Original Article | Musculoskeletal Imaging

eISSN 2005-8330
https://doi.org/10.3348/kjr.2020.1468
Korean J Radiol 2021;22(12):2017-2025

Korean Journal of Radiology
KJR

Clinical Validation of a Deep Learning-Based Hybrid (Greulich-Pyle and Modified Tanner-Whitehouse) Method for Bone Age Assessment

Kyu-Chong Lee[1]*, Kee-Hyoung Lee[2]*, Chang Ho Kang[1], Kyung-Sik Ahn[1], Lindsey Yoojin Chung[3], Jae-Joon Lee[4], Suk Joo Hong[5], Baek Hyun Kim[6], Euddeum Shim[6]

Departments of [1]Radiology and [2]Pediatrics, Korea University Anam Hospital, Seoul, Korea; [3]Department of Pediatrics, Myongji Hospital, Goyang, Korea; [4]Crescom, Seongnam, Korea; [5]Department of Radiology, Korea University Guro Hospital, Seoul, Korea; [6]Department of Radiology, Korea University Ansan Hospital, Ansan, Korea

Objective: To evaluate the accuracy and clinical efficacy of a hybrid Greulich-Pyle (GP) and modified Tanner-Whitehouse (TW) artificial intelligence (AI) model for bone age assessment.

Materials and Methods: A deep learning-based model was trained on an open dataset of multiple ethnicities. A total of 102 hand radiographs (51 male and 51 female; mean age ± standard deviation = 10.95 ± 2.37 years) from a single institution were selected for external validation. Three human experts performed bone age assessments based on the GP atlas to develop a reference standard. Two study radiologists performed bone age assessments with and without AI model assistance in two separate sessions, for which the reading time was recorded. The performance of the AI software was assessed by comparing the mean absolute difference between the AI-calculated bone age and the reference standard. The reading time was compared between reading with and without AI using a paired *t* test. Furthermore, the reliability between the two study radiologists' bone age assessments was assessed using intraclass correlation coefficients (ICCs), and the results were compared between reading with and without AI.

Results: The bone ages assessed by the experts and the AI model were not significantly different (11.39 ± 2.74 years and 11.35 ± 2.76 years, respectively, $p = 0.31$). The mean absolute difference was 0.39 years (95% confidence interval, 0.33–0.45 years) between the automated AI assessment and the reference standard. The mean reading time of the two study radiologists was reduced from 54.29 to 35.37 seconds with AI model assistance ($p < 0.001$). The ICC of the two study radiologists slightly increased with AI model assistance (from 0.945 to 0.990).

Conclusion: The proposed AI model was accurate for assessing bone age. Furthermore, this model appeared to enhance the clinical efficacy by reducing the reading time and improving the inter-observer reliability.

Keywords: *Artificial intelligence; Convolutional neural network; Bone age assessment; Greulich-Pyle method; Tanner-Whitehouse method*

INTRODUCTION

Bone age assessment is crucial for evaluating pediatric growth and maturity [1]. Bone age is an important parameter for the assessment of the progress and treatment of various pediatric endocrine diseases. Furthermore, it can be used to predict adult height [2,3]. Recently, the high prevalence of precocious puberty [4], increased interest in the height of children, and increased usage of growth hormone therapy emphasize the need for

Received: December 17, 2020 **Revised:** June 26, 2021 **Accepted:** June 28, 2021

*These authors contributed equally to this work.

Corresponding author: Chang Ho Kang, MD, PhD, Department of Radiology, Korea University Anam Hospital, 73 Goryeodae-ro, Seongbuk-gu, Seoul 02841, Korea.
• E-mail: mallecot@gmail.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://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 © 2021 The Korean Society of Radiology

assessing bone age [5].

In clinical practice, left hand and wrist radiography-based bone age assessments, such as the Greulich-Pyle (GP) [6] and the Tanner-Whitehouse 3 (TW3) methods [7], are widely used. The GP method is an atlas-based method, which determines bone age by comparing the radiographs of the hand and wrist with the most similar standard radiographs in the GP atlas. It is a simple method that is readily available in clinical practice. However, several studies have suggested possible issues, including inter- and intra-observer variability and the dependency of the accuracy on the experience of the clinician [8-10]. Furthermore, it is a semi-quantitative method because the GP atlas mainly covers one year. Recently, deep learning-based automatic bone age assessment models have been developed to overcome these issues [11-13]. According to Stanford University researchers, the mean absolute difference (MAD) between the automatic bone age assessment tools based on the GP atlas and standard bone age, which is determined by three radiologists, is 0.5 years [14].

The TW method is a scoring system that measures the individual bone maturity score and evaluates bone age by summing the scores. After two revisions, the current version (TW3) was proposed, which used the maturity of the radius, ulna, and short bones [7]. It is a quantified method and, therefore, is more accurate and has a higher reproducibility than the previous GP method [8-10]. However, the TW method also has certain limitations. First, it is a more complex method that takes a longer time than the GP method [8,15]. Second, it classifies individual bones based on nine maturity grades (A to I); thus, the classification is sometimes ambiguous and has inherent deviation because one particular bone shape can have two different pre-defined labels of the same feature [16,17].

