

RESEARCH

Open Access



Deep learning computer-aided detection system for pneumonia in febrile neutropenia patients: a diagnostic cohort study

Eui Jin Hwang¹, Jong Hyuk Lee¹, Jae Hyun Kim¹, Woo Hyeon Lim^{1,2}, Jin Mo Goo¹ and Chang Min Park^{1,3*}

Abstract

Background: Diagnosis of pneumonia is critical in managing patients with febrile neutropenia (FN), however, chest X-ray (CXR) has limited performance in the detection of pneumonia. We aimed to evaluate the performance of a deep learning-based computer-aided detection (CAD) system in pneumonia detection in the CXRs of consecutive FN patients and investigated whether CAD could improve radiologists' diagnostic performance when used as a second reader.

Methods: CXRs of patients with FN (a body temperature ≥ 38.3 °C, or a sustained body temperature ≥ 38.0 °C for an hour; absolute neutrophil count $< 500/\text{mm}^3$) obtained between January and December 2017 were consecutively included, from a single tertiary referral hospital. Reference standards for the diagnosis of pneumonia were defined by consensus of two thoracic radiologists after reviewing medical records and CXRs. A commercialized, deep learning-based CAD system was retrospectively applied to detect pulmonary infiltrates on CXRs. For comparing performance, five radiologists independently interpreted CXRs initially without the CAD results (radiologist-alone interpretation), followed by the interpretation with CAD. The sensitivities and specificities for detection of pneumonia were compared between radiologist-alone interpretation and interpretation with CAD. The standalone performance of the CAD was also evaluated, using area under the receiver operating characteristic curve (AUC), sensitivity, and specificity. Moreover, sensitivity and specificity of standalone CAD were compared with those of radiologist-alone interpretation.

Results: Among 525 CXRs from 413 patients (52.3% men; median age 59 years), pneumonia was diagnosed in 128 (24.4%) CXRs. In the interpretation with CAD, average sensitivity of radiologists was significantly improved (75.4% to 79.4%, $P=0.003$) while their specificity remained similar (75.4% to 76.8%, $P=0.101$), compared to radiologist-alone interpretation. The CAD exhibited AUC, sensitivity, and specificity of 0.895, 88.3%, and 68.3%, respectively. The standalone CAD exhibited higher sensitivity (86.6% vs. 75.2%, $P<0.001$) and lower specificity (64.8% vs. 75.4%, $P<0.001$) compared to radiologist-alone interpretation.

Conclusions: In patients with FN, the deep learning-based CAD system exhibited radiologist-level performance in detecting pneumonia on CXRs and enhanced radiologists' performance.

Keywords: Radiography, Thoracic, Deep learning, Artificial intelligence, Pneumonia, Febrile neutropenia

Background

Febrile neutropenia (FN) is observed in approximately 1% of patients receiving chemotherapy and has a mortality rate of approximately 10% [1–5]. Therefore, it is considered as a medical emergency that requires timely

*Correspondence: cmpark.morphius@gmail.com

¹ Department of Radiology, Seoul National University College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul 03080, Korea
Full list of author information is available at the end of the article



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

diagnosis and management [1, 5]. Investigation of the infection focus is the key component of diagnostic work-ups in these patients, and pneumonia is one of the most common and important causes of FN [6–9]. Indeed, pneumonia is associated with a higher rate of treatment failure, longer hospitalization, and increased mortality in FN patients [1–3, 10].

Chest X-ray (CXR) is the initial radiological examination for the evaluation of pneumonia at low cost and wide availability and thus considered a routine investigation in FN patients [1, 11–13]. However, a poor immune response may delay the development of radiographic infiltrates, causing difficulty in the sensitive detection of early pneumonia on CXR [14–16]. Furthermore, CXRs have high inter-observer variability among less-experienced readers [17, 18], and timely interpretation of CXRs by expert radiologists is not always possible. In this regard, a computer-aided detection (CAD) system that can identify pulmonary infiltrates suggestive of pneumonia on CXRs may help managing FN patients.

Recently, deep learning technologies showed promising performance regarding the detection of pneumonia on CXRs [19–22]. However, those studies tested the performance of CAD using conveniently collected datasets, with enriched prevalence of pneumonia [19–22] and limited radiologic diversity (i.e., composed of CXRs with pneumonia and clearly normal CXRs) [19–21], which is definitely far from the situation in the real-world setting. Thus, it remains difficult to believe that deep learning-based CAD truly proved their clinical usefulness in pneumonia detection [22]. In this context, external validation in consecutive patients in a real clinical setting is necessary to confirm the real-world accuracy and clinical usefulness of deep learning-based CAD.

Therefore, we evaluated the performance of deep learning-based CAD in pneumonia detection in the CXRs of consecutive FN patients and investigated whether CAD could improve radiologists' diagnostic performance when used as a second reader.

Methods

This single-center, retrospective study was approved by the institutional review board. The requirement for patients' informed consents was waived.

Study population

We retrospectively included patients from a single tertiary referral hospital if they met the following inclusion criteria: (a) patients with a body temperature ≥ 38.3 °C or a sustained body temperature ≥ 38.0 °C for an hour between January and December 2017 [13]; (b) patients with absolute neutrophil count (ANC) $< 500/\text{mm}^3$ in

peripheral blood on the day of fever [13]; and (c) patients who underwent CXRs on the day of fever.

In cases of multiple CXRs from a single patient, only the first CXR was included to represent the episode of FN. However, if there was an afebrile period of 30 days or longer between the days of fever, they were regarded as different episodes of FN and the first CXR in each episode was included.

CXR acquisition and CAD system

All CXRs were obtained using a fixed radiography system (Digital Diagnost, Philips Healthcare, Eindhoven, The Netherlands) or a portable radiography scanner (DRX-revolution, Carestream Health, Rochester, NY), depending on patients' condition. CXRs from a fixed radiography system were obtained in the erect position with posteroanterior projection, while CXRs from a portable radiography scanner were obtained in the supine or sitting position with anteroposterior projection.

We used a commercialized, regulatory-approved deep learning-based CAD (Lunit INSIGHT CXR 2, version 2.0.0.0, Lunit Inc., Seoul, Korea) to evaluate CXRs. The CAD was designed to detect radiographic abnormalities, including pulmonary nodules or masses, pulmonary infiltrates, and pneumothorax on a single frontal CXR. The CAD was initially trained using 54,221 normal CXRs and 35,613 abnormal CXRs including 6,903 CXRs from patients with pneumonia (see supplementary material for further information) [21].

