Computerized Bone Age Estimation Using Deep Learning–Based Program: Evaluation of the Accuracy and Efficiency

Jeong Rye Kim[1]
Woo Hyun Shim[1]
Hee Mang Yoon[1]
Sang Hyup Hong[1]
Jin Seong Lee[1]
Young Ah Cho[1]
Sangki Kim[2]

OBJECTIVE. The purpose of this study is to evaluate the accuracy and efficiency of a new automatic software system for bone age assessment and to validate its feasibility in clinical practice.

MATERIALS AND METHODS. A Greulich-Pyle method–based deep-learning technique was used to develop the automatic software system for bone age determination. Using this software, bone age was estimated from left-hand radiographs of 200 patients (3–17 years old) using first-rank bone age (software only), computer-assisted bone age (two radiologists with software assistance), and Greulich-Pyle atlas–assisted bone age (two radiologists with Greulich-Pyle atlas assistance only). The reference bone age was determined by the consensus of two experienced radiologists.

RESULTS. First-rank bone ages determined by the automatic software system showed a 69.5% concordance rate and significant correlations with the reference bone age ($r = 0.992$; $p < 0.001$). Concordance rates increased with the use of the automatic software system for both reviewer 1 (63.0% for Greulich-Pyle atlas–assisted bone age vs 72.5% for computer-assisted bone age) and reviewer 2 (49.5% for Greulich-Pyle atlas–assisted bone age vs 57.5% for computer-assisted bone age). Reading times were reduced by 18.0% and 40.0% for reviewers 1 and 2, respectively.

CONCLUSION. Automatic software system showed reliably accurate bone age estimations and appeared to enhance efficiency by reducing reading times without compromising the diagnostic accuracy.

Keywords: bone age, children, deep learning, neural network model

DOI:10.2214/AJR.17.18224

J. R. Kim and W. H. Shim contributed equally to this work.

Received March 12, 2017; accepted after revision July 7, 2017.

S. Kim is employed by Vuno, Inc., which created the deep learning–based automatic software system for bone age determination. J. R. Kim, W. H. Shim, H. M. Yoon, S. H. Hong, J. S. Lee, and Y. A. Cho are employed by Asan Medical Center, which holds patent rights for the deep learning–based automatic software system for bone age assessment.

[1]Department of Radiology and Research Institute of Radiology, Asan Medical Center, University of Ulsan College of Medicine, 88 Olympic-ro 43-gil, Songpa-gu, Seoul 05505, South Korea. Address correspondence to H. M. Yoon (espoirhm@gmail.com).

[2]Vuno Research Center, Vuno Inc., Seoul, South Korea.

Supplemental Data
Available online at www.ajronline.org.

AJR 2017; 209:1–7

0361–803X/17/2096–1

Bone age estimation is crucial for developmental status determinations and ultimate height predictions in the pediatric population, particularly for patients with growth disorders and endocrine abnormalities [1]. Two major left-hand wrist radiograph-based methods for bone age estimation are currently used: the Greulich-Pyle [2] and Tanner-Whitehouse [3] methods. The former is much more frequently used in clinical practice. Greulich-Pyle–based bone age estimation is performed by comparing a patient's left-hand radiograph to standard radiographs in the Greulich-Pyle atlas and is therefore simple and easily applied in clinical practice. However, the process of bone age estimation, which comprises a simple comparison of multiple images, can be repetitive and time consuming and is thus sometimes burdensome to radiologists. Moreover, the accuracy depends on the radiologist's experience and tends to be subjective.

Since 1992, concerns regarding interobserver variability in manual bone age estimation [4] have led to the establishment of several automatic computerized methods for bone age estimation, including computer-assisted skeletal age scores, computer-aided skeletal maturation assessment systems, and BoneXpert (Visiana) [5–14]. BoneXpert was developed according to traditional machine-learning techniques and has been shown to have a good performance for patients of various ethnicities and in various clinical settings [10–14]. The deep-learning technique is an improvement in artificial neural networks. Unlike traditional machine-learning techniques, deep-learning techniques allow an algorithm to program itself by learning from the images given a large dataset of labeled examples, thus removing the need to specify rules [15].