To overcome the limitations of both methods, we developed a hybrid GP and TW artificial intelligence (AI) bone age assessment software. This software was trained using hand and wrist radiography based on both the GP and modified TW methods, which use seven regions (radius, ulna, distal phalange, middle phalange, and proximal phalange, metacarpal of the third digit, and metacarpal of the first digit) instead of 13 regions for the TW3 method. The holistic hand image analysis based on the GP method in our software can cover all the regions that are not included in the TW3 method, and the holistic hand image analysis is further reinforced by minutely assessing the individual region of interest (ROI) based on the modified TW method, which improves the classification performance by zooming in more relevant regions, as discussed in a previous study [13]. Eventually, the final bone age in this software was obtained from the integrated analysis of both holistic image analysis and ROI analysis in a fully automatic manner. The purpose of this study was to evaluate the accuracy and clinical efficacy of a deep learning-based hybrid GP and modified TW AI model for assessing bone age.

MATERIALS AND METHODS

This study was approved by the Institutional Review Board and Ethics Committee of the Korea University Anam Hospital (IRB No. 2019AN0010). Informed consent was waived because the data were collected retrospectively and analyzed anonymously. The study complied with the ethical principles of the Helsinki Declaration of 1964, which was revised by the World Medical Organization in Edinburgh in 2000. This study involved three major steps: model development, external validation, and statistical analysis. Details about the steps and the study population are provided in Figure 1.

Model Development

The model consists of three steps: ROI detection, region maturity classification, and integrated bone age. The steps are summarized in Figure 2. First, seven regions based on TW3 were automatically detected using the convolutional neural network (CNN) algorithm. While TW3 used 13 ROIs for bone age assessment, the proposed modified method used seven regions (radius, ulna, distal phalange, middle phalange, proximal phalange, metacarpal of the third digit, and metacarpal of the first digit) to improve the labelling efficiency during the training steps. We postulated that the short bones of the first and fifth fingers, except the metacarpal of the first finger, correlate highly with those of the third finger, as reported in a previous study [18]. Second, each ROI and the holistic hand image were automatically classified for maturity using the CNN algorithm. While TW3 uses nine stages, from A to I, to assess the maturity of each ROI, our method used 34 stages with a 6-month gap from 1.5 to 18 years for maturity, which could be more accurate and intuitive. In addition, the holistic hand image is automatically classified and applied to our AI model because the holistic image can provide the maturity features of regions that are not included in the ROIs of TW3. In this ROI and the holistic hand image maturity classification procedure, we used the doctors' ratings as the reference standard for each CNN model training. Finally, the features from each ROI and the entire hand were integrated and classified to provide the final bone age estimate of the input image. Basically, the prediction probability distributions of the maturity

