Journal of Dharma 49, 3 (July–September 2024), 409–422

AUTOMATED BONE AGE ESTIMATION USING ARTIFICIAL INTELLIGENCE - Boneage.io® - IN HEALTHY CHILDREN

Jeong Eun Moon* **and Yong Jin Cho****

Abstract: Bone age assessment is vital for diagnosing and managing growth disorders. Traditional methods like the Greulich & Pyle Atlas and Tanner-Whitehouse 3 (TW3) method are either quick but less accurate or detailed but labor-intensive. This study evaluates the accuracy and reproducibility of Boneage.io®, a cloud-based AI solution using the TW3 method to estimate bone age in healthy Korean children aged 6–13. A total of 1,040 radiographs were analyzed; the results showed minimal deviation between estimated bone age and chronological age, with Cohen's D effect sizes of 0.021566 for boys and 0.026172 for girls. Boneage.io® provides reliable, real-time monthly bone age results, effectively addressing challenges of traditional methods and demonstrates high accuracy and reproducibility for clinical use.

Keywords: Bone Age, Boneage.io®, Healthy Children, Cloud-based AI, Reproducibility, Accuracy, Artificial Intelligence

1. Introduction

Bone age and chronological age provide distinct measures of an individual's maturity. Chronological age represents the time

---

* Moon, Jeong Eun (Principal Author) is a professor at the Department of Nursing, Honam University, Gwangju, Republic of Korea and is interested in data science and statistics.

** Cho, Yong Jin (Corresponding Author) is a professor cum pediatric orthopedic surgeon at Chosun University College of Medicine Gwangju, Republic of Korea and is interested in AI module development: E-mail: choisidoru@chosun.ac.kr

elapsed since birth, while skeletal or bone age reflects biological development, making it a vital parameter in medicine and forensics. Bone age assessment is particularly important for diagnosing and managing growth disorders. In healthy individuals, bone age typically corresponds to chronological age within a margin of six months (Buckler 761-763). Delayed bone age may point to conditions like constitutional growth retardation, malnutrition or genetic disorders, although the exact cause is not always clear. On the other hand, advanced bone age suggests accelerated growth, which may limit adult height. Despite advancements in bone age estimation techniques, the process still heavily relies on radiologists' subjective assessments.

Conventional manual methods for bone age assessment are complex and time-consuming, often leading to reduced accuracy due to fatigue and variability among evaluators. This highlights the growing need for automated, computer-assisted techniques in this field. Bone age assessment approaches differ based on the anatomical region being examined. For instance, Pyle & Herr's method focuses on the knee, Acheson's Oxford method on the pelvis, Greulich & Pyle's approach on the hand, Herr's on the foot, and Tanner & Whitehouse's method on the hand and wrist (Satoh 143-152; Shin et al. 237-243).

Among these, the Greulich & Pyle Atlas remains the most commonly used reference for bone age estimation. This method involves comparing left hand and wrist radiographs with age-specific reference images from the Atlas. While it is quick and convenient, its precision is limited. To enhance accuracy, Tanner & Whitehouse's method was introduced, which assigns scores to individual hand and wrist bones based on their maturity, with the scores summed to provide a more detailed measurement. Although more precise than the Greulich & Pyle Atlas method, the Tanner & Whitehouse approach requires greater expertise and time to implement effectively. The Tanner & Whitehouse's method also offers significant advantages over the Greulich & Pyle Atlas in the context of AI-based applications, making it particularly valuable for modern, automated bone age assessment techniques (Hamd et al. 199; Cundy et al. 513-515; Hwang et al. 683-691; Mehta et al. 3093-3096; Van Steenkiste et al. 674-677).

The demand for automated bone age determination has steadily increased over the years. Commercially, BoneXpert® software has been available in Europe since 2008, employing a layered segmentation approach to analyze 13 bones, including the ulna, radius and the first, third, and fifth phalanges and metacarpals. This analysis is based on shape, location and density distribution (Thodberg et al. 1338-1346). In Korea, a cloud-based service called Boneage.io®, offered by www.healthhub.kr, became available in 2022. This platform utilizes left hand and wrist radiographs and applies the Tanner & Whitehouse 3 (TW3) method to deliver real-time bone age estimations in monthly units. In this study, we aim to assess the accuracy and reproducibility of Boneage.io®'s monthly bone age estimations, which leverage left hand and wrist radiographs to provide automated determinations of bone age based on the TW3 method in healthy Korean children.

