Journal of Imaging Informatics in Medicine (2025) 38:1915–1923
https://doi.org/10.1007/s10278-024-01323-3

The Impact of Artificial Intelligence on Radiologists' Reading Time in Bone Age Radiograph Assessment: A Preliminary Retrospective Observational Study

Sejin Jeong[1] · Kyunghwa Han[2] · Yaeseul Kang[1] · Eun-Kyung Kim[1] · Kyungchul Song[3] · Shreyas Vasanawala[4] · Hyun Joo Shin[1]

Received: 30 May 2024 / Revised: 24 October 2024 / Accepted: 28 October 2024 / Published online: 11 November 2024

Abstract

To evaluate the real-world impact of artificial intelligence (AI) on radiologists' reading time during bone age (BA) radiograph assessments. Patients (<19 year-old) who underwent left-hand BA radiographs between December 2021 and October 2023 were retrospectively included. A commercial AI software was installed from October 2022. Radiologists' reading times, automatically recorded in the PACS log, were compared between the AI-unaided and AI-aided periods using linear regression tests and factors affecting reading time were identified. A total of 3643 radiographs (M:F=1295:2348, mean age 9.12 ± 2.31 years) were included and read by three radiologists, with 2937 radiographs (80.6%) in the AI-aided period. Overall reading times were significantly shorter in the AI-aided period compared to the AI-unaided period (mean 17.2 ± 12.9 seconds vs. mean 22.3 ± 14.7 seconds, $p < 0.001$). Staff reading times significantly decreased in the AI-aided period (mean 15.9 ± 11.4 seconds vs. mean 19.9 ± 13.4 seconds, $p < 0.001$), while resident reading times increased (mean 38.3 ± 16.4 seconds vs. 33.6 ± 15.3 seconds, $p = 0.013$). The use of AI and years of experience in radiology were significant factors affecting reading time (all, $p \leq 0.001$). The degree of decrease in reading time as experience increased was larger when utilizing AI (-1.151 for AI-unaided, -1.866 for AI-aided, difference =-0.715, $p<0.001$). In terms of AI exposure time, the staff's reading time decreased by 0.62 seconds per month (standard error 0.07, $p<0.001$) during the AI-aided period. The reading time of radiologists for BA assessment was influenced by AI. The time-saving effect of utilizing AI became more pronounced as the radiologists' experience and AI exposure time increased.

Keywords Artificial intelligence · Radiologists · Radiography · Time · Bone age measurement

✉ Hyun Joo Shin
lamer-22@yuhs.ac

[1] Department of Radiology, Research Institute of Radiological Science and Center for Clinical Imaging Data Science, Yongin Severance Hospital, Yonsei University College of Medicine, 363, Dongbaekjukjeon-daero, Giheung-gu, 16995 Yongin-si, Gyeonggi-do, Republic of Korea

[2] Department of Radiology, Research Institute of Radiological Science and Center for Clinical Imaging Data Science, Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea

[3] Department of Pediatrics, Yonsei University College of Medicine, Gangnam Severance Hospital, Seoul, Republic of Korea

[4] Department of Radiology, Stanford University, Stanford, CA, USA

Abbreviations

AI	Artificial intelligence
BA	Bone age
GP	Greulich-Pyle
TW	Tanner-Whitehouse
PACS	Picture archiving communicating system

Introduction

Bone age (BA) is one of the most crucial parameters in evaluating growth in pediatric patients with short stature, precocious puberty, metabolic diseases, or various genetic conditions [1]. The Greulich-Pyle (GP) method is the most commonly used approach, which involves comparing a patient's left wrist radiograph to an atlas derived from white children of high socioeconomic status in 1930s America [2, 3]. Alternatively, the Tanner-Whitehouse (TW) method assigns scores to each bone in the hand and wrist and then summing up the scores to determine BA [4]. However, these manual BA assessments are time-consuming, labor-intensive, and prone to subjective results [5]. With the emergence of artificial intelligence (AI) in radiology, BA assessment has become one of the most promising application for AI, as it involves processus such as detection and classification based on relatively standardized findings from a single left-hand wrist image [6]. Consequently, in the field of pediatric radiology, BA interpretation was among the first areas to clinically adopt AI and remains the most active area for AI research in pediatric radiology to date [7–10].

As AI becomes widely adopted in various fields of radiology, many studies have demonstrated its actual impact on improving radiology workflows [11, 12]. AI not only enhances diagnostic accuracy but it also supports tasks like image triaging and reducing reading time, positioning it as a valuable tool for radiologists [11, 13–15]. In terms of BA radiographs, several commercial AI products have already demonstrated their reliability and accuracy for interpretation [4, 10]. Studies have shown that AI-based BA assessment achieve a higher correlation with reference standards than manual assessments, while demonstrating superior inter- and intra-observer agreement [16–18].