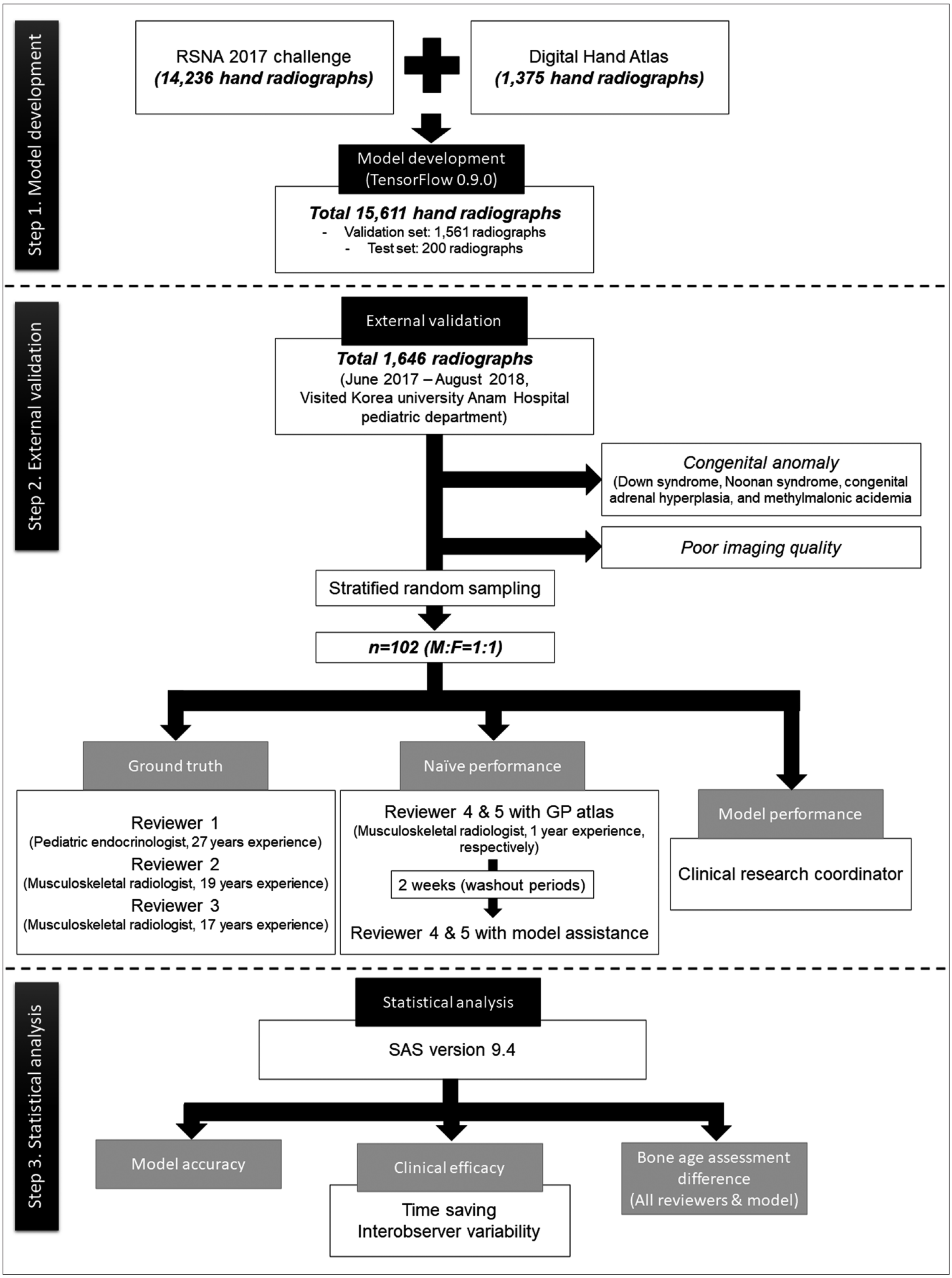

Step 1. Model development
RSNA 2017 challenge
(14,236 hand radiographs)
Digital Hand Atlas
(1,375 hand radiographs)
Model development
(TensorFlow 0.9.0)
Total 15,611 hand radiographs
- Validation set: 1,561 radiographs
- Test set: 200 radiographs
Step 2. External validation
External validation
Total 1,646 radiographs
(June 2017 – August 2018,
Visited Korea university Anam Hospital
pediatric department)
Congenital anomaly
(Down syndrome, Noonan syndrome, congenital adrenal hyperplasia, and methylmalonic acidemia
Poor imaging quality
Stratified random sampling
n=102 (M:F=1:1)
Ground truth
Reviewer 1
(Pediatric endocrinologist, 27 years experience)
Reviewer 2
(Musculoskeletal radiologist, 19 years experience)
Reviewer 3
(Musculoskeletal radiologist, 17 years experience)
Naïve performance
Reviewer 4 & 5 with GP atlas
(Musculoskeletal radiologist, 1 year experience, respectively)
2 weeks (washout periods)
Reviewer 4 & 5 with model assistance
Model performance
Clinical research coordinator
Step 3. Statistical analysis
Statistical analysis
SAS version 9.4
Model accuracy
Clinical efficacy
Time saving
Interobserver variability
Bone age assessment difference
(All reviewers & model)

Fig. 1. Overview of entire steps of the study including participant selection. The model was developed using two public datasets. A total of 15611 hand radiographs were used as training, validation, and test sets. A total of 102 hand radiographs were used for external validation. Finally, statistical analysis was performed. GP = Greulich-Pyle, RSNA = Radiological Society of North America

https://doi.org/10.3348/kjr.2020.1468

stages of the regions were concatenated and inputted to the final integration step, which used a fully connected neural network model. This entire procedure for bone age assessment, composed of three steps, is fully automatic. The model was implemented using an open-source machine learning library (TensorFlow version 0.9.0; Google). Figure 3 shows a sample image of the mediAI-BA, the automatic solution interface of this software.

We used two public datasets. The first is from the Radiological Society of North America (RSNA) 2017 challenge, which includes 14236 hand radiographs from Stanford University and Colorado University [19], and the