Deep-learning techniques permit higher levels of abstraction and improved predictions from data. Deep-learning techniques

TABLE 1: Distribution of Sex and Age of Subjects

Age (y)	Boys	Girls	Total
3	6	6	12
4	6	6	12
5	6	7	13
6	7	8	15
7	7	6	13
8	4	11	15
9	5	15	20
10	5	8	13
11	6	8	14
12	8	8	16
13	7	6	13
14	5	6	11
15	5	6	11
16	5	6	11
17	5	6	11
Total	87	113	200

have been introduced into the medical field in various ways, including classification of skin cancer [16], detection of diabetic retinopathy in retinal fundus photographs [17], and evaluation of lung nodules on chest CT [18]. Recently, Lee et al. [19] described an automated deep-learning system for bone age assessment, and its test accuracy was 57.32% for the female cohort and 61.40% for the male cohort. We also developed a new deep-learning technique–based automatic software system for bone age estimation using 18,940 left-hand radiographs labeled with the Greulich-Pyle method–estimated bone age that were collected from Asan Medical Center, one of the largest tertiary referring centers in South Korea.

The current study aimed to evaluate the accuracy and efficiency of this automatic software program for bone age assessment and to validate its feasibility in clinical practice.

Materials and Methods

Patients

The institutional review board of Asan Medical Center approved this retrospective study and waived the requirement for informed consent. We collected images from 200 pediatric patients who underwent left-hand radiography at our children's hospital between July and November 2016 and distributed the number of patients evenly by age. We excluded patients younger than 2 years, because the program that we used in this study was developed according to the Greulich-Pyle method, which is not suitable for the evaluation of bone age in patients younger than 2 years. The included pediatric patients' mean (± SD) age was 9.8 ± 4.1 years (range, 3–17 years). Table 1 shows the patients' sex and age distributions. The most common reasons for examination were short stature (n = 89) and precocious puberty (n = 53). The remaining 58 patients underwent evaluation because of known underlying disorders, including congenital adrenal hyperplasia (n = 13), Noonan syndrome (n = 4), hypergonadotrophic hypogonadism after irradiation for known tumor (n = 4), hypoparathyroidism (n = 4), rickets (n = 4), Prader-Willi syndrome (n = 3), Down syndrome (n = 3), methylmalonic acidemia (n = 2), CHARGE (coloboma of the eye, heart defect, atresia of the choanae, retardation of growth, and ear abnormalities) syndrome (n = 2), Crohn disease (n = 2), idiopathic genu valgum (n = 2), Turner syndrome (n = 2), Williams syndrome (n = 2), dyslipidemia (n = 2), glycogen storage disease (n = 2), suprasellar germinoma (n = 2), and glutaric aciduria, hypochondroplasia, Rubinstein-Taybi syndrome, hyperthyroidism, and Kallmann syndrome (n = 1 each).

Computer-Assisted Program for Bone Age Determination

The deep-learning technique was recently used to develop the automatic software program for bone age determination, as described in detail in the *AJR* electronic supplement to this article (available at www.ajronline.org). A total of 18,940 left-hand radiographs and radiologic reports containing Greulich-Pyle–estimated bone ages were used to develop a bone age measurement algorithm. This automatic software program is first activated when a radiologist wishes to read a left-hand radiograph for bone age estimation. The system displays the three most likely estimated bone ages (i.e., first-, second-, and third-rank artificial intelligence bone ages) in order of probability in simultaneously displayed percentages (Fig. 1).

Image Acquisition

Conventional left-hand wrist radiography was performed at the Department of Radiology using Definium 8000 (GE Healthcare) with a voltage of 55 kV and exposure of 160 mAs (with slight variations depending on patient age).

Image Interpretation

The bone age was estimated from 200 radiographs using three methods: the first-rank bone age, suggested by the automated software program; the computer-assisted bone age, as assessed by two radiologists with automated software program assistance; and the Greulich–Pyle atlas-assisted bone age, as assessed by two radiologists with paper-based Greulich-Pyle atlas assistance only (without automated software assistance). The three most likely estimated bone ages by the automatic software program (as described already) were recorded.

