Medical Science and Discovery
ISSN: 2148-6832

Artificial Intelligence in the endocrine clinic: Automated bone age analysis in children from UAE

Elham Ahmed Elgabaly Moustafa Ahmed[1], Ajay Prasanth D'Souza[1], Mireille El Bejjani[2], Nandu Kumar Thalange[2]*

1 Dept. of Radiology, Al Jalila Children's Hospital, Dubai, UAE
2 Dept. of Endocrinology, Al Jalila Children's Hospital, Dubai, UAE

*** Corresponding Author:** Nandu Thalange **E-mail:** nandu.thalange@ajch.ae

ABSTRACT

Objective: Artificial intelligence (AI) is playing an increasing role in patient assessment. AI bone age analysis is such a tool, but its value in Arabic children presenting to an endocrine clinic has not been explored. We compared results from an experienced pediatric radiologist and the AI bone age system, BoneXpert (BX), (Visiana, Denmark) to assess its utility in a cohort of children presenting to the Al Jalila Children's Specialty Hospital endocrine service.

Materials and Methods: We conducted a retrospective chart review of 47 children with growth disorders, initially assessed by a single experienced radiologist and subsequently by BX, to confirm the usefulness of the BX system in our population. The results of the analyses were analysed using a Bland-Altman plot constructed to compare differences between the radiologist's interpretation and BX across the available range of bone age.

Results: Forty-four of the patient x-ray images were analysed by BX. Three X-ray images were rejected by BX due to post-processing artifacts, which prevented computer interpretation. For the remaining 44 X-rays, there was a close correlation between radiologist and BX results ($r=0.93$; $p<0.00001$). Two radiographs were identified with a large discrepancy in the reported bone ages. Blinded, independent re-evaluation of the radiographs showed the original manually interpreted bone age to have been erroneous, with the BX results corresponding closely to the amended bone age. A small positive bias was noted in bone age (+0.39 years) in the BX analyses, relative to manual interpretation.

Conclusions: AI bone age analysis was of high utility in Arabic children from UAE presenting to an endocrine clinic, with results highly comparable to an experienced radiologist. In the two cases where a large discrepancy was found, independent re-evaluation showed AI analysis was correct.

Keywords: Bone Age, Greulich & Pyle, Growth, Puberty

Research Article

Received 04-07-2021

Accepted 20-07-2021

Available Online: 23-07-2021

Published 30-07-2021

INTRODUCTION

Assessment of bone age is a critical tool in the investigation of disorders affecting growth and puberty. (**1, 2, 3**) A child growing as expected for age will have a bone age close to his chronological age. However, a child with pathological short stature or delayed puberty is likely to have significant bone age delay. In contrast, a child who has precocious puberty will have a relatively advanced bone age. (**4**) Accuracy in the determination of bone age is crucial to correct diagnosis and may be used to determine the height prognosis, which is a major concern to parents and children. (**5**)

Additionally, serial bone age assessment is invaluable in disease monitoring, as in congenital adrenal hyperplasia, where excessive steroid replacement retards growth and delays bone age, whereas the inadequate replacement is reflected in the advanced bone age and compromised final height. (**6**) Similarly, the effect of treatments such as gonadotropin releasing hormone agonists (used for central precocious puberty) or growth hormone can be monitored through serial measurements of bone age. (**4**)

http://dx.doi.org/10.36472/msd.v8i7.572

In order for bone age measurement to have maximum utility, accuracy and precision of bone age assessment is critical. This is particularly true for serial measurements. There are two widely used approaches to determine bone age from a hand and wrist radiograph – the Greulich & Pyle method (G&P) – in which, most commonly, bone age is determined by comparing a hand-wrist radiograph of a child with the age-matched standard radiographs shown in the Greulich & Pyle atlas. (7) This method is straightforward and quick and hence widely used, but when bone age is assessed in this way, it is typically somewhat imprecise (e.g. 'the bone age is between 7 and 7.5 years'). The Tanner Whitehouse method, now in its third iteration (TW3) depends on assessing and scoring the skeletal maturity of each individual bone of the hand, (8) and hence is time-consuming and laborious compared with the G&P method and while more precise, it has less utility for determining height prognosis, which is a major objective of bone age assessment. (9) The G&P bone age method was originally compiled from bone age assessments of largely Caucasian children from Ohio, USA, of good socioeconomic status, but subsequently has been validated in many populations worldwide, although important ethnic differences do exist.(10)

