Original Article

Landmark annotation and mandibular lateral deviation analysis of posteroanterior cephalograms using a convolutional neural network

Saori Takeda [a†], Yuichi Mine [a*†], Yuki Yoshimi [b], Shota Ito [b], Kotaro Tanimoto [b], Takeshi Murayama [a]

[a] Department of Medical System Engineering, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan
[b] Department of Orthodontics and Craniofacial Developmental Biology, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan

Abstract *Background/purpose:* Facial asymmetry is relatively common in the general population. Here, we propose a fully automated annotation system that supports analysis of mandibular deviation and detection of facial asymmetry in posteroanterior (PA) cephalograms by means of a deep learning-based convolutional neural network (CNN) algorithm.
Materials and methods: In this retrospective study, 400 PA cephalograms were collected from the medical records of patients aged 4 years 2 months—80 years 3 months (mean age, 17 years 10 months; 255 female patients and 145 male patients). A deep CNN with two optimizers and a random forest algorithm were trained using 320 PA cephalograms; in these images, four PA landmarks were independently identified and manually annotated by two orthodontists.
Results: The CNN algorithms had a high coefficient of determination (R^2), compared with the random forest algorithm (CNN-stochastic gradient descent, $R^2 = 0.715$; CNN-Adam, $R^2 = 0.700$; random forest, $R^2 = 0.486$). Analysis of the best and worst performances of the algorithms for each landmark demonstrated that the right latero-orbital landmark was most difficult to detect accurately by using the CNN. Based on the annotated landmarks, reference lines were defined using an algorithm coded in Python. The CNN and random forest algorithms exhibited similar accuracy for the distance between the menton and vertical reference line.
Conclusion: Our findings imply that the proposed deep CNN algorithm for detection of facial asymmetry may enable prompt assessment and reduce the effort involved in orthodontic diagnosis.

* Corresponding author. Department of Medical System Engineering, Graduate School of Biomedical and Health Sciences, Hiroshima University, 1-2-3 Kasumi Minami-ku, Hiroshima 734-8553, Japan.
E-mail address: mine@hiroshima-u.ac.jp (Y. Mine).
† Saori Takeda and Yuichi Mine contributed equally to this work.

Introduction

Facial asymmetry is commonly encountered in the general population.[1] Generally, lateral deviation of the mandible is easily recognized in patients with severe facial asymmetry; this manifestation causes concern for patients and leads to functional impairment. Therefore, facial symmetry affects satisfaction with orthodontic treatment.[2,3]

Cephalometric radiography is a diagnostic tool used to quantitatively analyze the craniomaxillofacial skeleton during orthodontic treatment planning. Broadbent introduced cephalometric radiography in 1931;[4] cephalometric analysis has since been an important diagnostic procedure during orthodontic treatment. In orthodontic treatment planning, assessment of the midline and mandibular deviation in posteroanterior (PA) cephalograms are complex but clinically important steps. Notably, additional orthodontic training is needed to accurately evaluate the midline and mandibular deviation.

In recent years, deep learning (a branch of artificial intelligence) has developed rapidly and has shown potential for solving complicated medical tasks. Convolutional neural networks (CNNs), a type of deep learning inspired by human vision, have performed well in image classification tasks because of their robust abilities to automatically learn important features from images in healthcare,[5] including the dental field. Research involving artificial intelligence algorithms is ongoing with respect to the detection of caries[6,7] and apical lesions,[8] diagnosis of periodontal disease[9,10] and oral cancer,[11,12] and implementation of maxillofacial prosthetic rehabilitation.[13] Overall, lateral cephalometric analysis has been demonstrated to automatically identify anatomical landmarks.[14–18] In 2014 and 2015, public competitions for automatic lateral cephalometric landmark detection were held at the Institute of Electrical and Electronics Engineers International Symposium on Biomedical Imaging, with the goal of establishing an integrative approach to biomedical image analysis.[19,20] These competitions encouraged the continuing development of various approaches for automatic detection of lateral cephalometric landmarks.

