Korean Journal of Radiology

KJR

Check for updates

Conventional Versus Artificial Intelligence-Assisted Interpretation of Chest Radiographs in Patients With Acute Respiratory Symptoms in Emergency Department: A Pragmatic Randomized Clinical Trial

Eui Jin Hwang[1], Jin Mo Goo[1,2,3], Ju Gang Nam[1], Chang Min Park[1,2,3], Ki Jeong Hong[4], Ki Hong Kim[4]

[1]Department of Radiology, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Korea
[2]Cancer Research Institute, Seoul National University, Seoul, Korea
[3]Institute of Radiation Medicine, Seoul National University Medical Research Center, Seoul, Korea
[4]Department of Emergency Medicine, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Korea

Objective: It is unknown whether artificial intelligence-based computer-aided detection (AI-CAD) can enhance the accuracy of chest radiograph (CR) interpretation in real-world clinical practice. We aimed to compare the accuracy of CR interpretation assisted by AI-CAD to that of conventional interpretation in patients who presented to the emergency department (ED) with acute respiratory symptoms using a pragmatic randomized controlled trial.

Materials and Methods: Patients who underwent CRs for acute respiratory symptoms at the ED of a tertiary referral institution were randomly assigned to intervention group (with assistance from an AI-CAD for CR interpretation) or control group (without AI assistance). Using a commercial AI-CAD system (Lunit INSIGHT CXR, version 2.0.2.0; Lunit Inc.). Other clinical practices were consistent with standard procedures. Sensitivity and false-positive rates of CR interpretation by duty trainee radiologists for identifying acute thoracic diseases were the primary and secondary outcomes, respectively. The reference standards for acute thoracic disease were established based on a review of the patient's medical record at least 30 days after the ED visit.

Results: We randomly assigned 3576 participants to either the intervention group (1761 participants; mean age ± standard deviation, 65 ± 17 years; 978 males; acute thoracic disease in 472 participants) or the control group (1815 participants; 64 ± 17 years; 988 males; acute thoracic disease in 491 participants). The sensitivity (67.2% [317/472] in the intervention group vs. 66.0% [324/491] in the control group; odds ratio, 1.02 [95% confidence interval, 0.70–1.49]; *P* = 0.917) and false-positive rate (19.3% [249/1289] vs. 18.5% [245/1324]; odds ratio, 1.00 [95% confidence interval, 0.79–1.26]; *P* = 0.985) of CR interpretation by duty radiologists were not associated with the use of AI-CAD.

Conclusion: AI-CAD did not improve the sensitivity and false-positive rate of CR interpretation for diagnosing acute thoracic disease in patients with acute respiratory symptoms who presented to the ED.

Keywords: Chest radiography; Artificial intelligence; Deep learning; Computer-aided detection; Emergency radiology; Clinical trial; Diagnostic accuracy

INTRODUCTION

Chest radiograph (CR) is essential for the initial evaluation of acute respiratory diseases [1-4] in the emergency department (ED), where they pose a significant burden [5]. Although the number of CRs in the ED has increased significantly [6], An expert radiologist may find it challenging to interpret CRs promptly [7]. Consequently, a computer-

aided detection (CAD) device that assists physicians in identifying abnormalities in CRs has the potential to enhance the quality and efficiency of ED practice.

Recent studies have reported that deep learning-based artificial intelligence (AI) algorithms can improve physician's interpretation accuracy [8-17]. Eventually, these AI algorithms were implemented as CAD devices in clinical practice [18].

In a previous retrospective study [19], reinterpretation of CRs obtained in the ED using an AI algorithm enhanced the radiologist's sensitivity, indicating the potential of an AI-based CAD (AI-CAD) device. However, the performance and efficacy (i.e., enhancing the accuracy of physicians' interpretation) of AI-CAD have been evaluated primarily in retrospective, experimental settings, which cannot fully replicate the conditions of daily practice [12-17,20]. Although several studies have reported an increase in the accuracy of interpretation following the implementation of AI-CAD [21,22], the results of such "before-and-after" studies may be biased because of a lack of comparability between before and after implementation. Moreover, improving the accuracy of CR interpretation does not necessarily result in improvements in patient management, patient outcomes, or workflow efficiency [23]. Therefore, prospective, parallel, randomized clinical trials are necessary to determine the true impact of AI-CAD [24].

Therefore, this pragmatic, parallel, randomized controlled trial aimed to report the sensitivity and false-positive rate (FPRs) of chest radiograph interpretations with and without assistance from AI-CAD for diagnosing acute thoracic diseases in patients with acute respiratory symptoms who presented to the ED. Additionally, we aimed to investigate the effects of implementing AI-CAD on patient management in the ED.

MATERIALS AND METHODS

Trial Design
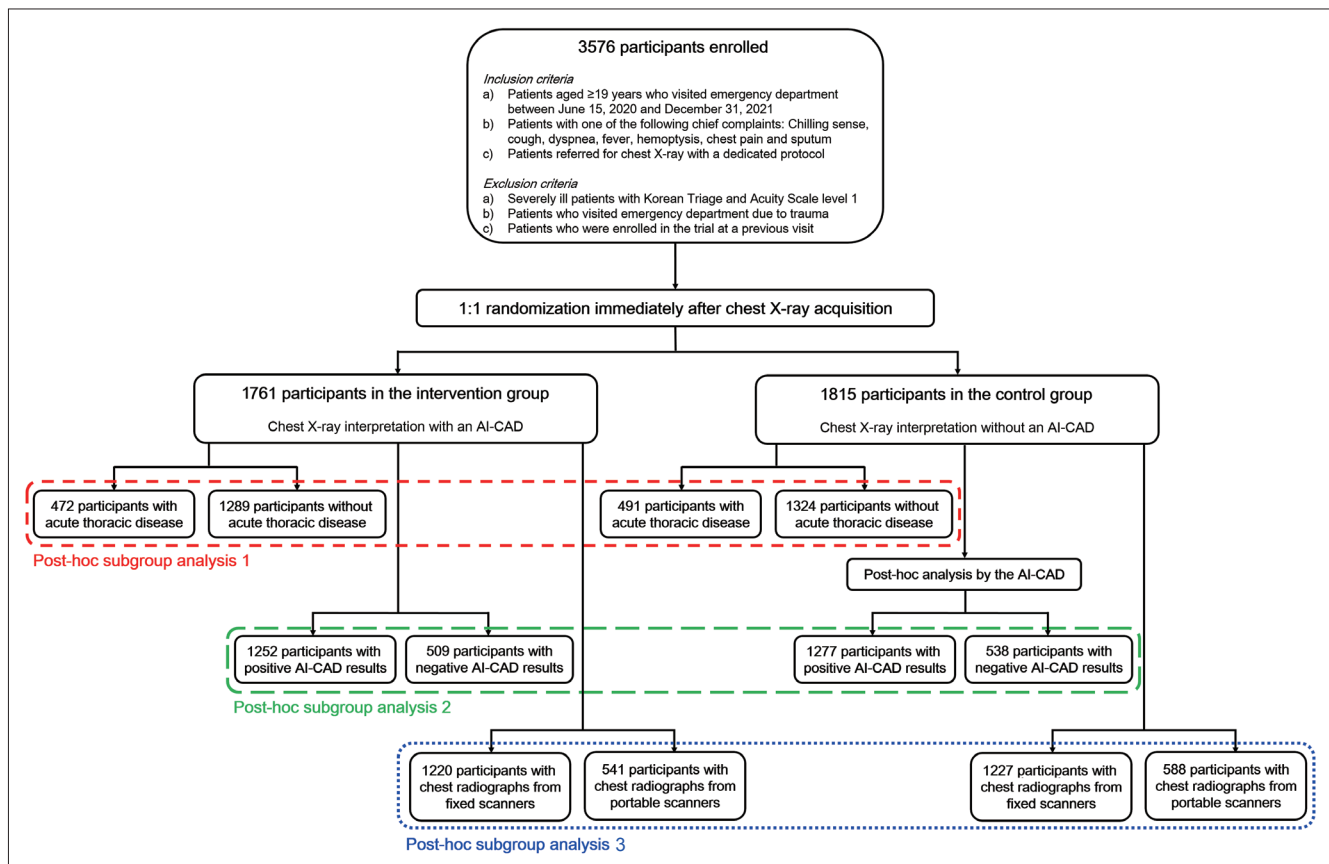We conducted a parallel, open-label trial at an academic

