Original Article

Deep focus approach for accurate bone age estimation from lateral cephalogram

Hyejun Seo [a,b,1], JaeJoon Hwang [c,d,1], Yun-Hoa Jung [c,d],
Eungyung Lee [a,d], Ok Hyung Nam [e*], Jonghyun Shin [a,d**]

[a] *Department of Pediatric Dentistry, School of Dentistry, Pusan National University, Yangsan, South Korea*
[b] *Department of Dentistry, Ulsan University Hospital, Ulsan, South Korea*
[c] *Department of Oral and Maxillofacial Radiology, School of Dentistry, Pusan National University, Yangsan, South Korea*
[d] *Dental and Life Science Institute & Dental Research Institute, School of Dentistry, Pusan National University, Yangsan, South Korea*
[e] *Department of Pediatric Dentistry, School of Dentistry, Kyung Hee University, Seoul, South Korea*

Abstract *Background/purpose:* Bone age is a useful indicator of children's growth and development. Recently, the rapid development of deep-learning technique has shown promising results in estimating bone age. This study aimed to devise a deep-learning approach for accurate bone-age estimation by focusing on the cervical vertebrae on lateral cephalograms of growing children using image segmentation.

Materials and methods: We included 900 participants, aged 4—18 years, who underwent lateral cephalogram and hand-wrist radiograph on the same day. First, cervical vertebrae segmentation was performed from the lateral cephalogram using DeepLabv3+ architecture. Second, after extracting the region of interest from the segmented image for preprocessing, bone age was estimated through transfer learning using a regression model based on Inception-ResNet-v2 architecture. The dataset was divided into train:test sets in a ratio of 4:1; five-fold cross-validation was performed at each step.

Results: The segmentation model possessed average accuracy, intersection over union, and mean boundary F1 scores of 0.956, 0.913, and 0.895, respectively, for the segmentation of cervical vertebrae from lateral cephalogram. The regression model for estimating bone age from segmented cervical vertebrae images yielded average mean absolute error and root mean squared error values of 0.300 and 0.390 years, respectively. The coefficient of determination

* Corresponding author. Department of Pediatric Dentistry, Kyung Hee University School of Dentistry, 26 Kyungheedae-ro, Dongdaemun-gu, Seoul 02447, South Korea.
** Corresponding author. Department of Pediatric Dentistry, Pusan National University, Geumo-ro 20, Mulgeum-eup, Yangsan-si, 50612, South Korea.
E-mail addresses: pedokhyung@gmail.com (O.H. Nam), jonghyuns@pusan.ac.kr (J. Shin).
[1] These authors contributed equally to this work.

of the proposed method for the actual and estimated bone age was 0.983. Our method visualized important regions on cervical vertebral images to make a prediction using the gradient-weighted regression activation map technique.

Conclusion: Results showed that our proposed method can estimate bone age by lateral cephalogram with sufficiently high accuracy.

Introduction

Age assessment is performed using various biological indicators, such as the face, bones, skeleton, and dental structures, because various skeletal structures of the human body undergo differential changes depending on the growth and development pattern.[1—3] The terms growth and development mainly refer to the skeleton, and radiation analysis is widely accepted as the gold standard for evaluating bone maturity.[1,4—7]

Various radiographic methods have been used to evaluate skeletal growth and developmental status. The hand-wrist radiograph method is based on the gradual change in the carpal bones and the degree of change and fusion of the epiphyses and diaphysis. This method has been studied the longest and is widely used clinically to determine the skeletal age, which is the principal type of physiological age used in the medical and dental fields.[5,6,8,9] In dentistry, hand-wrist radiograph analysis is used to evaluate the growth stage during orthodontic treatment of a growing child.[9]

Although growth evaluation based on hand-wrist radiograph is a reliable method, its clinical application is limited since the maturation and timing of ossification measured on radiographs may be subject to individual and sex-based differences, in addition to the need for exposure to radiation.[5,6,9] The possibility of assessing the growth stage using standardized lateral cephalogram, which is used for basic orthodontic evaluation, has currently garnered attention because it does not require additional radiation exposure resulting from the acquisition of hand-wrist radiograph.[6] Several studies have evaluated the growth stage using the cervical spine on standardized lateral cephalogram, which is a useful and predictable indicator for growth assessment. Studies have reported that the stage of pubertal growth, which is important for planning and determining the timing of orthodontic treatment, can be identified according to the maturational stage of segments 2—4 of the cervical spine.[6,10—12]

Various studies have revealed the relationship between the shape of the cervical vertebrae and the growth stage, and various methods have attempted to estimate age using this structure. A study has also found that the anterior height of the fourth cervical vertebra bears a strong correlation with hand-wrist bone age, concluding that lateral cephalogram is useful for bone-age estimation.[13] However, the general application of this method of age evaluation requires manual measurement of anatomical structures, which limits its applicability in busy clinical situations, since the process is manpower- and time-intensive.

The recent development of artificial intelligence in the medical field has resulted in the advent of a technique for automatically diagnosing radiographic images via deep-learning.[14,15] The accuracy of some results of this technique has already surpassed human accuracy.[16] Moreover, since this technique requires less time to diagnose an image compared to human assessors and artificial intelligence is not fatigued by repetitive work, it may provide wider application in the medical and dental fields.[17]

Currently, commercially available software can be used to predict bone age by learning hand-wrist radiographic images using deep-learning.[18] As growing children can be extremely sensitive to radiation doses, confirming the ability to evaluate bone age by applying deep-learning to cervical vertebrae images has important beneficial implications by reducing radiation exposure due to the acquisition of additional hand-wrist radiographic images. Recently, a fully automated deep-learning model that can perform cervical vertebrae maturation (CVM) classification using cervical segmentation from lateral cephalogram has also been studied.[19] However, predicting bone age may be more important clinically than classifying it into stages because growth is a continuous process. No study has investigated the prediction of bone age using the cervical vertebrae on lateral cephalogram with the deep-learning method.