Here, we propose a fully automated annotation system that supports analysis of mandibular deviation and detection of facial asymmetry in PA cephalograms by means of a deep learning-based CNN algorithm. In the first part of this study, four landmarks used for mandibular deviation analysis were annotated in PA cephalograms; the CNN algorithm with a stochastic gradient descent (SGD) optimizer (i.e., CNN-SGD) showed the best experimental performance. In the second part of this study, two reference lines were automatically defined and mandibular deviation was measured to aid in prompt detection of facial asymmetry.
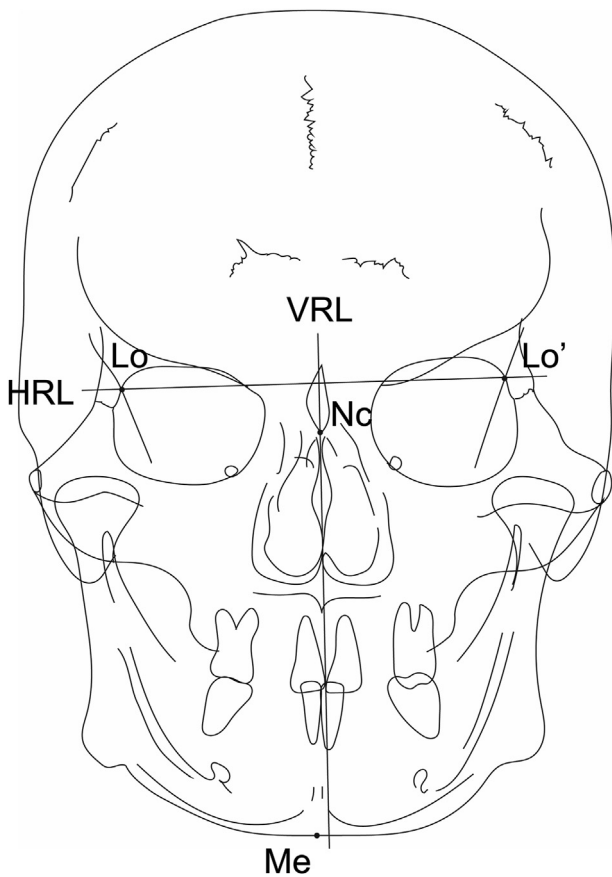
Materials and methods

Dataset

This study was approved by the Ethical Committee for Epidemiology of Hiroshima University (Approval Number: E−2119). Four hundred PA cephalograms were collected from the medical records of patients aged 4 years 2 months−80 years 3 months (mean age, 17 years 10 months; 255 female patients and 145 male patients). All images were recorded in DICOM format by using a cephalometric scanner (CX-150 W; Asahi Roentgen Ind. Co., Ltd, Kyoto, Japan). The original image resolution was 1648×1980 pixels with pixel spacing of 0.15 mm; images were resized to 256×256 pixels. Subsequently, the images were randomly divided into a training set (320 images) and a test set (80 images). Two orthodontists (12 and 6 years of experience, respectively) independently identified and manually annotated four PA cephalometric landmarks: neck of crista galli, right latero-orbital, left latero-orbital, and menton (Me) (Fig. 1). The X, Y coordinate values for each landmark were recorded as datasets and defined as the ground truth locations. Then, the horizontal reference line (a straight line connecting the right and left latero-orbital landmarks) and vertical reference line (VRL; a perpendicular line to the horizontal reference line through the neck of crista galli) were defined (Fig. 1). All landmark and line definitions employed Sassouni analysis, which is the most commonly used method for assessment of asymmetry.[21]

Deep learning-based CNN and random forest algorithms

All procedures were performed with an Intel Core i7-9750H 2.60 GHz CPU (Intel, Santa Clara, CA, USA), 16.0 GB RAM, and NVIDIA GeForce RTX 2070 MAX-Q 8.0 GB GPU (NVIDIA, Santa Clara, CA, USA). CNNs were constructed using Python and implemented using the Keras framework for deep learning, with TensorFlow as the backend.

