diagnostics

MDPI

Article

Role of an Automated Deep Learning Algorithm for Reliable Screening of Abnormality in Chest Radiographs: A Prospective Multicenter Quality Improvement Study

Arunkumar Govindarajan [1,*], Aarthi Govindarajan [1], Swetha Tanamala [2], Subhankar Chattoraj [2,*], Bhargava Reddy [2], Rohitashva Agrawal [2], Divya Iyer [2], Anumeha Srivastava [2], Pradeep Kumar [2] and Preetham Putha [2]

[1] Aarthi Scans & Labs, Chennai 600026, India
[2] Qure.ai, Mumbai 400063, India
* Correspondence: arunkumar@aarthiscans.com (A.G.); subhankar.chattoraj@qure.ai (S.C.)

Abstract: In medical practice, chest X-rays are the most ubiquitous diagnostic imaging tests. However, the current workload in extensive health care facilities and lack of well-trained radiologists is a significant challenge in the patient care pathway. Therefore, an accurate, reliable, and fast computer-aided diagnosis (CAD) system capable of detecting abnormalities in chest X-rays is crucial in improving the radiological workflow. In this prospective multicenter quality-improvement study, we have evaluated whether artificial intelligence (AI) can be used as a chest X-ray screening tool in real clinical settings. **Methods:** A team of radiologists used the AI-based chest X-ray screening tool (qXR) as a part of their daily reporting routine to report consecutive chest X-rays for this prospective multicentre study. This study took place in a large radiology network in India between June 2021 and March 2022. **Results:** A total of 65,604 chest X-rays were processed during the study period. The overall performance of AI achieved in detecting normal and abnormal chest X-rays was good. The high negatively predicted value (NPV) of 98.9% was achieved. The AI performance in terms of area under the curve (AUC), NPV for the corresponding subabnormalities obtained were blunted CP angle (0.97, 99.5%), hilar dysmorphism (0.86, 99.9%), cardiomegaly (0.96, 99.7%), reticulonodular pattern (0.91, 99.9%), rib fracture (0.98, 99.9%), scoliosis (0.98, 99.9%), atelectasis (0.96, 99.9%), calcification (0.96, 99.7%), consolidation (0.95, 99.6%), emphysema (0.96, 99.9%), fibrosis (0.95, 99.7%), nodule (0.91, 99.8%), opacity (0.92, 99.2%), pleural effusion (0.97, 99.7%), and pneumothorax (0.99, 99.9%). Additionally, the turnaround time (TAT) decreased by about 40.63% from pre-qXR period to post-qXR period. **Conclusions:** The AI-based chest X-ray solution (qXR) screened chest X-rays and assisted in ruling out normal patients with high confidence, thus allowing the radiologists to focus more on assessing pathology on abnormal chest X-rays and treatment pathways.

Keywords: chest X-rays; deep learning; qXR; computer aided diagnostic; neural network; multicenter prospective study

check for updates

Citation: Govindarajan, A.; Govindarajan, A.; Tanamala, S.; Chattoraj, S.; Reddy, B.; Agrawal, R.; Iyer, D.; Srivastava, A.; Kumar, P.; Putha, P. Role of an Automated Deep Learning Algorithm for Reliable Screening of Abnormality in Chest Radiographs: A Prospective Multicenter Quality Improvement Study. *Diagnostics* **2022**, *12*, 2724. https://doi.org/10.3390/diagnostics12112724

Academic Editor: Hajime Isomoto

Received: 3 October 2022
Accepted: 3 November 2022
Published: 7 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Chest radiography (chest X-ray or CXR) is the most common, economic, and low-radiation imaging modality, with a yearly estimate of over 2 billion imagings performed worldwide [1,2]. CXR serves as a primary and crucial screening modality in the diagnosis and management of cardiothoracic and pulmonary abnormalities, such as pneumothorax, pleural effusion, atelectasis, cardiac hypertrophy, hyperinflation, and consolidation, etc. [3]. CXR plays an important role in the diagnostic domain of medical practice. However, correctly identifying the abnormalities on CXR is often a significant challenge due to the limited number of well-trained radiologists and the heavy workload in extensive healthcare facilities [4]. Additionally, interpretations of CXR are subjective in nature. The complexity

of accurate CXR interpretation is largely increased by the presence of overlapping tissue structures. For example, when the lesion overlaps the ribs or large pulmonary blood arteries, or when there is minimal contrast between the lesion and the surrounding tissue, clear identification of the condition might be very difficult [5]. Sometimes, well-trained radiologists find it challenging to differentiate between the lesions or correctly identify very obscure pulmonary nodules [6]. Therefore, the chances of misdiagnosing lung abnormalities in CXR are high. To overcome these challenges, an automated system that can identify abnormalities on CXR with high accuracy and reliability is required. The automated CXR screening tool will aid in decreasing the workload and reducing the potentially missed findings in CXR. Recently, WHO also recommended using AI-based triaging as a potential solution in diagnostic pathways for tuberculosis [7]. The recent advancement of deep learning in various fields, such as computer vision, speech recognition, natural language processing, and bioinformatics, has been promising [8]. However, in contrast to deep learning, which is based on an automated feature learning process, machine learning relies on handcrafted, feature-based engineering that becomes vulnerable when applied to distinct patient groups and varying image qualities. Deep convolutional neural networks (CNNs) have proven to be more powerful in learning hierarchically rich deep features from labelled images. CNN performs a series of convolutional and non-linear operations by optimizing certain loss functions and mapping to the ground-truth labels to bridge the gap between high- and low-level visual feature representations [9].