Use of Artificial Intelligence (AI) to assess bone age has been attempted for over 30 years. The first step in the use of AI was the HANDX system developed in 1989. (11) HANDX was a semi-automated system able to detect skeletal growth abnormalities in children. The PROI system (1991) (12) and Computer-based Skeletal Aging Scoring (CASAS) system (1994) (13) allowed assessment of bone age in a highly reproducible and accurate way. However, although of increased accuracy, compared with manual assessment of bone age, CASAS was more labor-intensive than manual reading, and hence not viable for clinical practice. Improvements in computational power and AI analysis techniques finally led to the realization of viable commercial solutions, such as BoneXpert™ (BX) (Visiana, Copenhagen, Denmark) in 2008. (14) BX is an AI system that calculates bone age by analyzing the shape and density of 21 bones (ulna, radius, metacarpals and phalanges). The borders of the bones are detected using a machine-learning algorithm which has learned to locate landmarks on the bones, and the normal anatomy of each bone. The information is used to generate a G&P and TW3 bone age, and more recently, BX version 3.0 introduced in September 2019 also reports carpal bone age. BX is now widely used in Europe and has additionally been validated in multiple ethnic populations, worldwide. (15) This system uses AI assessment of bone age to derive G&P and TW bone ages, with a high degree of precision. It has been progressively refined since its introduction and has attained a level of precision and accuracy superior to conventional radiographical interpretation. Indeed, the current software (version 3.03) is equivalent to the combined assessment of five expert radiologists. (16)

An additional utility of the BX system is that it includes a measure of bone density – Bone Health Index – which has been found to correspond very closely to measurements by dual energy X-ray absorptiometry (DXA). (17, 18)

The use of BX to determine bone age also yields actionable results immediately, allowing real-time decision-making on the basis of the result, rather than awaiting a manual report, and aids the radiologist, the clinician and the patient. Having introduced BX, we conducted a retrospective validation of the previously obtained bone age results.

MATERIAL and METHODS

We examined bone age x-rays from 47 children with disorders of growth and puberty attending endocrine clinics between August 2017 until December 2018, who required bone age assessment. The children were aged between 3.75 and 14.95 years, (27 males) assessed by a single experienced radiologist (EA) before the introduction of BX, using the G&P method. We compared the results to those obtained with BoneXpert (BX), (version 2.0.1.3, Visiana, Denmark), in December 2018. Hand radiographs for bone age were identified retrospectively from the Picture Archiving and Communication System (PACS) in standard digital imaging (DICOM) format. **Figure 1** shows an example of an AI generated BX report.

