



A comparison of two artificial intelligence-based methods for assessing bone age in Turkish children: BoneXpert and VUNO Med-Bone Age

Evrim Özmen¹
 Hande Özen Atalay¹
 Evren Uzer¹
 Mert Veznikli²

¹Koç University Hospital, Department of Radiology, İstanbul, Türkiye

²Koç University Faculty of Medicine, Department of Computational Biology and Biostatistics, İstanbul, Türkiye

PURPOSE

This study aimed to evaluate the validity of two artificial intelligence (AI)-based bone age assessment programs, BoneXpert and VUNO Med-Bone Age (VUNO), compared with manual assessments using the Greulich–Pyle method in Turkish children.

METHODS

This study included a cohort of 292 pediatric cases, ranging in age from 1 to 15 years with an equal gender and number distribution in each age group. Two radiologists, who were unaware of the bone age determined by AI, independently evaluated the bone age. The statistical study involved using the intraclass correlation coefficient (ICC) to measure the level of agreement between the manual and AI-based assessments.

RESULTS

The ICC coefficients for the agreement between the manual measurements of two radiologists indicate almost perfect agreement. When all cases, regardless of gender and age group, were analyzed, a nearly perfect positive agreement was observed between the manual and software measurements. When bone age calculations were separated and analyzed separately for girls and boys, there was no statistically significant difference between the two AI-based methods for boys; however, ICC coefficients of 0.990 and 0.982 were calculated for VUNO and BoneXpert, respectively, and this difference of 0.008 was significant ($z = 2.528$, $P = 0.012$) for girls. Accordingly, VUNO showed higher agreement with manual measurements compared with BoneXpert. The difference between the agreements demonstrated by the two software packages with manual measurements in the prepubescent group was much more pronounced in girls compared with boys. After the age of 8 years for girls and 9 years for boys, the agreement between manual measurements and both AI software packages was equal.

CONCLUSION

Both BoneXpert and VUNO showed high validity in assessing bone age. Furthermore, VUNO has a statistically higher correlation with manual assessment in prepubertal girls. These results suggest that VUNO may be slightly more effective in determining bone age, indicating its potential as a highly reliable tool for bone age assessment in Turkish children.

CLINICAL SIGNIFICANCE

Investigating the most suitable AI program for the Turkish population could be clinically significant.

KEYWORDS

Bone age, BoneXpert, VUNO, artificial intelligence, deep learning

Corresponding author: Evrim Özmen

E-mail: evrimkilicdr@gmail.com

Received 06 April 2024; revision requested 02 May 2024;
last revision received 17 July 2024; accepted 28 July 2024.



Epub: 02.09.2024

Publication date: xx.xx.2024

DOI: 10.4274/dir.2024.242790

Bone age is a marker of skeletal maturation and is measured routinely by pediatricians, radiologists, and pediatric endocrinologists for the assessment of the maturation progress of children.¹ The most commonly used manual method for bone age measurement is the Greulich–Pyle (GP) method.² According to this method, the determination of bone age is based on the similarity between the image in the GP atlas and the patient's left-hand wrist radiography. Thus, the GP method is very subjective and has higher inter and intraobserver

variability in addition to inter and intraobserver variability.³ Besides, there is no standardized protocol for assessing bones, and it is unclear which bones should be included in the assessment.⁴ With the development of deep learning, which is a subclass of artificial intelligence (AI) that exploits artificial neural networks, several software programs have been developed to automate and standardize bone age assessment, thereby reducing interobserver variability. It has been reported previously that AI-based assessment methods have high accuracy, reproducibility, and time efficiency when compared with manual methods.⁴ Although BoneXpert version 2.4.5.1 (Visiana, Denmark) is one of the most frequently used methods of these, there are other AI-based bone age calculation software packages, including VUNO Med-Bone Age version 1.0.3 (VUNO) (VUNO, Seoul, Korea). The Turkish population is composed of various ethnic groups. As far as we know, no data compares these software packages, and no published report compares the manual method with these AI-based bone age assessment methods in Turkish children. This study aims to analyze the accuracy of two AI-based bone age assessment programs, namely BoneXpert and VUNO, in comparison with manual assessments using the GP bone age atlas.

