Artificial intelligence system for identification of overlooked lung metastasis in abdominopelvic computed tomography scans of patients with malignancy

Hye Soo Cho[1]
Eui Jin Hwang[1,2]
Jaeyoun Yi[3]
Boorym Choi[3]
Chang Min Park[1,2]

[1]Seoul National University Hospital, Seoul National University College of Medicine, Department of Radiology, Seoul, Republic of Korea

[2]Seoul National University College of Medicine, Department of Radiology, Seoul, Republic of Korea

[3]Coreline Soft Inc., Seoul, Republic of Korea

PURPOSE

This study aimed to evaluate whether an artificial intelligence (AI) system can identify basal lung metastatic nodules examined using abdominopelvic computed tomography (CT) that were initially overlooked by radiologists.

METHODS

We retrospectively included abdominopelvic CT images with the following inclusion criteria: a) CT images from patients with solid organ malignancies between March 1 and March 31, 2019, in a single institution; and b) abdominal CT images interpreted as negative for basal lung metastases. Reference standards for diagnosis of lung metastases were confirmed by reviewing medical records and subsequent CT images. An AI system that could automatically detect lung nodules on CT images was applied retrospectively. A radiologist reviewed the AI detection results to classify them as lesions with the possibility of metastasis or clearly benign. The performance of the initial AI results and the radiologist's review of the AI results were evaluated using patient-level and lesion-level sensitivities, false-positive rates, and the number of false-positive lesions per patient.

RESULTS

A total of 878 patients (580 men; mean age, 63 years) were included, with overlooked basal lung metastases confirmed in 13 patients (1.5%). The AI exhibited an area under the receiver operating characteristic curve value of 0.911 for the identification of overlooked basal lung metastases. Patient- and lesion-level sensitivities of the AI system ranged from 69.2% to 92.3% and 46.2% to 92.3%, respectively. After a radiologist reviewed the AI results, the sensitivity remained unchanged. The false-positive rate and number of false-positive lesions per patient ranged from 5.8% to 27.6% and 0.1% to 0.5%, respectively. Radiologist reviews significantly reduced the false-positive rate (2.4%–12.6%; all P values < 0.001) and the number of false-positive lesions detected per patient (0.03–0.20, respectively).

CONCLUSION

The AI system could accurately identify basal lung metastases detected in abdominopelvic CT images that were overlooked by radiologists, suggesting its potential as a tool for radiologist interpretation.

CLINICAL SIGNIFICANCE

The AI system can identify missed basal lung lesions in abdominopelvic CT scans in patients with malignancy, providing feedback to radiologists, which can reduce the risk of missing basal lung metastasis.

KEYWORDS

Artificial intelligence, computed tomography, image analysis, metastasis, radiology

Corresponding author: Eui Jin Hwang

E-mail: ken921004@hotmail.com

Abdominopelvic computed tomography (CT) is frequently performed in patients with cancer to evaluate various cancers of the abdominopelvic or extra-abdominopelvic organs. Lung metastasis frequently occurs in the advanced stages of various solid organ cancers, and abdominopelvic CT images inevitably capture the base of the lungs. Therefore, evaluating the presence of nodules suggestive of metastasis to the lung base is an important component in the interpretation of abdominopelvic CT scans in patients.[1-4] However, in a busy clinical environment, a radiologist may pay relatively less attention to the basal lungs compared with the abdominal organs, which are the main targets of evaluation.[4] Therefore, metastatic nodules in the basal lungs can be overlooked by interpreting radiologists, which may adversely affect a patient's treatment policy decisions or prognosis, leading to a medicolegal dispute.

The automatic detection of pulmonary nodules on chest CT images is one of the most widely investigated topics in artificial intelligence (AI)-based medical image analysis. Various studies have reported the radiologist-level performance of AI and the enhanced performance of radiologists using AI for lung nodule detection on CT scans.[5,6] Based on these impressive results, commercial AI-based software medical devices have begun to be utilized in daily clinical practice as computer-aided detection (CAD) tools.[7-10]

In addition to its use as a CAD tool, AI's utilization as a second reader–that is, for use in analyzing images after the radiologist's interpretation and providing feedback to the radiologist in case of suspected interpretation errors–can be another attractive scenario for applying AI in daily practice.[11-15] AI, as a second reader, can provide a safety net for radiologists against the risk of interpretation errors or medicolegal disputes without requiring the rigorous effort of scrutinizing the AI's results following every examination. The detection of pulmonary nodules in the basal lungs, as acquired using abdominopelvic CT, can serve as a compelling scenario for employing an AI second reader.[15] This is because it is beyond the primary focus of examination, yet carries a relatively high risk of interpretation errors, which could result in critical outcomes.

In consideration of the above, we aim to evaluate whether an AI system could detect metastatic pulmonary nodules in the basal lungs that have been overlooked by radiologists on abdominopelvic CT images.

Methods

This single-center, retrospective, diagnostic cohort study was approved by the Seoul National University Hospital Institutional Review Board on January 5, 2022 (approval number: 2112-142-1284). During the approved research period, patient data required for this study were accessed for research purposes. The requirement for informed consent was waived by the institutional review board.

Patients

Patients were consecutively included in a single tertiary referral institution in South Korea with the following criteria: a) patients diagnosed with solid organ cancers (International Statistical Classification of Diseases and Related Health Problems, 10th revision, C00 to C75); b) patients who underwent abdominopelvic CT between March 1 and March 31, 2019; and c) abdominopelvic CT scans interpreted as negative for basal lung metastasis in the formal reports of radiologists, based on a manual review of unstructured radiological reports by a thoracic radiologist. Patients who underwent chest CT on the same day as abdominopelvic CT and those lost to follow-up within 3 years without a clinical diagnosis of lung metastasis were excluded (Figure 1).

The first CT examination was performed on patients who underwent CT more than once. For multiphase CT examinations, images that captured the largest portion of the basal lungs were included in the analyses.

Diagnosis of pulmonary metastasis

To confirm the clinical diagnosis of pulmonary metastasis in patients, one thoracic radiologist (E.J.H., with 5 years of experience as a faculty thoracic radiologist) reviewed the medical records and CT images (including the index abdominopelvic CT and follow-up chest and abdominopelvic CT images). Pulmonary lesions that were pathologically confirmed as metastases, as well as lesions with persistent growth on follow-up CT images and a clinical impression of metastasis, were regarded as pulmonary metastases. Pulmo-

Main points

- An artificial intelligence (AI) system for pulmonary nodule detection on computed tomography (CT) images can be utilized as a second reader after the radiologist's interpretation, to identify overlooked pulmonary nodules.

