scientific reports

Check for updates

OPEN

Optimizing adult-oriented artificial intelligence for pediatric chest radiographs by adjusting operating points

Hyun Joo Shin[1], Kyunghwa Han[2], Nak-Hoon Son[3], Eun-Kyung Kim[1], Min Jung Kim[4], Sergios Gatidis[5] & Shreyas Vasanawala[5]

The purpose of this study was to evaluate whether the optimal operating points of adult-oriented artificial intelligence (AI) software differ for pediatric chest radiographs and to assess its diagnostic performance. Chest radiographs from patients under 19 years old, collected between March and November 2021, were divided into test and exploring sets. A commercial adult-oriented AI software was utilized to detect lung lesions, including pneumothorax, consolidation, nodule, and pleural effusion, using a standard operating point of 15%. A pediatric radiologist reviewed the radiographs to establish ground truth for lesion presence. To determine the optimal operating points, receiver operating characteristic (ROC) curve analysis was conducted, varying thresholds to balance sensitivity and specificity by lesion type, age group, and imaging method. The test set (4,727 chest radiographs, mean 7.2 ± 6.1 years) and exploring set (2,630 radiographs, mean 5.9 ± 6.0 years) yielded optimal operating points of 11% for pneumothorax, 14% for consolidation, 15% for nodules, and 6% for pleural effusion. Using a 3% operating point improved pneumothorax sensitivity for children under 2 years, portable radiographs, and anteroposterior projections. Therefore, optimizing operating points of AI based on lesion type, age, and imaging method could improve diagnostic performance for pediatric chest radiographs, building on adult-oriented AI as a foundation.

Keywords Child, Artificial intelligence, ROC curve, Radiologists, Pneumothorax

Abbreviations

AI	Artificial intelligence
ROC	Receiver operating characteristics
AUC	Area-under-the-curve
PA	Posteroanterior
AP	Anteroposterior

The American College of Radiology (ACR) Pediatric Artificial Intelligence (AI) Workgroup recently highlighted health equity issues, emphasizing the lack of pediatric-specific AI products, uneven development in AI research, and limited physician and industry engagement[1,2]. Several challenges contribute to this disparity, such as difficulties in collecting high quality datasets that reflect growth patterns, differences in pediatric and adult disease entities, limited access to diverse advanced imaging to establish gold standards, and regulatory hurdles unique to pediatric applications[1,3,4]. As a result, fewer dedicated research efforts, financial imbalances in industry involvement, and limited pediatric-specific software have hindered progress in pediatric AI, leading to skewed data and experiences in pediatric radiology applications[1,2].

[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, Yongin-si 16995, Gyeonggi-do, Republic of Korea. [2]Department of Radiology, Research Institute of Radiological Science, Severance Hospital, Yonsei University College of Medicine, 50 – 1 Yonsei-Ro, Seodaemun-Gu, Seoul 03722, Republic of Korea. [3]Department of Statistics, Keimyung University, 1095 Dalgubeol-daero, Dalseo-gu, Daegu 42601, Republic of Korea. [4]Department of Pediatrics, Yongin Severance Hospital, Yonsei University College of Medicine, 363, Dongbaekjukjeon-daero, Giheung-gu, Yongin-si 16995, Gyeonggi-do, Republic of Korea. [5]Department of Radiology, Stanford University, Lucile Packard Children's Hospital, 725 Welch Road, Palo Alto, CA 94304, USA. email: vasanawala@stanford.edu

However, AI has demonstrated significant benefits in radiology, and these advantages should extend beyond adult imaging to include pediatrics[2,5–10]. Several efforts have been made to implement AI in pediatric diseases, focusing on applications such as bone age assessment and pediatric emergency conditions, similar to those in adults, underscoring its importance[11–13]. Additionally, a few researchers have explored methods to enhance pediatric AI by leveraging foundations built in adult radiology[3,14,15]. This has fostered a growing consensus that adapting adult algorithms for pediatric populations, combined with careful validation, could offer a viable solution. However, such adaptations must address safety concerns, as these algorithms were not initially trained on pediatric images and require further enhancement[16].

For chest radiographs, there have been several suggestions to adapt adult AI software for pediatric use. One approach involves specifying age groups, imaging methods, or disease entities that require tailored validation or training for children[14,15]. Specifically, concerns have been raised that the operating points of adult AI may not be suitable for children and need to be validated to determine whether effective operating points exist based on the patient characteristics and imaging techniques[14,17]. Operating points are thresholds used by AI algorithms to classify lesions as positive or negative based on abnormality scores, which indicate the likelihood of a lesion. If the abnormality score exceeds the present threshold, the image is classified as positive. This threshold influences the balance between sensitivity and specificity. Modifying operating points based on the specific characteristics of pediatric patients can enhance the performance of adult-oriented algorithms for children.

This study assumes that customizing AI operating points can improve diagnostic performance in pediatric populations, addressing concerns that thresholds optimized for adults may not suit younger patients. Therefore, the purpose of this study was to evaluate whether the optimal operating points of adult-oriented AI software differ for pediatric chest radiographs and to assess its diagnostic performance by lesion type, age, and imaging methods to assess its potential as a solution for pediatric application of AI.