Fig. 1. Flow diagram. A total of 3576 participants who met the eligibility criteria were enrolled in the trial. As a result of random allocation, 1761 and 1815 participants were allocated to the intervention group and control group, respectively. We also conducted three post-hoc subgroup analyses for participants with and without acute thoracic diseases (post-hoc subgroup analysis 1), participants with positive and negative artificial intelligence-based computer-aided detection (AI-CAD) results (post-hoc subgroup anslysis 2), and participants with chest radiographs from fixed and portable scanners (post-hoc subgroup analysis 3).

tertiary referral hospital in South Korea (registered at the Clinical Research Information Service [https://cris.nih. go.kr]; registration number: KCT0005007). The Seoul National University Institutional Review Board (approval number: D-2002-169-1107) approved this trial and waived the participants' required informed consent.

Participants were enrolled according to the following inclusion criteria (Fig. 1): 1) patients aged ≥ 19 years who presented to the ED between June 15, 2020, and December 31, 2021; 2) patients with one of the following chief complaints at the time of the ED presentation: chills, cough, chest pain, dyspnea, fever, hemoptysis, and sputum; and 3) patients referred for CR using a dedicated examination protocol created for the trial. Patients who met the following criteria were excluded from the study: 1) severely ill patients with a Korean Triage and Acuity Scale (KTAS) level 1 [25], 2) patients who visited the ED due to trauma, and 3) patients who were previously enrolled in the trial. The participants were randomly assigned to the intervention and control groups in a ratio of 1:1 (CR interpretation with and without AI-CAD, respectively). The radiologists, ED physicians, and outcome evaluators were aware of the assigned groups.

Trial Implementation

The AI-CAD used in the trial (Lunit INSIGHT CXR, version 2.0.2.0; Lunit Inc.) was approved by the Korean Ministry of Food and Drug Safety of Korea as a tool to assist physicians with CR interpretation. AI-CAD analyzed a single frontal CR to determine the presence of pulmonary nodules, infiltration,

and pneumothorax. The AI-CAD provided a heat map overlaid on the input CR to visualize the location of the identified abnormality, in addition to providing a probability score (0%–100%) for the presence of the identified abnormality (Figs. 2-4) [11-19].

A dedicated examination protocol was developed for the participants' systematic selection and distribution. Before the start of the trial, ED physicians were instructed to request CRs for eligible patients using a specific protocol. Randomization of the examination was automatically performed when CR was obtained and transferred to the institutional picture archiving and communication system (PACS, Infinitt M6, Infinitt Healthcare). PACS generated a five-digit random number for each CR automatically. CRs with random numbers that were odd and even were assigned to the intervention and control groups to conceal the allocation sequence. Subsequently, the PACS automatically processed the AI-CAD analysis for the CRs assigned to the intervention group. Consequently, the original CRs and AI-CAD results were uploaded to the PACS for the intervention group, while only the original CRs were uploaded for the control group. The random assignment and AI-CAD evaluations were completed within a few minutes of image acquisition, allowing the radiologists and ED physicians to compare AI-CAD results with the original CR images without requesting additional AI-CAD analyses (Fig. 5).

During the study period, all CRs were interpreted in the ED by duty trainee radiologists in their third year of the residency, with each exam read by a single reader. CR interpretation used a standardized reporting form (Fig. 6). The standard report included a binary classification for
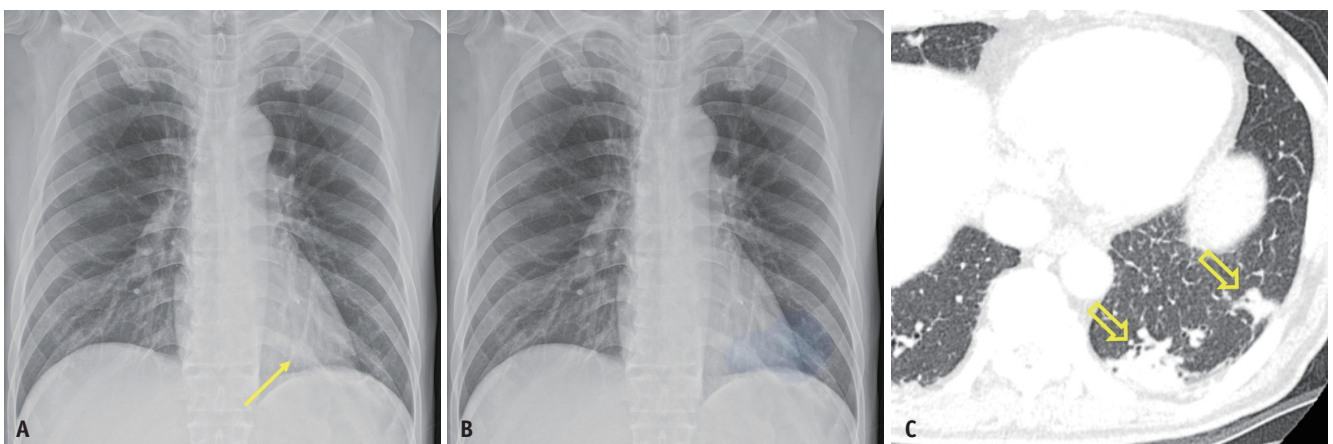