Methods

Study design and population

This retrospective cohort study was approved by the Ethics Committee of Koç University Faculty of Medicine (2024.050. IRB2.023) and conducted in accordance with the Declaration of Helsinki's ethical principles. Informed consent was not obtained

from the participants due to retrospective design of the study.

Pediatric cases who underwent left-hand X-ray imaging between January 2016 and December 2023 in the hospital due to suspicion of an endocrinological pathology and whose left-hand X-ray evaluation revealed that their chronological age and bone age were compatible were determined. Patients whose bone age was compatible with chronological age but who had known endocrinologic genetic or orthopedic disorders were excluded from the study list. Cases were also excluded if the radiological images were of poor quality, as this could make bone age estimation difficult.

After that, these cases were anonymized and grouped according to their age and gender, and the groups were randomized within themselves. Due to the limited number of male and female cases in the 1-year age group (aged 1–2 years), 6 cases for each gender were selected from this group. In the evaluation made for the other age groups, it was determined that the group of 15-year-old girls had the fewest case numbers, and there were 10 cases in this group. For this reason, in the other groups, the first 10 cases from the randomized list for both genders were selected. The specific age distribution included 6 boys and 6 girls aged 1–2 years, and 10 boys and 10 girls were included for

each subsequent age group (aged 2–16 years).

Radiological assessment

Left-hand wrist posterior to anterior X-ray images were used for the evaluation of bone age. Two radiologists with 15 and 5 years of experience and unaware of the results determined by AI independently evaluated bone ages according to the GP bone age atlas. Bone age was determined to be the midpoint when a case exhibited some, but not all, of the typical bone characteristics of a particular age (e.g., aged 8 years) and had all the characteristics of the previous age (e.g., aged 7 years). This approach was adopted to provide a more detailed and precise assessment of bone maturity. A third radiologist, aware of the cases' clinical details but blind to the manual bone age assessments, documented the AI assessments using BoneXpert version 3.0.3 and VUNO version 1.0.3 (Figure 1).

Statistical analysis

Correlation analysis was performed using the Statistical Package for the Social Sciences, version 28.0 (IBM SPSS Statistics, Armonk, NY, USA).⁵ Comparing correlation coefficients was done by the MedCalc Statistical Software version 12.7.7 (MedCalc Software bvba, Ostend, Belgium; <http://www.medcalc.org>; 2013). The test used by MedCalc is a z-test

Main points

- Our study reveals that both VUNO Med-Bone Age (VUNO) and BoneXpert correlated well with the manual assessment and Greulich–Pyle atlas.
- VUNO is significantly superior in assessing prepubertal girls. This suggests that VUNO may have a possible advantage in terms of sensitivity or algorithmic calibration by favoring bone age assessment in this specific group.
- The results of our study are particularly important given the lack of existing research on the applicability of these artificial intelligence (AI)-based bone age calculations in the Turkish pediatric population and highlight potential challenges with AI-driven assessments in younger ages, particularly among prepubertal girls.