The CAD provided probability scores for the presence of each target abnormality between 0 and 100% for a CXR. When the probability score was 15% or greater, the CAD also provided a heat map overlaid on the CXR for the localization of the abnormality (Fig. 1). In the present study, we used only probability scores for pulmonary infiltrate (see supplementary material for further information).

A CXR of a 68-year-old woman with febrile neutropenia (absolute neutrophil count, $366/\text{mm}^3$; body temperature, 39.1 °C) showing increased opacity in the right upper lung (A, arrow). A chest computed tomographic scan obtained on the same day showing the corresponding consolidation and ground-glass opacity lesions in the upper lobe of the right lung, suggestive of pneumonia (B). The CAD system correctly identified the abnormality with a probability score of 84% (C). In the reader test, three of five radiologists correctly identified the pulmonary infiltrate on the CXR in the radiologist-alone interpretation, and all five radiologists identified the lesion in the interpretation with CAD.

A CXR of a 69-year-old man with febrile neutropenia (absolute neutrophil count, $489/\text{mm}^3$; body temperature, 38.9 °C) showing subtle increased opacity in the right

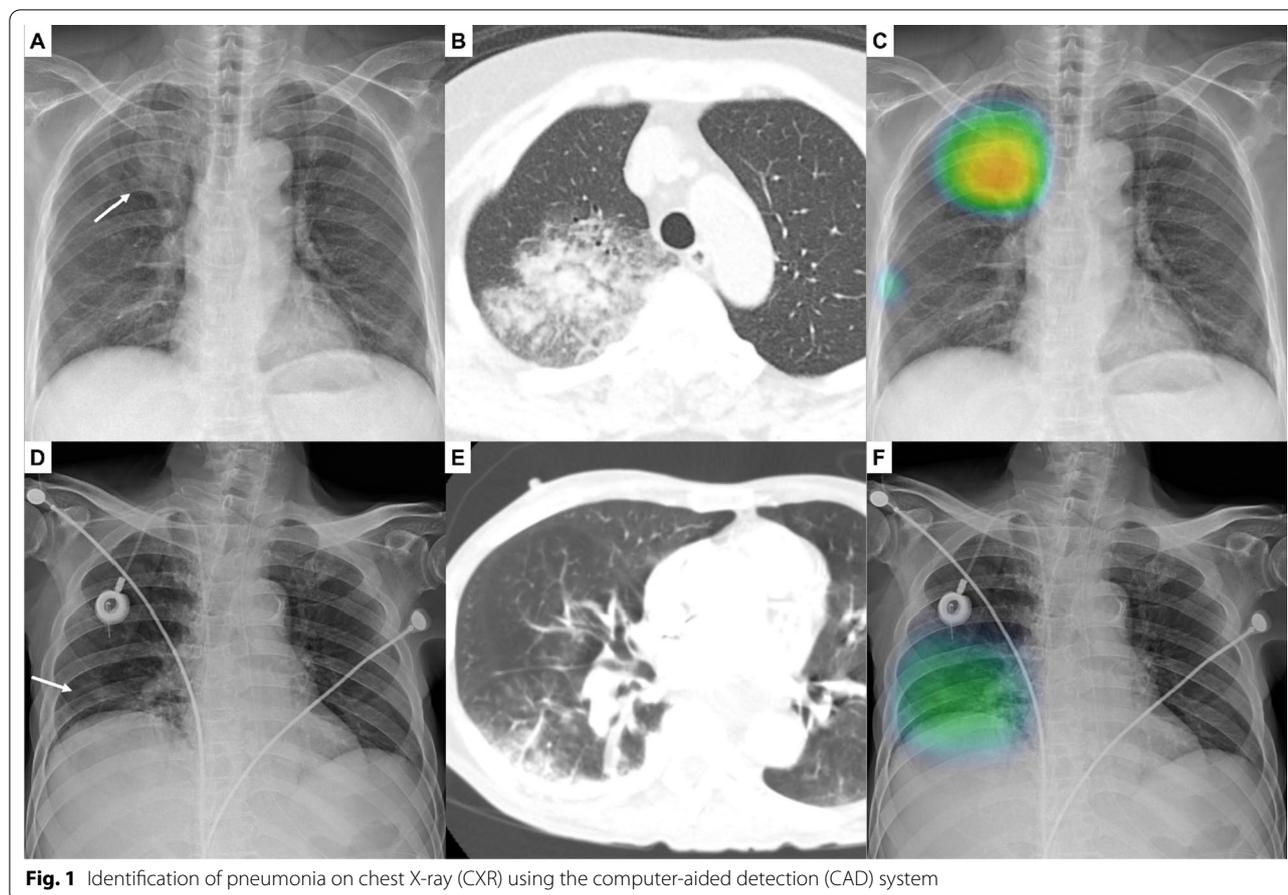


Fig. 1 Identification of pneumonia on chest X-ray (CXR) using the computer-aided detection (CAD) system

lower lung (D). A chest computed tomographic scan showing the corresponding consolidation and ground-glass opacity lesions in the lower lobe of the right lung, suggestive of pneumonia (E). The CAD system correctly identified the abnormality with a probability score of 51% (F). In the reader test, two of five radiologists correctly identified the pulmonary infiltrate on the CXR in the radiologist-alone interpretation, and four radiologists including two radiologists who initially could not recognize the lesion correctly identified the lesion in the interpretation with CAD.

Reference standard

Two thoracic radiologists (C.M.P. and E.J.H., 21 and 9 years of experience in interpreting CXR and chest CT) who were blinded to the CAD results reviewed patients' medical records and radiological, microbial, and laboratory examination results to define, in consensus, the clinical diagnosis of pneumonia at the time of CXR. The presence of clinical features suggestive of respiratory infection with demonstrable pulmonary infiltrate by CXR or CT was regarded as a clinical diagnosis of pneumonia, regardless of the microbiological identification of

the causative pathogen [23]. When a pulmonary infiltrate was identified on CT but invisible on CXR, such case was defined as positive for pneumonia based on CT findings.

Reader test

To compare the performance of CAD with that of radiologists and to evaluate whether CAD can enhance the performance of radiologists, we conducted a retrospective reader test. Five board-certified radiologists (general radiology, 1–3 years of experience after finishing residency) participated in the reader test, and 50% of CXRs among the entire cohort were randomly sampled for the test.

Each radiologist independently interpreted CXRs to classify them into those with or without pulmonary infiltrates suggestive of pneumonia. First, radiologists read the CXRs without CAD results (radiologist-alone interpretation). After finishing the first reading session for all CXRs, they re-interpreted the CXRs with the corresponding CAD results and were allowed to change their initial decision as needed (interpretation with CAD). The radiologists were informed that all CXRs were obtained

from FN patients; however, they were blinded to other clinical or laboratory information.

