Using AI to Identify Chest Radiographs with No Actionable Disease in Outpatient Imaging

Awais Mansoor (awais.mansoor@siemens-healthineers.com)
Siemens Healthineers https://orcid.org/0000-0001-6761-8030

Ingo Schmuecking
Siemens Healthineers

Florin-Cristian Ghesu
Siemens Healthineers Gmbh

Bogdan Georgescu
Siemens Healthineers

Sasa Grbic
Siemens Healthineers

R S Vishwanath
Siemens Healthineers

Oladimeji Farri
Siemens Healthineers https://orcid.org/0000-0002-1840-2062

Rikhiya Gosh
Siemens Healthineers

Ramya Vunikili
Siemens Healthineers

Mathis Zimmermann
Siemens Healthineers https://orcid.org/0000-0001-6170-6566

James Sutcliffe
Zwanger-Pesiri Radiology

Steven Mendelsohn
Zwanger-Pesiri Radiology

Warren Gefter
Department of Radiology, Penn Medicine, University of Pennsylvania

Dorin Comaniciu
Siemens Healthineers

Article

Keywords:

Posted Date: June 26th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2924070/v1

Additional Declarations: Yes there is potential Competing Interest. W.B.G. received consulting fees from Siemens Healthineers to support the research collaboration. Zwanger-Pesiri Radiology received funding from Siemens Healthineers to support the research collaboration. The remaining authors are employees of Siemens Healthineers. This study was funded by Siemens Healthineers. The authors have no other competing interests to disclose.

Using AI to Identify Chest Radiographs with No Actionable Disease in Outpatient Imaging

Background: Chest radiographs are one of the most frequently performed imaging examinations in radiology. Chest radiograph reading is characterized by a high volume of cases, leading to long worklists. However, a substantial percentage of chest radiographs in outpatient imaging are without actionable findings. Identifying these cases could lead to numerous workflow efficiency improvements.

Objective: To assess the performance of an AI system to identify chest radiographs with no actionable disease (NAD) in an outpatient imaging population in the United States.

Materials and Methods: The study includes a random sample of 15,000 patients with chest radiographs in posterior-anterior (PA) and optional lateral projections from an outpatient imaging center with multiple locations in the Northeast United States. The ground truth was established by manually reviewing procedure reports and classifying cases as non-actionable disease (NAD) or actionable disease (AD) based on predetermined criteria. The NAD cases include both completely normal chest radiographs without any abnormal findings and radiographs with non-actionable findings. The AI NAD Analyzer[1] trained on more than 1.3 million radiographs provides a binary case level output for the chest radiographs as either NAD or potential actionable disease (PAD). Two systems A (more specific) and B (more sensitive) were trained. Both systems were capable of processing either frontal only or frontal-lateral pair.

Results: After excluding patients < 18 years (n=861) as well as the cases not meeting the image quality requirements of the AI NAD Analyzer (n=82), 14057 cases (average age 56±16.1 years, 7722 women and 6328 men) remained for the analysis. The AI NAD Analyzer with input consisting of PA and lateral images, correctly classified 2891 cases as NAD with concordance between ground truth and AI, which is 20.6% of all cases and 29.1% of all ground truth NAD cases. The miss rate was 0.3% and included 0.06% significant findings. With a more specific version of the AI NAD Analyzer (System A), there were 12.2% of all NAD cases were identified correctly with a miss rate of 0.1%. No cases with critical findings were missed by either system.

Conclusion: The AI system can identify a meaningful number of chest radiographs with no actionable disease in an outpatient imaging population with a very low rate of missed findings.

[1] For research purposes only. Not for clinical use. Future commercial availability cannot be guaranteed.

Introduction

Chest radiographs (CXR) are one of the most frequently performed imaging examinations in radiology. Outpatient CXR may be acquired as an initial imaging study for a wide range of clinical indications including to confirm or rule out actionable disease. This results in a high percentage of clinically unremarkable exams, i.e., those which are completely normal as well as those with clinically non-actionable findings such as age-related changes.

However, the interpretation of CXR is a complex task and findings may be missed [1], [2]. This is primarily due to inherent limitations of the modality, i.e., the superimposition of structures on 2D projection images and the subtlety of many findings, coupled with required reading efficiency. Moreover, infrequent critical abnormalities among a large number of non-actionable cases raises the possibility of overlooking these important findings. Accurately performing radiological reading in a relatively short span of time without overlooking critical or significant findings is challenging. Unsurprisingly, chest radiography is one of the modalities with the highest number of malpractice lawsuits [3] [4]. Furthermore, the length of the reading worklist may not always allow for fast turnaround for time-critical studies; therefore, creating a workload reprioritization strategy to identify exams which can remain longer on the worklist with low risk of causing patient harm could create significant benefits in the clinical workflow optimization.

CXR frequently show a range of findings which are not considered actionable in a typical clinical context. This is especially true for the images acquired within an outpatient setting.. While non-actionable findings may be mentioned in the report for the purposes of completeness, the overall outcome of the exam can often be summarized under the term “No actionable disease” (NAD). The use of artificial intelligence (AI) to identify NAD radiographs can help with worklist optimization, enable efficient reading with added confidence, and increase reading accuracy. Recent advances in AI have enabled new models to learn the interpretation of entire chest radiographs using very large data volumes. Compared to earlier AI models that generally focused on detecting specific abnormalities, such as pulmonary nodules, these new AI models can be trained to learn from the entire image. That includes holistically looking at the image for ruling out abnormalities as well as exclusively at different organs or “compartments” (lungs, pleura, hila, mediastinum, trachea, cardiac, diaphragm, upper abdomen, bones, soft tissue, hardware) for a comprehensive assessment of subtle harder to rule out abnormalities. In recent studies, the use of AI to identify normal cases has gained traction [5], [6], [7] [8], [9].

In this retrospective study, we evaluate an AI NAD Analyzer to identify chest radiographs with NAD in an outpatient imaging population in the United States. The system uses the state-of-the-art self-supervised learning methodology based on contrastive learning and online feature clustering [10]. The backbone of the system has been trained on more than 1.3 million radiographs collected from diverse sites and demographics. The use of diverse big data allows the AI NAD Analyzer to learn variabilities of normal chest anatomy under a host of acquisition parameters.

The objective of this study is to assess the performance of the AI NAD Analyzer for the identification of normal or unremarkable chest radiographs in a representative outpatient population.