Fig. 2. Images of a 57-year-old male who visited the emergency department with a fever. **A.** Chest radiograph showing increased focal opacity in the left lower lung field (arrow). **B.** The artificial intelligence-based computer-aided detection identified the abnormality with a probability score of 21%. A duty radiologist also reported the abnormality and suggested the possibility of pneumonia. **C.** Chest computed tomography image obtained on the same day in the emergency department showing corresponding multifocal consolidations in the left lower lobe of the lung (open arrows). The patient was diagnosed with pneumonia.
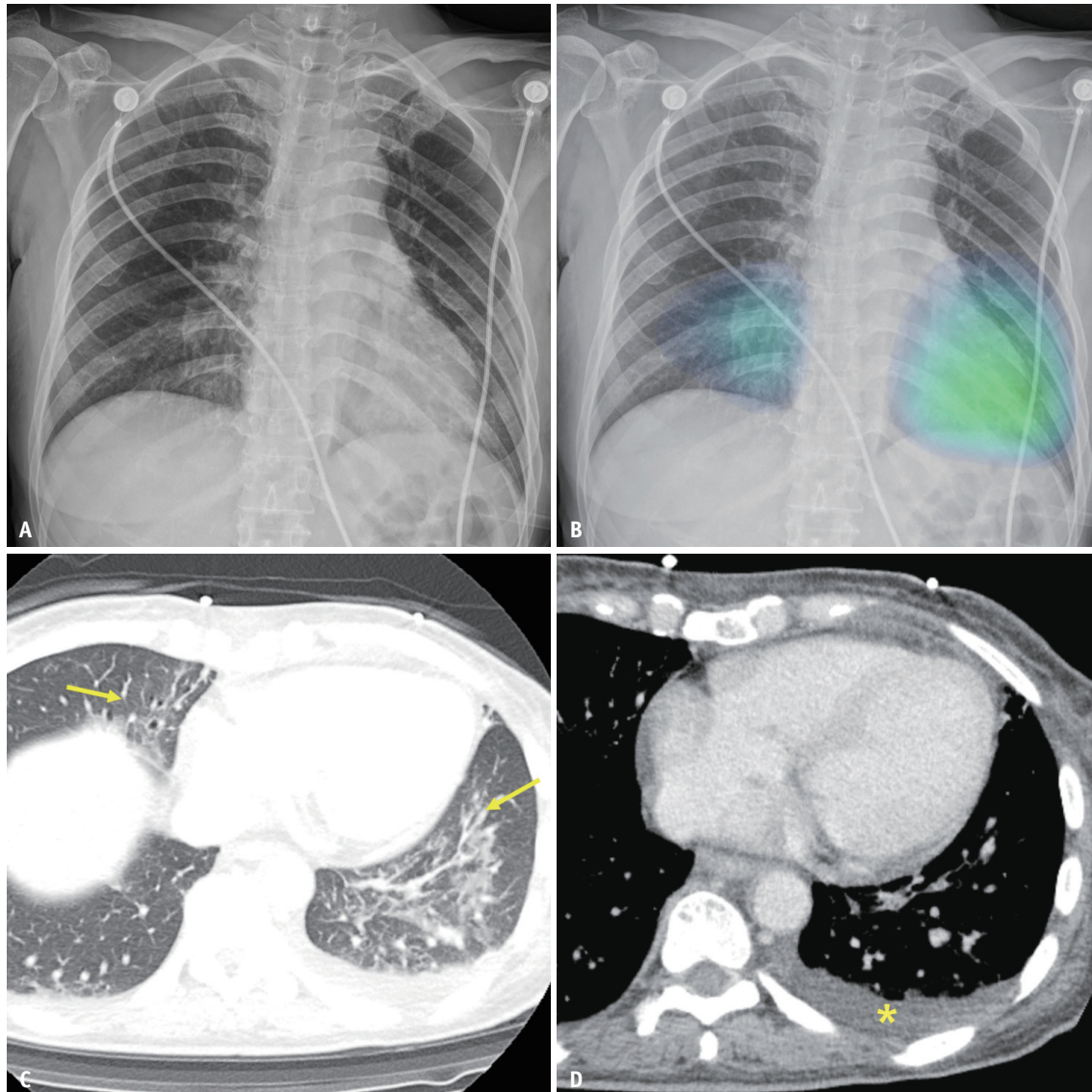
Fig. 3. Images of a 62-year-old female who visited the emergency department with a fever. **A.** Chest radiograph showing vaguely increased opacities in both lower lung fields. **B.** The artificial intelligence-based computer-aided detection identified the findings with a probability score of 52%. The duty radiologist rejected the results of the computer-aided detection and reported that there was no significant abnormality suggesting the presence of an acute thoracic disease. **C, D.** Chest computed tomography image obtained on the same day in the emergency department showing patchy ground-glass opacities in the right middle lobe and left lower lobe of the lung (arrows) **(C)** and a small amount of left pleural effusion (asterisk) **(D)**. The patient was diagnosed with pneumonia.

any abnormality indicating the presence of acute thoracic disease. The duty trainee radiologist can access the electronic report template directly from the PACS, and the generated standardized report can be transferred to the PACS. Before the start of the trial, radiology residents were instructed to use the standardized report for the trial-specific examinations and to review the AI-CAD results for CRs of the intervention group participants. As with the current process in the ED, duty trainee radiologists provide immediate interpretation if the ED physician requests formal interpretation. Otherwise, duty trainee radiologists interpreted the CRs at their own pace before the duty shift.