Materials and methods

Patients

The Institutional Review Board (IRB) of Yongin Severance Hospital approved this retrospective study (IRB number 9-2023-0072, Yongin Severance Hospital, Yonsei University College of Medicine). Informed consent was waived due to the study's retrospective nature. All methods were conducted in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines and relevant regulations. We specifically adhered to the STROBE guidelines for cohort studies, as our research followed a retrospective observational study design. All methods were performed in accordance with relevant guidelines and regulations.

All consecutive chest radiographs of patients under 19 years old, performed at our hospital from March to November 2021, were included in this study. The images were obtained using Innovision-EXII (DK medical systems, Seoul, Korea), GXR-82SD (DRGEM, Seoul, Korea) or ELMO-MX8 (Shimadzu, Kyoto, Japan). This study was conducted at a new university hospital established in 2020. The hospital serves a diverse pediatric population, providing care for a wide range of conditions across multiple specialties. Although the hospital's pediatric case volume is still growing, the disease spectrum includes a broad range of conditions, from mild illnesses to critically ill children and neonates requiring intensive care. The data gathered for this study represent a diverse sample of pediatric patients treated during this period. If patients had multiple repeated radiographs, each image was included in the analysis. Radiographs with artifacts caused by imaging errors, such as caregivers' hands or unnecessary external objects like necklaces or hairpins overlapping the thoracic region, were excluded. However, images containing medical lines and tubes necessary for patient care were included. Patient age, whether the radiograph was portable, and whether it was in posteroanterior (PA) or anteroposterior (AP) projection view were reviewed.

AI analysis of chest radiographs

A commercial AI software (Lunit INSIGHT for Chest Radiography, version 3.1.2, Lunit Inc, Republic of Korea) developed and approved for adult chest radiographs was used to analyze the pediatric chest radiographs. The images were sent to the AI server within the hospital to extract results without being transmitted outside, as the software operates exclusively within the hospital's system. This software detects lung lesions when the operating point is set above 15% and displays the lesion's location with its corresponding abbreviation, regardless of the amount of the lesions (Fig. 1).

The operating point represents the probability that the AI determines a lesion is present, with values ranging from 0 to 100%. The 15% cutoff value, based on the vendor's guidelines, has been validated in previous studies[18–20]. This threshold is used by the AI to classify whether a lesion is present. It is typically applied uniformly across different imaging conditions and lesion types in adults without adjustment.

We retrospectively evaluated the operating points for pneumothorax, consolidation, nodule, and pleural effusion in pediatric chest radiographs. These were all descriptive terms used for chest radiographs. For example, consolidation was used separately from atelectasis when the lung tissue was filled with fluid, pus, or other materials, causing it to appear opaque on imaging. This could result from various causes, including pneumonia, but was not limited to it. Since the AI detects lesion types based on descriptive rather than diagnostic terms (e.g., pneumonia), images were evaluated accordingly.

Setting ground truth and new optimal operating points

The ground truth for each radiograph was established by a board-certified pediatric radiologist (H.J.S.) with 12 years of experience in pediatric radiology. All radiographs were retrospectively reviewed by the radiologist specifically for research purposes, focusing on determining the presence or absence of each lesion type to ensure consistency and accuracy in line with the study's objectives. The radiologist classified all radiographs to establish

