cephalometric imaging. We hope that larger sample sizes, more accurate manual annotations, and better algorithms will contribute to its enhancement.

Materials and methods

Ethics

This study was approved by the Institutional Review Board of the University Hospital (IRB Number: KMUHIRB-E (II)-20210,143). The requirement for informed consent from

* Corresponding author. School of Dentistry, College of Dental Medicine, Kaohsiung Medical University, 100 Shih-Chuan 1st Road, Kaohsiung 80708, Taiwan.
E-mail addresses: tsengyc@kmu.edu.tw, yct79d@gmail.com (Y.-C. Tseng).

the participants was waived off by the ethics committee due to the retrospective nature of the study."

Data set

This retrospective study included data of 1002 patients (687 women and 315 men; mean age: 29.9 years [range: 18−76 years]). We included patients aged >18 years whose clear lateral cephalometric images were available and excluded those with damaged images, vague images, images of dental implants or other treated cases that could affect machine learning.

The images were initially captured in the JPEG format (resolution: 2600 pixels × 2304 pixels). Two physicians annotated the images using 14 landmarks (Fig. 1). The reason we selected these 14 points <u>was</u> because they encompass the commonly used anterior-posterior, vertical, and dental relationships in lateral cephalometric analysis. Inter- and intra-class correlation analyses between physicians showed coefficients exceeding 0.8, indicating a high level of qualification among the involved physicians. Each landmark was defined in detail (Table 1) and was subsequently used to measure 12 angles, distances, and ratios (Table 2) that are commonly used for orthodontic diagnosis.

Image grouping

We divided the samples into training datasets (700 images) and testing datasets (302 images). The entire training dataset was used to train Model 1 (700 patients). Subsequently, 91 images that could have affected machine learning, such as images of patients with excessive dental restorations, were excluded. The remaining dataset (609 patients) was then used as the training set for Model 2.
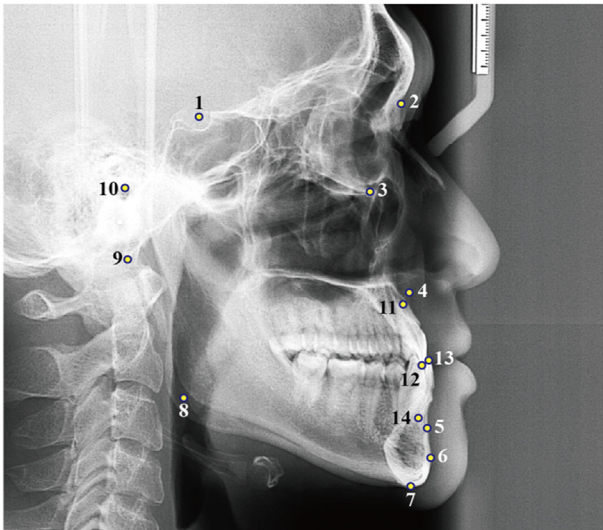
Figure 1 Labeled cephalogram example showing the 14 landmark positions used in this study. (Reference: Miyashita K. Contemporary cephalometric radiography. In Dixon AD, editor. Glossary of cephalometric terms and definitions. Tokyo: Quintessence publishing Co, Inc; 1996, p. 246−56).

Model 2 entailed a more stringent selection of images than Model 1 to assess whether stricter image selection can yield better results.

Technical aspects

Deep-learning models require considerable parallel and repetitive computational power.[21] CNNs require a larger display card memory capacity, which necessitates the use of suitable display cards.[22] The computer specifications used by the Department of Mechanical and Electrical Engineering at Sun Yat-Sen University, with whom we collaborated, were as follows: OS, Windows; CPU, Intel Core i9-10940X; RAM, DDR4-3200 128G; GPU, NVIDIA RTX 2080ti; and VRAM, 11 GB. Based on a previous study,[20] we set a 2-mm radius as the acceptable clinical range for landmark positions.

Convolutional neural network (CNN) architecture

CNNs comprise an input layer, multiple hidden layers, and an output layer. The input layer included radiographic images marked with 14 landmarks in ImageJ and saved as grayscale (256 bits) TIFF files. Each pixel corresponded to a distinct grayscale value, representing individual data points in the input layer.