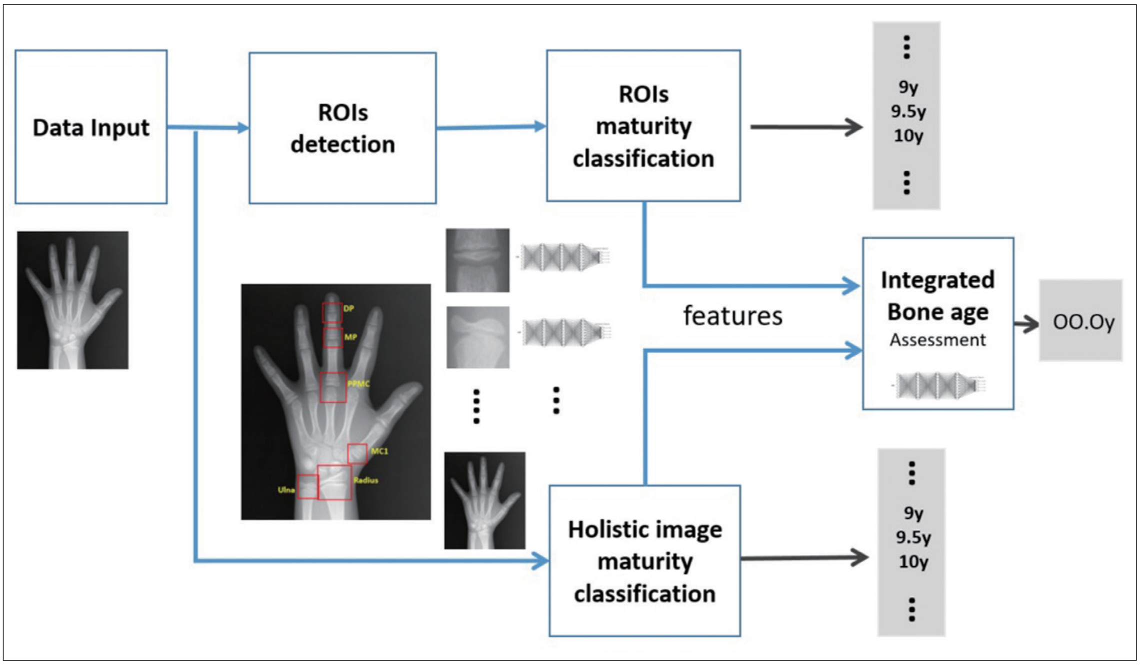
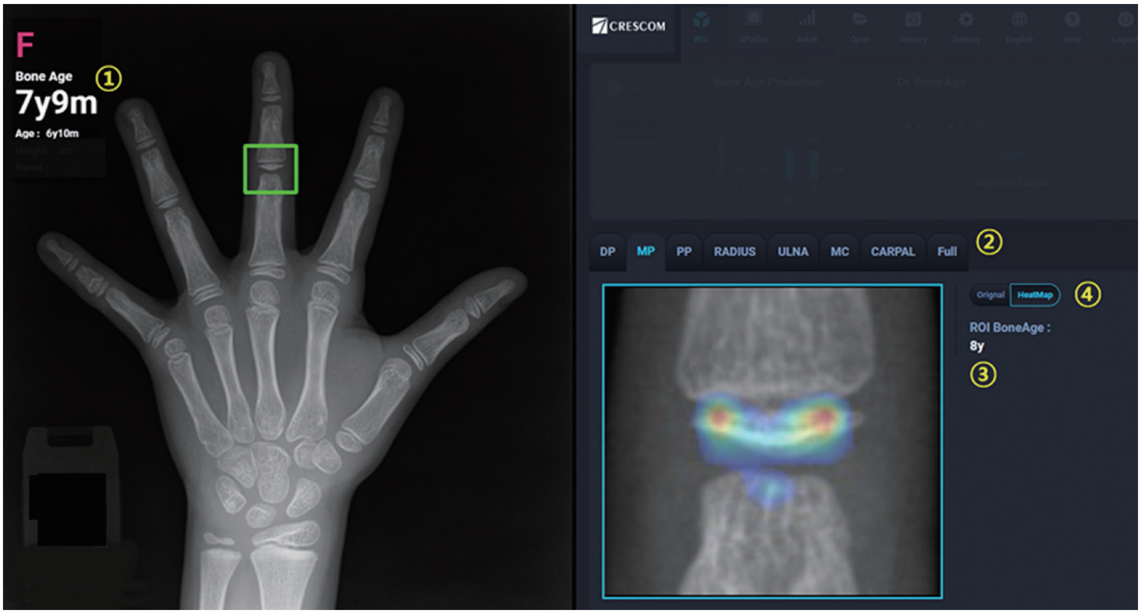

Data Input
ROIs detection
ROIs maturity classification
9y
9.5y
10y
DP
MP
PPMC
MC1
Ulna
Radius
features
Integrated Bone age Assessment
OO.Oy
Holistic image maturity classification
9y
9.5y
10y

Fig. 2. Overview of Greulich-Pyle and modified Tanner-Whitehouse hybrid bone age assessment models. ROI = region of interest


F

Bone Age ①

7y9m

Age : 6y10m

ROI BoneAge : 8y ③

Fig. 3. Result of automatic bone age assessment program (mediAI-BA) including analysis of detailed area of interest. Automatic hybrid method-derived bone age is observed in the left upper corner ①. The user can choose from ② the seven regions of interest (DP, MP, PP, and MC of the third finger; radius, ulna, and MC of the first finger), and third digit middle phalanx image with its respective maturity degree is shown ③. Heatmap overlay is selected ④ and is shown. DP = distal phalange, MC = metacarpal, MP = middle phalange, PP = proximal phalange

other is from the Digital Hand Atlas [20], which includes 1375 hand radiographs from the University of Southern California. We used 10% of the dataset as the validation set and 200 images as the test set.