2. Material and Methods

2.1. *Patient Selection and Study Design*

From January 1, 2020 to December 31, 2023, left hand and wrist radiographs were collected from Chosun University Hospital, matched with each patient's gender and date of birth for bone age assessment. This study targeted left hand and wrist radiographs of normal children between the ages of 6 and 13 years. Exclusion criteria were as follows: (1) left hand and wrist radiographs not taken according to protocol (2) cases where the exact monthly age at the time of imaging could not be verified (3) a confirmed diagnosis of a condition that could affect bone age (4) cases with unclear sex information. The underlying diseases and indications for exclusion were presented as follows: (1) Children with endocrine, nutritional, and metabolic diseases, such as hypopituitarism (2) Children have chromosomal anomalies such as Down syndrome or Turner syndrome (3) Children have mental and neurologic disorders, such as cerebral palsy (4) Children with known growth disturbances due to medication. (5) Children with abnormal perinatal conditions.

The local ethics committee granted ethical approval for this retrospective study (CHOSUN-2020-03-012) and the ethics board waived written and informed consent because of the study's retrospective nature. All methods and procedures were performed following the relevant guidelines and regulations by institutional review boards.

During the study period, a total of 2,049 left hand and wrist radiographs were taken at Chosun University Hospital for bone age assessment. Among these, 1,790 radiographs were of children aged 6 to 13 years. Excluding 726 images from patients with diagnoses potentially affecting bone age and 24 images that did not meet protocol standards due to poor quality or positioning, 1,040 radiographs were ultimately included in the study. No cases were identified where monthly age could not be verified or where gender was undetermined at the time of assessment.

Left hand and wrist radiographs were searched from the PACS (Picture Achieving and Communication System). The DICOM (Digital Imaging and Communications in Medicine) files were anonymized and de-identified to ensure confidentiality. After labeling, the DICOM files was removed, and only the patient's age and sex were anonymously matched and stored.

2.2. *Artificial Intelligence Model*

Bone age assessment was conducted using the cloud-based medical imaging AI solution "AI-based Bone Age Estimation and Height Prediction Report Service", named Boneage.io® provided by www.healthhub.kr. This service utilizes the TW3 method to analyze bone age from left hand and wrist radiographs images and delivers predicted values in monthly units in real-time. The software has received approval from both the Korean Ministry of Food and Drug Safety and European CE certification, and is characterized by cloud-based, autonomous AI-driven automation, high reliability through TW3 methodology, excellent accuracy with minimal variance from radiologist assessments, and comprehensive data presentation.

To access Boneage.io®, healthcare institutions must first register and verify on the website, after which they can log in with a provided ID and connect to the PACS system. After logging in, users can click the "New Request" button, follow prompts to upload Radiographs images, and obtain the result report within 5-10 seconds. For this study, only the predicted bone age in months from the provided report was used. While TW3 analysis generally applies to ages 3 and above, the www.healthhub.kr system restricts its application range to ages 7 through 15, based on clinical trials validating device safety and efficacy. However, this study focused on participants aged 6 to 13.

2.2. ***Statistical Analysis***

The paired t test was used to evaluate the mean changes by sex and age groups with calculated p-values. Cohen's D was calculated by dividing the absolute mean of the differences between the two comparison measurements with the chronological age's SD. Cohen's D as an index of standardized difference was considered as several levels of clinical significance (Table 1) (Wang et al. 937-943). Statistical analyses were performed using SAS 9.4 (SAS institute Inc., Cary, North Carolina).

3. Results

With the bone age assessment, mean bone age in boys was delayed by 1.7 to 1.27 months compared with mean chronological age in patients aged 6 to 9 years and advanced by 0.39 to 2.86 months in patients aged over 10 years. For girls, the bone age assessment was delayed by 1.15 to 3.22 months compared with the mean chronological ages in patients aged 6 to 10 years and advanced by 1.17 to 2.92 months in patients aged over 11 years (Figure 1). The effect size of Cohen's D using Boneage.io® methods for boys were 0.021566 and for girls they were 0.026172. Comparison of the estimated bone and chronological age in boys and girls were shown in Table 2.