Despite these benefits, the effect of AI on radiologists' reading time remains a subject of debate. AI has generally been expected to reduce radiologists' reading times by streamlining decision-making. However, some concerns have raised that AI might actually increase the image interpretation time by adding extra steps to the workflow [19, 20]. Various factors, such as image type, disease complexity, accuracy of AI, the method of AI integration, and the radiologists' experience, could influence how AI impacts reading time [15, 19]. Given these complexities, the effect of AI on reading time should be thoroughly investigated, taking into account various influencing factors. However, to our knowledge, limited research has focused on the real-world impact of AI on radiologists' reading time for BA assessments in real clinical environments, as opposed to simulated research settings, especially with respect to the radiologists' varying levels of experience.

Therefore, the purpose of this study was to examine the real-world impact of AI on radiologists' reading time during BA radiograph assessments in daily practice and to identify the factors, including radiologists' experience levels, that influence the effect of AI.

Materials and Methods

Patients

This retrospective study was approved by the Institutional Review Board of our institution. Informed consent was waived due to its retrospective nature. This study was conducted in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines and regulations. Patients under 19 years old who underwent left-hand radiographs for BA assessment between December 2021 and October 2023 due to clinical necessity were included. These patients underwent BA assessment due to conditions such as growth hormone deficiency, idiopathic short stature, precocious puberty, and metabolic diseases that could affect their growth. The need for BA assessment was determined by a pediatric endocrinologist. Patients were excluded if the radiologists' reading time of their BA radiographs could not be used for analysis, such as when it had a negative reading time on the Picture Archiving Communicating System (PACS) log records or if they were time outliers, as described in detail below.

BA Radiograph Acquisition and AI Application

Standard left-hand wrist radiographs for BA assessment were obtained using Innovision-EXII (DK medical systems, Seoul, Korea) or GXR-82SD (DRGEM, Seoul, Korea). A commercialized AI-based BA assessment software (VUNO Med-BoneAge, version 1.0.5, Vuno Inc., Seoul, South Korea), approved by the Korean Ministry of Food and Drug Safety for clinical use, was installed in October 2022 in our hospital. This deep learning-based software was trained using more than 50,000 left-hand wrist radiographs, including over 35,000 radiographs of Koreans using the GP method [17, 21, 22]. After installation in our institution, all BA radiographs were analyzed by the AI and attached to the PACS as secondary capture images. The analyzed result displays heatmap indicating areas the model considers important, along with the optimal BA and its probability within age ranges (Fig. 1). When the radiographs were taken, it automatically transferred to the AI server in our hospital as soon as the radiographers verified the images. After about 5–10 s, the analyzed images were generated automatically and attached to the original images in the PACS. Therefore, radiologists and clinicians could freely refer to the AI results by scrolling down the images during their reading whenever they wanted. The AI results were displayed next to the original image, allowing doctors to first open the original images through the worklist and then refer to the AI results by scrolling down the images using PACS. Although there was no requirement to review the original image first, it was necessary because the original image had to be opened first to access the AI results in our hospital. We did not use any additional user interface (UI) elements to present the AI results, as we seamlessly integrated them into clinical practice and presented the AI results in the most intuitive manner for all users, regardless of experience level.

Fig. 1 A left-hand radiograph from a 6-year-old girl was analyzed by AI for bone age assessment. The AI software determined the bone age to be 7 years and 11 months, considering a 90.4% probability of 94 months and a 7.5% probability of 106 months. This was read by reader 1 during the AI-aided period and the interpretation time was 37 seconds in PACS log record

When the AI software was not available in our hospital, radiologists usually interpreted BA radiographs using the GP atlas, patients' chronological age, sex, previous images, and medical history through electronic medical records (EMR). We defined this time as the 'AI-unaided period' (uninstalled period from December 2021 to September 2022 and server maintenance period in October 2023). When the AI software was available in our PACS after the installation or maintenance period, radiologists could refer to the AI results just by scrolling down the images on PACS when they interpreted the images, and we set this as the 'AI-aided period' (from October 2022 to September 2023). Radiologists could freely refer to the AI results and GP atlas, EMR, or previous images as usual to determine whether they would accept the AI results or not. For the BA assessment, radiologists, including residents, were able to read the images independently, without additional confirmation process that would extend the reading time.

Reading Time Measurement with or without AI

The reading time of all included BA radiographs during the study period was analyzed. Reading time was defined as the duration in seconds from the last time the radiologist opened the radiograph until the same radiologist first transcribed the dedicated image. The reading time of each radiograph was automatically measured through the PACS log record, and we could extract the values retrospectively. No intentional measurements were necessary to determine the actual reading time.

We excluded instances of negative reading time, which occurred when the image was unintentionally opened again after transcription, with the last opening time being later than the transcription time. Additionally, we excluded reading time outliers that could result from unexpected interruptions from other tasks after image opening. The time outliers were defined using the 1.5 interquartile range (IQR) method, as outlined in a previous study, with detailed formulas described in the statistical analysis section [15]. These exclusions were unavoidable because this was a retrospective study, and we relied solely on PACS log records for objective time measurement.

The interpreting radiologists and their years of experience in radiology at the time of interpreting the dedicated radiographs were evaluated. The number of radiographs and their reading time in seconds during the 'AI-unaided' and 'AI-aided periods' were assessed for each radiologist.

