



Deep Learning-based Diagnosis and Localization of Pneumothorax on Portable Supine Chest X-ray in Intensive and Emergency Medicine: A Retrospective Study

Chih-Hung Wang^{1,2} · Tzuching Lin³ · Guanru Chen³ · Meng-Rui Lee⁴ · Joyce Tay² · Cheng-Yi Wu² · Meng-Che Wu² · Holger R. Roth⁵ · Dong Yang⁵ · Can Zhao⁵ · Weichung Wang³ · Chien-Hua Huang^{1,2}

Received: 4 September 2023 / Accepted: 25 November 2023
© The Author(s) 2023

Abstract

Purpose To develop two deep learning-based systems for diagnosing and localizing pneumothorax on portable supine chest X-rays (SCXRs).

Methods For this retrospective study, images meeting the following inclusion criteria were included: (1) patient age ≥ 20 years; (2) portable SCXR; (3) imaging obtained in the emergency department or intensive care unit. Included images were temporally split into training (1571 images, between January 2015 and December 2019) and testing (1071 images, between January 2020 to December 2020) datasets. All images were annotated using pixel-level labels. Object detection and image segmentation were adopted to develop separate systems. For the detection-based system, EfficientNet-B2, DneseNet-121, and Inception-v3 were the architecture for the classification model; Deformable DETR, TOOD, and VFNet were the architecture for the localization model. Both classification and localization models of the segmentation-based system shared the UNet architecture.

Results In diagnosing pneumothorax, performance was excellent for both detection-based (Area under receiver operating characteristics curve [AUC]: 0.940, 95% confidence interval [CI]: 0.907–0.967) and segmentation-based (AUC: 0.979, 95% CI: 0.963–0.991) systems. For images with both predicted and ground-truth pneumothorax, lesion localization was highly accurate (detection-based Dice coefficient: 0.758, 95% CI: 0.707–0.806; segmentation-based Dice coefficient: 0.681, 95% CI: 0.642–0.721). The performance of the two deep learning-based systems declined as pneumothorax size diminished. Nonetheless, both systems were similar or better than human readers in diagnosis or localization performance across all sizes of pneumothorax.

Conclusions Both deep learning-based systems excelled when tested in a temporally different dataset with differing patient or image characteristics, showing favourable potential for external generalizability.

Keywords Chest Radiograph · Detection · Localization · Deep Learning · Pneumothorax

Tzuching Lin and Guanru Chen contributed equally to this work.

Weichung Wang PhD and Chien-Hua Huang MD, PhD contributed equally to this work.

✉ Weichung Wang
wwang@ntu.edu.tw

✉ Chien-Hua Huang
chhuang5940@ntu.edu.tw

¹ Department of Emergency Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan

² Department of Emergency Medicine, National Taiwan University Hospital, No. 7, Zhongshan S. Rd., Taipei, Zhongzheng Dist. 100, Taiwan

³ Institute of Applied Mathematical Sciences, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei 106, Taiwan

⁴ Department of internal medicine, National Taiwan University Hospital, Taipei, Taiwan

⁵ NVIDIA Corporation, Bethesda, USA

Introduction

A pneumothorax is an abnormal collection of air in the pleural space between the lung and the chest wall. The annual incidence rate of pneumothorax was approximately 7.3 cases per 100,000 individuals [1]; among hospitalized patients, the incidence of pneumothorax was estimated at 22.7 cases per 100,000 admissions every year [2].

Without prompt recognition and management, pneumothorax may evolve into life-threatening tension pneumothorax. Rapid and correct identification of pneumothorax can minimize the risk associated with tension pneumothorax [3] and thus improve patient outcomes.

Because of its advantage in mobility, the portable supine chest radiograph (SCXR) is one of the most common imaging studies performed in the emergency department (ED) and intensive care unit (ICU) [4, 5]. However, the reported sensitivity of SCXR for detecting pneumothorax varied widely, ranging from 9 to 75% [6], indicating the high rate of misses at initial encounters.

Several factors may explain the heterogeneous sensitivity of SCXR in detecting pneumothorax [7–9]. First, the imaging quality may be reduced because of limitations in the patient's positioning or body habitus. Second, the distribution of free air in the pleural space is variable and highly dependent on the intrathoracic anatomic structure and relevant pathology in the lung parenchyma and pleural space [10]. The subtle imaging findings of pneumothorax in SCXRs require expertise and cautious inspection to detect its presence.

In the current study, we hypothesized that artificial intelligence-based approaches for interpreting portable SCXRs may facilitate physicians in detecting pneumothorax with greater efficiency and accuracy. We aimed to develop and validate deep learning (DL)-based computer-aided diagnosis (CAD) systems that enable more efficient and accurate pneumothorax detection and localization by portable SCXR.

Materials and Methods

Study Design and Setting

We conducted a retrospective study to develop and test our CAD systems in chronologically differing image datasets. Local portable SCXRs were retrieved from the Picture Archiving and Communication System (PACS) database of the National Taiwan University Hospital (NTUH). This study was approved by the Research Ethics Committee of NTUH (reference number: 202003106RINC) and granted a consent waiver. Our results are reported according to the Checklist for Artificial Intelligence in Medical Imaging (CLAIM) [11].

Image Acquisition and Dataset Designation

As shown in Fig. 1, a Radiology Information System served to identify candidate images used in the building of training (NTUH-1519) and testing (NTUH-20) datasets.

Inclusion criteria for the candidate positive group of NTUH-1519 were as follows: (1) text report with clinical finding of pneumothorax; (2) patient age ≥ 20 years; (3) portable SCXR; (4) imaging obtained in ED or ICU; and (5) exam performed between January 1, 2015 and December 31, 2019. We only selected the first study as the representative image for each patient in the analysis. Inclusion criteria for the candidate negative group were the same as those above, except for text reports devoid of pneumothorax. We randomly selected qualifying images from negative candidates, populating positive and negative groups at an approximate ratio of 1:2. The tentative candidate list was then further scrutinized to avoid overlap of patients.