The images were output to the hidden layers, which are the most crucial parts of a CNN. The images were zero-padded from 2600 × 2304 to 2700 × 2700 pixels to make them compatible with machine learning. The images were then resized to 300 × 300 and 900 × 900 pixels (Fig. 2). Images with a resolution of 300 × 300 pixels were input into the global model, which used the U-Net architecture for image segmentation (Fig. 3).[5] The U-Net comprises an encoder and a decoder. Using the encoder part, we performed convolution, activation, and max pooling; using the decoder part, we performed convolution, activation, and deconvolution. Convolution is a set of kernels that are learnable and self-adapting filters. A kernel moves by a certain number of pixels, referred to as strides, performing an inner product with another area of an image and subsequently scanning the entire image to output a smaller matrix. For this study, we selected rectified linear units, a function that passes all positive values without any transformation but sets all negative values to 0. The resulting matrix was input into the max-pooling layer, where each 2 × 2 pixel block retained only its maximum value. Next, the matrix, which was now half the dimension of the input, was scanned again with the convolution layer to form a loop. Finally, the data were input into the decoder, which first processed the deconvolution and provided the final output in dimensions of 300 × 300 × 14 layers.

Most recent studies on keypoint prediction have been conducted through heatmap regression because of its suitability for image generalization.[23] Coordinate regression can predict keypoints by using the mean square error loss function, which does not consider peripheral points. In contrast, heatmap regression performs predictions using a Gaussian distribution (Table 3; Equation 1). Every pixel in the image has a predicted heatmap value ranging from 0 to 1; a shorter distance between the pixel and the actual point

Table 1 List of anatomical landmarks and their definitions.

Landmark number	Anatomical name	Definition
L1	Sella (S)	The center of sella turcica
L2	Nasion (N)	The suture between the frontal and nasal bones
L3	Orbitale (Or)	Lowest point on the lower margin of each orbit
L4	A-point (A)	Deepest point on midsagittal plane between ANS and prosthion. Usually around the level of, and anterior to, the apex of the upper central incisors
L5	B-point (B)	The most posterior point of bony curvature of the mandible below infradentale and above Pogonion
L6	Pogonion (Pog)	The most anterior point of the contour of the chin
L7	Menton (Me)	Lowermost point of the contour of the mandibular symphysis
L8	Gonion (Go)	A point on the curvature of the angle of the mandible located by bisecting the angle formed by lines tangent to the posterior ramus and the inferior border of the mandible (constructed landmark)
L9	Basion (Ba)	The most inferior posterior point in the sagittal plane on the anterior rim of the foramen magnum
L10	Porion (Po)	Highest point on the upper margin of the cutaneous auditory meatus
L11	Upper incisor apex (UIA)	The root apex of the most prominent upper incisor
L12	Upper incisor Edge (UIE)	Mid-point of the incisal edge of the most prominent upper central incisor
L13	Lower incisor Edge (LIE)	The incisal point of the most prominent lower central incisor
L14	Lower incisor apex (LIA)	The root apex of the most prominent lower incisor

Table 2 List of measurements and their definitions.

Number of measurements	Name of measurements	Definition
M1	∠ SNA	Angle between S (L1), N (L2) and A (L4)
M2	∠ SNB	Angle between S (L1), N (L2) and B (L5)
M3	∠ ANB	Angle between A (L4), N (L2) and B (L5)
M4	A-Nv	Distance from A (L4) to N (L2)-vertical line
M5	B-Nv	Distance from B (L5) to N (L2)-vertical line
M6	Pog-Nv	Distance from Pog (L6) to N (L2)-vertical line
M7	∠ SN-FH	Angle between the lines from S (L1) to N (L2) and from Po(L10) to Or (L3)
M8	∠ SN-GoMe	Angle between the lines from S (L1) to N (L2) and from Go (L8) to Me(L7)
M9	∠ U1-SN	Angle between the lines from UIE (L12) to UIA (L11) and from S (L1) to N (L2)
M10	∠ L1-MP	Angle between the lines from LIE (L13) to LIA (L14) and from Go (L8) to Me(L7)
M11	U1-APog	Shortest distance between the lines from A (L4) to Pog (L6) and UIE (L12)
M12	L1-APog	Shortest distance between the lines from A (L4) to Pog (L6) and LIE (L13)

indicates a heatmap value closer to 1, and vice versa. Moreover, a smaller standard deviation indicates a more centralized Gaussian distribution and vice versa. Our results revealed that the model performed better when the standard deviation was set at 5. Next, the heat map values were set at 255 to convert them from 0—1 to 0—255. The output of the global model, which contained AI-predicted points, was generated using Gaussian heatmap regression. Next, these two heat maps were combined into the mean square error loss function (Table 3; Equation 2) to configure the learning rate of the adapting parameters in the multistep mode. After approximately 300 epochs of adaptation, we regarded the coordinates of the predicted point as the center and captured the peripheral area with dimensions of 128 × 128 pixels. As we considered 14 AI-predicted landmarks, we input data with dimensions of 128 × 128 × 14 into the local model.

