 doi:10.20944/preprints202208.0189.v1

Article

Performance of a Chest Radiography AI Algorithm for Detection of Missed or Mislabeled Findings: A Multicenter Study

Parisa Kaviani [1], Subba R. Digumarthy [1], Bernardo Bizzo [1,2] Bhargava Reddy [3], Manoj Tadepalli [3], Preetham Putha [3], Ammar Jagirdar [3], Shadi Ebrahimian [1], Mannudeep K. Kalra [1,*] and Keith J. Dreyer[1,2]

[1] Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston, MA
[2] MGH & BWH Center for Clinical Data Science, Boston, MA, USA
[3] Qure.ai, Mumbai, India
*Corresponding author: mkalra@mgh.harvard.edu; TEL.: 617-643-4583; Fax: 617-724-4152

Abstract: Purpose: We assessed if a CXR AI algorithm can detect missed or mislabeled chest radiographs (CXRs) findings in radiology reports. **Methods**: We queried multi-institutional radiology reports search database of 13-million reports to identify all CXRs reports with addendums from 1999-2021. Of the 3469 CXR reports with an addendum, a thoracic radiologist excluded reports where addendum was created for typographic errors, wrong report template, missing sections, or uninterpreted signoffs. The remaining reports with addendum (279 patients) with errors related to side-discrepancy or missed findings such as pulmonary nodules, consolidation, pleural effusions, pneumothorax, and rib fractures. All CXRs were processed with an AI algorithm. Descriptive statistics were performed to determine the sensitivity, specificity, and accuracy of AI to detect missed or mislabeled findings. **Results**: AI had high sensitivity (96%), specificity (100%), and accuracy (96%) for detecting all missed and mislabeled CXR findings. The corresponding finding-specific statistics for AI were nodules (96%, 100%, 96%), pneumothorax (84%, 100%, 85%), pleural effusion (100%, 17%, 67%), consolidation (98%, 100%, 98%), and rib fractures (87%, 100%, 94%). **Conclusion**: The CXR AI could accurately detect mislabeled and missed findings. **Clinical Relevance**: The CXR AI can reduce the frequency of errors in detection and side-labeling of radiographic findings.

Keywords: chest-X ray; sddendum; missed finding; radiology

1. Introduction

Chest radiography (CXR) is the most common imaging test representing up to 20% of all types of imaging procedures.(1) Some studies report that 236 CXRs are performed per 1000 patients per year, representing up to 25% of the annual diagnostic imaging procedures.(2) In 2010 alone, of the 183 million radiographic procedures in the United States on 15,900 radiologic units, CXRs represented almost half of all radiographic images (44%).(3) Easy accessibility, portability, familiarity, and affordability (relative to other imaging tests) are all factors leading to its widespread use in medical practice for various cardiothoracic ailments. (4),(5) Despite their common use, CXRs are difficult to read and subject to substantial inter- and intra-reader variations. A retrospective study documented that inter-radiologist and physician concordance were 78% for CXRs.(6) Other studies report disagreements between radiologists on several CXR findings.(7) CXRs also have a high misinterpretation rate, reported as high as 30% in one study.(8) The impact of missed radiographic findings is non-trivial. A 1999 study reported that 19% of lung cancers which presented as pulmonary nodules on CXRs were missed.(9) Such missed findings can be catastrophic for the patients as well as for reporting physicians. The Institute of Medicine (IOM) states that 44,000-98,000 patients die in the United States every year because of preventable errors.(10)

 doi:10.20944/preprints202208.0189.v1

With the increasing use and availability of approved (such as from the US Food and Drug Administration) artificial intelligence (AI) algorithms for several CXR findings (11), we hypothesized that AI could help reduce the frequency of mislabeled or misinterpreted CXRs. Apart from improving the interpretation efficiency, several studies have reported improved accuracy of interpretation of several CXR findings (12, 13, 14). There are an increasing number of AI algorithms for triaging and detecting several CXR findings including pneumonia, pneumothorax, pleural effusion, and pulmonary nodules.(12),(13) However, to our best knowledge, there are no prior publications on the impact of an AI algorithm on the addended mislabeled or misinterpreted CXR reports in routine clinical practice. Therefore, we investigated if a CXR AI algorithm can detect missed or mislabeled CXR findings in radiology reports.

2. Materials and Methods

2.1. Approval and disclosures

The institutional review board at Massachusetts General Brigham approved (IRB protocol number: 2020P003950) our retrospective study with waiver of informed consent. A study coinvestigator (SRD) received research grant from Qure.AI but did not participate in data collection, study evaluation or statistical analysis. Another study coinvestigator (MKK) received research grants for unrelated projects (Siemens Healthineers, Riverain Tech, Coreline Inc.) The remaining coauthors have no financial disclosures. All study authors had equal and unrestricted access to study data.