Statistical Analysis

For statistical analysis, the R program (version 4.1.3; Foundation for Statistical Computing, Vienna, Austria) was utilized. The reading time outlier in this study was defined as any time exceeding 67.5 s, based on the 1.5 IQR method, as described in a previous study [15]. The determination formula is as follows; third quartile (33 s) + difference between the third and first quartiles (33–10 s) × 1.5 = 67.5 s. The Chi-square test and two-sample *t*-test were used for comparing radiographs and patient characteristics. The AI-unaided and AI-aided periods were compared in terms of reading times based on radiologists using the two-sample *t*-test. Univariate and multivariate linear regression tests were used to determine the factors affecting the reading time of radiographs, including AI use, reader group composed of residents and staff, and radiologists' experience in radiology. Given the small sample size and the distinct differences in experience among the readers, we opted for a linear regression model by treating the readers as a fixed effect. Interaction effect of the use of AI and experience in radiology on the reading time was also assessed. In addition, AI exposure time in months was evaluated during the AI-aided period, and a linear regression analysis was used to assess its impact on reading time for both resident and staff radiologists. Scatter plots with linear regression lines were generated to illustrate the changing pattern of reading time based on years of experience in radiology and AI exposure time. To better capture the trend between the AI-aided and AI-unaided periods, we assigned negative values, such as -1 month, -2 months, etc., to the AI-unaided months, using the AI implementation month as the reference point. A scatter plot and regression line were then created based on this adjustment for the AI-unaided period. A p-value less than 0.05 was considered statistically significant.

Result

Patient Characteristics

During the study period, a total of 5,989 BA radiographs were included and interpreted by three radiologists. According to the exclusion criteria, 1,818 radiographs were excluded due to the negative reading time, and 528 radiographs were excluded as reading time outliers according to the IQR method.

Finally, a total of 3,643 radiographs (M: F = 1,295:2,348, with a mean age of 9.12 ± 2.31 years) were included in this study. Among them, 706 radiographs were in the AI-unaided period, and the remaining 2,937 (80.6%) radiographs were in the AI-aided period. The most frequent indications for the radiographs in this study were idiopathic short stature (*n* = 1,658), precocity puberty (*n* = 1,577), obesity and overweight (*n* = 323), growth hormone deficiency (*n* = 65), and hypothyroidism (*n* = 20). Patients' age and sex did not show significant differences between the AI-unaided and AI-aided periods (Table 1).

Comparison of Reading Times

During the study period, a total of three radiologists interpreted the BA radiographs. Two radiologists were residents with up to one (reader 1) to two (reader 2) years of experience in radiology at the time of interpretation. The third radiologist was a board-certified pediatric radiologist staff with up to 14 years of experience in radiology when interpreting radiographs (reader 3). The total number of radiographs read by reader 1, 2, and 3 were 187 (AI-unaided vs. AI-aided periods = 51:136), 108 (AI-unaided vs. AI-aided periods = 71:37) and 3348 (AI-unaided vs. AI-aided periods = 584:2764), respectively. In the AI-aided period, AI exposure time were 1–2 months for residents and 1–13 months for staff radiologist. Since this was a retrospective study, the only radiologists who read the images were a dedicated pediatric radiologist and two residents who were on duty during the study period after data extraction. Therefore, it was inevitable to include only these three radiologists, as they were the only ones who read the images during the study period. About the difference in image numbers, a dedicated pediatric radiologist, who was the only

Table 1 Demographics of patients and reading time of BA radiographs according to the use of AI

Variables			All period (*n* = 3,643)	AI-unaided period (*n* = 706)	AI-aided period (*n* = 2,937)	*p*-value
Sex (M: F)			1,295 : 2,348	256 : 450	1,039 : 1,898	0.66*
Age of patients (year)			9.12 ± 2.31	9.17 ± 2.46	9.11 ± 2.27	0.528
Radiologists' experience in radiology (years)			12.48 ± 3.42	10.79 ± 4.49	12.88 ± 2.96	< 0.001
Reading time (seconds)	All	Reader 1&2&3	18.16 ± 13.39	22.3 ± 14.66	17.17 ± 12.88	< 0.001
	Residents	Reader 1	35.73 ± 16.31 (*n* = 187)	30.41 ± 14.4 (*n* = 51)	37.72 ± 16.59 (*n* = 136)	0.004
		Reader 2	37.41 ± 15.72 (*n* = 108)	35.89 ± 15.62 (*n* = 71)	40.32 ± 15.69 (*n* = 37)	0.167
		Reader 1 & 2	36.34 ± 16.09 (*n* = 295)	33.6 ± 15.3 (*n* = 122)	38.28 ± 16.39 (*n* = 173)	0.013
	Staff	Reader 3	16.56 ± 11.86 (*n* = 3348)	19.94 ± 13.38 (*n* = 584)	15.85 ± 11.39 (*n* = 2764)	< 0.001
	p-value between reader 1 and 2		0.384	0.051	0.393	.
	p-value between reader 1&2 (residents) and 3 (staff)		< 0.001	< 0.001	< 0.001	.