- As a second reader, the AI may analyze images after the radiologist's interpretation and provide feedback to the radiologist only when the AI suspects that the radiologist has overlooked a pulmonary nodule. In this scenario, the oversight of significant pulmonary nodules can be prevented without the need to review the AI results of all the examinations. In our study, the applied AI system could accurately identify basal lung metastases captured in abdominopelvic CT images that were overlooked by radiologists, suggesting its potential as a second reader after the radiologist's interpretation.

- We believe that our study contributes significantly to the literature by highlighting the effectiveness of AI in improving the accuracy of interpreting abdominopelvic CT images in patients with malignancies. Additionally, it underscores the importance of AI as a second reader to reduce interpretation errors.
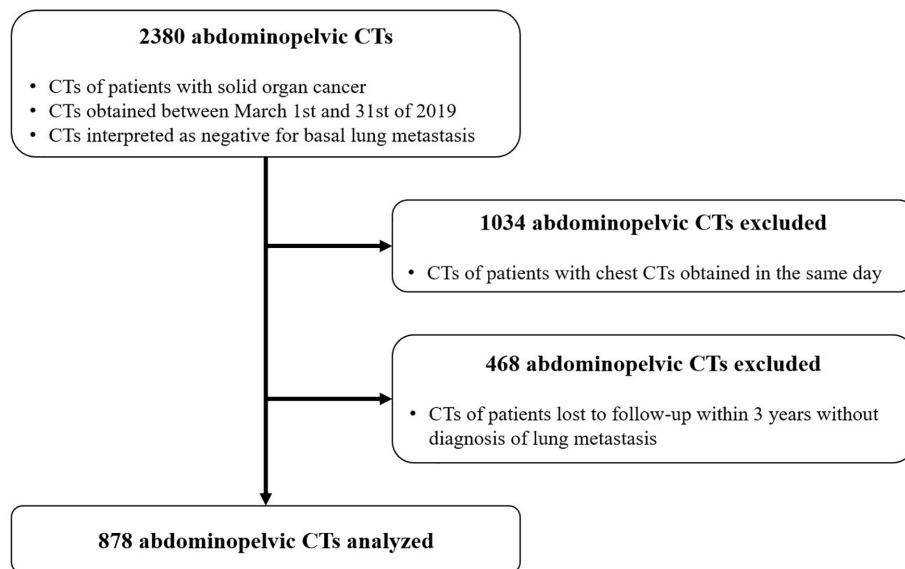

2380 abdominopelvic CTs

- CTs of patients with solid organ cancer
- CTs obtained between March 1st and 31st of 2019
- CTs interpreted as negative for basal lung metastasis

1034 abdominopelvic CTs excluded

- CTs of patients with chest CTs obtained in the same day

468 abdominopelvic CTs excluded

- CTs of patients lost to follow-up within 3 years without diagnosis of lung metastasis

878 abdominopelvic CTs analyzed

Figure 1. Flow diagram of the study. CT, computed tomography.

nary lesions that were stable for >3 years were considered benign. All individual pulmonary metastases present on the index abdominopelvic CT images but not documented in the radiologist's report were recorded as "overlooked metastases."

Artificial intelligence system

To detect pulmonary metastases in the basal lungs captured by abdominopelvic CT, an AI model based on a commercialized deep-learning-based CAD system (AVIEW Lung Nodule CAD, Coreline Soft, Seoul, Korea) was used. The CAD system was designed to detect pulmonary nodules in chest CT images and was approved for clinical use in Korea as an assistant tool for physicians in interpreting chest CT scans.

Since the original CAD system was optimized for low-dose chest CT images for lung cancer screening, the performance of the AI model may degrade when used for detecting pulmonary metastasis. Therefore, additional training of the AI model was conducted to optimize its performance in detecting small metastatic pulmonary nodules. A total of 3,558 CT scans were conducted, with 21,469 clinically diagnosed pulmonary metastases from a single institution (the same institution as where the present study was conducted). All of the abdominopelvic CT images were analyzed using an additionally trained AI model. Each pulmonary nodule was annotated by drawing three-dimensional bounding boxes on the CT images, along with a probability score (between 0 and 1) for the presence of a lesion (Figures 2-5). Then, these annotated CT images were used for the existing AI model for the original CAD system.