2.2. Chest Radiographs

We queried multi-institutional radiology search databases of 13-million reports to identify all CXRs reports with addenda from 1999-2021. We used the following keywords for search criteria: chest radiograph and addendum. The identified CXRs reports included both portable and posteroanterior upright CXRs. The two search engines used in our study were mPower (Nuance Inc) and Render (proprietary institutional search engine). Duplicate CXR reports were excluded when they had the same medical record and accession numbers. The search identified a total of 3469 unique CXR reports between January 2015 to March 2021, since the patients' medical information are not recorded electronically on our database before 2015. The inclusion criteria were: availability of DICOM CXR images from 7 hospitals in our healthcare enterprise (Massachusetts General Hospital [MGH], Brigham Women Hospital [BWH], Faulkner Health Center [FH], Martha's Vineyard Hospital [MVH], Salem Hospital [NSMC], Newton-Wellesley Hospital [NWH] and Spaulding Rehabilitation Hospital [SRH]) and an addendum with either missed or side-mislabeled finding. To protect institutional identity (since more number of addenda were linked to a greater volume of reported CXRs and not to the quality of reporting), we blinded the names of individual sites before analysis. Missed findings in CXR reports were defined as those reports where there was a missed radiographic finding in the original CXR report, which was subsequently corrected with addendum. Mislabeled findings included wrong side labels of CXR findings in the initial CXR reports, which were corrected with addenda.

Two study coinvestigators (MKK with 14 years of experience as a thoracic radiologist, PK with one year post-doctoral research in radiology) excluded CXR reports with addenda documenting typographical errors (n=782) , wrong report templates (n=341), missing section signoffs (n=289), and communication errors (n=1174). Duplicate reports of the same exam and patient were excluded as well (n=302). Other exclusion criteria were addenda about findings that cannot be assessed with the AI algorithm [such as cardiac calcification (n=27), diaphragmatic hernia (n=81), clavicle (n=29) or humeral (n=34) fractures or dislocations, lines (n=94) , and devices (n=37)]. The final sample size after the application of inclusion and exclusion criteria was 279 CXRs with addenda belonging to 279 patients (patient demographics are summarized in the results section). Figure 1 summarizes the distribution of missed and mislabeled CXR findings including in our study.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 August 2022 doi:10.20944/preprints202208.0189.v1

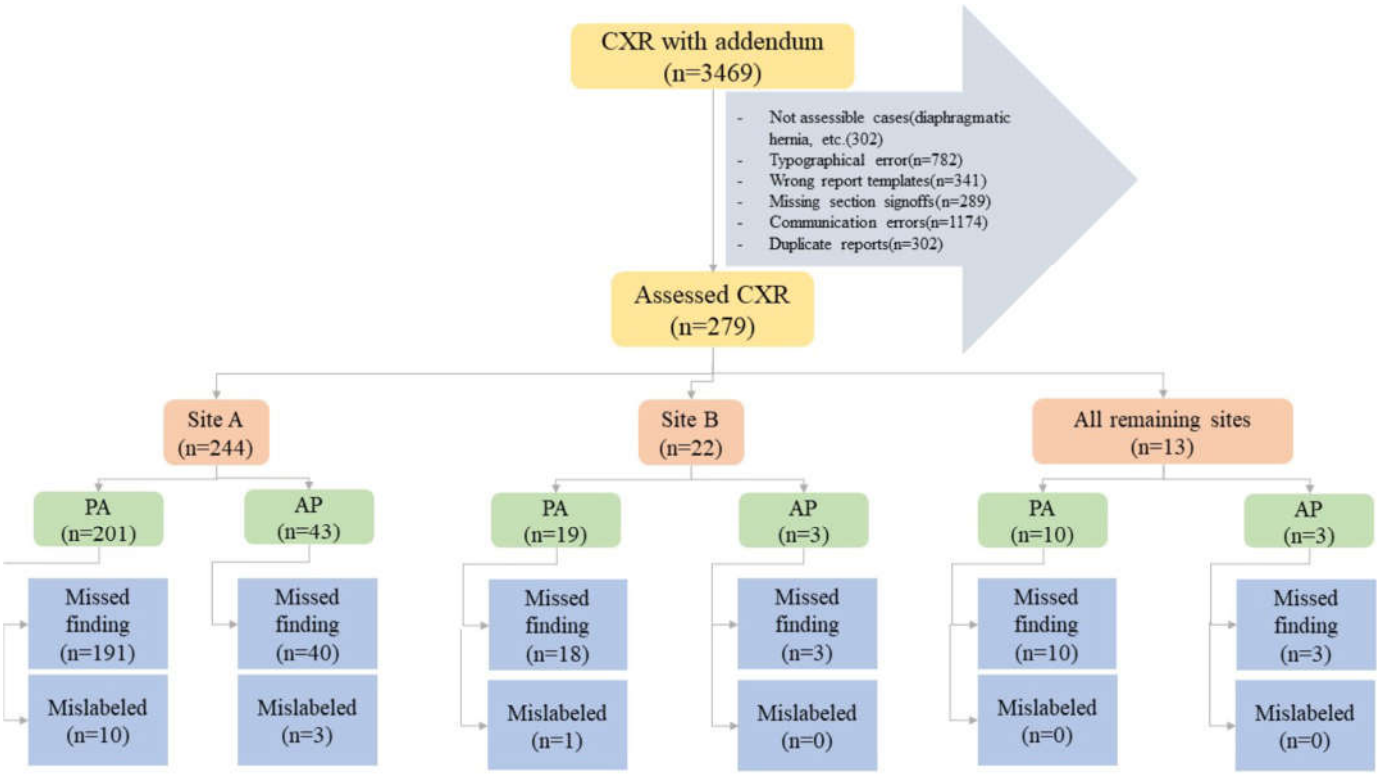

Figure 1. Flow diagram summarizing the study methods and distribution of specific missed and mislabeled findings at different sites and CXR types (PA- posteroanterior CXR; port- portable CXR).