4. Discussion

Bone age reflects a child's current development and sexual maturity and can help predict final height. Therefore, assessment of bone age is important for diagnosis and monitoring of treatment of growth disorders and endocrine disorders. The Greulich & Pyle Atlas method, which measures bone age by comparing radiographs images to a reference Greulich & Pyle Atlas, is quick and easy to use, but it has the disadvantage of being less accurate (Shah et al. 240-246). In contrast, the Tanner & Whitehouse's method, which assigns scores based on the maturity of bones in the hand and wrist and then sums these scores to determine bone age, is more detailed and accurate than the Greulich & Pyle Atlas method, but it requires more time to master and presents challenges for practical clinical use (Prokop-Piotrkowska et al. 251-262). To reflect the trend of faster growth and bone maturation in children, the Tanner & Whitehouse's 3 method was developed, following the earlier Tanner & Whitehouse's 2 method. The Tanner & Whitehouse's 3 method, applied to children in North America and Europe during the 1980s and 1990s, simplifies the calculation of maturity scores by focusing solely on the RUS (radius-ulna-short bones) score (Booz et al. 39-45).

Kim et al. compared bone age measurements in Korean children using the Greulich & Pyle Atlas method and the Tanner-Whitehouse 3 method, finding no significant differences between the two. Both methods yield relatively accurate results for prepubertal children, with minimal inter-reader variability and high reproducibility. However, bone age assessment becomes more challenging during puberty. In girls aged 11–13 and boys aged 13–15, rapid physical growth occurs, yet distinct skeletal changes in hand and wrist radiographs are not always apparent. At puberty onset, sesamoid bones emerge, while by its conclusion, fusion of the distal phalanx of the thumb and the first metacarpal becomes visible. During this phase, discrepancies between the Greulich & Pyle Atlas method and the Tanner-Whitehouse 3 method may arise due to the limited observable skeletal changes amid rapid growth (Kim, Lee, and Yu 201–205).

A study conducted by Yeon and Kim in 1999, using the Tanner & Whitehouse's 2 method, found a wide range of actual ages for the same bone age measurement. For example, a bone age of 132 months corresponded to an actual age range of 127-145 months (mean 133 months), while a bone age of 144 months had an actual age range of 139-155 months (mean 144 months), demonstrating that bone age measurements alone do not always accurately reflect chronological age (Yeon 9-16).

The Greulich & Pyle Atlas method also faces challenges during puberty, with changes in the radial and ulnar epiphyses, the hamate bone, and the metacarpal epiphyses that are difficult to distinguish. Cundy et al. reported that, during the pubertal age range, there was a discrepancy of more than two years in bone age readings among four radiologists who assessed 60 subjects using the Greulich & Pyle Atlas method (Cundy et al. 513-515). Little et al. also noted that the Greulich & Pyle Atlas method does not improve the accuracy of adult height prediction (Little, Nigo, and Aiona 173-179). To address these limitations, Sauvegrain et al. suggested a method for measuring bone age from the elbow in 1962, providing an alternative for pubertal individuals (Sauvegrain, Nahum, and Bronstein 542-550).

The Tanner-Whitehouse 3 method used here for bone age assessment involves analyzing the maturation of 13 specific bones visible on the left hand and wrist AP radiographs. Each bone's maturity is graded, and the resulting scores are summed to yield a RUS (Radius-Ulna-Short) score. This score is then converted to bone age using a standardized conversion chart. In contrast, the Greulich & Pyle Atlas method assesses bone age by comparing the radiographs to a set of standardized images, with bone age assigned based on the age of the reference image that most closely resembles the patient's bone maturity (Zhang et al. 1001-1015).

The Tanner-Whitehouse 3 method's advantages include reduced ethnic variability, high accuracy and reproducibility due to the detailed scoring of individual bones, and a more precise monthly bone age calculation, facilitating detailed assessments

(Lolli et al. 2683-2690). The Greulich & Pyle Atlas method, on the other hand, allows for simpler analysis with relatively less experience, offering faster calculation times due to its straightforward comparative approach (Tsehay, Afework, and Mesifin 631-640). The primary drawback of the Tanner-Whitehouse 3 method —longer analysis time compared to Greulich & Pyle Atlas method —has been addressed by the automation provided in the www.healthhub.kr AI solution, enabling efficient, computer-assisted assessments.