Definitions and Abbreviations

AD	Actionable Disease
AI	Artificial Intelligence
AP	Anterior-posterior projection
AVM	Arteriovenous malformation
CADe	Computer-aided detection
COPD	Chronic obstructive pulmonary disease
CXR	Chest radiographs
GT	Ground truth
ICD	Implantable cardioverter defibrillator
IDRI	Image Database Resource Initiative
LIDC	Lung Image Database Consortium
NAD	No Actionable Disease
NLP	Natural Language Processing
PA	Posterior-anterior projection
PAD	Potential actionable disease
PICC	Peripherally inserted central catheter
PTX	Pneumothorax
TB	Tuberculosis

Methods

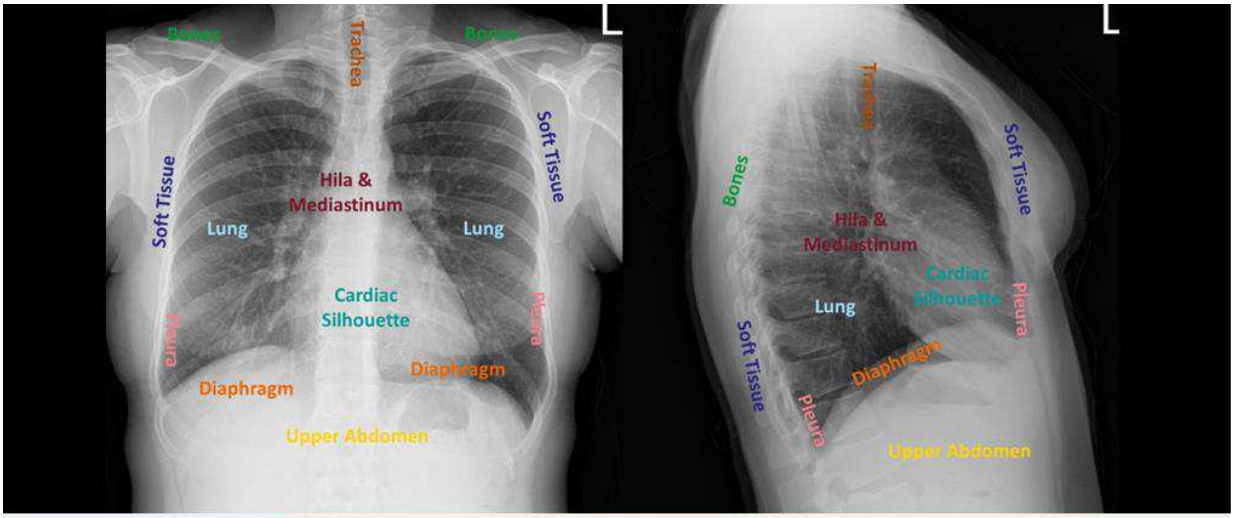
Data Selection

The cohort selected for performance evaluation consists of representative cases from Zwanger-Pesiri Radiology, LLP., an outpatient imaging center with multiple locations in the Northeast United States. Consecutive CXR between August 2020 and August 2022 were collected from the archive including lateral projections and corresponding procedure reports. The retrospectively collected data were anonymized per HIPAA guidelines at the source before secure transfer to Siemens Healthineers for the study. There was no intervention or interaction with individual patients or impact on patient management. IRB assessment has determined exempt status for the nature of the research project. A random sample of 15,000 cases was taken from the consecutive cases. The study includes patients 18 years and older. Images from multiple vendors were included. Follow-up exams from the same patient were not included in the study. Cases in the study were not used for the algorithm training.

Ground Truth

Ground Truth (GT) Definition: The criteria for the definition of chest radiographs with no actionable disease (NAD) are summarized in Figure 1. The CXR exam was divided into compartments and then a detailed analysis of abnormalities in each compartment was performed with reference to literature [11] [12]. Specifically, for each compartment, a list is compiled of abnormalities that can potentially occur in the compartment and can be seen on a chest radiograph. The abnormalities were then classified into NAD and actionable disease (AD) categories. Abnormal findings indicating AD were further grouped into categories related to clinical impact, e.g., critical findings which require urgent communication, significant findings which require clinical follow up, and other findings which were deemed neither urgent nor significant but clinically noteworthy.

Ground Truth Annotation: Manual review of procedure reports was performed using the classification scheme devised above (shown in Figure 1). Initial review was performed by at least one of three reviewers (AM, RV, IS). Cases with uncertain classification were sent to an expert radiologist (RSV) with more than seven (7) years of experience for adjudication. Finally, a handful of reports that remained uncertain after the radiologist review were assigned as AD out of an abundance of caution.



NO ACTIONABLE DISEASE	ACTIONABLE DISEASE

NO ABNORMALITIES

FINDINGS

- Platelike atelectasis, linear scar
- Aortic tortuosity, calcification
- Symmetric bi-apical pleural thickening (capping)
- Calcified granuloma
- Diaphragm eventration
- Epicardial fat pad
- Mild scoliosis/kyphosis
- Degenerative bone changes (without compression fractures)
- Healed fractures
- Normal variants
- Developmental variants, e.g azygos fissure

HARDWARE

- Necklace and earring artifacts
- ECG leads
- Nipple markers
- Surgical clips
- Sternotomy cerclage wires
- Metal plates/screws/surgical pins

CRITICAL

Lung abscess
Large pleural effusion
Findings suggesting active TB
Whole lung atelectasis
Diffuse airspace disease, e.g. pulmonary edema
Suspect pulmonary infarct
Subcutaneous emphysema

Pneumothorax
Pneumoperitoneum
Pneumomediastinum
Malpositioned hardware
Foreign body
Acute mediastinal widening
Acute aortic syndrome
Acute fracture/dislocation

SIGNIFICANT

Pulmonary nodule/mass
Pneumonia
Lobar atelectasis
Mediastinal/hilar mass
Neck mass
Lymphadenopathy
Bone metastasis

Cardiomegaly
Pulmonary venous congestion
Pericardial effusion
Moderate pleural effusion
Pleural tumor
Bone tumor
Aortic aneurysm / dilatation

COMMONLY MISSED

OTHER FINDINGS

Other abnormal lung opacities, COPD, bronchiectasis, small pleural effusion, pleural plaque, deviated trachea, elevated diaphragm, flattened diagram, bullae, vertebral compression fractures, abdominal calcifications, e.g. gallstones, kidney stones, benign bone sclerotic lesions (e.g. infarct), moderate or severe scoliosis/kyphosis, hardware (e.g. pacemakers, ICD, venous catheters, tubes, cardiac valves), hernia, calcified lymph nodes, cervical rib.