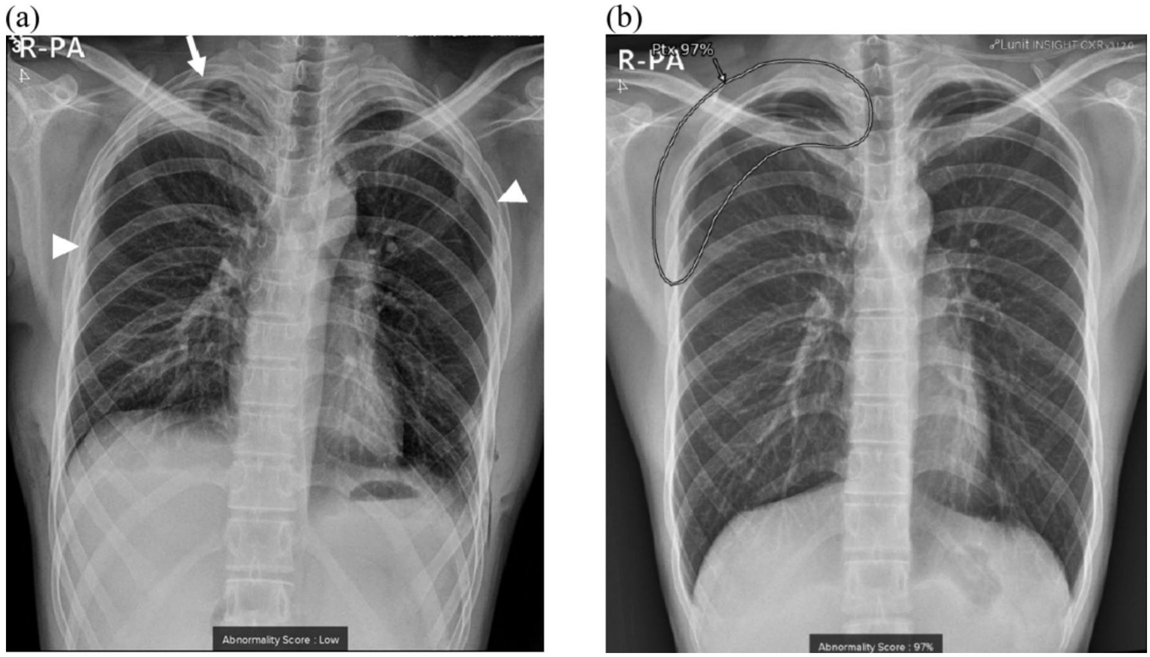

Fig. 1. Actual example of AI results for pneumothorax and pleural effusion. (**a**) A 17-year-old boy presented with a right pneumothorax (arrow) and a small amount of bilateral pleural effusion (arrowheads). However, the AI software did not correctly detect these findings because the operating points for pneumothorax and pleural effusion were 11% and 6%, respectively. (**b**) However, AI accurately identified the right pneumothorax (abbreviated as Ptx) with an operating point of 97% on his initial radiograph.

the ground truth, which was necessary for assessing diagnostic performance in both the test and exploration sets. The radiologist had access to relevant old and new images, medical history, and electronic medical records when assessing lesions.

Using this assessment as the ground truth for the four lesions, new operating points for each lesion and their diagnostic performance were evaluated using a test set of chest radiographs from March to August 2021. To determine the optimal operating point, Receiver Operating Characteristic (ROC) curves were generated for each lesion by plotting sensitivity against 1-specificity, and the Area Under the Curve (AUC) was calculated to assess the overall diagnostic performance. The optimal operating point was defined as the threshold that maximizes Youden's Index (sensitivity + specificity − 1), providing the best balance between sensitivity and specificity. This process involved varying the thresholds of AI-generated probabilities and comparing them with the radiologist's ground truth assessments to identify the threshold that maximizes diagnostic performance.

Testing of operating points

After assessing the operating points for all chest radiographs, subgroup analysis was performed based on age (≤ 2 years or >2 years), projection view (PA or AP view), and whether the radiographs were portable, because actual diagnostic performance varied not only by lesion type but also by imaging method and patient age, as demonstrated in previous studies[14,17]. In a previous study, when adult AI was applied to pediatric chest radiographs across ages 0–18 years, the diagnostic performance significantly declined in children under 2 years of age[14]. Additionally, another study showed that AI performance differed between AP and PA views[17]. The findings emphasized the need for additional adjustments based on these specific conditions. Subgroup analysis also allowed fine-tune of the operating points by identifying variations in diagnostic performance across patient characteristics and imaging methods. When a radiograph belonged to multiple subgroups, it was included in each relevant subgroup analysis independently, and the results were aggregated to determine the most consistent and robust operating point across all subgroups.

After identifying the best operating points for each subgroup, these points were applied to an independent set of radiographs (exploring set) from September to November 2021 to validate diagnostic performance. This two-step approach, involving both the test and exploring sets, ensures that the modified operating points are applicable across clinical settings and suitable for external validation.

Statistical analysis

The R program (version 4.1.3; Foundation for Statistical Computing, Vienna, Austria) was used for statistical analysis. Demographic characteristics of the chest radiographs were compared between the test and exploring sets using two-sample t-tests and Chi-square tests. ROC curve analysis was performed to determine the optimal operating points for pneumothorax, consolidation, nodule, and pleural effusion in all chest radiographs and across subgroup analyses. Diagnostic performance, including sensitivity and specificity, was evaluated in the test set based on the new operating points. For comparison, diagnostic performance using the conventional 15% operating point, commonly applied in adults, was also assessed. Finally, in the time-independent exploring set, diagnostic performance was evaluated using both the new operating points and the conventional 15% threshold. A p-value less than 0.05 was considered statistically significant.

Results

Subjects

During the study period, a total of 7,361 consecutive chest radiographs were initially included. Four radiographs were excluded due to artifacts, leaving 7,357 radiographs for the final analysis. The test set comprised 4,727 chest radiographs (mean 7.2 ± 6.1 years old, M:F = 2602:2125), with varying number of false positives (FP) and false negatives (FN) relative to the 15% operating point based on the ground truth. In the test set, consolidation had 88 FP and 5 FN, nodules had 133 FP, pleural effusion had 23 FP and 1 FN, and pneumothorax had 15 FP and 2 FN. In the exploring set, a total of 2,630 chest radiographs (mean 5.9 ± 6.0 years old, M:F = 1396:1234) were included. For lesion types, consolidation had 35 FP and 1 FN, nodules had 74 FP, pleural effusion had 7 FP, and pneumothorax had 4 FP and 2 FN.