The same 200 radiographs were also subjected to bone age assessment by two radiologists with different levels of bone age estimation expertise. Reviewer 1, a clinical pediatric radiology fellow, had clinical experience with more than 500 cases of Greulich-Pyle atlas–based bone age estimation. Reviewer 2, a second-year radiology resident, had no clinical experience in bone age estimation. Before this study, reviewer 2 participated in a 1-day training course involving 20 cases of bone age assessment, directed by an experienced pediatric radiologist. The reviewers independently assessed bone age estimation images in two different sessions. In session 1 (the Greulich-Pyle atlas–assisted session), the reviewers assessed the bone age using the paper form of the Greulich-Pyle atlas [2], without assistance from the automatic software system (Greulich-Pyle atlas–assisted bone age). In session 2 (computer-assisted session), the reviewers estimated the bone age (computer-assisted bone age) using reference bone ages and their individual probabilities as presented by the program. Once the reviewers ran a program and the program automatically displayed the results, including the three most likely estimated bone ages and each probability, the reviewers compared the atlas of the first-rank bone age presented by the program with radiograph of the patient. If the first-rank bone age was determined to be unsuitable in the light of the reviewer's experience, the reviewer repeated the same procedure in the order of second-rank and third-rank bone ages. If there was no appropriate bone age among three most likely bone ages, reviewers selected the most suitable bone age among other bone age classes. For 100 radiographs, the Greulich-Pyle atlas–assisted session was performed and then the computer-assisted session was done after a 1-week washout period. For the remaining 100 radiographs, the Greulich-Pyle atlas–assisted session and computer-assisted sessions were performed again with a 1-week washout period between them. We randomly mixed the order of radiographs for each session. The reviewers were blinded to their own results from the initial Greulich-Pyle atlas–assisted session, as well as the other reviewer's results, but were informed about each patient's sex and chronological age as per standard daily clinical practices.

Each reviewer's bone age assessment reading time was recorded in each session. The starting and finishing times for the interpretation of 200 left-hand radiographs were recorded. Break times were freely allowed, and exact times were

Fig. 1—Screen shot of bone age assessment by software program. Chronological age of this boy is 5 years and 1 month. On right column, estimated bone ages and each probability are shown. First-rank bone age of this patient is 5 years with probability of 71%. Second- and third-rank bone ages are 6 years with probability of 23% and 4 years and 6 months with probability of 6%, respectively.

recorded. The total bone age estimation reading time was calculated by subtracting the starting time from the ending time minus the break time. The computer-assisted session required additional time for activation of the software, preprocessing, analysis, and the display of the results. This additional time was added in the total reading time.

Reference Bone Age

The reference bone age was determined using the Greulich-Pyle atlas [2] via the consensus of two experienced pediatric radiologists (one with 18 years and one with 4 years of clinical experience as pediatric radiologists after fellowship training). These additional readers independently estimated bone ages in all cases without reading time limitations; if they disagreed on cases, the radiographs were reevaluated until a consensus was reached. The remaining disagreements were resolved by a third radiologist (with 24 years of clinical experience).

Statistical Analysis

Quantitative variables are presented as means (± SDs) or medians with ranges. The accuracies of all estimated bone ages (first-rank bone ages by automatic software system, and Greulich-Pyle atlas–assisted and computer-assisted bone ages by radiologists) were assessed using concordance rates (percentages) with the reference bone age. The concordance rate was defined as the proportion of cases for which the estimated bone age and the reference bone age showed perfect agreement. Pearson correlation analysis, scattered plots, and intraclass correlation analysis based on two-way random-effects model were also used to assess the accuracies of all estimated bone ages. In addition, agreement between the reference bone age and estimated bone ages determined by each method were quantified using the root mean square error (RMSE) rather than the SD, which hides overall bias. RMSE is the square root of the average of squared errors, and it is a frequently used measure representing the sample SD of the differences between predicted values and reference values. The smaller the RMSE is, the closer the predicted bone ages and reference bone ages are [20]. A Bland-Altman plot was used to evaluate mean biases and 95% limit of agreements of estimated bone ages. A scatterplot showing differences in bone age units between the first-ranked bone age and reference