ED physicians could evaluate CR images (with AI-CAD results for the intervention group) even before obtaining the radiologist's interpretation to make clinical decisions regarding participants' management. Workflow in the ED and participants' management adhered to the ED's standard procedure.
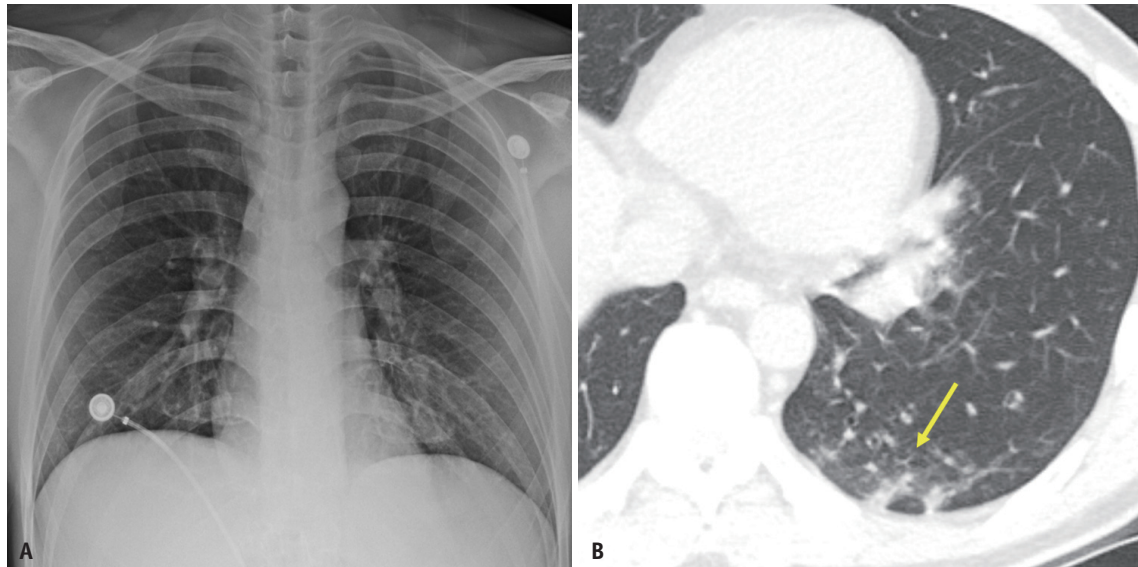
Fig. 4. Images of a 40-year-old male who visited the emergency department with a fever. **A.** Chest radiograph shows no definite abnormality, even after a retrospective correlation with chest CT. The artificial-intelligence-based computer-aided detection identified no abnormality on chest radiograph (probability score: 1%). **B.** Chest CT obtained on the same day in the emergency department showing patchy ground-glass opacity in the left lower lobe of the lung (arrow). The patient was diagnosed with pneumonia. CT = computed tomography
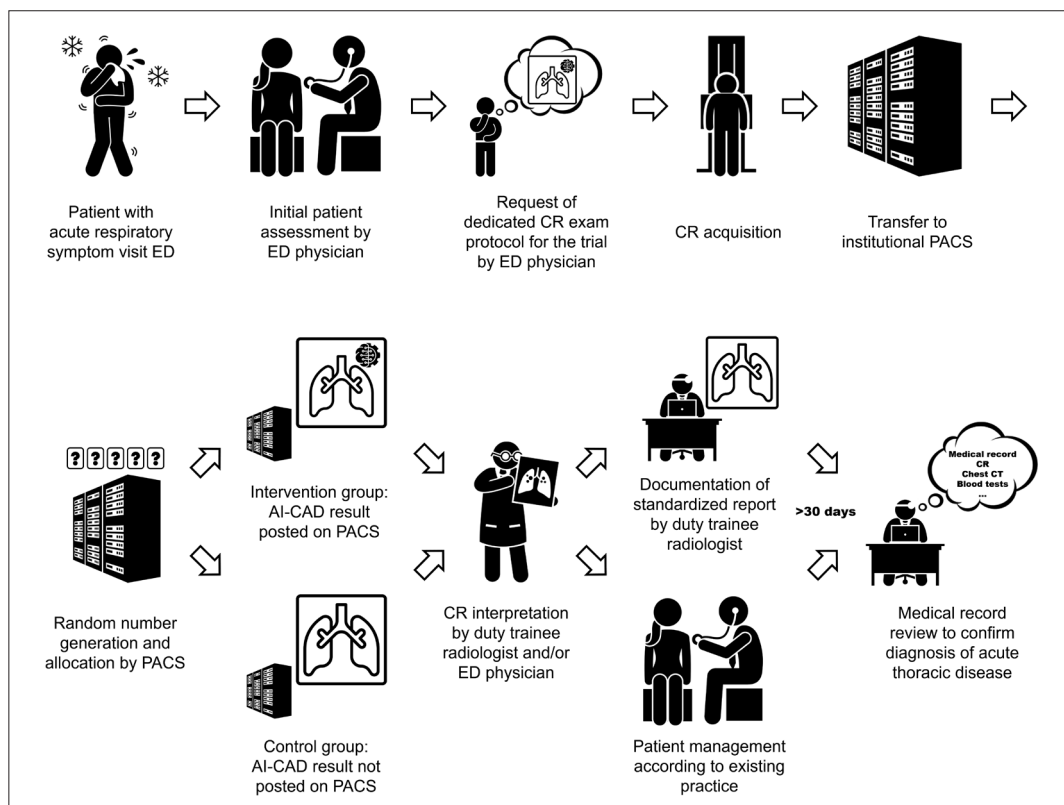

Fig. 5. Scheme of trial implementation. Patients who visited the emergency department (ED) with acute respiratory symptoms and underwent chest radiographs (CRs) with the dedicated exam protocol were systematically enrolled for the trial. After the transfer of the CR image to the institutional picture archiving and communication system (PACS), a random number for allocation of the CR was automatically generated by the PACS. For the CRs allocated to the intervention group, artificial intelligence-based computer-aided detection (AI-CAD) analyses were automatically processed. All CRs in both the intervention and control groups were evaluated by duty trainee radiologists and or ED physicians, and other patient managements followed existing practice in the ED. After at least 30 days after discharge from the ED, a thoracic radiologist reviewed medical records to confirm the diagnosis of any acute thoracic disease. CT = computed tomography

Fig. 6. Template for preparing standardized reports for chest radiographs (CRs). CRs were interpreted using a standardized reporting format for both the intervention and control groups. A duty radiologist could call a simple dedicated program to enter the standardized report from the picture archiving and communication system (PACS). The standardized report had the following components: 1) presence versus absence of any abnormality suggesting acute thoracic diseases; 2) description of an abnormality suggesting acute thoracic diseases in three subcategories: pulmonary parenchymal abnormalities, pleural abnormalities, and other abnormalities; and 3) any recommendations for emergency department physicians. The radiologist could easily transfer the generated report from the program to the PACS. ID = identification, RN = random number, CT = computed tomography

Outcome Definition

The primary outcome was the sensitivity of CR interpretation by duty trainee radiologists for any acute thoracic disease that could explain the participants' chief complaints during their ED visits. The FPR of the CR interpretation was investigated as a secondary outcome.

One investigator E.J.H (a thoracic radiologist with 11 years of experience), reviewed the participant's medical records at least 30 days after their ED visit to define the reference standard for the presence of any acute thoracic disease. The ED medical records, and any follow-up records, radiological, and laboratory test results, were comprehensively evaluated to determine whether acute thoracic disease could explain the participants' chief complaints during ED visits. The process of defining the standard of reference is described in details in (Supplementary Material 1).