Values are presented in number or mean ± standard deviation

**P*-value obtained from Chi-square test

BA = bone age, *AI* = artificial intelligence pediatric radiologist in our hospital, read most of the children's images. The other radiologists who read BA images during the study period were two residents during their training months, so differences in the number of images were inevitable. In addition, unlike the staff radiologist who was consistently assigned to the pediatric section, each resident rotated through the pediatric part only 1–2 times per year, making the variation in AI exposure months an inevitable outcome.

The overall reading time for all readers was a mean of 18.2 ± 13.4 s. Reading time was significantly shorter for reader 3 compared to reader 1 and 2 (mean 16.6 ± 11.9 s vs. mean 36.3 ± 16.1 s, $p < 0.001$), while there was no significant difference in time between readers 1 and 2 (mean 35.7 ± 16.3 s vs. mean 37.4 ± 15.7 s, $p = 0.384$).

When comparing the reading times of radiologists between the AI-aided and AI-unaided periods, total reading times were significantly decreased in the AI-aided period compared to the AI-unaided period (mean 17.2 ± 12.9 s vs. mean 22.3 ± 14.7 s, $p < 0.001$). While there was a significant decrease in reading time for reader 3 in the AI-aided period (mean 15.9 ± 11.4 s vs. mean 19.9 ± 13.4 s, $p < 0.001$), the total reading time for readers 1 and 2 increased in the AI-aided period (mean 38.3 ± 16.4 s vs. 33.6 ± 15.3 s, $p = 0.013$), with reader 1 exhibiting a more significant increase. The reading time between readers 1 and 2 was not significantly different during the AI-unaided ($p = 0.051$) and AI-aided periods ($p = 0.393$), but significantly decreased for reader 3 in each period (all, $p < 0.001$). The results regarding reading times are presented in Table 1.

Factors Affecting the Reading Time of BA Radiographs

In the univariable linear regression test, the use of AI (estimate − 5.134, $p < 0.001$), whether the reader was a staff member (estimate − 19.781, $p < 0.001$), and the total years of experience in radiology (estimate − 1.661, $p < 0.001$) were significantly associated with reading time. When we separately analyzed the effect of the presence of AI on each reader, the presence of AI was significantly associated with decreased reading time in reader 3 (estimate − 4.095, $p < 0.001$). However, in readers 1 and 2, the presence of AI was a significant factor for increasing reading time (estimate 4.679, $p = 0.014$), especially in reader 1 (estimate 7.309, $p = 0.006$) (Table 2).

When we performed multivariable linear regression test for the significant variables, variance inflation factors (VIF) of whether the reader was a staff member and total years of experience in radiology were 646 and 644, respectively. Considering multicollinearity, these two variables could not simultaneously serve as independent variables in a multivariable regression. Therefore, only two variables including the use of AI could be chosen for multivariate regression. As a result, the use of AI and whether the reader was a staff were significant factors affecting the reading time (estimate − 2.961, $p < 0.001$ and estimate − 19.073, $p < 0.001$, respectively). Similarly, the use of AI and total years of experience in radiology were significant factors affecting the reading time (estimate − 1.764, $p = 0.001$ and estimate − 1.612, $p < 0.001$, respectively). When we assessed the interaction effect of the use of AI and experience in radiology, the degree of decrease in reading time as experience increased was larger when utilizing AI (-1.151 for AI-unaided, -1.866 for AI-aided, difference =-0.715, $p < 0.001$) (Table 2).

In terms of AI exposure time, the residents' reading time increased by 5.914 s for each month of AI exposure (standard error 2.21, $p = 0.008$), while the staff's reading time decreased by 0.62 s per month (standard error 0.07, $p < 0.001$) during the AI-aided period. As shown in Fig. 2, there were distinct patterns in reading times between the AI-aided and AI-unaided periods, depending on radiologists' experience and AI exposure time. In Fig. 2a, reading times decreased as years of experience in radiology increased for both AI-aided and AI-unaided periods. However, less experienced radiologists (residents with 1–2 years of experience) showed higher variability and slower reading times, while more experienced staff (12–14 years of experience)

Table 2 Factors affecting the reading time of BA radiographs using linear regression analysis

Variables		Univariable linear regression			Multivariable linear regression					
		estimates	se	*p*-value	estimates	se	*p*-value	estimates	se	*p*-value
Use of AI		-5.134	0.555	< 0.001	-2.961	0.519	< 0.001	-1.764	0.523	0.001
Staff		-19.781	0.745	< 0.001	-19.073	0.752	< 0.001	.	.	.
Radiologists' experience in radiology (years)		-1.661	0.059	< 0.001	.	.	.	-1.612	0.061	< 0.001
Use of AI	Reader 1	7.309	2.632	0.006						
	Reader 2	4.437	3.172	0.165						
	Reader 1 & 2 (residents)	4.679	1.886	0.014						
	Reader 3 (staff)	-4.095	0.536	< 0.001						

BA = bone age, *se* = standard error, *AI* = artificial intelligence exhibited consistently faster reading times. Figure 2b illustrates the effect of AI exposure time on reading times. For staff radiologists who continuously interpreted BA radiographs before and after AI implementation, and who had ongoing exposure to the AI software after the implementation, a gradual decrease in reading times was observed after AI implementation, in contrast to the flat, horizontal trend seen before AI was introduced. This demonstrates a clear visual effect of prolonged AI usage on reducing reading times (Fig. 2b).