When divided by age, among the test set, 766 patients (16.2%) were neonates and infants (< 1 years old), 1093 patients (23.1%) were toddlers (< 3 years old), 1634 patients (34.6%) were children (< 12 years old), and 1234 patients (26.1%) were adolescents (< 18 years old). In the exploring set, 587 patients (22.3%) were neonates and infants, 764 patients (29%) were toddlers, 758 patients (28.8%) were children, and 521 patients (19.8%) were adolescents. Among the test set, 3060 patients (64.7%) were inpatients, while 1841 out of 2630 patients (70%) in the exploring set were inpatients. Demographic details of the included chest radiographs for each dataset are presented in Table 1, and a confusion matrix showing the FP and FN results is provided in Supplementary file 1.

Diagnostic performance of AI when applying adults' operating point of 15% in the test set

When applying the conventional adult operating points of 15% in the test set, the overall AUC values were 0.996 for pneumothorax, 0.973 for consolidation, 0.985 for nodules, and 0.996 for pleural effusion. However, in the subgroup analysis of pneumothorax, the sensitivity was 0.5 for children younger than 2 years old and 0.8 for portable radiographs and AP projection views, while other categories had sensitivities over 0.94.

For consolidation and pleural effusion, sensitivities were approximately 0.7–0.8 for nonportable radiographs or radiographs in PA projection views. Other results, including nodules, showed overall sensitivities of about 0.9-1.0 using the 15% operating point. The detailed results are summarized in Table 2.

Identifying the best operating points in the test set

When evaluating new operating points with the best diagnostic performances in the test set, the optimal thresholds were determined as 11% for pneumothorax, 14% for consolidation, 15% for nodules, and 6% for pleural effusion. Despite variations in overall AUC values during subgroup analysis, sensitivities for pneumothorax reached 1.0 in children younger than 2 years old, portable radiographs, and AP projection views when using an operating point of 3%. For consolidation and pleural effusion, applying operating points of 2–3% led to sensitivities exceeding 90% for nonportable radiographs or radiographs in PA projection views. The operating point for nodules remained consistent at about 15% across subgroup analyses. Detailed results are presented in Table 3.

Demonstrating diagnostic performance of new operating points in the exploring set

The diagnostic performances of conventional and new operating points in the exploring set are presented in Tables 4 and 5. Compared to the conventional operating points, the new thresholds demonstrated improved sensitivity for pneumothorax in younger children (Fig. 2), portable radiographs, and AP projection views. No significant differences were observed in the diagnostic performances for other results.

Discussion

In this study, we aimed to evaluate the diagnostic performance of adult AI software by applying conventional and new optimal operating points to pediatric chest radiographs. We also sought to determine whether the best operating points differed based on age, image acquisition methods, and lesion types, and to explore whether this approach could enhance the pediatric application of AI from developed adult software. Despite having a large cohort spanning nine months from a single hospital, diagnostic performance and operating points varied by lesion type, age, and image acquisition method. The overall best operating points were lower than those used for adults, particularly for pneumothorax in children younger than 2 years, portable radiographs, and in AP projection views. Although the AUC values were not strikingly different, sensitivities improved with the new operating points in both the test and exploring sets. However, the optimal operating points for nodules were not significantly different across our dataset.

There is a significant disparity in dedicated research efforts and software development for pediatric AI[1,21]. The ACR noted that only 3% of FDA-cleared AI algorithms are intended for pediatric use[1,2]. Many studies suggest that an effective approach to enhancing pediatric AI utilization is by leveraging adult AI technologies for pediatric applications[1,2]. Validating and retraining adult AI models could be one solution, emphasizing the identification and resolution of weaknesses when applied to pediatric images[2]. This is essential because AI developed and approved for adults requires adjustments to perform optimally with pediatric data[1]. Addressing these weaknesses could expedite the adoption of existing AI technologies into pediatric radiology.

In a previous study, the diagnostic performance of adult-oriented AI software on pediatric radiographs was demonstrated, and the authors mentioned that further research on the operating point may be necessary[14]. This is because, in most studies, the operating point was set at 15% for that software and applied uniformly regardless