Subgroup analyses

For a more solid reference standard for the presence of pulmonary infiltrates, we separately evaluated the performance of CAD in patients with available chest CT obtained within 3 days from the CXRs with reference to CTs for the presence of pulmonary infiltrates.

To investigate the performance of CAD in patients with different clinical characteristics, we evaluated the performance of CAD in the following subgroup populations: (a) male vs. female patients; (b) patients aged < 60 years vs. \geq 60 years; (c) CXRs from a fixed radiography system vs. CXRs from a portable radiography scanner.

Statistical analyses

The discriminative performance of CAD (ability to separate CXRs with and without pneumonia) was evaluated using the area under the receiver operating characteristic curve (AUC). The sensitivity and specificity of CAD were determined at the pre-defined threshold (probability score of 15%) recommended by the manufacturer. A comparison of AUCs of CAD between subgroups was performed according to the method by DeLong [24], while the chi-square test was used to compare the sensitivities and specificities of CAD between subgroups.

To compare the performance between radiologists and CAD, the sensitivity and specificity of CAD at the pre-defined threshold were compared with those of individual radiologists using generalized estimating equations, to consider the clustering effects caused by multiple CXRs obtained from a single patient and multiple interpretations by radiologists for a single CXR [25]. The sensitivities and specificities of radiologists in radiologist-alone interpretation and interpretation with CAD were also compared to investigate whether or not CAD can enhance radiologists' detection performance.

Calibration (degree of agreement between the predicted probabilities by CAD versus the observed probability of pneumonia) of CAD was evaluated by constructing a calibration plot. Inter-reader agreement among radiologists was evaluated using the Fleiss' kappa coefficient.

All statistical analyses were performed using R (version 3.6.3, R project for statistical computing, Vienna, Austria). A *P* value of less than 0.05 was considered to indicate a statistically significant difference.

Results

Demographic and clinical information

A total of 525 CXRs obtained from 413 patients (216 men and 197 women; median age, 59 years [interquartile

range (IQR), 48–67]) were analyzed in the present study. Based on the information in medical records, pneumonia was diagnosed in 128 (24.4%) FN episodes. Detailed demographic and clinical information are described in Table 1.

Standalone performance of the CAD

The performance of CAD is summarized in Figure three. The CAD exhibited an AUC of 0.892 (95% confidence interval [CI], 0.851–0.926) for the identification of pneumonia on CXRs. At the pre-defined threshold, CAD exhibited a sensitivity and specificity of 88.3% (95% CI, 82.7–93.9%) and 68.3% (95% CI, 63.7–72.8%), respectively (Fig. 2).

Regarding calibration, CAD tended to overestimate the probability of pneumonia (Fig. 2). Data of the sensitivities and specificities of CAD at different thresholds are

Table 1 Demographic and clinical information

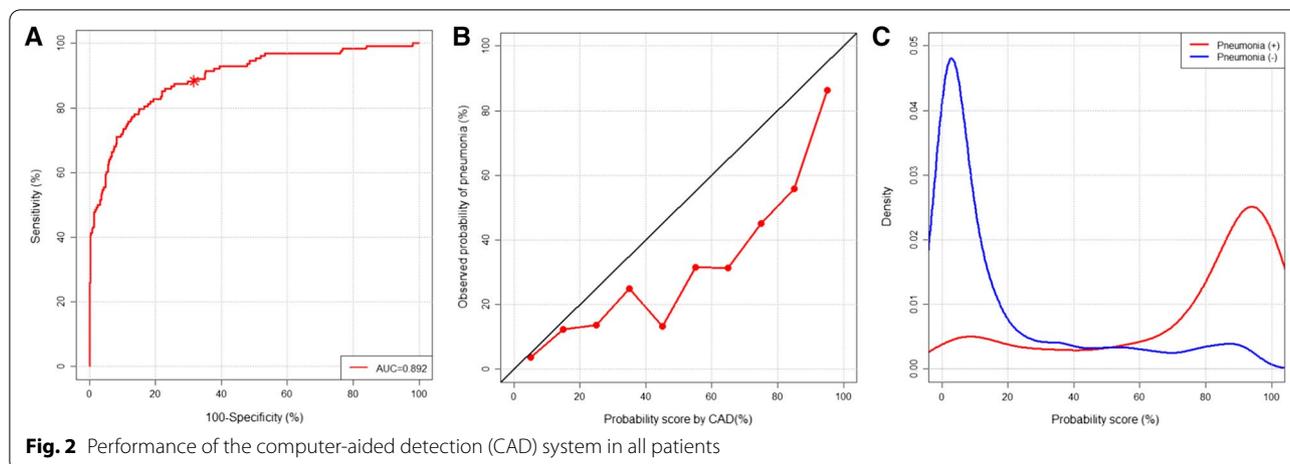
Variable	Data
Male patient (%) ^a	52.3% (216/413)
Age (year) ^b	59 (48–67)
Number of CXR per patient ^a	
1 CXR per patient	79.9% (330/413)
2 CXRs per patient	13.6% (56/413)
3 CXRs per patient	6.1% (25/413)
4 CXRs per patient	0.5% (2/413)
Underlying disease of medical history ^a	
Hematologic malignancy or lymphoma	60.8% (251/413)
Solid organ malignancy	26.4% (109/413)
Hematopoietic stem cell transplantation	17.4% (72/413)
Solid organ transplantation	1.7% (7/413)
ANC (count per mm ³) ^b	36 (0–166)
Peak body temperature (°C) ^b	38.9 (38.5–39.3)
Site of infection ^c	
Pneumonia	24.4% (128/525)
Gastrointestinal tract	6.9% (36/525)
Bone or soft tissue	4.2% (22/525)
Oral cavity	1.9% (10/525)
Bloodstream infection	1.7% (9/525)
Urinary tract	1.3% (7/525)
Biliary tree	0.8% (4/525)
Unidentifiable infection focus	62.1% (326/525)
CXRs from fixed radiography unit	27.8% (146/525)
CXRs from portable radiography scanner	72.2% (379/525)

CAD, computer-aided detection; CXR, chest X-ray; ANC, absolute neutrophil count

^a Data indicate proportions among entire patients (numerators/denominators)

^b Data indicate median (interquartile ranges) values

^c Data indicate proportions among entire episodes of febrile neutropenia (numerators/denominators)



listed in Additional file 1: Table S1 in the supplementary material.