The following participant management efficiency and

effectiveness outcomes were investigated: 1) turn-around time (TAT) of CR interpretation, 2) frequency of chest CT acquisition in the ED, 3) TAT to decide on CT acquisition, 4) TAT to decide on antibiotic administration, 5) TAT to refer participants to departments other than the emergency medicine department, 6) length of stay in the ED, and 7) rate of revisiting the ED within 30 days with the same complaint. Each study's endpoint is described in Supplementary Material 2.

Sample Size Estimation

Based on the results of our previous retrospective study, a sample size estimation was conducted before participant enrollment [19]. The target sample size was 4862 participants (2431 participants in each group). The procedure for estimating sample size is described in Supplementary Material 3. We intended to enroll participants

until the sample size was reached or until December 31, 2021, whichever came first. We did not conduct any intermediate analysis.

Statistical Analyses

To compare study outcomes between the intervention and control groups, we first compared the crude outcome values using the chi-square test and t-test. We used mixed-effect models to evaluate the association between CR interpretation using AI-CAD and study outcomes while adjusting for confounding variables. Radiologists who interpreted CR were included as a random effect; however, participants' age and sex, chief complaint of visiting the ED, triage results by the KTAS [25], time of ED visit (weekdays vs. weekends [Saturday, Sunday, national holidays, and closed days of the institution], daytime [8:00 AM to 6:00 PM] vs. nighttime), and type of scanner (fixed vs. portable scanner) were included as fixed effects.

We performed post-hoc subgroup analyses to investigate the study outcomes in subgroups of participants divided according to the reference standard (with and without thoracic disease), AI-CAD results (positive and negative results), and CR scanner types (fixed and portable). We compared the sensitivity and FPR of standalone AI-CAD to those of radiologists' interpretations using McNemar's tests. After stopping participant enrollment, the AI-CAD results of the CRs of the control group participants were obtained for post-hoc analyses and outside participant management.

R software (version 3.6.3, R Project for Statistical Computing) was used to perform all statistical analyses. $P < 0.05$ was considered to be statistically significant.

RESULTS

Participant Characteristics

In this trial we enrolled 3576 patients (male-to-female ratio, 1966:1610; mean age, 64 years) between June 15, 2020, and December 31, 2021. Of the 3576 patients, 1761 and 1815 were allocated to the intervention and control groups, respectively (Fig. 1). Table 1 shows the demographic and clinical information of the participants.

CR Interpretation

The participants' CRs were interpreted by one of the 20 trainee radiologists. The median number of CRs, each radiologist, interpreted was 243 (range, 1–328). Table 1 shows abnormalities suggesting any acute thoracic diseases were

noted in 32.1% and 31.3% of the CRs in the intervention and control groups, respectively.

According to the reference standard, acute thoracic diseases were present in 26.8% and 27.1% of participants in the intervention and control groups, respectively. The most common acute thoracic disease was pneumonia (n = 669) (Supplementary Table 1) The sensitivities of CR interpretation, the primary outcome of the trial, was 67.2% (317/472) and 66.0% (324/491) in the intervention and control groups, respectively. After adjusting for potential confounding variables, there was no association between the use of AI-CAD for interpretation and sensitivity (adjusted odds ratio [OR], 1.02; 95% confidence interval [CI], 0.70–1.49; $P = 0.917$). The FPRs for CXR interpretations were 19.3% (249/1289) and 18.5% (245/1324) in the intervention and control groups, respectively. The FPR was not significantly associated with the use of AI-CAD (adjusted OR, 1.00 [95% CI, 0.79–1.26]; $P = 0.985$) (Tables 2, 3). Supplementary Table 2 shows the sensitivity and FPR of each radiologist on duty.

Management in ED

Approximately 44.7% and 46.8% of the participants in the intervention and control groups underwent chest CT, respectively (adjusted OR, 0.90 [95% CI, 0.76–1.06]; $P = 0.194$). The TAT to decide on CT acquisition was slightly longer in the intervention group (159 vs. 138 min; $P = 0.043$ [for comparison of crude values]); however, the association with AI-CAD use was not statistically significant ($P = 0.090$) after adjustment for confounding variables. AI-CAD was not associated with TATs to decide on antibiotic administration and refer participants to other departments (Table 2).

After ED visits, 42.5%, 39.1%, and 17.8% of participants were discharged home, admitted to the hospital, and transferred to other institutions, respectively (Table 1). AI-CAD was not associated with the length of stay in the ED (968 [intervention group] vs. 972 min [control group]; $P = 0.919$) and the rate of revisiting the ED within 30 days (6.7% [intervention group] vs. 6.6% [control group]; adjusted OR, 1.02 [95% CI, 0.78–1.32]; $P = 0.912$) (Table 2).

Subgroup Analyses

In the subgroup of participants with acute thoracic diseases (472 vs. 491 participants in the intervention and control groups, respectively) No study outcome was associated with using AI-CAD, while the crude TAT to decide on CT acquisition (167 vs. 130 min; $P = 0.027$) and to refer participants to other departments (347 vs. 296 min; $P = 0.034$)

Table 1. Participant Characteristics

Variables	All Participants (n = 3576)	Intervention Group (n = 1761)	Control Group (n = 1815)	P
Age, yr*	64 ± 17	65 ± 17	64 ± 17	0.891
Male participants[†]	55.0% (1966)	55.5% (978)	54.4% (988)	0.530
Chief complaint[†]				0.863
Fever	43.7% (1561)	43.3% (763)	44.0% (798)	
Dyspnea[‡]	35.0% (1250)	34.9% (615)	35.0% (635)	
Chest pain[§]	13.6% (486)	14.0% (247)	13.2% (239)	
Hemoptysis[ǁ]	3.9% (140)	3.5% (64)	4.2% (76)	
Cough	1.6% (57)	1.8% (32)	1.4% (25)	
Chilling sense	1.2% (43)	1.1% (20)	1.3% (23)	
Sputum	1.1% (39)	1.1% (20)	1.0% (19)	
Korean triage and acuity scale[†]				0.418
Level 2[¶]	32.2% (1153)	33.4% (589)	31.1% (564)	
Level 3**	54.1% (1933)	52.7% (928)	55.4% (1005)	
Level 4[††]	12.4% (444)	12.6% (222)	12.2% (222)	
Level 5[‡‡]	1.3% (46)	1.2% (22)	1.3% (24)	
Time of visit[†,§§]				0.953
Weekdays, daytime	45.9% (1640)	45.5% (802)	46.2% (838)	
Weekdays, nighttime	25.5% (913)	25.4% (448)	25.6% (465)	
Weekends, daytime	15.9% (570)	16.1% (283)	15.8% (287)	
Weekends, nighttime	12.7% (453)	12.9% (228)	12.4% (225)	
CRs from fixed scanner[†]	68.4% (2447)	69.3% (1220)	67.6% (1227)	0.298
Positive CR interpretation[†]	31.7% (1135)	32.1% (566)	31.3% (569)	0.637
Interpretations reporting pulmonary abnormalities[†]	24.0% (857)	23.7% (417)	24.2% (440)	0.723
Interpretations reporting pleural abnormalities[†]	11.6% (416)	12.1% (213)	11.2% (203)	0.425
Interpretations reporting other abnormalities[†]	0.5% (18)	0.5% (8)	0.6% (10)	0.863
Positive AI-CAD result[†]	70.8% (2529)	71.1% (1252)	70.4% (1277)	0.672
Participants with acute thoracic disease[†]	26.9% (963)	26.8% (472)	27.1% (491)	0.896
Disposition after visit[†]				0.601
Discharge to home	42.5% (1519)	42.9% (755)	42.1% (764)	
Admission	39.1% (1400)	39.3% (692)	39.0% (708)	
Transfer	17.8% (635)	17.1% (301)	18.4% (334)	
Death	0.6% (22)	0.7% (13)	0.5% (9)	