Sets		Test set					Exploring set										
Variables		n (%)	Age (years)	Sex (M: F)	Portable (n, %)	AP (n, %)	n (%)	p-value	Age (years)	p-value	Sex (M: F)	p-value	Portable (n, %)	p-value	AP (n, %)	p-value
Overall		4,727	7.2 ± 6.1	2602:2125	1601 (33.9)	2138 (45.2)	2,630	.	5.9 ± 6.0	< 0.001	1396:1234	0.105	1112 (42.3)	< 0.001	1432 (54.5)	< 0.001
Pneumothorax	Positive	192 (4.1)	15.7 ± 3.6	173:19	20 (10.4)	21 (10.9)	32 (1.2)	< 0.001	13.8 ± 6.2	0.015	27:5	0.333	6 (18.8)	0.174	6 (18.8)	0.21
	Negative	4535	6.9 ± 5.9	2429:2106	1581 (34.9)	2117 (46.7)	2598		5.8 ± 5.9	< 0.001	1369:1229	0.48	1106 (42.6)	< 0.001	1426 (54.9)	< 0.001
Consolidation	Positive	526 (11.1)	4.4 ± 5.8	310:216	391 (74.3)	447 (85)	235 (8.9)	0.003	2.9 ± 4.6	< 0.001	126:109	0.171	158 (67.2)	0.044	204 (86.8)	0.508
	Negative	4201	7.6 ± 6.1	2292:1909	1210 (28.8)	1691 (40.3)	2395		6.2 ± 6.0	< 0.001	1270:1125	0.23	954 (39.8)	< 0.001	1228 (51.3)	< 0.001
Nodule	Positive	132 (2.8)	7.8 ± 6.7	65:67	95 (72)	105 (79.6)	58 (2.2)	0.128	4.0 ± 5.3	< 0.001	38:20	0.039	37 (63.8)	0.261	47 (81)	0.814
	Negative	4595	7.2 ± 6.1	2537:2058	1506 (32.8)	2033 (44.2)	2572		5.9 ± 6.0	< 0.001	1358:1214	0.049	1075 (41.8)	< 0.001	1385 (53.9)	< 0.001
Pleural effusion	Positive	122 (2.6)	12.8 ± 5.8	59:63	73 (59.8)	76 (62.3)	32 (1.2)	< 0.001	6.4 ± 7.2	< 0.001	18:14	0.428	23 (71.9)	0.212	24 (75)	0.182
	Negative	4605	7.1 ± 6.0	2543:2062	1528 (33.2)	2062 (44.8)	2598		5.9 ± 5.9	< 0.001	1378:1220	0.074	1089 (41.9)	< 0.001	1408 (54.2)	< 0.001

Table 1. Demographics of the test and exploring sets. Values are presented in mean ± standard deviation or number (%). *AP* anteroposterior.

Lesions	Pneumothorax			Consolidation			Nodule			Pleural effusion		
Diagnostic performance	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC
All	0.948	0.997	0.996	0.873	0.979	0.973	1	0.97	0.985	0.918	0.995	0.996
≤ 2 years old	0.5	0.998	0.978	0.858	0.96	0.942	1	0.922	0.963	0.947	0.987	0.994
> 2 years old	0.967	0.996	0.997	0.898	0.987	0.987	1	0.994	0.998	0.913	0.999	0.998
No portable	0.965	1	0.999	0.733	0.995	0.973	1	0.99	0.995	0.816	0.998	0.996
Portable	0.8	0.991	0.982	0.921	0.94	0.963	1	0.93	0.963	0.986	0.989	0.999
PA	0.965	1	0.999	0.76	0.999	0.983	1	0.998	0.999	0.826	0.998	0.997
AP	0.81	0.993	0.987	0.893	0.948	0.952	1	0.934	0.966	0.974	0.991	0.998

Table 2. Diagnostic performance of AI using adults' operating point of 15% in the test set. *AUC* area-under-the-curve, *PA* posteroanterior, *AP* anteroposterior.

of lesion type, age, or imaging methods[9,22]. It remains essential to determine whether this uniform application of the operating point is suitable for children, and even for adults. For example, adult AI has shown decreased accuracy when applied to images of children younger than 2 years[14]. Additionally, a recent study demonstrated that the positive predictive value of AI for detecting pneumothorax differed significantly between PA and AP views (88.2% vs. 20.1%, $p < 0.001$) in adults[17]. This highlights the importance of optimizing AI for practical use, considering variations in patients, imaging techniques, lesion types, and hospital settings[17].

In the same context, understanding whether using the best operating points based on age or imaging methods can enhance pediatric AI usage is crucial. Currently, a consistent cutoff value is applied regardless of factors such as age, hospital characteristics, or imaging features. However, we demonstrated that AI performance varies depending on whether the image is AP or PA, and that the optimal cutoff changes based on factors such as age and lesion type. This underscores the need for future efforts to apply tailored algorithm settings for each specific situation to enhance AI accuracy. Based on these results, we suggest that a different operating point, lower than that used for adults, may be needed to optimize sensitivity for detecting pneumothorax in younger patients or when radiographs are portable or in AP projection views. However, for nodules, it appears that setting different operating points from those used for adults may not be necessary.

In this study, we prioritized optimizing sensitivity because the primary goal was to detect as many true positive cases as possible. Missing a positive case could have serious clinical consequences, especially in situations where detection is critical. In pediatric imaging, AI still requires further development and validation. We believe that efforts should first focus on increasing sensitivity and then work to reduce FP. Therefore, in this initial study, we chose sensitivity as our primary metric while assessing overall diagnostic performance through ROC curve analysis. Optimizing for positive predictive value (PPV) would be more appropriate if the focus were on reducing FP, which could lead to unnecessary follow-up tests or treatments. Ultimately, the choice of metric depends on the priorities of each study. In this study, we emphasized enhancing patient safety by minimizing missed diagnoses through the initial use of AI for pediatric chest radiographs. Thus, we chose to prioritize sensitivity over PPV. In addition, the actual incidence of diseases may vary between adults and children. For example, consolidation due to infection may be more prevalent in children, while nodules, such as those caused by cancer, are much less common in pediatric populations. These differences in disease characteristics could impact AI performance based on the training data. Therefore, further research is needed to reduce FP cases, exploring how disease incidence differs between pediatric and adult datasets, and evaluate its impact on AI diagnostic performance.