Unlike the global model, the local model replaced the encoder part of U-Net with MobileNetV2 (Figs. 4 and 5).[24] The local model produced outputs in the dimensions of 128 × 128 × 14, which were subsequently used for heatmap prediction. These data were also input into the loss function, and the parameters of the local model were updated. After approximately 100 adaptation epochs, the coordinates of the AI-predicted points were generated.

Statistical analysis

In this study, we evaluated the performance and clinical acceptability of CNN models, specifically AI-predicted points, and compared them with the performance of manually labeled points using three approaches: (1) average absolute distance, (2) average absolute errors
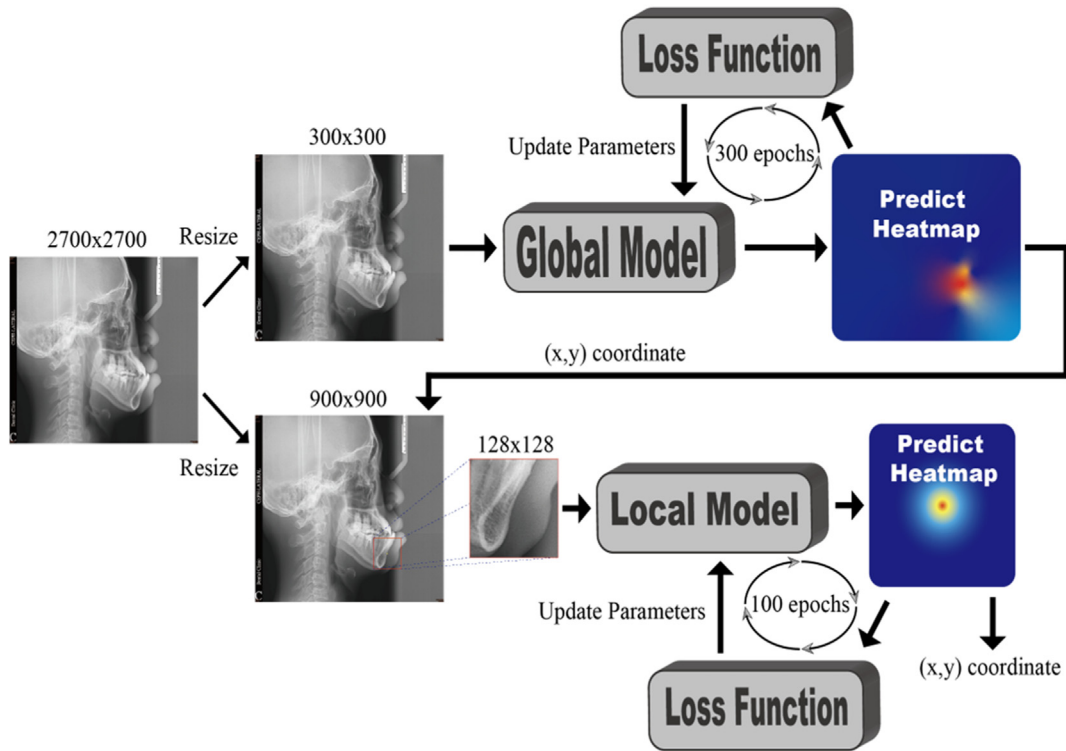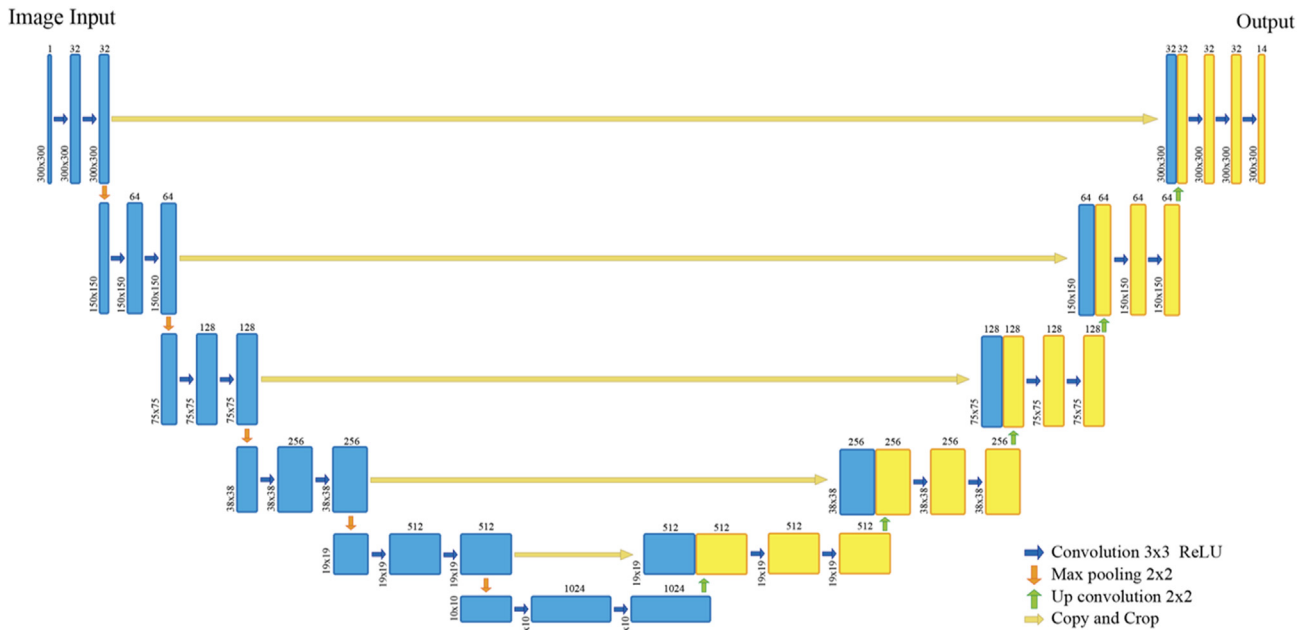
Figure 2 Architecture of the AI.