*Numbers indicate mean ± standard deviation, [†]Numbers in parentheses indicate the number of participants, [‡]Includes respiratory difficulty and dyspnea on exertion, [§]Includes chest wall pain and pleuritic chest pain, [ǁ]Includes blood-tinged sputum, [¶]Potential threats to life, limb, or body function, and quick intervention is needed, **Conditions that can lead to serious problems that potentially require emergency intervention. Significant discomfort or influence on physical function that affects work or everyday life, [††]Patient age, condition associated with the possibility of pain or worsening, and complications, [‡‡]Conditions caused by chronic diseases, [§§]Daytime indicates 8:00 AM to 6:00 PM. Weekends include Saturday, Sunday, national holidays, and the official closed day of the institution. AI-CAD = artificial intelligence-based computer-aided detection; CR = chest radiograph

were slightly longer in the intervention group. No study outcome was associated with using AI-CAD in the subgroup of participants without acute thoracic diseases (n = 1289 and 1324 participants in the intervention and control groups, respectively) (Supplementary Table 3).

AI-CAD was significantly associated with a longer TAT than conventional interpretation to decide on CT acquisition (162 vs. 138 min; P = 0.047) in the subgroup of participants

with positive AI-CAD results (1252 and 1277 participants in the intervention and control groups, respectively). No study outcome was associated with using AI-CAD In the subgroup of patients with negative AI-CAD results (509 and 538 participants in the intervention and control groups, respectively) (Supplementary Table 4).

In the subgroup of participants who underwent CRs with fixed scanners (1220 and 1227 participants in the

Table 2. Comparison of Study Outcomes

Outcomes	All Participants (n = 3576)	Intervention Group (n = 1761)	Control Group (n = 1815)	P (Crude)*	P (Adjusted)[†]
Sensitivity of CR interpretation[‡]	66.6% (641/963)	67.2% (317/472)	66.0% (324/491)	0.751	0.917
FPR of CR interpretation[‡]	18.9% (494/2613)	19.3% (249/1289)	18.5% (245/1324)	0.631	0.985
TAT of CR interpretation (min)[§]	405 ± 386	406 ± 415	403 ± 354	0.338	0.265
Frequency of CT acquisition[‖]	45.8% (1637)	44.7% (788)	46.8% (849)	0.236	0.194
TAT to decide on CT acquisition (min)[§]	148 ± 198	159 ± 231	138 ± 160	0.043	0.090
TAT to decide on antibiotic administration (min)[§]	188 ± 255	190 ± 256	187 ± 255	0.800	0.823
TAT to refer to another department (min)[§]	274 ± 275	279 ± 270	269 ± 280	0.395	0.473
Length of stay in the ED (min)[§]	970 ± 968	968 ± 1002	972 ± 934	0.900	0.919
Rate of revisiting ED[‖]	6.7% (238)	6.7% (118)	6.6% (120)	0.968	0.912

*P-values are the results of the comparison of crude values between the intervention and control groups, [†]P-values are the results of mixed-effect models including age, sex, chief complaints, Korean triage and acuity scale (KTAS), time of visit, and scanner type as fixed effects and interpreting radiologists as a random effect, [‡]Numbers in parentheses indicate numerators/denominators, [§]Numbers indicate mean ± standard deviation, [‖]Numbers in parentheses indicate the number of participants. CR = chest radiograph, FPR = false-positive rate, TAT = turn-around time, CT = computed tomography, ED = emergency department

Table 3. Summary of Mixed Effect Models for Sensitivity and False-Positive Rate of CR Interpretation

Variables	Sensitivity of CR Interpretation		FPR of CR Interpretation	
	Odds Ratio*	P	Odds Ratio*	P
Intervention group (reference: Control group)	1.02 (0.70, 1.49)	0.917	1.00 (0.79, 1.26)	0.985
Male sex (reference: Female sex)	1.31 (0.98, 1.75)	0.073	0.74 (0.60, 0.92)	0.008
Age (for 1 year increase)	1.01 (1.00, 1.02)	0.005	1.03 (1.02, 1.04)	< 0.001
CR from fixed scanner (reference: CR from portable scanner)	1.33 (0.91, 1.94)	0.136	1.60 (1.22, 2.10)	0.001
Chief complaint				
Fever	Reference	Reference	Reference	Reference
Dyspnea	1.33 (0.96, 1.87)	0.091	6.25 (4.88, 8.01)	< 0.001
Chest pain	2.05 (0.91, 4.61)	0.082	1.19 (0.80, 1.78)	0.390
Hemoptysis	0.50 (0.32, 0.80)	0.004	0.81 (0.10, 6.32)	0.840
Couth	2.25 (0.91, 5.55)	0.079	2.58 (0.90, 7.39)	0.077
Chilling sense	3.53 (0.41, 30.77)	0.253	0.9 (0.27, 3.04)	0.868
Sputum	0.94 (0.39, 2.24)	0.887	6.42 (1.92, 21.44)	0.003
Korean triage and acuity scale				
Level 2	Reference	Reference	Reference	Reference
Level 3	0.85 (0.62, 1.17)	0.317	0.71 (0.56, 0.91)	0.007
Level 4	1.11 (0.60, 2.06)	0.731	0.45 (0.29, 0.69)	< 0.001
Level 5	1.56 (0.17, 14.3)	0.694	1.59 (0.65, 3.89)	0.309
Time of visit				
Weekdays, daytime	Reference	Reference	Reference	Reference
Weekdays, nighttime	0.81 (0.57, 1.16)	0.245	0.86 (0.65, 1.14)	0.304
Weekends, daytime	0.72 (0.48, 1.08)	0.114	1.16 (0.85, 1.58)	0.339
Weekends, nighttime	0.88 (0.53, 1.47)	0.633	0.81 (0.56, 1.16)	0.243

Interpreting radiologist was included in the models as a random effect. *Numbers in parentheses indicate 95% confidence intervals. CR = chest radiograph, FPR = false-positive rate

intervention and control groups, respectively), using AI-CAD was significantly associated with a longer TAT than conventional interpretation to decide on CT acquisition (176 vs. 149 min; P = 0.018). In the subgroup of participants who underwent CRs with portable scanners (541 vs. 588

in the intervention and control groups, respectively), no study outcome was associated with using AI-CAD, although the crude length of stay in the ED was slightly shorter in the intervention group (690 vs. 829 min; P = 0.034) (Supplementary Table 5).

