Original Article

Yonsei Med J 2023 Nov;64(11):679-686
https://doi.org/10.3349/ymj.2023.0244

Yonsei Medical Journal
YMJ
pISSN: 0513-5796 · eISSN: 1976-2437

Bone Age Estimation and Prediction of Final Adult Height Using Deep Learning

Junghwan Suh[1*], Jinkyoung Heo[2*], Su Jin Kim[1], Soyeong Park[1], Mo Kyung Jung[3], Han Saem Choi[4], Youngha Choi[5], Jun Suk Oh[6], Hae In Lee[7], Myeongseob Lee[8], Kyungchul Song[1], Ahreum Kwon[1], Hyun Wook Chae[8], and Ho-Seong Kim[1]

[1]Department of Pediatrics, Severance Children's Hospital, Endocrine Research Institute, Yonsei University College of Medicine, Seoul;
[2]Department of University Industry Foundation, Yonsei University, Seoul;
[3]Department of Pediatrics, CHA Bundang Medical Center, CHA University, Seongnam;
[4]Department of Pediatrics, International St. Mary's Hospital, Catholic Kwandong University, Incheon;
[5]Department of Pediatrics, Kangwon National University Hospital, Chuncheon;
[6]Department of Pediatrics, Konyang University College of Medicine, Daejeon;
[7]Department of Pediatrics, CHA Gangnam Medical Center, CHA University, Seoul;
[8]Department of Pediatrics, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Korea.

Purpose: The appropriate evaluation of height and accurate estimation of bone age are crucial for proper assessment of the growth status of a child. We developed a bone age estimation program using a deep learning algorithm and established a model to predict the final adult height of Korean children.

Materials and Methods: A total of 1678 radiographs from 866 children, for which the interpretation results were consistent between two pediatric endocrinologists, were used to train and validate the deep learning model. The bone age estimation algorithm was based on the convolutional neural network of the deep learning system. The test set simulation was performed by a deep learning program and two raters using 150 radiographs and final height data for 100 adults.

Results: There was a statistically significant correlation between bone age interpreted by the artificial intelligence (AI) program and the reference bone age in the test set simulation (r=0.99, p<0.001). In the test set simulation, the AI program showed a mean absolute error (MAE) of 0.59 years and a root mean squared error (RMSE) of 0.55 years, compared with reference bone age, and showed similar accuracy to that of an experienced pediatric endocrinologist (rater 1). Prediction of final adult height by the AI program showed an MAE of 4.62 cm, compared with the actual final adult height.

Conclusion: We developed a bone age estimation program based on a deep learning algorithm. The AI-derived program demonstrated high accuracy in estimating bone age and predicting the final adult height of Korean children and adolescents.

Key Words: Bone age, final height, artificial intelligence, deep learning, convolutional neural network

Received: June 26, 2023 **Revised:** August 8, 2023
Accepted: August 14, 2023 **Published online:** October 17, 2023
Corresponding author: Ho-Seong Kim, MD, PhD, Department of Pediatrics, Severance Children's Hospital, Endocrine Research Institute, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea
E-mail: kimho@yuhs.ac

*Junghwan Suh and Jinkyoung Heo contributed equally to this work.
•The authors have no potential conflicts of interest to disclose.

INTRODUCTION

Owing to increasing social interest in height and quality of life, many children visit hospitals to evaluate their linear growth status.[1] It is necessary to differentiate between pathological short stature and short stature within a normal range. Appropriate examination and evaluation of children are important for accurate assessment of their growth status and identification of various growth problems.

Bone age, which represents skeletal maturation of the body, is one of the most important aspects of growth status evaluation. Bone age is usually estimated from a left-hand radiograph, including the wrist. Bone age is mainly interpreted using the Greulich-Pyle (GP) method[2] and Tanner-Whitehouse (TW) method.[3] The GP method is used to intuitively estimate bone age by comparing it with a standard atlas of the reference population, and the TW method is used to calculate the scores and sums of the number and shape of bones in the hand and wrist. The GP method is more widely used in clinical practice because it takes relatively less reading time than the TW method. However, the GP method may yield different results depending on the experience of the interpreter, and inter- and intra-person variation could exist in the interpretation. Therefore, clinicians need a faster, more accurate, and consistent method with which to assess bone age.

With recent advances in artificial intelligence (AI) technology, deep learning algorithms are thought to be applicable in bone age estimation.[4] A convolutional neural network (CNN), which is a type of deep learning architecture, has received attention for its strength in image processing. Bone age estimation programs based on deep learning have been developed and used worldwide;[5,6] however, there are relatively few programs for the Korean population, owing to racial and ethnic differences in skeletal maturation. In addition, although there are some commercialized bone age estimation programs that are of use in predicting final adult height, the prediction results are not always constant or accurate.[7,8] Additionally, there may be some discrepancies in bone age interpretation results between radiologists and pediatricians who treat actual patients.[9] Accurate and rapid estimation of bone age and precise prediction of final adult height using deep learning would be significantly beneficial in clinical settings to promote children and public health and reduce social costs.