There are several limitations in this study. First, although we included all consecutive chest radiographs taken during the study period to avoid selection bias, the small number of certain lesions, such as nodules, may have impacted diagnostic accuracy. Additionally, differences in the basic characteristics between the two groups, as well as discrepancies in the number of cases between the test and exploring sets, were observed. These differences were likely due to the incidence of these lesions and the characteristics of the hospital population. While we could have included more diseased cases than normal images to balance the datasets, doing so might have influence the results. Therefore, we chose to include all images despite the discrepancy in diseased cases. The differences in the characteristics of the two groups were inevitable as this was a retrospective study, and our goal was to include all available data during the study period to minimize bias. Second, the determination of ground truth relied on a single pediatric radiologist who retrospectively reviewed all images, which could be a limitation. While a consensus interpretation involving multiple radiologists could have reduced bias and improved the reliability of the ground truth, this approach was not feasible due to the large number of radiographs included in the study and the limited number of pediatric radiologists specialized in this area. To minimize discrepancies and ensure consistency, we focused on achieving uniformity in the results. Additionally, we utilized all relevant medical images and records to establish the ground truth as comprehensively as possible. Future studies may benefit from incorporating consensus interpretation to achieve greater accuracy and consistency. Third, including repeated exams for the same patient could have influenced the results. However, we aimed to include as many cohort studies as possible within a specific time period, and due to the relatively low incidence and variability of lesions over long periods, it was necessary to include repeated exams. Fourth, the use of a single commercial software may limit the generalizability of our findings. Lastly, the variation in operating points, both increases and decreases, affected both sensitivity and the FP rate. Future studies should focus on optimizing both sensitivity and PPV to balance accurate detection with minimizing unnecessary follow-up procedures. Given the

Lesions	Pneumothorax				Consolidation				Nodule				Pleural effusion			
Diagnostic performance	Operating point (%)	Sensitivity	Specificity	AUC	Operating point (%)	Sensitivity	Specificity	AUC	Operating point (%)	Sensitivity	Specificity	AUC	Operating point (%)	Sensitivity	Specificity	AUC
All	11	0.964	0.993	0.996	14	0.882	0.973	0.973	15	1	0.971	0.985	6	0.975	0.977	0.996
≤2 years old	3	1	0.943	0.978	16	0.858	0.961	0.942	15	1	0.925	0.963	8	1	0.966	0.994
> 2 years old	9	0.984	0.992	0.997	12	0.929	0.971	0.987	15	1	0.995	0.998	3	0.99	0.972	0.998
No portable	2	0.994	0.989	0.999	2	0.926	0.9	0.973	15	1	0.99	0.995	3	0.98	0.973	0.996
Portable	3	1	0.903	0.982	16	0.921	0.941	0.963	15	1	0.934	0.963	17	0.986	0.992	0.999
PA	2	0.994	0.989	0.999	2	0.924	0.965	0.983	14	1	0.998	0.999	3	0.978	0.979	0.997
AP	3	1	0.925	0.987	16	0.893	0.949	0.952	15	1	0.938	0.966	8	1	0.969	0.998

Table 3. Identifying the optimal operating points of AI in the test set. *AUC* area-under-the-curve, *PA* posteroanterior, *AP* anteroposterior.

Lesions	Pneumothorax			Consolidation			Nodule			Pleural effusion		
Diagnostic performance	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC
All	0.906	0.998	0.996	0.821	0.986	0.965	1	0.971	0.987	0.918	0.995	0.996
≤ 2 years old	0.4	0.997	0.959	0.888	0.967	0.967	1	0.934	0.971	0.875	0.995	0.962
> 2 years old	1	0.999	1	0.652	1	0.971	1	0.998	0.999	1	1	1
No portable	1	0.999	1	0.805	0.994	0.984	1	0.989	0.995	0.889	1	0.983
Portable	0.5	0.997	0.954	0.829	0.974	0.938	1	0.945	0.974	0.957	0.995	0.996
PA	1	1	1	0.936	1	0.999	1	1	1	0.875	1	0.989
AP	0.5	0.997	0.963	0.804	0.972	0.933	1	0.946	0.973	0.958	0.996	0.997

Table 4. Diagnostic performance of AI using adults' operating point of 15% in the exploring set. *AUC* area-under-the-curve, *PA* posteroanterior, *AP* anteroposterior.

promising results of this study, further research involving a larger number of hospitals, diverse disease severities, and multiple readers is needed to validate and generalize our findings.

In conclusion, the optimal operating points for pediatric chest radiographs using adult-oriented AI software vary based on lesion type, age, and image acquisition methods. Customizing these operating points is necessary to optimize AI performance and applicability in pediatric imaging. Starting with adult AI could serve as an initial approach to enhance its application in pediatric radiographs; however, further validation and retraining are essential to ensure optimal performance and reliability.