Figure 3 Global model.

between the values of the 12 parameters generated, and (3) confusion matrices of classifications.

Results

The results of the following three approaches are showed in the table: (1) average absolute distance (Table 4), (2)

average absolute errors between the values of the 12 parameters generated (Table 5), and (3) confusion matrices of classifications (Table 6).

Based on the previous study,[19] the desired estimation accuracy was set at <2 mm. The point-to-point error for landmark ℓ ($\mathscr{PE}_\ell$) is the average absolute distance between the AI-predicted points and manually labeled points (Table 3; Equation 3). Table 4 presents a comparison of

Table 3 Point-to-point error of Model 1, 2 and SDR of Model 2.

Landmark	Model 1 $\mathscr{PE}$ (mm)	Model 2 $\mathscr{PE}$ (mm)	Model 2 SDR (%)			
			2.0 mm	2.5 mm	3.0 mm	4.0 mm
L1 (Sella)	1.87	1.2	92.38	97.35	98.01	99.01
L2 (Nasion)	1.41	0.97	90.07	94.37	97.02	99.34
L3 (Orbitale)	1.26	1.09	90.4	95.36	97.68	99.34
L4 (A-point)	1.37	1.57	84.77	92.38	96.36	99.01
L5 (B-point)	2.17	1.28	80.79	87.75	92.72	97.68
L6 (Pogonion)	1.66	0.9	91.72	96.03	98.01	99.67
L7 (Menton)	0.96	0.81	99.34	99.67	99.67	99.67
L8 (Gonion)	2.21	1.77	65.89	75.5	85.1	95.36
L9 (Basion)	1.66	1.52	76.16	86.42	93.38	97.68
L10 (Porion)	1.89	1.73	78.48	87.75	91.72	97.68
L11 (Upper incisor apex)	2.32	2.35	52.65	63.91	78.48	93.71
L12 (Upper incisor Edge)	0.91	0.65	96.36	97.68	98.01	99.01
L13 (Lower incisor Edge)	0.95	0.89	91.06	95.36	97.02	99.34
L14 (Lower incisor apex)	1.93	1.58	73.84	85.1	92.38	98.68
Average	1.61	1.31	83.14	89.62	93.97	98.23