## Radiologist's evaluation of artificial intelligence findings

All abdominopelvic CT images with corresponding AI results were reviewed by a fellowship trainee in thoracic radiology (H.S.C., 1st year of fellowship training) who was blinded to the diagnosis of pulmonary metastasis. The radiologist classified all lesions identified by the AI into three groups: those with the potential for pulmonary metastasis, clearly benign lesions, and pseudo-lesions. Subsequently, the radiologist checked the diagnoses of pulmonary metastasis to confirm that the lesions detected by AI were overlooked pulmonary metastases and classified the individual AI-detected lesions as either true-positives or false-positives.
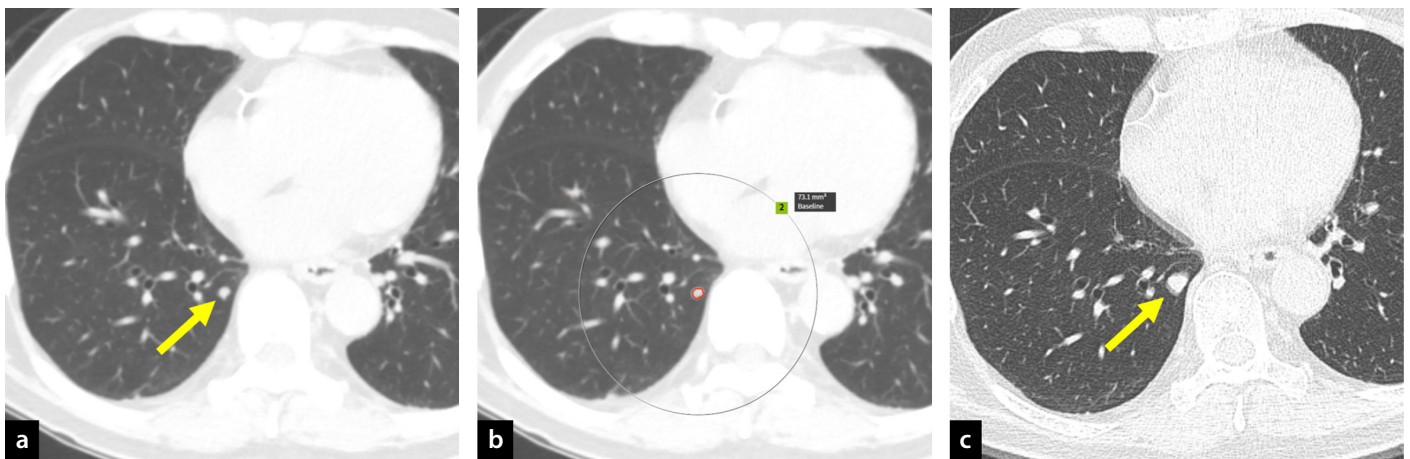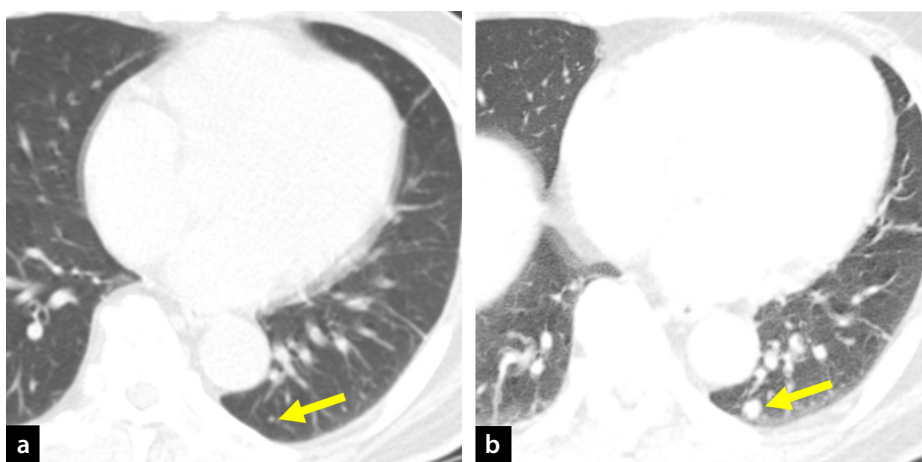

Figure 2. (a) This abdominal computed tomography (CT) image of a 72-year-old male patient with hepatocellular carcinoma shows a small nodule in the right lower lobe (arrow), a feature that was overlooked in the initial interpretation. **(b)** The artificial intelligence system detected the nodule with a probability score of 0.79. **(c)** A chest CT image obtained 167 days later shows growth of the nodule (arrow), which was clinically diagnosed as metastasis.
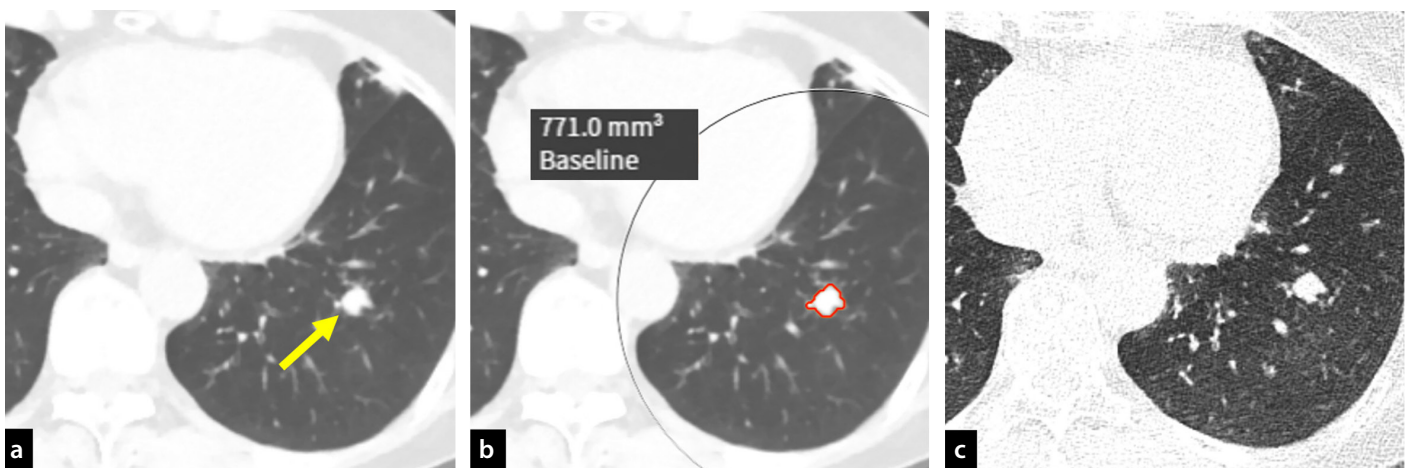
Figure 3. (a) This abdominopelvic computed tomography (CT) image of a 65-year-old male patient with colon cancer shows a nodule in the left lower lobe (arrow), a feature that was overlooked in the initial interpretation. **(b)** The artificial intelligence system detected the nodule with a probability score of 0.87. The radiologist who reviewed the AI results interpreted the lesion as a true nodule with the possibility of metastasis. **(c)** A chest CT image obtained 28 months later shows that the lesion remains unchanged, suggesting benignancy. AI, artificial intelligence.
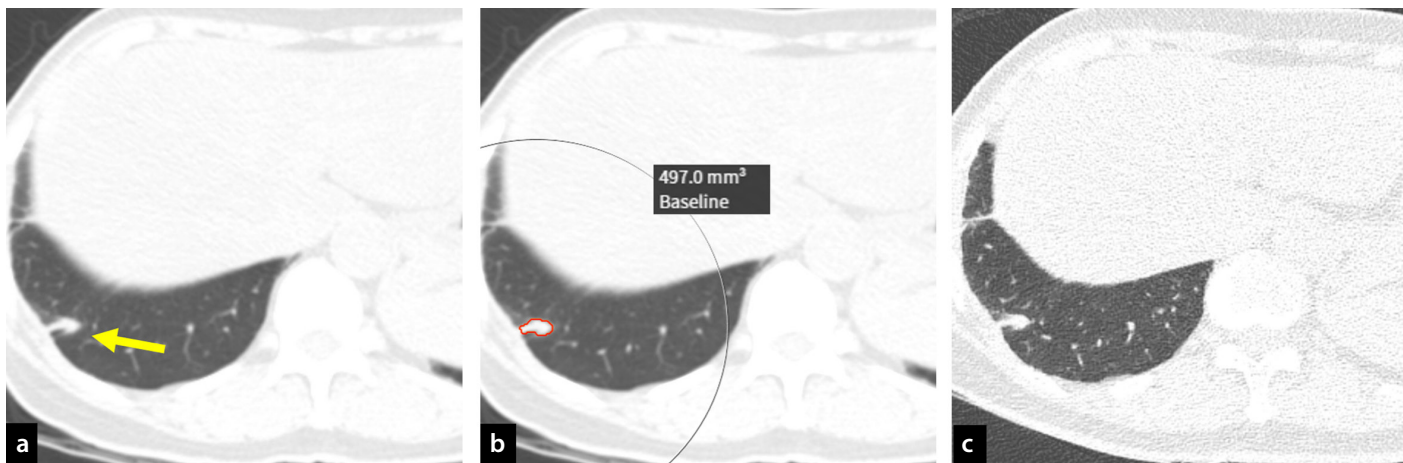
Figure 4. (a) This abdominopelvic computed tomography (CT) image of a 58-year-old female patient with colon cancer shows a nodular lesion in the right lower lobe (arrow), an observation that was not described in the initial interpretation. **(b)** The artificial intelligence (AI) system identified the lesion with a probability score of 0.50. The radiologist who reviewed the AI results interpreted the lesion as focal atelectasis rather than a true nodule. **(c)** A chest CT image obtained 44 months later shows that the lesion remained stable, suggesting benignancy.

