Original Article
Yonsei Med J 2022 Jul;63(7):683-691
https://doi.org/10.3349/ymj.2022.63.7.683
Yonsei Medical Journal
YMJ
pISSN: 0513-5796 · eISSN: 1976-2437

Re-Assessment of Applicability of Greulich and Pyle-Based Bone Age to Korean Children Using Manual and Deep Learning-Based Automated Method

Jisun Hwang[1], Hee Mang Yoon[2], Jae-Yeon Hwang[3], Pyeong Hwa Kim[2], Boram Bak[4], Byeong Uk Bae[5], Jinkyeong Sung[5], Hwa Jung Kim[6], Ah Young Jung[2], Young Ah Cho[2], and Jin Seong Lee[2]

[1]Department of Radiology, Hallym University Dongtan Sacred Heart Hospital, Hwaseong;
[2]Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, Seoul;
[3]Department of Radiology, Research Institute for Convergence of Biomedical Science and Technology, Pusan National University Yangsan Hospital, College of Medicine, Pusan National University, Yangsan;
[4]University of Ulsan Foundation for Industry Cooperation, Ulsan;
[5]VUNO, Inc., Seoul;
[6]Department of Clinical Epidemiology and Biostatistics, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Korea.

Purpose: To evaluate the applicability of Greulich-Pyle (GP) standards to bone age (BA) assessment in healthy Korean children using manual and deep learning-based methods.

Materials and Methods: We collected 485 hand radiographs of healthy children aged 2–17 years (262 boys) between 2008 and 2017. Based on GP method, BA was assessed manually by two radiologists and automatically by two deep learning-based BA assessment (DLBAA), which estimated GP-assigned (original model) and optimal (modified model) BAs. Estimated BA was compared to chronological age (CA) using intraclass correlation (ICC), Bland-Altman analysis, linear regression, mean absolute error, and root mean square error. The proportion of children showing a difference >12 months between the estimated BA and CA was calculated.

Results: CA and all estimated BA showed excellent agreement (ICC ≥0.978, $p<0.001$) and significant positive linear correlations (R^2 ≥0.935, $p<0.001$). The estimated BA of all methods showed systematic bias and tended to be lower than CA in younger patients, and higher than CA in older patients (regression slopes ≤-0.11, $p<0.001$). The mean absolute error of radiologist 1, radiologist 2, original, and modified DLBAA models were 13.09, 13.12, 11.52, and 11.31 months, respectively. The difference between estimated BA and CA was >12 months in 44.3%, 44.5%, 39.2%, and 36.1% for radiologist 1, radiologist 2, original, and modified DLBAA models, respectively.

Conclusion: Contemporary healthy Korean children showed different rates of skeletal development than GP standard-BA, and systemic bias should be considered when determining children's skeletal maturation.

Key Words: Age determination by skeleton, radiography, hand bones, child, deep learning

Received: November 15, 2021 **Revised:** April 4, 2022 **Accepted:** April 5, 2022

Co-corresponding authors: Hee Mang Yoon, MD, PhD, Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, 88 Olympic-ro 43-gil, Songpa-gu, Seoul 05505, Korea.
Tel: 82-2-3010-0906, Fax: 82-2-476-4719, E-mail: espoirhm@gmail.com and
Jae-Yeon Hwang, MD, PhD, Department of Radiology, Research Institute for Convergence of Biomedical Science and Technology, Pusan National University Yangsan Hospital, College of Medicine, Pusan National University, 20 Geumo-ro, Mulgeum-eup, Yangsan 50612, Korea.
Tel: 82-55-360-1840, Fax: 82-55-360-1848, E-mail: jyhwang79@gmail.com

•The authors have no potential conflicts of interest to disclose.

INTRODUCTION

The determination of skeletal maturation in children is important for the assessment of growth disorders, endocrine problems, planning for orthopedic surgery, and non-clinical legal or forensic issues.[1,2] The Greulich-Pyle (GP) method[3] is the most commonly used method in clinical practice. The GP method utilizes the assessment of predictable serial changes of ossification centers on left hand radiographs, and is preferred by pediatric radiologists.[4] Recently, deep learning-based bone age (BA) assessment techniques have been developed to improve the low efficiency and reproducibility of manual BA reading, with a similar accuracy compared to experienced readers.[5-7] In South Korea, GP-based automated deep learning software has been developed, and is being used in the real-world clinical practice.[7]

The GP atlas was based on the data collected from North American Caucasians of good socioeconomic status between 1931 and 1942, and its applicability to modern children with improvable nutritional status and to children of different ethnicities has been questioned by many researchers. In general, racial differences in the estimated BA by GP method were observed.[8-13] Regarding contemporary and healthy Korean children, the applicability of GP method has been investigated in only a few studies to date, with children of a limited age range.[8,14] In addition, there has been no study regarding the applicability of GP-based deep learning software in healthy Korean children.

The present study aimed to evaluate the applicability of GP method to BA assessment in contemporary healthy Korean children by using manual and deep learning-based automated methods. Our hypothesis was that if clinically meaningful difference would exist between GP-based BA and CA in healthy Korean children, physicians should be aware of the limitation when assessing the developmental status of pediatric patients.