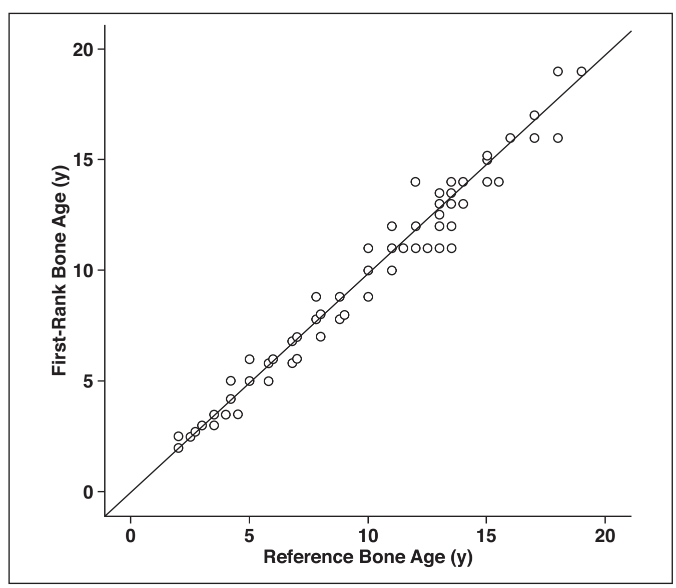

Fig. 2—Correlation between reference bone age and first-rank bone age as calculated by automated software system.

TABLE 2: Agreement Between the Reference Bone Age and Each Estimated Bone Age

Bone Age	Concordance Rate, Percentage (No./Total)	Root Mean Square Error (y)	Intraclass Correlation Coefficient	r
First-rank bone age	69.5 (139/200)	0.60	0.996[a]	0.992[a]
Reviewer 1				
Greulich-Pyle atlas–assisted bone age	63.0 (126/200)[b]	0.53	0.996[a]	0.993[a]
Computer-assisted bone age	72.5 (145/200)[b]	0.42	0.998[a]	0.996[a]
Reviewer 2				
Greulich-Pyle atlas–assisted bone age	49.5 (99/200)[c]	0.79	0.992[a]	0.984[a]
Computer-assisted bone age	57.5 (115/200)[c]	0.70	0.994[a]	0.988[a]

[a] $p < 0.05$.
[b] $p = 0.054$.
[c] $p = 0.109$.

bone age according to the probability (percentage) of the first-rank bone age was also evaluated. We determined the cutoff probability (percentage) of the first-rank bone age to evaluate the reliability of the automatic software system. The first-rank bone age was considered acceptable when a comparison with the reference bone age showed the same result or a difference of only one unit. The Fisher exact test was used to compare differences in the proportions of reliable results between two groups that were categorized by the probability (percentage) of the first-rank bone age. Statistical differences were accepted to be significant at $p < 0.05$. SPSS software (version 20.0, IBM) was used for the analyses. MedCalc (version 16.8, MedCalc) was used to generate Bland-Altman plots.

Results

Accuracy of the Estimated Bone Age by the Automatic Software Program

The first-rank bone age exhibited a 69.5% (139/200) concordance rate with the reference bone age. The second-rank bone age exhibited a 17.0% concordance rate with the reference bone age; therefore, the results of 86.0% of cases listed in terms of first-rank and second-rank bone ages matched the reference bone ages. The third-rank bone age exhibited a 7.0% concordance rate with the reference bone age. Therefore, the results of 93.0% of cases listed in terms of the first-, second-, and third-rank bone ages matched the reference bone age.

The first-rank bone age significantly correlated with the reference bone age (r = 0.992; p < 0.001; Fig. 2; intraclass correlation coefficient, 0.996; p < 0.001; Table 2). The Bland-Altman plot of agreement between the reference bone age and first-rank bone age showed a mean difference of –0.20 years (95% limit of agreement, ± 1.20; Fig. 3). The RMSE of the first-rank bone age was 0.60 year (Table 2).