Image acquisition for NTUH-20 differed in the time frame (January 1, 2020 to December 31, 2020) but was otherwise the same. In addition, the candidate negative group

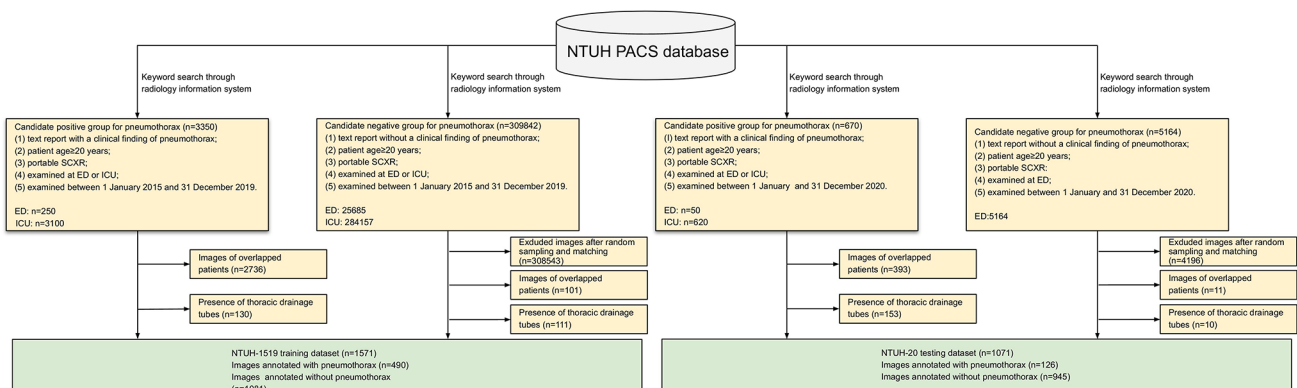


Fig. 1 Flow chart of image inclusion process and dataset designation. SCXR, supine chest X-ray; ED, emergency department; ICU, intensive care unit; NTUH, National Taiwan University Hospital; PACS, Picture Archiving and Communication System

included only images obtained in EDs, and the image ratio for positive and negative groups was approximately 1:10.

We exported all eligible de-identified images in Digital Imaging and Communications in Medicine (DICOM) format, including corresponding text reports for analysis. The radiological reports were generated by various board-certified radiologists for clinical purposes.

Image Annotation, Ground Truth, and CXR Report Extraction

Each image was first split into 10×10 grids of equal size. Bounding boxes were then used to cover the pneumothorax visible in each grid, utilizing the least area. Each image was randomly assigned to two emergency physicians, blinded to each other's efforts, for image annotation. A total of six board-certified and four board-eligible emergency physicians were involved, each with at least 4 years of clinical experience. All images were ultimately reviewed by an experienced (10 years) board-certified pulmonologist and intensivist who adjusted annotations as necessary. The reviewed annotations served as ground truth in model training and testing. Any images harbouring thoracic drainage tubes were picked up and excluded from further analysis. CXR findings and diagnoses [12, 13] were extracted manually by research assistants blinded to annotations according to the radiology reports.

Development of Algorithm

We designed two separate CAD systems (Fig. 2), each including a classification model and a localization model and jointly yielding the following variables: (1) diagnosis output, indicating the presence or absence of pneumothorax, and (2) localization output, indicating the pneumothorax lesion site. Our CAD systems were designated as detection- or segmentation-based systems according to the localization method applied (i.e., object detection or image segmentation).

Supplemental Figs. 1 and 2 show the training pipelines in detail. In brief, NTUH-1519 was first split into different subsets to train the CAD systems. This partition process ensured similar proportions of images (with vs. without pneumothoraces) and eliminated patient overlap across all subsets. All images underwent preprocessing before analysis. The annotated bounding boxes were transformed into segmentation masks for the segmentation-based system. The segmentation masks would also be replaced with one or several larger bounding boxes, covering all adjoining masks at minimal area and serving as input for the detection-based system.

For the detection-based system, EfficientNet-B2 [14], DneseNet-121 [15], and Inception-v3 [16] were selected as the architecture for the classification model; Deformable DETR [17], TOOD [18], and VFNet [19] were selected as the architecture for the localization model. Both classification and localization models of the segmentation-based

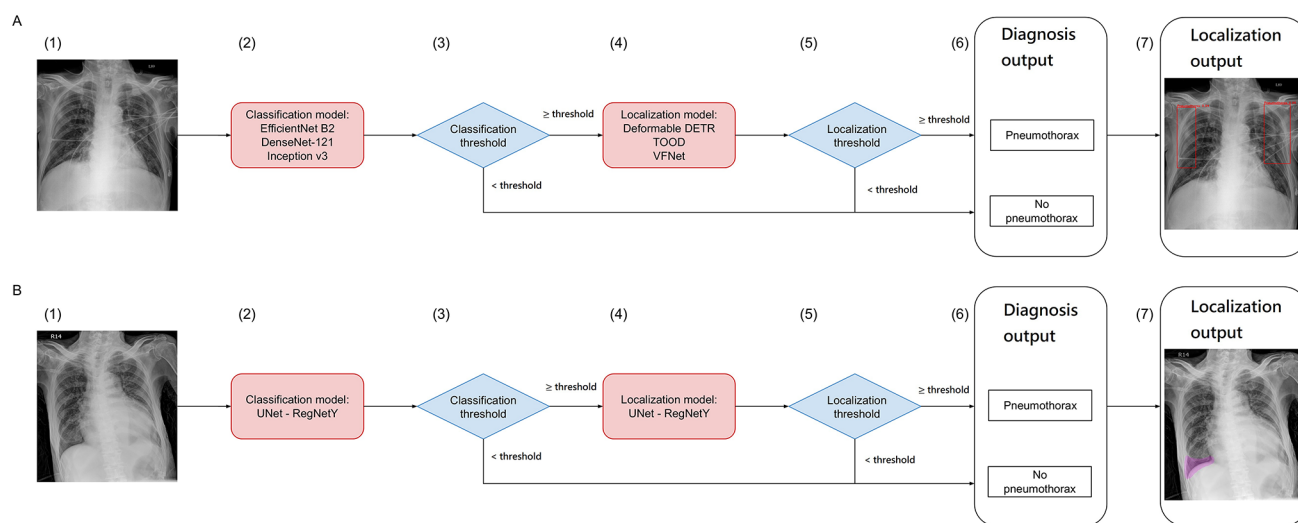


Fig. 2 Structures of (A) detection-based and (B) segmentation-based CAD systems: Each system incorporates a classification model and a localization model, the major difference being the localization method (object detection vs. image segmentation). A CXR image (1) is first passed to the classification model (2) to derive the probability regarding presence of pneumothorax. If the output probability exceeds the classification threshold (3), the image is passed to the localization