Figure 5. (a) This abdominopelvic computed tomography (CT) image of a 72-year-old male patient with colon cancer shows a tiny nodule in the left lower lobe (arrow), an observation that was overlooked in the initial interpretation. The artificial intelligence system did not detect the lesion. **(b)** A chest CT image obtained 204 days later shows the growth of the nodule, suggesting a diagnosis of lung metastasis (arrow).

Performance metrics and statistical analysis

First, the discriminative performance of the AI model in identifying patients with overlooked metastases was evaluated using an area under the receiver operating characteristic curve (AUC-ROC) analysis. Subsequently, the performance and efficacy of the AI model were evaluated using metrics at threshold probability scores of 0.4, 0.5, 0.6, and 0.7.

• Patient-level sensitivity = number of patients with true-positive detection of overlooked metastases/number of patients with overlooked metastases.

• Patient-level false-positive rate = number of patients with false-positive detection of overlooked metastases/number of patients without overlooked metastases.

• Patient-level positive predictive value (PPV) = number of patients with true-positive detection of overlooked metastases/number of patients with positive AI results.

• Lesion-level sensitivity = number of true-positive detections of overlooked metastases/number of all overlooked metastases.

• Number of false-positive lesions per patient = number of false-positive detections of overlooked metastases/number of patients.

• Lesion-level PPV = number of true-positive detections of overlooked metastases/number of all lesions detected by AI.

All metrics were obtained for both the AI results and the radiologist's review of the AI results (following the exclusion of clearly benign lesions or pseudo-lesions). The performance metrics of the AI results and the radiologist's review of the AI results were compared using McNemar's tests, chi-squared tests, and paired t-tests.

Decision curve analysis was conducted to evaluate the net benefit of using the AI tool as a second reader for detecting overlooked pulmonary metastasis, considering the benefit of true-positive results and the cost of false-positive results.

Statistical analysis

All statistical analyses were performed using MedCalc statistical software (MedCalc Software Ltd, Ostend, Belgium, 22.006 version). Statistical significance was set at $P < 0.05$.

Results

Patient characteristics

A total of 878 abdominopelvic CT images from 878 patients (580 men; mean age ± standard deviation: 62 ± 11 years) were included in the study (Figure 1). The most common primary malignancy was hepatocellular carcinoma (411, 47%), followed by stomach cancer (169, 19%) and colorectal cancer (96, 11%). A total of 707 CT examinations (81%) were obtained after the administration of intravenous contrast media. Table 1 presents the demographic information of the patients and their CT imaging characteristics.

Sixty-nine (7.8%) patients were diagnosed with lung metastases within 3 years of an abdominopelvic CT, including 5 patients who had already been diagnosed with lung metastases at the time of the CT. In a retrospective evaluation of abdominopelvic CT

Table 1. Patients and computed tomography characteristics

Variables	All patients (n = 878)	Patients with overlooked lung metastases (n = 13)	Patients without overlooked lung metastases (n = 865)
Age, mean ± SD (years)	62 ± 11	65 ± 13	62 ± 11
Male-to-female patient ratio	580:298	7:6	573:292
Primary malignancy			
Hepatocellular carcinoma	411 (47%)	3 (25%)	408 (47%)
Stomach cancer	169 (19%)	1 (8%)	168 (19%)
Colorectal cancer	96 (11%)	2 (15%)	94 (11%)
Biliary tree or pancreatic cancer	46 (5%)	5 (42%)	41 (5%)
Uterus or ovary cancer	25 (3%)	0	25 (3%)
Urinary tract cancer	16 (2%)	1 (8%)	15 (2%)
Breast cancer	15 (2%)	1 (8%)	14 (2%)
Prostate cancer	9 (1%)	0	9 (1%)
Others	91 (10%)	0	91 (11%)
CT examination with intravenous contrast media	844 (96%)	10 (83%)	834 (96%)
Multiphase CT examination	707 (81%)	9 (75%)	699 (81%)

Numbers in parentheses indicate proportions among all patients. CT, computed tomography; SD, standard deviation.
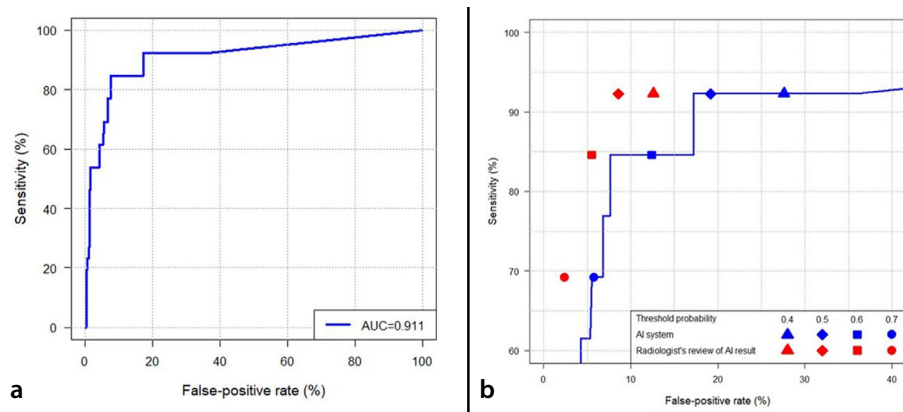
Figure 6. Receiver operating characteristic (ROC) curve for the identification of abdominopelvic computed tomography (CT) scans with overlooked basal lung metastases. **(a)** A ROC curve show that the artificial intelligence (AI) system identified abdominopelvic CT scans with overlooked basal lung metastases with an area under the ROC curve of 0.911. **(b)** The modified ROC curve shows that the sensitivity and false-positive rate of the AI system ranged from 69.2–92.3% and 46.2–92.3%, respectively at thresholds between 0.4 and 0.7. The radiologist's review significantly reduced the false-positive rate (2.4–12.6%) while preserving the sensitivity. AUC, area under the curve.