The scatterplot in Figure 4 depicts differences in bone age units between the first-rank bone age and reference bone age according to the probability (percentage) of the first-rank bone age. The difference in bone age units between the first-rank bone age and reference bone age increased in cases with a probability of greater than or equal to 40%. After setting 40.0% as a cutoff, 161 cases (80.5%) were found to have the probability of the first-rank bone age of greater than or equal to 40%. Among them, 155 (96.3%) exhibited reliable results relative to the reference bone age (i.e., identical results [n = 118 cases, 73.3%] or difference of only one unit [n = 37 cases, 23.0%]; Table 3). Six cases (3.7%) exhibited a difference of two bone age units, and no cases had a difference exceeding two units. On the other hand, among 39 cases with a probability of less than 40%, four (10.3%) exhibited differences of two units and four (10.3%) exhibited differences of three or more bone age units relative to the reference bone age. The number of cases with reliable results was higher among cases with a first-rank bone age probability of greater than or equal to 40% (155/161 for ≥ 40% vs 31/39 for < 40%; p < 0.001).

Application of the Automatic Software Program in Clinical Practice

Effect of the automatic software program on radiologists' performances—Bland-Altman plots showed agreement of the reference bone

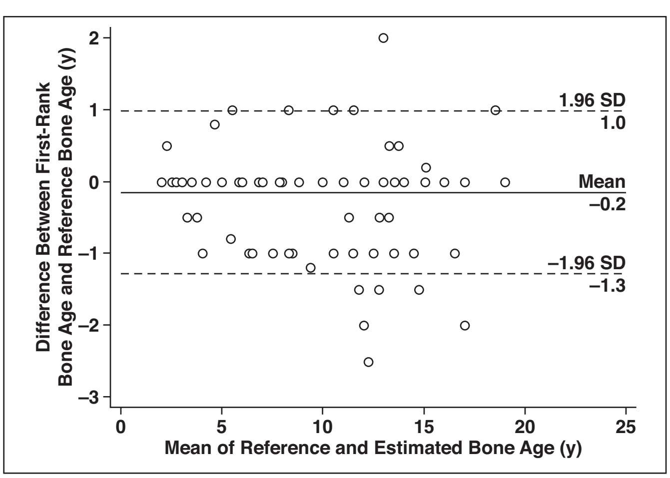

Fig. 3—Bland-Altman plot of reference bone age and first-rank bone age as calculated by automated software system.

TABLE 3: Difference in Bone Age Units Between the First-Rank Estimated Bone Age Suggested by the Automated Software System and the Reference Bone Age

Difference in Bone Age Units	Probability of First-Rank Estimated Bone Age	
	< 40% (n = 39)	≥ 40% (n = 161)
0	21 (53.8)	118 (73.3)
1	10 (25.6)	37 (23.0)
2	4 (10.3)	6 (3.7)
3	3 (7.7)	0
4	1 (2.6)	0

Note—Data are number (%) of subjects.

age with the Greulich-Pyle atlas–assisted or computer-assisted bone ages (see Fig. S1, which can be viewed in the *AJR* electronic supplement to this article, available at www.ajronline.org). Relative to the reference bone age, the mean difference in the Greulich-Pyle atlas–assisted bone age as determined by reviewer 1 was –0.03 year (95% limit of agreement, ± 1.05 year; Fig. S1A). In the computer-assisted session, the mean difference was –0.04 year (95% limit of agreement, ± 0.83 year; Fig. S1B). For reviewer 2, the mean difference in the Greulich-Pyle atlas–assisted session was –0.04 year (95% limit of agreement, ± 1.56 years; Fig. S1C). In the computer-assisted session, the mean difference was –0.10 year (95% limit of agreement, ± 1.38 years; Fig. S1D). The 95% limits of agreement for both reviewers tended to decrease in the computer-assisted sessions. Significant positive correlations were observed between the reference bone age and estimated bone ages determined by the two reviewers in each session ($r = 0.984–0.996$; $p < 0.001$; Table 2; see also Fig. S2, which can be viewed in the *AJR* electronic supplement to this article, available at www.ajronline.org).

The agreements of reference bone age with each estimated bone age according to the results of the two reviewers were quantified using concordance rates and RMSE (Table 2). For reviewer 1, the concordance rates of the Greulich-Pyle atlas–assisted and computer-assisted sessions with the reference bone age were 63.0% and 72.5%, respectively. Moreover, the RMSE was slightly lower for the computer-assisted session (0.42 year) than for the Greulich-Pyle atlas–assisted session (0.53 year). For reviewer 2, the concordance rate was slightly higher for the computer-assisted session (57.5%) than for the Greulich-Pyle atlas–assisted session (49.5%), but both concordance rates still remained lower than those for reviewer 1 as well as first-rank bone age. RMSE was also slightly lower for the computer-assisted session (0.70 year) than the Greulich-Pyle atlas–assisted session (0.79 year).