For all patients, the CAD system showed an area under the receiver operating characteristic curve of 0.892 for the identification of pneumonia in the chest X-ray. At a pre-defined threshold probability score of 15%, the sensitivity and specificity of the CAD system were 88.3% and 68.3%, respectively (A). The calibration plot (B) shows that the CAD system tended to overestimate the probability of pneumonia. The density plot of the probability score from the CAD system according to the presence of pneumonia (C) shows that the CAD system can relatively clearly separate CXRs with pneumonia from those without pneumonia.

Reader test

A total of 263 (50.1%) CXRs were randomly sampled for the reader test. Among them, 67 (25.5%) were obtained from patients with pneumonia. Data of the performances of CAD and each radiologist are listed in Table 2 and Fig. 3.

In the radiologist-alone interpretation, radiologists exhibited average sensitivity and specificity of 75.2% and 75.4%, respectively. At the pre-defined threshold, the CAD exhibited significantly higher sensitivity (86.6% vs. 75.2%, $P < 0.001$) and lower specificity (64.8% vs. 75.4%, $P < 0.001$) than radiologist-alone interpretation. Regarding individual radiologists, the CAD showed significantly higher sensitivity and specificity than four and one radiologist, respectively. Meanwhile, the CAD showed significantly lower specificity than four radiologists (Table 2).

In the interpretation with CAD, average sensitivity of radiologists was significantly improved (75.4% to 79.4%, $P = 0.003$) while their specificity remained similar (75.4% to 76.8%, $P = 0.101$), compared to radiologist-alone interpretation. Regarding individual radiologists, significant

improvements in sensitivity and specificity were observed in two and one radiologist, respectively (Table 2).

Analyses after exclusion of radiologists who showed outlier sensitivity and specificity exhibited similar results with those including all radiologists (Additional file 1: Table S2 in the supplementary material).

Regarding the inter-reader agreement among radiologists, interpretation with CAD exhibited better agreement (Fleiss' kappa coefficient, 0.638 [95% CI, 0.600–0.676]) than the radiologist-alone interpretation (Fleiss' kappa coefficient, 0.546 [95% CI, 0.507–0.584]).

In the reader test, the CAD system showed an area under the receiver operating characteristic curve of 0.895 for the identification of pneumonia in the chest X-ray. At a pre-defined threshold probability score of 15%, the sensitivity and specificity of the CAD system were 86.6% and 64.8%, respectively. In the radiologist-alone interpretation, radiologists exhibited average sensitivity and specificity of 75.2% and 75.4%, respectively. In the interpretation with CAD, average sensitivity of radiologists was significantly improved (79.4%), at the similar specificity (76.8%).

Subgroup analyses

The discriminative performance of CAD in different subgroups is shown in Table 3 and Fig. 4.

Chest CTs obtained within 3 days from the CXRs were available in 122 (23.2%) CXRs. Based on CT findings, pulmonary infiltrates suggestive of pneumonia were identified in 68 of 122 cases (55.7%). Among the pulmonary infiltrates, 57 were retrospectively visible on CXRs, while the other 11 were invisible on CXRs, even after correlation with chest CT findings. With the reference standard of chest CT, the CAD system exhibited an AUC of 0.852 (95% CI, 0.776–0.909). At the pre-defined threshold, CAD exhibited sensitivity and specificity of

Table 2 Performances of the computer-aided detection system and radiologists in the reader test

Reader	Sensitivity	P value	Specificity	P value
CAD system	86.6% (58/67, 76.2–92.9%) ^a	Reference	64.8% (127/196, 58.1–71.5%) ^a	Reference
Radiologist-alone interpretation				
Average	75.2% (252/335, 65.7–82.8%)	< 0.001 ^b	75.4% (739/980, 70.9–79.4%)	< 0.001 ^b
Radiologist A (8 years experience)	79.1% (53/67, 67.7–87.2%)	0.019 ^b	76.0% (149/196, 69.5–81.5%)	< 0.001 ^b
Radiologist B (6 years experience)	71.6% (48/67, 59.8–81.1%)	0.001 ^b	81.1% (159/196, 75.0–86.0%)	< 0.001 ^b
Radiologist C (7 years experience)	64.2% (43/67, 52.1–74.7%)	< 0.001 ^b	88.3% (173/196, 83.0–92.1%)	< 0.001 ^b
Radiologist D (7 years experience)	86.6% (58/67, 79.2–92.9%)	> 0.999 ^b	44.9% (88/196, 38.1–51.9%)	< 0.001 ^b
Radiologist E (5 years experience)	74.6% (50/67, 62.3–83.6%)	0.003 ^b	86.7% (170/196, 81.2–90.8%)	< 0.001 ^b
Interpretation with CAD				
Average	79.4% (266/335, 69.9–86.5%)	0.003 ^c	76.8% (753/980, 72.1–81.0%)	0.101 ^c
Radiologist A (8 years experience)	85.1% (57/67, 74.4–91.8%)	0.039 ^c	75.5 (148/196, 69.0–81.0%)	0.841 ^c
Radiologist B (6 years experience)	76.1% (51/67, 64.5–84.8%)	0.174 ^c	79.6% (156/196, 73.4–84.7%)	0.365 ^c
Radiologist C (7 years experience)	73.1% (49/67, 61.3–82.4%)	0.010 ^c	87.8% (172/196, 82.4–91.7%)	0.763 ^c
Radiologist D (7 years experience)	86.6% (58/67, 79.2–92.9%)	> 0.999 ^c	53.6% (105/196, 46.6–60.4%)	< 0.001 ^c
Radiologist E (5 years experience)	76.1% (51/67, 64.5–84.8%)	0.314 ^c	87.8% (172/196, 82.4–91.7%)	0.316 ^c

Numbers in parentheses indicate numerators/denominators, 95% confidence intervals

The experience of each radiologist indicates the length of their experiences in the interpretation of chest X-rays

CAD, computer-aided detection

^a Performance of the CAD system at the predefined threshold (probability score of 15%)

^b Comparison of performance between the CAD system and the radiologist-alone interpretation

^c Comparison of performance between radiologist-alone interpretation and interpretation with CAD

