PAEDIATRIC

Assessment of rapidly advancing bone age during puberty on elbow radiographs using a deep neural network model

Kyung-Sik Ahn[1] · Byeonguk Bae[2] · Woo Young Jang[3] · Jin Hyuck Lee[4] · Saelin Oh[1] · Baek Hyun Kim[5] ·
Si Wook Lee[6] · Hae Woon Jung[7] · Jae Won Lee[2] · Jinkyeong Sung[2] · Kyu-Hwan Jung[2] · Chang Ho Kang[1] ·
Soon Hyuck Lee[3]

Abstract

Objectives Bone age is considered an indicator for the diagnosis of precocious or delayed puberty and a predictor of adult height. We aimed to evaluate the performance of a deep neural network model in assessing rapidly advancing bone age during puberty using elbow radiographs.

Methods In all, 4437 anteroposterior and lateral pairs of elbow radiographs were obtained from pubertal individuals from two institutions to implement and validate a deep neural network model. The reference standard bone age was established by five trained researchers using the Sauvegrain method, a scoring system based on the shapes of the lateral condyle, trochlea, olecranon apophysis, and proximal radial epiphysis. A test set (n = 141) was obtained from an external institution. The differences between the assessment of the model and that of reviewers were compared.

Results The mean absolute difference (MAD) in bone age estimation between the model and reviewers was 0.15 years on internal validation. In the test set, the MAD between the model and the five experts ranged from 0.19 to 0.30 years. Compared with the reference standard, the MAD was 0.22 years. Interobserver agreement was excellent among reviewers (ICC: 0.99) and between the model and the reviewers (ICC: 0.98). In the subpart analysis, the olecranon apophysis exhibited the highest accuracy (74.5%), followed by the trochlea (73.7%), lateral condyle (73.7%), and radial epiphysis (63.1%).

Conclusions Assessment of rapidly advancing bone age during puberty on elbow radiographs using our deep neural network model was similar to that of experts.

Key Points

• *Bone age during puberty is particularly important for patients with scoliosis or limb-length discrepancy to determine the phase of the disease, which influences the timing and method of surgery.*

• *The commonly used hand radiographs–based methods have limitations in assessing bone age during puberty due to the less prominent morphological changes of the hand and wrist bones in this period.*

• *A deep neural network model trained with elbow radiographs exhibited similar performance to human experts on estimating rapidly advancing bone age during puberty.*

Keywords Puberty · Elbow · Artificial intelligence

Kyung-Sik Ahn and Byeong-Uk Bae contributed equally to this work.

✉ Woo Young Jang
  opmanse@gmail.com

1 Department of Radiology, Korea University Anam Hospital, Seoul, Republic of Korea

2 VUNO, Inc., Seoul, Republic of Korea

3 Department of Orthopedic Surgery, Korea University Anam Hospital, 73, Goryeodae-ro, Seongbuk-gu, Seoul 02841, Republic of Korea

4 Department of Sports Medicine, Anam Hospital, Korea University College of Medicine, Seoul, Republic of Korea

5 Department of Radiology, Korea University Ansan Hospital, Gyeonggi-do, Republic of Korea

6 Department of Orthopedic Surgery, Keimyung University, School of Medicine, Dongsan Medical Center, Daegu, Republic of Korea

7 Department of Pediatrics, Kyung Hee University Hospital, Seoul, Republic of Korea

⌂ Springer

Abbreviations

AP	Anteroposterior
DICOM	Digital Imaging and Communications in Medicine
GP	Greulich–Pyle
ICC	Intraclass coefficient
LDL	Label distribution learning
MAD	Mean absolute difference
PHV	Peak height velocity
RMSE	Root mean square error
ROI	Region of interest
SD	Standard deviation

Introduction

Bone age has been widely used to diagnose and treat patients with precocious or delayed puberty and is a predictor of adult height [1]. In the pubertal period, bone age advances rapidly during the growth spurt. During puberty, it is particularly important to determine bone age in patients with scoliosis or limb-length discrepancy, since it is associated with the phase of the disease (i.e., acceleration or deceleration), which influences the timing and method of surgery. The Greulich–Pyle (GP) and Tanner–Whitehouse methods applied to hand radiographs are the most commonly used in this field; however, they are limited in assessment of bone age during puberty due to less prominent morphological changes in hand and wrist bones during this period. Little et al. reported that the use of a GP atlas would not improve the accuracy of the prediction of limb-length discrepancy [2]. Furthermore, the GP atlas is not regularly divided into 6-month intervals and exhibits a relatively large interobserver error during the pubertal growth spurt [3].

Sauvegrain et al. introduced a method to determine bone age using elbow anteroposterior (AP) and lateral radiographs based on the shapes of the lateral condyle, trochlea, olecranon apophysis, and proximal radial epiphysis [4]. The Sauvegrain method is used on elbow bones, which undergo prominent changes during puberty; it has excellent interobserver correlation and reproducibility in assessing bone age during puberty and is useful in estimating peak height velocity (PHV), an important factor for predicting disease progression and determining surgery timing for scoliosis or limb-length discrepancy [3]. Additionally, the Sauvegrain method can be used to assess skeletal maturity in regular 6-month intervals during the PHV phase [5].