Supervised learning was implemented by means of two machine learning approaches: a deep learning-based CNN and a random forest algorithm (i.e., a robust decision tree-based machine learning algorithm).[22] The overall CNN architecture is shown in Fig. 2. In the CNN learning process, optimizers play crucial roles in model training.[23,24] Therefore, two CNN optimizers were employed in this study: SGD and Adam,[25] with learning rates of 1.8×10^{-3} and 1.8×10^{-6}, respectively. The total epoch number was 5000. The output layer had eight nodes, defined as individual pairs of X, Y coordinate values for right latero-orbital, left latero-orbital, neck of crista galli, and Me landmarks. The random forest hyper parameters were implemented with

measurements, in accordance with previous reports.[19,20] The successful detection rate was defined as the proportion of corresponding landmarks within 2 mm, 2.5 mm, 3 mm, and 4 mm from the ground truth location. The distance from the VRL to the Me was measured to determine mandibular deviation. To evaluate algorithm performance, mean absolute error (MAE) was calculated using the following formula:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |D_i - D_i'|,$$

where D_i is the ground truth distance from the VRL to the Me and D_i' is the predicted distance from the predicted VRL to the predicted Me.

Results

Table 1 shows the algorithm performances during prediction of the four landmarks. The CNN algorithms had high R^2 values, compared with the random forest algorithm (CNN-SGD, $R^2 = 0.715$; CNN-Adam, $R^2 = 0.700$; random forest, $R^2 = 0.486$). Table 2 summarizes the best and worst performances of the algorithms for each landmark. Notably, the right latero-orbital landmark was most difficult to detect accurately by using the CNN. The successful detection rates of the CNN and random forest algorithms are shown in Fig. 3. Compared with the random forest algorithm, the CNN-SGD algorithm exhibited an approximately 5% higher successful detection rate across all precision ranges.

Based on the annotated landmarks, the horizontal reference line and VRL were defined using an algorithm coded in Python. Representative predicted reference lines are shown in Fig. 4. The distances between the Me and VRL were determined for ground truth and predicted

Figure 1 Landmarks and prediction points on posteroanterior cephalograms. NC, neck of crista galli; Lo, right latero-orbital; Lo', left latero-orbital; Me, menton; HRL, horizontal reference line; VRL, vertical reference line.

default settings. Following landmark annotation, the horizontal reference line and VRL were automatically defined with an algorithm coded in Python.

Performance metrics

Eighty test images were used to validate accuracy and computational efficacy. The CNN and random forest performances were assessed based on the coefficient of determination (R^2). All annotation results were analyzed as the successful detection rate (in %) for four precision

Table 1 Performance evaluation of proposed algorithms for landmark prediction.