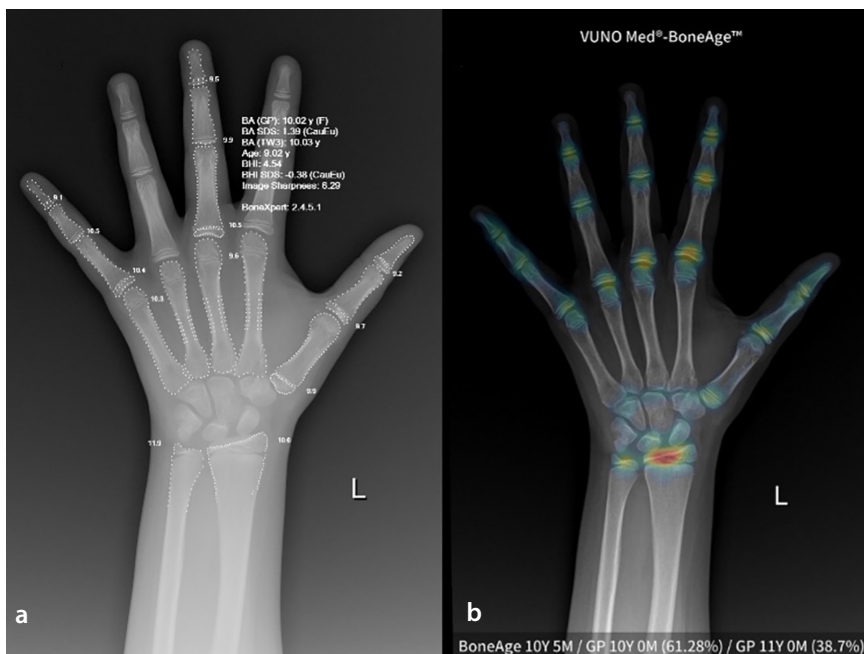


Figure 1. BoneXpert and VUNO images of patients. In the 9.02-year-old female case, BoneXpert showed the bone age as 10.02 years (a) and VUNO indicated the bone age as 10 years 5 months (b). As illustrated in the accompanying image, the BoneXpert system is not designed to evaluate carpal bones, whereas the VUNO system assesses carpal bones as well as tubular bones, as shown with numbers for BoneXpert and colors for VUNO. Upon receipt of the image to be analyzed, VUNO Med-Bone Age suggests three probable estimated bone ages (i.e., first- and second-rank artificial intelligence bone ages) based on probabilities, accompanied by similar images for comparison.

on Fisher's z-transformed correlation coefficients.⁶ The inter-reader agreement between the manual evaluations of two radiologists was measured to ensure consistency in the manual evaluation process. Intraclass correlation coefficients (ICC) were calculated for agreement between two radiologists using a two-way random-effects model, assessing absolute agreement. According to Shrout and Fleiss⁷ (1979), this corresponds to ICC (2,1) for single measures and ICC (2,2) for average measures. Since the agreement was very high, manual evaluation was calculated with the arithmetic mean of these two measurements. The ICC values were used for assessing the agreement between software measurements and the mean radiologist measurements using a two-way random-effects model. According to Shrout and Fleiss⁷ (1979), this corresponds to ICC (2,1) for single measures. Bland–Altman analysis was used to further evaluate the agreement between manual and AI-based assessments. To also see the effect of gender and age on the measurements, all analyses were repeated for all combinations of subgroups: girls, boys, and different age groups. Boys over the age of 9 years and girls over the age of 8 years were considered to be prepubescent.⁸ The statistical significance level was accepted as 0.05.

Results

All pediatric patients aged 1–15 years with left-hand X-ray images generated in our institution were included in the study. Thirty-six patients with poor-quality radiological images and 54 patients with known endocrinologic genetic or orthopedic disorders were excluded from the study. The final study cohort included 292 cases with an equal distribution of genders across all age groups, ranging from 1 to 15 years (Figure 2). The ICC coefficients for the agreement between the manual measurements of two radiologists were calculated as 0.961 for ICC (2,1) and 0.980 for ICC (2,2) (Table 1). These values indicate almost perfect agreement. Based on these measurements, the average of the two observer values was taken and accepted as the manual measurement.

For the manual vs. software comparison, the ICC (2,1) values were calculated for single measurements. When all cases, regardless of gender and age group, were analyzed, a nearly perfect positive agreement was observed between the manual and software measurements. The ICC was calculated as 0.988 for VUNO and 0.984 for BoneXpert. Although the correlation values were not sta-

tistically different, the difference was close to significance ($z = 1.741$, $P = 0.082$).

When bone age calculations were analyzed separately for girls and boys, an ICC co-

efficient of 0.987 and 0.986 was calculated for VUNO and BoneXpert, respectively, for boys, and this difference of 0.001 was not significant ($z = 0.312$, $P = 0.755$). For girls, ICC coeffi-

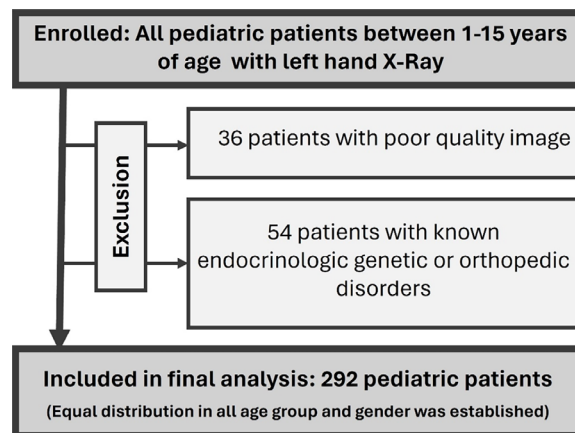


Figure 2. Flowchart of the study.