MATERIALS AND METHODS

Patients

This study was approved by the Institutional Review Board of Asan Medical Center (No. 2018-0692), and informed consent was waived due to the study's retrospective nature. The inclusion criteria were as follows: healthy children 1) aged between 2 and 17 years; and 2) who visited the emergency department and underwent left wrist and hand radiographs for trauma evaluation. We assumed that skeletal development of these children was likely to be within the normal range, and that they could represent relatively healthy children compared to children who visited our endocrinologic department to take left hand radiographs with a specific request for BA assessment. The data were consecutively collected from two tertiary hospitals, Asan Medical Center and Pusan National University Yangsan Hospital (PNUYH), in South Korea, between January 2013 and December 2017, and between December 2008 and December 2017, respectively. The exclusion criteria were as follows: 1) presumed metabolic disease (n=1); 2) bony abnormalities, including fracture (n=3), congenital anomalies (n=3), and tumors (n=1); 3) poor image quality (n=1); and/or 4) foreign children (n=0). Finally, a total of 485 radiographs were included in this study.

Deep learning-based automated bone age assessment

BA was assessed by a deep learning-based BA assessment (DLBAA) system (VUNO Med-BoneAge, version 1.1.0, VUNO, Seoul, Korea). This system has been commercially available in South Korea since May 2018 and in Europe since June 2020. The DLBAA system is designed with the convolutional neural networks (CNNs) to assess BA by months for hand radiographs. The input image is normalized by two image pre-processing methods. First, the hand region is segmented from the input image using CNNs, and the remaining background region is removed. Second, the hand pose estimation network is built to normalize diverse hand positions using a geometric transformation matrix. After pre-processing the image, the BA assessment network predicts probability values for each BA. This original model provided the probability of top three GP-assigned BAs (i.e., age intervals equal to GP atlas from 3 months to 1 year) in the order of probability.[7] The modified model is able to provide the optimal BA by using all GP-assigned BAs and their probabilities, rather than just displaying the top three GP-assigned BAs. To convert the GP-assigned BA results to the optimal BA, BA expectation regression with the softmax output for GP standards was performed. The softmax output represents the bone age distribution (the probability of belonging to all of the different BAs of GP standards), which is used to calculate the expectation of BA. The optimal BA was then calculated by weighted sum of GP-assigned BAs using the predicted probabilities as weights in the modified DLBAA model. In this study, we assessed GP-assigned BA that showed the highest probability by the original DLBBA model and optimal BA estimated by the modified DLBBA model.

Bone age assessment by radiologists

Two board-certified pediatric radiologists (J.H. with 7 years of experience and H.M.Y. with 11 years of experience) independently rated the BA of all of the hand radiographs based on the GP method without time limitation. Both radiologists had training sessions with 40 cases before starting the BA reading. The radiologists were blinded to the CA of the children examined.

Statistical analysis

The outcome of this study was to assess the difference between GP-based BA and CA in healthy Korean children. Due to a lack of the perfect ground truth in normal skeletal development, CA was inevitably set to be as a reference standard, although

https://doi.org/10.3349/ymj.2022.63.7.683

normal skeletal development may show a wide range of difference. The BAs were estimated manually by two radiologists and automatically by using two different DLBAA methods based on GP standard. The estimated BAs were compared to CA of each patient. First, to investigate the agreement between CA and BA, intraclass correlation (ICC) analysis, linear regression, and Bland-Altman analysis were performed. Second, the mean absolute error and root mean square error were calculated to estimate the difference between BA and CA. Additionally, proportions of BA showing a difference of >12 months, >18 months, and >24 months compared to CA were analyzed. The ICC values were categorized as poor (ICC <0.40), fair (ICC= 0.40–0.59), good (ICC=0.60–0.74), and excellent agreement (ICC=0.75–1.0).[15] The mean absolute error was computed as the average over the absolute differences between the estimated BA and CA of each patient.[6] The root mean square error was computed by the square root of the average of squared errors.[16] A p-value<0.05 was considered statistically significant. Repeated measures analysis of variance and pairwise comparisons were performed to compare the mean absolute error and root mean square error of the radiologists and DLBAA methods. The comparison of proportion of BA estimations >12, 18, and 24 months was done by Cochran's Q test, followed by multiple comparisons using the McNemar test with Bonferroni correction. Statistical analyses were performed with R software version 4.1.0 (R Foundation for Statistical Computing) and MedCalc software version 20.009 (MedCalc Software, Ostend, Belgium).

RESULTS

Patient characteristics

The patients' sex and age distributions are summarized in Fig. 1. In total, 485 radiographs (226 and 259 radiographs from Asan Medical Center and PNUYH, respectively) from 223 girls and 262 boys were included in this study. The mean (± SD) age of the included pediatric patients was 10.0±4.3 years (range, 2–17 years).

Concordance between chronological age and estimated bone ages

The ICC values were calculated from the data of CA and estimated BA by radiologist 1, radiologist 2, and the original and modified DLBAA models. All of the ICC values showed excellent agreement (ICC ≥0.978, all p<0.001) (Table 1).