model (4) to assess pneumothorax position. Once the largest predicted confidence score of the bounding boxes (detection-based CAD system) or predicted areas (segmentation-based CAD system) of the pneumothorax is above the localization threshold (5), the CAD system yields diagnosis (6) and localization (7) outputs. Red rectangles and purple areas denote pneumothorax locations. CAD, computer-aided diagnosis; CXR, chest X-ray

system shared the UNet [20] architecture, using RegNetY [21] as encoder.

Evaluation Metrics of Algorithm

Diagnosis output performance was determined by the area under receiver operating characteristics curve (AUC), area under precision-recall curve, sensitivity, specificity, positive predictive value, and negative predictive value. Youden's index [22] acquired from the training (NTUH-1519) dataset indicated the optimal threshold for testing.

Localization output performance was measured by Dice coefficient, calculated as twice the area of overlap divided by total pixel count in predicted and ground-truth masks. Dice coefficients were only computed in images positive for both predicted and ground-truth pneumothoraces (i.e., images classified as true positives), referred to as prediction-ground truth TP-Dice. TP-Dice coefficients were also calculated to evaluate the consistency shown by two annotators (inter-annotator TP-Dice).

Table 1 Comparison of training (NTUH-1519) and testing (NTUH-20) datasets

Variables	NTUH-1519 (n = 1571)	NTUH-20 (n = 1071)	p-value
Age, year	65.1 (16.1)	58.6 (19.9)	< 0.001
Male, n	946 (60.2)	580 (54.2)	0.002
Qualitative findings in radiology reports, n			
Atelectasis	38 (2.4)	26 (2.4)	0.99
Cardiomegaly	656 (41.8)	286 (26.7)	< 0.001
Consolidation	331 (21.1)	77 (7.2)	< 0.001
Emphysema	6 (0.4)	4 (0.4)	0.97
Endotracheal intubation	558 (35.5)	135 (12.6)	< 0.001
Haziness	192 (12.2)	72 (6.7)	< 0.001
Infiltration	255 (16.2)	204 (19.0)	0.06
Nodularity	55 (3.5)	46 (4.3)	0.30
Opacification	587 (37.4)	379 (35.4)	0.30
Pleural effusion	418 (26.6)	200 (18.7)	< 0.001
Pneumothorax	484 (30.8)	124 (11.6)	< 0.001
Annotation, n			
Pneumothorax	490 (31.2)	126 (11.8)	< 0.001
Large pneumothorax	164 (10.4)	44 (4.1)	< 0.001
Medium pneumothorax	162 (10.3)	35 (3.3)	< 0.001
Small pneumothorax	164 (10.4)	47 (4.4)	< 0.001
Manufacturer, n			
Agfa	1405 (89.4)	113 (10.6)	
Canon	128 (8.1)	30 (2.8)	
Carestream	12 (0.8)	0 (0)	
Philips	5 (0.3)	0 (0)	
Samsung	21 (1.3)	928 (86.6)	
Image pixels	8055937.4 (2898803.0)	7538921.0 (2337194.0)	< 0.001

Data expressed as mean (standard deviation) values or as counts (proportions)

Statistical Analysis

Continuous variables were expressed as means with standard deviations while categorical variables as counts and proportions. Continuous variables were compared with Student's *t*-test, and categorical variables were compared with the Chi-squared test. All statistics were determined as point estimates, with 95% confidence intervals (CIs), through a bootstrap technique at 1,000 repetitions. Prediction-ground truth and inter-annotator TP-Dice coefficients were compared by paired *t*-test. Subgroup analysis was performed to explore the influence of the pneumothorax size on the model performance. The pneumothorax was categorized into large, medium, and small sizes based on the 33rd and 66th percentiles of areas of segmentation masks. A two-tailed *p*-value < 0.05 was considered statistically significant. All computations were driven by open-source freeware (SciPy v1.8.1) [23].

Results

As shown in Figs. 1 and 2642 images were acquired from the PACS database, (training, 1571; testing, 1071). Significant differences between NTUH-1519 and NTUH-20 datasets are shown in Table 1, with 490 (31.2%) and 126 (11.8%) patients, respectively annotated as pneumothorax. Aside from pneumothorax, other patient characteristics and image findings were numerically similar for images annotated for presence or absence of pneumothorax in NTUH-1519 and NTUH-20 datasets (Supplemental Tables 3 and 4).

Table 2 indicates that pneumothorax was accurately diagnosed by detection-based (AUC: 0.940, 95% CI: 0.907–0.967) and segmentation-based (AUC: 0.979, 95% CI: 0.963–0.991) systems, both achieving levels similar to those of radiology reports. Figure 3 demonstrates four representative imaging sets. The overlain predicted bounding boxes or segmentation masks served to assist clinicians in verifying the diagnosis and position of pneumothorax. As shown in Table 3, prediction-ground truth TP-Dice coefficients for detection- and segmentation-based systems were 0.758 (95% CI: 0.707–0.806) and 0.681 (95% CI: 0.642–0.721), respectively, both values significantly surpassing inter-annotator TP-Dice values. Supplemental Table 5 lists the required computational resources for both systems.