Therefore, this study aimed to devise a deep-learning approach for accurate bone age estimation in growing children using lateral cephalogram via automatic region of interest (ROI) segmentation of the cervical vertebrae. Delineating the ROI is essential for enhancing the performance of diagnostic deep-learning models; therefore, adequate preprocessing is needed to obtain as small an ROI as possible that provides sufficient context to improve accuracy. For accurate analysis of the cervical vertebrae, intensive focus on this region can be accomplished by the image segmentation process.

Materials and methods

Ethics statement

This study was approved by the Institutional Review Board (IRB) of the Pusan National University Dental Hospital (IRB approval number: PNUDH-2021-032). The IRB of Pusan National University Dental Hospital waived the need for individual informed consent, as this study had a non-interventional retrospective design, and all the data were anonymized before analysis.

Participants

This study included 900 patients (chronological age: 4–18 years) who underwent both lateral cephalogram and hand-wrist radiograph on the same day between 2017 and 2021 at the Department of Pediatric Dentistry of Pusan National University Dental Hospital. All radiographs using a Proline PM 2002 CC machine (Planmeca, Helsinki, Finland) were stored in the DICOM format. The radiographs of patients with bone growth disorders and congenital or acquired disorders of the cervical vertebrae, hand, or wrist were excluded from the study.

Data annotation

The Med-BoneAge version 1.0.3 software program (VUNO, Seoul, South Korea) was used for automated bone age analysis. Each patient's hand–wrist radiograph was entered into the deep-learning-based automatic software for bone age determination, focusing on the shape and density of each bone. The software displayed the three most likely estimated bone age values as percentages in order of probability, and the first-rank bone age value (i.e., the estimation with the highest probability) was chosen as the deep-learning estimation. The ROIs around the cervical vertebrae were manually cropped and labeled bone-age values for the manual ROI method (Fig. 1A). The Image Labeler application present in the MATLAB 2022a software (MathWorks Inc., Natick, MA, USA) was used for the labeling process in the automatic ROI method. Each patient's cervical vertebrae observed on the lateral cephalogram were manually delineated for the pixel-based ground truth data before the bone-age labeling process (Fig. 1B).

Deep-learning algorithms

Our proposed automatic method for bone age estimation is composed of two networks, i.e., DeepLabv3+, the semantic segmentation network for delineated cervical vertebral region, and Inception-ResNet-v2, a classification network modified to a regression model for age estimation.

DeepLabv3+[20] was employed to train and test the pixel-labeled image data. DeepLabv3+ network introduces a common encoder-decoder structure of semantic segmentation, with atrous separable convolution. The whole data set comprised 900 lateral cephalograms, and semantic segmentation was performed using the DeepLabv3+ network (Fig. 2). The images were not subjected to further image processing, such as filtering or enhancement, to allow all information containing soft tissue to be learned.[21] Inception-ResNet-v2 pretrained using over a million images from the ImageNet database integrates the advantages of the Inception module and residual blocks of a ResNet backbone architecture.[22,23] During the preprocessing stage, each ROI was extracted from the bounding box of the binary mask image on the segmentation network to yield the input data. The classification network was converted into a regression network by modifying the layers. The convolutional layers of the classification networks extract the image features that are used by the last learnable and the final classification layer to classify the input image. These two layers contain information on the method to combine the features extracted by the network into class probabilities. These two layers were replaced with new layers that were adapted to the task to retrain the pretrained network for regression. We replaced the final fully connected layer, the softmax layer, and the classification output layer with a fully connected layer of size 1 (number of responses) and a regression layer (see Fig. 3). Transfer learning using pre-trained features was employed to estimate bone age using the segmented cervical vertebral images.

Five-fold cross-validation and data augmentation

The 900 images included in this study were subjected to five-fold cross-validation for accurate performance comparison. The dataset was divided into the training and test sets in a ratio of 4:1, which rendered all available datasets as the training and test sets. In the manual ROI method, bone age was estimated without ROI segmentation. Various data augmentation techniques were used to prevent overfitting of the deep-learning models on small datasets.
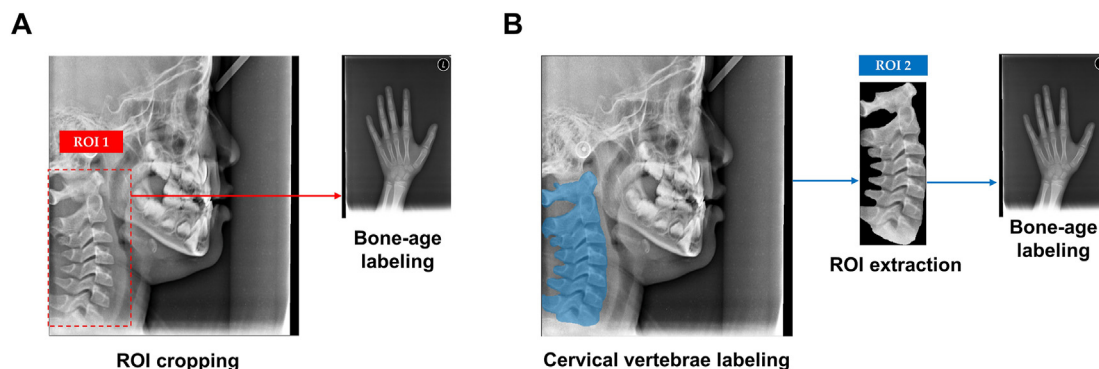