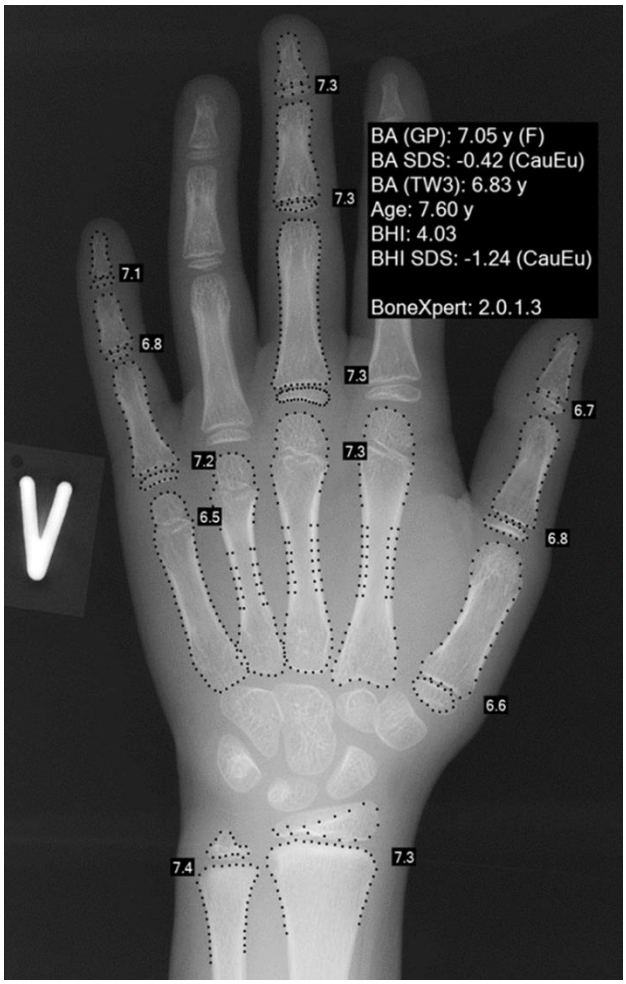

Figure 1. An example BX report. The annotated image is the result of the analysis, containing the following: **BA (GP):** Greulich-Pyle bone age (gender, M or F), **BA SDS:** Bone Age Standard deviation score of GP bone age, **BA (TW3):** Tanner-Whitehouse bone age; Chronological age, **BHI:** Bone health index – a measure of bone density, **BHI SDS:** Standard deviation score of BHI

http://dx.doi.org/10.36472/msd.v8i7.572

Statistical analysis was performed using Microsoft Excel (Microsoft Corp., Redmond, WA, USA). The bone ages obtained by the two methods were compared and the correlation coefficient and statistical significance were determined. A Bland-Altman plot was constructed to determine analogy between the manual bone age and BX results and to identify any systematic bias.

The analysis was a retrospective review of routine radiographic data obtained as part of normal clinical practice and as such, ethical approval was deemed unnecessary.

RESULTS

Of the 47 X-rays analysed, three could not be evaluated by BX due to image processing artifacts. Image processing ('edge enhancement') is commonly employed to improve the clarity of radiographs for manual interpretation, but the resultant digital "noise" may render the X-ray unsuitable for AI analysis. (**14**)

Of the 44 analyses where comparison was possible, there was a high correlation between the two bone age measures ($r= 0.93$, $p<0.00001$). There were 2 outliers, which differed by more than 3 years. The outliers were independently reviewed by co-author, APD, who was blinded to the original bone age assessments. He reported bone ages that corresponded closely to the bone age determined by BX. The range excluding these two outliers was -1.22 years to +2.30 years, which were within the expected range of bone age, as shown by the Bland-Altman plot (**Figure 2**).

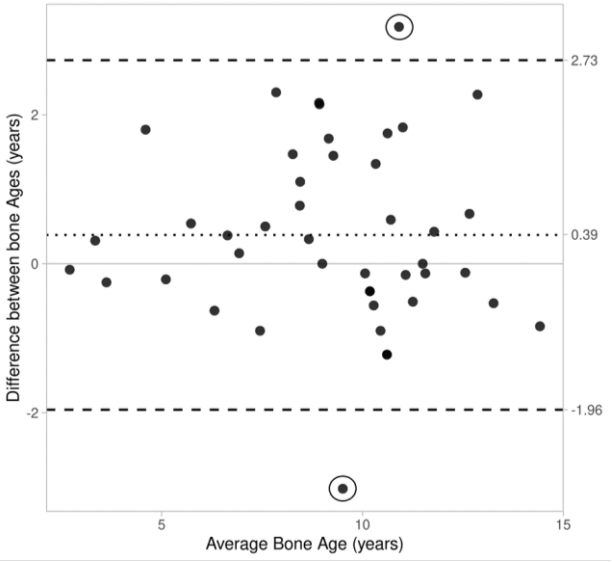

Figure 2. Bland-Altman plot demonstrating difference between bone age measurements plotted against the average bone age. The dashed lines indicate the Limits of Agreement between which 95% of measurements are expected to lie. Two outliers (highlighted with circles) are clearly evident – one above, and one below the upper and lower limits of agreement, respectively. There is a small positive bias of 0.39 years indicating that, on average, the bone age determined by BX is 0.39 years greater than manual reading

This clearly demonstrated the two outlying values, (circled) but the remaining observations were in close agreement, all lying within the limits of agreement, albeit that there was a systematic positive bias of +0.39 years, indicating that BX systematically scored bone ages 0.39 years higher than manual reading. The data spread and bias were similar when the results were analysed by gender (data not shown).

DISCUSSION

In our retrospective analysis of bone ages of children attending an endocrine clinic, we found a close correlation between an experienced radiologist's interpretations and AI Bone age analysis. Two bone age results differed markedly from those determined by BX and following independent review of the images, blinded to the original bone age estimation, the original bone age was amended and was in agreement with that given by BX. With the exception of these two outliers, there was close agreement between the bone ages, indicating that the discrepancy arose from erroneous manual reading of bone age and not from software error. This suggests that for practical purposes, AI bone age analysis using BX is a suitable tool for the evaluation of bone age in Arabic children attending an endocrine clinic in UAE.