Several artificial intelligence systems for bone age assessment have been reported. Most of these systems have been developed based on hand and wrist radiographs [6–10]. Although these systems can improve efficiency in clinical routines by reducing reading time and supporting decision making, they are limited in the accurate assessment of bone age during puberty due to less prominent morphological changes in hand and wrist bones during this period compared to those of the elbow, as well as the lack of correlation between artificial intelligence systems and PHV. To our knowledge, this is the first study to estimate bone age based on elbow radiographs using a deep neural network model.

We aimed to evaluate the performance of our deep neural network model in assessing rapidly advancing bone age during puberty on elbow radiographs.

Materials and methods

Training/validation set

The institutional review board of two tertiary hospitals approved this study. Elbow AP and lateral pairs of radiographs of pubertal individuals (9–16 years for boys and 8–14 years for girls) from March 2004 to September 2018 were obtained from the two institutions. Exclusion criteria included clinical history of delayed or precocious puberty; radiographic evidence of fracture; tumor, deformity, or surgery around the elbow; and poor image quality, including inappropriately positioned elbows and images containing splints. Experienced radiologists judged the appropriateness of the image quality. In all, 4437 AP and lateral pairs of elbow radiographs were used for training and internal validation. The images were anonymized and in Digital Imaging and Communications in Medicine (DICOM) format.

The reference standard bone age was established by consensus interpretation of five trained researchers (two physical therapists with doctoral degrees and three nurses with master's degrees) under the supervision of an experienced pediatric orthopedic surgeon using the Sauvegrain method, which is a scoring system based on the shapes of the lateral condyle (1–9 points), trochlea (1–5 points), olecranon apophysis (1–7 points), and proximal radial epiphysis (1–6 points) [8]. The scoring system was the same irrespective of sex; however, the conversion ratio of the score to bone age differed by sex. All labelers were blinded to the sex and chronological age of the subjects. A score of 0 was applied for each portion of the bone in which the secondary ossification center was not visible, slightly broadening the age range. During labeling, different colored lines were drawn along the boundary of the secondary ossification centers of the four portions of the elbow to help improve identification of each bone portion using the algorithm. Online point and click templates that support the paint brush tool and automatic total score calculation were introduced for efficient labeling (Fig. 1).

Test set

An anonymized test set in the DICOM format was obtained from an outside tertiary hospital. A total of 141 pairs of AP
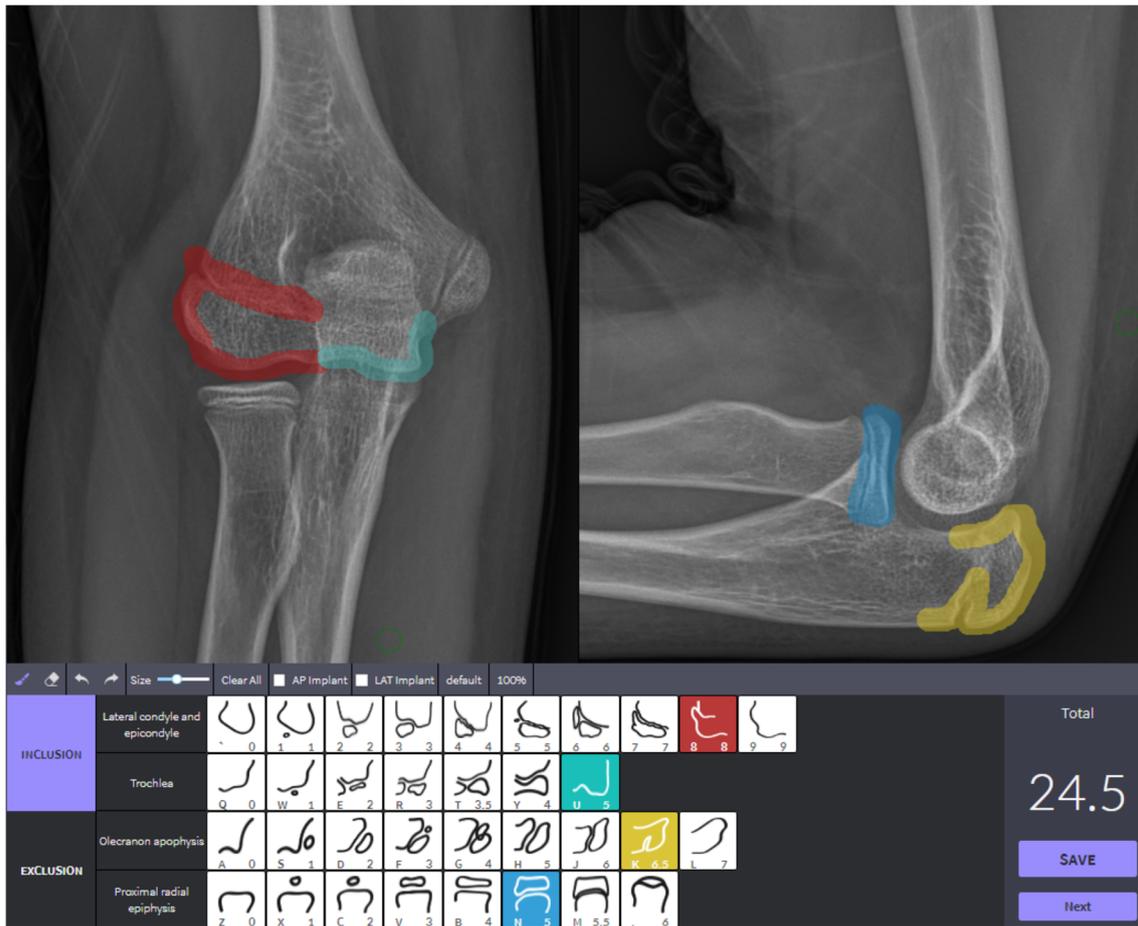
Fig. 1 Online point and click template for efficient labeling. After authorized log-in, the labeler examined the anteroposterior and lateral pairs of the radiographs and clicked on the appropriate score for each portion of the bone at the lower side of the screen. By clicking the score, the paint brush tool of the corresponding color (red, lateral epicondyle; green, trochlea; yellow, olecranon apophysis; and blue, radial epiphysis) was activated, and the boundary of the location was marked roughly by the labeler. After assigning scores for the four parts, the total score was automatically calculated and displayed on the right lower side of the screen