Efforts to automate the repetitive and time-consuming task of bone age estimation have been ongoing for years. In 1994, Tanner and Gibbons explored the use of a computer system for analyzing Tanner-Whitehouse 2 skeletal maturity scores from hand-wrist radiographs (Tanner and Gibbons 282-287). Their method involved positioning the relevant x-ray area under a video camera's light box, allowing the computer to process the image automatically. They reported that the computer-assisted skeletal age scoring system provided more consistent, repeatable, and highly reproducible results compared to manual assessments. That same year, Drayer and Cox conducted a study on tall Dutch girls, utilizing a computer-aided system based on Fourier analysis to estimate bone age. This system matched a template to the scanned radiograph and identified the most appropriate stage of bone maturity. Their findings indicated that the computer-assisted assessments showed no significant differences from the original Tanner-Whitehouse 2 reference values (Drayer and Cox 77-80).

Researchers continued to explore clinical assessments of skeletal maturity using atlas-based patterns. In 2001, Pietka et al. introduced the concept of epiphyseal/metaphyseal regions of interest (EMROIs), outlining a three-step image analysis process. They reported on the accuracy of detecting the phalangeal tip, extracting the EMROIs, and identifying their diameters and lower edges. Their findings suggested that these extracted features provided a more objective assessment compared to traditional visual comparisons (Pietka et al. 715-729).

After 2020, as deep learning methods gained prominence in computer analysis, research on deep learning-based automated bone age estimation also emerged. Jang Hyung Lee et al. conducted a preliminary study using a deep learning network architecture, refining the training images by eliminating unnecessary reference points and retaining only the relevant regions for age estimation. When tested on 400 image sets, the mean absolute difference error was reported as 8.890 months (Lee, Kim, and Kim 323-331). More recently, Zuhal Y. Hamd et al. applied deep learning-based automated bone age estimation to 473 Saudi children. Their findings highlighted the effectiveness of estimating bone age from left-hand radiographs, emphasizing its potential utility for clinicians while considering the model's margin of error (Hamd et al. 199).

Our study has several limitations. First, the sample size was relatively small, and age groups could not be further subdivided into half-year intervals. Second, children under three years old were not included in the analysis. Third, the study population primarily consisted of children from the Gwangju and Chonnam regions, limiting the generalizability of the findings to other areas in Korea.

5. Conclusion

The cloud-based bone age estimation service, Boneage.io®, provided by www.healthhub.kr, offers a highly user-friendly solution that streamlines the complex and time-consuming Tanner-Whitehouse 3 method through computerization. By delivering bone age estimation results in months, it ensures high reproducibility and reliability. In this study, the effect size of Cohen's D for boys using the Boneage.io® method was 0.021566, while for girls, it was 0.026172. These findings indicate that the standardized differences between the Boneage.io® method and chronological ages are minimal. Therefore, the Boneage.io® method is considered both acceptable and suitable for evaluating skeletal maturation in Korean children.

Cohen's D effect size	Interpretation
0.01 ≤ Cohen's D < 0.2	Very small
0.2 ≤ Cohen's D < 0.5	Small
0.5 ≤ Cohen's D < 0.8	Moderate
0.8 ≤ Cohen's D < 1.2	Large
1.2 ≤ Cohen's D < 2.0	Vary large
Cohen's D ≥ 2.0	Huge

Table 1. Interpretation of the results on D according to Cohen's (Wang et al. 937-943)

			CA (months)		e-BA (months)		CA - e-BA (months)		
Sex	**Age Group(y)**	**N**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**P-value**
Boys	6	10	80.20	2.974	78.50	4.478	1.70	2.541	0.063
	7	30	89.90	3.387	88.63	6.322	1.27	5.219	0.194
	8	61	101.11	3.337	99.61	7.433	1.51	6.358	0.069
	9	79	112.71	3.336	111.09	7.934	1.62	6.903	0.040
	10	88	125.13	3.400	125.51	8.120	-0.39	6.763	0.593
	11	81	137.17	3.232	139.42	6.650	-2.25	5.631	0.001
	12	65	148.89	3.128	151.75	5.121	-2.86	4.000	0.000
	13	30	160.07	3.084	162.87	4.659	-2.80	3.316	0.000
	14	6	170.33	1.211	172.17	2.927	-1.83	2.401	0.120
Girls	6	9	78.00	3.354	74.78	4.944	3.22	2.489	0.005
	7	48	90.83	2.956	89.69	6.858	1.15	6.126	0.201
	8	106	102.01	3.441	99.69	7.620	2.32	7.025	0.001
	9	126	113.48	3.295	111.02	7.793	2.46	7.005	0.000
	10	104	125.65	3.511	124.05	8.472	1.61	7.197	0.025
	11	75	128.11	3.543	139.28	8.427	-1.17	6.703	0.134
	12	62	149.98	3.091	152.90	5.133	-2.92	3.842	0.000
	13	42	161.57	3.255	164.40	3.970	-2.83	2.498	0.000
	14	18	173.33	3.378	176.00	3.757	-2.67	2.142	0.000