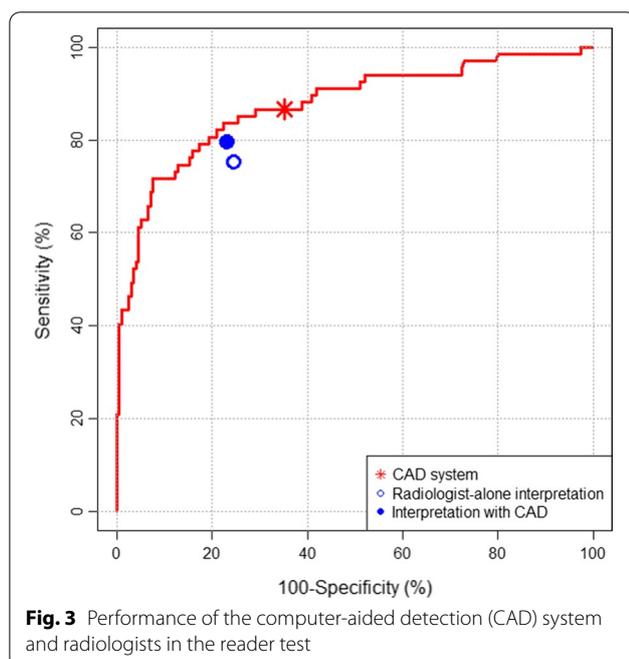


Fig. 3 Performance of the computer-aided detection (CAD) system and radiologists in the reader test

83.8% (95% CI, 75.1–92.6%) and 59.3% (95% CI, 46.2–72.4%), respectively. When compared with patients without chest CT (AUC, 0.917 [$P=0.099$]; sensitivity, 93.3% [$P=0.095$]; specificity, 69.7% [$P=0.126$]), CAD showed

a slightly lower performance in patients with chest CT, without any significant difference.

The CAD showed similar AUCs between men vs. women (0.892 vs. 0.899; $P=0.854$), patients aged < 60 years vs. ≥ 60 years (0.906 vs. 0.867; $P=0.300$), and CXRs from a fixed radiography system vs. CXRs from a portable radiography scanner (0.899 vs. 0.899; $P=0.831$). However, the specificity of CAD was significantly higher in patients aged < 60 years than in those aged ≥ 60 years (79.8% vs. 52.7%; $P < 0.001$) and in CXRs from the fixed radiography system than in CXRs from the portable radiography scanner (79.5% vs. 63.1%; $P=0.001$).

The CAD system exhibited a slightly lower area under the receiver operating characteristic curve (AUC: 0.852 vs. 0.917; $P=0.10$), sensitivity (83.8% vs. 93.3%, $P=0.10$), and specificity (59.3% vs. 69.7%, $P=0.13$) in patients with a chest CT scan than in those without a chest CT scan, without a significant difference (A). The CAD system showed similar AUCs between men and women (0.892 vs. 0.899; $P=0.854$; B), patients aged < 60 years and ≥ 60 years (0.906 vs. 0.867; $P=0.300$; C), and chest X-rays (CXRs) from a fixed radiography system and CXRs from a portable radiography scanner (0.899 vs. 0.899; $P=0.83$; D). However, the specificity of the CAD system was significantly higher in patients aged < 60 years than in those aged ≥ 60 years (79.8% vs. 52.7%; $P < 0.001$;

Table 3 Performance of the computer-aided detection system in different subgroups

Patient groups	AUC	Sensitivity	Specificity
All patients (n = 525)	0.892 (0.851–0.926)	88.3% (113/128, 82.7–93.9%)	68.3% (271/397, 63.7%–72.8%)
Patients with chest CT scan (n = 122)	0.852 (0.776–0.909)	83.8% (57/68, 75.1–92.6%)	59.3% (32/54, 46.2–72.4%)
Patients without chest CT scan (n = 403)	0.917 (0.886–0.942)	93.3% (56/60, 87.0–99.6%)	69.7% (239/343, 64.8–74.5%)
<i>P</i> value ^a	0.099	0.095	0.126
Men (n = 291)	0.892 (0.851–0.926)	88.9% (64/72, 81.6–96.1%)	69.4% (152/219, 63.3–75.5%)
Women (n = 234)	0.899 (0.853–0.934)	87.5% (49/56, 78.8–96.2%)	66.9% (119/178, 59.9–73.8%)
<i>P</i> value ^b	0.854	0.809	0.587
Age < 60 years (n = 286)	0.906 (0.866–0.937)	86.2% (50/58, 77.3–95.1%)	79.8% (182/228, 74.6–85.0%)
Age ≥ 60 years (n = 239)	0.867 (0.818–0.908)	90.0% (63/70, 83.0–97.0%)	52.7% (89/169, 45.1–60.2%)
<i>P</i> value ^c	0.300	0.507	< 0.001
CXRs from a fixed scanner (n = 146)	0.899 (0.838–0.943)	87.5% (21/24, 74.3–100%)	79.5% (97/122, 72.3–86.7%)
CXRs from a portable scanner (n = 379)	0.889 (0.853–0.919)	88.5% (92/104, 82.3–94.6%)	63.3% (174/275, 57.3–68.9%)
<i>P</i> value ^d	0.831	0.895	0.001

Numbers in parentheses indicate numerators/denominators, 95% confidence intervals

AUC, area under the receiver operating characteristic curve; CT, computed tomography; CXR, chest X-ray

^a Comparison of performance between patients with and without a chest CT scan

^b Comparison of performance between men and women

^c Comparison of performance between patients aged < 60 years and those aged ≥ 60 years

^d Comparison of performance between CXRs from a fixed scanner and CXRs from a portable scanner

C), and CXRs from a fixed radiography system than CXRs from a portable radiography scanner (79.5% vs. 63.1%; $P = 0.001$; D).