Subgroup analysis showed that in diagnosing pneumothorax, performances of both systems declined according to the size (large to small) of pneumothorax, consistent with the trend for radiology reports (Table 2). In terms of pneumothorax localization, diminishing pneumothorax size (large to small) corresponded with similar declines in prediction-ground truth TP-Dice coefficients for both systems,

Table 2 Diagnostic performances of computer-aided diagnosis (CAD) systems and radiology reports

Readers	AUC	AUPRC	Sensitivity	Specificity	PPV	NPV
<i>Diagnosing pneumothorax</i>						
Detection-based CAD system	0.940 (0.907–0.967)	0.833 (0.775–0.884)	0.687 (0.598–0.776)	0.991 (0.984–0.996)	0.892 (0.823–0.954)	0.966 (0.955–0.977)
Segmentation-based CAD system	0.979 (0.963–0.991)	0.910 (0.863–0.948)	0.926 (0.874–0.972)	0.926 (0.909–0.942)	0.584 (0.510–0.657)	0.991 (0.985–0.997)
Radiology reports	0.953 (0.927–0.976)	0.866 (0.802–0.923)	0.914 (0.862–0.961)	0.992 (0.986–0.997)	0.936 (0.889–0.976)	0.989 (0.981–0.995)
<i>Diagnosing large pneumothorax</i>						
Detection-based CAD system	0.997 (0.993–1.000)	0.950 (0.895–0.987)	0.946 (0.861–1.000)	0.991 (0.984–0.996)	0.788 (0.659–0.903)	0.998 (0.995–1.000)
Segmentation-based CAD system	0.999 (0.998–1.000)	0.977 (0.933–1.000)	1.000 (1.000–1.000)	0.926 (0.909–0.942)	0.336 (0.248–0.426)	1.000 (1.000–1.000)
Radiology reports	0.996 (0.993–0.998)	0.820 (0.697–0.927)	1.000 (1.000–1.000)	0.992 (0.985–0.997)	0.820 (0.697–0.927)	1.000 (1.000–1.000)
<i>Diagnosing medium pneumothorax</i>						
Detection-based CAD system	0.947 (0.880–0.991)	0.779 (0.630–0.907)	0.763 (0.600–0.917)	0.991 (0.984–0.996)	0.718 (0.556–0.864)	0.993 (0.987–0.998)
Segmentation-based CAD system	0.976 (0.925–0.999)	0.897 (0.792–0.973)	0.967 (0.889–1.000)	0.926 (0.909–0.942)	0.290 (0.203–0.381)	0.999 (0.997–1.000)
Radiology reports	0.996 (0.993–0.998)	0.791 (0.657–0.923)	1.000 (1.000–1.000)	0.992 (0.985–0.997)	0.791 (0.657–0.923)	1.000 (1.000–1.000)
<i>Diagnosing small pneumothorax</i>						
Detection-based CAD system	0.886 (0.822–0.943)	0.478 (0.326–0.637)	0.407 (0.257–0.565)	0.991 (0.985–0.996)	0.652 (0.462–0.833)	0.975 (0.965–0.985)
Segmentation-based CAD system	0.964 (0.942–0.981)	0.636 (0.476–0.778)	0.833 (0.711–0.941)	0.926 (0.909–0.942)	0.330 (0.242–0.420)	0.992 (0.986–0.998)
Radiology reports	0.984 (0.955–0.998)	0.818 (0.698–0.911)	0.975 (0.917–1.000)	0.992 (0.986–0.996)	0.837 (0.723–0.923)	0.999 (0.997–1.000)

Data expressed as point estimates (95% confidence intervals)

AUC, area under receiver operating characteristics curve; AUPRC, area under precision-recall curve; NPV, negative predictive value; PPV, positive predictive value

again aligned with the observed trend for inter-annotator TP-Dice values (Table 3).

Discussion

Main Findings

Both detection- and segmentation-based systems achieved excellent performance, which was comparable to radiology reports or human annotators. Like human readers, the diagnosis and localization performance of the CAD systems might be influenced by the size of the pneumothorax.

Annotation of Pneumothorax on SCXR

Most public datasets [24, 25] rely on chest X-rays with image-level labels of common thoracic diseases that are text-mined from radiology reports and are inherently inaccurate [26, 27]. For example, for ChestX-ray14, a study suggested the agreement regarding pneumothorax diagnosis between the image-level label and radiologist review was

only about 60% [28], which may lead to poor model generalizability [29].

On the other hand, pixel-based annotation may effectively facilitate the development of pneumothorax-detecting algorithms [30]. For standing CXR, the pneumothorax lesion could usually be delineated [31, 32] by the visceral pleural line in the apicolateral space [33]. Nonetheless, when patients are in the supine position, the spaces where the air is trapped differ from those in the standing position [34]. Adopting segmentation masks to delineate the pneumothorax lesion on SCXR might raise a concern that only those images with clear pleural lines were annotated, leading to selection bias.

Consequently, we used bounding boxes for annotation, allowing for localization of pneumothoraces without distinct pleural lines. Nevertheless, in some lesions, such as those spanning lung apices and basal aspects, the use of bounding boxes might encompass nearly an entire unilateral lung region. This problem was overcome by dividing images into 10×10 grids, permitting bounding boxes to accommodate lesions of varying shapes.