External Validation

Among 1646 the participants aged 2–18 years who visited the pediatric department and underwent left hand radiography at Korea University Anam Hospital between June 2017 and August 2018, those with a congenital anomaly (Down syndrome, Noonan syndrome, congenital adrenal hyperplasia, and methylmalonic acidemia) and poor imaging quality were excluded. According to stratified random sampling, we selected 102 participants (51 male and 51 female; mean age ± standard deviation [SD] = 10.95 ± 2.37 years) for external validation.

Three reviewers independently estimated the bone age based on the GP atlas with the first digit after the decimal point (year). They intended to judge the bone age according to the GP standard bone age if possible; however, they were allowed to use the median age according to their level of experience. Reviewer 1 is a pediatric endocrinologist with 29 years of clinical experience who is familiar with bone age estimation based on radiographs. Reviewers 2 and 3 are musculoskeletal radiologists with 19 and 12 years of clinical experience, respectively. The average of the independent bone age estimates of the three reviewers was then used as the reference standard for this study. If there was a discrepancy of more than 2 years, the image was re-evaluated until a consensus was reached.

Two different one-year fellowship-trained musculoskeletal radiologists (reviewers 4 and 5) conducted bone age assessments in two different sessions. During the first session, they independently estimated the bone age based on the GP atlas. The time was measured in seconds using a stopwatch. Two weeks after the washout period, they repeated the bone age assessment with model assistance and the GP atlas shown in the model. Time was measured using the same method. Finally, the clinical research coordinator conducted the bone age assessment using the model.

Statistical Analyses

We checked the intraclass correlation coefficient (ICC) for the three reviewers (reviewers 1–3) who participated in generating reference standards for the validation of the reference standard. The results of the bone age assessment by the model and the reviewers are summarized as the mean, SD, median, minimum, and maximum values. To evaluate model performance, the model estimates were compared with the reference standard. The results of the bone age assessment were compared using a paired *t* test. The difference in bone ages between the model assessment and the reference standard was also evaluated by the MAD with its 95% confidence interval (CI). The upper limit of the 95% CI (< 0.5) indicated no statistically significant systematic bias in bone age assessment. This value (0.5 years) was adopted as the same equivalence limit in previously published articles [8,14]. Furthermore, we calculated the root mean squared error (RMSE).

We determined the amount of time required to evaluate the clinical efficacy of the model. A paired *t* test was used to compare the time required for reading with and without model assistance for reviewers 4 and 5. Furthermore, we calculated the MAD and ICC for reviewers 4 and 5 for the reading with and without model assistance.

All analyses were conducted using SAS version 9.4 (SAS Institute Inc.). Statistical significance was set at $p < 0.05$.

RESULTS

Characteristics of the Participants

Table 1 shows the demographic data of the participants. The mean chronological age ± SD of the participants was 10.95 ± 2.37 years (male: 11.18 ± 2.88, female: 10.72 ± 1.71). The age range was 4.92 to 17.00 years. According to the age distribution, the largest group included 63 participants aged 10 or more years but under 15 years, while the second largest group included 33 participants aged 5 or more years but under 10 years.

The most common causes for examination were precocious puberty (n = 50), followed by short stature (n = 40). The remaining 12 participants had an endocrine disease, including diabetes mellitus, obesity, thyroiditis, vitamin D deficiency, and growth hormone deficiency.

Validation of Reference Standard

The ICC (95% CI) of the three reviewers (reviewers 1–3) was 0.993 (0.990–0.995). This value was sufficiently high to use the average bone age assessment value as a reference standard.

Model Accuracy in Bone Age Assessment

Table 2 shows the results of the bone age assessment using the model and the reference standard determined by three human experts. The mean bone ages ± SDs assessed with the model and the reference standard were 11.35 ± 2.76 and 11.39 ± 2.74 years, respectively, without a statistically significant difference (p = 0.31). The MAD between the model and the reference standard was 0.39 years (95% CI, 0.33–0.45), which is less than 0.5 years. The RMSE was 0.498 years.

Clinical Efficacy: Effect on Reading Time and Inter-Observer Reliability

Table 3 summarizes the results related to the clinical efficacy. The mean interpretation times (seconds) of reviewer 4 with and without model assistance were 31.72 seconds and 56.81 seconds, respectively. The mean difference was 25.10 seconds (95% CI, 21.41–28.79), which was significantly different (p < 0.001). The mean interpretation times (seconds) of reviewer 5 with and without model assistance were 38.82 seconds and 51.76 seconds, respectively. The mean difference was 12.1 seconds (95% CI, 7.07–17.1), which was statistically different (p = 0.001). Combining the two readers, he bone age assessment with model assistance took 35.27 seconds, while the initial bone age assessment time was 54.29 seconds, which was 1.54 times more. The mean difference was 18.6 (95% CI, 12.9–24.3), which indicated a significant reduction (p < 0.001).