To a certain extent, bone age is an artificial construct. The assessment of bone age using a largely Caucasian reference population from the 1940s takes no account of secular changes in bone age over time, (**19**) nor of important ethnic differences. (**10**) The mean age of puberty is falling worldwide, in both boys and girls and this is reflected by a relative advance in bone age compared with chronological age. (**19**) Nonetheless, BX has been validated in multiple ethnic populations and despite such systematic differences, has nevertheless been found suitable for clinical use. (15, 20, 21, 22)

The finding of a small systematic bias in bone age assessment between the software and an experienced radiologist (+0.39 years) could simply represent the skills of the assessor, secular trend in bone age or a true ethnic difference in the application of the G&P bone age to an Arabic population. Ethnic variation in bone age using the G&P method is well-recognised (10) and this seems the most likely explanation.

Recently, the validity of the G&P and TW3 methods of bone age measurement was assessed in a large, predominantly Arabic Saudi population of children attending an emergency Dept. for reasons unrelated to growth or puberty. (23) The authors found that there were small systematic differences in bone age estimation using either TW3 or G&P. They manually reviewed hand and wrist radiographs from 420 children and compared the bone age with the patient's chronological age. They concluded that G&P bone age consistently underestimated true chronological age in girls but overestimated it in boys. Analysis using BX was possible in only 210 children (50%) of the cohort and they found the same pattern, with a mean difference of -2 months in girls and +2.5 months in boys. H owever, we would observe that although the bone age significantly differed from the chronological age in statistical terms, in clinical practice this difference is inconsequential.

Moreover, the endocrinologist is not seeking to estimate chronological age, but to identify discrepancies between bone age and chronological age, which may highlight the presence of an endocrine disorder.

Potentially, a significant limitation of our study was the relatively small number of radiographs available for analysis, which was inevitable in a newly established children's hospital – the endocrine service was established only in August 2017. However, the correlation between assessment by an experienced radiologist and BX was highly statistically significant ($p<0.00001$) and corroborated by the Bland-Altman analysis (**Figure 2**).

Clearly, children attending an endocrine clinic with concerns regarding growth and puberty are not representative of all children in our population. However, our primary purpose in conducting the study was to assure ourselves that AI bone age analysis produced comparable results to manual reading in a cohort of Arabic children undergoing bone age analysis for the evaluation of endocrine disorders, and in this respect it excelled. Our data gave us confidence that BX is indeed a suitable tool in the assessment of bone age in our local population. Since its introduction, we have come to rely on BX and at the time of writing (July 2021) we have analysed a further 931 x-rays with BX. The added efficiency of the AI system which provides results in moments, means we can make immediate decisions based on the bone age. This convenience and speed aids clinical management, to the benefit of the patient clinician and radiologist alike. Moreover, serial bone age assessments may be performed to monitor progress, with the added confidence of a precise bone age measurement. This is in contrast to relying on often conflicting measurements of different human observers, based on subjective impressions from a bone age atlas. While experts may achieve intra-observer error as low as 0.25 years (**24**), Bull et al. found that in clinical practice, intra-observer error averaged 0.82 years for G&P bone age. (**25**) In comparison, even the first iteration of BX achieved a precision of 0.17 years, and this has been progressively improved with subsequent versions. Indeed, with the latest version (3.0.3), BX is equivalent to the combined assessment of 5 expert raters. (**16**) Thus, for routine clinical use, bone age assessment by BX is superior to manual bone age interpretation. (**26**) The additional benefit of an estimate of bone density (Bone Health Index) provides a valuable extra dimension to the assessment. We are now at a point where AI bone age analysis is so clearly superior to manual reading that it should become the tool of choice for bone age analysis. (**27**) For the present, BX cannot replace radiological review of bone age X-rays as it is not capable of identifying morphological abnormalities such as rickets or features of skeletal dysplasia, but it has replaced the onerous and burdensome task of manual bone age evaluation.