Fig. 1 Labeling process (A) the manually drawn region of interest (ROI) 1 (red-dashed box) around the cervical vertebrae is labeled with the patient's hand-wrist bone age. (B) The cervical vertebrae (blue color) are labeled as the ground truth on each lateral cephalogram. ROI 2 from the ROI extraction process is labeled with the same patient's hand-wrist bone age. (For interpretation of the references to color/color in this figure legend, the reader is referred to the Web version of this article.)
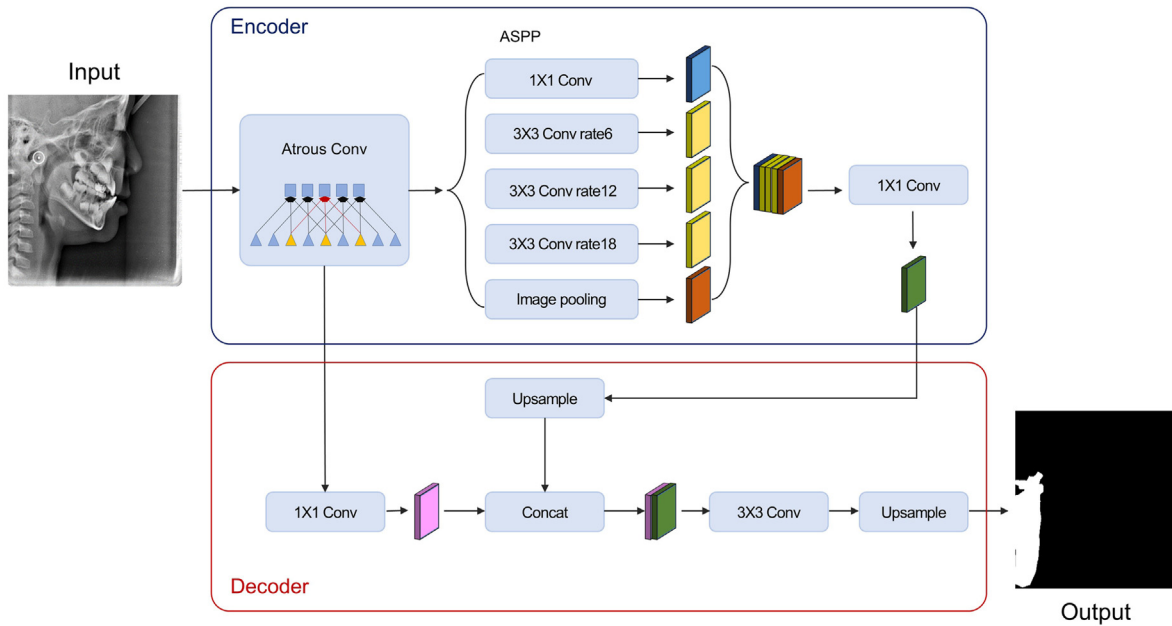
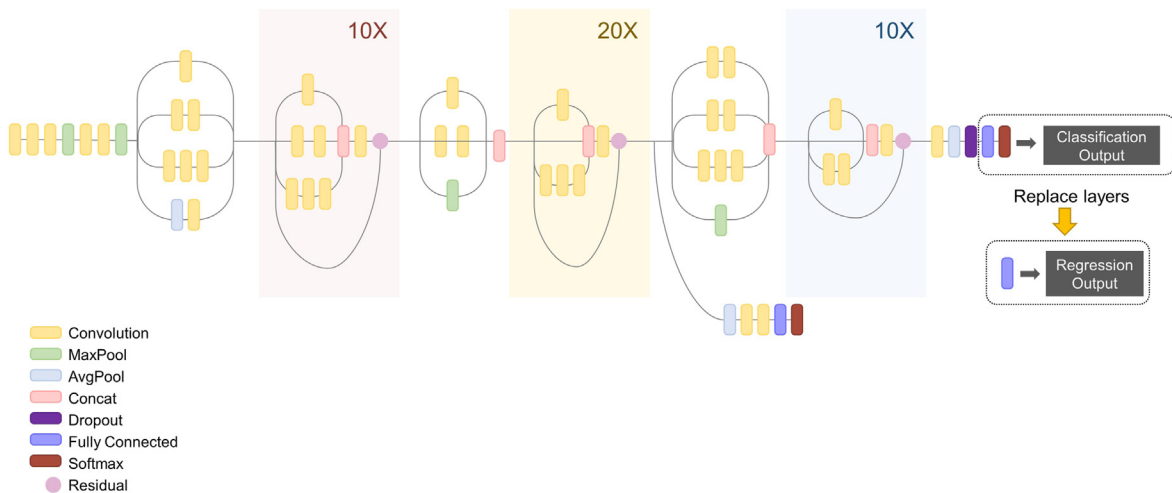**Fig. 2** DeepLabv3+ architecture for the semantic segmentation network.

Fig. 3 Transfer learning of Inception-ResNet-v2 architecture for the regression network.

Augmentation was achieved via rotation from −7 to 7, horizontal and vertical scaling from 0.8 to 1.2, and horizontal and vertical translation from −5 to 5 pixels.