The aim of this study was to develop a bone age estimation program using a deep learning algorithm based on bone age data of Korean children interpreted by pediatricians and to establish a model to predict the final adult height of children.

MATERIALS AND METHODS

Image selection and interpretation

Bone age view radiographs (left-hand anteroposterior view including the wrist) were obtained from children who visited the pediatric endocrinology clinic at Severance Children's Hospital from March 2011 to March 2020. In total, 71466 radiographs were obtained from 21614 children. We excluded radiographs of patients with underlying conditions that could affect growth status, such as growth hormone deficiency, precocious puberty, diabetes, intracranial tumors, and adrenal diseases. Patients treated with medications that could affect linear growth, including recombinant human growth hormones, gonadotropin-releasing hormone agonists, steroids, and thyroid hormones, were also excluded. Also, children under 3 years of age were excluded from the study as the GP method is not accurate for a younger age. Subsequently, two board-certified pediatric endocrinologists (each with 34 and 12 years of clinical experience) interpreted the remaining 4526 radiographs from a total of 1961 children based on the GP method,[2] and we only included the results of readings that matched between the two reviewers. If the difference between the results of the two reviewers was less than 3 months, the reading results were considered to be matched. Finally, 1678 radiographs (877 from males and 801 from females) from 866 children were used to obtain the reference bone age with which to train and validate the deep learning model (Fig. 1). 1678 radiographs were ran-

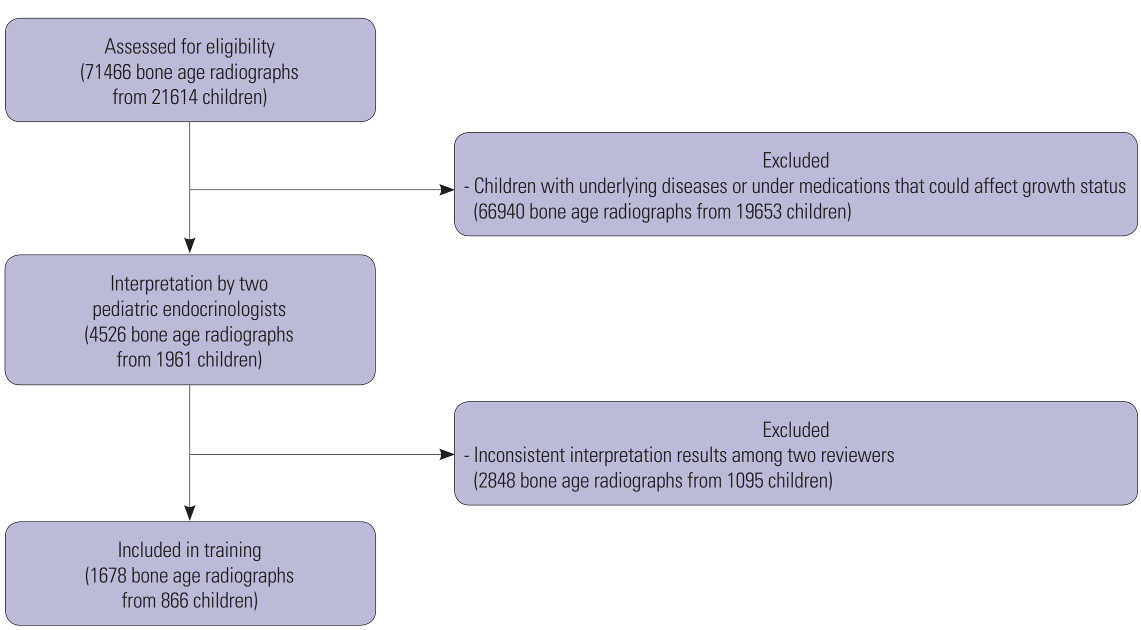

Fig. 1. Flow chart of the selection of bone age radiographs.

https://doi.org/10.3349/ymj.2023.0244

domly divided into training and validation sets at a ratio of 7:3, using hierarchical sampling. Finally, 1237 radiographs (667 from males and 570 from females) were used to train the deep learning model, and 441 radiographs (210 from males and 231 from females) were used as the validation set. Table 1 shows the distributions of sex and chronological age according to the radiographs used in the training set.

This study was approved by the Institutional Review Board of Severance Hospital, Yonsei University College of Medicine, Seoul, Korea (No. 4-2020-0192). The requirement for informed consent was waived because of the retrospective nature of the study.

