

# Accuracy of predicted adult height using the Greulich-Pyle method and artificial intelligence medical device

Dongho Cho<sup>1</sup>, Yun Sun Choi<sup>2</sup>, Hayun Oh<sup>2</sup>, Young min Ahn<sup>3</sup>, Ji-Young Seo, MD<sup>1</sup>

<sup>1</sup>Department of Pediatrics, Nowon Eulji Medical Center, Eulji University School of Medicine, Seoul, Korea; <sup>2</sup>Department of Radiology, Nowon Eulji Medical Center, Eulji University School of Medicine, Seoul, Korea; <sup>3</sup>Jangs hospital, Seoul, Korea

## To the editor

It is expected that artificial intelligence (AI) medical devices would be introduced in medical images earlier than in other fields, and in fact, there are many cases where AI-based bone age (BA) medical devices are used in primary medical institutions. Besides, there are few studies comparing BA and final adult height (FAH) prediction of humans and AI.<sup>1-3)</sup>

Medical image reading by AI is by deep learning based on the Greulich-Pyle (GP) and tanner-white house methods, and this study aimed to compare the accuracy of BA and FAH prediction of VUNO Med-BoneAge (VUNO Inc., Seoul, Korea), the most commonly used AI program in Korea and specialist.

Our study included 190 children and adolescents (73 males and 117 females) aged 8–12 years who visited the Growth Clinic of Pediatric Endocrinology in 2012 to evaluate their BA and

predicted adult height (PAH). The height, weight, body mass index, and height of the father and mother were retrospectively reviewed, and their BA and PAH were predicted based on the GP method by pediatric endocrinologist and musculoskeletal specialist of radiology in 2012 and parents' and subject's height in 2021 were collected by a telephone survey.<sup>4,5)</sup>

Those who had chronic diseases, treatment to improve growth, height below 3 percentile or above 97 percentile in 2012, the difference in chronological age and BA is above 2 years, did not reach FAH by 2021 were excluded. Of the total 961 individuals (322 males and 639 females), 190 (73 males and 117 females) were studied.

In this study, VUNO Med-BoneAge was used as an AI medical device, and its principle was based on deep learning to find the atlas of the most statistically similar BA and to provide the final

**Table 1. Baseline characteristics of study population**

Variable	Male (n=73)		Female (n=117)	
	Value	z	Value	z
Chronologic age (yr)	10.3±1.3	-	9.2±1.3	-
Height (cm)	141.9±9.2	-0.15±1.18	136.6±8.4	0.03±1.04
Weight (cm)	38.7±8.8	-0.07±1.01	33.4±7.7	0±1.09
BMI (kg/m <sup>2</sup> )	19±2.8	0.06±1.03	17.8±2.9	0.06±1.03
Midparental height (cm)	172±3.6	-0.45±0.66	160.2±3.7	-0.18±0.75
Final adult height (cm)	173±5.2	-0.28±0.93	160.5±5.2	-0.14±1.04

Values are presented as mean±standard deviation.

BMI, body mass index.

**Table 2. Difference of bone age, final adult height, and predicted adult height between specialist and AI group**

Variable	Male (n=73)			Female (n=117)		
	Specialist	AI	P value	Specialist	AI	P value
Bone age (yr)	10.7±2	11±1.7	<b>&lt;0.001</b>	10.6±1.2	9.8±1.7	<b>&lt;0.001</b>
Predicted adult height (cm)	174.8	175.8	<b>&lt;0.001</b>	159.3	160.5	<b>&lt;0.001</b>
Predicted adult height (z)	0.04	0.22	<b>&lt;0.001</b>	-0.38	-0.15	<b>&lt;0.001</b>
FAH-PAH (cm)	-1.8	-2.8	<b>&lt;0.001</b>	1.2	0	<b>&lt;0.001</b>

Values are presented as the mean±standard deviation.

AI, artificial intelligence; FAH, final adult height; PAH, predicted adult height.

Boldface indicates a statistically significant difference with  $P<0.05$ .