images, 13 (1.5%) patients had pulmonary metastases that were overlooked during interpretation. Of these 13 patients, 3 had already been diagnosed with lung metastases at the time of the CT. For the other 10 patients, the time interval between the abdominopelvic CT with overlooked lung metastases and the clinical diagnosis of lung metastasis was 141 days (interquartile range, 78–195 days).

Performance of the artificial intelligence system

For the discrimination of CT examinations with and without overlooked pulmonary metastases, the AI system exhibited an AUC-ROC value of 0.911 [95% confidence interval (CI), 0.890–0.929; Figure 6]. The results of the AI analyses and their performances for different thresholds are listed in Table 2 and Table 3. At the lowest threshold (0.4), the AI system detected 475 lesions (0.54 per examination) in 251 patients (positive rate, 28.7%). In contrast, it detected 100 lesions (0.11 per examination) in 59 (positive rate, 6.7%) patients at the highest threshold (0.7). The sensitivities of the AI system for the identification of patients with overlooked metastases were 92.3% (12/13; 95% CI, 64.0%–99.8%) at the lowest threshold and 69.2% (9/13; 95%

CI, 38.6%–90.9%) at the highest threshold. Correspondingly, the patient-level false-positive rates ranged from 5.8% (50/865; 95% CI, 4.3%–7.6%) to 27.6% (239/865; 95% CI, 24.7%–30.7%), and the PPVs ranged from 4.8% (12/251; 95% CI, 2.5%–8.2%) to 15.3% (9/59; 95% CI, 7.2%–27.0%). The accuracy of the AI system ranged from 72.7% (638/878; 95% CI, 69.6%–75.6%) to 93.8% (824/878; 95% CI, 92.1%–95.4%).

Among 26 overlooked pulmonary metastases in eight patients, the sensitivities of the AI system were 92.3% (24/26; 95% CI, 74.5%–99.1%) at the lowest threshold and 46.2% (12/26; 95% CI, 26.6%–66.6%) at the highest. Correspondingly, the number of false-positive detections per examination ranged from 0.10 (88/878; 95% CI, 0.03–0.17) to 0.51 (451/878; 95% CI, 0.35–0.69), and the PPVs ranged from 5.1% (24/475; 95% CI, 3.3%–7.4%) to 12.0% (12/100; 95% CI, 6.4%–20.0%).

In the decision curve analysis, using the AI system as a second reader for detecting overlooked pulmonary metastases exhibited a higher net benefit than the default scenario without AI when the risk threshold was ≤3.7% (Figure 7). In other words, using the AI would be beneficial if the ratio of the cost from false-positive results to the benefit from true-positive results is ≤3.7:96.3 (1:26).

Review of the artificial intelligence results by the radiologist

Following the review of the AI results by the radiologist, 57.9% (275/475) of the

Table 2. Patient-level performance of the artificial intelligence system and the radiologist's review of the artificial intelligence results

Performance metric	Threshold probability 0.4	Threshold probability 0.5	Threshold probability 0.6	Threshold probability 0.7
AI system				
Number of true-positive results	12	12	11	9
Number of false-positive results	239	164	107	50
Number of true-negative results	1	1	2	4
Number of false-negative results	626	701	758	815
Positive rate	28.7% (251/878; 25.6%, 31.7%)	20.0% (176/878; 17.4, 22.9)	13.4% (118/878; 11.3, 15.9)	6.7% (59/878; 5.2, 8.6)
Number of detections per patient	0.54 (475/878; 0.37, 0.71)	0.36 (319/878; 0.22, 0.50)	0.23 (205/878; 0.12, 0.35)	0.11 (100/878; 0.05, 0.18)
Sensitivity	92.3% (12/13; 64.0, 99.8)	92.3% (12/13; 64.0, 99.8)	84.6% (11/13; 54.6, 98.1)	69.2% (9/13; 38.6, 90.9)
False-positive rate	27.6% (239/865; 24.7, 30.7)	19.2% (164/865; 16.4, 21.7)	12.4% (107/865; 10.3, 14.8)	5.8% (50/865; 4.3, 7.6)
PPV	4.8% (12/251; 2.5, 8.2)	6.8% (12/176; 3.6, 11.6)	9.3% (11/118; 4.8, 16.1)	15.3% (9/59; 7.2, 27.0)
Accuracy	72.7% (638/878; 69.6, 75.6)	81.2% (713/878; 78.5, 83.7)	87.6% (769/878; 85.2, 89.7)	93.8% (824/878; 92.1, 95.4)
Radiologist's review of the AI results				
Number of true-positive results	12	12	11	9
Number of false-positive results	109	74	48	21
Number of true-negative results	1	1	2	4
Number of false-negative results	756	791	817	844
Positive rate	13.8% (121/878; 11.6, 16.2)	9.8% (86/878; 7.9, 12.0)	6.7% (59/878; 5.2, 8.6)	3.4% (30/878; 2.3, 4.8)
P value	<0.001	<0.001	<0.001	<0.001
Number of detections per patient	0.23 (200/878; 0.18, 0.28)	0.15 (128/878; 0.11, 0.18)	0.09 (78/878; 0.06, 0.11)	0.04 (35/878; 0.02, 0.06)
P value	<0.001	<0.001	<0.001	<0.001
Sensitivity	92.3% (12/13; 64.0, 99.8)	92.3% (12/13; 64.0, 99.8)	84.6% (11/13; 54.6, 98.1)	69.2% (9/13; 38.6, 90.9)
P value	NA	NA	NA	NA
False-positive rate	12.6% (109/865; 10.5, 15.0)	8.6% (74/865; 6.8, 10.6)	5.5% (48/865; 4.1, 7.3)	2.4% (21/865; 1.5, 3.7)
P value	<0.001	<0.001	<0.001	<0.001
PPV	9.9% (12/121; 5.2, 16.7)	14.0% (12/86; 7.4, 23.1)	18.6% (11/59; 9.7, 30.9)	30.0% (9/30; 14.7, 49.4)
P value	0.059	0.061	0.077	0.104
Accuracy	87.5% (768/878; 85.1, 89.6)	91.5% (803/878; 89.4, 93.2)	94.3% (828/878; 92.6, 95.7)	97.2% (853/878; 95.8, 98.2)
P value	<0.001	<0.001	<0.001	<0.001