Figure 1. Criteria for no actionable disease (NAD) and actionable disease (AD) covering all compartments in chest radiographs.

AI NAD Analyzer

The flow diagram of the AI NAD Analyzer[2] is shown in Figure 2. The AI NAD Analyzer is based on self-supervised pretraining using 100 million multimodal images and finetuned on 1.3 million radiographs [10]. The AI system can process both PA and lateral chest radiographs. The processing starts with a module that verifies the input image quality (within domain, proper field-of-view, inspiration, penetration, and rotation). Any image failing on any of the quality metrics is rejected for further processing. The quality assessment check is followed by multiple classifiers in tandem system (image-level rule in, compartment-level rule out). The detection systems for individual compartments enumerated previously are trained in a multi-class setting, i.e., a single parametric learning model supports the joint classification and detection of these abnormalities. As such, during training there is an

[2] For research purposes only. Not for clinical use. Future commercial availability cannot be guaranteed.

active and implicit transfer of class conditional information in the learning model. Specifically, different classifiers work in tandem to provide holistic image-level and more in-depth compartment-level assessment. The aggregated assessment from various classifiers produces a unified binary case-level assessment by the system. The AI NAD Analyzer provides a binary case-level output: “No Actionable Disease (NAD)” or “Potential Actionable Disease (PAD)”. Further details of the network architecture, input/output, and pre/post-processing steps can be found in the Appendix.

Two systems, System A and System B, were tuned with the same backbone architecture with System A tuned to being more specific than System B. Both systems are capable of processing PA-alone or PA+lateral pair.

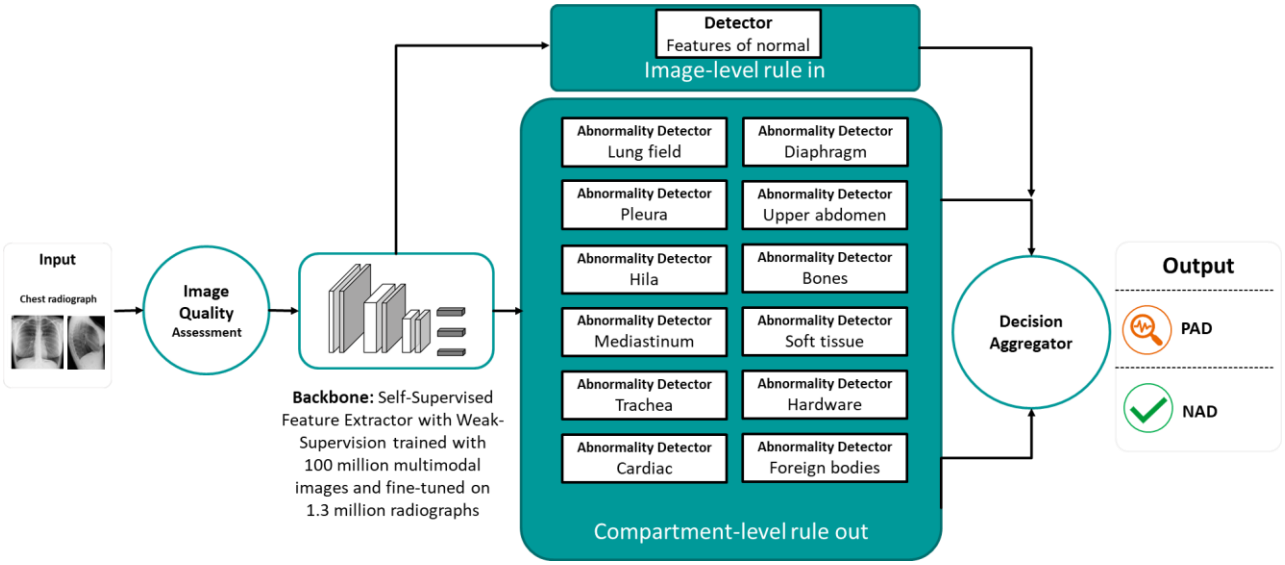


Figure 2. Flow diagram of the AI NAD Analyzer. PAD = Potential Actionable Disease, NAD = No Actionable Disease.

Stress Testing and System Validation

To ensure system reliability in ruling out abnormalities, the AI NAD Analyzer underwent stress testing on a dataset (not used in training) of 1905 cases across 17 finding classes. This dataset does not include cases with NAD. The objective is to cover a broad range of critical and significant findings to assure their accurate detection by the algorithm and to ensure system performance on rare findings. Each case was reviewed and annotated by two radiologists (EK with 24 years, RSV with over 7 years of experience) unless otherwise noted.

The finding class ‘Pulmonary Nodules and Masses’ includes cases from The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI) [13]. The LIDC/IDRI dataset is constructed by the collaborative effort of seven academic centers and eight medical imaging companies to create a robust dataset for computer-aided diagnostic (CAD) methods validation for lung nodule detection. Seventy-five (75) positive cases from the database were chosen by two radiologists (EK, RSV)

to cover a wide range of location, size, and conspicuity levels. The positive findings of the 75 cases were confirmed by chest CT.

The finding class 'Findings concerning for TB' includes 146 positive cases from The TB Portals [14], a repository of socioeconomic/geographic, clinical, laboratory, radiological, and genomic data from patient cases of drug-resistant tuberculosis backed by shareable, physical samples. The portal includes 373 cases which contain chest radiographs. A radiologist (EK) analyzed them individually to identify cases with the positive imaging biomarkers of TB.

Cases for the rest of the finding classes were acquired from multiple outpatient sites and were reviewed and confirmed by two radiologists (EK, RSV).

Results of the stress testing with NAD AI Analyzer are shown in Table 1. Cases with multiple findings (e.g., pulmonary nodule and scoliosis) are counted in each finding class.

None of the AD cases in the stress testing cohort were classified as NAD by AI.