TABLE 2. Mean Chronological Age and estimated Bone Age according to the Boneage.Io® Methods

ABBREVIATIONS: TABLE 1 and TABLE 2
2D Two-Dimensional
CA: Chronological Age
CT Computed Tomography
DICOM Digital Imaging and Communications in Medicine
E-BA: Estimated Bone Age
GP Greulich and Pyle
N: Number
PACS Picture Achieving and Communication System
SD: Standard Deviation
TW3 Tanner Whitehouse 3 Methods

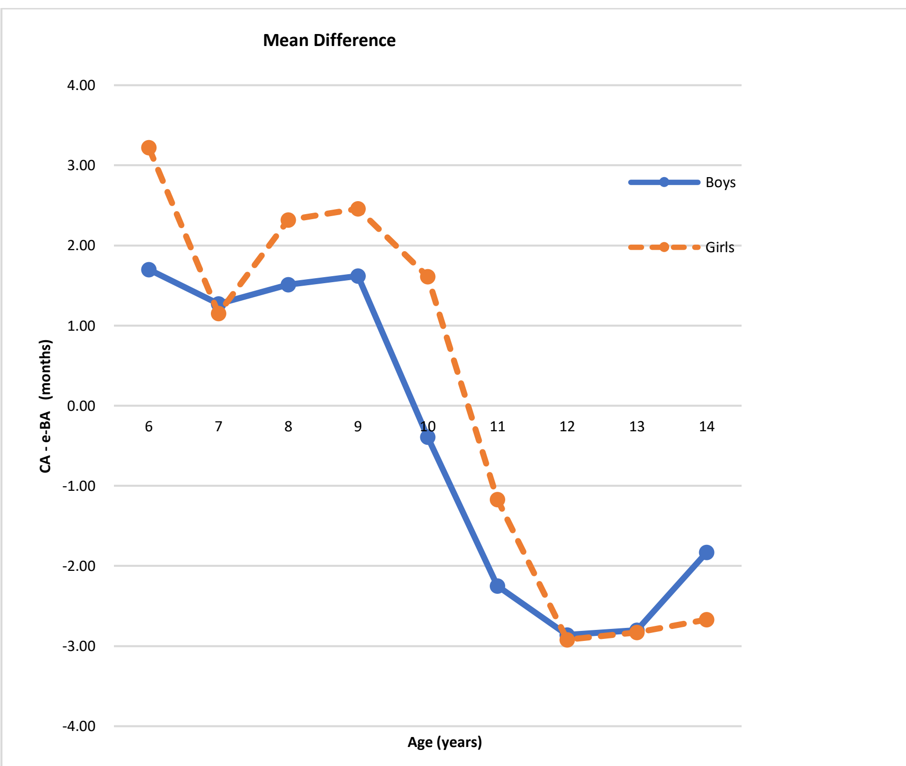

Figure 1. Mean differences between bone ages estimated by using the Boneage.io® methods and chronological ages for boys and girls.

The authors declare that there is no conflict of interest regarding the publication of this paper. This study was supported by research fund from Chosun University, 2019. YJC and JEM contributed to conceptualization, methodology, writing the first draft, editing and reviewing the final draft. YJC contributed to data curation, supervision, and funding acquisition. JEM contributed to the methodology, analyzing the results, writing the methods and results sections, and revising the manuscript for important intellectual content. All authors contributed to the manuscript revision and approved the submitted version.

ETHICAL APPROVAL AND CONSENT TO PARTICIPATE

This study was approved by the Institutional Review Board and Ethics Committee of Chosun University Hospital (IRB No. CHOSUN-2020-03-012). The need for consent to participate was waived by an Institutional Review Board (IRB No. CHOSUN-2020-03-012).