Training configurations

The MATLAB deep-learning and parallel computing toolboxes (MathWorks Inc.) were used for network training on the Windows 10 operating system, which was accelerated using the NVIDIA Titan RTX graphical processing unit. The first model for segmentation was trained for up to 50 epochs using stochastic gradient descent with a momentum optimizer. The size of the mini-batch was 8, and the initial learning efficiency was e^{-3}. The second regression model was trained for up to 50 epochs using the Adam optimizer.[24]

The size of the mini-batch was 8, and the initial learning efficiency was e^{-4}.

Performance evaluation

The performance of the proposed method in the test set was evaluated using the following indices:

(1) Accuracy, intersection over union (IoU), and mean boundary F1 (BF) scores were calculated as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

$$BF^* = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

$$IoU = \frac{TP}{TP + FP + FN}$$

where TP: true positive, FP: false positive, FN: false negative, TN: true negative.

The mean BF score is the average BF score of the overall images of that class. The BF score ranges between [0, 1], where 1 indicates that the contours of objects in the corresponding class, prediction, and ground truth, are a perfect match.[25]

(2) The mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R^2) were calculated as follows:

$$MAE = \frac{\sum_{i=1}^{n} |y_i - \widehat{y}_i|}{n}$$

$$RMSE = \left(\sqrt{\frac{\sum_{i=1}^{n} (y_i - \widehat{y}_i)^2}{n}} \right)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2}$$

where y represents the real data, $\overline{y}$ is the mean value of the real data, $\widehat{y}$ represents the predicted data, and n is the number of samples.

(3) The Bland—Altman plot was used for visual assessment of the accuracy and precision of the predicted bone age compared to the original bone age.

The effect of ROI segmentation on bone-age estimation was investigated by comparing the automatic and manual ROI methods.

Visualization

We applied the gradient-weighted regression activation mapping (Grad-RAM) technique to both the automatic and manual ROI methods. Grad-RAM performs visualization by weighing the contribution of the image area for the regression result using the heatmap, which borrows the concept from gradient-weighted classification activation mapping.[26] Consequently, the heatmap shows the correspondence between the input images and the last regression layer of the network using the Grad-RAM technique. Therefore, the heatmap can localize the key determinant region, which results in the regression outcomes. In general, regions that contribute substantially to prediction are highlighted in red (warm color), whereas regions with low contribution are highlighted in blue (cool color). As a visual aid to help explain the results produced by our deep-learning methods, the heatmaps indicate the important regions in each image.

Results

Segmentation performance

Table 1 shows the results of segmentation of the cervical vertebrae on lateral cephalogram via five-fold cross-validation. The mean performance accuracy was 0.956, IoU was 0.913, and mean BF score was 0.895.

Regression performance

The data in Table 2 indicate the comparison of the chronological age and bone age in relation to the gender. The accuracy of bone age estimation for each method is enumerated in Table 3. After five-fold cross-validation, the

Table 1 Segmentation performance on lateral cephalogram using cross-validation.

Fold	Accuracy (background)	IoU (background)	Mean BF score (background)
1	0.958 (0.992)	0.916 (0.983)	0.906 (0.948)
2	0.956 (0.992)	0.916 (0.983)	0.903 (0.945)
3	0.961 (0.991)	0.913 (0.983)	0.890 (0.940)
4	0.954 (0.992)	0.912 (0.983)	0.892 (0.943)
5	0.950 (0.991)	0.909 (0.982)	0.884 (0.940)
Average	0.956 (0.992)	0.913 (0.983)	0.895 (0.945)

IoU: intersection over union, BF: boundary F1.

Table 2 Chronological age and bone age according to the gender.

Gender	Number	Chronological age		Bone age	
		Mean (years)	Standard deviation	Mean (years)	Standard deviation
Male	456	9.017	1.816	8.695	2.450
Female	444	9.214	2.080	9.305	2.485
Total	900				

Table 3 Regression performances through five-fold cross-validation.

Fold	Automatic ROI method			Manual ROI method		
	RMSE (years)	MAE (years)	R^2	RMSE (years)	MAE (years)	R^2
1	0.363	0.282	0.985	0.579	0.540	0.956
2	0.389	0.308	0.985	0.741	0.687	0.924
3	0.367	0.280	0.986	0.629	0.578	0.948
4	0.420	0.324	0.982	0.563	0.522	0.952
5	0.413	0.307	0.977	0.742	0.651	0.919
Average	0.390	0.300	0.983	0.651	0.595	0.934

ROI: region of interest, RMSE: root mean squared error, MAE: mean absolute error.