The study coinvestigators also noted the organization name for addended CXR reports and the specific missed finding name (such as pneumothorax, nodule, atelectasis, rib fractures, and pleural effusion). A chest radiologist (MKK) reviewed all CXRs with missed findings/labeled nodules and assessed their size and clinical importance on a three-point scale (1: not significant since the nodule is definitely benign such as granuloma; 2: indeterminate clinical importance; 3: definitely of clinical importance). Information on request for further imaging or follow-up evaluation was recorded from the reports.

2.3. AI algorithm

All 279 frontal CXRs were deidentified, exported as DICOM images and processed with an offline AI algorithm (qXR, Qure.AI, Mumbai, India) installed on a personal computer within our institutional firewall to protect patient privacy. Although approved in several countries in Asia, Africa, and Europe, the AI algorithm used in our study is not an US FDA cleared product. The algorithm was trained on over 3.7 million CXRs and radiology reports from various healthcare sites from different parts of the world. The algorithm uses a series of convolutional neural networks (CNNs) trained to identify different abnormalities on frontal CXRs. The algorithm first resizes and normalizes CXRs to decrease variations in the acquisition process, and then applies modifications in either densenets or resnets network architectures to separate CXRs from radiographs of other anatomies. Subsequently, multiple networks, including densenets and resnets are applied for individual CXR findings. Further technical details of the algorithms have been described in prior publications (15). The algorithm was validated on a separate dataset of over 93,000 CXRs from multiple sites in India. Neither training nor validation datasets included CXRs from any test sites. The image algorithms (qXR v3) were trained and tuned on the training set of 3.7 million chest X-rays with the corresponding reports. Optimal thresholds were selected using a proprietary method developed at Qure.ai in addition to standard methods like Youden's Index. These thresholds, additionally validated on a test set of over

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 August 2022 doi:10.20944/preprints202208.0189.v1

93,000 CXRs which was not used during training. The process used has been described in the past work done at Qure.ai with qXR in: https://arxiv.org/abs/1807.07455. The threshold values used were part of the commercial version of qXR and were frozen before the start of the study.

In its current version, the algorithm can identify abnormalities such as pulmonary opacities (consolidation, fibrosis, nodules), emphysema, cavity, pleural effusions, pneumothorax; tracheal deviation, hilar enlargement, cardiac silhouette enlargement, elevated hemidiaphragm, pneumoperitoneum, scoliosis and rib fractures. Neither the training nor validation datasets included any CXRs from the 7 test sites in our study. Following post-processing of test datasets, the algorithm created a de-identified EXCEL file containing information on model outputs for individual CXR findings. Separately, the algorithm also provided heat map of CXRs with annotated findings and automatically generated radiology reports with information on the presence or absence of the AI-assessable findings.

2.4. Statistical analyses

Statistical analysis was performed with Microsoft EXCEL (Microsoft Inc., Redmond, Washington). To assess the performance of the AI algorithm, we predefined true positive (specific missed finding is identical in addendum and AI output, for example, both addendum and AI output document pneumothorax), true negative (addendum and AI output agreed on the absence of specific findings, for example, both addendum and AI agreed on the absence of pneumothorax), false positive (addendum or the original radiology report did not document a finding identified by the AI algorithm, for example, AI identified pneumothorax was not present in radiology report or the CXR), and false negative (addendum's description of missed finding did not correspond to AI detected finding, for example, addendum documented the presence of pneumothorax which AI did not detect) The sensitivity, specificity, accuracy, and receiver operating characteristics (ROC) with the area under the curve (AUC) were calculated using Microsoft Excel and SPSS.

3. Results

Of the 279 CXRs with addenda performed in 279 patients (mean age 59±20 years), 143 belonged to male patients and 136 to female patients. There were 230 PA CXRs and 49 portable CXRs in the dataset. As the algorithm labeled both pneumonia and atelectasis as consolidation; we reported the sum of these two findings as consolidation. Table 1 summarizes the distribution of missed and mislabeled CXR findings in our study according to CXR types (posteroanterior or portable). Regardless of the sites, most missed and mislabeled findings in the addenda were present in reports of posteroanterior CXRs compared to portable CXRs ($p<0.001$). Documentation of missed findings greatly outnumbered mislabeled findings. The most common missed findings included pneumothoraces (100/279; 35.8%), consolidation (62/279; 22.2%), pulmonary nodules (54/279; 19.4%), rib fractures (48/279; 17.2%), and pleural effusions (15/279; 5.4%).

Table 1. Summary of missed (MS) and mislabeled (ML) findings in posteroanterior (PA) and portable CXRs at different sites included in our study.