Compartment	Finding Class	Description	Cases	NAD
Lung	Pulmonary nodules and masses	Nodules and masses, incl. subtle lung nodules and challenging locations (blind spots)	140	0
Lung	Airspace opacities (consolidation)	Includes challenging areas such as retrocardiac and overlying the diaphragms	67	0
Lung	Atelectasis	Range from lobar to whole lung atelectasis	95	0
Lung	Suggest COPD (emphysema)	Flattened diaphragms, definite hyperinflation, decreased pulmonary vascularity	245	0
Lung	Findings concerning for TB	Range of findings which may suggest possible TB	146	0
Pleura	Pleural effusion	Range from small to large effusions	85	0
Pleura	Pneumothorax	Range of sizes (small to large) and locations	154	0
Pleura	Pneumomediastinum	Range from subtle to large	15	0
Pleura	Pneumoperitoneum	Range from subtle to large	21	0
Cardiac	Cardiomegaly	Range from mild to marked	172	0
Mediastinum	Hilar enlargement	Lymphadenopathy / mass, dilated central pulmonary arteries (e.g. arterial pulmonary hypertension)	7	0
Mediastinum	Hernia	Hiatal hernia	104	0
Mediastinum	Aortic aneurysm or dilatation	Aneurysm (focal) or dilatation	76	0
Soft tissue	Subcutaneous emphysema	Range from mild to marked	11	0
Bones	Scoliosis / Kyphosis	More than mild scoliosis or kyphosis	305	0
Bones	Fracture	Vertebra, ribs, humerus	121	0
Hardware	Devices	Pacemaker, ICD, central venous lines (PICCs and port catheters), enteric tubes, metallic prosthetic cardiac valves, cardiac loop monitor, vertebroplasty	245	0
Total findings			**2009**	**0**
Total cases			**1905**	**0**

Table 1. Stress testing across 17 finding classes using System B

Results

The dataset consisting of a random sample of 15,000 patients with chest radiographs in posterior-anterior (PA) and optional lateral projections were used for testing the performance of the NAD AI Analyzer. After excluding patients with age < 18 years (n=861) and cases not meeting the image quality requirements for the algorithm (n=82), there were 14,057 cases remaining for analysis. The average age of the subjects within the remaining set was 56±16.1 years. The population is gender balanced with 7,722 women and 6,328 men. All cases had PA and lateral chest radiographs. See Table 2 for further details about the study cases.

Parameter	
Age	Average **56±16.1** years
Gender	Female: **7722** (55%) Male: **6328** (45%)
Projection	PA: **14057** (100%) Lateral: **14057** (100%)
Manufacturer	Samsung Electronics: **13495** (96.0%) Swissray: **404** (2.9%) Fujifilm Corporation: **55** (0.4%) Carestream Health: **11** (0.1%) Other manufacturers: **10** (0.1%) Missing: 82 (0.6%)

Table 2. Study cases

As shown in Figure 3, the ground truth was NAD for 9,940 of 14,057 cases (70.7%) and AD for 4,117 of 14,057 cases (29.3%).

With a more sensitive system B and input of PA and lateral image, there were 2,891 cases with NAD concordance between GT and AI, which is 20.6% of all cases and 29.1% of cases with NAD as GT. Representative example cases are shown in Figure 5. There were 47 cases with GT as AD which were classified as NAD by AI. This represents 0.3% of all cases, 1.1% of the GT AD cases and 1.6% of cases with NAD by AI. Analysis for these 47 cases is shown in Figure 4 and Table 3Table 3. Based on our analysis, there were no missed critical findings. However, there were 9 significant findings which is 0.06% of all cases. Other findings missed by AI include many cases with mild abnormalities, e.g., increased interstitial markings. Example cases with missed findings are shown in Figure 6.

When using the PA chest radiograph only as input for the AI NAD Analyzer with System B, there were 2,290 cases correctly classified as NAD (21.3% of total) at a higher miss rate of 56 cases (0.4% of total cases).

When using a more specific System A and input of PA and lateral image, the miss rate of AI decreased from 0.3% to 0.1% of all cases. However, the cases with NAD concordance between GT and AI also decreased from 20.6% to 12.2% of all cases.

Figure 7 shows the distribution of ground truth and AI results by patient age for 13,975 cases where the age information was available.

		AI NAD Analyzer					
		PA + Lateral			PA Only		
Ground Truth		NAD	PAD		NAD	PAD	
	System A (more specific)						
	NAD	1714 (12.2%)	8226 (58.5%)	9940 (70.7%)	1778 (12.6%)	8162 (58.1%)	9940 (70.7%)
	AD	19 (0.1%)	4098 (29.2%)	4117 (29.3%)	23 (0.2%)	4094 (29.1%)	4117 (29.3%)
		1733 (12.3%)	12324 (87.7%)	14057 (100%)	1801 (12.8%)	12256 (87.2%)	14057 (100%)
	System B (more sensitive)						
	NAD	2891 (20.6%)	7049 (50.1%)	9940 (70.7%)	2990 (21.3%)	6950 (49.4%)	9940 (70.7%)
	AD	47 (0.3%)	4070 (29.0%)	4117 (29.3%)	56 (0.4%)	4061 (28.9%)	4117 (29.3%)
		2938 (20.9%)	11119 (79.1%)	14057 (100%)	3046 (21.7%)	11011 (78.3%)	14057 (100%)

Figure 3. Case level analysis for System A and System B. AI result as NAD or PAD. Ground truth as NAD or AD.

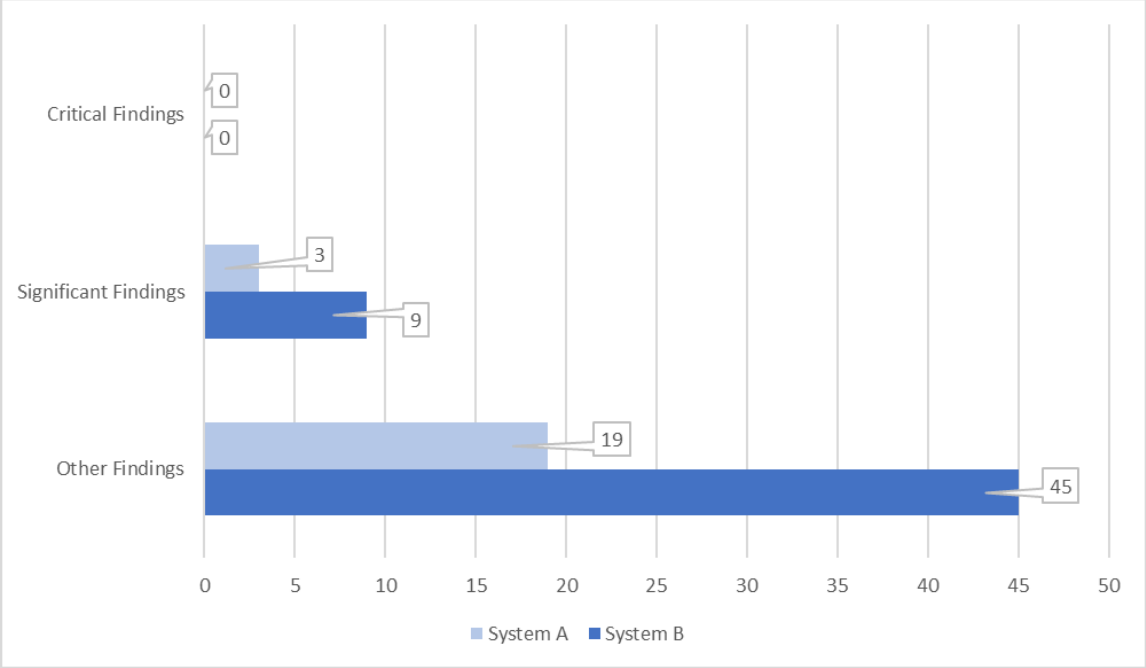


Figure 4. Findings level analysis of cases with AD in the ground truth and NAD by System A and System B. PA and lateral as input. No cases with critical findings were missed by AI.