There was no significant difference between the diagnostic accuracies of the bone ages assessed by reviewer

Table 1. Demographic Data of Subjects

	Total (n = 102)	Male (n = 51)	Female (n = 51)
Age, year			
Mean ± SD	10.95 ± 2.37	11.18 ± 2.88	10.72 ± 1.71
Median	10.88	11.17	10.67
Range, min–max	4.92–17.00	4.92–17.00	7.67–14.58
Age distribution, years			
< 5	1 (0.98)	1 (1.96)	0 (0)
≥ 5 and < 10	33 (32.35)	14 (27.45)	19 (37.25)
≥ 10 and < 15	63 (61.76)	31 (60.78)	32 (62.75)
≥ 15	5 (4.90)	5 (9.80)	0 (0)

Data are number of patients with % in parentheses, unless specified otherwise. max = maximin, min = minimum, SD = standard deviation

Table 2. Results of Bone Age Assessment

	Total (n = 102)	Male (n = 51)	Female (n = 51)
Automatic bone age assessment by model			
Mean ± SD	11.35 ± 2.76	11.58 ± 3.47	11.11 ± 1.80
Median	11.30	12.10	11.10
Range, min–max	3.60–16.90	3.60–16.90	6.90–14.80
Reference standard bone age reference by three reviewers			
Mean ± SD	11.39 ± 2.74	11.42 ± 3.52	11.37 ± 1.61
Median	11.50	11.83	11.33
Range, min–max	3.17–17.00	3.17–17.00	7.60–14.93

Data are years. max = maximin, min = minimum, SD = standard deviation

Table 3. Results of Clinical Efficacy Evaluation

	MAD (95% CI), Year*	Mean Interpretation Time, Sec	ICC (95% CI)
First session: without model			
Reviewer 4	0.42 (0.35–0.50)	56.81	0.945 (0.919–0.963)
Reviewer 5	0.88 (0.75–1.01)	51.76	
Second session: with model			
Reviewer 4	0.42 (0.35–0.50)	31.72	0.990 (0.985–0.993)
Reviewer 5	0.32 (0.27–0.37)	38.82	

*MAD between each reviewer's estimated bone age and reference standard. CI = confidence interval, ICC = intraclass correlation coefficient, MAD = mean absolute deviation

4 with and without the model. The MAD (95% CI) for the assessment of reviewer 4 based on the model and the reference standard was 0.42 (0.35–0.50), while that (95% CI) for the assessment of reviewer 4 not based on the model and the reference standard was 0.42 (0.35–0.499). However, the MAD (95% CI) for the assessment of reviewer 5 not based on the model and the reference standard was 0.88 (0.75–1.01) while the MAD (95% CI) for the assessment of reviewer 5 based on the model and the reference standard was 0.32 (0.27–0.37). Therefore, the diagnostic accuracy of reviewer 5 was significantly improved by model assistance (p < 0.001).

Furthermore, the ICC (95% CI) for reviewers 4 and 5 without the model was 0.945 (0.919–0.963); that (95% CI) for reviewers 4 and 5 with the assistance of the model was 0.990 (0.985–0.993).

DISCUSSION

Our study verified the accuracy and clinical efficacy of the newly developed GP and modified TW hybrid AI bone age assessment model. Our model had an accuracy similar to that of human experts, with the upper limit of the 95% CI of the MAD between the AI bone age assessment measurement and the reference standard being less than 0.5 years. Our model shortened the reading time by approximately 35% for two additional radiologists. Kim et al. [21] showed that reading times were reduced by 18.0% and 40.0% for each of the two reviewers. Other studies have also reported that AI-assisted bone age assessment can reduce the interpretation time [22]. Furthermore, the accuracy of bone age assessment was significantly improved in the case of one additional radiologist (reviewer 5). Additionally, the ICC of two additional radiologists was somewhat improved during model assistance. This means that this model could help improve the inter-observer reliability, as in previous studies [12,14,23].

Recently, AI with deep learning has been applied to musculoskeletal radiology, including image interpretation, such as fracture detection and bone age assessment [24]. Several studies have suggested that CNN bone age assessment is as accurate as that of experts and has clinical efficacy [11,12,14,23]. Tajmir et al. [23] showed that AI assistance improves the performance of radiologists. Our study also showed that the AI tool improved the performance of the radiologist in terms of reducing the interpretation time and improving the inter-observer reliability. Furthermore, our model improved the diagnostic accuracy of bone age assessment for less experienced radiologists.

Contrary to previous studies [22], our model has the following advantages. First, our hybrid model complemented the limitations of GP and TW by focusing on the regions that are highly related to bone maturity changes and by applying finer-grained maturity stages than TW3, which resulted in a reliable and accurate bone age estimate. Second, our model reflects the comprehensive human decision-making process in a clinical setting whereby experts exploit the ROIs of bone rather than strictly use the GP atlas. Finally, the black box nature of CNNs [25,26], we thought, could be partially resolved by integrating two different methods (GP and TW) within the medical domain during the model development process.