Discussion

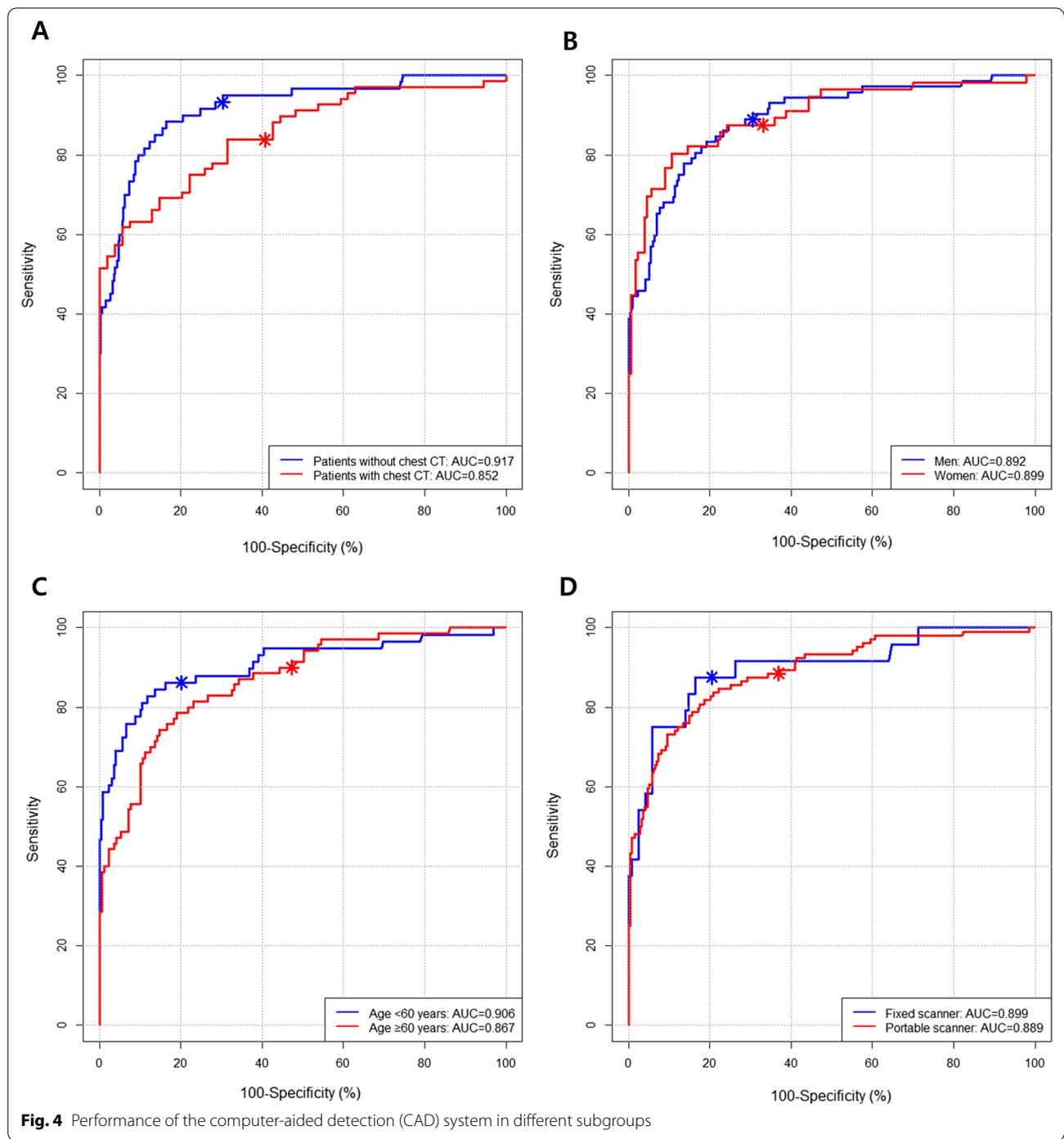
In our study, the deep learning-based CAD system showed good performance (sensitivity, 88.3%; specificity, 68.3%) in the identification of CXRs with pneumonia in consecutive febrile neutropenia patients. When compared with radiologists, the CAD system exhibited significantly higher sensitivity (86.6% vs. 75.2%) but lower specificity (64.8% vs. 75.4%). In the interpretation assisted by the CAD system, radiologists exhibited improvement in sensitivity (75.2% to 79.4%) at similar specificity (75.4% to 76.8%).

Symptoms suggestive of respiratory infection and pulmonary infiltrates on CXR or CT are the two main pillars for diagnosing pneumonia [23]. In patients with neutropenia, however, symptoms of pneumonia may not be apparent due to their compromised host immune responses [8, 11]; thus, diagnosis of pneumonia through imaging can be particularly critical [11, 12]. However, CXRs are reported to have limited diagnostic performance for pneumonia in FN patients [14, 15]. Furthermore, interpretation of CXR by an expert radiologist may not be readily available for timely management. In this regard, CAD may help in the accurate and timely diagnosis of pneumonia in these patients. Indeed, CAD used in our study could identify pneumonia on CXRs with higher

sensitivity than radiologists (higher sensitivity than four of five radiologists at the pre-defined threshold). High sensitivity of the CAD in a consecutively-collected CXRs reflecting the prevalence of pneumonia and the spectrum of CXR findings of the actual clinical situation suggests this CAD can accurately identify pneumonia among CXRs of febrile neutropenia patients in the situation of daily practice.

Contrarily, CAD showed lower specificity than radiologist at the predefined threshold. The results suggest that the standalone interpretation by CAD without modification of the threshold may produce more false-positive results than radiologists' interpretation. However, sensitivity may have priority over specificity for diagnosing pneumonia in FN patients, considering that pneumonia is frequently overlooked on CXR [11, 12, 26] and is associated with poor prognosis in FN [2, 3]. At the same sensitivity level as that of radiologists' interpretation, CAD exhibited higher specificity than most radiologists (four of five). The result suggests that adjustment of the threshold for positive results can improve specificity while maintaining radiologist-level sensitivity or higher.

Calibration of a deep learning-based algorithm is an important factor for its interpretability and credibility, but this aspect is frequently overlooked [27]. In our study, we found that CAD tended to overestimate the probability of pneumonia (Fig. 2B). The same CAD system exhibited a similar tendency of overestimation for the identification of referable abnormalities in CXRs



obtained from patients in the emergency department [28]. Recalibration of the CAD result may correct the overestimation and improve the interpretability and credibility of CAD [28].

Assisting radiologists to enhance their performance is another key role of CAD [29]. In our study, average sensitivity of radiologists was significantly improved at similar

specificity in the interpretation with CAD. The results suggest that CAD can help in improving the quality of radiologists' interpretation and management of patients. Preserved specificity while improving sensitivity suggests that false-positive detection by CAD can be appropriately arbitrated by radiologists. Improvement of inter-reader agreement in the interpretation with CAD suggests that

the CAD can also reduce variability across radiologists in CXR interpretations. Prioritization of CXRs with high suspicion for critical finding is another potential application of CAD [30]. The relatively clear separation of CAD results between CXRs with and without pneumonia (Fig. 2C) suggests that such prioritization of CXR requiring early interpretation by a radiologist might be feasible with CAD. However, our retrospective study was not conducted for that purpose; further studies on this topic are warranted in the near future.