Model 1: A model contains training data composed of 700 patients.
Model 2: A model contains training data composed of 609 patients, which excluded images that could have affected machine learning, such as images of patients with excessive dental restorations.
Model 2 SDR: The successful detection rate of Model 2.
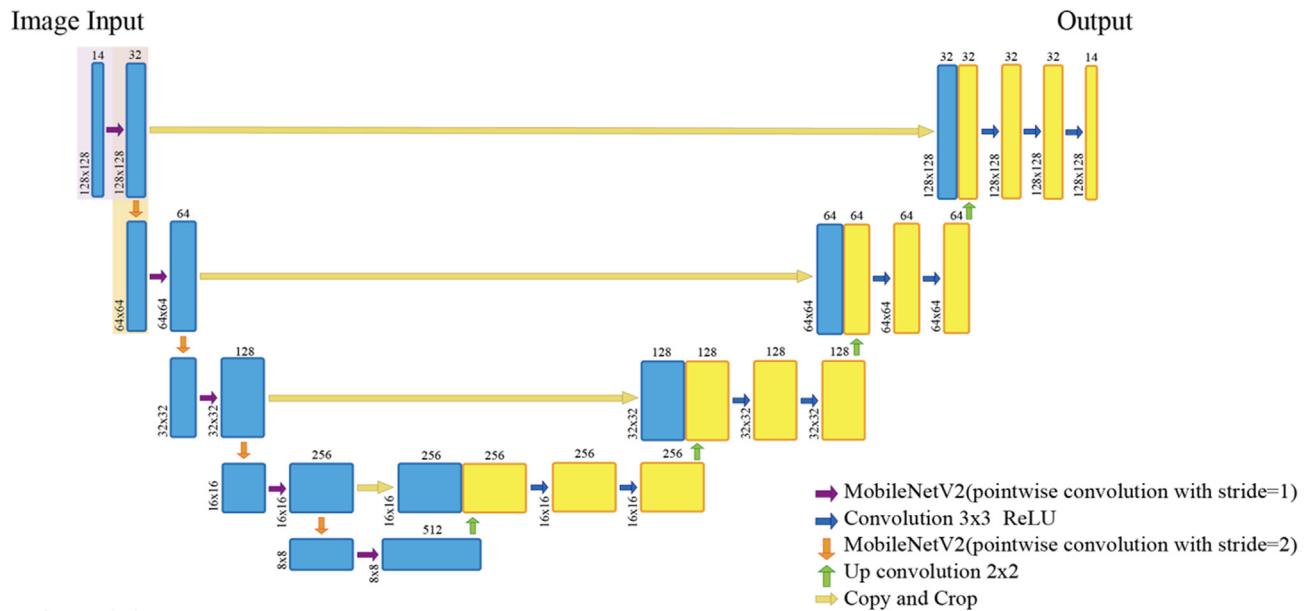
Figure 4 Lobal model.

Models 1 and 2. Model 2, which included the data of only those patients who had not undergone any treatment, significantly outperformed Model 1. We also calculated the successful detection rate (SDR) of Model 2, which represents the percentage of AI-predicted points within the precision range.

$$x \in \mathscr{PE}\{2.0\,mm, 2.5\,mm, 3.0\,mm, 4.0\,mm\}: SDR_{\ell} = \#\{\|\mathscr{A}_{\ell_i} - m_{\ell_i}\| \leq x\}/n \times 100.$$

Except for the upper incisor apex, the errors for all landmarks were within 2 mm, and the average SDRs for

distances of 2.0, 2.5, 3.0, and 4.0 mm were 83.14 %, 89.62 %, 93.97 %, and 98.23 %, respectively.

The standard classification (Table 5) was based on other studies.[25,26] The results are summarized in Table 6. The diagonal averages of the ANB, SN-GoMe, U1-SN, and L1-MP were 84.44 %, 90.73 %, 80.79 %, and 80.79 %, respectively.

Discussion

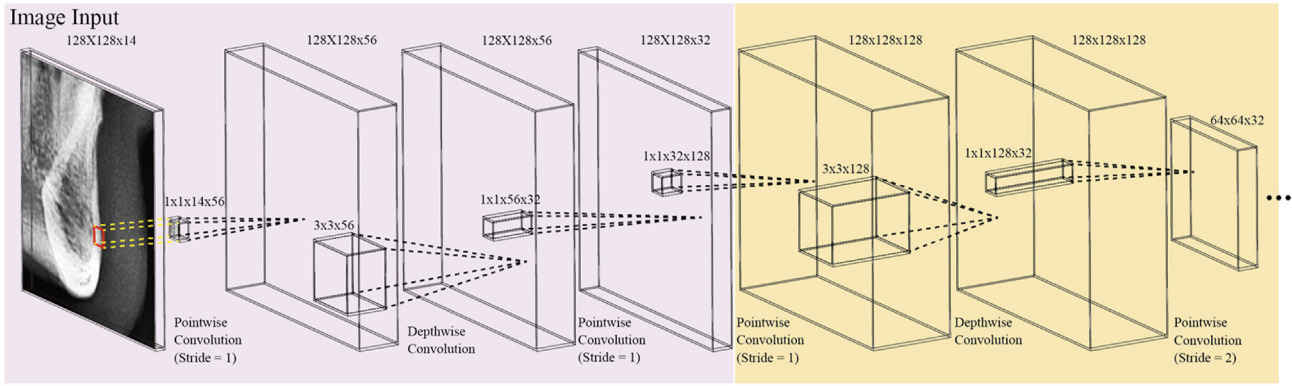We selected 14 lateral cephalometric landmarks that are commonly used as diagnostic landmarks in orthodontics.

Figure 5 MobileNetV2.

Table 4 Absolute mean errors of 12 parameters of clinical measurements and clinical standard deviation of adult male and female.