	Site A				**Site B**				**All remaining sites**			
CXRs	PA		Portable		PA		Portable		PA		Portable	
Findings	MS	ML	MS	ML	MS	ML	MS	ML	MS	ML	MS	ML
Consolidation	49	3	3	0	4	0	0	0	3	0	0	0
Pulmonary nodule	28	1	8	1	8	1	3	0	2	0	2	0
Pneumothorax	68	6	20	1	2	0	0	0	3	0	0	0
Pleural effusion	10	0	0	1	1	0	0	0	2	0	1	0
Rib fracture	36	0	9	0	3	0	0	0	0	0	0	0

 doi:10.20944/preprints202208.0189.v1

The sensitivity, specificity, specificity, accuracy and AUCs of the AI algorithm for different findings are summarized in Table 2. AI's highest performance was in the detection of pulmonary nodules and consolidation and lowest in pleural effusion (low specificity). Table 3 illustrates the frequency of true positive, true negative, false positive and false negative for each finding. Table 4. Demonstrates the distribution of each finding based on different vendors.

Table 2. Summary statistics of AI performance in detection of missed findings on CXRs. The numbers in parenthesis represent 95% confidence interval for the area under the curve (AUC).

Findings	Sensitivity	Specificity	Accuracy	AUC
Pulmonary nodule	96	100	96	0.98 (0.94 - 1.00)
Consolidation	98	100	98	0.99 (0.97 - 1.00)
Rib fracture	87	100	94	0.94 (0.85-1.00)
Pleural effusion	100	17	67	0.82 (0.54-1.00)
Pneumothorax	84	100	85	0.92 (0.86-0.98)

 doi:10.20944/preprints202208.0189.v1

Table 3. Summary frequencies of true positive, true negative, false positive and false negative of each finding.

Findings	True positive	True negative	False positive	False negative
Pulmonary nodule	51	1	0	2
Consolidation	62	62	0	1
Rib fracture	20	25	0	3
Pleural effusion	9	1	0	5
Pneumothorax	80	5	0	15

Figure 3 summarizes the AUCs of the algorithm for different radiographic findings. There was no significant difference in the performance of AI on CXRs from sites A and B, or between the portable and posteroanterior CXRs ($p>0.5$). Data from the remaining sites were not compared due to low sample size (<5 data points per finding) for missed and mislabeled findings. The AI algorithm had moderate to high AUC for all missed findings. For true positive on the AI algorithm, the AI annotated heat map showed the lesion as on the correct side instead of the mislabeled findings reported in addended radiology reports. Figure 4 displays different missed findings which were correctly detected (true positive) by the AI algorithm.

Among 279 CXR, 41 CXR addenda requested for additional assessment [rib fractures (n=4), pneumothoraces (n=3), and pulmonary nodules (n=34)]. All patients with missed pneumothoraces had chest tube placement. All 34 CXRs with missed nodules underwent chest CT, three patients underwent lung nodule biopsy (one benign nodule; 2 malignant nodules). The 31 remaining pulmonary nodules remained stable on follow-up imaging (CXR and/or CT). Figure 2 illustrates the distribution of missed nodule size and further evaluation.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 August 2022 doi:10.20944/preprints202208.0189.v1