Stand-Alone Performance of the AI-CAD

AI-CAD identified abnormal findings in 70.8% of the CRs. Positive AI-CAD results demonstrated a sensitivity of 95.3% and an FPR of 61.7%. The sensitivity and FPR of AI-CAD were significantly higher than those of radiologists on duty (sensitivity, 66.6% [P < 0.001]; FPR, 18.9% [P < 0.001]), respectively. The differences between AI-CAD and radiologists were consistent in the intervention and control groups and in the CRs from the fixed and portable scanners (Supplementary Table 6).

DISCUSSION

In this single-center pragmatic randomized controlled trial, we found no association between using AI-CAD for CR interpretation and the sensitivity (67.2% [intervention group] vs. 66.0% [control group]; odds ratio, 1.02; P = 0.917) or FPR (19.3% vs. 18.5%; OR, 1.00; P = 0.985) of duty trainee radiologists' interpretations for the diagnosis of acute thoracic diseases in patients who presented at the ED with acute respiratory symptoms. Furthermore, although ED physicians were provided with AI-CAD results, using AI-CAD was not associated with workflow efficiency and clinical decision-making outcomes.

Several retrospective studies have suggested that using AI-CAD can improve the diagnostic accuracy of physicians' interpretations [12-17,20]. In our previous study, reinterpretation of CRs using AI-CAD significantly increased sensitivity from 65.6% to 73.4% for detecting referable abnormalities in consecutive baseline CRs obtained in the ED [19]. However, in this prospective, pragmatic clinical trial, conducted in the same institution with the same AI-CAD, the sensitivity and FPR of CR interpretation by radiologists on duty did not differ between interpretation with and without AI-CAD.

CR interpretation environments could be the primary cause of this difference. In a retrospective reader test, AI-CAD tended to influence the reader's interpretation, particularly when they were informed that they were participating in a study investigating the efficacy of AI-CAD. However, in real-world practice, radiologists may tend to maintain their own opinions because they are accountable for the outcomes of their interpretations [26].

The diminished performance of AI-CAD in real-world practice may be another reason for its reduced efficacy. In our trial, AI-CAD showed a high FPR (61.7%), which was higher than that in the retrospective study (9.7% or 30.4% according to the threshold definition) [19]. This high FPR may

have negatively impacted the interaction between AI-CAD and radiologists despite the high sensitivity (95.3%) [27].

Because patient management in the ED and patient outcomes are influenced by factors other than CR results [28], it may be challenging to demonstrate an improvement in workflow efficiency or patient outcomes by using AI-CAD for CR interpretation alone [21]. In our study, most workflow efficiency secondary outcomes did not differ between the intervention and control groups. In the subgroup of participants with positive AI-CAD results and those who underwent CRs with fixed scanners, the TAT to decide on CT acquisition was slightly longer in the intervention group than in the control group (Supplementary Tables 4, 5). Further studies are required to confirm the impact of AI-CAD on clinical decision-making and its efficacy.

There are some limitations to this study. First, because this was a single-center trial conducted at an academic tertiary referral institution, the generalization of our findings remain uncertain. Second, our trial was conducted during the coronavirus pandemic when the emergency department's workflow was significantly impacted. Consequently, the reproducibility of our results in the ED workflow following the end of the pandemic is uncertain. Thirdly, the medical record defined the reference standard for the presence of acute thoracic disease, which might have been influenced by the practice of ED physicians and the subjective opinion of radiologists during the medical record review. Fourth, the CRs were interpreted by trainee radiologists on duty. Therefore, the reproducibility of the results when the CRs are interpreted by readers with various years of experience remains uncertain. Finally, the number of participants included in our trial fell short of the targeted sample size (3576/4862; 73.5%), thereby reducing the statistical power of the study.

In conclusion, using AI-CAD for interpretating chest radiographs did not improve the sensitivity and FPR for diagnosing acute thoracic diseases in patients who presented to the ED with acute respiratory symptoms. Furthermore, the use of AI-CAD by ED physicians did not affect workflow efficiency and clinical decision-making in the ED. To enhance the value of AI-CAD in real-world practice, a more specific real-world optimized strategy for implementing AI-CAD in practice and an improvement in the performance of AI-CAD may be required.

Korean Journal of Radiology

KJR

Supplement

The Supplement is available with this article at https://doi.org/10.3348/kjr.2022.0651.

Availability of Data and Material

The datasets generated or analyzed during the study are available from the corresponding author on reasonable request.

Conflicts of Interest

Jin Mo Goo who is on the editorial board of the *Korean Journal of Radiology* was not involved in the editorial evaluation or decision to publish this article.
Eui Jin Hwang reports research grants from Lunit, Coreline Soft, and Monitor corporation, outside the present study; Jin Mo Goo reports research grants from Infinitt Healthcare, Dongkook Lifescience, and LG Electronics, outside the present study; Ju Gang Nam reports a research grant from Vuno, outside the present study; Chang Min Park reports a research grant from Lunit, outside the present study, and stock of Promedius and stock options of Lunit and Coreline Soft. All remaining authors have declared no conflicts of interest.

Author Contributions

Conceptualization: Eui Jin Hwang, Jin Mo Goo. Data curation: Eui Jin Hwang, Ju Gang Nam. Formal analysis: Eui Jin Hwang. Funding acquisition: Jin Mo Goo. Investigation: Eui Jin Hwang, Jin Mo Goo. Methodology: Eui Jin Hwang, Jin Mo Goo. Project administration: Jin Mo Goo, Ki Jeong Hong. Resources: Jin Mo Goo. Software: Eui Jin Hwang, Chang Min Park. Supervision: Jin Mo Goo. Validation: Ju Gang Nam, Chang Min Park, Ki Jeong Hong, Ki Hong Kim. Visualization: Eui Jin Hwang. Writing—original draft: Eui Jin Hwang, Jin Mo Goo. Writing-review—editing: Ju Gang Nam, Chang Min Park, Ki Jeong Hong, Ki Hong Kim.