Measurements	Unit	Model 2	SD (male)	SD (female)
∠ SNA	°	1.73	3.7	3.2
∠ SNB	°	1.13	3.5	3.5
∠ ANB	°	1.09	2.1	2.2
A-Nv	mm	1.69	3.5	4.3
B-Nv	mm	2.33	4.5	6.9
Pog-Nv	mm	2.36	5.1	7.5
∠ SN-FH	°	1.9	3	2.5
∠ SN-GoMe	°	1.63	5.2	5.6
∠ U1-SN	°	3.36	5.6	8.0
∠ L1-MP	°	3.25	6.9	7.0
U1-APog	mm	0.66	2.6	2.2
L1-APog	mm	0.75	2.4	2.2

Model 2: A model contains training data composed of 609 patients, which excluded images that could have affected machine learning, such as images of patients with excessive dental restorations.

Some studies have stated that AI-based marking has poor accuracy than that of manual marking. In particular, AI-based detection of the S point (Sella, i.e., L1 in this study) is relatively poor, possibly because the S point is a virtual point and the geometric center of the sella turcica.[27] However, our CNN model accurately detected S (L1), with a point-to-point error of only 1.2 mm (Model 2) and an SDR within 2 mm of 92.38 %. This discrepancy may be attributed to the differences in the marking methods used for the S point, which affect the initial accuracy, and the differences in the detection performance of the AI model for the virtual points.

Another landmark where AI performed poorly was the gonion (L8 in this study), possibly because it is a construction point that requires more complex considerations during the machine-learning process using bilateral mandibular planes as a reference. Consistent with the models used in other studies,[19,20] our model performed relatively poorly in gonion detection. The point-to-point error was 1.77 mm (Model 2), which was higher than the average value of the 14 landmarks considered in this study (1.31 mm). Other landmarks with relatively poor performances included A (L4), B (L5), and Po (L10), with A (L4) having point-to-point errors as high as 1.57 mm (Model 2). The SDRs within 2 mm for B and Po were 80.79 % and 78.48 %, respectively, both of which were lower than the average value of 83.14 % in this study. Certain points that are difficult to predict include bilateral structures such as Po (L10), which are difficult to judge, and points defined on broad curves, such as A (L4) and B (L5), which are prone to initial errors during manual marking.[28]

In this study, the performance of the AI models in detecting dental landmarks was slightly inferior to that when detecting skeletal landmarks, particularly for the apices of upper (L11) and lower (L14) incisors where the SDRs within 2 mm were 52.65 % and 73.84 %, respectively, both of which were lower than 83.14 % (Model 2). This may be because of the blurry images (obtained using the lateral cephalometric machine in this study) of the root tips of the

Table 5 Four clinical measurements and their standard of classifications.

∠ ANB		∠ SN-GoMe	
Class I (C1)	0°−4°	Hypo-divergent (C1)	<28°
Class II (C2)	>4°	Normo-divergent (C2)	28°−38°
Class III (C3)	<0°	Hyper-divergent (C3)	>38°

∠ U1-SN		∠ L1-MP	
U1 proclination (C1)	>114.78°	L1 proclination (C1)	>103.98°
Neutral (C2)	102.8°−114.78°	Neutral (C2)	91.12°−103.98°
U1 retroclination (C3)	<102.8°	L1 retroclination (C3)	<91.12°

C1: Classification 1, C2: Classification 2, C3: Classification 3.

Table 6 Confusion matrices of classifications based on AI-predicted points and manual-labeled points.

Parameter	Confusion matrix (%)				Diagonal average (proposed)	Diagonal average
∠ ANB		C1 (Manual)	C2 (Manual)	C3 (Manual)	84.44 %	79.90 %
	C1 (AI)	26.5	5.0	3.3		
	C2 (AI)	4.0	41.1	0.0		
	C3 (AI)	3.3	0.0	16.9		
∠ SN-GoMe		C1 (Manual)	C2 (Manual)	C3 (Manual)	90.73 %	81.92 %
	C1 (AI)	13.2	1.0	0.0		
	C2 (AI)	1.3	47.0	2.3		
	C3 (AI)	0.0	4.6	30.5		
∠ U1-SN		C1 (Manual)	C2 (Manual)	C3 (Manual)	80.79 %	x
	C1 (AI)	10.9	4.0	0.3		
	C2 (AI)	3.0	35.8	9.9		
	C3 (AI)	0.0	2.0	34.1		
∠ L1-MP		C1 (Manual)	C2 (Manual)	C3 (Manual)	80.79 %	x
	C1 (AI)	6.3	4.6	0.3		
	C2 (AI)	4.6	39.4	7.3		
	C3 (AI)	0.0	2.3	35.1		

C1: Classification 1, C2: Classification 2, C3: Classification 3.