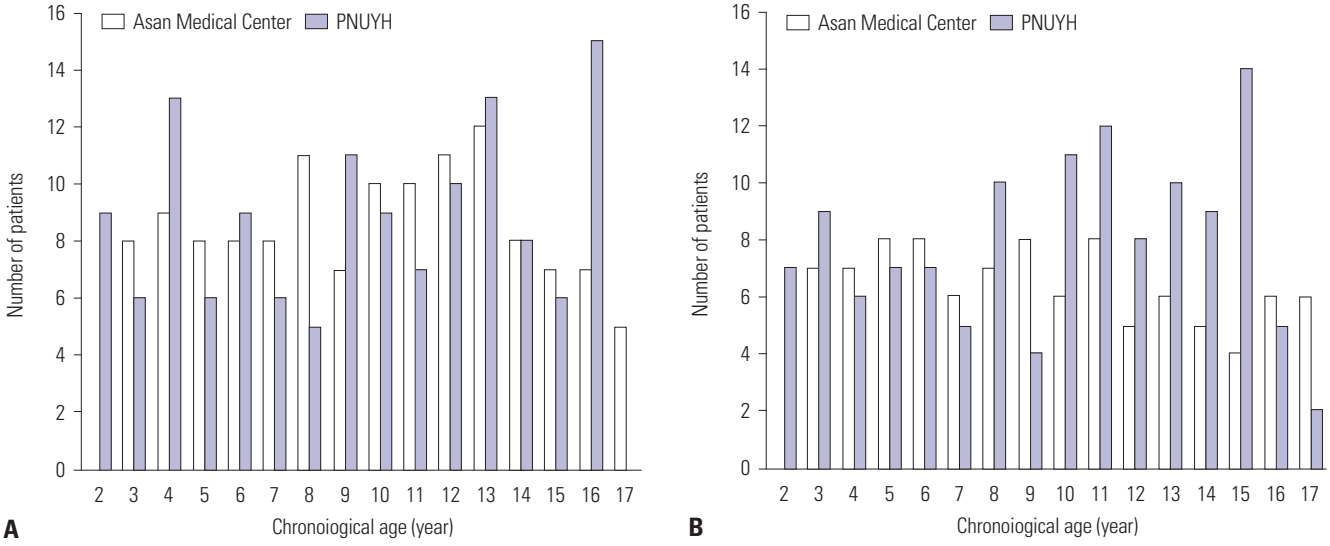

Fig. 1. Number of included children per age and sex (A: boys, B: girls) from two hospitals. PNUYH, Pusan National University Yangsan Hospital.

Table 1. ICC Values of the Comparison between Chronological Age and Bone Age Assessment Methods

Parameter	Estimated bone age			
	Radiologist 1	Radiologist 2	Original DLBAA model	Modified DLBAA model
Chronological age	0.978	0.978	0.982	0.982
Estimated bone age				
Radiologist 1		0.995	0.994	0.994
Radiologist 2			0.993	0.994
Original DLBAA model				0.999

DLBAA, deep learning-based bone age assessment; ICC, intraclass correlation.
All p-values were <0.001 by ICC analysis.

The Bland-Altman plots revealed negative trend curves (all slopes ≤-0.11, all p<0.001) showing proportional negative bias (Table 2 and Fig. 2). These results indicated that, compared to CA, the radiologists and DLBAA methods tended to underestimate BA in younger children and overestimate BA in older children. The mean differences were -2.24 months, -0.48 months, -1.64 months, and - 1.40 months for radiologist 1, radiologist 2, and the original and modified DLBAA models, respectively. When the analyses were conducted according to each sex and each hospital, the Bland-Altman plots and trend curve revealed similar results (Supplementary Table 1 and Sup-

Table 2. Bland-Altman Analysis with Slope from the Linear Regression Between Estimated Bone Ages and Chronological Age

Measurements	Mean difference	Standard deviation	Slope	Intercept
Chronological age vs.				
Radiologist 1	-2.24	16.30	-0.16	17
Radiologist 2	-0.48	16.55	-0.15	18
Original DLBAA model	-1.64	14.62	-0.11	12
Modified DLBAA model	-1.40	14.43	-0.11	12

DLBAA, deep learning-based bone age assessment.