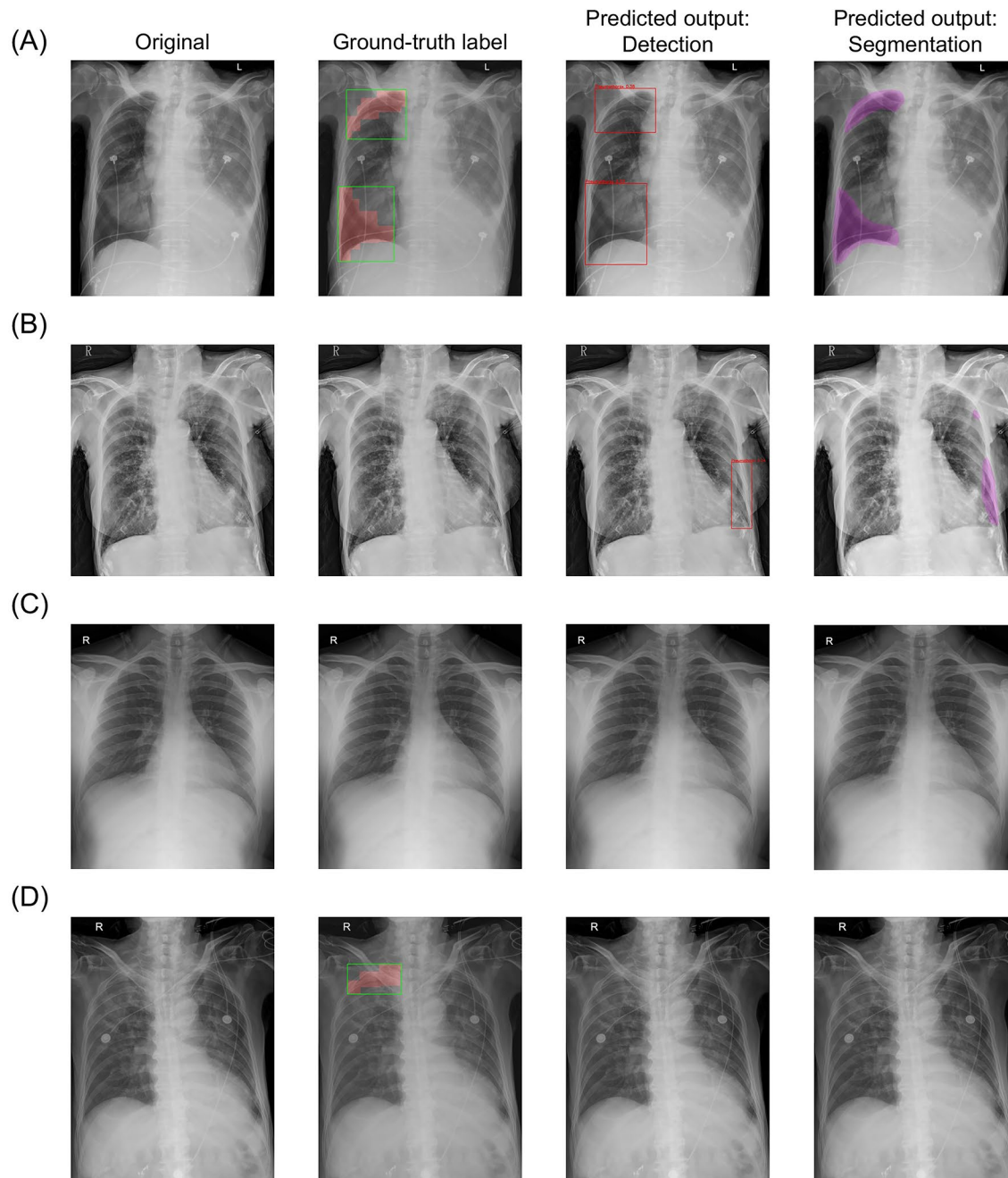


Fig. 3 Sample images stratified by predicted results of diagnosis outputs, including (A) true-positive, (B) false-positive, (C) true-negative, and (D) false-negative results. The first column at left displays original images. In the second column from left, preprocessed bounding boxes (green rectangles) and segmentation masks (red areas) of detection-

and segmentation-based CAD systems are shown. The third column from left demonstrates bounding boxes (red rectangles) predicted by detection-based CAD system, with segmentation masks (purple areas) predicted by segmentation-based CAD system appearing in the fourth column. CAD, computer-aided diagnosis

Dataset Selection for Training and Testing Models

Considering the low (0.5–3%) incidence of pneumothorax cited in epidemiologic data [35, 36], use of a consecutive random SCXR sampling for model development may result in class imbalance. Such imbalance may bias CAD systems

towards learning features of a more common class (i.e., pneumothorax-negative images) and distort various evaluation metrics [37]. Thus, we employed a case-controlled design [38, 39] to achieve greater balance in training and testing datasets. As shown in Table 1, the higher proportion (31.2%) of images annotated as pneumothorax in the

Table 3 Localization performances of computer-aided diagnosis (CAD) systems and annotators

Annotations	Detection-based CAD system			Segmentation-based CAD system		
	Prediction-ground truth TP-Dice	Inter-annotator TP-Dice	<i>p</i> -value	Prediction-ground truth TP-Dice	Inter-annotator TP-Dice	<i>p</i> -value
Pneumothorax	0.758 (0.707–0.806)	0.717 (0.665–0.766)	< 0.001	0.681 (0.642–0.721)	0.669 (0.653–0.654)	< 0.001
Large pneumothorax	0.883 (0.840–0.921)	0.767 (0.690–0.836)	< 0.001	0.769 (0.715–0.823)	0.756 (0.739–0.741)	< 0.001
Medium pneumothorax	0.708 (0.625–0.798)	0.677 (0.574–0.771)	< 0.001	0.693 (0.628–0.757)	0.639 (0.651–0.653)	< 0.001
Small pneumothorax	0.574 (0.446–0.701)	0.686 (0.596–0.771)	> 0.99	0.558 (0.489–0.627)	0.616 (0.589–0.591)	> 0.99

Data expressed as point estimates (95% confidence intervals)

Prediction-ground truth TP-Dice, coefficient indicating agreement between predicted and ground-truth results; inter-annotator TP-Dice, coefficient indicating agreement between two annotators

NTUH-1519 dataset may enable our CAD systems to better learn pneumothorax-related features; whereas the lower proportion (11.8%) in NTUH-20 fostered performance testing on a plane approaching real-world prevalence [35, 36].

In a previous study [29], the accuracy of DL-based pneumothorax detection was shown to significantly decline when testing the algorithm in an external dataset. Concerns over accuracy overestimation and limited generalizability of such algorithms may be mitigated by model evaluation in an independent dataset. However, no datasets dedicated to portable SCXRs were available for our purposes. According to the Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement [40], external validation may use data collected by the same researchers, using the same predictors and outcome definitions and assessments, but typically sampled from a later period (temporal or narrow validation). In our study, the NTUH-20 dataset consisted of SCXRs taken during 2020 at NTUH. Compared with NTUH-1519, NTUH-20 was a chronologically different dataset (2015–2019 vs. 2020), with significant differences (Table 1). According to the TRIPOD statement, the chronologically different testing dataset can be used to verify the external generalizability of the CAD system.