automatic ROI method exhibited the average RMSE value of 0.390 years and MAE value of 0.300 years. The manual ROI method exhibited the average RMSE value of 0.651 years and MAE value of 0.595 years. Figs. 4 and 5 depict linear regression plots and the Bland—Altman plots for the difference versus the mean between the predicted and actual bone ages over the mean of the two estimates. The average values of R-squared (coefficient of determination) were 0.983 (Figs. 4A) and 0.934 (Fig. 5A), respectively. The mean differences for all predicted ages and the ground truth were close to 0 years (0.13 and 0.21, respectively; $P < 0.001$), with a standard deviation of 0.95 (Figs. 4B and 5B).
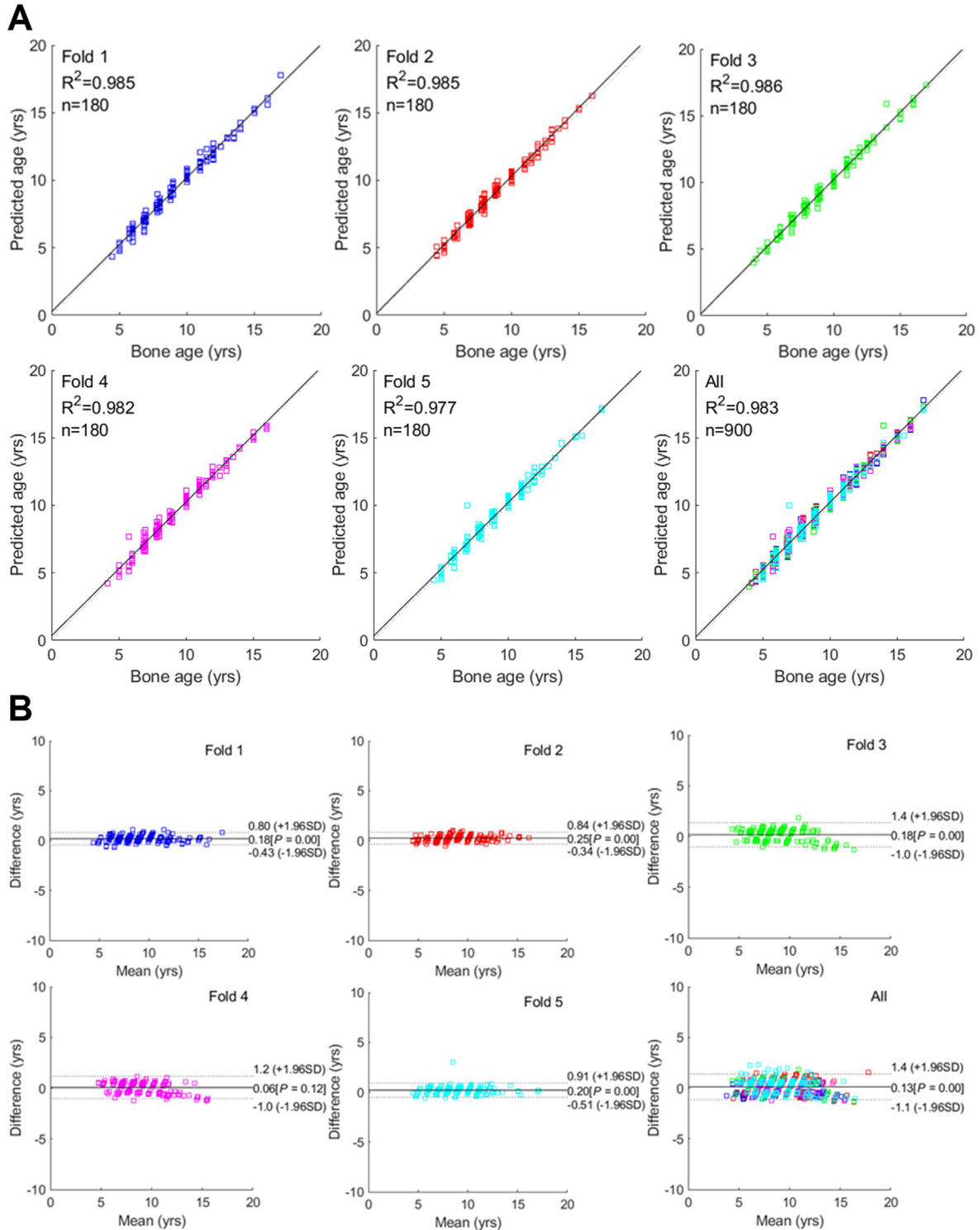

Fig. 4 Linear regression and Bland—Altman plots for comparison between the actual and predicted bone age by the automatic region of interest (ROI) method.