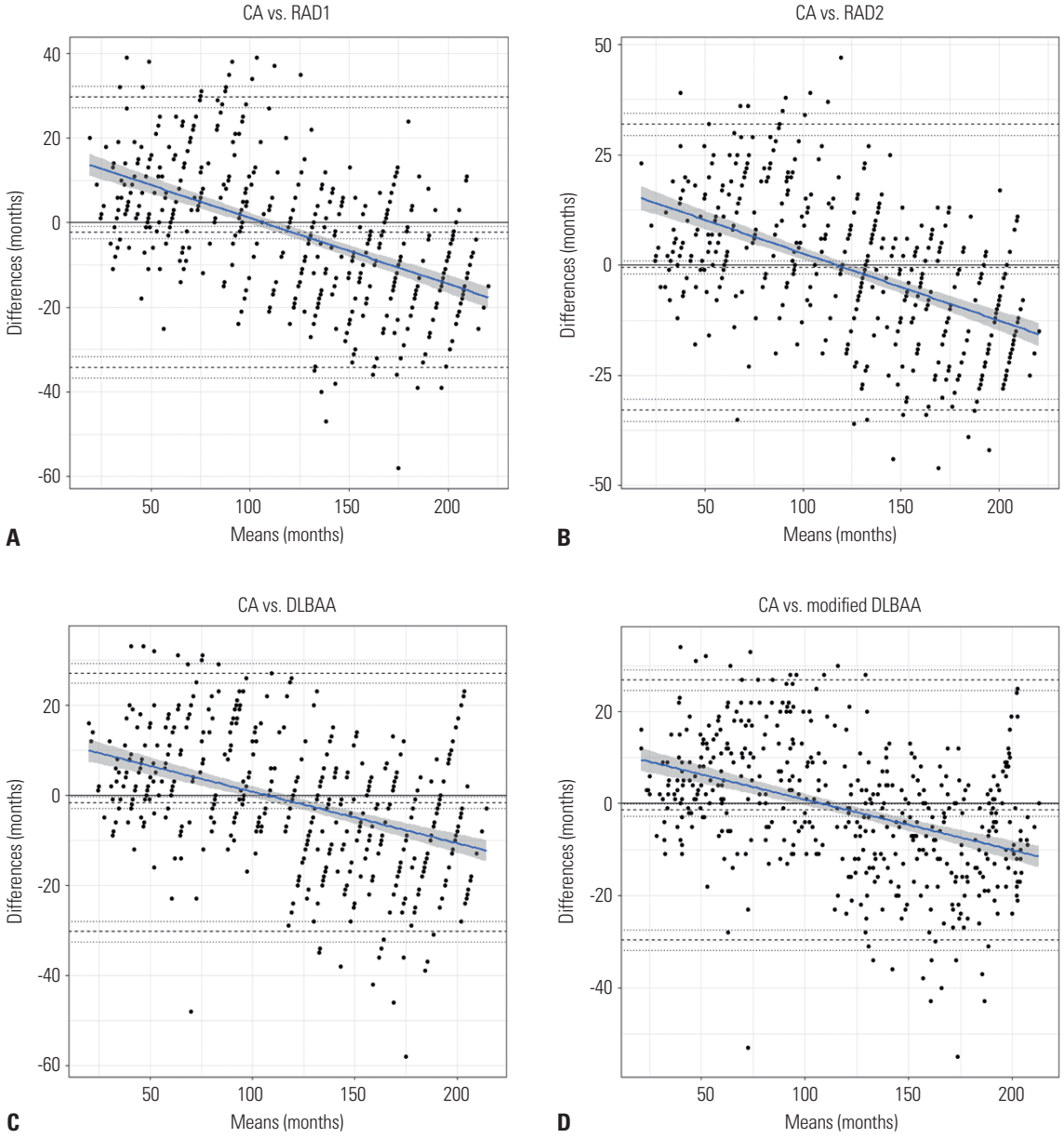

Fig. 2. Bland-Altman plots and trend curve for comparison between chronological age (CA) and estimated bone age by radiologist 1 (A), radiologist 2 (B), original model of deep learning-based bone age assessment (DLBAA) system (C), and modified model of DLBAA system (D). Limits of agreement are shown as the top and bottom dashed lines and average bias (the center dashed line) with 95% confidence intervals of each value (dotted line). The regression fit of the differences on the means are shown as solid blue lines with 95% confidence intervals (gray shaded area).

https://doi.org/10.3349/ymj.2022.63.7.683

plementary Figs. 1-4, only online).

In linear regression analysis, there were significant positive linear correlations between CA and estimated BA by the radiologists and DLBBA methods ($R^2 \geq 0.935$, $p<0.001$) (Table 3 and Fig. 3). The regression lines of all of the estimates showed an underestimation of BA in younger children (up to 102.8 months by radiologist 1, 116.8 months by radiologist 2, 101.8 months by the original DLBAA model, and 103.1 months by the modified

Table 3. Linear Regression Results for Bone Age Estimation by Radiologists and Deep Learning-Based Software Compared to Chronological Age

Measurements	Regression coefficient	R^2 value	Intercept	SD	*p* value
Chronological age vs.					
Radiologist 1	1.130	0.939	-13.370	14.878	<0.001
Radiologist 2	1.123	0.935	-14.363	15.290	<0.001
Original DLBAA model	1.086	0.942	-8.751	13.936	<0.001
Modified DLBAA model	1.082	0.942	-8.452	13.809	<0.001

DLBAA, deep learning-based bone age assessment; SD, standard deviation of residuals of the regression.