Discussion

Our study demonstrated that the overall reading time of radiologists for BA radiographs was significantly reduced with the use of AI, even after adjusting for experience in radiology. However, the time-saving effect was not consistent across all radiologists. Specifically, the time-saving effect from utilizing AI was more pronounced for experienced radiologist, while radiology residents exhibited prolonged reading time when utilizing AI. This trend was particularly evident among resident with less experience. While reducing overall reading time by using AI was consistent with prior studies [16, 17, 23], the specific impact of AI assistance varied depending on the radiologists' experience and AI exposure time in this study.

Previous studies have suggested that AI can help alleviate radiologist burnout by reducing workload [24]. However, it remains uncertain whether AI can consistently reduce the workload of radiologists, particularly in real clinical settings [19]. Several studies have explored the impact of AI on reading time in different imaging modalities [19]. For example, in chest radiograph, conflicting results exist. One prospective observational study found that AI significantly reduced radiologists' reading times, but when abnormalities were detected, the reading time increased [15]. Conversely, a randomized crossover study demonstrated that AI significantly reduced reading times, regardless of abnormality detection [25]. Other studies showed that reading times decreased with AI assistance, regardless of radiologist's experience level, including residents, general radiologists, and thoracic radiologists [25, 26]. However, in health screening environments with low disease prevalence, AI increased reading times for less experienced residents [27]. These discrepancies highlight the complex interaction between AI and reading times, which may depend on several factors such as experience, disease prevalence, AI integration methods, and AI exposure time. Therefore, research focused on various factors is required to assess AI's effectiveness in optimizing clinical workflows.

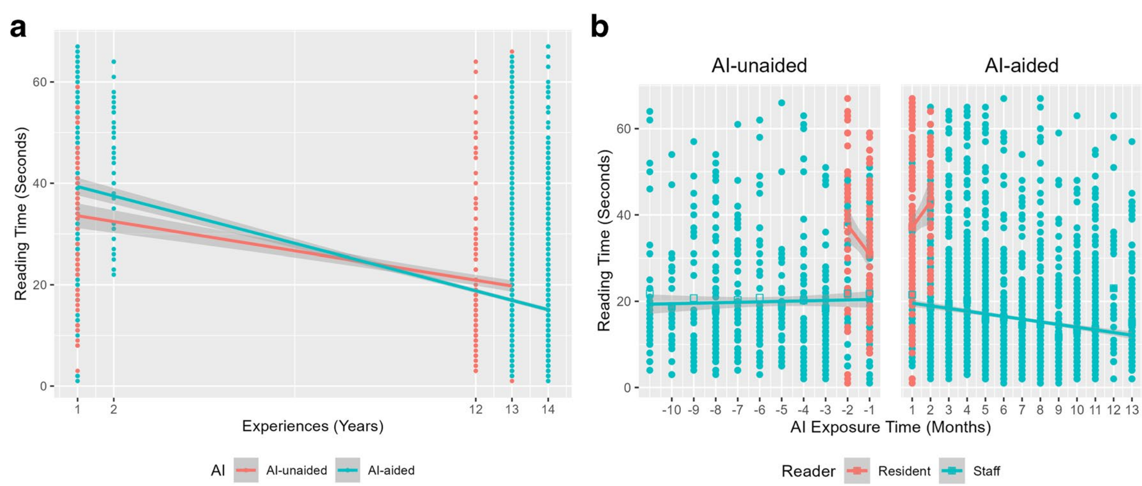

Fig. 2 Scatter plots with linear regression lines and shaded confidence intervals illustrating the impact of radiologists' experience and AI exposure time on reading times. **a** Scatter plot displaying reading times (in seconds) as a function of years of experience in radiology, with separate regression lines for the AI-aided (blue) and AI-unaided (red) periods. Residents (1–2 years of experience) exhibit higher reading times compared to staff radiologists (12–14 years of experience), but there is a gradual decrease in overall reading time as years of experience increase for both AI-aided and AI-unaided periods. The gap between AI-aided and AI-unaided periods is more noticeable in residents, but diminishes as experience grows. **(b)** Scatter plot comparing reading times (in seconds) as a function of AI exposure time (in months) for residents (red) and staff (blue) radiologists during the AI-aided and AI-unaided periods. The left panel now displays the AI-unaided period with negative AI exposure times (e.g., -1, -2 months) to represent the months prior to AI implementation, using the AI implementation date as a reference point, while the right panel shows the AI-aided period with positive AI exposure times. Staff radiologist exhibits a gradual decrease in reading times after AI implementation, with increasing AI exposure, in contrast to the flat, horizontal trend observed before AI was introduced. In comparison, residents show more variability, with some initial increases in reading times during the early stages of AI exposure

In the context of BA radiographs, one prospective randomized controlled trial demonstrated that AI significantly reduced interpretation times [23]. However, this study did not assess the impact of AI on less experienced radiologists [23]. Another study involving a pediatric radiology fellow and a second-year resident showed a consistent trend of reduced reading time with AI assistance, by 18% and 40%, respectively [17]. This contrasts with our findings, which might be due to differences in the interpretation environments. The previous study was conducted in a controlled research setting, whereas our study took place in a real clinical environment where radiologists had access to the GP atlas, medical records, and previous images. This actual workflow likely influence the observed reading times, especially for residents, who may have revisited their interpretations due to their limited experience.