Corresponding author: Ji-Young Seo, MD, PhD. Department of Pediatrics, Nowon Eulji Medical Center, Eulji University School of Medicine, 68 Hangeulbiseong-ro, Nowon-gu, Seoul 01830, Korea

✉ Email: pedseo@eulji.ac.kr, <https://orcid.org/0000-0003-3471-7803>

Received: 25 October 2022, Revised: 6 December 2022, Accepted: 28 December 2022

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2023 by The Korean Pediatric Society

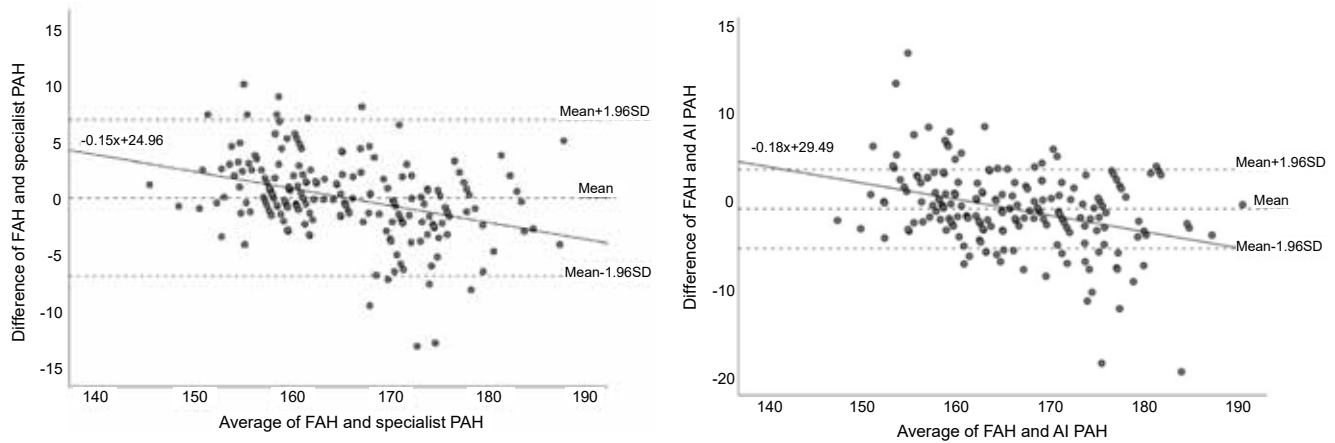


Fig. 1. Bland-Altman plot of specialist and artificial intelligence (AI). FAH, final adult height; PAH, predicted adult height; SD, standard deviation.

value to the first decimal place through the matching rate

The average adult height predicted by a specialist was 174.8 cm for male, 159.3 cm for female, and by AI was 175.8 cm for male, 160.5 cm for female. The FAH surveyed by phone was 173 cm for male and 160.5 cm for female. When subjects were divided by sex, the BA and PAH values differed significantly to the FAH and PAH values in both the specialist and AI groups, especially in male. This difference was smaller in female (Tables 1, 2).

When comparing the specialist and AI's Bland-Altman plot, 93% of the GP method (mean  $\pm$  1.96 standard deviation [SD] = -6.88, 6.96) and 78% of AI (mean  $\pm$  1.96 SD = -5.56, 3.42) were within the agreement limits, so the predictive accuracy of the specialist was 93%, and AI was 78% (Fig. 1).

However, when the subjects were divided by sex and puberty, the P value of the difference between PAH and FAH was not statistically significant in pubertal male and prepubertal female for specialists and prepubertal female for AI.

The largest difference was observed in prepubertal and pubertal male, in the case of specialists and AI, respectively. Other studies comparing BA using another AI (BoneXpert, Hørsholm, Denmark) showed that both male and female tended to measure BA younger in prepubertal age (male, 0.001–0.61 years of age; female, 0.02–0.76 years of age), and older in pubertal age (male, 0.43–1.64 years of age; female, 0.03–1.24 years of age), and the study also showed the greatest difference in pubertal male.<sup>1,6)</sup>

In most studies, both specialists and AI have high predictive rates when measuring FAH in female, which is presumed to be because most children visiting the growth clinic are female who are worried about precocious puberty.<sup>7,8)</sup> According to the report, the prevalence of precocious puberty in Korea is 40 times higher in female than in male, and thus, both specialists and AI have more experience in measuring female growth plates, which makes it possible to predict more accurately.<sup>9)</sup>

In this study, the Bland-Altman plot was used to evaluate the accuracy of the prediction. Other studies related to BA and PAH also used the Bland-Altman plot for accuracy comparison with each method or AI. Jeong et al.<sup>10)</sup> confirmed that the difference

between the expected and FAHs calculated using the BP method falls into the limits of agreement; Kim et al.<sup>2)</sup> showed that most values are located within the limits of agreement in specialist's predictions and BoneXpert predictions, and there are not many differences between the 2 methods.