Table 1. Correlations for agreement between two radiologists

	Intraclass correlation ^b
Single measures ICC (2,1)	0.961 ^a
Average measures ICC (2,2)	0.980

Two-way random effects model where both people effects and measures effects are random. ^a, the estimator is the same, whether the interaction effect is present or not; ^b, type A intraclass correlation coefficients using an absolute agreement definition; ICC, intraclass correlation coefficient.

Table 2. Intraclass correlations for agreement between software and manual measurements

Group		ICC (2,1) ^a		Difference between two correlation coefficients ^b	
		VUNO	BoneXpert	z	P
Overall	ICC	0.988	0.984	1.741	0.082
	95% CI	0.985–0.991	0.980–0.987		
Boys	ICC	0.987	0.986	0.312	0.755
	95% CI	0.981–0.991	0.979–0.991		
Girls	ICC	0.990	0.982	2.528	0.012*
	95% CI	0.985–0.993	0.975–0.987		
Prepubescent	ICC	0.963	0.938	2.229	0.026*
	95% CI	0.948–0.973	0.915–0.955		
Boys aged >9 years and girls aged >8 years	ICC	0.966	0.966	0.000	1.000
	95% CI	0.953–0.975	0.951–0.976		
Prepubescent boys	ICC	0.957	0.953	1.941	0.052
	95% CI	0.933–0.973	0.927–0.970		
Boys aged >9 years	ICC	0.967	0.967	0.000	1.000
	95% CI	0.778–0.989	0.873–0.986		
Prepubescent girls	ICC	0.972	0.917	3.202	0.001*
	95% CI	0.942–0.985	0.867–0.948		
Girls aged >8 years	ICC	0.965	0.965	0.000	1.000
	95% CI	0.946–0.978	0.946–0.977		

*, significant P value at 0.05 level ^a, intraclass correlation coefficient (ICC) estimates and their 95% confident intervals based on a single measures, absolute-agreement, 2-way random effects model; ^b, z-test on Fisher z-transformed correlation coefficients; CI, confidence interval.

cients of 0.990 and 0.982 were calculated for VUNO and BoneXpert, respectively, and this difference of 0.008 was significant ($z = 2.528$, $P = 0.012$). Accordingly, VUNO showed higher agreement with manual measurements compared with BoneXpert.

Upon categorization of all cases by age, a decrease in the software–manual agreement was observed for measurements of the pre-pubescent group. While the ICC coefficient was 0.963 for VUNO, it was calculated as 0.938 for BoneXpert. Accordingly, it was evaluated that, in the measurements of prepubescent children, although there was a slight difference, VUNO exhibited a higher agreement with manual measurements compared with BoneXpert ($z = 2.229$, $r = 0.026$). However, after the age of 8 years for girls and 9 years for boys, the compliance of both software and manual measurements was calculated as 0.966 for VUNO and BoneXpert, and no significant difference was detected between the software (Table 2).

The difference between the agreements demonstrated by the two software packages with manual measurements in the prepubescent group was much more pronounced in girls than in boys. The ICC values in pre-

pubescent girls were calculated as 0.972 for VUNO and 0.917 for BoneXpert, and the difference of 0.055 was significant ($z = 3.202$, $r = 0.001$). In prepubescent boys, the ICC value was 0.957 for VUNO and 0.953 for BoneXpert; the difference was not statistically different, but it was very close to significance ($z = 1.941$, $P = 0.052$).

For girls aged >8 years and boys aged >9 years, the agreement between manual measurements and both AI software packages was equal. While the ICC values were 0.967 in boys aged >9 years, these values were 0.965 in girls aged >8 years (Table 2).

When examining Bland–Altman plot graphs, higher variability is observed on the left side of the graphs. Therefore, it can be seen that both AI-based bone age calculations tend to diverge more from manual measurements in the prepubescent group and are inclined to make lower age estimates than manual measurements (Figures 3, 4). This tendency is particularly evident in the BoneXpert measurements of prepubescent girls.