Image pre-processing

Image pre-processing can improve the efficiency of machine learning. First, a masked area was created around the hand.

Table 1. Distribution of Sex and Chronological Age According to Radiographs Used in the Training Set of Deep Learning Model

Age, years	Male	Female	Total
3–4	31	21	52
4–5	55	32	87
5–6	48	64	112
6–7	52	51	103
7–8	50	43	93
8–9	50	49	99
9–10	50	52	102
10–11	50	52	102
11–12	50	47	97
12–13	49	49	98
13–14	49	46	95
14–15	54	33	87
15–16	40	19	59
16–17	30	12	42
17–18	9	0	9
Total	667	570	1237

Data are presented as n.

Except for the hand, all other parts, such as letters or white borders, were removed so that only the information on the hand area could be extracted. Using the original image and masked area, blurry backgrounds, characters, and areas other than the hands were removed through bitwise mask operations. Next, the image was rotated to make the angle of the radiograph consistent with that of others. The center point of the wrist and the starting point of the wrist were set to x and y, respectively. After the linear regression line was found, the image was rotated so that the linear regression line was perpendicular. To better distinguish bones from faint X-ray images, the brightness value was adjusted by varying the weight of the rotated image according to the degree of brightness. Then, a top-hat morphological operation was performed to emphasize the area where the brightness value greatly changed, thus extracting the bone area from the image. To minimize noise, blur processing was performed, and contrast enhancement was conducted with different weights set according to the image average value. Finally, contour detection was performed to detect the borders of the hand. We used YOLOv5 as an algorithm for joint detection, and the joints of the wrist, thumb, middle finger, and little finger were annotated using a bounding box.[10]

AI program modeling

For modeling the AI program, the CNN of the deep learning algorithm was used as the basis, and it was modified using the attention module. For hyper-parameter tuning, the activation function of the backbone of YOLOx and YOLOm model of YOLOv5 was tested on convolutional layer and bottleneck structures. We selected the model with the best detection performance through activation functions and optimizer tuning. The learning technique of our program was based on TjNet,[11] which is based on VGG16 and the convolutional block attention module (CBAM).[12,13] However, instead of CBAM, we applied a bottleneck attention module (BAM) in which the channel attention and spatial attention modules were arranged in parallel.[14] Fig. 2 shows a schema of the process of recognizing and interpreting bone age radiographs using our deep learning model.

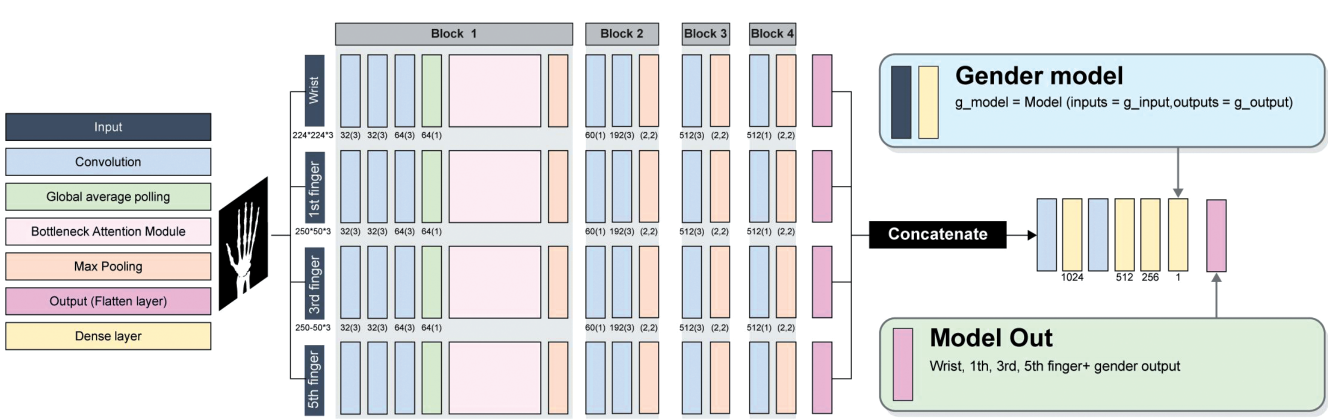

Fig. 2. Process of interpreting bone age radiographs in the deep learning model.

When the wrist and the first, third, and fifth fingers obtained through image pre-processing are input, they first go through the convolutional layer and global average pooling and then enter the BAM layer. After going through three more blocks, the four outputs are concatenated into one, and then the result is derived through a fully connected layer and gender classification layer.