Reading times for bone age assessment—The total reading times of reviewer 1 were 188 minutes 22 seconds for the Greulich-Pyle atlas–assisted session and 154 minutes 31 seconds for the computer-assisted session, saving 33 minutes 51 seconds for the latter (i.e., 18.0% of the Greulich-Pyle atlas–assisted session reading time). Reviewer 2 required 180 minutes 55 seconds for the Greulich-Pyle atlas–assisted session and 108 minutes 33 seconds for the computer-assisted session, saving 72 minutes 22 seconds (i.e., 40.0% of the Greulich-Pyle atlas–assisted session reading time). Additional data may be seen in Figures S3 and S4 and Table S5 in the *AJR* electronic supplement to this article (available at www.ajronline.org).

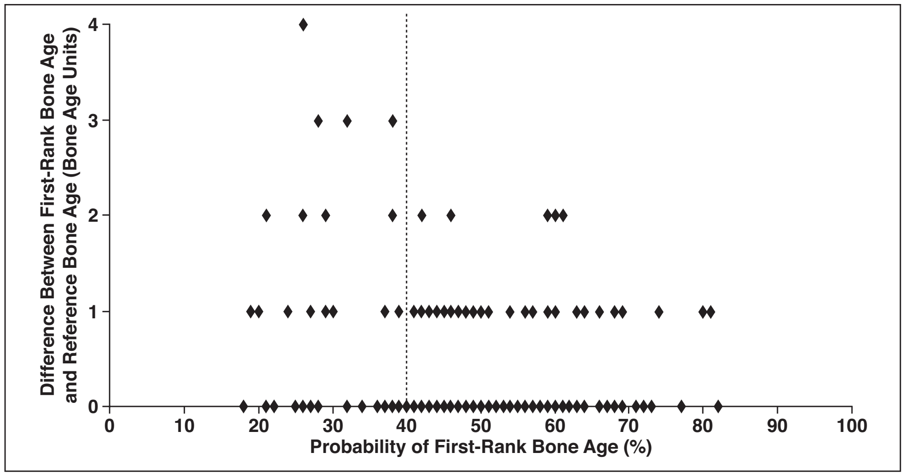

Fig. 4—Scatterplot of differences in bone age units between first-rank bone ages and reference bone ages according to probability of first-ranked bone age being concordant with reference bone age. Vertical dotted line indicates 40% cutoff value.

Discussion

The results of our study show the accuracy and efficiency of a newly developed deep-learning technique–based automatic software program for bone age estimation. We determined that the concordance rate between the automatic software system and reference bone age was 69.5%. Among cases with a first-rank bone age probability of greater than or equal to 40%, the estimated bone age by automatic software program was identical to or had a one unit difference from the reference bone age in 96.3% of cases. When the automatic software system was implemented as the second opinion in daily clinical practice, the radiologists' reading times were reduced by 29%, on average, without compromising the accuracy of bone age estimation.

The first-rank bone age significantly correlated with the reference age ($r = 0.992$; $p < 0.05$). The concordance rate of the first-rank bone age with the reference age was 69.5%. This rate increased to 93% when the first-, second-, and third-rank bone ages were combined. In previous validation studies on BoneXpert, the RMSE values between automatic ratings by BoneXpert and the reference bone age ranged from 0.61 to 0.80 year [10–14]. The RMSE of the first-rank bone ages using the new automatic software system was 0.60 year, similar to that of BoneXpert. Therefore, we believe that this new automatic software system shows relatively accurate results, with a strong possibility that the most appropriate bone age exists among three most likely bone ages (i.e., first-, second-, and third-rank bone ages). Because our study is preliminary and uses a deep learning–based bone age estimation program that is still under development, the accuracy including the concordance rate is expected to increase through further developments of the deep-learning technique and the addition of more data input.