Discussion

This study represents the inaugural investigation into the comparative efficacy of AI-based systems, namely BoneXpert and VUNO, in the determination of bone age among a Turkish pediatric population. The results of our study indicate that both AI-based systems demonstrated a high level of agreement with each other and with manual methods in all our subgroups, including both genders and age groups. This is consistent with the findings of previous studies in the field. This highlights the potential for integrating AI-based bone age calculation into clinical practice, with the aim of enhancing the effectiveness of bone age assessment.

The GP method is the most widely used and well-known manual method, and according to Martin et al.⁹ it is the method preferred by 76% of pediatric endocrinologists and radiologists.¹⁰ The GP method is based on the comparison of the cases' hand and wrist X-rays, with a standardized radiographic atlas compiled and standardized according to age and gender from birth to 18 years of age for girls and 19 years of age for boys.¹⁰ However, bone age is influenced by ethnicity, gender, genetic factors, socioeconomic level,

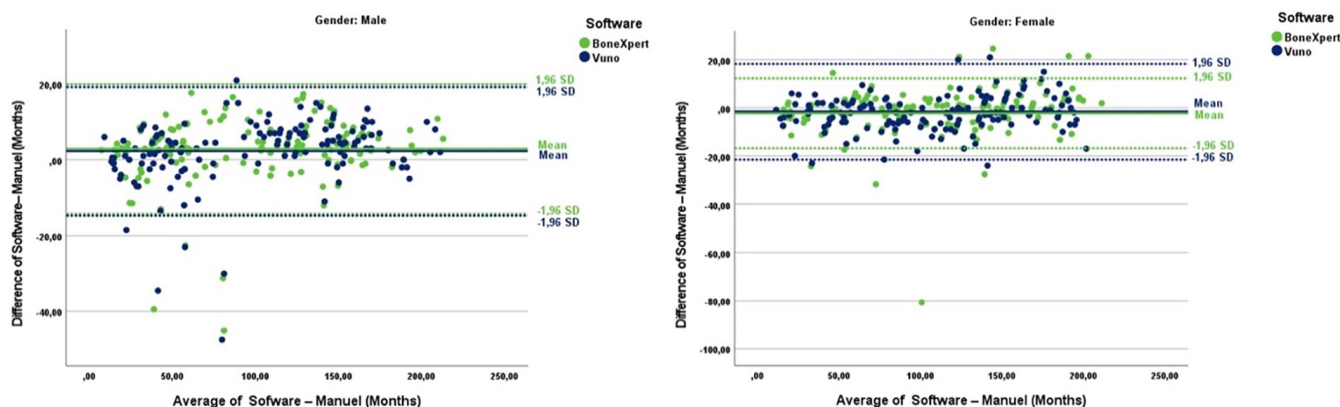


Figure 3. Bland–Altman plot. Difference between radiological bone age and automated bone age in boys and girls.