In recent years, automated CXR screening using CAD systems based on CNNs for the detection of thoracic abnormalities has been an active area of research [10,11]. The CNN-based CAD system is generally trained on a large-scale database of labelled images. This helps in improving the learning of underlying complex semantic features for superior generalization in disease detection. Prior work based on large-scale open-source CXR datasets such as "NIH ChestX-ray 14", "ChestXpert", and "MIMIC-CXR" has been conducted by using the CAD system to detect different lung abnormalities i.e., pleural effusion, cardiomegaly, tuberculosis, lung cancer, pulmonary edema, pneumothorax, and pneumonia [12–17]. However, in most earlier studies, triaging AI for detecting chest abnormalities from open-source CXR was designed as a retrospective study. In this, ground truth was established based on the already available labelled images or based on the opinion of the minimum number of radiologists. Generally, in such a retrospective setting, the inclusion of CXR is based on a particular disease of interest and a specified number of patients with and without the condition. As a result, the true disease status of the patient at the time of study is generally not known. However, in a prospective design, a patient sample is better characterized in terms of clinical characteristics, standardized techniques for carrying out and interpreting the test(s) and establishing the gold standard process [18]. Moreover, in terms of hierarchy in the importance of evidence, prospective study designs are ranked higher than retrospective study designs as the level of evidence is of higher relevance [19].

The commercially available deep learning-based AI algorithm qXR (Qure.ai Technologies, Mumbai, India) [20] has been used in multiple studies previously in CXR screening for diagnosis of tuberculosis [21], missed or mislabelled findings [22], severity assessment of pneumonia with the need for mechanical ventilation [23], and identification of malignant nodules [24]. We conducted a prospective multicentre quality-improvement study. The study aimed to evaluate the quality of CXR analysis software (qXR) for predicting normal and abnormal CXR in routine screening. The study compared the performance of qXR with the radiologist in triaging and interpreting CXR in a high-CXR volume facility. The qXR also provided an AI-generated report detailing the abnormalities and localization of the abnormality present on the CXR, which assisted the frontline radiologists in patient diagnosis and reduced interreader variability across the readers. To the best of our knowledge, there is no prior publication on triaging AI in routine clinical practice for normal and abnormal reporting from CXR in a multi-centers prospective design.

2. Materials and Methods

The CXR and their corresponding radiological reports were obtained prospectively from the 35 centres of large radiology network in India from June 2021 to March 2022. A total of 10 radiologists with varying levels of experience read the CXR. The inclusion criteria of the study included CXR of patients with posterior–anterior (PA) or anterior–posterior (AP) views, image resolution of 1440 × 1440 pixels, and de-identified DICOMs(.dcm) format and more than six years of age. Patients were included in the final analysis if the CXR meet the inclusion criteria. The exclusion criteria included lateral view of the chest, incomplete view of lung, incomplete metadata, non-chest radiographs, and CXR with postoperative defects and excessive motion artifacts. Technical integration was tested prior to the start of the study to establish a connection between the picture archiving and communication system (PACS) and CAD software. This ensured that CXRs are auto-pushed from PACS. qXR was set up and installed on the radiologists' workstations after a successful testing phase. qXR is compatible and can integrate with any PACS system to maintain the data integrity without the substantial change in existing workflow. In this study, qXR classified CXR into three category settings—normal, abnormal, and to be reviewed (TBR)—as the aim was to filter the normal scans with high confidence. In Figure 1, the process followed at different study sites is given.

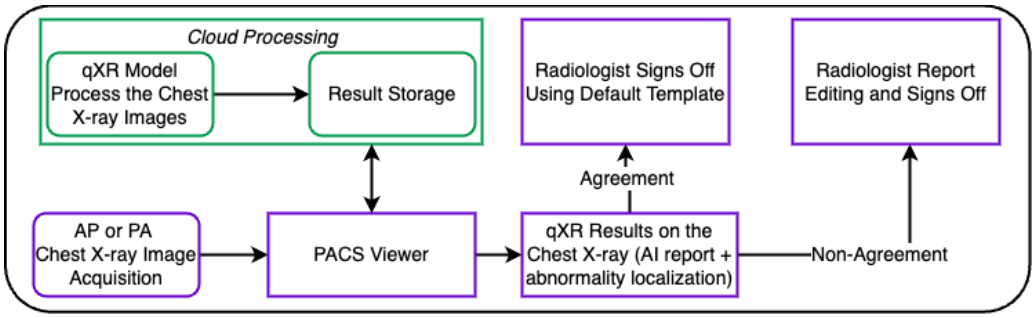

Figure 1. Flow diagram illustrating qXR assisted reporting process followed during this study across different study sites.