Compartment	Finding	Category	Missed System A	Missed System B
Mediastinum	Aortic aneurysm	Significant		**1**
Mediastinum	Dilated/Ectatic aorta	Significant	**1**	**2**
Lung	Arteriovenous malformation	Significant	**1**	**1**
Lung	Subtle density right upper lobe	Significant		**1**
Lung	Small density	Significant		**1**
Lung	Nodular density which may represent nipple shadow	Significant	**1**	**1**
Upper abdomen	Calcified mass in the region of the spleen	Significant		**1**
Lung	Increased interstitial markings, initial manifestation of Covid pneumonia	Significant		**1**
Bone	Vertebral compression	Other Finding	**1**	**2**
Lung	Bronchiectasis	Other Finding		**1**
Diaphragm	Elevated diaphragm	Other Finding	**2**	**5**
Mediastinum	Hiatal hernia	Other Finding		**2**
Lung	Hyperinflated lungs	Other Finding	**4**	**7**
Lung	Increased interstitial markings	Other Finding	**8**	**16**
Lung	Slight stable scarring	Other Finding		**1**
Bones	Kyphosis	Other Finding	**3**	**3**
Bones	Moderate scoliosis	Other Finding	**1**	**2**
Bone	Scoliosis	Other Finding		**4**
Hardware	Hardware	Other Finding		**2**
Total			**22**	**54**
Critical Findings			0	0
Significant Findings			3	9
Other Findings			19	45

Table 3. Findings level analysis of cases with AD GT and AI NAD by System A and System B. PA and lateral as input.

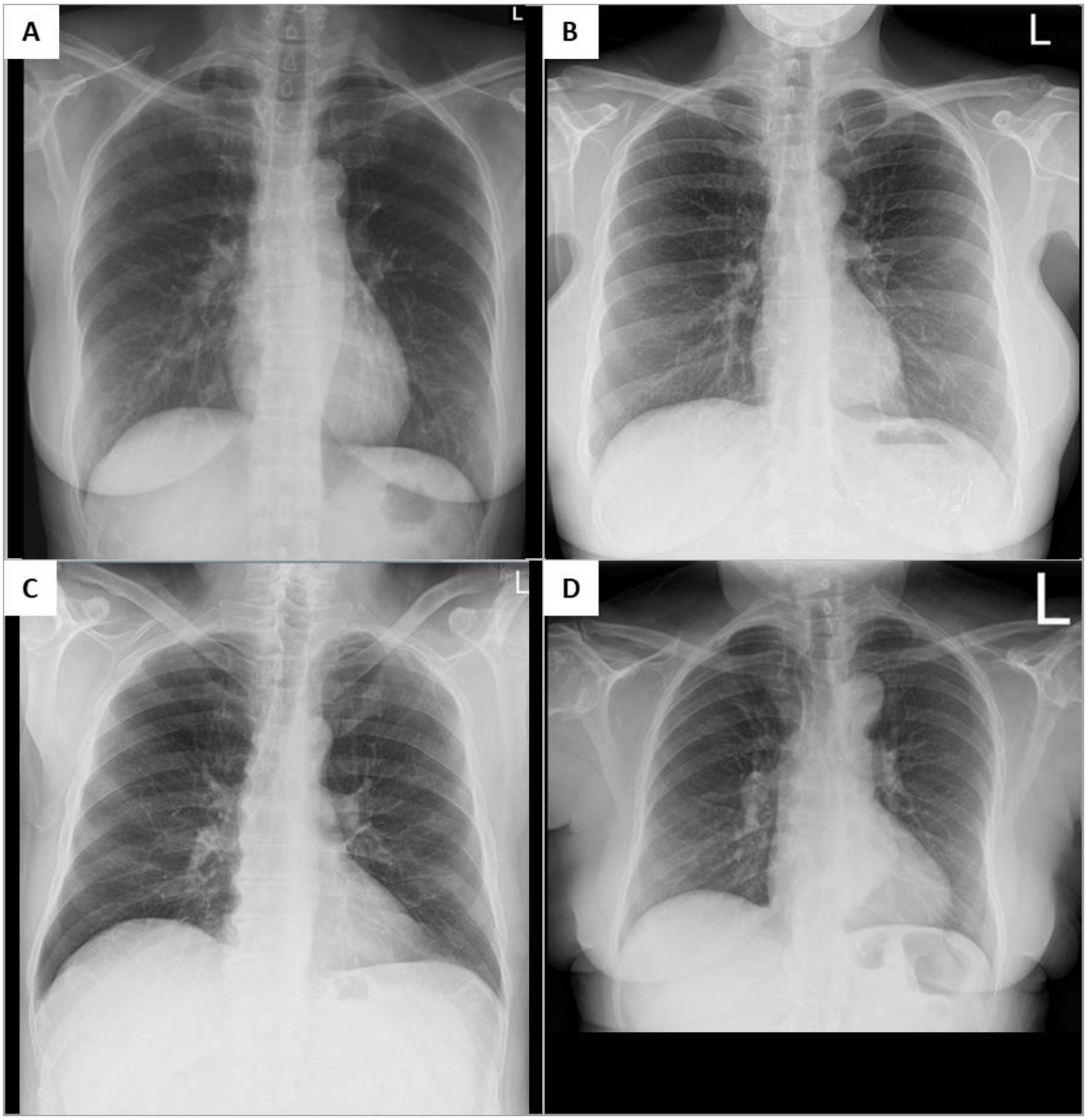


Figure 5. Cases with NAD in the ground truth and NAD classification by System B. PA and lateral pair as input. Normal chest radiograph without findings (A), and with findings included in the NAD definition (B-D). Presence of surgical clips (B), degenerative bone changes (C), and aortic tortuosity (D).