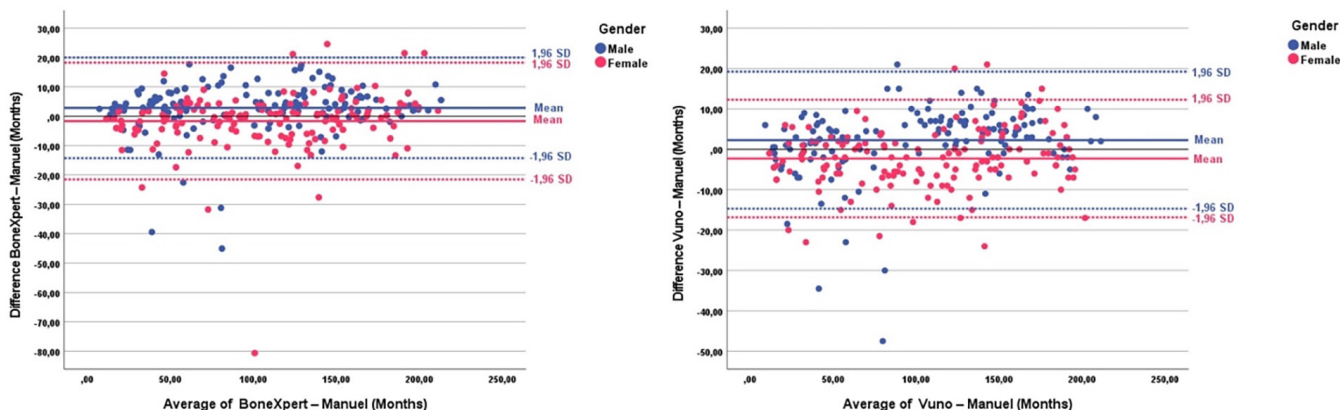


Figure 4. Bland–Altman plot. Difference between radiological bone age and automated bone age in BoneXpert and VUNO.

nutritional metabolic status, and bone disorders.⁹⁻¹² The standardized radiographic images of the atlas were derived from healthy North American and Western European-originated children.¹³ They had good reliability in Australian and Middle Eastern ethnicity but were less reliable in Asian people. In addition to this, the evaluation of bone age with the GP method is also time-consuming; it takes a lot of time to evaluate the age of the bones individually with high accuracy when performed manually.¹⁴ Furthermore, one of the major disadvantages of manual bone age assessment with the GP method is the possible risk of high inter and intraobserver error.¹⁵ Therefore, before the comparison of manual bone age assessment with an AI-based system, the interobserver agreement between manual assessments performed by two radiologists was calculated and yielded an ICC of 0.961, thus establishing a solid basis for comparison of the AI-based measurements.

AI-based bone age calculation systems, developed to overcome all these disadvantages of manual calculation, can identify the morphological features of bone ossification automatically and provide rapid information about the patient's bone age. Therefore, this has resulted in a more objective and efficient method for assessing bone age.¹⁶

Numerous studies have demonstrated that newly developed AI technologies and software can accurately perform bone age assessments, surpassing the accuracy of the GP method.^{1,4,9,15} Furthermore, these studies have shown that AI-based assessments exhibit excellent agreement with assessments made by experienced human observers.¹ In their study to compare deep learning systems, including AlexNet, GoogleNet and Vgg19, in performing age estimation with the Turkish population, Senel et al.¹⁷ reported a success rate of 98.39%.

Similarly, we found a high level of agreement between manual assessments (using GP) and both AI-based systems, with an ICC of 0.988 for VUNO and 0.984 for BoneXpert when the entire cohort was considered. This high correlation is particularly important given the lack of existing research on the applicability of AI-based bone age calculations in the Turkish pediatric population.

BoneXpert is an AI-based automated bone age assessment system and is known as the first AI radiology system.¹³ This method, which is based on traditional machine learning methodology, predicts bone age by considering bone shape, density, and the degree of epiphyseal fusion.^{18,19} Image analysis

predicts bone age by measuring shape, density, and texture scores at specific locations.¹⁴ If a bone's appearance falls outside the range covered by the machine learning process or if its bone age value deviates above the threshold value compared with the average of all tubular bones, it will not be included in the calculation. The final bone age is calculated using the evaluated bones. If fewer than eight bones are evaluated, the X-ray is not assessed due to possible inaccurate calculations, which is a major disadvantage of BoneXpert.²⁰ It has been validated for bone age calculation in eutrophic North American, Caucasian, African American, Hispanic, and Asian children and has also been reported to be applicable in various ethnic groups.^{19,21,22} Many published reports show a notable distinction between bone ages determined by the GP method and chronological ages in Asian children.^{23,24} Similarly, Ontell et al.²⁵ reported delayed bone age in preadolescence and increased bone age in adolescence in Asian boys. The process of skeletal maturation in Korean children is initiated at a later age and completed at an earlier age than in Caucasian children. The VUNO Korean bone age assessment method, which is based on deep learning, has demonstrated superior performance compared with the manual assessment from the GP atlas. Compared with the manual assessment with the GP atlas, the Korean model has a lower root mean square error and lower mean absolute error. VUNO is the first AI-based bone age assessment system approved by the Korean Food and Drug Administration. The system was developed by analyzing 18,940 left-hand wrist radiographs using the GP method.^{25,26} VUNO provides the most likely estimated bone ages based on the examined wrist radiography.