and lateral elbow radiographs were obtained after applying the same exclusion criteria as in the training set. Five reviewers (a musculoskeletal radiologist with 16 years of experience, an orthopedic surgeon with 12 years of experience, a pediatric radiologist with 7 years of experience, a fourth-year orthopedic resident, and an orthopedic researcher with 10 years of experience) independently graded bone age on elbow radiographs using the Sauvegrain method. For the final comparison, the reference standard was re-established based on the majority decision of the five reviewers. If more than three reviewers assigned the same score, the image was consolidated as a reference standard. In cases with no agreed-upon score, a consensus interpretation was obtained from the three reviewers.

Model implementation

Based on the Sauvegrain method, the proposed bone age assessment framework applied to AP and lateral elbow radiograph images uses region of interest (ROI) extractors to initially extract four bone patches and local classifier models to estimate their corresponding scores. Subsequently, the sum of these scores is converted to bone age.

ROI extractor The configuration of the ROI extractor was inspired by the pose estimation network HRNet-32 by Escobar [11] that uses hand detection and hand pose estimation of various hand positions to assess bone age. The model is trained to localize key points indicating the centers of the ROI patches from the elbow images using patch annotations, as illustrated in Fig. 1. The bounding box containing the segmentation area was used as the patch ROI annotation. Each output of the ROI extractor is a spatial map, with model confidence describing the likelihood of a key point being located at each pixel. After training, ROIs were automatically extracted using a fixed window centered at the maximum probability key points obtained from each heatmap.

Local classifier The local classifiers showed a convolutional neural network architecture and were trained to estimate the maturity stage, which can be mapped using a rule-based bijective function to compute the scores. Each bone is related to an ordered set of maturity stages, and the output of each local classifier is a distribution over the possible maturity scores. The models were obtained by minimizing the loss with a target one-hot vector (or dirac-delta distribution), indicating the maturity stage of each bone. In our experiments, we used the state-of-the-art classification network EfficientNet-B4 [12] for each local classifier by modifying the number of output nodes to a set number. The estimated maturity stage was then mapped to its corresponding score, and the sum of the scores was converted to the predicted bone age.

Implementation details All models were implemented using an open-source machine library (PyTorch version 1.2.0) [13]. All elbow images for the ROI extractor were extracted from DICOM files, resized to 512×512 with padding, and the aspect ratio was maintained. Input ROI patches for local classifiers were cropped to 512×512 size from the raw elbow image to maintain high-resolution information. To normalize variable pixel intensity scales within the X-ray radiographs, we performed per image standardization so that each image had a mean of 0 and a variance of 1. During training, each training example was rotated randomly between $-15°$ and $15°$, shifted randomly to -32 and 32 pixels, and flipped horizontally, with 50% probability for both ROI extractor and local classifier. The model parameters were optimized using the RAdam optimizer [14], with the hyperparameter weight decay set to 0.0001, beta1 set to 0.9, and beta2 set to 0.9.

The ROI extractor was trained using a pixel-wise binary classification loss, which was computed by comparing the predicted probability of the corresponding pixel as the key point and the true label of the key point. For the local classifier, we used joint learning of label distribution and expectation regression, which showed state-of-the-art performance in facial age estimation [15]. Label distribution learning (LDL) is used to utilize the correlation among adjacent labels, as in the case of age or maturity estimation [16, 17], and the expectation regression is used to directly minimize the discrepancy between the estimated score and the label. The Kullback–Leibler divergence loss and regression loss were used for two learning objectives with equal weights.