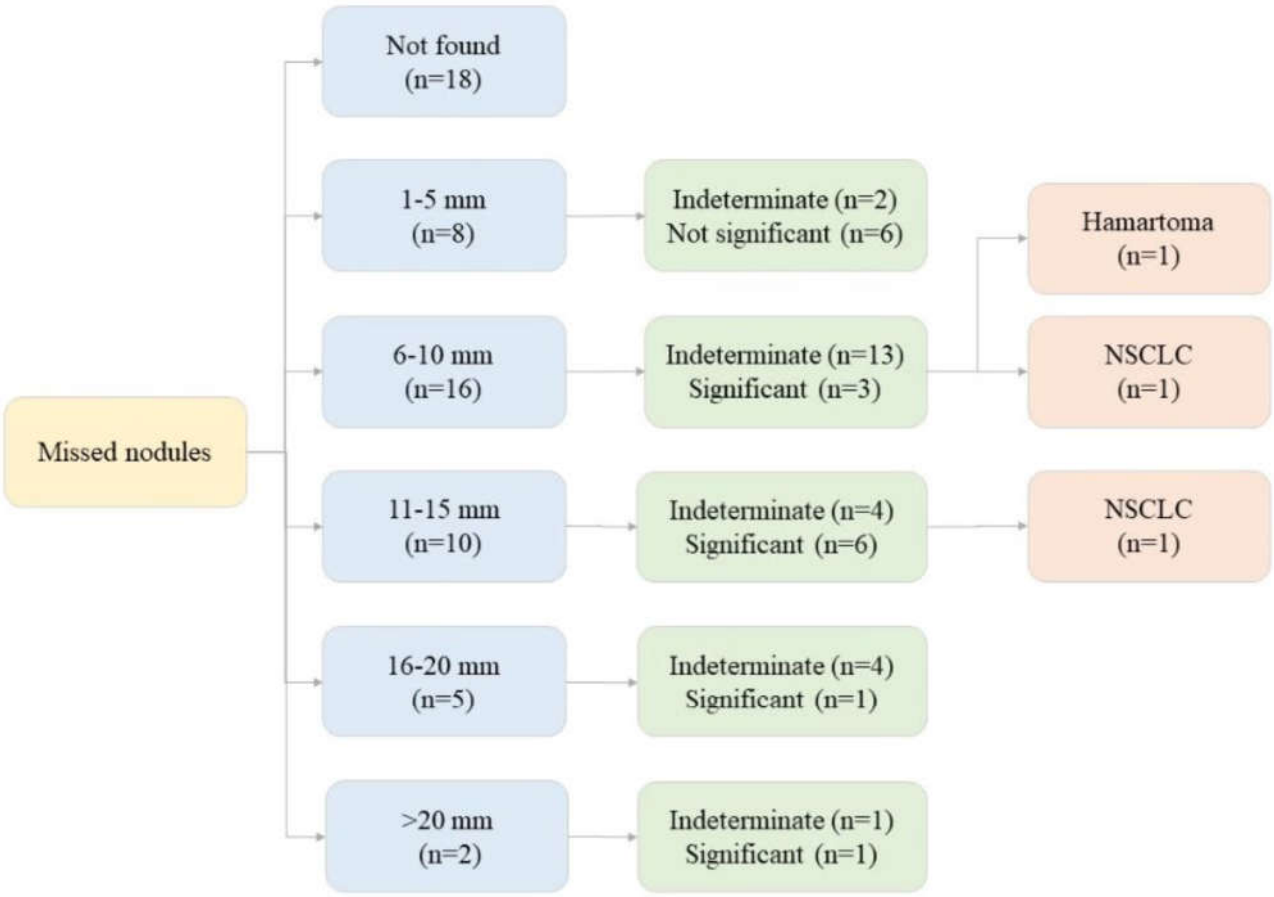

Figure 2. Flow diagram illustrating the missed nodules distribution based on size and significance.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 August 2022 doi:10.20944/preprints202208.0189.v1