Numbers in parentheses indicate numerators/denominators and 95% confidence intervals. *P* values indicate a comparison between the AI system and the radiologist's review of the AI's result. AI, artificial intelligence; PPV, positive predictive value; NA, not applicable.

lesions detected by the AI were regarded as false-positive detections at the lowest threshold, while 65.0% (65/100) were regarded as false-positive detections at the highest threshold. As a result, the positivity rate after the radiologist's review was 13.8% (121/878) at the lowest threshold and 3.4% (30/878) at the highest threshold.

The sensitivities in the identification of patients with overlooked metastases were 92.3% (12/13; 95% CI, 64.0%–99.8%) at the lowest threshold and 69.2% (9/13; 95% CI, 38.6%–90.9%) at the highest threshold, consistent with the initial analyses by the AI. Meanwhile, the patient-level false-positive rates ranged from 2.4% (21/865; 95% CI, 1.5%–3.7%) to 12.6% (109/865; 95% CI,

10.5%–15.0%), representing a significant reduction compared with the initial analyses by the AI (all *p* < 0.001). Additionally, the patient-level PPVs ranged from 9.9% (12/121; 95% CI, 5.2%–16.7%) to 30.0% (9/30; 95% CI, 14.7%–49.4%) (Table 2) and were increased from the initial analyses by the AI, although the difference was not statistically significant. The accuracy ranged from 87.5% (768/878;

Table 3. Lesion-level performance of the artificial intelligence system and the radiologist's review of the artificial intelligence results

Performance metric	Threshold probability 0.4	Threshold probability 0.5	Threshold probability 0.6	Threshold probability 0.7
AI system				
Number of true-positive results	24	18	16	12
Number of false-positive results	451	301	189	88
Number of true-negative results	401	547	663	764
Number of false-negative results	2	12	10	14
Number of detection per patient	0.54 (475/878; 0.37, 0.71)	0.36 (319/878; 0.22, 0.50)	0.23 (205/878; 0.12, 0.35)	0.11 (100/878; 0.05, 0.18)
Sensitivity	92.3% (24/26; 74.5, 99.1)	69.2% (18/26; 48.2, 85.7)	61.5% (16/26; 40.6, 79.8)	46.2% (12/26; 26.6, 66.6)
Number of false-positive lesions per patient	0.51 (451/878; 0.35, 0.69)	0.34 (301/878; 0.20, 0.48)	0.22 (189/878; 0.10, 0.33)	0.10 (88/878; 0.03, 0.17)
PPV	5.1% (24/475; 3.3, 7.4)	5.6% (18/319; 3.4, 8.8)	7.8% (16/205; 4.5, 12.4)	12.0% (12/100; 6.4, 20.0)
Radiologist's review of the AI results				
Number of true-positive results	24	18	16	12
Number of false-positive results	176	110	62	23
Number of true-negative results	676	738	790	829
Number of false-negative results	2	12	10	14
Number of detection per patient	0.23 (200/878; 0.18, 0.28)	0.15 (128/878; 0.11, 0.18)	0.09 (78/878; 0.06, 0.11)	0.04 (35/878; 0.02, 0.06)
Sensitivity	92.3% (24/26; 74.5, 99.1)	69.2% (18/26; 48.2, 85.7)	61.5% (16/26; 40.6, 79.8)	46.2% (12/26; 26.6, 66.6)
P value	NA	NA	NA	NA
Number of false-positive lesions per patient	0.20 (176/878; 0.17, 0.25)	0.12 (110/878; 0.10, 0.16)	0.07 (62/878; 0.05, 0.09)	0.03 (23/878; 0.02, 0.04)
P value	<0.001	<0.001	<0.001	<0.001
PPV	12.0% (24/200; 7.8, 17.3)	14.1% (18/128; 0.09, 0.21)	20.5% (16/78; 0.12, 0.31)	34.3% (12/35; 19.1, 52.2)
P value	0.001	0.003	0.003	0.003

Numbers in parentheses indicate numerators/denominators and 95% confidence intervals. *P* values indicate a comparison between the AI system and the radiologist's review of the AI's result. AI, artificial intelligence; PPV, positive predictive value; NA, not applicable.

95% CI, 85.1%–89.6%) to 97.2% (853/878; 95% CI, 95.8%–98.2%). The accuracies exhibited significant improvement compared with the initial analyses by the AI (all *P* < 0.001).

The lesion-level sensitivities after the radiologist's review also remained similar to those following the initial analyses by the AI [92.3% (24/26; 95% CI, 74.5%–99.1%) at the lowest threshold; 46.2% (12/26; 95% CI, 26.6%–66.6%) at the highest threshold]. Meanwhile, the number of false-positive detections per examination ranged from 0.03 (23/878; 95% CI, 0.02–0.04) to 0.20 (176/878; 95% CI, 0.17–0.25), representing a significant reduction compared with the initial analyses by the AI (all *P* < 0.001). In addition, the lesion-level PPVs exhibited a significant increase compared with the initial analyses by the AI [*P* ≤ 0.001; 12.0% (24/200; 95% CI, 7.8%–17.3%) at the lowest threshold; 34.3% (12/35; 95% CI, 19.1%–52.2%) at the highest threshold].

Table 4 displays the patterns of false-positive detections by the AI system. The most common cause of false-positive detection was pulmonary nodules with the possibility of metastasis, based on the radiologist's review. Among clearly benign lesions that were regarded as false-positive detections by the radiologist's review, findings of infection or inflammation were the most common causes of false-positive detections, followed by calcified nodules.

Clinical significance

The AI system may identify missed basal lung lesions in abdominopelvic CT scans in patients with malignancy, providing feedback to radiologists, which can reduce the risk of missing basal lung metastasis.