In general, it is difficult to optimize a convolutional neural network model from scratch using a small amount of data. For this reason, we initialized the weights of the encoder layers of HRNet-32 and EfficientNet using COCO pose estimation and ImageNet pretrained weights, respectively.

Each ROI extractor was trained for 200 epochs with a learning rate of 0.001 over. The learning rate was scheduled using the cosine annealing scheduler implemented in PyTorch with a hyperparameter Tmax set to total epochs. Local classifiers can also be learned in the same scheduler; however, the best results are obtained at 100 epochs and a learning rate of 0.0003. The best model was selected using the validation set with the best performance. A diagram of model implementation is shown in Fig. 2.

Statistical analysis

The difference between model estimation and expert results was reviewed with Bland–Altman plots and compared using the mean absolute difference (MAD) or root mean square error (RMSE). The results were separately analyzed with the scores of each elbow part and used to calculate the final bone age. Subsequently, the accuracy of each portion of the elbow was derived. Interobserver agreement among human experts and between the model and human reviewers was evaluated using the intraclass coefficient (ICC). Interobserver agreement was categorized as follows: 0–0.20, poor; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–1.00, excellent.

Results

Demographic data of the subjects

The subjects comprised 3162 boys and 1275 girls. Labeled bone age was 12.54 ± 2.76 years in the training/validation set and 12.07 ± 1.48 years in the test set. The distribution of the data is shown in Fig. 3.

Internal validation results

Internal validation was performed using fivefold cross-validation (80% training and 20% validation). On internal validation, the MAD of bone age between model estimation and reviewers was 0.15 years, with a score of 0.44. Examining the scores of each portion of the elbow, accuracy was best at the olecranon (90.73%), followed by the lateral condyle (87.62%), trochlea (86.45%), and radial epiphysis (81.21%), with a MAD of 0.108, 0.153, 0.144, and 0.167, respectively. The radial epiphysis was assessed on lateral views since it is associated with a more accurate assessment than the AP view (78.88% and 77.88%, respectively) in preliminary internal validation.

Test set results

The MAD of bone age between the model and the five human experts ranged from 0.19 to 0.30 years. The MAD of the total score ranged from 0.79 to 1.11, and RMSE ranged from 1.33 to 1.73. The MAD among reviewers ranged from 0.19 to 0.28
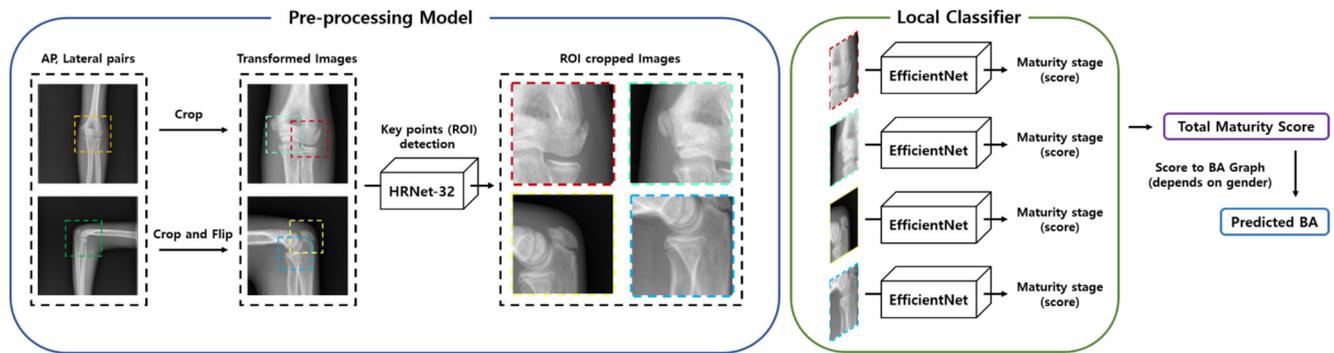
Fig. 2 Diagrammatic illustration of pre-processing and maturity score prediction model

years, with excellent interobserver agreement (ICC: 0.993, p = 0.000). The MADs of the reviewers and the model are presented in Table 1. For the lateral epicondyle, trochlea, olecranon apophysis, and radial epiphysis, agreement was also high among reviewers (ICC = 0.988, 0.985, 0.993, and 0.967, respectively; all $p < 0.001$).

In the comparison between the model and reference standard derived by the majority decision, MAD was 0.22 in years, with a score of 0.81. Agreement between the model and the reviewers was also high (ICC: 0.98, $p < 0.001$). The performance of the model is listed in Table 2. The Bland–Altman plot of model estimation on the reference standard bone age is presented in Fig. 4.