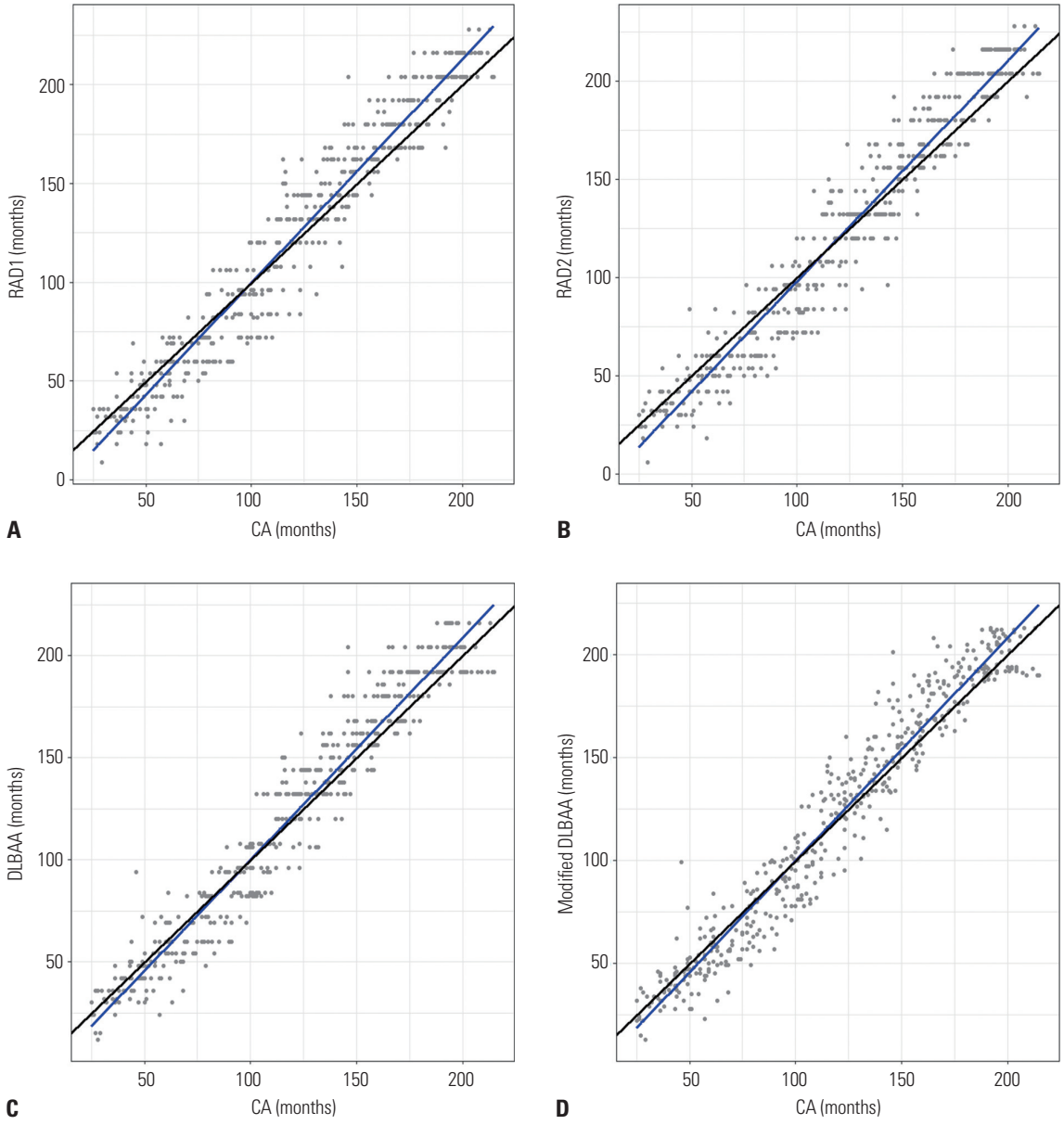

Fig. 3. Linear regression scatter plots between chronological age (CA) and estimated bone age by radiologist 1 (A), radiologist 2 (B), original model of deep learning-based bone age assessment (DLBAA) system (C), and modified model of DLBAA system (D). Lines represent the line of linear regression (blue line) and identity line (black line).

DLBAA model) and overestimation in older children (Fig. 4).

Difference between chronological age and estimated bone ages

The mean absolute error of radiologist 1, radiologist 2, and the original and modified DLBAA models were 13.09, 13.12, 11.52, and 11.31 months, respectively (Table 4). The root mean square error of radiologist 1, radiologist 2, and the original and modified DLBAA models were 16.44, 16.54, 14.69, and 14.48 months, respectively (Table 4). The differences between radiologists vs. DLBAA models were significant for both mean absolute error (p<0.001) and root mean square error (p≤0.018). No significant difference was found in both the mean absolute error (p=0.81) and root mean square error (p>0.999) between the two DLBAA models.

The difference between estimated BA and CA was >12 months in 44.3%, 44.5%, 39.2%, and 36.1% of the patients; >18 months in 27.0%, 28.9%, 21.0%, and 20.0% of the patients; >24 months in 14.2%, 15.3%, 8.0%, and 8.7% of the patients by radiologist 1, radiologist 2, and the original and modified DLBAA models, respectively. Cochran's Q test showed a significant difference in the percentage of BA estimations >12 months, 18 months, and 24 months in reference to CA among the radiologists and DLBAA methods (p<0.001). The post-hoc test results are shown in Table 5. The differences in the percentage of BA estimations >12 months were significant between radiologists vs. modified DLBAA model (p<0.001). The differences in the percentage of BA estimations >18 and 24 months were significant between radiologists vs. DLBAA models (p≤0.002). There was no significant difference in the percentage of BA estimations >12 months (p=0.028), 18 months (p=0.487), and 24 months (p=0.678) compared to CA between the two DLBAA models.