Visual explanation

Fig. 6 shows the original image for age estimation in the test set, and their corresponding heatmaps for each image. The heatmap produced by the manual ROI method sporadically focused on multiple regions around the cervical vertebrae. In contrast, the heatmap mainly focused on the cervical vertebrae in the automatic ROI method, which is distinct from the manual ROI method, also focusing on the black background without information.
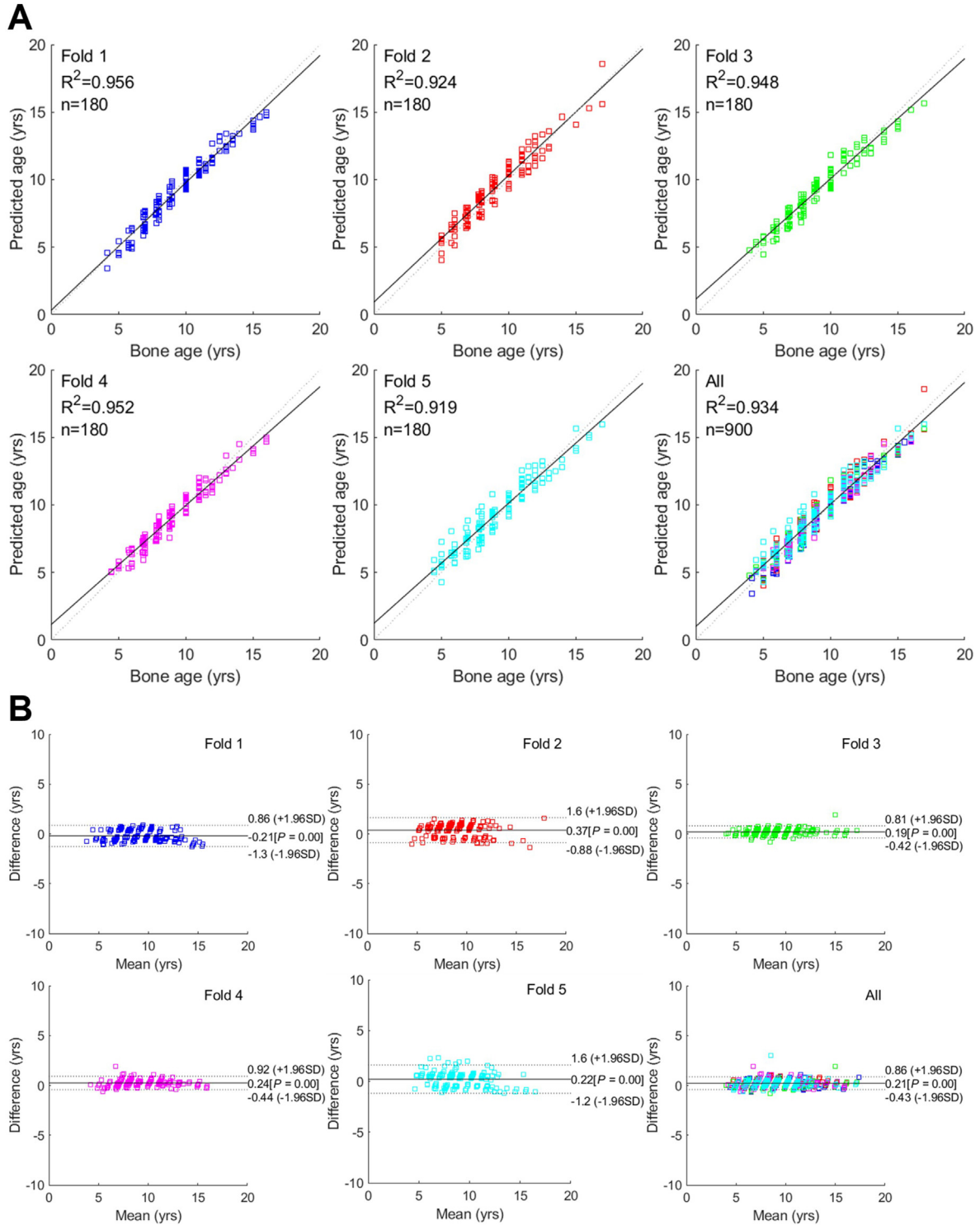

Fig. 5 Linear regression and Bland—Altman plots for comparison between the actual and predicted bone age by the manual region of interest (ROI) method.

Discussion

In the field of dentistry, the CVM method described by Baccetti et al. (2005) has been used to analyze the morphology of the second through fourth cervical vertebrae at six maturational stages using lateral cephalogram.[7] However, ordinal division of the six maturational stages and the presence of inter- and intra-observer variabilities has limitations for measuring individual growth, since growth is a continuous process.[27]

The bone age of growing children has been evaluated using hand-wrist radiograph. Studies have also reported that it is possible to evaluate the bone age using the cervical vertebrae with lateral cephalogram, which is usually acquired during orthodontic evaluation of the growing child.[9] Numerous studies have found that skeletal evaluation using the cervical vertebrae is highly correlated to chronological age, and thus can be an extremely useful tool for forensic age evaluation.[28]

The current study developed a deep-learning model to estimate bone age from the lateral cephalogram without the need for hand-wrist radiograph, through automatic ROI segmentation of the cervical vertebrae. ROI segmentation task in lateral cephalogram increased performance of estimating CVM using convolutional neural network.[19] In this manner, we compared the automatic and manual ROI methods to analyze the effect of ROI segmentation on bone-age estimation. Better performance of our automatic
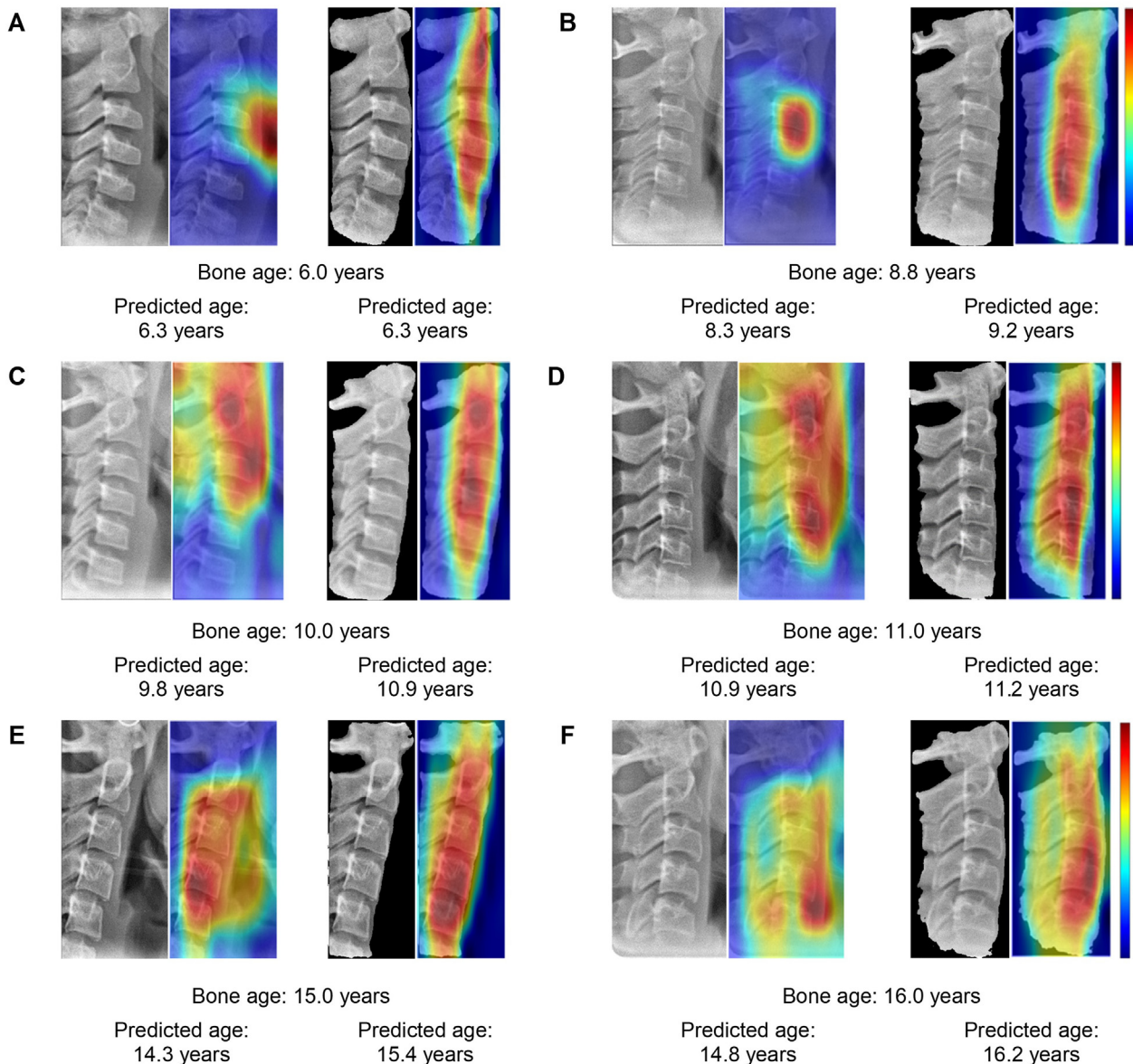