Our model showed the integrated bone age, as well as two different results, based on the detailed ROIs and a holistic image. The current commercial automated bone age assessment system, BoneXpert (Visiana Aps, Holte, Denmark, http://www.boneexpert.com) with its recently launched version 3.0 (September 2019), is accurate [27], and it evaluates the bone age according to both the GP and alternative TW2 methods [28], which is similar to our model. BoneXpert is based on a feature extraction technique that reconstructs the border of 15 bones (including metacarpal, phalangeal bones, distal radius, and ulna) [29], which is different from our model. Several AI bone age assessment models have been developed using all the bones included in the radiograph and sometimes display the sensitive region of the image like a heat map [14,21]. The mean sensitive regions of the image were determined by the model, not by clinicians, and were different during the serial follow-up.

The MAD and RMSE of our model were 0.39 and 0.50, respectively. This value shows that the accuracy of our hybrid model is similar to that of previous studies [14,15,18,29]. The MAD and RMSE of the GP method-based AI model MAD and RMSE were 0.50 and 0.63 [14]. Another GP method-based model called "HH-boneage" has a MAD of 0.46 and an RMSE of 0.62 [15]. The 9-stage TW method showed the MAD and RMSE for 7 ROIs of 0.59 and 0.76 [18]. BoneXpert version 3.0 showed an RMSE of 0.63 [29]. It was observed that the test data sets, including the number of images, differ among these studies, and the values for comparing the automated assessed bone age with the standard reference were obtained by the different expert groups. Therefore, the results cannot be directly compared with the RMSE and MAD values. However, these values corroborate the accuracy of the proposed model.

However, our study had some limitations. First, this was a single-center retrospective study with only a few participants; in particular, only six participants aged < 5 years and > 15 years were enrolled. In the future, prospective multicenter large-sample studies are needed. Second, our study included a single-ethnicity external validation. Previous studies have shown racial differences in certain age bone growth patterns, which can affect bone age assessment [30,31]. However, our model was trained on open data, including those on ethnicity; therefore, we believe that it could be used globally. Third, we only compared the accuracy of our model and the GP method. A recent article suggested that the question of whether AI bone age assessment should be compared with other bone age assessment methods, including the TW method or using other imaging modalities such as MRI or ultrasonography instead of the left hand and wrist radiography persists [15]. Therefore, further comparative studies are required to confirm this. Finally, our model could not detect disorders such as congenital syndrome or rickets, a similar limitation to that of other recent AI models. However, the purpose of AI bone age assessment models is to assist the radiologist and not to use it independently in a clinical setting.

In conclusion, this new hybrid GP and modified TW AI bone age assessment model was accurate for bone age assessment. Furthermore, it appeared to improve the clinical efficacy by reducing the interpretation time and improving the inter-observer reliability.

Conflicts of Interest

Chan Ho Kang; Research funded, Crescom Inc Stockholder, Crescom Inc. Jae-Joon Lee is a founder of Crecsom Inc., a startup company, the eventual products and services of which will be related to the subject matter of the article. Other authors declare no conflict on interest or industry support for the study.

Acknowledgments

Thanks to Ji-Eun Lee, Statistician, whose statistical expertise was invaluable during the statistical analysis and data interpretation.

Author Contributions

Conceptualization: Kee-Hyoung Lee, Chang Ho Kang, Kyu-Chong Lee. Data curation: Chang Ho Kang, Kyung-Sik Ahn, Jae-Joon Lee. Formal analysis: Kyu-Chong Lee, Chang Ho Kang. Funding acquisition: Chang Ho Kang, Jae-Joon Lee. Investigation: Chang Ho Kang, Kyu-Chong Lee, Kyung-Sik Ahn, Lindsey Yoojin Chung. Methodology: Kee-Hyoung Lee, Chang Ho Kang, Jae-Joon Lee. Resources: Chang Ho Kang, Jae-Joon Lee. Software: Jae-Joon Lee. Supervision: Chang Ho Kang. Validation: Kyu-Chong Lee, Lindsey Yoojin Chung. Visualization: Kyu-Chong Lee. Writing—original draft: Kyu-Chong Lee. Writing—review & editing: Suk Joo Hong, Baek Hyun Kim, Euddeum Shim, Chang Ho Kang, Lindsey Yoojin Chung, Kyung-Sik Ahn.

ORCID iDs

Kyu-Chong Lee
https://orcid.org/0000-0002-4518-8567
Kee-Hyoung Lee
https://orcid.org/0000-0002-4319-9019
Chang Ho Kang
https://orcid.org/0000-0003-2385-7245
Kyung-Sik Ahn
https://orcid.org/0000-0001-9354-5699
Lindsey Yoojin Chung
https://orcid.org/0000-0002-7447-6250
Jae-Joon Lee
https://orcid.org/0000-0002-6948-8230
Suk Joo Hong
https://orcid.org/0000-0002-3923-1426
Baek Hyun Kim
https://orcid.org/0000-0002-3284-1803
Euddeum Shim
https://orcid.org/0000-0002-0983-0209