Discussion

An AI system for pulmonary nodule detection on CT images can be utilized as a second reader after the radiologist's interpretation to prevent radiologists from overlooking clinically relevant pulmonary nodules. In the present study, we used an AI system to detect metastatic pulmonary nodules in the basal lungs captured by abdominopelvic CT images that were overlooked by radiologists. The results showed that the AI system could identify CT images with overlooked pulmonary metastases, with an AUC-ROC value of 0.911 and maximum patient-level and lesion-level sensitivity of 92.3%, respectively.

Although the AI generated several false-positive detections (maximum false-positive rate of 27.6%, 0.51 false-positive detections per patient), the radiologist's review of the AI results could effectively reduce the rate and number of false-positive detections (maximum false-positive rate of 12.6%, 0.20 false-positive detections per patient; *P* < 0.001, respectively).

Multiple studies have reported good performance of AI in the detection of pulmonary nodules on chest CT images.[16-19] In this study, the performance of the AI reached a level similar to that of radiologists. However, considering that AI cannot replace a radiologist's interpretation, its efficacy needs to be investigated based on its method of utilization. Since the most widely accepted methods of utilization involve CAD tools,[20-25] many studies have reported that AI can improve the performance of radiologists in lung nodule detection.[5,6,15-17] In addition to the use of AI as a CAD tool, several other utilization methods may also be feasible.[15] For instance, one promising method is its use as a second reader. In this context, the AI may analyze images after the radiologist's interpretation and provide feedback to the radiologist only when

the AI suspects that the radiologist has overlooked a pulmonary nodule. In this scenario, the oversight of significant pulmonary nodules can be prevented without the need to review the AI results of all the examinations.

We performed decision curve analyses to evaluate the net benefit of applying the AI system for true-positive and false-positive identifications. The scenario with AI as a second reader showed a higher net benefit than the scenario without AI when the ratio between the harm of false-positive interpretations to the benefit of true-positive interpretations is ≤1:26. In most clinical situations, overlooking pulmonary metastases could have significant consequences,
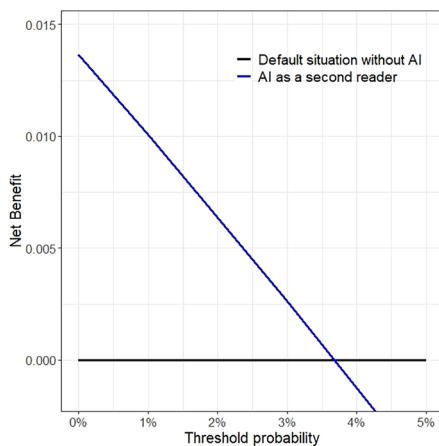
Figure 7. Decision curve for the identification of overlooked basal lung metastases in abdominopelvic computed tomography scans. The artificial intelligence (AI) as a second reader scenario (blue line) exhibited a higher net benefit than the default scenario without AI (black line) when the risk threshold is 3.7% or smaller. In other words, using the AI tool would be beneficial if the ratio of cost from false-positive results to benefit from true-positive results is 3.7:96.3 (1:26) or smaller.

potentially depriving the patient of timely systemic treatment. Meanwhile, false-positive detections by AI may lead to a review by the radiologist, and the associated costs would be relatively much smaller compared with the risks of overlooking pulmonary metastases. Therefore, we believe that using the AI as a second reader would be a reasonable scenario.

In our study, an AI system was applied to the abdominopelvic CT scans of patients with cancer who were interpreted as negative for basal lung metastasis. In a retrospective evaluation of available follow-up examinations, overlooked pulmonary metastases were identified in 1.5% of patients, a frequency that should not be ignored. In this context, the AI could accurately discriminate between CT images with and without overlooked pulmonary metastases (AUC-ROC, 0.911). Furthermore, at a sensitive operating threshold, the AI could identify most CT scans with overlooked metastases (sensitivity: 92.3%). Notably, the identification of false-negative interpretations by radiologists using AI has been investigated in the field of chest radiography. Specifically, Nam et al.[26] and Jang et al.[27] reported that AI can identify lung cancers overlooked by radiologists on chest X-rays. In addition, Hwang et al.[28] reported that AI can identify various clinically relevant abnormalities on chest radiographs that were previously interpreted as normal.

Because benign pulmonary nodules and pulmonary metastases are often difficult to differentiate, false-positive detection by AI is inevitable. When used as a second reader,[15] false-positive detection may lead to unnecessary feedback to the radiologist, followed

by reinterpretation by the radiologist. In our study, the maximum false-positive rate was 27.6%, indicating that the AI may generate false-positive feedback in 27.6% of CT images without overlooking metastases. Based on the review of the AI results by a radiologist, more than half of the AI detections were regarded as clearly benign nodules (findings of pulmonary infection and calcified nodules). Notably, the radiologist's review was effective because it significantly reduced the rate of false positives while maintaining a similar sensitivity for metastasis. The results also suggest that further improvements in AI performance may reduce the false-positive rate and the frequency of unnecessary reinterpretation by radiologists.

Pulmonary metastases and benign pulmonary nodules are often indistinguishable, even when evaluated by a radiologist. Therefore, as expected, there were considerable false-positive detections even after the radiologist's review (maximum false-positive rate: 12.6%). Moreover, the identification of benign nodules may lead to the requirement of chest CT examinations for further evaluation or follow-up of the pulmonary nodules. Considering that all patients were under follow-up for malignancies, we believe that additional chest CT scans may not significantly harm the patients.

Our study has several limitations. First, since our study was conducted at a single tertiary medical institution, the reproducibility of our results remains uncertain. Future studies may be required to confirm the reproducibility of our results in other clinical situations. Second, although we consecutively included 878 abdominopelvic CT scans, the absolute number of overlooked pulmonary metastases is quite small (n = 13), limiting the statis-

Table 4. Detection patterns of the artificial intelligence system				
Variable	Threshold probability 0.4	Threshold probability 0.5	Threshold probability 0.6	Threshold probability 0.7
Lesions with the possibility of metastasis based on the radiologist's review				
Metastasis	24 (5.1%)	18 (5.6%)	16 (7.8%)	12 (12%)
Benign lung nodules	176 (37.1%)	110 (34.5%)	62 (30.2%)	23 (23%)
Clearly benign lesions based on the radiologist's review				
Findings of infection/inflammation	116 (24.4%)	88 (28%)	66 (32.2%)	40 (40%)
Calcified nodules	90 (18.9%)	72 (27.6%)	49 (24.0%)	23 (23%)
Ground-glass nodules	33 (6.9%)	21 (6.6%)	9 (4.4%)	1 (1%)
Pulmonary vessels	24 (5.1%)	3 (0.9%)	1 (0.5%)	0 (0%)
Atelectasis	7 (1.5%)	4 (1.3%)	1 (0.5%)	1 (1%)
Others	5 (1.1%)	3 (0.9%)	1 (0.5%)	0 (0%)
Total	475 (100%)	319 (100%)	205 (100%)	100 (100%)

Numbers in parentheses indicate the proportions among the total detections by the artificial intelligence system.

tical power. A multicenter study with a larger sample size may be required to confirm the efficacy of AI as a second reader. Third, in this study, AI was retrospectively applied to abdominopelvic CT scans. Therefore, the practical efficacy of AI systems remains unknown. A prospective study following the integration of AI into the workflow may be required to investigate its real-world efficacy. Finally, the effect of AI beyond the detection of overlooked metastases, including its effects on patient outcomes and changes in treatment decision-making, remains unknown.