Diagnosis Output Performance

Niehues et al. [41] used portable SCXR to develop a CAD algorithm with excellent performance in identifying pneumothorax (AUC: 0.92, 95% CI: 0.89–0.95). Nonetheless, the thoracic drains were concomitantly present in approximately half of the images with pneumothorax [41]. It is thus conceivable that these drains were misconstrued in the algorithm as a feature of pneumothorax [42]. Rueckel et al. [30] also collected 3062 SCXRs, including 760 images with pixel-level annotations of pneumothorax and thoracic drain. This model also performed well overall (AUC: 0.877) for unilateral pneumothorax detection.

For the present study, however, we excluded images with thoracic drains and used bounding boxes for pixel-level annotation. Both of detection- and segmentation-based

systems delivered excellent performances (AUC values > 0.94) in pneumothorax detection. In our study, the architectures of the classification models differed between the two CAD systems as the UNet-based model [20] itself could output both classification results and localization information.

Routine portable SCXR exams are common practice in critical care [43, 44]. Such regular use of portable SCXR exams may partly account for the prolonged turnaround time from image acquisition to interpretation by a radiologist [45]. Our systems may help prioritize portable SCXRs within queues, flagging those to be checked upfront by a radiologist or earmarking treating clinicians for notifications. As shown in Table 1, there was a high percentage of patients receiving tracheal intubation. Early detection of pneumothorax may facilitate prompt life-saving procedures for these patients to prevent serious complications, such as tension pneumothorax.

Localization Output Performance

Using standing chest X-rays, a model devised by Lee et al. [46] has achieved a Dice coefficient of 0.798 in pneumothorax localization. Feng et al. [47] also derived a model able to localize the pneumothorax lesion (Dice coefficient: 0.69). Nevertheless, even though Feng et al. [47] included portable SCXRs in the analysis, the researchers excluded films with only supine signs of pneumothorax, e.g., deep sulcus sign. Another model by Zhou et al. [48], based on frontal chest X-rays alone (no portable SCXRs), could detect pneumothorax with a Dice coefficient of 0.827.

The images of portable SCXR are generally deemed suboptimal for diagnosis. The patients are often unable to cooperate during image acquisition, leading to poor bodily orientation or inspiratory efforts. Compared with standing chest X-rays, they are also inferior in image quality, hindering the diagnosis of pneumothorax due to lesser degrees of resolution and luminance [49]. Furthermore, classic findings of pneumothorax on standing chest X-rays are often lacking on portable SCXRs. Given the more challenging

interpretation of SCXRs, past models [46–48; 50] may not be suitable for pneumothorax localization on these images.

Both CAD systems we developed (based on object detection or image segmentation) performed excellently in pneumothorax localization, comparable to the level of annotators (Table 3). To the best of our knowledge, our CAD systems may be the first ones capable of localizing pneumothoraces on portable SCXRs. Although the detection- and segmentation-based systems performed similarly in testing, their required computational resources differed substantially (Supplemental Table 5). The detection-based system only outputs approximate positional information with several coordinates of bounding boxes. Logically, its computational demands should be less than those of the segmentation-based system, which provides accurate pixel-wise lesion information. However, the detection-based system must integrate several models for ensemble and thus is more demanding of resources by comparison. Users must take into account specific computational requirements when choosing a preference.

Influence of Pneumothorax Size

Previous studies [42, 51, 52] have demonstrated that model performance (as with human readings) may be influenced by extent of pneumothorax. A model that Taylor et al. [52] devised correctly identified 100% of large pneumothoraces but only 39% of small ones. Similarly, performance levels of our CAD systems declined as pneumothorax size diminished. This is not surprising, because inter-annotator TP-Dice coefficients also fell as pneumothorax size decreased, underscoring the problematic model learning of small-volume lesions. This phenomenon was more obvious for the detection-based CAD system as its lower prediction-ground truth TP-DICE than the inter-annotator TP-DICE (Table 3) may lead to the lower diagnostic performance for small pneumothorax than the radiology reports (Table 2).

Unlike large pneumothoraces, small pneumothorax is apt to be overlooked by clinicians, especially on portable SCXRs, necessitating assistance by CAD systems. Because most patients subjected to portable SCXRs are those susceptible to complications caused by pneumothorax, especially those receiving mechanical ventilation, timely detection is critical to prevent a small pneumothorax from progressing into tension pneumothorax [53].

Future Applications

The CAD system can serve two primary functions: (1) prioritizing the SCXRs and selecting those in question to be checked first by the radiologist or (2) issuing notifications to attending clinicians. When the clinicians examine the

diagnosis results, the localization outputs of pneumothorax may pop up to facilitate verification of the results. We present the requirements of computational resources for these two CAD systems (Supplemental Table 5), which can assist healthcare institutions in selecting the most suitable model for deployment. Moreover, in future studies, it is warranted to examine the feasibility of adapting these CAD systems for edge computing and their integration into portable chest X-ray machines, which holds the potential to broaden the CAD systems' applicability.

Study Limitations

First, because we only have de-identified images available for analysis, we did not know whether patients' clinical comorbidities may influence the performance of the CAD system. Nonetheless, Table 1 shows there were diverse findings or diagnoses on SCXRs, which might somewhat mitigate this concern. Second, given the low prevalence for pneumothorax [35, 36], we used a case-controlled study design for image collection to ensure sufficient numbers of pneumothorax-positive patients. This design may result in an artificially elevated pneumothorax prevalence in our datasets, compared with real-world settings. We therefore relied on radiology reports or annotators as reader reference points by which to judge CAD system performance. Further prospective studies are warranted to better test performance with real-life pneumothorax prevalence by enrolling consecutive patients from EDs or ICUs on a manageable scale [54].