Fig. 6 Examples for the age estimation and gradient-weighted regression activation map (Grad-RAM) results. The bone age and predicted age are presented below each image. The input image for the manual region of interest (ROI) method (first column), Grad-RAM for the manual ROI method prediction (second column), input image for the automatic ROI method (third column), and Grad-RAM for automatic ROI method prediction (fourth column) are shown.

ROI method reflected the positive effect of the ROI, focusing on enhancing the accuracy of the deep-learning model.

Five-fold cross-validation was performed to generalize the results of our age evaluation model. An IoU score above 0.5 is generally considered to denote good prediction. A previous study conducted cervical segmentation for CVM classification with lateral cephalogram using an ROI detector and U-Net model in conjunction, yielding an average IoU score of 0.918, which was slightly higher than the 0.913 value of this study.[19] As precise ROI segmentation improves image reliability,[29] segmentation of the cervical vertebrae in this study could have improved the prediction results. Our study, which used the cervical vertebrae, showed a relatively more accurate result than that of existing deep-learning studies that used pelvic X-rays (RMSE: 1.30 years)[30] and dental panoramic radiograph (MAE: 0.826 years for 2−11 years, and 1.229 years for 12−18 years) for age estimation.[31] A strong correlation was also observed between the real and predicted bone age obtained using the automatic and manual ROI methods in this study (the coefficients of determination were 0.983 and 0.934, respectively).

Deep-learning has shown excellent performance in image classification, detection, and segmentation; however, it is not easy to understand the decision-making process that takes place inside a deep neural network. Thus, the deep-learning process is often compared to a "black box". Several approaches have been proposed to investigate the optimal information for classification using neural networks.[32,33] We used the Grad-RAM technique[26] to generate heatmaps to determine the part of an image that was locally discriminative for age regression. This confirmed that there is a difference in the heatmap generation pattern according to the ROI segmentation.

This study had several limitations. First, the small sample size affects the performance of the deep-learning model, and participants are limited to growing Korean children. It is well known that the median bone age of patients of the same sex in a group is identical. Based on this notion, the Greulich−Pyle method, which evaluates bone age based on median hand X-ray images, has been used worldwide.[34] Therefore, based on this study, if a large multi-center data is trained in the future, automatic evaluation of bone age will be possible using reference radiographs of the cervical vertebrae corresponding to the median age. Second, although our model segmented whole cervical vertebrae including spinous processes shown in lateral cephalogram to estimate bone age, there is a difference in each lateral cephalogram. In addition, two-dimensional lateral cephalogram may be afflicted by artefacts such as magnification errors, patients with craniofacial asymmetry and the superimposition of anatomical structures.[35] This can lead to difficulty in evaluating craniofacial morphology. In the future, reliable three-dimensional cephalometric computed tomography analysis can aid in the creation of a deep-learning model with better performance. Furthermore, it is necessary to validate the results using comparative studies of deep-learning models and humans.

In this study, we proposed a novel method to estimate bone age through automatic cervical vertebrae segmentation from lateral cephalograms. We found that bone age can be estimated with high accuracy from the cervical vertebrae using lateral cephalogram. The results of this study may encourage clinicians in the medical and dental fields to use lateral cephalogram for skeletal maturity evaluation without the routine use of hand-wrist radiograph.

Declaration of competing interest

All authors have no conflict of interest relevant to this article.

Acknowledgements

References

1. Byun B-R, Kim Y-I, Yamaguchi T, et al. Quantitative skeletal maturation estimation using cone-beam computed tomography-generated cervical vertebral images: a pilot study in 5-to 18-year-old Japanese children. *Clin Oral Invest* 2015;19: 2133−40.

2. Thevissen P, Kaur J, Willems G. Human age estimation combining third molar and skeletal development. *Int J Leg Med* 2012;126:285−92.