The automatic software system displays three ranked bone ages along with probability percentages. We found that a first-rank bone age probability of greater than or equal to 40% could be used as a cutoff for identifying reliable results (i.e., the same as or only one unit different from the reference bone age). We observed a significant difference in the proportion of reliable results between the two groups categorized using a cutoff probability of 40%. Our results suggest that cases with an automatic software system–based first-rank bone age probability of less than 40% should be carefully handled.

RMSEs decreased with the use of this program as a decision supporting system for both reviewer 1 (from 0.53 to 0.42 year) and reviewer 2 (from 0.79 to 0.70 year), which is considered to be an increased in accuracy. The concordance rates increased with the use of this program for both reviewer 1 (from 63.0% to 72.5%) and reviewer 2 (from 49.5% to 57.5%), which is also considered to be an increased in accuracy; however, the differences were not statistically significant ($p = 0.054$ for reviewer 1; $p = 0.109$ for reviewer 2). For reviewer 2, however, the experience of bone age estimation during the Greulich-Pyle atlas–assisted session may have affected that reviewer's accuracy during the computer-assisted session. Because the reviewers conducted the computer-assisted sessions after Greulich-Pyle atlas–assisted sessions, 200 cases of experience during the Greulich-Pyle atlas–assisted sessions were added at the beginning of the computer-assisted sessions. Considering that reviewer 2 had only 20 cases of experience at the beginning of Greulich-Pyle atlas–assisted sessions by the 1-day training course, this order of two sessions might have positively affected higher concordance rates of the computer-assisted session for reviewer 2.

The efficiency of the automatic software system must be considered. The automatic software system allowed us to reduce the total reading times of the experienced (18.0%) and novice (40.0%) radiologists, along with slight improvements in bone age estimation accuracy. Time saving was more noticeable for reviewer 2 than for reviewer 1; however, this result might have been affected by the self-learning effect of reviewer 2. The 29%, on average, time-saving rate may be critical, given the clinical burden of bone age estimation in daily practice. Thus, the automatic software system could enhance clinical efficiency.

BoneXpert is the most successful automatic computerized method for bone age estimation up to now. BoneXpert automatically reconstructs the borders of 15 bones (including metacarpal and phalangeal bones, the distal radius, and the ulna) from radiographs of hands using a generative model (active appearance model), and then computes intrinsic bone ages from the shape, intensity, and texture scores derived from principal component analysis for each bones. Finally, it transforms the intrinsic bone ages into Greulich-Pyle or Tanner-Whitehouse bone age [2]. Although BoneXpert has been found to have a good performance for patients of various ethnicities and in various clinical settings, it has important limitations. BoneXpert uses only 15 bones, and several short bones and carpal bones are not included in bone age evaluation. It also automatically rejects images with poor image quality or abnormal bone structure, and the rejection rate was reported to be up to 4.5% [21]. In contrast, our program was developed according to the convolutional neural network, which is one of various deep-learning architectures that uses all bones included in radiographs for the evaluation of bone age and also automatically preprocesses radiographs for the evaluation without rejecting it. Deep learning has been introduced into the field of medical imaging in various ways. Recently, Lee et al. [19] described an automated deep-learning system for bone age assessment and its test accuracy (57.32% for the female cohort and 61.40% for the male cohort). Compared with their results, our automatic software system showed better accuracy, with a concordance rate of 69.5%.

Our study had several limitations. First, it included a small sample size (200 cases) at a single center. Furthermore, cases were limited to patients of a single ethnicity, although previous studies reported racial differences in growth patterns at certain ages [22, 23]. Second, most cases had various clinical indications. Further investigations on the accuracy and feasibility of this automatic software system should include healthy populations. Third, the program we used in this study was developed on the basis of the Greulich-Pyle method, which is not suitable for the evaluation of bone age in patients younger than 2 years. Therefore, patients younger than 2 years are not eligible for this program. Fourth, the automatic software program showed a predicted bone age that was significantly different from the reference bone age at certain ages. Bland-Altman plots of the reference bone age and first-rank bone age by the automated system (Fig. 3) showed that patients with large differences between the reference bone age and first-rank bone age by the automated system were between 12 and 15 years old. Bone ages at this age group are classified by minute changes in epiphyses of the phalangeal bones, distal radius, and ulna. In this respect, large differences were observed in this age group. More data regarding these age groups should be included in the additional learning dataset to increase the accuracy of this software program. Another limitation of our study is that the efficiency of the program was identified through two radiologists with relatively little experience in interpreting bone age and did not include a senior radiologist. A further study of the assessment of differences in efficiency of this program between a senior radiologist and a less-experienced radiologist is needed. Finally, both reviewers conducted computer-assisted sessions after Greulich-Pyle atlas–assisted sessions. Despite the 1-week washout period between the sessions, this order might have positively affected concordance rates and reduced reading times during the computer-assisted session, particularly for reviewer 2, because of the self-training effect.