The limitation of this study is that only about 40% of all respondents answered the FAH by phone because the call was made 10 years after the outpatient visit. Therefore, a selection bias may occur because a group that the FAH does not reach the PAH is more likely to be excluded by themselves, and due to the nature of the telephone survey, the given height value can be larger than the actual measured value.<sup>4,5)</sup> Specialist's BA and AI predictions were performed in 2012 and 2021, respectively, therefore the latter could have an advantage.

In addition, the study compared the results of BA with one pediatric endocrinologists and AI, and most of the current studies are single-center studies, so later studies should include pediatric endocrinologists and results in multiple centers.

Along with the AI medical devices, doctors, not pediatric endocrinologists, rely on AI to diagnose growth problems only with BA, and accordingly, there are cases where proper evaluation and treatment of diseases are delayed or missed.

In conclusion, in this study, the accuracy of the PAH assessed by a specialist was 93%, while that assessed by AI was 78%. This suggests that AI prediction may still need a monitoring by specialists.

This paper is conducted on a small group randomly selected from a single center, and it should not be interpreted as AI can be a tool to predict BA or PAH on behalf of pediatric endocrinologists, and AI companies should avoid using the paper commercially.

## Footnotes


Conflicts of interest: No potential conflict of interest relevant to this article was reported.

Funding: This study received no specific grant from any funding


agency in the public, commercial, or not-for-profit sectors.


ORCID :

Dongho Cho  <https://orcid.org/0000-0002-8829-4626>

Yun Sun Choi  <https://orcid.org/0000-0001-7082-3991>

Hayun Oh  <https://orcid.org/0000-0002-0874-7286>

Young min Ahn  <https://orcid.org/0000-0002-1697-8041>

Ji-Young Seo  <https://orcid.org/0000-0003-3471-7803>

## References

1. Ahn KS, Bae B, Jang WY, Lee JH, Oh S, Kim BH, et al. Assessment of rapidly advancing bone age during puberty on elbow radiographs using a deep neural network model. *Eur Radiol* 2021;31:8947-55.
2. 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.
3. Wang YM, Tsai TH, Hsu JS, Chao MF, Wang YT, Jaw TS. Automatic assessment of bone age in Taiwanese children: a comparison of the Greulich and Pyle method and the Tanner and Whitehouse 3 method. *Kaohsiung J Med Sci* 2020;36:937-43.
4. Flegal KM, Ogden CL, Fryar C, Afful J, Klein R, Huang DT. Comparisons of self-reported and measured height and weight, BMI, and obesity prevalence from national surveys: 1999-2016. *Obesity (Silver Spring)* 2019;27:1711-9.
5. Ezzati M, Martin H, Skjold S, Vander Hoorn S, Murray CJ. Trends in national and state-level obesity in the USA after correction for self-report bias: analysis of health surveys. *J R Soc Med* 2006;99:250-7.
6. Pose Lepe G, Villacrés F, Silva Fuente-Alba C, Guiloff S. Correlation in radiological bone age determination using the Greulich and Pyle method versus automated evaluation using BoneXpert software. *Rev Chil Pediatr* 2018;89:606-611. Spanish.
7. 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.
8. Choukair D, Hückmann A, Mittnacht J, Breil T, Schenk JP, Alrajab A, et al. Near-adult heights and adult height predictions using automated and conventional Greulich-Pyle bone age determinations in children with chronic endocrine diseases. *Indian J Pediatr* 2022;89:692-8.
9. Kim YJ, Kwon A, Jung MK, Kim KE, Suh J, Chae HW, et al. Incidence and Prevalence of central precocious puberty in Korea: an epidemiologic study based on a national database. *J Pediatr* 2019;208:221-8.
10. Jeong SW, Cho JH, Jung HW, Shim KS. Near final height in Korean children referred for evaluation of short stature: clinical utility and analytical validity of height prediction methods. *Ann Pediatr Endocrinol Metab* 2018;23:28-32.

**How to cite this article:** Cho D, Choi YS, Oh H, Ahn YM, Seo JY. Accuracy of predicted adult height using the Greulich-Pyle method and artificial intelligence medical device. *Clin Exp Pediatr* 2023;66:145-7.