Subpart analysis was performed to investigate the accuracy of the model for each portion of the elbow. Considering the score with the highest output probability as the only correct answer (top 1 accuracy), the olecranon apophysis exhibited

the highest accuracy (74.5%), followed by the trochlea (73.7%), lateral condyle (73.7%), and proximal radial epiphysis (63.1%); also, the percentage accuracy increased to 90.7%, 90.7%, 86.5%, and 78.7%, respectively, after including the first and second choices as correct answers (top 2 accuracy). The top 1 to top 3 accuracies are presented in Table 3.

To ensure that the models had learning pertinent features, we generated a class activation map using gradCAM [15]. By aggregating intermediate feature maps using the gradient from the final layer of the network as the weighting factor, the heatmap had higher values in the portion of the bone associated with high impact on the model prediction. The heatmap was then converted to a color image using a color scheme, upsampled to 512×512 pixels, and overlaid on the input image. A sample heatmap of the subparts from the test set generated by gradCAM is shown in Fig. 5.

Fig. 3 Distributions of bone ages of the training/validation (n = 4437) and test (n = 141) sets
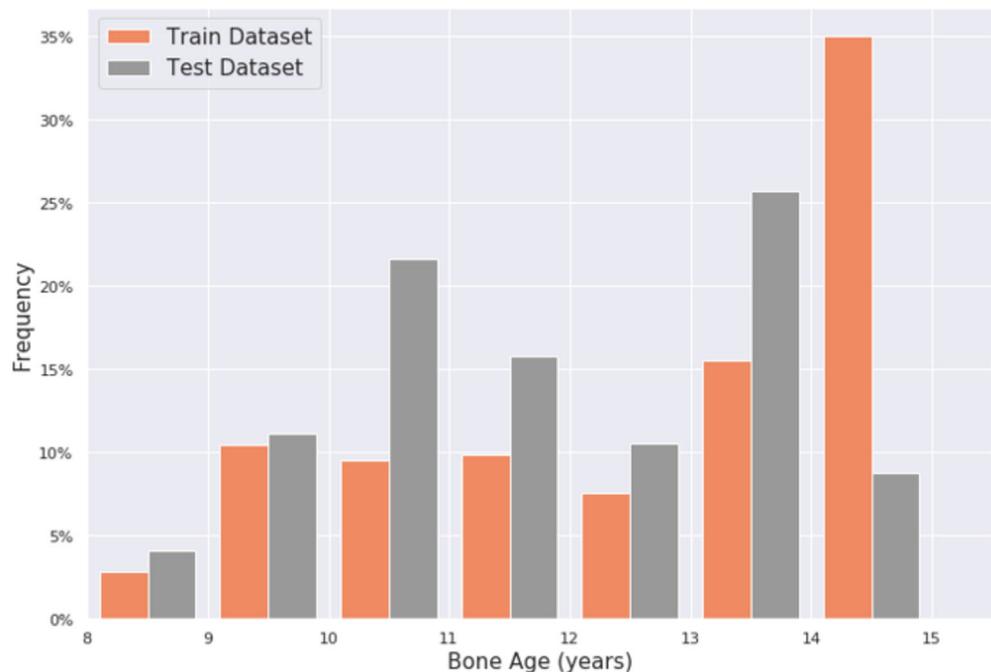
Table 1 The mean absolute differences among the five reviewers and the model

	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Model
Reviewer 1		0.2480	0.2392	0.2214	0.2800	0.3020
Reviewer 2			0.2061	0.2039	0.1894	0.2292
Reviewer 3				0.2033	0.2356	0.2467
Reviewer 4					0.2670	0.2736
Reviewer 5						0.1942

Note: Numbers are years

Outlying case review

Ten of 141 cases were outliers, with a 1.96 standard deviation (SD) in the test set (4 girls and 6 boys). Five cases involved underestimation (four girls and one boy), and five cases involved overestimation (five boys). Regarding the total scores, the differences were 5 points in 3 cases, 3 points in 3 cases, 2 points in 3 cases, and 1 point in 1 case. For age in years, differences ranged from 0.84 to 1.14 years (mean ± SD, 0.88 ± 0.18). In the subpart review, > 2 points difference at the lateral epicondyle was observed in 2 cases (2 points), at the trochlea in 1 case (4 points), at the olecranon in 2 cases (2 and 3 points, respectively), and at the radial epiphysis in 1 case (2 points). There was no misinterpretation case, with a score of 0 to the full score, or vice versa.

Discussion

We developed a deep-learning model for estimating bone age on elbow radiographs using the Sauvegrain method and obtained results comparable to those of experts. The consensus MAD of reviewers was 0.22 in years, and the agreement between model and reviewers was high (ICC: 0.98, $p < 0.001$).