Algorithm	Optimizer	R^2
CNN	SGD	0.715
	Adam	0.700
Random forest	—	0.486

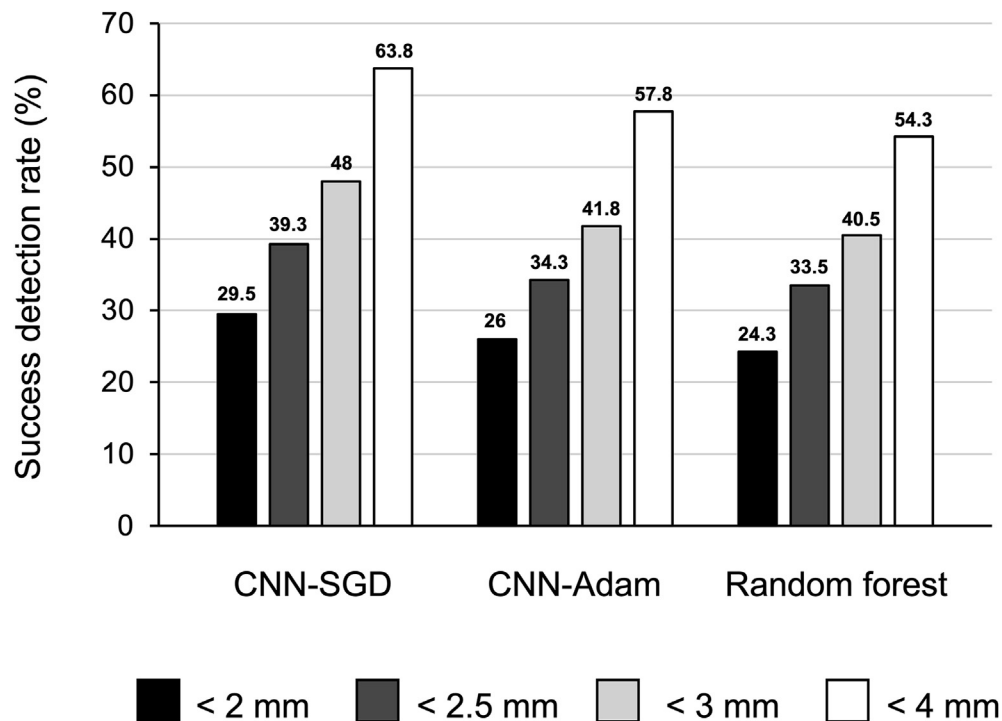CNN, convolutional neural network; SGD, stochastic gradient descent.
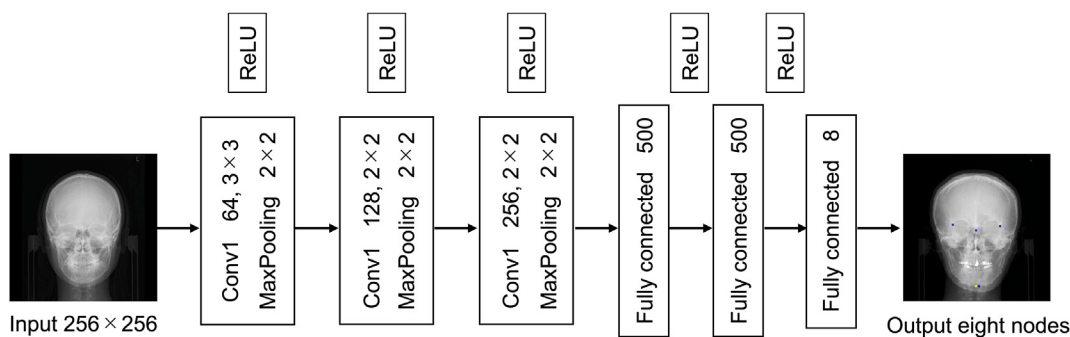
Figure 2 Architecture of the convolutional neural network used in this study. ReLU, rectified linear unit.

Table 2   Successful detection rates for four landmarks at four precision ranges.

Algorithm	Optimizer	Landmark	SDR (%)			
			<2.0 mm	<2.5 mm	<3.0 mm	<4.0 mm
CNN	SGD	Mn	26	33	45	63
		Nc	30	41	50	62
		Lo	26	35	42	60
		Lo'	36	48	55	70
	Adam	Mn	27	41	45	57
		Nc	32	38	47	65
		Lo	22	27	37	51
		Lo'	23	31	38	58
Random forest	—	Mn	12	13	26	35
		Nc	37	48	53	67
		Lo	22	36	40	55
		Lo'	26	37	43	60

SDR, successful detection rate; CNN, convolutional neural network; SGD, stochastic gradient descent.


Figure 3   Success detection rates of the proposed convolutional neural network and random forest algorithms for 2.0-mm, 2.5-mm, 3.0-mm, and 4.0-mm precision ranges for the four landmarks assessed in this study. CNN, convolutional neural network; SGD, stochastic gradient descent.

landmarks, respectively. MAEs between ground truth and predicted distances were similar for CNN and random forest algorithms (Table 3).