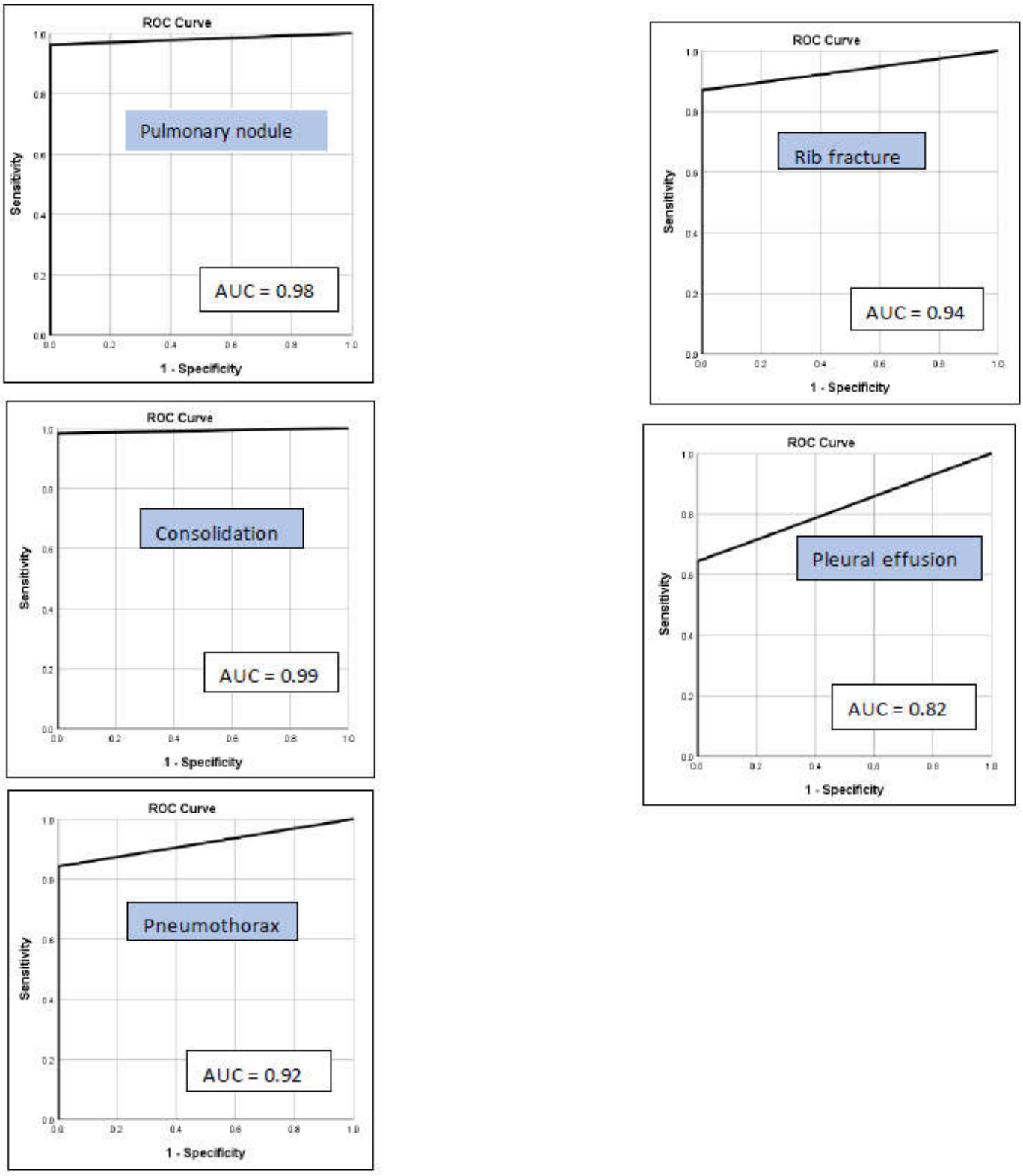

Figure 3. Receiver operating characteristic analyses with area under the curve (AUC) for different missed findings detected with the AI algorithm.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 August 2022 doi:10.20944/preprints202208.0189.v1

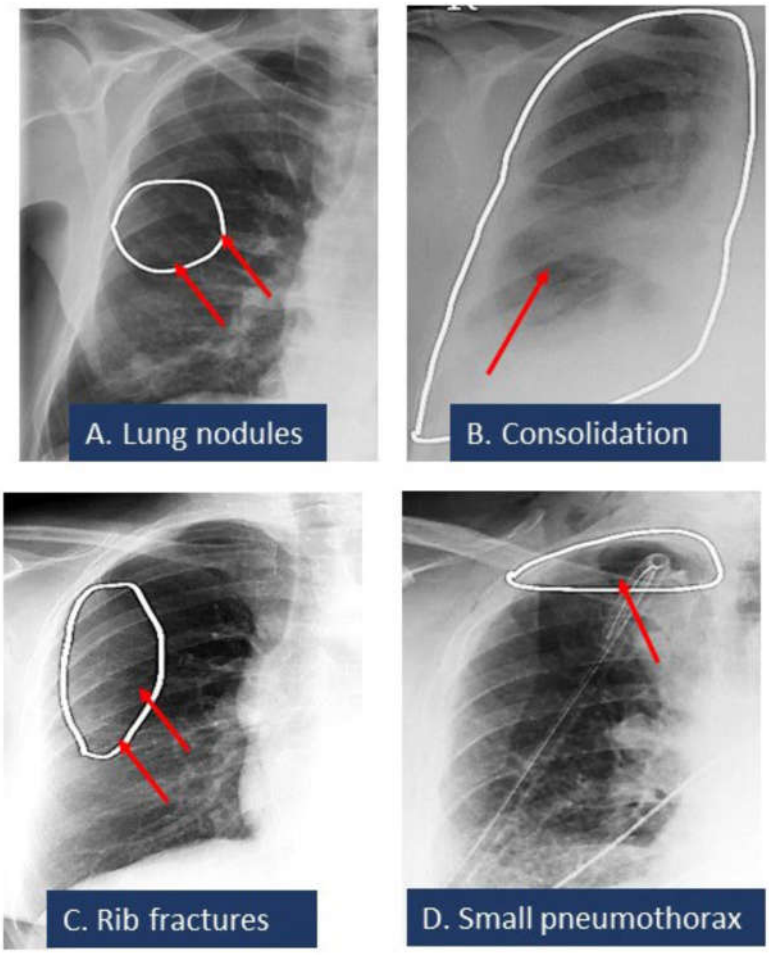

Figure 4. Spectrum of missed CXR findings (pulmonary nodule [A], pneumonia [B], pneumothorax [C], rib fractures [D]) for which radiologists issued addenda to their original radiology reports. These findings were detected with the AI algorithm.

4. Discussion

We report the frequency of different CXR findings which are either missed or mislabeled and later corrected with an addendum. The assessed AI algorithm can help identify such findings and errors with high performance (AUC 0.82-0.99). Although prior studies have described the comparable performance of AI either as standalone or a second reader (17), most studies evaluated consecutive or selected CXRs without specific attention to the clinical significance of detected or missed radiographic findings.

Prior research publications have investigated the frequency of misdiagnosis of CXR findings (18). In a recent study, Wu et al describe an AI algorithm that reaches and exceeds the performance of third-year radiology residents for detecting findings on frontal chest radiographs with a mean AUC of 0.772 for the assessed AI algorithm.(14) Such AI algorithms can improve accuracy while improving workflow efficiency of reporting. In our institution, addenda are issued for final signed off radiology reports by the attending radiologists or a fully licensed trainee (such as a clinical fellow). Therefore, high accuracy (up to 0.99) of the assessed AI algorithm pertains to findings missed by interpreting physicians beyond residency training.

High accuracy and AUCs of the AI algorithm for detecting consolidation and pulmonary nodules in our study also correspond to those reported in recent studies. Behzadi-Khormouji et al reported accuracy of 94.67% for detection of consolidation on CXRs with the use of their AI model.(17) Likewise, the performance of our AI algorithm for detecting

 doi:10.20944/preprints202208.0189.v1

all-cause pulmonary nodule is comparable to the overall performance of another AI algorithm (Lunit Inc., Seoul, South Korea) . Yoo et al reported a sensitivity of 86.2% and specificity of 85% for all-cause nodules, and a higher sensitivity of AI (up to 100%) compared to 94.1% for radiologists in detecting malignant nodules on digital CXRs (16). Similarly, high sensitivity, specificity, accuracy, and AUC of the AI algorithm for detection of pneumothorax compares well with other multicenter studies with other AI algorithms such as from Thian et al who reported an AUC of 0.91-0.97 for detection of pneumothorax detection with their AI algorithm (19).