CONCLUSIONS

AI bone age assessment gives a speedy, accurate and precise result and does away with subjective visual interpretation and obviates the issue of inter- and intra-observer variability, thereby reducing the reporting burden on radiologists and facilitating patient care, through enabling a "one stop" visit. Our experience shows that AI bone age analysis is of high utility in evaluating bone age in Arabic children from UAE presenting to our endocrine service, with results highly comparable to those obtained by an experienced consultant pediatric radiologist.

Acknowledgments: We would like to acknowledge the invaluable practical assistance and ongoing support provided by Mr Sarfaraz Ulde, PACS Administrator in the Diagnostic Imaging Dept. at Al Jalila Children's Hospital in implementing and managing the BoneXpert system.

Author Contributions: EAEMA, APD, ME, NKT: Data collection, Formal analysis, Methodology, Project administration, Statistical Analyses, **NKT:** Article writing and revisions

Financial & competing interest's disclosure: The authors have no relevant affiliations or financial involvement with any organisation or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Ethical approval: Conflict of interest: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. This research did not receive and specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Conflict of interest: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. This research did not receive and specific grant from funding agencies in the public, commercial or not-for-profit sectors.

REFERENCES

1. Creo AL, Schwenk WF 2nd. Bone Age: A Handy Tool for Pediatric Providers. Pediatrics. 2017 Dec;140(6):e20171486. https://doi.org/10.1542/peds.2017-1486 PMID:29141916

2. 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):1–9. https://doi.org/10.1159/000329372 PMID:21691054

3. Labarta JI, Ranke MB, Maghnie M, Martin D, Guazzarotti L, Pfäffle R, et al. Important Tools for Use by Pediatric Endocrinologists in the Assessment of Short Stature. J Clin Res Pediatr Endocrinol. 2021 Jun;13(2):124–35. https://doi.org/10.4274/jcrpe.galenos.2020.2020.0206 PMID:33006554

4. Vargas Trujillo M, Dragnic S, Aldridge P, Klein KO. Importance of individualizing treatment decisions in girls with central precocious puberty when initiating treatment after age 7 years or continuing beyond a chronological age of 10 years or a bone age of 12 years. J Pediatr Endocrinol Metab. 2021 Apr;34(6):733–9. https://doi.org/10.1515/jpem-2021-0114 PMID:33856747

5. Martin DD, Schittenhelm J, Thodberg HH. Validation of adult height prediction based on automated bone age determination in the Paris Longitudinal Study of healthy children. Pediatr Radiol. 2016 Feb;46(2):263–9. https://doi.org/10.1007/s00247-015-3468-8 PMID:26573823

6. Martin DD, Wit JM, Hochberg Z, van Rijn RR, Fricke O, Werther G, et al. The use of bone age in clinical practice - part 2. Horm Res Paediatr. 2011;76(1):10–6. https://doi.org/10.1159/000329374 PMID:21691055

7. Greulich W, Pyle I. Radiographic atlas of skeletal development of the hand and wrist. London: Stanford University Press; 1959. https://doi.org/10.1097/00000441-195909000-00030.

8. Tanner JM, Healy MJ, Goldstein H, Cameron N. Assessment of skeletal maturity and prediction of adult height TW3 method. London: WB Saunders; 2001.

9. Thodberg HH, Jenni OG, Caflisch J, Ranke MB, Martin DD. Prediction of adult height based on automated determination of bone age. J Clin Endocrinol Metab. 2009 Dec;94(12):4868–74. https://doi.org/10.1210/jc.2009-1429 PMID:19926715

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 Jun;29(6):2910–23. https://doi.org/10.1007/s00330-018-5792-5 PMID:30617474

11. Michael DJ, Nelson AC. HANDX: a model-based system for automatic segmentation of bones from digital hand radiographs. IEEE Trans Med Imaging. 1989;8(1):64–9. https://doi.org/10.1109/42.20363 PMID:18230501

12. Pietka E, McNitt-Gray MF, Kuo ML, Huang HK. Computer-assisted phalangeal analysis in skeletal age assessment. IEEE Trans Med Imaging. 1991;10(4):616–20. https://doi.org/10.1109/42.108597 PMID:18222868

13. Tanner JM, Oshman D, Lindgren G, Grunbaum JA, Elsouki R, Labarthe D. Reliability and validity of computer-assisted estimates of Tanner-Whitehouse skeletal maturity (CASAS): comparison with the manual method. Horm Res. 1994;42(6):288–94. https://doi.org/10.1159/000184211 PMID:7698726