Discussion

Various artificial intelligence algorithms have been proposed and applied to medical research. Here, we applied two artificial intelligence algorithms (i.e., CNN and random forest) to orthodontic diagnosis; our proposed system was able to automatically annotate landmarks and measure mandibular lateral deviation by means of PA cephalograms. CNNs typically consist of three types of layers: convolution, pooling, and fully connected. Convolution layers are composed of multiple feature maps, whereas pooling layers are inserted periodically between convolution layers to reduce the number of parameters in the network. These two types of layers perform feature map extraction of the images; extracted features are then transformed into the fully connected layer. The algorithm does not require manual feature extraction and does not necessarily require
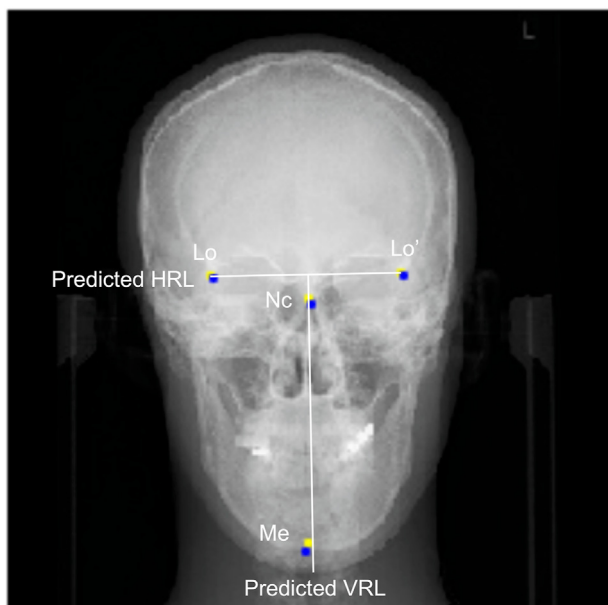
Figure 4 Representative annotations of landmarks and definitions of reference lines. Ground truth landmarks (blue), convolutional neural network-stochastic gradient descent predictions (yellow), and predicted reference lines are shown. NC, neck of crista galli; Lo, right latero-orbital; Lo', left latero-orbital; Me, menton; HRL, horizontal reference line; VRL, vertical reference line.

Table 3 Performance evaluation of proposed algorithm in terms of mean absolute error for distance from the vertical reference line.

Algorithm	Optimizer	MAE	Max	Min	Median
CNN	SGD	1.67 ± 1.77	7.957	0.038	1.147
	Adam	1.69 ± 1.63	8.374	0.013	1.174
Random forest	—	1.80 ± 1.81	8.298	0.005	1.295

CNN, convolutional neural network; SGD, stochastic gradient descent; MAE, mean absolute error.

manual segmentation of the target (e.g., tumor or organ) by human experts.[26] However, CNNs are computationally intensive because they require large amounts of data to estimate millions of trainable parameters.[26] Random forest is an ensemble algorithm that builds randomized decision trees and incorporates a variety of features into its classification process; importantly, it can deter overfitting by a "majority voting" approach. Thus, the random forest algorithm has crucial advantages relative to other algorithms; these include effectiveness in multiclass classification and regression tasks, rapid training speed, ease of implementation in parallel computation, tuning simplicity, robustness to noise, and ability to handle highly non-linear biological data.[27] Although prior studies have used lateral cephalograms, rather than PA cephalograms, machine learning-based algorithms such as random forest and CNN have been proposed for landmark annotation. Random

forest algorithms have been proposed for some landmark annotation systems.[19,20] In a recent study, You-Only-Look-Once version 3,[28] a CNN specialized for real-time object detection, was applied to lateral cephalograms for landmark detection.[17] These systems are effective for a broad range of landmark annotation.

Selection of an optimizer is an important step in deep learning-based CNNs that influences model performance.[23,24] The SGD and Adam optimizers were employed in this study because of their widespread use in prior investigations.[7,8,13,23,24] However, there is no established guideline for optimizer selection. Thus, researchers rely on empirical studies and comparative benchmarking.[24] In the present study, the SGD optimizer showed the best experimental performance, in terms of the successful detection rate.