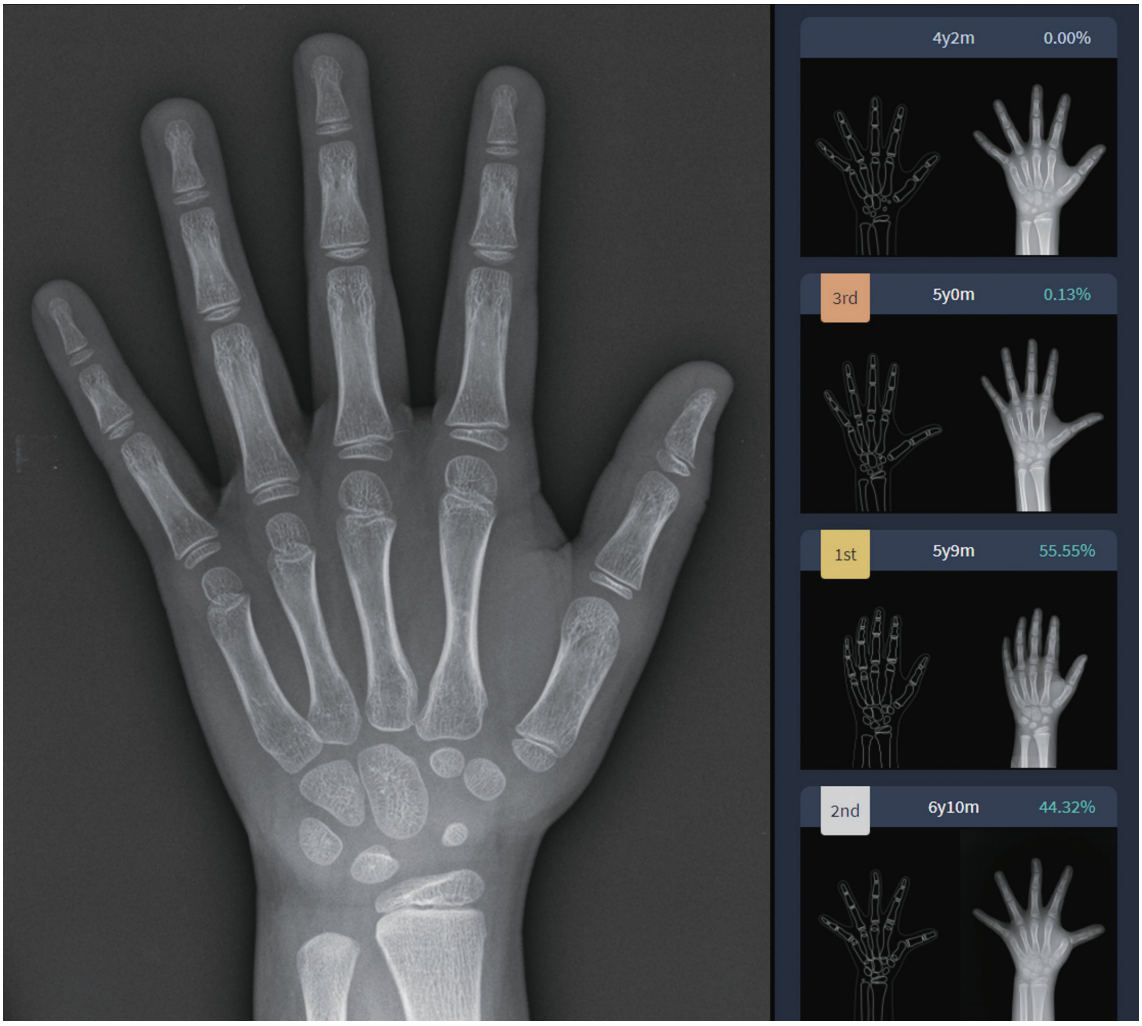

Fig. 4. Screenshot result of the original model of DLBAA system in a girl with chronological age of 6 years 9 months. Among the top three GP-assigned bone ages, the estimated bone age with the highest probability was 5 years 9 months. In this patient, the estimated bone ages by radiologist 1, radiologist 2, and modified model of DLBAA system were 5 years 9 months, 5 years, and 6 years 3 months, respectively. DLBAA, deep learning-based bone age assessment.

https://doi.org/10.3349/ymj.2022.63.7.683

Table 4. Results of Pair-Wise Comparison of the Mean Absolute Error and Root Mean Square Error between Radiologists and Deep Learning-Based Software When Using Chronological Age as a Reference Standard

	Reader				Bonferroni-corrected *p* value					
	R1	R2	Original DLBAA model	Modified DLBAA model	R1 vs. R2	R1 vs. Original DLBAA model	R1 vs. Modified DLBAA model	R2 vs. Original DLBAA model	R2 vs. Modified DLBAA model	Original vs. Modified DLBAA model
MAE (month)	13.09	13.12	11.52	11.31	>0.999	<0.001*	<0.001*	<0.001*	<0.001*	0.81
RMSE (month)	16.44	16.54	14.69	14.48	>0.999	<0.001*	0.018*	<0.001*	<0.001*	>0.999

R1, radiologist 1; R2, radiologist 2; DLBAA, deep learning-based bone age assessment; MAE, mean absolute error; RMSE, root mean square error.
Difference is statistically significant at the 0.05 level.
*Significant differences by the repeated measures of analysis of variance.

Table 5. Comparison of Proportions of Bone Age Estimations >12, 18, and 24 Months Compared to Chronological Age between Radiologists and Deep Learning-Based Software

	Proportions (%)					Post-hoc (McNemar Test)					
	R1	R2	Original DLBAA model	Modified DLBAA model	Cochran's Q Test	R1 vs. R2	R1 vs. Original DLBAA model	R1 vs. Modified DLBAA model	R2 vs. Original DLBAA model	R2 vs. Modified DLBAA model	Original vs. Modified DLBAA model
>12 months	44.3	44.5	39.2	36.1	<0.001	>0.999	0.022	<0.001*	0.016	<0.001*	0.028
>18 months	27.0	28.9	21.0	20.0	<0.001	0.28	0.002*	<0.001*	<0.001*	<0.001*	0.487
>24 months	14.2	15.3	8.0	8.7	<0.001	0.511	<0.001*	<0.001*	<0.001*	<0.001*	0.678

R1, radiologist 1; R2, radiologist 2; DLBAA, deep learning-based bone age assessment.
*Statistically significant differences by post-hoc tests using Bonferroni correction (p<0.0083).