A subgroup analysis of the data revealed subtle differences between the calculated bone ages by BoneXpert and VUNO, particularly when examining data based on gender and age subgroups. Both VUNO and BoneXpert demonstrated a high level of agreement with manual assessments in boys. However, in girls, the correlation between VUNO and manual assessment was found to be statistically significantly higher. This finding suggests that VUNO may have an advantage in terms of sensitivity in bone age assessment in girls. All age groups provided further insights that were instrumental in evaluating the efficacy of AI-based bone age assessments. Both BoneXpert and VUNO demonstrated a high degree of correlation with manual assessment in boys over the age of 9 years and girls over the age of 8 years, al-

though a significant decrease in agreement was observed among prepubertal individuals. Nevertheless, it is noteworthy that VUNO exhibits a distinct advantage over BoneXpert in this regard.

The observed decrease in correlation was particularly evident within the prepuberty female group. This result indicates that AI-based assessments present challenges in younger age groups, particularly in prepuberty girls. Bland-Altman plots serve to illustrate this observation, demonstrating that both BoneXpert and VUNO correlate with manual assessments during prepuberty, resulting in reduced estimations of bone age. The decreased correlation of BoneXpert and VUNO with manual assessments in the prepubertal group, particularly in prepubertal girls, may be attributed to technical differences between the two methods. In contrast to VUNO, BoneXpert is designed to measure tubular bones exclusively, with the capacity to automatically exclude those with suboptimal quality or anomalous bone structure. The BoneXpert technique does not analyze carpal bones in the calculation of bone age, whereas VUNO uses all bones included in radiographs for the evaluation of bone age and automatically preprocesses radiographs for evaluation without rejecting them. Tubular bones exhibit endochondral ossification, while carpal bones exhibit intramembranous ossification; membranous ossification is less reliant on growth hormones than endochondral ossification, and the maturation of carpal bones occurs earlier than in tubular bones. As a result, the maturation of carpal bones varies significantly between individuals during the prepubertal period in comparison with tubular bones.¹⁸ The exclusion of carpal bones from the assessment of bone age may result in a slight reduction in the BoneXpert success in the prepubertal period in comparison with VUNO.

Our study had some limitations, including a small sample size and the fact that it focused on a single, heterogeneous ethnicity. Additionally, the study did not include participants aged <2 years or >15 years due to the unsuitability of the GP manual method for evaluating bone age in these age groups.

In conclusion, our study confirms that BoneXpert and VUNO are reliable AI-based systems for assessing bone age in the Turkish pediatric population. VUNO has a higher agreement with manual correlation than BoneXpert, particularly in prepubescent girls.

Conflict of interest disclosure

Evrin Özmen, MD, is Section Editor in Diagnostic and Interventional Radiology. She had no involvement in the peer-review of this article and had no access to information regarding its peer-review. Other authors have nothing to disclose.