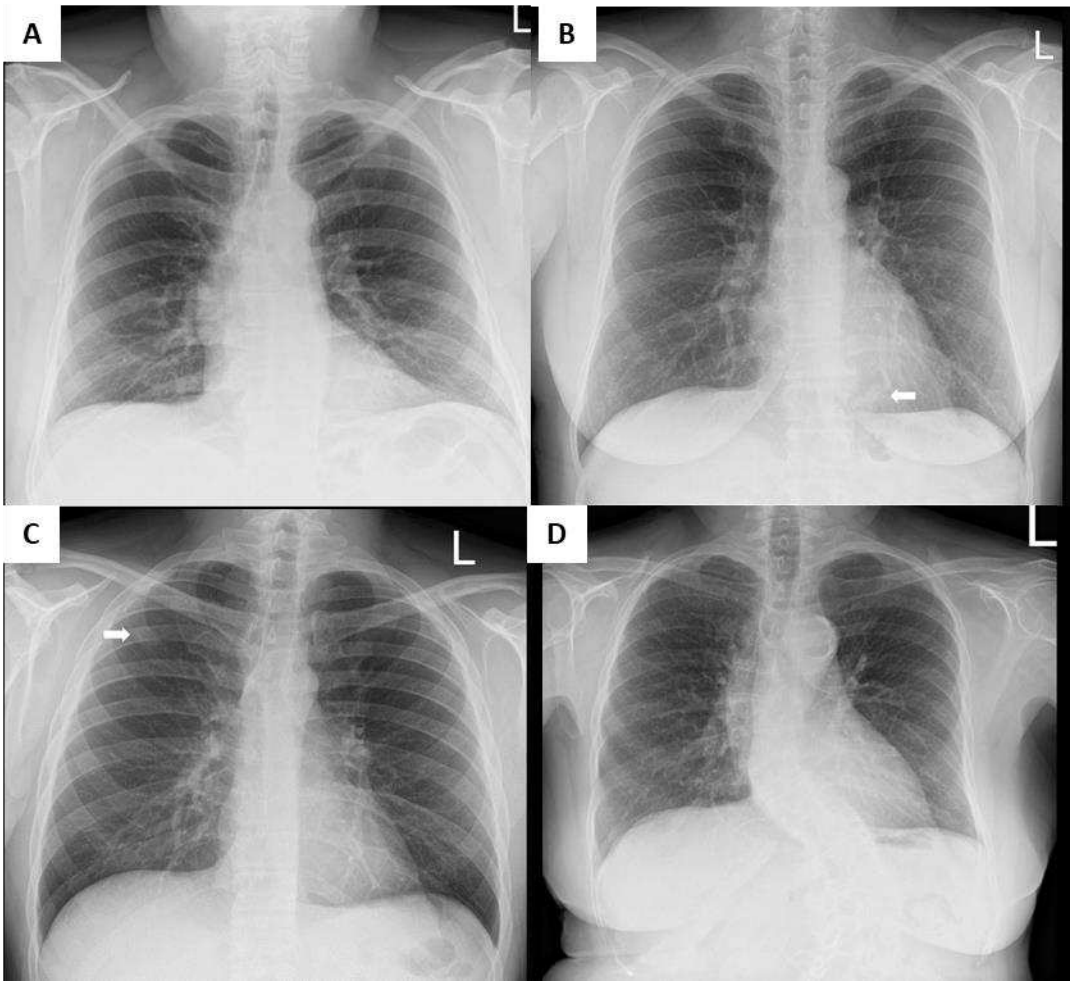


Figure 6. Cases with AD in ground truth which were classified as NAD by System B. PA and lateral pair as input. Examples for Significant Findings (A-C) and Other Findings category (D). Prominent bulge in the region of the ascending aorta compatible with aortic aneurysm (A), arteriovenous malformation (arrow) in the left lower lobe (B), subtle increased density (arrow) overlying the anterior second rib at its junction with the posterior fifth rib with chest CT recommended for further assessment (C), and moderate S-shaped thoraco-lumbar scoliosis (D).