Conclusions

We developed two DL-based CAD systems to diagnose and localize pneumothoraces on portable SCXRs, using detection and segmentation methods, respectively. Performances of both systems proved excellent, comparable to those of radiologists or human annotators when tested in a dataset of differing time frame, with differing patient or image characteristics. Hence, the potential for external generalizability seems favourable. Although each performed similar in testing, the detection-based system may demand more in terms of computational resources.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10916-023-02023-1>.

Acknowledgements We thank the staff of the 3rd Core Lab, Department of Medical Research, National Taiwan University Hospital for technical support. We thank the Integrated Medical Database, National Taiwan University Hospital for assisting in acquiring images for analysis.

Author Contributions CHW: Conceptualization, Methodology, Validation, Resources, Formal analysis, Investigation, Data curation, Writing – original draft, Project administration; TL: Conceptualization, Methodology, Validation, Resources, Formal analysis, Investigation, Data curation, Writing – original draft, Project administration; GC: Conceptualization, Methodology, Validation, Resources, Formal analysis, Investigation, Data curation, Writing – original draft, Project administration; MRL: Resources, Formal analysis, Investigation, Data curation, Writing – review & editing; JT: Resources, Formal analysis, Investigation, Data curation, Writing – review & editing; CYW: Resources, Formal analysis, Investigation, Data curation, Writing – review & editing; MCW: Resources, Formal analysis, Investigation, Data curation, Writing – review & editing; HRR: Formal analysis, Investigation, Data curation; DY: Formal analysis, Investigation, Data curation; CZ: Formal analysis, Investigation, Data curation; WW: Conceptualization, Methodology, Validation, Resources, Formal analysis, Writing – review & editing, Supervision; CHH: Conceptualization, Methodology, Validation, Resources, Formal analysis, Writing – review & editing, Supervision.

Funding Author Chih-Hung Wang received a grant (112-UN0022) from the National Taiwan University Hospital. Authors Chih-Hung Wang, Weichung Wang, and Chien-Hua Huang received a grant (MOST 111-2634-F-002-015-, Capstone project) from the National Science and Technology Council, Taiwan. National Taiwan University Hospital and National Science and Technology Council had no involvement in designing the study, collecting, analysing or interpreting the data, writing the manuscript, or deciding whether to submit the manuscript for publication.

Declarations

Ethics Approval This study was approved by the Research Ethics Committee of the National Taiwan University Hospital (NTUH; reference number: 202003106RINC) and granted a consent waiver.

Competing Interests The authors declare no competing interests.

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/>.

References

- Olesen WH, Titlestad IL, Andersen PE, Lindahl-Jacobsen R, Licht PB (2019) Incidence of primary spontaneous pneumothorax: a validated, register-based nationwide study. *ERJ open research* 5:00022–02019
- Bobbio A, Dechartres A, Bouam S et al (2015) Epidemiology of spontaneous pneumothorax: gender-related differences. *Thorax* 70:653–658
- Tocino IM, Miller MH, Fairfax WR (1985) Distribution of pneumothorax in the supine and semirecumbent critically ill adult. *AJR Am J Roentgenol* 144:901–905
- Trotman-Dickenson B (2003) Radiology in the intensive care unit (Part I). *J Intensive Care Med* 18:198–210
- Hill JR, Horner PE, Primack SL (2008) ICU imaging. *Clin Chest Med* 29:59–76, vi
- Chan KK, Joo DA, McRae AD et al (2020) Chest ultrasonography versus supine chest radiography for diagnosis of pneumothorax in trauma patients in the emergency department. *Cochrane Database Syst Rev* 7:Cd013031
- Spillane RM, Shepard JO, Deluca SA (1995) Radiographic aspects of pneumothorax. *Am Fam Physician* 51:459–464
- Chiles C, Ravin CE (1986) Radiographic recognition of pneumothorax in the intensive care unit. *Crit Care Med* 14:677–680
- Ball CG, Kirkpatrick AW, Laupland KB et al (2005) Factors related to the failure of radiographic recognition of occult post-traumatic pneumothoraces. *Am J Surg* 189:541–546; discussion 546
- Tocino IM (1985) Pneumothorax in the supine patient: Radiographic anatomy. *RadioGraphics* 5:557–586
- Mongan J, Moy L, Kahn CE, Jr. (2020) Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. *Radiol Artif Intell* 2:e200029
- Chou EH, Wang CH, Chou FY et al (2022) Development and validation of a prediction model for estimating one-month mortality of adult COVID-19 patients presenting at emergency department with suspected pneumonia: a multicenter analysis. *Intern Emerg Med* 17:805–814
- Zhou W, Cheng G, Zhang Z et al (2022) Deep learning-based pulmonary tuberculosis automated detection on chest radiography: large-scale independent testing. *Quant Imaging Med Surg* 12:2344–2355
- Tan M, Le Q (2019) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: Kamalika C, Ruslan S, (eds) *Proceedings of the 36th International Conference on Machine Learning*. PMLR, *Proceedings of Machine Learning Research*, pp 6105–6114
- Huang G, Liu Z, Maaten LVD, Weinberger KQ (2017) Densely Connected Convolutional Networks 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 2261–2269
- Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the Inception Architecture for Computer Vision 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 2818–2826
- Zhu X, Su W, Lu L, Li B, Wang X, Dai J (2020) Deformable detr: Deformable transformers for end-to-end object detection. *arXiv preprint arXiv:201004159*
- Feng C, Zhong Y, Gao Y, Scott MR, Huang W (2021) Toood: Task-aligned one-stage object detection 2021 IEEE/CVF International Conference on Computer Vision (ICCV). *IEEE Computer Society*, pp 3490–3499
- Zhang H, Wang Y, Dayoub F, Sunderhauf N (2021) Varifocalnet: An iou-aware dense object detector Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 8514–8523
- Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation International Conference on Medical image computing and computer-assisted intervention. Springer, pp 234–241
- Radosavovic I, Kosaraju RP, Girshick R, He K, Dollár P (2020) Designing Network Design Spaces 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 10425–10433
- Youden WJ (1950) Index for rating diagnostic tests. *Cancer* 3:32–35