DISCUSSION

Our study compared the CA of contemporary healthy children in Korea with the BA determined by radiologists and the DL-BAA system based on the GP method. Although the estimated BA and CA showed excellent agreement, a systemic bias was present in all of the estimated BA methods in our study population. Specifically, BA tended to be lower than CA in younger patients, and higher than CA in older patients (approximately below and above 102–117 months, respectively). This tendency was seen in both boys and girls, and in children of two tertiary hospitals located in two major Korean cities.

The systemic bias noted in our study was in concordance with the findings from previous studies. Ontell, et al.[9] evaluated GP-based BA in children of diverse ethnicities using hand radiograph of healthy children. They concluded that compared to CA, BA was lower in the ages of 4–8 years, and higher in adolescent ages in Asian boys. Zhang, et al.[11] assessed BA based on the GP method using a large number of digital hand atlases obtained from healthy children in California, which included 331 Asian children. The authors concluded that the BAs of Asian girls (aged 10–13 years) and boys (aged 11–15 years) were significantly advanced than those of white children in the same age group. One study with a population of 212 heathy children in Korea showed a strong correlation between GP-based BA and CA, and the estimated BA tended to be lower than CA among boys.[8] However, their study only included prepubertal aged children (7–12 years). Our systemic bias was consistently observed in the evaluation by all of the radiologists and deep learning-based software; therefore, it may be assumed that this finding is a reliable reflection of existing difference between the GP atlas and contemporary Korean children, rather than rater variability.

Previous studies demonstrated a reduced interpretation time, improved accuracy, and/or decreased variability with the assistance of deep learning-based software.[7,17] The aforementioned results support that deep learning software can reliably assist the assessment of BA in children and function as a time-saving tool when used in clinical practice. In our study, the differences between CA and BA were higher in manual reading compared to the automated method. However, even after applying the deep learning-based software for BA assessment in current healthy Korean children, the systemic bias remained unresolved. This issue should be clarified to pediatric radiologists or pediatrician who use the DLBAA system for BA assessment. We first validated the modified model of DLBAA system that can calculate optimal BA, not limited to BA intervals used in the GP atlas. We confirmed feasibility of the modified DLBAA model that showed comparable results to the original DLBAA model in BA estimation. The accuracy of this modified DLBAA model must be validated in a larger population and various ethnic groups.

In practice, advanced BA can be considered in children showing a difference of >2 SD[18] or 12 months[19] between BA and CA, and a delay >2 years has arbitrarily been used for the diagnosis of the constitutional delay of growth and puberty.[20] Approxi-

mately half of the healthy children in our study showed a difference of more than 12 months between the estimated BA and CA, and this difference can be important in the clinical context or in forensic science. Considering the observed difference and systemic bias between the estimated BA and CA in the children of our study, the GP-independent and Korean-specific deep learning model trained upon the normal bone morphology of contemporary children should be explored in the near future. The result of a recent study by Pan, et al.[5] is encouraging in that the GP-independent deep learning model showed a significantly better performance than did the GP-dependent model (mean absolute error of 11.1 months vs. 12.9 months, respectively) in the children in the United States.

This study had a few limitations. First, we retrospectively reviewed hand trauma radiographs and considered the patients as healthy children, compared to those who underwent radiographs for typical BA estimation in the endocrinology department. We could not evaluate the patients' physical development, such as the height or Tanner scale; therefore, a small number of included patients might not have shown normal skeletal development. Second, we could not collect data on the socioeconomic status of the included patients. Third, we used the CAs of children as a reference standard, but a wide range of normal variation can be present in the pattern of ossification of the hand and wrist;[21] thus, it should be noted that CA may not be a "perfect" gold standard.

In conclusion, contemporary healthy Korean children showed different rates of skeletal development than the GP standard-BA, which was lower in younger children and higher in older children. This issue remained unresolved when applying deep learning-based automated software, and physicians should be aware of the limitation when assessing the developmental status of pediatric patients.

ACKNOWLEDGEMENTS

This study was supported by a grant (2020IE0016-1) from Asan Medical Center Children's Hospital (Heart Institute), Seoul, Korea. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2020R1G1A1004591).

AUTHOR CONTRIBUTIONS

Conceptualization: Hee Mang Yoon and Jae-Yeon Hwang. Data Curation: Hee Mang Yoon and Jae-Yeon Hwang. Formal Analysis: Jisun Hwang and Pyeong Hwa Kim. Funding Acquisition: Hee Mang Yoon. Investigation: Hee Mang Yoon, Jae-Yeon Hwang, and Jisun Hwang. Methodology: Hee Mang Yoon, Jae-Yeon Hwang, Hwa Jung Kim, and Pyeong Hwa Kim. Project Administration: Hee Mang Yoon and Jin Seong Lee. Resources: Hee Mang Yoon, Jae-Yeon Hwang, Jisun Hwang, Boram Bak, Byeong Uk Bae, and Jinkyeong Sung. Software: Byeong Uk Bae and Jinkyeong Sung. Supervision: Hee Mang Yoon, Jae-Yeon Hwang, Ah Young Jung, Young Ah Cho, and Jin Seong Lee. Validation: Jae-Yeon Hwang. Visualization: Jisun Hwang and Pyeong Hwa Kim. Writing—original draft: Jisun Hwang. Writing—review & editing: Hee Mang Yoon, Jae-Yeon Hwang, Ah Young Jung, Young Ah Cho, Jin Seong Lee, Byeong Uk Bae, Jinkyeong Sung, and Hwa Jung Kim. Approval of final manuscript: all authors.