Orthodontists may choose any of several methods for setting the facial midline, such as drawing a perpendicular line at the midpoint between two landmarks on either side, or connecting the left and right landmarks with a horizontal line and drawing a perpendicular line passing through a landmark located near the midline of the face. It is challenging to define facial midline due to the influences of vertical and horizontal errors in setting each landmark.[29] Thus, automatic detection of the neck of crista galli and other PA landmarks by means of artificial intelligence may be a promising clinical technique that facilitates definition of the facial midline. A limitation of this study was that the successful detection rates were relatively moderate, compared with previously reported landmark detection systems that used lateral cephalograms; this might have had substantial effects on the final measurement of mandibular deviation. Despite the use of symmetrical landmarks, the right latero-orbital landmark showed lower successful detection rates, compared with the left latero-orbital landmark. The difference in annotations between the two experts may have been greater for the right latero-orbital landmark in our dataset. In addition, craniofacial growth continues with advancing age in humans, a phenomenon widely accepted in current medical literature.[30,31] We included 400 patients in our small dataset; their ages ranged from 4 to 80 years. Given the changes that occur in the facial skeleton over time, this is a fairly wide age range. We presume that many sophisticated datasets evaluated by consensuses involving several experts will contribute to the improvement of successful detection rates in the future.

The Sassouni analysis employed in this study is a widely used method in which lines connecting the left and right latero-orbital landmarks are used as the horizontal reference plane, while the perpendicular line passing through the neck of crista galli is regarded as the facial midline.[21] Some investigators have concluded that the neck of crista galli is among the landmarks with the greatest inter-inspector error in PA cephalometric analysis.

It is important to emphasize that the final evaluation of facial symmetry requires hard tissue evaluation by PA cephalometric analysis, as well as soft tissue evaluation with facial photographs.[32,33] Recently, several methods have been reported for evaluating facial symmetry via simultaneous analysis of hard and soft tissue characteristics

in computed tomography images.[34,35] The use of CNNs for three-dimensional evaluation may provide an important diagnostic tool in the future.

In this study, we annotated PA cephalometric landmarks that contribute to the determination of reference lines in the Sassouni analysis, using deep learning-based CNN algorithms; we evaluated the precision of this annotation. Additionally, we described systems that could automatically measure mandibular deviation to aid in the detection of facial asymmetry. Although further improvement may be necessary for clinical implementation, the proposed application of deep CNNs for detection of facial asymmetry offers a promising technique that might reduce the effort involved in orthodontic diagnosis. Future studies should focus on building a comprehensive diagnostic system that includes lateral cephalometric analysis and three-dimensional evaluation.

Declaration of competing interest

All authors declare no conflicts of interest.