In patients with available chest CT as a firm reference standard for pneumonic infiltrates, CAD exhibited similar performance as that reported in a previous study, where the same CAD system was evaluated in patients suspected of having coronavirus disease (sensitivity 83.8% and specificity 59.3% in our study; sensitivity 81.5% and specificity 52.3% in the previous study) [31], suggesting the robustness of performance regardless of population characteristics. CAD exhibited slightly lower performances in patients who underwent chest CT than in patients who did not undergo chest CT although significant difference was not shown. This result is probably because chest CTs were obtained in patients with a questionable diagnosis of pneumonia or other causes of fever [1, 11], while patients with a definite diagnosis of pneumonia on CXRs or those with clearly normal CXRs and no clinical suspicion of pneumonia may not have undergone CT.

Our study also showed that CAD missed a substantial proportion of pulmonary infiltrates (16.2%) on CXRs, which could be identified on CT. Although sensitivity improvement occurred in some radiologists, they remained below 90%. Therefore, the negative results by the CAD or radiologist may not necessarily indicate the absence of pneumonia, and CAD cannot be a substitute for chest CT. Chest CT should be performed in patients with persistent fever despite empirical antibiotics, or those with clinical suspicions of pneumonia but negative CXRs [1, 11, 32, 33].

Consistency of performance in patients with different characteristics is an important factor for the reliability of CAD [29]. In our study, CAD exhibited similar performances between men and women. However, it exhibited lower specificities in older patients (age ≥ 60 years) than in younger patients, and CXRs obtained using a portable scanner than those obtained using a fixed radiography unit. Suboptimal image qualities of portable radiography and patient-related artifacts (e.g., limited inspiration and inappropriate positioning of patients) might have led to lower specificities. Previous studies reported that the identification of pulmonary infiltrates and differentiation from pleural effusion or atelectasis was difficult for human readers [34, 35]. A higher prevalence of existing

pulmonary abnormalities (e.g., scarring, interstitial fibrosis, and emphysema) may also have contributed to more false-positive detection by CAD. A higher rate of false-positive detection should be considered when using CAD in clinical practice.

Our study has several limitations. First, because CAD was validated on a retrospectively collected dataset, the effect of CAD on the management and outcome of patients could not be evaluated. Second, since the reader test was conducted outside the daily practice, the reading behaviors of radiologists may be different from those in their daily practices. Third, since our study was performed in a single tertiary medical center, the generalizability of the results remains questionable. Finally, interpreting same CXRs twice in the reader test may have induced improvement of radiologists' performance in the second reading session, regardless of using the CAD. Therefore, it cannot be excluded that the added value of the CAD has been overestimated.

Conclusions

In conclusion, the deep learning-based CAD system exhibited radiologist-level performance in the identification of CXRs with pneumonia in consecutive FN patients. It enhanced the performance of radiologists for the identification of pneumonia. We believe that CAD system helps diagnose and manage pneumonia in FN patients.

Abbreviations

ANC: Absolute neutrophil count; AUC: Area under receiver operating characteristic curve; CAD: Computer-aided detection; CI: Confidence interval; CT: Computed tomography; CXR: Chest X-ray; FN: Febrile neutropenia; IQR: Interquartile range.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12890-021-01768-0>.

Additional file 1. Supplementary information. Development and function of the computer-aided detection system. **Figure S1.** Example input and output of the computer-aided detection system. **Table S1.** Performance of the computer-aided detection system at different thresholds. **Table S2.** Performances of the computer-aided detection system and radiologists in the reader test.

Additional file 2. Raw data analyzed in the study.

Acknowledgements

Not applicable.

Authors' contributions

EJH and CMP conceptualized and designed the study. EJH, JHL, JHK, WHL collected clinical data. EJH performed statistical analyses. EJH and CMP drafted the manuscript. All authors revised and approved the final manuscript.

Funding

This work was supported by the Seoul National University Hospital Research Fund [Grant Numbers 06-2016-3000, 03-2019-0190]. The funding source had no role in the design of the study, collection, analysis, and interpretation of data, and in writing the manuscript.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary material, except for chest X-ray images. Chest X-ray image data analyzed during the current study are not publicly available due to private health information policy of the institution but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The present study was approved by the institutional review board of Seoul National University Hospital. The requirement for patients' informed consents was waived by the institutional review board of Seoul National University Hospital. The present study was carried out in accordance with the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

EJH reports research grant from Lunit Inc., outside the present study. JHL reports no competing interest. JHK reports no competing interest. WHL reports no competing interest. JMG reports no competing interest. CMP reports research grant from Lunit Inc., outside the present study, and holds stock options of Lunit Inc. and Coreline Soft.

Author details

¹Department of Radiology, Seoul National University College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul 03080, Korea. ²Namwon Medical Center, 365 Chungjeong-ro, Namwon 55726, Jeollabuk-do, Korea. ³Institute of Medical and Biological Engineering, Medical Research Center, Seoul National University, 103 Daehak-ro, Jongno-gu, Seoul 03080, Korea.