In terms of how AI results are presented, the AI-generated data is attached directly to the PACS, displayed as the next image following the original DICOM file, in our hospital. The AI-determined BA and the corresponding heatmap location are displayed in an intuitive format, identical to what is presented in Fig. 1. This user-friendly format allows radiologists to easily access the AI output without additional UI elements. Given the intuitive nature of these results, we believe that even residents could interpret them without difficulty, and the interpretation method did not significantly affect the reading time in this study. In addition, regarding the education level with this AI software, residents received sufficient training on interpreting BA images and were capable of independent assessments during their pediatric radiology rotation. Certain images with bone abnormalities or skeletal dysplasia may have required additional review. We minimized the impact of such outliers by using the IQR method to exclude extended reading times. It should also be noted that this AI software was new to both pediatric radiologist and residents alike.

The varying number of interpreted images among readers could have influenced their exposure to the AI software during the study period. In addition, differences in AI exposure time arose from the clinical settings. The duty time for interpreting BA images differed between the dedicated staff radiologists and rotational residents, making these variations inevitable in this retrospective observational study, which reflects the actual clinical practices of radiologists. Although our findings for experienced radiologist are consistent with previous research, the trend for less experienced radiologists suggests that AI may increase reading times in real clinical settings. This contrasts with some earlier studies and highlights the importance of considering user-dependent variability in AI's impact on clinical workflows. In mammography, for instance, Lee et al. found that AI assistance increased reading times for general radiologists [28]. The authors proposed that this could be due to increased attention to suspicious findings flagged by the AI [28]. However, in our study, the AI used for BA interpretation did not suggest positive findings or simply provided predicted BA ranges with saliency maps. In our study, residents interpreting images in real clinical settings, where their readings were used for clinical decisions by physicians, might have felt increased pressure to provide precise interpretations compared to a research environment. Less experienced radiologists might have struggled to incorporate the predicted BA range from AI into their interpretations efficiently, prompting them to revisit the GP atlas or medical records for confirmation during the reading process. This additional step in the workflow of residents could contribute to increased reading times compared to experienced radiologist. Not only the radiologists' experience level but also AI exposure time was found to have an impact. For staff radiologists, in contrast to the horizontal trend in reading time before AI implementation, a gradual decrease in reading time was observed as AI exposure increased. While it is difficult to separate the effects of these two factors, our study highlights that continuous AI exposure may be necessary to fully realize the benefits of AI, and this could be recommended to maximize its effectiveness.

This study had several limitations. Firstly, this was a single-center retrospective study with a limited number of readers. Due to the nature of the hospital and study design, we could not include a broad range of readers, with varying experience levels. We initially expected that other radiologists would also read the BA images during the study period. However, the only radiologists who read the images were a dedicated pediatric radiologist and two residents who were on duty during the study period after data extraction. No other board-certified radiologists or experts interpreted the BA images during this period. In addition, because of the retrospective design, most of the radiographs were interpreted by dedicated pediatric radiologists, while residents only read images during specific periods when they were in the pediatric radiology division as part of their training rotation schedule. This led to an uneven distribution of interpreted images and AI exposure time and this variation made it difficult to observe a sharp decrease in reading time after AI implementation. This could introduce bias into the results, but it was inevitable in this retrospective preliminary study design and impossible to implement case matching for this study. However, we aimed to validate it through real-world practice data without any intervention. We analyzed the results that naturally occurred during our actual practice, both before and after implementing AI in the context of BA assessment. Pediatric radiology specialists are rare in any country, and it is common practice for residents to interpret pediatric imaging only during a few designated months of their rotation each year. This study, based on data from a limited number of readers, particularly residents with restricted months of exposure, may have some limitations. However, we believe this approach holds unique value as it reflects the real-world conditions of a radiology department without simulating or intervening in the environment. The post-deployment influence of AI is a significant concern today, and we felt that our hospital's experience was worth demonstrating, particularly its actual impact in reading environment. The differing trends in reading time with AI between residents and experienced radiologist could potentially be attributed to the residents' relatively limited experience in BA interpretation and in utilizing AI software. Further large-scale research with a prospective multi-center study design, involving more pediatric radiologists and residents at various stages of training, is needed to build upon these findings. Second, we only utilized one commercialized AI software, raising concerns about generalizability. However, this AI software was proven representative through previous research, and we could only use one software in a clinical setting [17, 21]. In addition, the UI used to present the AI results could also affect the reading time. Using different software or UI methods might yield different results; however, we believe that our findings reflect the actual implications of AI application in real-world clinical practice. Lastly, we did not include whether AI influenced diagnostic accuracy in this study because it was outside the scope of our study's purpose. Further prospective multicenter studies, involving participants with varying levels of experience and a larger, more balanced number of cases with evenly matched readers and various type of AI software, are needed to overcome these limitations and provide a more comprehensive understanding of AI's role in BA interpretation.