Prediction of final adult height

Standard growth charts of Korean children and adolescents were used to construct a regression model to predict final adult height.[15] The height z-score for bone age was calculated using the modified least mean square (LMS) method, where L stands for the power of box-cox transformation, M is the median, and S is the coefficient of variation.[16] The height at the age of 18 years was estimated by calculating the height z-score for bone age and the L, M, and S values of each male and female at the age of 18 years, which were regarded as indicative of the final adult height. The following equation expresses the formula for predicting the final adult height:

$$g(z)=M_{18}\times[1+(L_{18}\times S_{18}\times Z)]^{(1/L_{18})}$$

g(z): Predicted final adult height
M_{18}: M value at the age of 18 years
L_{18}: L value at the age of 18 years
S_{18}: S value at the age of 18 years
Z: height z-score for bone age

Simulation for bone age estimation and prediction of final adult height

The AI program and two raters participated in the simulation using the test set. Rater 1 is a board-certified pediatric endocrinologist with 11 years of clinical experience, and rater 2 is a fourth-year resident in pediatrics. Bone ages were assessed with the assistance of a GP atlas. The test set consisted of 150 bone age radiographs taken from a total of 150 children, 75 each for males and females. The 150 radiographs used for simulation were prepared not to overlap with the 1678 radiographs used for machine learning and validation. In addition, 100 of these (50 each for males and females) were from children who had reached their final adult height. Therefore, 150 radiographs were used for bone age estimation, and a final adult height of 100 children was predicted by the AI program and the two raters.

Statistical analysis

Statistical analyses were performed using R version 4.2.1 (The R Foundation for Statistical Computing, Vienna, Austria). Pearson correlation analysis and scatter plots were used to assess correlation between bone age determined using the AI program and the reference bone age. To compare the accuracy and precision of the AI program and the two raters, the mean absolute error (MAE) and root mean square error (RMSE) were analyzed, and the values were compared using Student's t-test. Bonferroni adjustment was applied for adjustment of multiple testing. Additionally, intraclass correlation coefficient (ICC) and concordance correlation coefficient (CCC) values were assessed to determine the agreement between the simulation results and reference values.[17,18] Bland–Altman plots were used to evaluate the correlation in bone age estimation and prediction of the final adult height between the simulation results and the references. Statistical significance was set at $p<0.05$.

RESULTS

Accuracy of the estimated bone age by AI program

Internal validation was performed using 441 bone age radiographs (210 from males and 231 from females) to assess the accuracy of the AI program. An MAE of 0.39 years (4.7 months) was observed between the AI program and reference bone age.

The scatterplot in Fig. 3 shows the correlation between bone age determined by the AI program and the reference bone age in the test set simulation. We noted a statistically significant correlation between bone age interpreted by the AI program and reference values ($r=0.99$, $p<0.001$). Table 2 presents the simulation results of the bone age estimation using a test set of 150 bone age radiographs. The AI program showed an MAE of 0.59 years and RMSE of 0.55 years, compared to the reference bone age. Rater 1 showed an MAE of 0.60 years and RMSE of 0.52 years, which were similar to those of the AI program. Rater 2 showed slightly inferior results, compared with those of the AI program and rater 1, but the differences were not statistically significant. When the data were divided on the basis of sex, compared to raters 1 and 2, the AI program showed superior

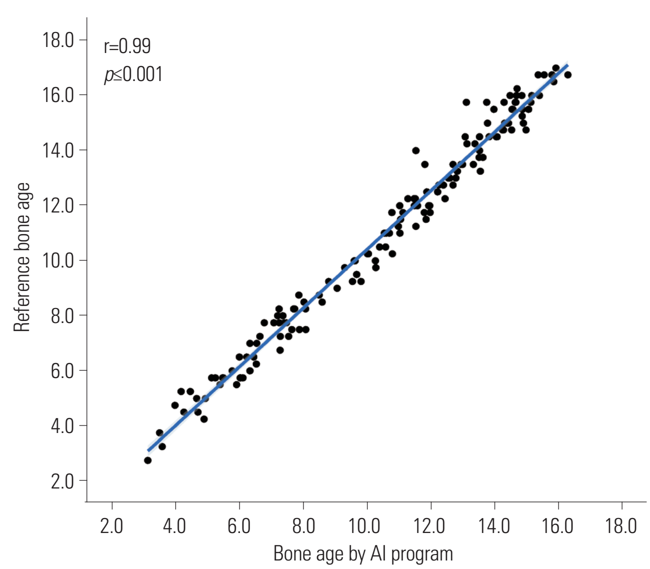

Fig. 3. Correlation between bone age determined by the AI program and reference bone age. AI, artificial intelligence.

https://doi.org/10.3349/ymj.2023.0244

Table 2. Accuracy and Agreement of Bone Age Estimation between Simulation Results and Reference Values