Received: 17 March 2021 Accepted: 19 November 2021

Published online: 07 December 2021

References

- Klastersky J, de Naurois J, Rolston K, Rapoport B, Maschmeyer G, Aapro M, Herrstedt J, Committee EG. Management of febrile neutropenia: ESMO Clinical Practice Guidelines. *Ann Oncol*. 2016;27(suppl 5):v111–8.
- Chindaprasit J, Wanitpongpun C, Limpawattana P, Thepsuthammarat K, Sriprakdee W, Sookprasert A, Wirasorn K. Mortality, length of stay, and cost associated with hospitalized adult cancer patients with febrile neutropenia. *Asian Pac J Cancer Prev*. 2013;14(2):1115–9.
- Kuderer NM, Dale DC, Crawford J, Cosler LE, Lyman GH. Mortality, morbidity, and cost associated with febrile neutropenia in adult cancer patients. *Cancer*. 2006;106(10):2258–66.
- Crawford J, Dale DC, Lyman GH. Chemotherapy-induced neutropenia: risks, consequences, and new directions for its management. *Cancer*. 2004;100(2):228–37.
- Taplitz RA, Kennedy EB, Bow EJ, Crews J, Gleason C, Hawley DK, Langston AA, Nastoupil LJ, Rajotte M, Rolston K, et al. Outpatient management of fever and neutropenia in adults treated for malignancy: American Society of Clinical Oncology and Infectious Diseases Society of America Clinical Practice Guideline Update. *J Clin Oncol*. 2018;36(14):1443–53.
- Yoshida M, Akiyama N, Fujita H, Miura K, Miyatake J, Handa H, Kito K, Takahashi M, Shigeno K, Kanda Y, et al. Analysis of bacteremia/fungemia and pneumonia accompanying acute myelogenous leukemia from 1987 to 2001 in the Japan Adult Leukemia Study Group. *Int J Hematol*. 2011;93(1):66–73.
- Lanoix JP, Schmit JL, Douadi Y. Bacterial lung sepsis in patients with febrile neutropenia. *Curr Opin Pulm Med*. 2012;18(3):175–80.
- Evans SE, Ost DE. Pneumonia in the neutropenic cancer patient. *Curr Opin Pulm Med*. 2015;21(3):260–71.
- Maschmeyer G, Link H, Hiddemann W, Meyer P, Helmerking M, Eisenmann E, Schmitt J, Adam D. Pulmonary infiltrations in febrile patients with neutropenia. Risk factors and outcome under empirical antimicrobial therapy in a randomized multicenter study. *Cancer*. 1994;73(9):2296–304.
- Turkoglu M, Mirza E, Tunccan OG, Erdem GU, Dizbay M, Yagci M, Aygenel G, Turkoz Sucak G. Acinetobacter baumannii infection in patients with hematologic malignancies in intensive care unit: risk factors and impact on mortality. *J Crit Care*. 2011;26(5):460–7.
- Heussel CP, Kauczor HU, Ullmann AJ. Pneumonia in neutropenic patients. *Eur Radiol*. 2004;14(2):256–71.
- Heussel CP. Importance of pulmonary imaging diagnostics in the management of febrile neutropenic patients. *Mycoses*. 2011;54(Suppl 1):17–26.
- Freifeld AG, Bow EJ, Sepkowitz KA, Boeckh MJ, Ito JI, Mullen CA, Raad II, Rolston KV, Young JA, Wingard JR, et al. Clinical practice guideline for the use of antimicrobial agents in neutropenic patients with cancer: 2010 update by the infectious diseases society of America. *Clin Infect Dis*. 2011;52(4):e56–93.
- Heussel CP, Kauczor HU, Heussel GE, Fischer B, Begrich M, Mildnerberger P, Thelen M. Pneumonia in febrile neutropenic patients and in bone marrow and blood stem-cell transplant recipients: use of high-resolution computed tomography. *J Clin Oncol*. 1999;17(3):796–805.
- Yolin-Raley DS, Dagogo-Jack I, Niell HB, Soiffer RJ, Antin JH, Alyea EP 3rd, Glotzbecker BE. The utility of routine chest radiography in the initial evaluation of adult patients with febrile neutropenia patients undergoing HSCT. *J Natl Compr Canc Netw*. 2015;13(2):184–9.
- Zornoza J, Goldman AM, Wallace S, Valdivieso M, Bodey GP. Radiologic features of gram-negative pneumonias in the neutropenic patient. *AJR Am J Roentgenol*. 1976;127(6):989–96.
- Hopstaken RM, Witbraad T, van Engelshoven JM, Dinant GJ. Interobserver variation in the interpretation of chest radiographs for pneumonia in community-acquired lower respiratory tract infections. *Clin Radiol*. 2004;59(8):743–52.
- Melbye H, Dale K. Interobserver variability in the radiographic diagnosis of adult outpatient pneumonia. *Acta Radiol*. 1992;33(1):79–81.
- Hashmi MF, Katiyar S, Keskar AG, Bokde ND, Geem ZW. Efficient pneumonia detection in chest X-ray images using deep transfer learning. *Diagnostics (Basel)*. 2020;10(6):417.
- Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, McKeown A, Yang G, Wu X, Yan F, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*. 2018;172(5):1122–31.
- Hwang EJ, Park S, Jin KN, Kim JI, Choi SY, Lee JH, Goo JM, Aum J, Yim JJ, Cohen JG, 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(3):e191095.
- Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS Med*. 2018;15(11):e1002683.
- Metlay JP, Waterer GW, Long AC, Anzueto A, Brozek J, Crothers K, Cooley LA, Dean NC, Fine MJ, Flanders SA, et al. Diagnosis and treatment of adults with community-acquired pneumonia. An official clinical practice guideline of the American Thoracic Society and Infectious Diseases Society of America. *Am J Respir Crit Care Med*. 2019;200(7):e45–67.
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a non-parametric approach. *Biometrics*. 1988;44(3):837–45.
- Genders TS, Spronk S, Stijnen T, Steyerberg EW, Lesaffre E, Hunink MG. Methods for calculating sensitivity and specificity of clustered data: a tutorial. *Radiology*. 2012;265(3):910–6.
- Maschmeyer G. Pneumonia in febrile neutropenic patients: radiologic diagnosis. *Curr Opin Oncol*. 2001;13(4):229–35.
- Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology*. 2018;286(3):800–9.

28. Hwang EJ, Kim H, Lee JH, Goo JM, Park CM. Automated identification of chest radiographs with referable abnormality with deep learning: need for recalibration. *Eur Radiol.* 2020;30(12):6902–12.
29. Petrick N, Sahiner B, Armato SG 3rd, Bert A, Correale L, Delsanto S, Freedman MT, Fryd D, Gur D, Hadjiiski L, et al. Evaluation of computer-aided detection and diagnosis systems. *Med Phys.* 2013;40(8):087001.
30. Hwang EJ, Park CM. Clinical implementation of deep learning in thoracic radiology: potential applications and challenges. *Korean J Radiol.* 2020;21(5):511–25.
31. Hwang EJ, Kim H, Yoon SH, Goo JM, Park CM. Implementation of a deep learning-based computer-aided detection system for the interpretation of chest radiographs in patients suspected for COVID-19. *Korean J Radiol.* 2020;21(10):1150–60.
32. Gerritsen MG, Willeminck MJ, Pompe E, van der Bruggen T, van Rhenen A, Lammers JW, Wessels F, Sprengers RW, de Jong PA, Minnema MC. Improving early diagnosis of pulmonary infections in patients with febrile neutropenia using low-dose chest computed tomography. *PLoS ONE.* 2017;12(2):e0172256.
33. Patsios D, Maimon N, Chung T, Roberts H, Disperati P, Minden M, Paul N. Chest low-dose computed tomography in neutropenic acute myeloid leukaemia patients. *Respir Med.* 2010;104(4):600–5.
34. Kunz WG, Patzig M, Crispin A, Stahl R, Reiser MF, Notohamiprodjo M. The value of supine chest X-ray in the diagnosis of pneumonia in the basal lung zones. *Acad Radiol.* 2018;25(10):1252–6.
35. Kitazono MT, Lau CT, Parada AN, Renjen P, Miller WT Jr. Differentiation of pleural effusions from parenchymal opacities: accuracy of bedside chest radiography. *AJR Am J Roentgenol.* 2010;194(2):407–12.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