Lesions	Pneumothorax				Consolidation				Nodule				Pleural effusion			
Diagnostic performance	Operating point (%)	Sensitivity	Specificity	AUC	Operating point (%)	Sensitivity	Specificity	AUC	Operating point (%)	Sensitivity	Specificity	AUC	Operating point (%)	Sensitivity	Specificity	AUC
All	11	0.938	0.997	0.996	14	0.834	0.98	0.965	15	1	0.971	0.987	6	0.975	0.977	0.996
≤2 years old	3	0.6	0.969	0.959	16	0.888	0.967	0.967	15	0.97	0.94	0.971	8	0.875	0.962	0.962
> 2 years old	9	1	0.997	1	12	0.682	0.997	0.971	15	1	0.998	0.999	3	1	0.985	1
No portable	2	1	0.993	1	2	0.948	0.882	0.984	15	1	0.989	0.995	3	0.889	0.982	0.983
Portable	3	0.667	0.954	0.954	16	0.829	0.974	0.938	15	1	0.946	0.974	17	0.87	0.995	0.996
PA	2	1	0.997	1	2	1	0.956	0.999	14	1	1	1	3	0.875	0.991	0.989
AP	3	0.667	0.961	0.963	16	0.804	0.972	0.933	15	0.979	0.947	0.973	8	0.958	0.965	0.997

Table 5. Diagnostic performance of AI using new operating points in the exploring set. *AUC* area-under-the-curve, *PA* posteroanterior, *AP* anteroposterior.

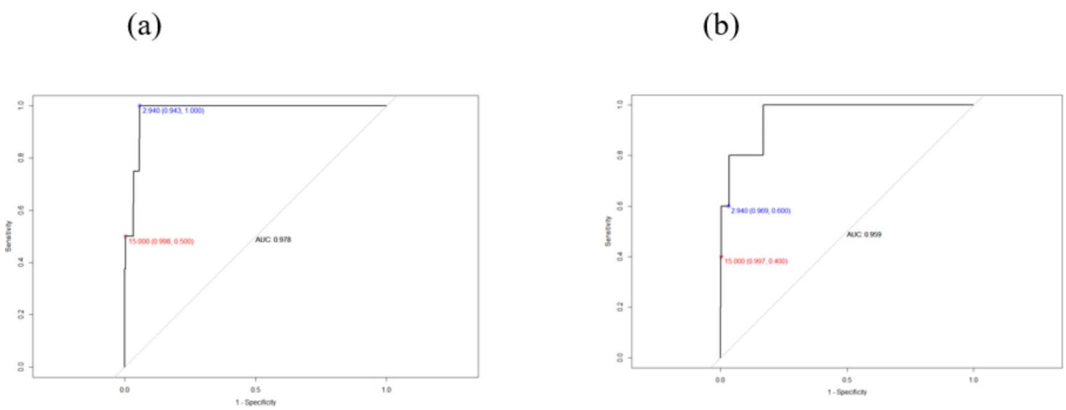

Fig. 2. ROC curves for pneumothorax detection in patients aged 2 years and under in (**a**) test and (**b**) exploring sets. Sensitivities were improved by applying the new optimal operating point (blue) compared to the conventional operating point of 15% (red). The specificity and sensitivity are presented in parentheses.

Data availability

The part of the datasets generated and analyzed during the current study are available in the Supplementary file 2.

Received: 30 August 2024; Accepted: 9 December 2024
Published online: 28 December 2024