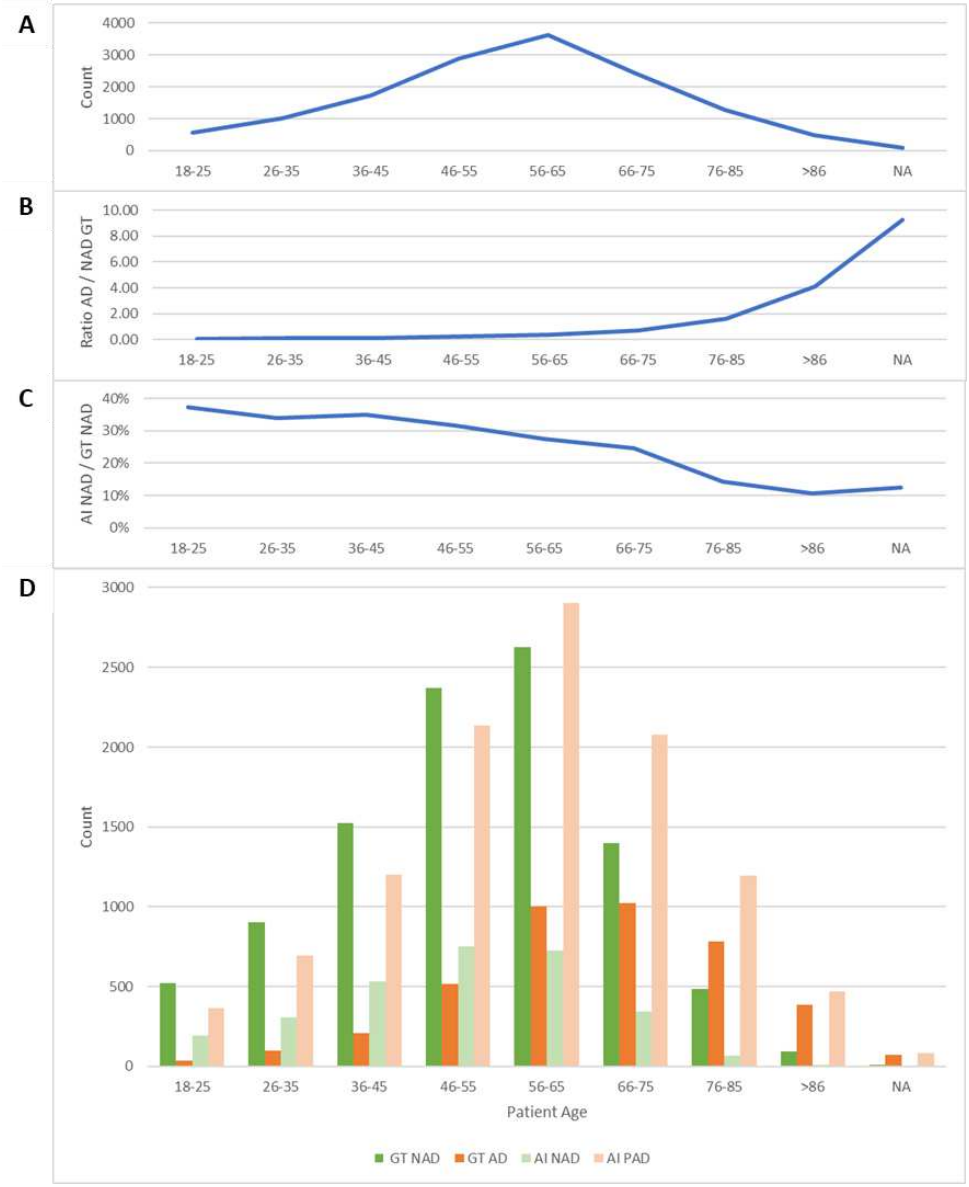


Figure 7. Ground truth and AI results by patient age. Age distribution (A). Ratio of AD to NAD ground truth (B). Percentage of AI NAD by GT NAD (C). Ground truth and AI results by age (D). System B with PA and lateral pair as input. NA=age information not available.

Discussion

Using AI to identify chest radiographs with no actionable disease (NAD) has the potential to benefit workflow in outpatient radiology, especially in situations with high case volumes which often lead to long worklists. Accurate identification of NAD cases may also provide enhanced confidence to the radiologist and enable efficient review. These cases normally exhibit lower risk when remaining on the worklist in a triage setting. Outpatient imaging is likely to have the highest prevalence of unremarkable exams among the care settings in which chest radiographs are acquired. A recent study by Plesner et al. [8] showed that an AI system can identify up to 28% of normal chest radiographs which represented

only 7.8% of all PA chest radiographs in a population of inpatients, emergency department patients and outpatients.

This paper is a novel attempt at establishing a comprehensive definition of what constitutes an NAD case for chest radiographs. Specifically, our study emphasized that in addition to normal chest radiographs, the NAD definition also includes selected non-actionable findings. This represents an important difference to prior studies which focus mostly on the analysis of completely normal chest radiographs. While both terms may not have universally accepted definitions, we believe that NAD is the more relevant concept for clinical practice in outpatient imaging. Unlike a vaguely constructed criteria for classification adopted by prior approaches, our study not only provided an exhaustive list of abnormalities seen in a standard chest radiograph, but also their classification as AD or NAD. This provides a blueprint for how a device to identify NAD cases can be designed and validated with results that can be reproduced. The criteria established in this study can be used to validate future such systems and can be extended to other imaging modalities.

Our study shows that 20.6% of chest radiographs and 29.1% of NAD chest radiographs in an outpatient imaging population in the United States can be correctly identified as NAD with a high degree of accuracy at a miss rate of only 0.3% (47 of 14,057). This is a low miss rate when put in the context of human errors while interpreting chest radiographs [1]. It is important to note here that the cases missed by the AI NAD Analyzer (System B) did not include any critical findings. However, there were 9 significant findings (0.06% of all cases). This includes 3 potential findings for which the procedure report recommended additional imaging to confirm or rule out the finding. Additionally, there were also 2 findings that were already known from a prior exam. To ensure performance and reliability, the AI NAD Analyzer underwent stress testing across 17 commonly encountered critical or significant finding classes with no cases incorrectly classified as NAD. Stress testing on large datasets as done in this study is required because some critical and significant findings have low prevalence in outpatient imaging populations; however, these need to be detected with high accuracy when present. The size of the stress test dataset may need to be further expanded. For example, while the AI NAD Analyzer (System B) did not miss any of the 104 hiatal hernia findings in stress testing, there were still 2 findings missed in the study.

All cases in our study included chest radiographs in PA and lateral projection. However, the practice of acquiring lateral CXR varies across healthcare providers, geographical regions, as well as dictated by concerns related to the exposure of some subjects to unnecessary radiation dose (pediatrics, pregnant women, etc.). Access to lateral chest radiographs showed a slightly lower miss rate of the AI NAD Analyzer (0.3% with lateral compared to 0.4% without lateral radiographs for System B). When lateral radiographs are available in clinical practice, they contain important information contributing to accurate interpretation of chest radiographs. The AI system may not have yet realized the full potential of analyzing lateral radiographs to improve accuracy.

The more specific System A has a lower yield of NAD cases, i.e., 12.2% of all cases but also a lower miss rate of 0.1% (down from 0.3% of system B). Over time, the percentage of NAD cases detected by the system as NAD will increase as the future iterations of the system learn to distinguish non-actionable from actionable findings.

It is important to also highlight the shortcomings of the current study. Specifically, the standalone performance testing does not evaluate the impact on radiologists using the AI system. Clinical context,

such as the reason for the exam, and comparison to prior exams are important components of the chest radiograph interpretation which are not considered by the AI NAD Analyzer.

Furthermore, the ground truth was based on the written procedure report, which is an efficient way to include large case volumes but is also prone to inter-reader variabilities. While reporting styles vary across radiologists (e.g., concise vs detailed reports, free text *vs* structured reporting), the procedure report represents an important medical record and does distinguish NAD from AD cases in a population of cases. More granular analysis of findings would require image interpretation by multiple radiologists, which is less scalable but more robust in establishing the ground truth.

The interpretation of chest radiographs is prone to errors and findings missed by the reader could affect the ground truth which is based on the procedure report. Cases which are identified as NAD by both procedure report and AI should have a lower likelihood of missed findings.

Another shortcoming of our study is that the test cases in the study came from only one geographic area in the United States, albeit with multiple locations in the region. However, the cases used for AI stress testing came from different sources which supports the generalizability of the AI system.

When analyzing the ratio of NAD to AD cases, our study included more cases with NAD than AD until the age group of 66 – 75 years. The results also show that the AI identifies a higher rate of NAD cases in younger patients than older patients. The ratio of AI NAD (System B) to GT NAD declined by patient age from 37% in patients 18 – 25 years to 11% in patients over 86 years. Given that patient populations are aging in many geographic regions, patient age will need to be considered. We plan to further investigate this issue in the future iteration of our AI system.