ORCID iDs

Jisun Hwang https://orcid.org/0000-0002-7593-2246
Hee Mang Yoon https://orcid.org/0000-0001-6491-5734
Jae-Yeon Hwang https://orcid.org/0000-0003-2777-3444
Pyeong Hwa Kim https://orcid.org/0000-0003-4276-8803
Boram Bak https://orcid.org/0000-0003-0409-6292
Byeong Uk Bae https://orcid.org/0000-0003-2309-8517
Jinkyeong Sung https://orcid.org/0000-0003-3546-6081
Hwa Jung Kim https://orcid.org/0000-0003-1916-7014
Ah Young Jung https://orcid.org/0000-0002-7427-6240
Young Ah Cho https://orcid.org/0000-0001-6722-121X
Jin Seong Lee https://orcid.org/0000-0002-8470-4595

REFERENCES

1. Kelly PM, Diméglio A. Lower-limb growth: how predictable are predictions? J Child Orthop 2008;2:407-15.
2. Creo AL, Schwenk WF 2nd. Bone age: a handy tool for pediatric providers. Pediatrics 2017;140:e20171486.
3. Greulich WW, Pyle SI. Radiographic atlas of skeletal development of the hand and wrist. 2nd ed. California: Stanford University Press; 1959.
4. Breen MA, Tsai A, Stamm A, Kleinman PK. Bone age assessment practices in infants and older children among Society for Pediatric Radiology members. Pediatr Radiol 2016;46:1269-74.
5. Pan I, Baird GL, Mutasa S, Merck D, Ruzal-Shapiro C, Swenson DW, et al. Rethinking Greulich and Pyle: a deep learning approach to pediatric bone age assessment using pediatric trauma hand radiographs. Radiol Artif Intell 2020;2:e190198.
6. Larson DB, Chen MC, Lungren MP, Halabi SS, Stence NV, Langlotz CP. Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs. Radiology 2018;287:313-22.
7. Kim JR, Shim WH, Yoon HM, Hong SH, Lee JS, Cho YA, 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.
8. Kim JR, Lee YS, Yu J. Assessment of bone age in prepubertal healthy Korean children: comparison among the Korean standard bone age chart, Greulich-Pyle method, and Tanner-Whitehouse method. Korean J Radiol 2015;16:201-5.
9. Ontell FK, Ivanovic M, Ablin DS, Barlow TW. Bone age in children of diverse ethnicity. AJR Am J Roentgenol 1996;167:1395-8.
10. Alshamrani K, Messina F, Offiah AC. Is the Greulich and Pyle atlas applicable to all ethnicities? A systematic review and meta-analysis. Eur Radiol 2019;29:2910-23.
11. Zhang A, Sayre JW, Vachon L, Liu BJ, Huang HK. Racial differences in growth patterns of children assessed on the basis of bone age. Radiology 2009;250:228-35.
12. Hackman L, Black S. The reliability of the Greulich and Pyle atlas when applied to a modern Scottish population. J Forensic Sci 2013;58:114-9.
13. Alshamrani K, Offiah AC. Applicability of two commonly used bone age assessment methods to twenty-first century UK children. Eur Radiol 2020;30:504-13.
14. Kim SY, Oh YJ, Shin JY, Rhie YJ, Lee KH. Comparison of the

https://doi.org/10.3349/ymj.2022.63.7.683

Greulich-Pyle and Tanner Whitehouse (TW3) methods in bone age assessment. J Korean Soc Pediatr Endocrinol 2008;13:50-5.
15. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. Tutor Quant Methods Psychol 2012; 8:23-34.
16. Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. Geosci Model Dev 2014;7:1247-50.
17. Tajmir SH, Lee H, Shailam R, Gale HI, Nguyen JC, Westra SJ, et al. Artificial intelligence-assisted interpretation of bone age radiographs improves accuracy and decreases variability. Skeletal Radiol 2019;48:275-83.
18. Oh MS, Kim S, Lee J, Lee MS, Kim YJ, Kang KS. Factors associated with advanced bone age in overweight and obese children. Pediatr Gastroenterol Hepatol Nutr 2020;23:89-97.
19. Kim D, Cho SY, Maeng SH, Yi ES, Jung YJ, Park SW, et al. Diagnosis and constitutional and laboratory features of Korean girls referred for precocious puberty. Korean J Pediatr 2012;55:481-6.
20. Martin DD, Wit JM, Hochberg Z, Sävendahl L, van Rijn RR, Fricke O, et al. The use of bone age in clinical practice-part 1. Horm Res Paediatr 2011;76:1-9.
21. Gilsanz V, Ratib O. Hand bone age: a digital atlas of skeletal maturity. Berlin: Springer; 2005.