Conclusion

Our study showed that while the overall reading time for BA radiographs was reduced with AI in daily practice, the effect of AI varied based on the radiologists' experience levels and AI exposure time. AI's time-saving effect was more pronounced for experienced radiologists and as AI exposure time increased, whereas less experienced residents exhibited increased reading times, highlighting the need for further research.

Acknowledgements The authors would like to thank Jun Tae Kim for his dedicated help in our research.

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Hyun Joo Shin and Kyunghwa Han. The first draft of the manuscript was written by Sejin Jeong and Hyun Joo Shin. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This research was supported by grants from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: RS-2023-00269924). The funder played no role in the study design; collection, analysis, or interpretation of data; or in the writing of this manuscript.

Data Availability The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics Approval This retrospective study was approved by the Institutional Review Board of our institution. Informed consent was waived due to its retrospective nature. This study was conducted in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines and regulations.

Competing Interests There is no conflict of interest to declare.

References

1. Creo AL, Schwenk WF, 2nd: Bone Age: A Handy Tool for Pediatric Providers. Pediatrics 140(6), 2017. https://doi.org/10.1542/peds.2017-1486
2. Breen MA, Tsai A, Stamm A, Kleinman PK: Bone age assessment practices in infants and older children among Society for Pediatric Radiology members. Pediatr Radiol 46(9):1269–1274, 2016. https://doi.org/10.1007/s00247-016-3618-7
3. Greulich WW, Pyle SI: Radiographic Atlas of Skeletal Development of the Hand and Wrist. Stanford University Press, 1959
4. Prokop-Piotrkowska M, Marszałek-Dziuba K, Moszczyńska E, Szalecki M, Jurkiewicz E: Traditional and New Methods of Bone Age Assessment-An Overview. J Clin Res Pediatr Endocrinol 13(3):251–262, 2021. https://doi.org/10.4274/jcrpe.galenos.2020.2020.0091
5. Thodberg HH, Thodberg B, Ahlkvist J, Offiah AC: Autonomous artificial intelligence in pediatric radiology: the use and perception of BoneXpert for bone age assessment. Pediatr Radiol 52(7):1338–1346, 2022. https://doi.org/10.1007/s00247-022-05295-w
6. Lee H, Tajmir S, Lee J, Zissen M, Yeshiwas BA, Alkasab TK, Choy G, Do S: Fully Automated Deep Learning System for Bone Age Assessment. J Digit Imaging 30(4):427–441, 2017.https://doi.org/10.1007/s10278-017-9955-8
7. Choy G, Khalilzadeh O, Michalski M, Do S, Samir AE, Pianykh OS, Geis JR, Pandharipande PV, Brink JA, Dreyer KJ: Current Applications and Future Impact of Machine Learning in Radiology. Radiology 288(2):318–328, 2018. https://doi.org/10.1148/radiol.2018171820
8. Halabi SS, Prevedello LM, Kalpathy-Cramer J, Mamonov AB, Bilbily A, Cicero M, Pan I, Pereira LA, Sousa RT, Abdala N, Kitamura FC, Thodberg HH, Chen L, Shih G, Andriole K, Kohli MD, Erickson BJ, Flanders AE: The RSNA Pediatric Bone Age Machine Learning Challenge. Radiology 290(2):498–503, 2019. https://doi.org/10.1148/radiol.2018180736
9. Siegel EL: What Can We Learn from the RSNA Pediatric Bone Age Machine Learning Challenge? Radiology 290(2):504–505, 2019. https://doi.org/10.1148/radiol.2018182657
10. Offiah AC: Current and emerging artificial intelligence applications for pediatric musculoskeletal radiology.