ORCID iDs

Eui Jin Hwang
  https://orcid.org/0000-0002-3697-5542
Jin Mo Goo
  https://orcid.org/0000-0003-1791-7942
Ju Gang Nam
  https://orcid.org/0000-0003-3991-4523
Chang Min Park
  https://orcid.org/0000-0003-1884-3738
Ki Jeong Hong
  https://orcid.org/0000-0003-3334-817X
Ki Hong Kim
  https://orcid.org/0000-0003-2115-0078

REFERENCES

1. Expert Panel on Thoracic Imaging: Jokerst C, Chung JH, Ackman JB, Carter B, Colletti PM, Crabtree TD, et al. ACR Appropriateness Criteria(R) Acute respiratory illness in immunocompetent patients. *J Am Coll Radiol* 2018;15(11S):S240-S251
2. Hoffmann U, Akers SR, Brown RK, Cummings KW, Cury RC, Greenberg SB, et al. ACR appropriateness criteria acute nonspecific chest pain-low probability of coronary artery disease. *J Am Coll Radiol* 2015;12(12 Pt A):1266-1271
3. Heitkamp DE, Albin MM, Chung JH, Crabtree TP, Iannettoni MD, Johnson GB, et al. ACR Appropriateness Criteria(R) acute respiratory illness in immunocompromised patients. *J Thorac Imaging* 2015;30:W2-W5
4. Ketai LH, Mohammed TL, Kirsch J, Kanne JP, Chung JH, Donnelly EF, et al. ACR appropriateness criteria(R) hemoptysis. *J Thorac Imaging* 2014;29:W19-W22
5. Cairns C, Kang K, Santo L. National Hospital Ambulatory Medical Care Survey: 2018 emergency department summary tables. Centers for Disease Control and Prevention.com Web site. https://www.cdc.gov/nchs/data/nhamcs/web_tables/2018-ed-web-tables-508.pdf. Published May 7, 2021. Accessed August 30, 2022
6. Chung JH, Duszak R Jr, Hemingway J, Hughes DR, Rosenkrantz AB. Increasing utilization of chest imaging in US emergency departments from 1994 to 2015. *J Am Coll Radiol* 2019;16:674-682
7. Sellers A, Hillman BJ, Wintermark M. Survey of after-hours coverage of emergency department imaging studies by US academic radiology departments. *J Am Coll Radiol* 2014;11:725-730
8. Hwang EJ, Park CM. Clinical implementation of deep learning in thoracic radiology: potential applications and challenges. *Korean J Radiol* 2020;21:511-525

9. Lee SM, Seo JB, Yun J, Cho YH, Vogel-Claussen J, Schiebler ML, et al. Deep learning applications in chest radiography and computed tomography: current state of the art. *J Thorac Imaging* 2019;34:75-85

10. Dunnmon JA, Yi D, Langlotz CP, Ré C, Rubin DL, Lungren MP. Assessment of convolutional neural networks for automated classification of chest radiographs. *Radiology* 2019;290:537-544

11. Hwang EJ, Park S, Jin KN, Kim JI, Choi SY, Lee JH, et al. Development and validation of a deep learning-based automated detection algorithm for major thoracic diseases on chest radiographs. *JAMA Netw Open* 2019;2:e191095

12. Hwang EJ, Park S, Jin KN, Kim JI, Choi SY, Lee JH, et al. Development and validation of a deep learning-based automatic detection algorithm for active pulmonary tuberculosis on chest radiographs. *Clin Infect Dis* 2019;69:739-747

13. Majkowska A, Mittal S, Steiner DF, Reicher JJ, McKinney SM, Duggan GE, et al. Chest radiograph interpretation with deep learning models: assessment with radiologist-adjudicated reference standards and population-adjusted evaluation. *Radiology* 2020;294:421-431

14. Nam JG, Kim M, Park J, Hwang EJ, Lee JH, Hong JH, et al. Development and validation of a deep learning algorithm detecting 10 common abnormalities on chest radiographs. *Eur Respir J* 2021;57:2003061

15. Park S, Lee SM, Lee KH, Jung KH, Bae W, Choe J, et al. Deep learning-based detection system for multiclass lesions on chest radiographs: comparison with observer readings. *Eur Radiol* 2020;30:1359-1368

16. Sim Y, Chung MJ, Kotter E, Yune S, Kim M, Do S, et al. Deep convolutional neural network-based software improves radiologist detection of malignant lung nodules on chest radiographs. *Radiology* 2020;294:199-209

17. Sung J, Park S, Lee SM, Bae W, Park B, Jung E, et al. Added value of deep learning-based detection system for multiple major findings on chest radiographs: a randomized crossover study. *Radiology* 2021;299:450-459

18. Hwang EJ, Goo JM, Yoon SH, Beck KS, Seo JB, Choi BW, 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* 2021;22:1743-1748

19. Hwang EJ, Nam JG, Lim WH, Park SJ, Jeong YS, Kang JH, et al. Deep learning for chest radiograph diagnosis in the emergency department. *Radiology* 2019;293:573-580

20. Nam JG, Park S, Hwang EJ, Lee JH, Jin KN, Lim KY, et al. Development and validation of deep learning-based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. *Radiology* 2019;290:218-228

21. Hong W, Hwang EJ, Lee JH, Park J, Goo JM, Park CM. Deep learning for detecting pneumothorax on chest radiographs after needle biopsy: clinical implementation. *Radiology* 2022;303:433-441

22. Hwang EJ, Lee JS, Lee JH, Lim WH, Kim JH, Choi KS, et al. Deep learning for detection of pulmonary metastasis on chest radiographs. *Radiology* 2021;301:455-463

23. Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology* 2018;286:800-809

24. Park SH, Choi JI, Fournier L, Vasey B. Randomized clinical trials of artificial intelligence in medicine: why, when, and how? *Korean J Radiol* 2022;23:1119-1125

25. Park J, Lim T. Korean Triage and Acuity Scale (KTAS). *J Korean Soc Emerg Med* 2017;28:547-551

26. Price WN 2nd, Gerke S, Cohen IG. Potential liability for physicians using artificial intelligence. *JAMA* 2019;322:1765-1766

27. Gaube S, Suresh H, Raue M, Merritt A, Berkowitz SJ, Lermer E, et al. Do as AI say: susceptibility in deployment of clinical decision-aids. *NPJ Digit Med* 2021;4:31

28. Van den Bruel A, Cleemput I, Aertgeerts B, Ramaekers D, Buntinx F. The evaluation of diagnostic tests: evidence on technical and diagnostic accuracy, impact on patient outcome and cost-effectiveness is needed. *J Clin Epidemiol* 2007;60:1116-1122