23. Virtanen P, Gommers R, Oliphant TE et al (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 17:261–272
24. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM (2017) ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 3462–3471
25. Irvin J, Rajpurkar P, Ko M et al (2019) CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. *Proceedings of the AAAI Conference on Artificial Intelligence* 33:590–597
26. Filice RW, Stein A, Wu CC et al (2020) Crowdsourcing pneumothorax annotations using machine learning annotations on the NIH chest X-ray dataset. *J Digit Imaging* 33:490–496
27. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM (2017) ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 3462–3471
28. Oakden-Rayner L (2017) Exploring the ChestXray14 dataset: problems. *Wordpress: Luke Oakden Rayner* 1
29. Kitamura G, Deible C (2020) Retraining an open-source pneumothorax detecting machine learning algorithm for improved performance to medical images. *Clinical imaging* 61:15–19
30. Rueckel J, Huemmer C, Fieselmann A et al (2021) Pneumothorax detection in chest radiographs: optimizing artificial intelligence system for accuracy and confounding bias reduction using in-image annotations in algorithm training. *Eur Radiol* 31:7888–7900
31. Cho Y, Kim JS, Lim TH, Lee I, Choi J (2021) Detection of the location of pneumothorax in chest X-rays using small artificial neural networks and a simple training process. *Sci Rep* 11:13054
32. Mosquera C, Diaz FN, Binder F et al (2021) Chest x-ray automated triage: A semiologic approach designed for clinical implementation, exploiting different types of labels through a combination of four Deep Learning architectures. *Comput Methods Programs Biomed* 206:106130
33. Sakai M, Hiyama T, Kuno H et al (2020) Thoracic abnormal air collections in patients in the intensive care unit: radiograph findings correlated with CT. *Insights into Imaging* 11:35
34. Tocino IM, Miller MH, Fairfax W (1985) Distribution of pneumothorax in the supine and semirecumbent critically ill adult Presented at the annual meeting of the American,
35. Anzueto A, Frutos-Vivar F, Esteban A et al (2004) Incidence, risk factors and outcome of barotrauma in mechanically ventilated patients. *Intensive Care Med* 30:612–619
36. Surleti S, Famà F, Murabito L, Villari SA, Bramanti C, Florio G (2011) Pneumothorax in the Emergency Room: personal caseload. *Il giornale di chirurgia* 32:473–478
37. Lever J, Krzywinski M, Altman N (2016) Classification evaluation. *Nat Methods* 13:603–604
38. Grimes DA, Schulz KF (2005) Compared to what? Finding controls for case-control studies. *Lancet* 365:1429–1433
39. Schulz KF, Grimes DA (2002) Case-control studies: research in reverse. *Lancet* 359:431–434
40. Collins GS, Reitsma JB, Altman DG, Moons KG (2015) Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 162:55–63
41. Niehues SM, Adams LC, Gaudin RA et al (2021) Deep-Learning-Based Diagnosis of Bedside Chest X-ray in Intensive Care and Emergency Medicine. *Invest Radiol* 56:525–534
42. Rueckel J, Trappmann L, Schachtner B et al (2020) Impact of Confounding Thoracic Tubes and Pleural Dehiscence Extent on Artificial Intelligence Pneumothorax Detection in Chest Radiographs. *Invest Radiol* 55:792–798
43. Ganapathy A, Adhikari NK, Spiegelman J, Scales DC (2012) Routine chest x-rays in intensive care units: a systematic review and meta-analysis. *Crit Care* 16:R68
44. Hooper KP, Anstey MH, Litton E (2021) Safety and efficacy of routine diagnostic test reduction interventions in patients admitted to the intensive care unit: A systematic review and meta-analysis. *Anaesth Intensive Care* 49:23–34
45. Rachh P, Levey AO, Lemmon A et al (2018) Reducing STAT Portable Chest Radiograph Turnaround Times: A Pilot Study. *Current problems in diagnostic radiology* 47:156–160
46. Lee SY, Ha S, Jeon MG et al (2022) Localization-adjusted diagnostic performance and assistance effect of a computer-aided detection system for pneumothorax and consolidation. *npj Digital Medicine* 5:107
47. Feng S, Liu Q, Patel A et al (2022) Automated pneumothorax triaging in chest X-rays in the New Zealand population using deep-learning algorithms. *J Med Imaging Radiat Oncol*. <https://doi.org/10.1111/1754-9485.13393>
48. Zhou L, Yin X, Zhang T et al (2021) Detection and Semiquantitative Analysis of Cardiomegaly, Pneumothorax, and Pleural Effusion on Chest Radiographs. *Radiol Artif Intell* 3:e200172
49. Herron JM, Bender TM, Campbell WL, Sumkin JH, Rockette HE, Gur D (2000) Effects of luminance and resolution on observer performance with chest radiographs. *Radiology* 215:169–174
50. Wang H, Gu H, Qin P, Wang J (2020) CheXLocNet: Automatic localization of pneumothorax in chest radiographs using deep convolutional neural networks. *PLoS One* 15:e0242013
51. Lee SY, Ha S, Jeon MG et al (2022) Localization-adjusted diagnostic performance and assistance effect of a computer-aided detection system for pneumothorax and consolidation. *NPJ Digit Med* 5:107
52. Taylor AG, Mielke C, Mongan J (2018) Automated detection of moderate and large pneumothorax on frontal chest X-rays using deep convolutional neural networks: A retrospective study. *PLoS Med* 15:e1002697
53. Kollef MH (1991) Risk factors for the misdiagnosis of pneumothorax in the intensive care unit. *Crit Care Med* 19:906–910
54. Park SH (2019) Diagnostic case-control versus diagnostic cohort studies for clinical validation of artificial intelligence algorithm performance. *Radiology* 290:272–273
55. Zuiderveld KJ (1994) Contrast Limited Adaptive Histogram Equalization *Graphics Gems*,
56. Solovyev R, Wang W, Gabruseva T (2021) Weighted boxes fusion: Ensembling boxes from different object detection models. *Image and Vision Computing* 107:104117
57. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation International Conference on Medical image computing and computer-assisted intervention. Springer, pp 234–241

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