Pediatr Radiol 52(11):2149–2158, 2022.https://doi.org/10.1007/s00247-021-05130-8
11. Rajpurkar P, Lungren MP: The Current and Future State of AI Interpretation of Medical Images. N Engl J Med 388(21):1981–1990, 2023. https://doi.org/10.1056/NEJMra2301725
12. Van Leeuwen KG, de Rooij M, Schalekamp S, van Ginneken B, Rutten M: How does artificial intelligence in radiology improve efficiency and health outcomes? Pediatr Radiol, 2021.https://doi.org/10.1007/s00247-021-05114-8
13. Nam JG, Park S, Hwang EJ, Lee JH, Jin KN, Lim KY, Vu TH, Sohn JH, Hwang S, Goo JM, Park CM: Development and Validation of Deep Learning-based Automatic Detection Algorithm for Malignant Pulmonary Nodules on Chest Radiographs. Radiology 290(1):218–228, 2019. https://doi.org/10.1148/radiol.2018180237
14. Lee S, Shin HJ, Kim S, Kim EK: Successful Implementation of an Artificial Intelligence-Based Computer-Aided Detection System for Chest Radiography in Daily Clinical Practice. Korean J Radiol, 2022. https://doi.org/10.3348/kjr.2022.0193
15. Shin HJ, Han K, Ryu L, Kim EK: The impact of artificial intelligence on the reading times of radiologists for chest radiographs. NPJ Digit Med 6(1):82, 2023.https://doi.org/10.1038/s41746-023-00829-4
16. Booz C, Yel I, Wichmann JL, Boettger S, Al Kamali A, Albrecht MH, Martin SS, Lenga L, Huizinga NA, D'Angelo T, Cavallaro M, Vogl TJ, Bodelle B: Artificial intelligence in bone age assessment: accuracy and efficiency of a novel fully automated algorithm compared to the Greulich-Pyle method. Eur Radiol Exp 4(1):6, 2020. https://doi.org/10.1186/s41747-019-0139-9
17. Kim JR, Shim WH, Yoon HM, Hong SH, Lee JS, Cho YA, Kim S: Computerized Bone Age Estimation Using Deep Learning Based Program: Evaluation of the Accuracy and Efficiency. AJR Am J Roentgenol 209(6):1374–1380, 2017. https://doi.org/10.2214/ajr.17.18224
18. Zhao K, Ma S, Sun Z, Liu X, Zhu Y, Xu Y, Wang X: Effect of AI-assisted software on inter- and intra-observer variability for the X-ray bone age assessment of preschool children. BMC Pediatr 22(1):644, 2022. https://doi.org/10.1186/s12887-022-03727-y
19. Alexander R, Waite S, Bruno MA, Krupinski EA, Berlin L, Macknik S, Martinez-Conde S: Mandating Limits on Workload, Duty, and Speed in Radiology. Radiology 304(2):274–282, 2022. https://doi.org/10.1148/radiol.212631
20. Kwee TC, Kwee RM: Workload of diagnostic radiologists in the foreseeable future based on recent scientific advances: growth expectations and role of artificial intelligence. Insights Imaging 12(1):88, 2021. https://doi.org/10.1186/s13244-021-01031-4
21. Lee BD, Lee MS: Automated Bone Age Assessment Using Artificial Intelligence: The Future of Bone Age Assessment. Korean J Radiol 22(5):792–800, 2021.https://doi.org/10.3348/kjr.2020.0941
22. Hwang J, Yoon HM, Hwang JY, Kim PH, Bak B, Bae BU, Sung J, Kim HJ, Jung AY, Cho YA, Lee JS: 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(7):683–691, 2022. https://doi.org/10.3349/ymj.2022.63.7.683
23. Eng DK, Khandwala NB, Long J, Fefferman NR, Lala SV, Strubel NA, Milla SS, Filice RW, Sharp SE, Towbin AJ, Francavilla ML, Kaplan SL, Ecklund K, Prabhu SP, Dillon BJ, Everist BM, Anton CG, Bittman ME, Dennis R, Larson DB, Seekins JM, Silva CT, Zandieh AR, Langlotz CP, Lungren MP, Halabi SS: Artificial Intelligence Algorithm Improves Radiologist Performance in Skeletal Age Assessment: A Prospective Multicenter Randomized Controlled Trial. Radiology 301(3):692–699, 2021. https://doi.org/10.1148/radiol.2021204021
24. Bailey CR, Bailey AM, McKenney AS, Weiss CR: Understanding and Appreciating Burnout in Radiologists. Radiographics 42(5):E137-E139, 2022. https://doi.org/10.1148/rg.220037
25. Bennani S, Regnard NE, Ventre J, Lassalle L, Nguyen T, Ducarouge A, Dargent L, Guillo E, Gouhier E, Zaimi SH, Canniff E, Malandrin C, Khafagy P, Koulakian H, Revel MP, Chassagnon G: Using AI to Improve Radiologist Performance in Detection of Abnormalities on Chest Radiographs. Radiology 309(3):e230860, 2023. https://doi.org/10.1148/radiol.230860
26. Sung J, Park S, Lee SM, Bae W, Park B, Jung E, Seo JB, Jung KH: Added Value of Deep Learning-based Detection System for Multiple Major Findings on Chest Radiographs: A Randomized Crossover Study. Radiology 299(2):450–459, 2021.https://doi.org/10.1148/radiol.2021202818
27. Kim C, Yang Z, Park SH, Hwang SH, Oh YW, Kang EY, Yong HS: Multicentre external validation of a commercial artificial intelligence software to analyse chest radiographs in health screening environments with low disease prevalence. Eur Radiol 33(5):3501–3509, 2023.https://doi.org/10.1007/s00330-022-09315-z
28. Lee JH, Kim KH, Lee EH, Ahn JS, Ryu JK, Park YM, Shin GW, Kim YJ, Choi HY: Improving the Performance of Radiologists Using Artificial Intelligence-Based Detection Support Software for Mammography: A Multi-Reader Study. Korean J Radiol 23(5):505–516, 2022.https://doi.org/10.3348/kjr.2021.0476