Bone age determination using elbow radiography is more precise during puberty due to prominent morphological changes in elbow bones compared to hand bones. The Sauvegrain method uses scores of five ossification centers of the elbow based on the provided visual score template, irrespective of sex. Then, the score is converted to bone age based on the conversion ratio, which is dependent on sex [4]. The results provide data on bone age, with excellent reproducibility and a regular 6-month scale. Therefore, the Sauvegrain method is reliable, simple, and reproducible [3, 5]. Another advantage of this method is its reliable correlation with the timing of the PHV. For orthopedic surgical treatment, evaluation of PHV and exact bone age assessment at 6-month intervals in adolescence are important [18, 19]. In patients with idiopathic scoliosis, the timing of PHV is useful, since it is indicative of the likelihood of disease progression to a phase in which spinal arthrodesis is required [20]. In patients with lower limb length discrepancy, since there is a short time period of remaining growth after onset of puberty (approximately 2.5 years), accurate evaluation of bone age with elbow and hand radiographs is necessary to determine the timing of epiphysiodesis [18].

Considering the complex and time-consuming nature of bone age assessment, automation of the process has been attempted in the past [21, 22]. Many convolutional neural network algorithms have been developed with the advent of deep-learning technology to assess bone age using hand radiographs [6–10]. Many attempts have been made to improve the accuracy of the model. Recent results demonstrated a MAD of 0.33 years when combining the models; however,

Table 2 Comparison of the differences between the model and reviewers

	Years		Score				
	MAD	RMSE	Total	Lateral condyle	Trochlea	Olecranon apophysis	Radial epiphysis
Reviewer 1	0.1691	0.3189	0.6277	0.2553	0.1135	0.1277	0.2163
Reviewer 2	0.1467	0.2785	0.5355	0.1348	0.1809	0.1844	0.1773
Reviewer 3	0.1683	0.3154	0.6135	0.2979	0.2021	0.1631	0.1206
Reviewer 4	0.1614	0.3024	0.6135	0.2199	0.2270	0.1489	0.1418
Reviewer 5	0.1594	0.3061	0.5461	0.0851	0.3050	0.0887	0.2872
Model	0.2253	0.3576	0.812	0.3120	0.3049	0.3049	0.3297

Note: Numbers are years or Sauvegrain scores. *MAD*, mean absolute difference; *RMSE*, root mean square error
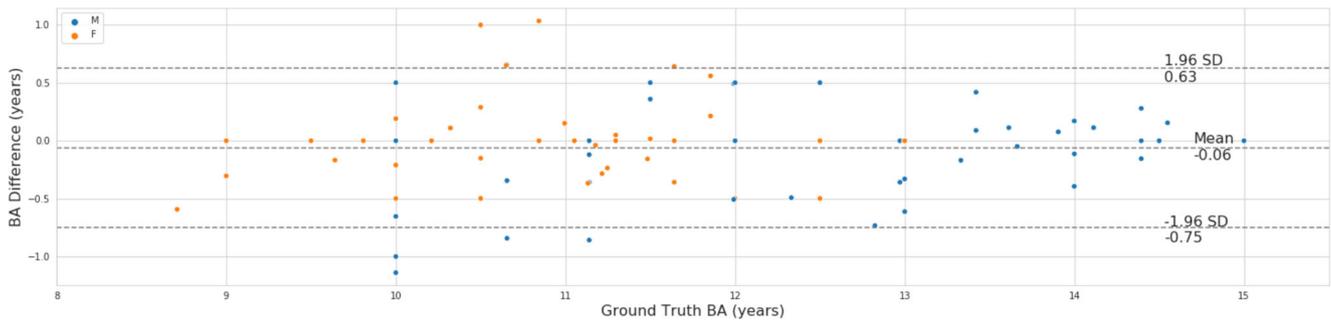
Fig. 4 Bland–Altman plot showing the difference between model and reference standards derived from reviewers (M, male; F, female)

results obtained during the pubertal period were not specified [23, 24]. To our knowledge, there is no automated bone age assessment model using elbow radiographs that target the pubertal period [10]. Our model using elbow radiographs showed a MAD of 0.22 years and RMSE of 0.36 years. Although the Sauvegrain method supports a narrow age range in the pubertal period, its accuracy appears to be high. This indicates that our model is helpful in identifying the appropriate time for spinal arthrodesis or lower limb epiphysiodesis in adolescents. Although the overall performance was similar to that of human experts, there were several cases with incorrect estimations greater than 1 SD. We attributed misinterpretations of undetected apophysis/epiphysis (score 0) as full score ossification or vice versa, since inexperienced reviewers may sometimes confuse these two portions of the bone. However, reviewers can rectify this error by referring to the status of other ossification centers of the elbow. In our outlying case review, the difference between point 0 and the full score was clear. Only gradual 1- to 4-point differences were found in the subparts of the bone. The model may detect the status of other ossification centers based on some of the heatmaps tagged on other ossification centers included in the cropped ROI (Fig. 5). In our subpart accuracy analysis, the olecranon exhibited the highest top 1 and top 2 accuracy values. This may contribute to the evidence supporting the olecranon method, a simplified version of the Sauvegrain method, using only the olecranon during the accelerating phase of PHV [25]. Although the radial epiphysis could be evaluated on both AP and lateral views, we assessed the score based on the lateral view due to its high accuracy in internal validation.