	AI program	Rater 1	Rater 2
Overall			
MAE, yr	0.59	0.60	0.64
RMSE, yr	0.55	0.52	0.62
ICC (95% CI)	0.980 (0.914, 0.992)	0.983 (0.945, 0.992)	0.978 (0.904, 0.991)
CCC (95% CI)	0.980 (0.973, 0.985)	0.983 (0.977, 0.987)	0.978 (0.971, 0.984)
Males			
MAE, yr	0.59	0.53	0.61
RMSE, yr	0.58	0.42	0.56
ICC (95% CI)	0.979 (0.909, 0.991)	0.987 (0.978, 0.992)	0.980 (0.897, 0.992)
CCC (95% CI)	0.978 (0.968, 0.986)	0.987 (0.981, 0.991)	0.980 (0.969, 0.986)
Females			
MAE, yr	0.59	0.67	0.67
RMSE, yr	0.51	0.62	0.67
ICC (95% CI)	0.982 (0.914, 0.993)	0.979 (0.616, 0.994)	0.977 (0.906, 0.991)
CCC (95% CI)	0.981 (0.972, 0.987)	0.979 (0.969, 0.986)	0.977 (0.964, 0.985)

AI, artificial intelligence; MAE, mean absolute error; RMSE, root mean square error; ICC, intraclass correlation coefficient; CI, confidence interval; CCC, concordance correlation coefficient.
Among raters 1 and 2, there was no statistically significant difference, compared to the AI program.

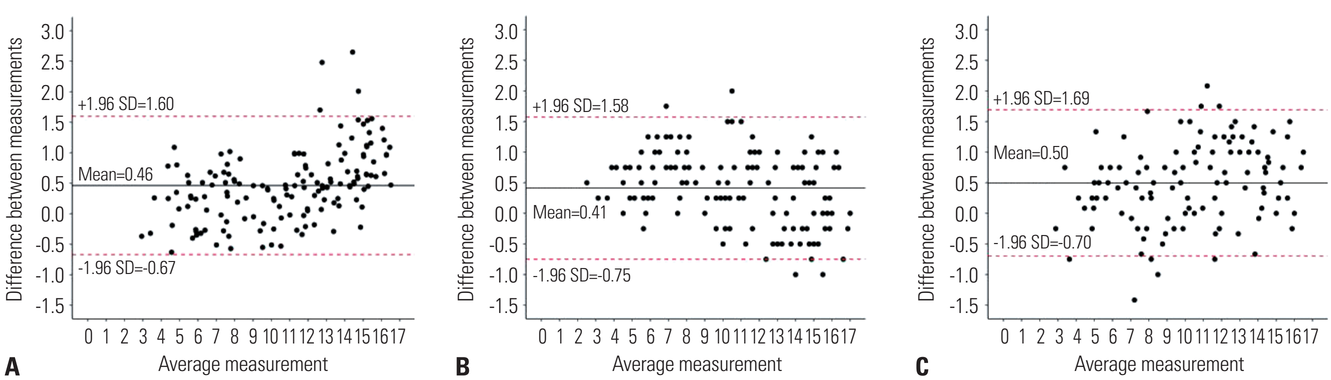

Fig. 4. Bland-Altman plots of the bone age estimation results and reference bone age. (A) Reference bone age-bone age by the AI program. (B) Reference bone age-bone age by rater 1. (C) Reference bone age-bone age by rater 2. AI, artificial intelligence.

outcomes in females. However, interpretation results from rater 1 were the best in estimating male bone age. The ICC and CCC values showed substantial agreement between the simulation results and reference values.

Bland–Altman plots of the bone age estimation results and the reference bone age depicted in Fig. 4 illustrate a high level of correlation, with a tendency to slightly underestimate bone age, compared to the reference values in all three participants. In addition, the accuracy of the AI program decreased slightly as age increased. Sex-stratified Bland–Altman plots for bone age estimation are presented in Supplementary Figs. 1 and 2 (only online).

Accuracy of the AI program in the prediction of final adult height

The prediction results of final adult height by the AI program and the two raters were compared using the final adult height data from 100 children (Table 3). The accuracy of the AI program showed an MAE 4.62 cm and an RMSE of 37.49 cm. Rater 1 showed significantly better accuracy in predicting the final adult height in both sexes (MAE of 3.17 cm, RMSE of 18.62 cm), whereas the accuracy of rater 2 was similar to that of the AI program (MAE of 4.43 cm, RMSE of 33.96 cm). The overall ICC and CCC values showed relatively high agreement among all three participants, although the strength of the correlation decreased when the data were stratified by sex.

Bland–Altman plots of the prediction results and reference values of final adult height also depicted the highest correlation for rater 1, and similar results for the AI program and rater 2 (Fig. 5). In addition, a tendency to overestimate the final adult height, compared to the reference values, was observed for the AI program and both raters. Sex-stratified Bland–Altman plots for the prediction of final adult height are presented in Supplementary Figs. 3 and 4 (only online).