14. Thodberg HH, Kreiborg S, Juul A, Pedersen KD. The BoneXpert method for automated determination of skeletal maturity. IEEE Trans Med Imaging. 2009 Jan;28(1):52–66. https://doi.org/10.1109/TMI.2008.926067 PMID:19116188

15. Thodberg HH, Sävendahl L. Validation and reference values of automated bone age determination for four ethnicities. Acad Radiol. 2010 Nov;17(11):1425–32. https://doi.org/10.1016/j.acra.2010.06.007 PMID:20691616

16. 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 Feb;290(2):498–503. https://doi.org/10.1148/radiol.2018180736 PMID:30480490

17. Schündeln MM, Marschke L, Bauer JJ, Hauffa PK, Schweiger B, Führer-Sakel D, et al. A Piece of the Puzzle: The Bone Health Index of the BoneXpert Software Reflects Cortical Bone Mineral Density in Pediatric and Adolescent Patients. PLoS One. 2016 Mar;11(3):e0151936. https://doi.org/10.1371/journal.pone.0151936 PMID:27014874

18. Leijten AD, Hampsink B, Janssen M, Klein WM, Draaisma JM. Can digital X-ray radiogrammetry be an alternative for dual-energy X-ray absorptiometry in the diagnosis of secondary low bone quality in children? Eur J Pediatr. 2019 Sep;178(9):1433–41. https://doi.org/10.1007/s00431-019-03425-5 PMID:31352546

19. Boeyer ME, Sherwood RJ, Deroche CB, Duren DL. Early Maturity as the New Normal: A Century-long Study of Bone Age. Clin Orthop Relat Res. 2018 Nov;476(11):2112–22. https://doi.org/10.1097/CORR.0000000000000446 PMID:30179948

20. De Sanctis V, Di Maio S, Soliman AT, Raiola G, Elalaily R, Millimaggi G. Hand X-ray in pediatric endocrinology: skeletal age assessment and beyond. Indian J Endocrinol Metab. 2014 Nov;18(7 Suppl 1):S63–71. https://doi.org/10.4103/2230-8210.145076 PMID:25538880

21. Zhang SY, Liu G, Ma CG, Han YS, Shen XZ, Xu RL, et al. Automated determination of bone age in a modern chinese population. ISRN Radiol. 2013 Feb;2013:874570. https://doi.org/10.5402/2013/874570 PMID:24967289

22. Koc U, Taydaş O, Bolu S, Elhan AH, Karakas SP. The Greulich-Pyle and Gilsanz-Ratib atlas method versus automated estimation tool for bone age: a multi-observer agreement study. Jpn J Radiol. 2021 Mar;39(3):267–72. https://doi.org/10.1007/s11604-020-01055-8 PMID:33067733

23. Alshamrani K, Hewitt A, Offiah AC. Applicability of two bone age assessment methods to children from Saudi Arabia. Clin Radiol. 2020 Feb;75(2):156.e1–9. https://doi.org/10.1016/j.crad.2019.08.029 PMID:31706569

24. Berst MJ, Dolan L, Bogdanowicz MM, Stevens MA, Chow S, Brandser EA. Effect of knowledge of chronologic age on the variability of pediatric bone age determined using the Greulich and Pyle standards. AJR Am J Roentgenol. 2001 Feb;176(2):507–10. https://doi.org/10.2214/ajr.176.2.1760507 PMID:11159105

25. Bull RK, Edwards PD, Kemp PM, Fry S, Hughes IA. Bone age assessment: a large scale comparison of the Greulich and Pyle, and Tanner and Whitehouse (TW2) methods. Arch Dis Child. 1999 Aug;81(2):172–3. https://doi.org/10.1136/adc.81.2.172 PMID:10490531

26. van Rijn RR, Thodberg HH. Bone age assessment: automated techniques coming of age? Acta Radiol. 2013 Nov;54(9):1024–9. https://doi.org/10.1258/ar.2012.120443 PMID:24179234

27. Lee BD, Lee MS. Automated Bone Age Assessment Using Artificial Intelligence: The Future of Bone Age Assessment. Korean J Radiol. 2021 May;22(5):792–800. https://doi.org/10.3348/kjr.2020.0941 PMID:33569930