maxillary and mandibular anterior teeth. Moreover, the root tips of the lower central incisors may be easily confused with those of the lateral incisors or canines in lateral cephalometric images, particularly in patients with mandibular anterior crowding.[28]

For most landmarks, such as N (L2), Or (L3), Pog (L6), Me (L7), Go (L8), and UIE (L12), the average error was lower in our study than that reported previously.[20] The SDRs for distances within 2, 2.5, 3, and 4 mm were higher in most ranges than in those noted in other studies.[20]

The error of initial landmark placement can influence the subsequent analysis of lateral cephalogram.[29] Therefore, we used AI to calculate commonly used clinical data, including angle values such as ∠SNA, ∠SNB, and ∠ANB in Steiner analysis, as well as linear distances such as A-Nv and B-Nv in the McNamara analysis. We observed that the errors were within the acceptable ranges. Our findings indicate that all landmark placements and measurement data can be derived and developed for clinical diagnosis. We used confusion matrices to present ∠ANB, ∠SN-GoMe, ∠U1-SN, and ∠L1-MP values as diagnostic criteria and evaluated whether these diagnoses differed from human judgments. ∠ANB is a measure of the skeletal horizontal relationship, whereas ∠SN-GoMe is a measure of the skeletal vertical relationship. ∠U1-SN and ∠L1-MP are diagnostic measurements of the angles of the upper and lower anterior teeth, respectively. The results revealed that the success rates for all four diagnoses were >80 %, reaching as high as 90.73 % for the skeletal vertical relation (∠SN-GoMe). This indicates that the proposed AI framework can be successfully applied to various types of skeletal and dental analyses and diagnoses.

In contrast, the diagonal averages of the previous study, which improved the system and led to it winning first prize at the 2015 ISBI Grand Challenge in Dental X-ray Image Analysis, were 79.90 % and 81.92 % for ANB and SN-GoMe, respectively.[20]

This study has some limitations. As mentioned in a systematic review,[30] the challenges include standardized labeling and positioning. Labeling is crucial for training deep learning models because it directly affects model performance. However, owing to the lack of standardized labeling and positioning guidelines, labels and positions vary across studies, leading to differences in model performance. Differences in the reproducibility of lateral cephalometric landmarks may be attributed to different measurement methods, patient characteristics, and operator skill levels, resulting in variations in data or diagnosis.[31]

Based on a previous study,[29] we propose strategies to improve the accuracy of manually labeled points in the future. First, using multiple-point localization methods may improve labeling accuracy because different methods yield better performances for certain features. Second, recruiting multiple labelers and standardizing the definitions of each point may reduce individual subjectivity and improve labeling consistency and accuracy. Additionally, labelers should be trained to help them better understand labeling requirements and positions.[32] Finally, quality control of input images, such as seeking the assistance of experienced orthodontists to supervise labeling accuracy and promptly detecting and adjusting any issues, may improve input data quality, leading to better model performance.

Another limitation was the difficulty of getting approved to do experiments on patients under eighteen; our training data are all adults, which means that this model only works when the subject is an adult. Also, since those patients who have received previous orthognathic surgery or other treatments may affect machine learning, our model performs better when the subjects have no treatment.

In conclusion, our CNN model achieved excellent results, like those of the manual marking group, during the analysis of lateral cephalometric images for orthodontic treatment. By utilizing MobileNetV2, we can reduce huge portion of calculations and still make the model perform efficaciously. These findings highlight the predictive ability and reliability of the proposed model. The standardization of manual point placement, image quality, and image count may

influence model performance. In the future, our focus will be on improving the AI architecture and upgrading the hardware equipment.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