References

1. Martin DD, Calder AD, Ranke MB, Binder G, Thodberg HH. Accuracy and self-validation of automated bone age determination. *Sci Rep*. 2022;12(1):6388. [\[CrossRef\]](#)
2. Greulich WW, Pyle IS. Radiographic atlas of skeletal development of the hand and wrist. 2nd ed. Stanford: Stanford University Press; 1959. [\[CrossRef\]](#)
3. Prokop-Piotrkowska M, Marszałek-Dziuba K, Moszczyńska E, Szalecki M, Jurkiewicz E. Traditional and new methods of bone age assessment-an overview. *J Clin Res Pediatr Endocrinol*. 2021;13(3):251-262. [\[CrossRef\]](#)
4. Kim PH, Yoon HM, Kim JR, et al. Bone age assessment using artificial intelligence in Korean pediatric population: a comparison of deep-learning models trained with healthy chronological and greulich-Pyle ages as labels. *Korean J Radiol*. 2023;24(11):1151-1163. [\[CrossRef\]](#)
5. IBM Corp. Released 2021. IBM SPSS Statistics for Windows, Version 28.0. Armonk, NY: IBM Corp. [\[CrossRef\]](#)
6. Kinckle DE. Applied Statistics for the behavioral sciences. 2nd ed. Houghton Mifflin Harcourt; 1988. [\[CrossRef\]](#)
7. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;82(2):420-428. [\[CrossRef\]](#)
8. National Pediatric Society of Turkey. Promed-Mail. Accessed July 21, 2024. [\[CrossRef\]](#)
9. Martin DD, Wit JM, Hochberg Z, et al. The use of bone age in clinical practice - part 1. *Horm Res Paediatr*. 2011;76(1):1-9. [\[CrossRef\]](#)
10. Artioli TO, Alvares MA, Carvalho Macedo VS, et al. Bone age determination in eutrophic, overweight and obese Brazilian children and adolescents: a comparison between computerized BoneXpert and Greulich-Pyle methods. *Pediatr Radiol*. 2019;49(9):1185-1191. [\[CrossRef\]](#)
11. Kaplowitz P, Srinivasan S, He J, McCarter R, Hayeri MR, Sze R. Comparison of bone age readings by pediatric endocrinologists and pediatric radiologists using two bone age atlases. *Pediatr Radiol*. 2011;41(6):690-693. [\[CrossRef\]](#)
12. Halabi SS, Prevedello LM, Kalpathy-Cramer J, et al. The RSNA pediatric bone age machine learning challenge. *Radiology*. 2019;290(2):498-503. [\[CrossRef\]](#)
13. Alshamrani K, Hewitt A, Offiah AC. Applicability of two bone age assessment methods to children from Saudi Arabia. *Clin Radiol*. 2020;75(2):156. [\[CrossRef\]](#)
14. Thodberg HH, Thodberg B, Ahlkvist J, Offiah AC. Autonomous artificial intelligence in pediatric radiology: the use and perception of BoneXpert for bone age assessment. *Pediatr Radiol*. 2022;52(7):1338-1346. [\[CrossRef\]](#)
15. Gräfe D, Beeskow AB, Pfäffle R, Rosolowski M, Chung TS, DiFranco MD. Automated bone age assessment in a German pediatric cohort: agreement between an artificial intelligence software and the manual Greulich and Pyle method. *Eur Radiol*. 2024;34(7):24407-24413. [\[CrossRef\]](#)
16. Dallora AL, Anderberg P, Kvist O, et al. Bone age assessment with various machine learning techniques: a systematic literature review and meta-analysis. *PLoS One*. 2019;14:e0220242. [\[CrossRef\]](#)
17. Senel FA, Dursun A, Ozturk K, Ayyildiz VA. Determination of bone age using deep convolutional neural networks. *Ann Med Res*. 2021;28(7):1381-1386. [\[CrossRef\]](#)
18. Acheson RM, Vicinus JH, Fowler GB. Studies in the reliability of assessing skeletal maturity from x-rays. 3. Greulich-Pyle Atlas and Tanner-Whitehouse method contrasted. *Hum Biol*. 1966;38(3):204-218. [\[CrossRef\]](#)
19. Thodberg HH, Kreiborg S, Juul A, Pedersen KD. The BoneXpert method for automated determination of skeletal maturity. *IEEE Trans Med Imaging*. 2009;28(1):52-66. [\[CrossRef\]](#)
20. Booz C, Yel I, Wichmann JL, 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(1):6. [\[CrossRef\]](#)
21. Thodberg HH, Savendahl L. Validation and reference values of automated bone age determination for four ethnicities. *Acad Radiol*. 2010;17(11):1425-1432. [\[CrossRef\]](#)
22. Satoh M. Bone age: assessment methods and clinical applications. *Clin Pediatr Endocrinol*. 2015;24(4):143-152. [\[CrossRef\]](#)
23. 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(6):2910-2923. [\[CrossRef\]](#)
24. 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(1):228-235. [\[CrossRef\]](#)
25. Ontell FK, Ivanovic M, Ablin DS, Barlow TW. Bone age in children of diverse ethnicity. *AJR Am J Roentgenol*. 1996;167(6):1395-1398. [\[CrossRef\]](#)
26. Kim JR, Shim WH, Yoon HM, et al. Computerized bone age estimation using deep learning based program: evaluation of the accuracy and efficiency. *AJR Am J Roentgenol*. 2017;209(6):1374-1380. [\[CrossRef\]](#)