References

1. Sammer, M. B. K. et al. Use of Artificial Intelligence in Radiology: Impact on Pediatric patients, a White Paper from the ACR Pediatric AI Workgroup. *J. Am. Coll. Radiology: JACR* **20**, 730–737. https://doi.org/10.1016/j.jacr.2023.06.003 (2023).
2. Tierradentro-Garcia, L. O., Sotardi, S. T., Sammer, M. B. K. & Otero, H. J. Commercially available Artificial Intelligence algorithms of interest to Pediatric Radiology: the growing gap between potential Use and Data Training. *J. Am. Coll. Radiology: JACR* **20**, 748–751. https://doi.org/10.1016/j.jacr.2023.04.017 (2023).
3. Moore, M. M., Iyer, R. S., Sarwani, N. I. & Sze, R. W. Artificial intelligence development in pediatric body magnetic resonance imaging: best ideas to adapt from adults. *Pediatr. Radiol.* **52**, 367–373. https://doi.org/10.1007/s00247-021-05072-1 (2022).
4. Alkhulaifat, D., Rafful, P., Khalkhali, V., Welsh, M. & Sotardi, S. Implications of Pediatric Artificial Intelligence Challenges for Artificial Intelligence Education and Curriculum Development. *J. Am. Coll. Radiology: JACR*. https://doi.org/10.1016/j.jacr.2023.04.013 (2023).
5. van Leeuwen, K. G., de Rooij, M., Schalekamp, S., van Ginneken, B. & Rutten, M. How does artificial intelligence in radiology improve efficiency and health outcomes? *Pediatr. Radiol. (2022)* **52**, 2087–2093. https://doi.org/10.1007/s00247-021-05114-8 (2021).
6. Shelmerdine, S. C., Rosendahl, K. & Arthurs, O. J. Artificial intelligence in paediatric radiology: international survey of health care professionals' opinions. *Pediatr. Radiol.* **52**, 30–41. https://doi.org/10.1007/s00247-021-05195-5 (2022).
7. Shin, H. J., Han, K., Ryu, L. & Kim, E. K. The impact of artificial intelligence on the reading times of radiologists for chest radiographs. *NPJ Digit. Med.* **6**, 82. https://doi.org/10.1038/s41746-023-00829-4 (2023).
8. Hwang, S. H., Shin, H. J., Kim, E. K., Lee, E. H. & Lee, M. Clinical outcomes and actual consequence of lung nodules incidentally detected on chest radiographs by artificial intelligence. *Sci. Rep.* **13**, 19732. https://doi.org/10.1038/s41598-023-47194-6 (2023).
9. Lee, S., Shin, H. J., Kim, S. & Kim, E. K. Successful implementation of an Artificial Intelligence-based computer-aided detection system for chest radiography in Daily Clinical Practice. *Korean J. Radiol.* https://doi.org/10.3348/kjr.2022.0193 (2022).
10. Shin, H. J., Lee, S., Kim, S., Son, N. H. & Kim, E. K. Hospital-wide survey of clinical experience with artificial intelligence applied to daily chest radiographs. *PloS One* **18**, e0282123. https://doi.org/10.1371/journal.pone.0282123 (2023).
11. Kim, S. et al. Performance of deep learning-based algorithm for detection of ileocolic intussusception on abdominal radiographs of young children. *Sci. Rep.* **9**, 19420. https://doi.org/10.1038/s41598-019-55536-6 (2019).
12. Dillman, J. R., Somasundaram, E., Brady, S. L. & He, L. Current and emerging artificial intelligence applications for pediatric abdominal imaging. *Pediatr. Radiol.* https://doi.org/10.1007/s00247-021-05057-0 (2021).
13. Choi, J. W. et al. Deep learning-assisted diagnosis of Pediatric Skull fractures on plain radiographs. *Korean J. Radiol.* **23**, 343–354. https://doi.org/10.3348/kjr.2021.0449 (2022).
14. Shin, H. J., Son, N. H., Kim, M. J. & Kim, E. K. Diagnostic performance of artificial intelligence approved for adults for the interpretation of pediatric chest radiographs. *Sci. Rep.* **12**, 10215. https://doi.org/10.1038/s41598-022-14519-w (2022).
15. Morcos, G., Yi, P. H. & Jeudy, J. Applying Artificial Intelligence to Pediatric chest imaging: reliability of leveraging adult-based Artificial Intelligence models. *J. Am. Coll. Radiology: JACR* **20**, 742–747. https://doi.org/10.1016/j.jacr.2023.07.004 (2023).
16. Ciet, P. et al. The unintended consequences of artificial intelligence in paediatric radiology. *Pediatr. Radiol.* https://doi.org/10.1007/s00247-023-05746-y (2023).
17. Lee, S. et al. Factors for increasing positive predictive value of pneumothorax detection on chest radiographs using artificial intelligence. *Sci. Rep.* **14**, 19624. https://doi.org/10.1038/s41598-024-70780-1 (2024).
18. Nam, J. G. et al. Development and Validation of Deep Learning-based Automatic Detection Algorithm for malignant pulmonary nodules on chest radiographs. *Radiology* **290**, 218–228. https://doi.org/10.1148/radiol.2018180237 (2019).
19. Hwang, E. J. et al. Use of Artificial Intelligence-Based Software as Medical devices for chest radiography: a position paper from the Korean Society of Thoracic Radiology. *Korean J. Radiol.* **22**, 1743–1748. https://doi.org/10.3348/kjr.2021.0544 (2021).
20. Kim, E. Y. et al. Concordance rate of radiologists and a commercialized deep-learning solution for chest X-ray: real-world experience with a multicenter health screening cohort. *PloS One* **17**, e0264383. https://doi.org/10.1371/journal.pone.0264383 (2022).

21. Padash, S., Mohebbian, M. R., Adams, S. J., Henderson, R. D. E. & Babyn, P. Pediatric chest radiograph interpretation: how far has artificial intelligence come? A systematic literature review. *Pediatr. Radiol.* **52**, 1568–1580. https://doi.org/10.1007/s00247-022-05368-w (2022).
22. Jin, K. N. et al. Diagnostic effect of artificial intelligence solution for referable thoracic abnormalities on chest radiography: a multicenter respiratory outpatient diagnostic cohort study. *Eur. Radiol.* https://doi.org/10.1007/s00330-021-08397-5 (2022).

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

Author contributions
Data curation: H.J.S. Formal analysis: H.J.S., K.H., N-H.S. Funding acquisition: H.J.S. Investigation: H.J.S. Methodology: H.J.S., K.H., N-H.S., E-K.K., S.V. Writing-original draft: H.J.S. Writing-review and editing: H.J.S., K.H., N-H.S., E-K.K., M.J.K., S.G., S.V.

Funding
This research was supported by a grant of 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.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-82775-z.

Correspondence and requests for materials should be addressed to S.V.