1. Nebauer C. Evaluation of convolutional neural networks for visual recognition. *IEEE Trans Neural Network* 1998;9:685—96.
2. Ciresan DC, Meier U, Masci J, Gambardella LM, Schmidhuber J. Flexible, high performance convolutional neural networks for image classification. In: *Twenty-second international joint conference on artificial intelligence*; 2011, June.
3. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process* 2012;25:1097—105.
4. Yang X, Wu N, Cheng G, et al. Automated segmentation of the parotid gland based on atlas registration and machine learning: a longitudinal MRI study in head-and-neck radiation therapy. *Int J Radiat Oncol Biol Phys* 2014;90:1225—33.
5. U-net: convolutional networks for biomedical image segmentation. In: Ronneberger O, Fischer P, Brox T, eds. *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, Part III 18*. Springer International Publishing, 2015:234—41.
6. LeCun Y, Bengio Y, Hinton G. Google scholar google scholar cross ref cross ref. Deep learning. *Nature* 2015;521:436—44.
7. Belhumeur PN, Jacobs DW, Kriegman DJ, Kumar N. Localizing parts of faces using a consensus of exemplars. *IEEE Trans Pattern Anal Mach Intell* 2013;35:2930—40.
8. Face detection, pose estimation, and landmark localization in the wildZhu X, Ramanan D, eds. *IEEE conference on computer vision and pattern recognition* 2012:2879—86.
9. Huval B, Wang T, Tandon S, et al. An empirical evaluation of deep learning on highway driving. *arXiv preprint arXiv* 2015; 150401716.[.
10. Vial A, Stirling D, Field M, et al. The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: a review. *Transl Cancer Res* 2018;7:803—16.
11. Cardillo J, Sid-Ahmed MA. An image processing system for locating craniofacial landmarks. *IEEE Trans Med Imaging* 1994; 13:275—89.
12. Grau V, Alcaniz M, Juan M, Monserrat C, Knoll C. Automatic localization of cephalometric landmarks. *J Biomed Inform* 2001;34:146—56.
13. Kahn Jr CE. From images to actions: opportunities for artificial intelligence in radiology. *Radiology* 2017;285:719—20.
14. Ibragimov B, Prince JL, Murano EZ, et al. Segmentation of tongue muscles from super-resolution magnetic resonance images. *Med Image Anal* 2015;20:198—207.
15. Kafieh R, Sadri S, Mehri A, Raji H. Discrimination of bony structures in cephalograms for automatic landmark detection. In: *Advances in computer science and engineering: 13th international CSI computer conference, CSICC 2008 kish island, Iran, March 9-11, 2008 revised selected papers*; 2008:609—20.
16. Ibragimov B, Likar B, Pernus F. A game-theoretic framework for landmark-based image segmentation. *IEEE Trans Med Imaging* 2012;31:1761—76.
17. Dunbar AC, Bearn D, McIntyre G. The influence of using digital diagnostic information on orthodontic treatment planning-a pilot study. *J Healthc Eng* 2014;5:411—28.
18. Durão APR, Morosolli A, Pittayapat P, Bolstad N, Ferreira AP, Jacobs R. Cephalometric landmark variability among ortho-dontists and dentomaxillofacial radiologists: a comparative study. *Imaging Sci Dent* 2015;45:213—20.
19. Arik SÖ, Ibragimov B, Xing L. Fully automated quantitative cephalometry using convolutional neural networks. *J Med Imaging* 2017;4:014501.
20. 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;6:1—10.
21. Blaiech AG, Khalifa KB, Valderrama C, Fernandes MA, Bedoui MH. A survey and taxonomy of FPGA-based deep learning accelerators. *J Syst Archit* 2019;98:331—45.
22. Gerchberg RW. A practical algorithm for the determination of plane from image and diffraction pictures. *Optik* 1972;35: 237—46.
23. Bulat A, Tzimiropoulos G, eds. *Human pose estimation via convolutional part heatmap regression. Computer Vision--ECCV.* vol. 7; 2016:717—32.
24. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC, eds. *Mobilenetv2: inverted residuals and linear bottlenecks. Proc IEEE conf comput vis pattern recognit*; 2018:4510—20.
25. Riedel RA. The relation of maxillary structures to cranium in malocclusion and in normal occlusion. *Angle Orthod* 1952;22: 142—5.
26. Hwang S, Jeong S, Choi YJ, Chung CJ, Lee HS, Kim KH. Three-dimensional evaluation of dentofacial transverse widths of adults with various vertical facial patterns. *Am J Orthod Dentofacial Orthop* 2018;153:692—700.
27. Schwendicke F, Chaurasia A, Arsiwala L, et al. Deep learning for cephalometric landmark detection: systematic review and meta-analysis. *Clin Oral Invest* 2021;25:4299—309.
28. Schlicher W, Nielsen I, Huang JC, Maki K, Hatcher DC, Miller A. Consistency and precision of landmark identification in three-dimensional cone beam computed tomography scans. *Eur J Orthod* 2012;34:263—75.
29. Perillo M, Beideman R, Shofer F, et al. Effect of landmark identification on cephalometric measurements: guidelines for cephalometric analyses. *Clin Orthod Res* 2000;3:29—36.
30. Junaid N, Khan N, Ahmed N, et al. Development, application, and performance of artificial intelligence in cephalometric landmark identification and diagnosis: a systematic review. *Healthcare* 2022;10:2454.
31. Trpkova B, Major P, Prasad N, Nebbe B. Cephalometric landmarks identification and reproducibility: a meta analysis. *Am J Orthod Dentofacial Orthop* 1997;112:165—70.
32. Haynes S, Chau MN. Inter-and intra-observer identification of landmarks used in the Delaire analysis. *Eur J Orthod* 1993;15: 79—84.