Acknowledgments

References

1. Thiesen G, Gribel BF, Freitas MP. Facial asymmetry: a current review. *Dental Press J Orthod* 2015;20:110–25.
2. Proffit WR, Phillips C, Dann 4th C. Who seeks surgical-orthodontic treatment? *Int J Adult Orthod Orthognath Surg* 1990;5:153–60.
3. Chen YJ, Yao CC, Chang ZC, Lai HH, Lu SC, Kok SH. A new classification of mandibular asymmetry and evaluation of surgical-orthodontic treatment outcomes in Class III malocclusion. *J Cranio-Maxillo-Fac Surg* 2016;44:676–83.
4. Broadbent BH. A new X-ray technique and its application to orthodontia. *Angle Orthod* 1931;1:45–66.
5. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;42:60–88.
6. Lee JH, Kim DH, Jeong SN, Choi SH. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *J Dent* 2018;77:106–11.
7. Schwendicke F, Elhennawy K, Paris S, Friebertshäuser P, Krois J. Deep learning for caries lesion detection in near-infrared light transillumination images: a pilot study. *J Dent* 2020;92:103260.
8. Ekert T, Krois J, Meinhold L, et al. Deep learning for the radiographic detection of apical lesions. *J Endod* 2019;45:917–22.
9. Papantonopoulos G, Takahashi K, Bountis T, Loos BG. Artificial neural networks for the diagnosis of aggressive periodontitis trained by immunologic parameters. *PloS One* 2014;9:e89757.
10. Krois J, Ekert T, Meinhold L, et al. Deep learning for the radiographic detection of periodontal bone loss. *Sci Rep* 2019;9:8495.
11. Kim DW, Lee S, Kwon S, Nam W, Cha IH, Kim HJ. Deep learning-based survival prediction of oral cancer patients. *Sci Rep* 2019;9:6994.
12. Kar A, Wreesmann VB, Shwetha V, et al. Improvement of oral cancer screening quality and reach: the promise of artificial intelligence. *J Oral Pathol Med* 2020;49:727–30.
13. Mine Y, Suzuki S, Eguchi T, Murayama T. Applying deep artificial neural network approach to maxillofacial prostheses coloration. *J Prosthodont Res* 2020;64:296–300.
14. Ibragimov B, Likar B, Pernus F, Vrtovec T. Automatic cephalometric X-ray landmark detection by applying game theory and random forests. *Proc. ISBI Int. Symp. on Biomedical Imaging* 2014:1–8.
15. 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;6:33581.
16. Arık SÖ, Ibragimov B, Xing L. Fully automated quantitative cephalometry using convolutional neural networks. *J Med Imaging* 2017;4:014501.
17. Park JH, Hwang HW, Moon JH, et al. Automated identification of cephalometric landmarks: Part 1-Comparisons between the latest deep-learning methods YOLOV3 and SSD. *Angle Orthod* 2019;89:903–9.
18. 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;81:52–68.
19. Wang CW, Huang CT, Hsieh MC, et al. Evaluation and comparison of anatomical landmark detection methods for cephalometric X-Ray images: a grand challenge. *IEEE Trans Med Imag* 2015;34:1890–900.
20. Wang CW, Huang CT, Lee JH, et al. A benchmark for comparison of dental radiography analysis algorithms. *Med Image Anal* 2016;31:63–76.
21. Sassouni V. Diagnosis and treatment planning via roentgenographic cephalometry. *Am J Orthod* 1958;44:433–63.
22. Breiman L. Random forests. *Mach Learn* 2001;45:5–32.
23. Ruder S. *An overview of gradient descent optimization algorithms*. arXiv preprint, 2016. arXiv:1609.04747.
24. Choi D, Shallue CJ, Nado Z, Lee J, Maddison CJ, Dahl GE. *On empirical comparisons of optimizers for deep learning*. arXiv preprint, 2019, 05446. arXiv:1910.
25. Kingma DP, Ba J. *Adam: a method for stochastic optimization*. arXiv preprint, 2014. arXiv:1412.6980.
26. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights Imag* 2018;9:611–29.
27. Lebedev AV, Westman E, Van Westen GJ, et al. Alzheimer's Disease Neuroimaging Initiative and the AddNeuroMed consortium. Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness. *Neuroimage Clin* 2014;6:115–25.
28. Redmon J, Farhadi A. *Yolov3: an incremental improvement*. arXiv preprint, 2018. arXiv:1804.02767.
29. Trpkova B, Prasad NG, Lam EW, Raboud D, Glover KE, Major PW. Assessment of facial asymmetries from posteroanterior cephalograms: validity of reference lines. *Am J Orthod Dentofacial Orthop* 2003;123:512–20.
30. Enlow DH. A morphogenetic analysis of facial growth. *Am J Orthod* 1966;52:283–99.
31. Mendelson B, Wong CH. Changes in the facial skeleton with aging: implications and clinical applications in facial rejuvenation. *Aesthetic Plast Surg* 2020;44:1151–8.

32. Maeda M, Katsumata A, Ariji Y, et al. 3D-CT evaluation of facial asymmetry in patients with maxillofacial deformities. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2006;102: 382—90.

33. Masuoka N, Muramatsu A, Ariji Y, Nawa H, Goto S, Ariji E. Discriminative thresholds of cephalometric indexes in the subjective evaluation of facial asymmetry. *Am J Orthod Dentofacial Orthop* 2007;131:609—13.

34. Lonic D, Sundoro A, Lin HH, Lin PJ, Lo LJ. Selection of a horizontal reference plane in 3D evaluation: identifying facial asymmetry and occlusal cant in orthognathic surgery planning. *Sci Rep* 2017;7:2157.

35. Dobai A, Markella Z, Vízkelety T, Fouquet C, Rosta A, Barabás J. Landmark-based midsagittal plane analysis in patients with facial symmetry and asymmetry based on CBCT analysis tomography. *J Orofac Orthop* 2018;79:371—9.