Our study has several limitations. First, the 4437 pairs of radiographs used in the training set may be small compared to the amount used in other radiograph-based models, regardless of the narrow age range selected. We modified the method of Demeglio et al [3] by adding a 0 point in four anatomic locations to maximize the data size. However, we only used cases with a total score greater than 8 points based on the Sauvegrain method, which allows for expansion of the age range. In addition, we included available right elbow radiographs, despite the fact that the left elbow is recommended when using the Sauvegrain method. Based on image analysis, we assumed that side differences could be resolved by applying various image augmentations, including horizontal flipping. Additionally, there was an imbalance in data distribution, and specific scores of the subparts were relatively limited, possibly influencing the accuracy of the model. However, it is uncertain whether this resulted from insufficient data or the natural distribution of advancing bone age. Based on the Bland–Altman plot (Fig. 3), outlying cases over 1 SD mostly presented in the period between 10 and 11 years of age, during which the data were relatively sufficient. Finally, two test set reviewers indirectly and directly participated in training set labeling, potentially influencing the test results.

In conclusion, the results obtained using a deep neural network model trained with elbow radiographs were similar to those of experts in estimating rapidly advancing bone age during puberty. Automated determination of bone age of the elbow may be useful in estimating bone age during puberty based on hand bones and could potentially be useful to determine the timing of surgery in scoliosis or limb-length discrepancy.

Table 3 Accuracy of the model based on the portion of the elbow

	Lateral condyle	Trochlea	Olecranon apophysis	Radial epiphysis
Top 1 accuracy	73.7	73.7	74.5	63.1
Top 2 accuracy	86.5	90.7	90.7	78.7
Top 3 accuracy	92.9	95.7	95.0	91.4

Note: Numbers are percentages (%). Top N accuracy refers to the frequency at which the reference standard grade is among the highest N prediction probabilities for each location

Fig. 5 Heatmaps of elbow radiographs at various bone ages for each ossification center (left to right is low to high score). Some of the heatmaps tagged in other ossification centers were included in the cropped region of interest
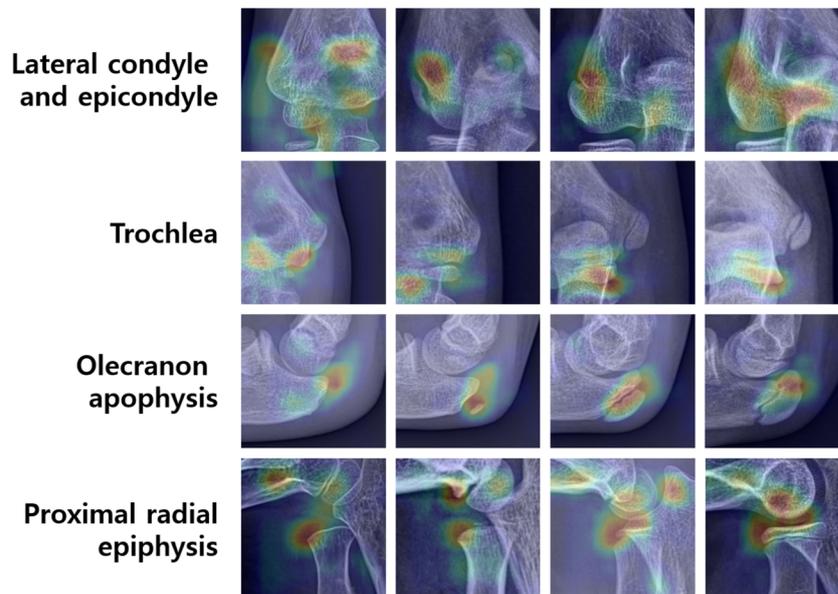
Lateral condyle and epicondyle

Trochlea

Olecranon apophysis

Proximal radial epiphysis

Declarations

Guarantor The scientific guarantor of this publication is Woo Young Jang.

Conflict of interest The authors of this manuscript declare relationships with the following company: VUNO Inc.

Statistics and biometry Two of the authors (Byeong-Uk Bae and Kyu-Hwan Jung) have statistical expertise.

Informed consent Written informed consent was waived by the Institutional Review Board because this study is a retrospective analysis.

Ethical approval Institutional Review Board approval was obtained.