DATA AVAILABILITY

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

Booz C., et al. "Evaluation of a Computer-Aided Diagnosis System for Automated Bone Age Assessment in Comparison to the Greulich-Pyle Atlas Method." *Journal of Computer Assisted Tomography*, 43 (2019), 39-45.

Buckler J.M. "How to make the most of bone ages." Archives of Disease in Childhood, vol. 58, 1983, 761-763.

Cundy P., et al. "Skeletal age estimation in Leg Length Discrepancy." J Pediatr Orthop, 8 (1988): 513-515.

Drayer N.M. and Cox L.A. "Assessment of Bone Ages by the Tanner-Whitehouse Method Using a Computer-aided System." Acta Paediatrica (Oslo: 1992) Supplement, 406 (1994): 77-80.

Hamd Z.Y., et al. "Deep Learning-based Automated Bone Age Estimation for Saudi Patients on Hand Radiograph Images." BMC Med Imaging, 24 (2024): 199.

Hwang J., et al. "Re-Assessment of Applicability of Greulich and Pyle-Based Bone Age to Korean Children Using Manual and Deep Learning-Based Automated Method." Yonsei Med J, 63 (2022): 683-691.

Kim J.R., Lee Y.S. and Yu J. "Assessment of Bone Age in Prepubertal Healthy Korean Children." Korean J Radiol, 16 (2015): 201-205.

Lee J.H., Kim Y.J. and Kim K.G. "Bone age Estimation Using Deep Learning and Hand X-ray Images." Biomedical Engineering Letters, 10 (2020): 323-331.

Little D.G., Nigo L. and Aiona M.D. "Deficiencies of Current Methods for the Timing of Epiphysiodesis." J Pediatr Orthop, 16 (1996): 173-179.

Lolli L., et al. "Tanner-Whitehouse and Modified Bayley-Pinneau Adult Height Predictions in Elite Youth Soccer Players from the Middle East." Medicine and Science in Sports and Exercise, 53 (2021): 2683-2690.

Pietka E., et al. "Computer-assisted bone age assessment." IEEE Transactions on Medical Imaging, 20 (2001): 715-729.

Prokop-Piotrkowska M., et al. "Traditional and New Methods of Bone Age Assessment." Journal of Clinical Research in Pediatric Endocrinology, 13 (2021): 251-262.

Satoh M. "Bone Age: Assessment Methods and Clinical Applications." Clinical Pediatric Endocrinology, 4 (2015): 143-152.

Sauvegrain J., Nahum H. and Bronstein H. "Study of Bone Maturation of the Elbow." Annales de Radiologie, 5 (1962): 542-550.

Shah N., et al. "Comparison of Bone Age Assessments by Gruelich-Pyle, Gilsanz-Ratib, and Tanner Whitehouse Methods in Healthy Indian Children." Indian Journal of Endocrinology and Metabolism, 25 (2021): 240-246.

Shin N.Y., et al. "Evaluation of the Clinical Efficacy of a TW3-

Based Fully Automated Bone Age Assessment System Using Deep Neural Networks." Imaging Science in Dentistry, 50 (2020): 237-243.

Van Steenkiste T., et al. "Automated Assessment of Bone Age Using Deep Learning and Gaussian Process Regression." Annu Int Conf IEEE Eng Med Biol Soc, 2018: 674-677.

Tanner J.M. and Gibbons R.D. "A Computerized Image Analysis System for Estimating Tanner-Whitehouse 2 Bone Age." Hormone Research, 42 (1994): 282-287.

Thodberg H.H., et al et al. "Autonomous Artificial Intelligence in Pediatric Radiology." Pediatr Radiol, 52 (2022): 1338-1346.

Tsehay B., Afework M. and Mesifin M. "Assessment of Reliability of Greulich and Pyle (GP) Method for Determination of Age of Children at Debre Markos Referral Hospital, East Gojjam Zone." Ethiopian Journal of Health Sciences, 27 (2017): 631-640.

Wang Y.M., et al. "Automatic Assessment of Bone Age in Taiwanese Children." The Kaohsiung Journal of Medical Sciences, 36 (2020): 937-943.

Yeon K.M. "Standard bone-age of infants and children in Korea." Journal of Korean Medical Science, 12 (1997): 9-16.

Zhang D., et al. "An Automated TW3-RUS Bone Age Assessment Method with Ordinal Regression-Based Determination of Skeletal Maturity." J Digit Imaging, 36 (2023): 1001-1015.