The main implication of our study pertains to the performance and potential use of AI when reporting CXRs. Given the high volume of CXR use in hospital settings, relatively low reimbursement for CXR interpretation, and pressure for rapid and efficient reporting compounded by the highly subjective nature of projectional radiography, reporting errors on CXRs are common. In such circumstances, as noted from our study and supported by other investigations (12, 13, 14) AI algorithms can reduce the frequency of commonly missed and mislabeled CXR findings. Furthermore, routine use of CXR AI in interpretation has the potential to avoid common reporting errors and therefore, reduce the need for issuing addenda to previously reported exams. At the same time, AI algorithms can potentially shift the focus from under- or non-reporting of radiographic findings to over-reporting of findings due to high false positive outputs. Such challenges can be addressed with robustly trained AI models and selection of appropriate cut-off values that maintain a good balance of sensitivity and specificity across different radiography units and radiographic quality. Although we did not compare the performance with other models in the literature, the AUCs from of our AI algorithm was similar to those reported for other models assessed with open access CXR datasets (https://nihcc.app.box.com/v/ChestXray-NIHCC/file/220660789610) (20). Apart from detection of radiographic findings assessed in our study with an AI algorithm, other studies have also assessed applications of AI for prioritizing interpretation of CXRs to expedite reporting of abnormal CXRs and specific findings (21). Baltruschat et. al. reported use of AI-based worklist prioritization for substantial reduction in reporting turnaround time for critical CXR findings (22). Similar improvements in reporting time with worklist prioritization with AI have been reported for other body regions as well, such as head CTs for intracranial hemorrhage (23).

There are some limitations to our study. Although we queried over 13 million radiology reports from 1999 to March 2021 to identify 3469 CXR reports with addenda, the stringent inclusion and exclusion criteria made our sample size small (n= 279 CXRs) – the primary limitation of our study. While a larger dataset for AI algorithm is ideal, our study still provides a representative snapshot of missed and mislabeled findings on CXRs on consecutive eligible CXRs from multiple sites. While addended radiology reports describing missed and mislabeled findings pertain to identified or recognized errors in reporting, they underestimate the true incidence of missed or mislabeled findings since most findings might not be discovered or corrected in subsequent follow-up CXRs or other imaging tests (such as CT). The purpose of our study was to not uncover the true incidence of missed CXR findings, but to investigate the performance of the AI algorithm in missed or mislabeled findings deemed important by the referring physicians and/or radiologists, and therefore addressed with addenda. Due to the limited sample size, we were unable to determine the performance of AI for missed rare findings such as mediastinal and hilar abnormalities, cavities, and pulmonary fibrosis. In addition, the performance of AI can vary based on the type of findings and therefore our results may not be generalizable to sites with different distribution of CXR findings.

Another limitation of our study pertains to findings that are currently not detected with the AI algorithm, such as placement of lines and devices, which was a significant contributor of excluded portable CXR reports with addenda. Although CXRs included in our study belonged to real-world CXRs with addended reports to add missed findings, we did not include real-world randomized CXRs datasets without addended reports since prior studies have reported on performance (sensitivity, specificity, AUC, and accuracy for individual CXR findings) of the AI algorithm used in our study on such datasets (24).

We also did not specifically perform a systematic analysis of the algorithm for understanding domain bias at the study sites. However, the inclusion of different study site types (including quaternary, community, cottage, and rehabilitation hospitals), different radiographic equipment, and a large group of interpreting radiologists would have minimized such bias in our study. Finally, we did not process the CXRs with other AI algorithms, and therefore, we cannot comment on the relative performance of different AI algorithms.

In conclusion, our study demonstrates that AI can help identify several missed and mislabeled findings on CXRs. As a secondary reader, the assessed AI algorithm can help radiologists identify and avoid common mistakes in detection, description, and labeling of specific radiographic findings, including consolidation, pulmonary nodules, pneumothorax, rib fractures, and to a lesser extent, pleural effusions. Future studies with a larger number of reporting errors can help assess the effectiveness and relative performance of AI algorithms in improving the accuracy of radiology reporting for chest radiographs.