Funding Statement

None

REFERENCES

1. Zerin JM, Hernandez RJ. Approach to skeletal maturation. *Hand Clin* 1991;7:53-62
2. Satoh M. Bone age: assessment methods and clinical applications. *Clin Pediatr Endocrinol* 2015;24:143-152
3. Manzoor Mughal A, Hassan N, Ahmed A. Bone age assessment methods: a critical review. *Pak J Med Sci* 2014;30:211-215
4. 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;208:221-228

5. 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;16:201-205
6. Greulich WW, Pyle SI. *Radiographic atlas of skeletal development of the hand and wrist*. California: Stanford University Press, 1971
7. Tanner JM, Healy H, Goldstein H, Cameron N. *Assessment of skeletal maturity and prediction of adult height (TW3 method)*, 3rd ed. London: W.B Saunders Company, 2001
8. Kim SY, Oh YJ, Shin JY, Rhie YJ, Lee KH. Comparison of the Greulich-Pyle and Tanner Whitehouse (TW3) methods in bone age assessment. *J Korean Soc Pediatr Endocrinol* 2008;13:50-55
9. 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;176:507-510
10. 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;81:172-173
11. Spampinato C, Palazzo S, Giordano D, Aldinucci M, Leonardi R. Deep learning for automated skeletal bone age assessment in X-ray images. *Med Image Anal* 2017;36:41-51
12. Lee H, Tajmir S, Lee J, Zissen M, Yeshiwas BA, Alkasab TK, et al. Fully automated deep learning system for bone age assessment. *J Digit Imaging* 2017;30:427-441
13. Fu J, Zheng H, Mei T. Look closer to see better: recurrent attention convolutional neural network for fine-grained image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 Jul 21-26; Honolulu, HI, USA: IEEE; 2017; p. 4438-4446
14. Larson DB, Chen MC, Lungren MP, Halabi SS, Stence NV, Langlotz CP. Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs. *Radiology* 2018;287:313-322
15. Lee BD, Lee MS. Automated bone age assessment using artificial intelligence: the future of bone age assessment. *Korean J Radiol* 2021;22:792-800
16. Oh YJ. *Development and evaluation of semi-automatic bone age estimation method*. Seoul: The Graduate School of Ewha Womans University, 2011
17. Aja-Fernández S, De Luis-García R, Martín-Fernández MA, Alberola-López C. A computational TW3 classifier for skeletal maturity assessment. A computing with words approach. *J Biomed Inform* 2004;37:99-107
18. Bui TD, Lee JJ, Shin J. Incorporated region detection and classification using deep convolutional networks for bone age assessment. *Artif Intell Med* 2019;97:1-8
19. Halabi SS, Prevedello LM, Kalpathy-Cramer J, Mamonov AB, Bilbily A, Cicero M, et al. The RSNA pediatric bone age machine learning challenge. *Radiology* 2019;290:498-503
20. University of Southern California. Digital hand atlas – Image processing and informatics lab. Ipilab.usc.edu Web site. https://ipilab.usc.edu/research/baaweb/. Published July, 2017. Accessed March 6, 2020
21. 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;209:1374-1380
22. Dallora AL, Anderberg P, Kvist O, Mendes E, Diaz Ruiz S, Sanmartin Berglund J. Bone age assessment with various machine learning techniques: a systematic literature review and meta-analysis. *PLoS One* 2019;14:e0220242
23. Tajmir SH, Lee H, Shailam R, Gale HI, Nguyen JC, Westra SJ, et al. Artificial intelligence-assisted interpretation of bone age radiographs improves accuracy and decreases variability. *Skeletal Radiol* 2019;48:275-283
24. Gyftopoulos S, Lin D, Knoll F, Doshi AM, Rodrigues TC, Recht MP. Artificial intelligence in musculoskeletal imaging: current status and future directions. *AJR Am J Roentgenol* 2019;213:506-513
25. Park SH. Artificial intelligence in medicine: beginner's guide. *J Korean Soc Radiol* 2018;78:301-308
26. Castelvecchi D. Can we open the black box of AI? Nature.com Web site. https://www.nature.com/news/can-we-open-the-black-box-of-ai-1.20731. Published October 5, 2016. Accessed November 20, 2019
27. Thodberg HH, Martin DD. Validation of a new version of BoneXpert bone age in children with congenital adrenal hyperplasia (CAH), precocious puberty (PP), growth hormone deficiency (GHD), Turner syndrome (TS), and other short stature diagnoses. Proceedings of the 58th Annual ESPE Meeting; 2019 Sep 19-21; Vienna, Austria: European Society for Paediatric Endocrinology; 2019; p. FC2.6
28. Thodberg HH, Kreiborg S, Juul A, Pedersen KD. The BoneXpert method for automated determination of skeletal maturity. *IEEE Trans Med Imaging* 2008;28:52-66
29. van Rijn RR, Thodberg HH. Bone age assessment: automated techniques coming of age? *Acta Radiol* 2013;54:1024-1029
30. 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;250:228-235
31. Ontell FK, Ivanovic M, Ablin DS, Barlow TW. Bone age in children of diverse ethnicity. *AJR Am J Roentgenol* 1996;167:1395-1398

https://doi.org/10.3348/kjr.2020.1468