In conclusion, the applied AI system could accurately identify basal lung metastases captured in abdominopelvic CT images that were overlooked by radiologists, suggesting its potential as a second reader after the radiologist's interpretation. Further prospective studies are warranted to investigate the real-world efficacy of AI as a second reader as well as the impact of AI beyond the detection of metastases.

Conflict of interest

Funding

References

1. Tan JR, Cheong EHT, Chan LP, Tham WP. Implementation of an artificial intelligence-based double read system in capturing pulmonary nodule discrepancy in CT studies. *Curr Probl Diagn Radiol.* 2021;50(2):119-122. Erratum in: *Curr Probl Diagn Radiol.* 2022;51(4):673. [Crossref]

2. Li F, Armato SG, Giger ML, MacMahon H. Clinical significance of noncalcified lung nodules in patients with breast cancer. *Breast Cancer Res Treat.* 2016;159(2):265-271. [Crossref]

3. Rubin GD. Lung nodule and cancer detection in computed tomography screening. *J Thorac Imaging.* 2015;30(2):130-138. [Crossref]

4. Lee B, Lim A, Lalvani A, et al. The clinical significance of radiologically detected silent pulmonary nodules in early breast cancer. *Ann Oncol.* 2008;19(12):2001-2006. [Crossref]

5. Hwang EJ, Park S, Jin KN, 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. [Crossref]

6. Hwang EJ, Park S, Jin KN, 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(5):739-747. [Crossref]

7. Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med.* 2019;25(6):954-961. Erratum in: *Nat Med.* 2019;25(8):1319. [Crossref]

8. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA.* 2016;316(22):2402-2410. [Crossref]

9. Ting DSW, Cheung CY, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA.* 2017;318(22):2211-2223. [Crossref]

10. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA.* 2017;318(22):2199-2210. [Crossref]

11. Hinton G. Deep learning-a technology with the potential to transform health care. *JAMA.* 2018;320(11):1101-1102. [Crossref]

12. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019;25(1):44-56. [Crossref]

13. Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: assessing new tests against existing diagnostic pathways. *BMJ.* 2006;332(7554):1089-1092. Erratum in: *BMJ.* 2006;332(7554):1368. [Crossref]

14. Tang A, Tam R, Cadrin-Chênevert A, et al. Canadian association of radiologists white paper on artificial intelligence in radiology. *Can Assoc Radiol J.* 2018;69:120-135. [Crossref]

15. Hwang EJ, Park CM. Clinical Implementation of deep learning in thoracic radiology: potential applications and challenges. *Korean J Radiol.* 2020;21(5):511-525. [Crossref]

16. Nam JG, Park S, Hwang EJ, et al. Development and validation of deep learning-based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. *Radiology.* 2019;290:218-228. [Crossref]

17. Harris M, Qi A, Jeagal L, et al. A systematic review of the diagnostic accuracy of artificial intelligence-based computer programs to analyze chest X-rays for pulmonary tuberculosis. *PloS One.* 2019;14:e0221339. [Crossref]

18. Liu Z, Li Z, Qu J, et al. Radiomics of multiparametric MRI for pretreatment prediction of pathologic complete response to neoadjuvant chemotherapy in breast cancer: a multicenter study. *Clin Cancer Res.* 2019;25:3538-3547. [Crossref]

19. Sim Y, Chung MJ, Kotter E, et al. Deep convolutional neural network-based software improves radiologist detection of malignant lung nodules on chest radiographs. *Radiology.* 2020;294(1):199-209. [Crossref]

20. Cha MJ, Chung MJ, Lee JH, Lee KS. Performance of deep learning model in detecting operable lung cancer with chest radiographs. *J Thorac Imaging.* 2019;34(2):86-91. [Crossref]

21. Park S, Lee SM, Lee KH, et al. Deep learning-based detection system for multiclass lesions on chest radiographs: comparison with observer readings. *Eur Radiol.* 2020;30(3):1359-1368. [Crossref]

22. Majkowska A, Mittal S, Steiner DF, et al. Chest radiograph interpretation with deep learning models: assessment with radiologist-adjudicated reference standards and population-adjusted evaluation. *Radiology.* 2020;294(2):421-431. [Crossref]

23. Pande T, Cohen C, Pai M, Ahmad Khan F. Computer-aided detection of pulmonary tuberculosis on digital chest radiographs: a systematic review. *Int J Tuberc Lung Dis.* 2016;20:1226-1230. [Crossref]

24. Schalekamp S, van Ginneken B, Karssemeijer N, Schaefer-Prokop CM. Chest radiography: new technological developments and their applications. *Semin Respir Crit Care Med.* 2014;35:3-16. [Crossref]

25. Kim DW, Jang HY, Kim KW, Shin Y, Park SH. Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: results from recently published papers. *Korean J Radiol.* 2019;20:405-410. [Crossref]

26. Nam JG, Hwang EJ, Kim DS, et al. Undetected lung cancer at posteroanterior chest radiography: potential role of a deep learning-based detection algorithm. *Radiol Cardiothorac Imaging.* 2020;2(6):e190222. [Crossref]

27. Jang S, Song H, Shin YJ, et al. Deep Learning-based automatic detection algorithm for reducing overlooked lung cancers on chest radiographs. *Radiology.* 2020;296(3):652-661. Erratum in: *Radiology.* 2022;302(2):E10. [Crossref]

28. Hwang EJ, Park J, Hong W, et al. Artificial intelligence system for identification of false-negative interpretations in chest radiographs. *Eur Radiol.* 2022;32(7):4468-4478. [Crossref]