Table 3. Accuracy and Agreement of Prediction of Final Adult Height between Simulation Results and Reference Values

	AI program	Rater 1	Rater 2
Overall			
MAE, cm	4.62	3.17*	4.43
RMSE, cm	37.49	18.62*	33.96
ICC (95% CI)	0.790 (0.704, 0.854)	0.844 (0.770, 0.894)	0.792 (0.689, 0.861)
CCC (95% CI)	0.789 (0.715, 0.845)	0.843 (0.776, 0.891)	0.790 (0.712, 0.849)
Males			
MAE, cm	4.77	3.28	5.23
RMSE, cm	37.70	21.52	43.10
ICC (95% CI)	0.287 (0.000, 0.528)	0.309 (0.045, 0.536)	0.111 (-0.093, 0.331)
CCC (95% CI)	0.282 (0.084, 0.459)	0.305 (0.044, 0.527)	0.109 (-0.083, 0.293)
Females			
MAE, cm	4.46	3.05	3.63
RMSE, cm	37.28	15.72	24.82
ICC (95% CI)	0.281 (0.023, 0.509)	0.513 (0.278, 0.691)	0.468 (0.218, 0.660)
CCC (95% CI)	0.277 (0.028, 0.493)	0.508 (0.299, 0.671)	0.463 (0.218, 0.653)

AI, artificial intelligence; MAE, mean absolute error; RMSE, root mean square error; ICC, intraclass correlation coefficient; CI, confidence interval; CCC, concordance correlation coefficient.
*$p<0.05$, compared to the AI program.

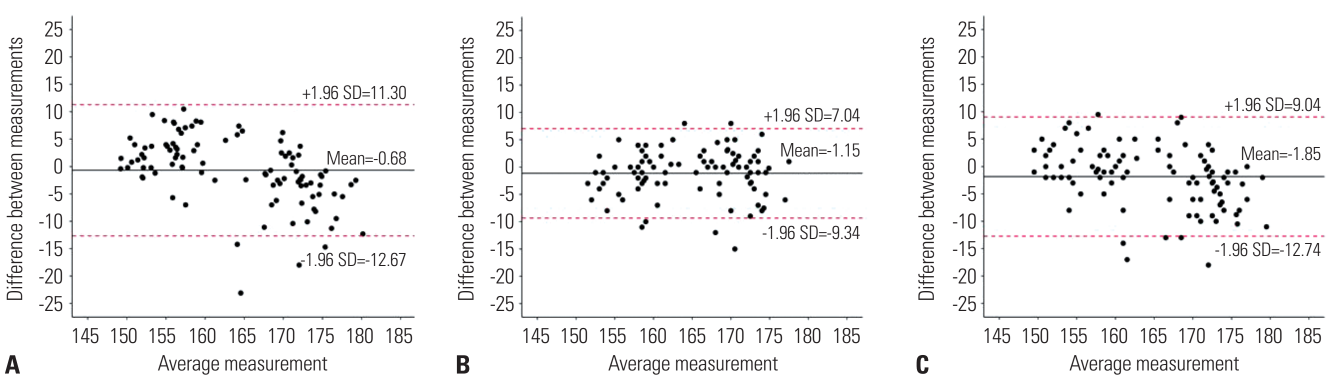

Fig. 5. Bland-Altman plots of the prediction results and reference values of final adult height. (A) Reference final adult height-final adult height by the AI program. (B) Reference final adult height-final adult height by rater 1. (C) Reference final adult height-final adult height by rater 2. AI, artificial intelligence.

DISCUSSION

In this study, we developed a bone age estimation program using a CNN via a deep learning algorithm based on bone age data of Korean children and adolescents. Additionally, a regression model was established to predict final adult height. To focus on accuracy, our study only included bone age radiographs, where the interpretation results were consistent between two experienced pediatric endocrinologists. Although the total number of bone age radiographs used in our study was relatively small, the efficiency of machine learning was increased by excluding potentially controversial radiographs. The AI program showed accurate and reliable results in bone age estimation and prediction of final adult height, comparable with results from two pediatricians.

Based on test set simulation results, the accuracy of the AI program in bone age estimation was relatively high and consistent, and there were no noticeable differences in estimates, compared with those of raters 1 and 2. However, the degree of error gradually increased with increasing chronological age; therefore, further machine learning is necessary for bone age radiographs of older individuals. An MAE of 0.59 years and RMSE of 0.55 years from the simulation by the AI program in this study are similar to those shown by previous AI-based bone age programs. BoneXpert, which was based on data from European children, showed an accuracy within 0.72 years (standard deviation) for Asians,[19] and VUNO Med-BoneAge based on the Korean population had an RMSE of 0.60 years.[8] In addition, MediAI-BA solution, an AI bone age estimation program based on the TW3 method, showed an MAE of 0.59 years,[20] and HH-boneage.io solution showed an MAE of 0.46 years and RMSE of 0.62 years,[21] confirming that the performance of the AI program in our study, in general, was similar to those of previous releases.