In conclusion, this new automatic software system showed reliably accurate bone age estimations and appeared to enhance efficiency by reducing reading times without compromising the diagnostic accuracy.

Acknowledgment

We thank Seon-Ok Kim, a statistician at Asan Medical Center, for her assistance with statistical analysis, which greatly improved the manuscript.

References

1. Zerin JM, Hernandez RJ. Approach to skeletal maturation. *Hand Clin* 1991; 7:53–62
2. Greulich WW, Pyle SI. *Radiographic atlas of skeletal development of hand wrist*, 1st ed. Stanford, CA: Stanford University Press, 1959
3. Tanner JM, Healy MJ, Goldstein H, Cameron N. *Assessment of skeletal maturity and prediction of adult height (TW3 method)*, 3rd ed. London, UK: Saunders Company, 2001
4. 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* 2001; 176:507–510
5. Tanner J, Gibbons R, Bock R. An image analysis system for TW skeletal maturity. *Horm Res* 1992; 37:11–15
6. Tanner JM, Gibbons RD. Automatic bone age measurement using computerized image analysis. *J Pediatr Endocrinol* 1994; 7:141–145
7. 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; 42:288–294
8. Sato K, Ashizawa K, Anzo M, et al. Setting up an automated system for evaluation of bone age. *Endocr J* 1999; 46(suppl):S97–S100
9. Thodberg HH, Kreiborg S, Juul A, Pedersen KD. The BoneXpert method for automated determination of skeletal maturity. *IEEE Trans Med Imaging* 2009; 28:52–66
10. Martin DD, Deusch D, Schweizer R, Binder G, Thodberg HH, Ranke MB. Clinical application of automated Greulich-Pyle bone age determination in children with short stature. *Pediatr Radiol* 2009; 39:598–607
11. Martin DD, Meister K, Schweizer R, Ranke MB, Thodberg HH, Binder G. Validation of automatic bone age rating in children with precocious and early puberty. *J Pediatr Endocrinol Metab* 2011; 24:1009–1014
12. Thodberg HH, Savendahl L. Validation and refer-

ence values of automated bone age determination for four ethnicities. *Acad Radiol* 2010; 17:1425–1432
13. Martin DD, Sato K, Sato M, Thodberg HH, Tanaka T. Validation of a new method for automated determination of bone age in Japanese children. *Horm Res Paediatr* 2010; 73:398–404
14. van Rijn RR, Lequin MH, Thodberg HH. Automatic determination of Greulich and Pyle bone age in healthy Dutch children. *Pediatr Radiol* 2009; 39:591–597
15. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; 521:436–444
16. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; 542:115–118
17. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016; 316:2402–2410
18. Setio AA, Ciompi F, Litjens G, et al. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. *IEEE Trans Med Imaging* 2016; 35:1160–1169
19. Lee H, Tajmir S, Lee J, et al. Fully automated deep learning system for bone age assessment. *J Digit Imaging* 2017; 30:427–441
20. Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. *Geosci Model Dev* 2014; 7:1247–1250
21. Zhang J, Lin F, Ding X. Maturation disparity between hand-wrist bones in a Chinese sample of normal children: an analysis based on automatic BoneXpert and manual Greulich and Pyle atlas assessment. *Korean J Radiol* 2016; 17:435–442
22. Ontell FK, Ivanovic M, Ablin DS, Barlow TW. Bone age in children of diverse ethnicity. *AJR* 1996; 167:1395–1398
23. 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

FOR YOUR INFORMATION

A data supplement for this article can be viewed in the online version of the article at: www.ajronline.org.