3. Hauspie RC, Cameron N, Molinari L, eds. *Methods in human growth research*. Cambridge: Cambridge University Press, 2004.

4. Cameriere R, Giuliodori A, Zampi M, et al. Age estimation in children and young adolescents for forensic purposes using fourth cervical vertebra (C4). *Int J Leg Med* 2015;129:347−55.

5. Caldas MP, Ambrosano GMB, Neto FH. New formula to objectively evaluate skeletal maturation using lateral cephalometric radiographs. *Braz Oral Res* 2007;21:330−5.

6. Cericato G, Bittencourt M, Paranhos L. Validity of the assessment method of skeletal maturation by cervical vertebrae: a systematic review and meta-analysis. *Dentomaxillofacial Radiol* 2015;44:20140270.

7. Baccetti T, Franchi L, McNamara Jr JA. The cervical vertebral maturation (CVM) method for the assessment of optimal treatment timing in dentofacial orthopedics. *Semin Orthod* 2005;11:119−29.

8. Predko-Engel A, Kaminek M, Langova K, Kowalski P, Fudalej P. Reliability of the cervical vertebrae maturation (CVM) method. *Bratisl Lek Listy* 2015;116:222−6.

9. Santiago RC, de Miranda Costa LF, Vitral RWF, Fraga MR, Bolognese AM, Maia LC. Cervical vertebral maturation as a biologic indicator of skeletal maturity: a systematic review. *Angle Orthod* 2012;82:1123−31.

10. Lamparski DG. *Skeletal age assessment utilizing cervical vertebrae*. MSD thesis. Pittsburgh: University of Pittsburgh, 1972.

11. O'Reilly MT, Yanniello GJ. Mandibular growth changes and maturation of cervical vertebrae—a longitudinal cephalometric study. *Angle Orthod* 1988;58:179−84.

12. Hassel B, Farman AG. Skeletal maturation evaluation using cervical vertebrae. *Am J Orthod Dentofacial Orthop* 1995;107: 58–66.

13. Liu N. *Chronological age estimation of lateral cephalometric radiographs with deep learning.* 2021. arXiv:2101.11805.

14. Lee J, Jun S, Cho Y, et al. Deep learning in medical imaging: general overview. *Korean J Radiol* 2017;18:570–84.

15. Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans Med Imag* 2016;35:1207–16.

16. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115–8.

17. Kim J, Shim W, Yoon H, et al. Computerized bone age estimation using deep learning based program: evaluation of the accuracy and efficiency. *AJR Am J Roentgenol* 2017;209: 1374–80.

18. Lee BD, Lee MS. Automated bone age assessment using artificial intelligence: the future of bone age assessment. *Korean J Radiol* 2021;22:792–800.

19. Kim EG, Oh IS, So JE, et al. Estimating cervical vertebral maturation with a lateral cephalogram using the convolutional neural network. *J Clin Med* 2021;10:5400.

20. Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. *Encoder-decoder with atrous separable convolution for semantic image segmentation.* 2018. arXiv:1802.02611.

21. Makaremi M, Lacaule C, Mohammad-Djafari A. Deep learning and artificial intelligence for the determination of the cervical vertebra maturation degree from lateral radiography. *Entropy* 2019;21:1222.

22. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-resnet and the impact of residual connections on learning. *AAAI* 2012;4:4278–84.

23. Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis* 2015;115:211–52.

24. Kingma DP, Ba J. *Adam: a method for stochastic optimization.* 2014. arXiv:1412.6980.

25. Csurka G, Larlus D, Perronnin F, Meylan F. What is a good evaluation measure for semantic segmentation? *Bmvc* 2013;2013.

26. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. *Grad-cam: visual explanations from deep networks via gradient-based localization.* 2016. arXiv:1610.02391v4.

27. McNamara JA, Franchi L. The cervical vertebral maturation method: a user's guide. *Angle Orthod* 2018;88:133–43.

28. Gulsahi A, Cehreli SB, Galić I, Ferrante L, Cameriere R. Age estimation in Turkish children and young adolescents using fourth cervical vertebra. *Int J Leg Med* 2020;134:1823–9.

29. Lee H, Tajmir S, Lee J, et al. Fully automated deep learning system for bone age assessment. *J Digit Imag* 2017;30:427–41.

30. Li Y, Huang Z, Dong X, et al. Forensic age estimation for pelvic X-ray images using deep learning. *Eur Radiol* 2019;29:2322–9.

31. Kim J, Bae W, Jung K, Song I. Development and validation of deep learning-based algorithms for the estimation of chronological age using panoramic dental x-ray images. *Proc Mach Learn Res* 2019. https://openreview.net/revisions?id=BJg 4tI2VqV [Accessed 1 May 2022].

32. Zeiler MD, Fergus R. *Visualizing and understanding convolutional networks.* 2013. arXiv:1311.2901.

33. Varshosaz M, Ehsani S, Nouri M, Tavakoli MA. Bone age estimation by cervical vertebral dimensions in lateral cephalometry. *Prog Orthod* 2012;13:126–31.

34. Bull R, Edwards P, Kemp P, Fry S, Hughes I. Bone age assessment: a large scale comparison of the Greulich and Pyle, and Tanner and Whitehouse (TW2) methods. *Arch Dis Child* 1999;81:172–3.

35. Pinheiro M, Ma X, Fagan MJ, et al. A 3D cephalometric protocol for the accurate quantification of the craniofacial symmetry and facial growth. *J Biol Eng* 2019;13:42.