References

1. Ekpo EU, Egbe NO, Akpan BE. Radiographers' performance in chest X-ray interpretation: the Nigerian experience. Br J Radiol. 2015;88(1051):20150023.
2. Speets AM, van der Graaf Y, Hoes AW, Kalmijn S, Sachs AP, Rutten MJ, et al. Chest radiography in general practice: indications, diagnostic yield and consequences for patient management. Br J Gen Pract. 2006;56(529):574-8.
3. Forrest JV, Friedman PJ. Radiologic errors in patients with lung cancer. West J Med. 1981;134(6):485-90.
4. Kelly B. The chest radiograph. Ulster Med J. 2012;81(3):143-8.
5. Schaefer-Prokop C, Neitzel U, Venema HW, Uffmann M, Prokop M. Digital chest radiography: an update on modern technology, dose containment and control of image quality. Eur Radiol. 2008;18(9):1818-30.
6. Satia I, Bashagha S, Bibi A, Ahmed R, Mellor S, Zaman F. Assessing the accuracy and certainty in interpreting chest X-rays in the medical division. Clin Med (Lond). 2013;13(4):349-52.
7. Fancourt N, Deloria Knoll M, Barger-Kamate B, de Campo J, de Campo M, Diallo M, et al. Standardized Interpretation of Chest Radiographs in Cases of Pediatric Pneumonia From the PERCH Study. Clin Infect Dis. 2017;64(suppl_3):S253-s61.
8. Berlin L. Reporting the "missed" radiologic diagnosis: medicolegal and ethical considerations. Radiology. 1994;192(1):183-7.
9. Quekel LG, Kessels AG, Goei R, van Engelshoven JM. Miss rate of lung cancer on the chest radiograph in clinical practice. Chest. 1999;115(3):720-4.
10. Institute of Medicine (US) Committee on Quality of Health Care in America; Kohn LT CJ, Donaldson MS, editors. To Err Is Human: Building a Safer Health System.; 2000.
11. Ebrahimian S., Kalra MK., Agarwal S., Bizzo B., Elkholy M., Wald C., et al. FDA-regulated AI algorithms: Trends, strengths, and gaps of validation studies. Academic Radiology. 2021 (In Press)
12. Li B, Kang G, Cheng K, Zhang N. Attention-Guided Convolutional Neural Network for Detecting Pneumonia on Chest X-Rays. Annu Int Conf IEEE Eng Med Biol Soc. 2019;2019:4851-4.
13. Li X, Shen L, Xie X, Huang S, Xie Z, Hong X, et al. Multi-resolution convolutional networks for chest X-ray radiograph based lung nodule detection. Artif Intell Med. 2020;103:101744.
14. Wu JT, Wong KCL, Gur Y, Ansari N, Karargyris A, Sharma A, et al. Comparison of Chest Radiograph Interpretations by Artificial Intelligence Algorithm vs Radiology Residents. JAMA Netw Open. 2020;3(10):e2022779.
15. Engle E, Gabrielian A, Long A, Hurt DE, Rosenthal A. Performance of Qure. ai automatic classifiers against a large annotated database of patients with diverse forms of tuberculosis. PLoS One. 2020 Jan 24;15(1):e0224445.
16. Yoo H, Kim KH, Singh R, Digumarthy SR, Kalra MK. Validation of a deep learning algorithm for the detection of malignant pulmonary nodules in chest radiographs. JAMA network open. 2020 Sep 1;3(9):e2017135-.
17. Behzadi-Khormouji H, Rostami H, Salehi S, Derakhshande-Rishehri T, Masoumi M, Salemi S, et al. Deep learning, reusable and problem-based architectures for detection of consolidation on chest X-ray images. Comput Methods Programs Biomed. 2020;185:105162.
18. Itri JN, Tappouni RR, McEachern RO, Pesch AJ, Patel SH. Fundamentals of diagnostic error in imaging. Radiographics. 2018 Oct;38(6):1845-65.
19. Thian YL, Ng D, Hallinan JT, Jagmohan P, Sia DS, Tan CH, Ting YH, Kei PL, Pulickal GG, Tiong VT, Quek ST. Deep Learning Systems for Pneumothorax Detection on Chest Radiographs: A Multicenter External Validation Study. Radiology: Artificial Intelligence. 2021 Apr 14:e200190.
20. Arora R, Bansal V, Buckchash H, et al. AI-based diagnosis of COVID-19 patients using X-ray scans with stochastic ensemble of CNNs. *Phys Eng Sci Med*. 2021;44(4):1257-1271. doi:10.1007/s13246-021-01060-9
21. Nabulsi Z, Sellergren A, Jamshy S, Lau C, Santos E, Kiraly AP, Ye W, Yang J, Pilgrim R, Kazemzadeh S, Yu J. Deep learning for distinguishing normal versus abnormal chest radiographs and generalization to two unseen diseases tuberculosis and COVID-19. Scientific reports. 2021 Sep 1;11(1):1-5.

 doi:10.20944/preprints202208.0189.v1

22. Baltruschat I, Steinmeister L, Nickisch H, Saalbach A, Grass M, Adam G, Knopp T, Ittrich H. Smart chest X-ray worklist prioritization using artificial intelligence: a clinical workflow simulation. European radiology. 2021 Jun;31(6):3837-45.
23. O'neill TJ, Xi Y, Stehel E, Browning T, Ng YS, Baker C, Peshock RM. Active reprioritization of the reading worklist using artificial intelligence has a beneficial effect on the turnaround time for interpretation of head CT with intracranial hemorrhage. Radiology: Artificial Intelligence. 2020 Nov 18;3(2):e200024.
24. Engle E, Gabrielian A, Long A, Hurt DE, Rosenthal A. Performance of Qure. ai automatic classifiers against a large annotated database of patients with diverse forms of tuberculosis. PLoS One. 2020 Jan 24;15(1):e0224445.