Predicting final adult height is a difficult process in actual clinical practice, and a doctor's experience plays a major role.

https://doi.org/10.3349/ymj.2023.0244

Rater 1, who had more patient experience, showed significantly better results than the AI program and rater 2 (a fourth-year resident) with relatively little clinical experience. The results predicted by the AI program and rater 2, even though not as precise as those by rater 1, were similar and accurate. In our study, each machine-learned bone age radiograph was tagged with the child's height, weight, body mass index, and chronological age at the time of examination, and only data from healthy children were included, excluding cases of diseases or specific treatments that might affect growth. Currently, the prediction of final adult height in this study represents projection using growth charts and the LMS method; however, we plan to include serially measured longitudinal data to increase accuracy in the future.

In the test set simulation conducted in this study, there were sex differences in both bone age estimation and final height prediction. Even in the same bone age radiograph, the estimated bone age would be different because the degree and speed of skeletal maturation vary depending on sex, and it is very important to accurately identify this difference when managing patients. Importantly, the two raters in this study showed difference in interpreting results according to sex; however, the AI program demonstrated very consistent performance regardless of sex. The predicted adult heights were also different according to sex. Generally, when predicting the final adult height through plotting and projection methods using growth curves, height for boys tends to be overestimated and height for girls tends to be underestimated.[22] In particular, the bone age of boys progresses rapidly as they enter puberty, making it more difficult to predict their final height. Our simulation results showed a higher accuracy of the predicted height for girls for both raters. On the contrary, the AI program did not show a large difference in accuracy between boys and girls in predicting final adult height and showed steady results.

This study has some limitations. To increase accuracy, we only included healthy children who did not receive any specific treatment related to linear growth. In addition, as we included only bone age radiographs that were rated similarly by the two endocrinologists, the total number of radiographs used in machine learning was relatively small. In addition, because the GP method is not accurate for children of younger age and a small number of radiographs, children under 3 years of age were excluded from the study. Moreover, as final adult height is rarely predicted at a young age owing to its low accuracy, we did not perform final height prediction in simulation for children under 6 years of age. In addition, the reading time for bone age estimation was not measured in the test set simulation. As the reading time was not measured, raters with little experience would have benefited from spending more time to increase the accuracy of interpretation and height prediction. Additionally, one of the main limitations of this study is that external validation was not performed. Finally, this study used data from a single tertiary institution. In future, we plan to increase the accuracy of the program by additional machine learning of bone age radiographs, comparing it with multicenter data and readings from other clinicians.

In conclusion, we developed a bone age estimation program based on the CNN of a deep learning system solely using imaging data and interpretation results of Korean pediatric population. The AI program showed high accuracy in estimating bone age and predicting final adult height in Korean children and adolescents.

ACKNOWLEDGEMENTS

This study was supported by a grant from Severance Children's Hospital (2020-102). The authors thank Medical Illustration & Design, part of the Medical Research Support Services of Yonsei University College of Medicine, for all the artistic support related to this work.

AUTHOR CONTRIBUTIONS

Conceptualization: Junghwan Suh and Ho-Seong Kim. Data curation: Han Saem Choi, Youngha Choi, Jun Suk Oh, Hae In Lee, and Myeongseob Lee. Formal analysis: Junghwan Suh and Jinkyoung Heo. Funding acquisition: Junghwan Suh. Investigation: Junghwan Suh and Jinkyoung Heo. Methodology: Junghwan Suh, Jinkyoung Heo, and Ho-Seong Kim. Project administration: Kyungchul Song, Ahreum Kwon, Hyun Wook Chae, and Ho-Seong Kim. Resources: Su Jin Kim, Soyeong Park, and Mo Kyung Jung. Software: Jinkyoung Heo. Supervision: Hyun Wook Chae and Ho-Seong Kim. Validation: Junghwan Suh, Jinkyoung Heo, and Ho-Seong Kim. Visualization: Junghwan Suh and Jinkyoung Heo. Writing—original draft: Junghwan Suh and Jinkyoung Heo. Writing—review & editing: Junghwan Suh, Jinkyoung Heo, and Ho-Seong Kim. Approval of final manuscript: all authors.