Methodology
• Retrospective
• Cross-sectional study
• Multicenter study

References

1. Satoh M (2015) Bone age: assessment methods and clinical applications. Clin Pediatr Endocrinol 24:143–152. https://doi.org/10.1297/cpe.24.143
2. Little DG, Nigo L, Aiona MD (1996) Deficiencies of current methods for the timing of epiphysiodesis. J Pediatr Orthop 16:173–179. https://doi.org/10.1097/00004694-199603000-00007
3. Diméglio A, Charles YP, Daures JP, de Rosa V, Kaboré B (2005) Accuracy of the Sauvegrain method in determining skeletal age during puberty. J Bone Joint Surg Am 87:1689–1696. https://doi.org/10.2106/jbjs.D.02418
4. Sauvegrain J, Nahum H, Bronstein H (1962) Study of bone maturation of the elbow. Ann Radiol (Paris) 5:542–550
5. Hans SD, Sanders JO, Cooperman DR (2008) Using the Sauvegrain method to predict peak height velocity in boys and girls. J Pediatr Orthop 28:836–839. https://doi.org/10.1097/BPO.0b013e31818ee3c4
6. Lee H, Tajmir S, Lee J et al (2017) Fully automated deep learning system for bone age assessment. J Digit Imaging 30:427–441. https://doi.org/10.1007/s10278-017-9955-8
7. Kim JR, Shim WH, Yoon HM et al (2017) Computerized bone age estimation using deep learning based program: evaluation of the accuracy and efficiency. AJR Am J Roentgenol 209:1374–1380. https://doi.org/10.2214/ajr.17.18224
8. Larson DB, Chen MC, Lungren MP, Halabi SS, Stence NV, Langlotz CP (2018) Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs. Radiology 287:313–322. https://doi.org/10.1148/radiol.2017170236
9. Reddy NE, Rayan JC, Annapragada AV et al (2020) Bone age determination using only the index finger: a novel approach using a convolutional neural network compared with human radiologists. Pediatr Radiol 50:516–523. https://doi.org/10.1007/s00247-019-04587-y
10. Dallora AL, Anderberg P, Kvist O, Mendes E, Diaz Ruiz S, Sanmartin Berglund J (2019) Bone age assessment with various machine learning techniques: a systematic literature review and meta-analysis. PLoS One 14:e0220242. https://doi.org/10.1371/journal.pone.0220242
11. Escobar M, González C, Torres F, Daza L, Triana G, Arbeláez P (2019). Hand pose estimation for pediatric bone age assessment. Paper presented at the Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, Cham, pp 531–539
12. Tan M, Le Q (2019) EfficientNet: rethinking model scaling for convolutional neural networks. In: Kamalika C, Ruslan S (eds) Proceedings of the 36th International Conference on Machine Learning. PMLR, Proceedings of Machine Learning Research, pp 6105–6114

13. Paszke A, Gross S, Massa F et al (2019) Pytorch: an imperative style, high-performance deep learning library. arXiv preprint arXiv: 1912.01703

14. Liu L, Jiang H, He P et al (2019) On the variance of the adaptive learning rate and beyond. arXiv preprint arXiv:1908.03265

15. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-cam: visual explanations from deep networks via gradient-based localization. Paper presented at the Proceedings of the IEEE International Conference on Computer Vision, pp 618–626

16. Pan H, Han H, Shan S, Chen X (2018) Mean-variance loss for deep age estimation from a face. Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 5285–5294

17. Rothe R, Timofte R, Van Gool L (2015) Dex: deep expectation of apparent age from a single image. Paper presented at the Proceedings of the IEEE international conference on computer vision workshops, pp 10–15

18. Kelly PM, Diméglio A (2008) Lower-limb growth: how predictable are predictions? J Child Orthop 2:407–415. https://doi.org/10.1007/s11832-008-0119-8

19. Chazono M, Tanaka T, Marumo K, Kono K, Suzuki N (2015) Significance of peak height velocity as a predictive factor for curve progression in patients with idiopathic scoliosis. Scoliosis 10:S5. https://doi.org/10.1186/1748-7161-10-s2-s5

20. Little DG, Song KM, Katz D, Herring JA (2000) Relationship of peak height velocity to other maturity indicators in idiopathic scoliosis in girls. J Bone Joint Surg Am 82:685–693. https://doi.org/10.2106/00004623-200005000-00009

21. van Rijn RR, Thodberg HH (2013) Bone age assessment: automated techniques coming of age? Acta Radiol 54:1024–1029. https://doi.org/10.1258/ar.2012.120443

22. Thodberg HH, Kreiborg S, Juul A, Pedersen KD (2009) The BoneXpert method for automated determination of skeletal maturity. IEEE Trans Med Imaging 28:52–66. https://doi.org/10.1109/tmi.2008.926067

23. Halabi SS, Prevedello LM, Kalpathy-Cramer J et al (2019) The RSNA pediatric bone age machine learning challenge. Radiology 290:498–503. https://doi.org/10.1148/radiol.2018180736

24. Pan I, Thodberg HH, Halabi SS, Kalpathy-Cramer J, Larson DB (2019) Improving Automated pediatric bone age estimation using ensembles of models from the 2017 RSNA machine learning challenge. Radiol Artif Intell 1:e190053. https://doi.org/10.1148/ryai.2019190053

25. Charles YP, Diméglio A, Canavese F, Daures JP (2007) Skeletal age assessment from the olecranon for idiopathic scoliosis at Risser grade 0. J Bone Joint Surg Am 89:2737–2744. https://doi.org/10.2106/jbjs.G.00124