While autonomous interpretation of unremarkable chest radiographs may become a possibility in the future, the NAD definition in this paper is made with the intention to work in a triage setting, i.e., support radiology workflow with all cases ultimately being reviewed by a radiologist. Regardless, cases classified as PAD by the system will go through the standard review workflow.

Finally, workflow integration will be key for adoption in clinical practice. Incorrect NAD results by AI may bias the radiologist towards expecting normal or unremarkable exams and radiologists need to be aware of this. There may also be workflow options that can help to reduce or suppress the bias. Options include the worklist, reading and reporting software. Workflow impact needs to be tested (combined reading AI and radiologist) to assess efficiency, accuracy, and reader confidence benefits. Ongoing quality control should be considered as part of a systems-based approach. Future extensions of such an AI system could include automated prepopulating of the procedure report for efficiency gains.

Finally, the objective of the study was to demonstrate AI's immense potential in identifying unremarkable cases so that resources can be focused over time to cases with significant findings. The AI NAD Analyzer can be combined with other AI systems providing detection of specific abnormal findings as both approaches are complementary. This enables comprehensive decision support where cases with NAD are identified and specific findings in AD cases are marked. Comprehensive support by AI systems has the potential to provide real gains for chest radiograph interpretation in outpatient imaging.

References

[1] Gefter WB, Post BA and Hatabu H, "Commonly missed findings on chest radiographs: Causes and consequences," *CHEST,* vol. 163(3), pp. 650-661, 2023.

[2] Gefter WB and Hatabu H, "Reducing errors resulting from commonly missed chest radiography findings," *CHEST,* vol. 163(3), pp. 634-649, 2023.

[3] Whang JS, Baker SR, Patel R, Luk L and Castro III A, "The causes of medical malpractice suits against radiologists in the United States.," *Radiology,* vol. 266(2), pp. 548-554, 2013.

[4] Cannavale A, Santoni M, Mancarella P, Passariello R and Arbarello P, "Malpractice in radiology: what should you worry about?," *Radiology research and practice,* 2013.

[5] Yoo H and et al, "Artificial Intelligence-Based Identification of Normal Chest Radiographs: A Simulation Study in a Multicenter Health Screening Cohort," *Korean J Radiol,* vol. 23(10), pp. 1009-1018, 2022.

[6] Keski-Filppula T, Nikki M, Haapea M and et al, "Using artificial intelligence to detect chest X-rays with no significant findings in a primary health care setting in Oulu, Finland," *arXiv,* 2022 https://doi.org/10.48550/arXiv.2205.08123.

[7] Dyer T and et al, "Diagnosis of normal chest radiographs using an autonomous deep-learning algorithm," *clinical Radiology ,* vol. 76, no. 6, pp. P473.E9-473, 2021.

[8] Plesner LL, Mueller FC, Nybing JD and et al, "Autonomous Chest Radiograph Reporting Using AI: Estimation of Clinical Impact," *Radiology ,* 2023 00:e222268.

[9] Annarumma M and et al, "Automated Triaging of Adult Chest Radiographs with Deep Artificial Neural Networks," *Radiology ,* vol. 291, p. 196–202, 2019.

[10] Ghesu FC and et al, "Contrastive self-supervised learning from 100 million medical images with optional supervision," *Journal of Medical Imaging,* vol. 9, no. 6, 064503, 2022.

[11] Collins J and Stern EJ, "Chest radiology: the essentials," *Lippincott Williams & Wilkins,* 2008.

[12] Reed JC, "Chest Radiology: Patterns and Differential Diagnoses," *Elsevier Health Sciences,* 2017.

[13] Armato SG 3rd, McLennan G, Bidaut L and et al, "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans.," *Med Phys.,* vol. 38(2), pp. 915-931, 2011.

[14] Rosenthal A, Gabrielian A and Engle E, "The TB Portals: an Open-Access, Web-Based Platform for Global Drug-Resistant-Tuberculosis Data Sharing and Analysis.," *J Clin Microbiol,* vol. 55(11), pp. 3267-3282, 2017.

[15] Lin TY, Goyal P, Girshick R and et al, "Focal loss for dense object detection," *Proceedings of the IEEE international conference on computer vision,* pp. 2980-2988, 2017.

[16] Homayounieh F, Digumarthy S, Ebrahimian S, Rueckel J and et al, "An artificial intelligence–based chest X-ray model on human nodule detection accuracy from a multicenter study," *JAMA Network Open,* Vols. 4(12) e2141096-e2141096, 2021.

[17] Rudolph J, Huemmer C, Ghesu FC, Mansoor A and et al, "Artificial intelligence in chest radiography reporting accuracy: Added clinical value in the emergency unit setting without 24/7 radiology coverage," *Investigative Radiology,* vol. 57(2), pp. 90-98, 2022.

Authors information

Awais Mansoor[1]

Ingo Schmuecking[1]

Florin C. Ghesu[1]

Bogdan Georgescu[1]

Sasa Grbic[1]

R S Vishwanath[2]

Dimeji Farri[1]

Rikhiya Gosh[1]

Ramya Vunikili[1]

Mathis Zimmermann[3]

James Sutcliffe[4]

Steven L Mendelsohn[4]

Warren B Gefter[5]

Dorin Comaniciu[1]

[1]Siemens Healthineers, Digital Technology and Innovation, Princeton, NJ, USA.

[2]Siemens Healthineers, Digital Technology and Innovation India, Bengaluru, India

[3]Siemens Healthineers, Digital & Automation, Malvern, PA, USA.

[4]Zwanger-Pesiri Radiology, Lindenhurst, NY, USA.

[5]Department of Radiology, Penn Medicine, University of Pennsylvania, Philadelphia, PA

Corresponding authors:

Correspondence to Awais Mansoor or Ingo Schmuecking.

Data availability statement

The data that support the findings of this study were used under license and are not publicly available. Data are however available from the authors upon reasonable request and with permission of the licensor, if applicable.

The code used for training the models has a large number of dependencies on internal tooling, infrastructure and hardware, and its release is therefore not feasible.

Ethics declarations

W.B.G. received consulting fees from Siemens Healthineers to support the research collaboration.

Zwanger-Pesiri Radiology received funding from Siemens Healthineers to support the research collaboration.

The remaining authors are employees of Siemens Healthineers.

This study was funded by Siemens Healthineers.

The authors have no other competing interests to disclose.

Acknowledgements

We thank Pranjal Sahu for critical review of the manuscript.

We thank Dr. Eileen Krieg for clinical expertise and annotations of stress test data.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- UsingAItoIdentifyChestRadiographswithNADappendixforsubmission.pdf