ORCID iDs

Junghwan Suh	https://orcid.org/0000-0002-2092-2585
Jinkyoung Heo	https://orcid.org/0000-0001-7279-0994
Su Jin Kim	https://orcid.org/0000-0003-0907-9213
Soyeong Park	https://orcid.org/0000-0002-6160-3839
Mo Kyung Jung	https://orcid.org/0000-0001-5904-4946
Han Saem Choi	https://orcid.org/0000-0003-4140-0345
Youngha Choi	https://orcid.org/0000-0002-1607-6048
Jun Suk Oh	https://orcid.org/0000-0002-5765-2347
Hae In Lee	https://orcid.org/0000-0002-7644-1617
Myeongseob Lee	https://orcid.org/0000-0001-7055-3100
Kyungchul Song	https://orcid.org/0000-0002-8497-5934
Ahreum Kwon	https://orcid.org/0000-0002-9692-2135
Hyun Wook Chae	https://orcid.org/0000-0001-5016-8539
Ho-Seong Kim	https://orcid.org/0000-0003-1135-099X

REFERENCES

1. Allen DB. Growth promotion ethics and the challenge to resist cosmetic endocrinology. Horm Res Paediatr 2017;87:145-52.
2. Greulich WW, Pyle SI. Radiographic atlas of skeletal development of the hand and wrist. Redwood City, CA: Stanford University Press; 1959.

3. Tanner JM, Whitehouse RH, Cameron N, Marshall WA, Healy MJR, Goldstein H. Assessment of skeletal maturity and prediction of adult height (TW2 method). 2nd ed. London: Academic Press; 1983.
4. Halabi SS, Prevedello LM, Kalpathy-Cramer J, Mamonov AB, Bilbily A, Cicero M, et al. The RSNA pediatric bone age machine learning challenge. Radiology 2019;290:498-503.
5. Lee H, Tajmir S, Lee J, Zissen M, Yeshiwas BA, Alkasab TK, et al. Fully automated deep learning system for bone age assessment. J Digit Imaging 2017;30:427-41.
6. Booz C, Yel I, Wichmann JL, Boettger S, Al Kamali A, Albrecht MH, et al. Artificial intelligence in bone age assessment: accuracy and efficiency of a novel fully automated algorithm compared to the Greulich-Pyle method. Eur Radiol Exp 2020;4:6.
7. Lee BD, Lee MS. Automated bone age assessment using artificial intelligence: the future of bone age assessment. Korean J Radiol 2021;22:792-800.
8. 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.
9. Lea WW, Hong SJ, Nam HK, Kang WY, Yang ZP, Noh EJ. External validation of deep learning-based bone-age software: a preliminary study with real world data. Sci Rep 2022;12:1232.
10. Ultralytics. YOLOv5 [accessed on 2023 January 18]. Available at: https://github.com/ultralytics/yolov5.
11. Shastry AS, Mervyn BA, Rizvi BZ, Menon V, Girisha GS. Automatic bone age assessment of radiographs using deep learning. Proceedings of the 2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES); 2021 Sep 24-25; Chennai, India: IEEE; 2021; p.1-7.
12. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv [Preprint]. 2014 [accessed on 2023 June 27]. Available at: https://doi.org/10.48550/arXiv.1409.1556.
13. Woo S, Park J, Lee JY, Kweon IS. Cbam: convolutional block attention module. Proceedings of the European Conference on Computer Vision (ECCV); 2018 Sep 8-14; Munich, Germany: ECCV; 2018; p.3-19.
14. Park J, Woo S, Lee JY, Kweon IS. BAM: bottleneck attention module. arXiv [Preprint]. 2018 [accessed on 2023 June 27]. Available at: https://doi.org/10.48550/ARXIV.1807.06514.
15. Kim JH, Yun S, Hwang SS, Shim JO, Chae HW, Lee YJ, et al. The 2017 Korean national growth charts for children and adolescents: development, improvement, and prospects. Korean J Pediatr 2018;61:135-49.
16. Cole TJ. The LMS method for constructing normalized growth standards. Eur J Clin Nutr 1990;44:45-60.
17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
18. McBride GB. Using statistical methods for water quality management: issues, problems and solutions. Hoboken, NJ: John Wiley & Sons; 2005.
19. Martin DD, Sato K, Sato M, Thodberg HH, Tanaka T. Validation of a new method for automated determination of bone age in Japanese children. Horm Res Paediatr 2010;73:398-404.
20. Bui TD, Lee JJ, Shin J. Incorporated region detection and classification using deep convolutional networks for bone age assessment. Artif Intell Med 2019;97:1-8.
21. Son SJ, Song Y, Kim N, Do Y, Kwak N, Lee MS, et al. TW3-based fully automated bone age assessment system using deep neural networks. IEEE Access 2019;7:33346-58.
22. Sperling MA. Pediatric endocrinology. 4th ed. Philadelphia, PA: Saunders; 2014.

https://doi.org/10.3349/ymj.2023.0244