The qXR processed all 65,604 CXR and categorised them as normal, abnormal, or TBR by the radiologist. TBR are those cases wherein qXR flagged CXR for escalated radiologist's review. A total of 16,664 CXR were categorized as TBR. A total of 48,940 CXR were either classified as normal (qXR predicted no significant abnormality) and abnormal when qXR predicted significant abnormality. These cases were used for final analysis to determine the performance of the qXR. In Figure 2 a detailed semantic diagram on the final study population is given. The natural language processing (NLP) was implemented to parse the radiological report. The normal CXR comprised of no detectable abnormalities in the airways, lung fields, cardiac region, mediastinum, pleura, and visible thoracic skeleton and diaphragm. CXR which were not normal were categorized as abnormal. The major abnormalities which were clubbed in the abnormal category included blunted costo phrenic angle, hilar dysmorphism, cardiomegaly, reticulonodular pattern, rib fracture, scoliosis, atelectasis, calcification, consolidation, emphysema, fibrosis, nodule, opacity, pleural effusion, and pneumothorax. All the radiologists reading the CXR across multiple radiological centres had access to automated qXR diagnostic report on reader's computers. The radiologist's reports were finally used as a reference criterion to compare qXR performance in classifying normal and abnormal CXR by using standard evaluation metrics.

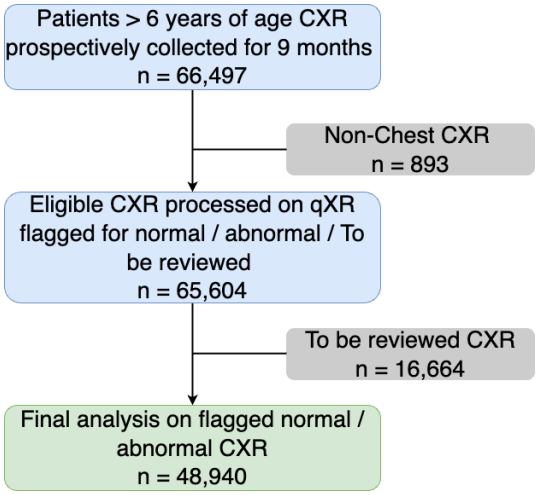

Figure 2. A consort diagram determining the exclusion criteria and establishing the final study population.

Statistical Analysis

Statistical analyses were performed by using Python version 3.9.7, Programming Language. The performance of qXR in categorizing normal and abnormal CXR was evaluated by using standard evaluation metrics, i.e., sensitivity, specificity, negative predictive value (NPV), and area under the ROC curve (AUC). These 95% confidence intervals for all performance evaluating metrics have been estimated by using the Wilson score method [25,26] without continuity correction.

3. Results

The final study population resulted in 65,604 CXR from 35 study centers in six different states of India after excluding 893 CXR as per exclusion criteria. Among the 16,664 TBR cases, 15,019 (90.13%) were marked as normal and 1645 (9.87%) were abnormal by the radiologists. To reduce potential bias in assessing the performance of the qXR, the final analysis is based upon 48,940 CXR after excluding all the TBR cases. The median age of the study population was 42 years. The mean, standard deviation, and interquartile range of patient age were 43.36 and 16.32, and 24. There were 61.7% male and 38.3% female patients included in the study. The automated qXR successfully sent all the processed CXR results to the PACS for easy visualization of outputs for the radiologists. The qXR achieved an overall mean sensitivity and specificity of 87.9% and 82.9%, respectively, in predicting normal and abnormal CXR. The high mean NPV of 98.9% suggests the efficacy of qXR in ruling out normal patients, thus allowing the radiologists to focus more on assessing pathology on abnormal CXR and treatment pathways. In Table 1, the detailed performance of qXR in categorizing normal and abnormal CXR is given.

Table 1. Performance of qXR in categorizing normal or abnormal CXR using radiologist as reference standard. negative predictive value (NPV); confidence interval (CI).

Evaluation Metrics	Point Estimate (95% CI)
Sensitivity	0.879 (0.867–0.889)
Specificity	0.829 (0.825–0.832)
NPV	0.989 (0.981–0.990)
AUC	0.871 (0.866–0.877)

The robustness of the qXR is further analyzed by stratified performance comparison in terms of NPV, sensitivity, specificity, and AUC across three different manufacturers, age in three different subgroups (16 years and less, 16 to 45 years, and more than 45 years of age), and gender is given in Table 2. It is observed from Table 2 that the NPV for different

manufacturers obtained by qXR in categorizing normal or abnormal is 98.8%, 98.7%, and 99.1%. From three different age groups, the values are 99.5%, 99.3% and 98.1%. From gender, the values are male, 99% and female, 98.7%. The stratified analysis conveys the high-performance stability of qXR across subgroups and usability in ruling out normal in routine CXR screening. In Figure 3 the ROC curves for qXR overall performance, stratified by manufacturer and patient's demographic is also given.

Table 2. Performance of qXR in categorizing normal or abnormal CXR stratified by patient's demographic and manufacturer using radiologist as reference standard. Negative predictive value (NPV); Area under the curve (AUC).

Attributes		NPV	Sensitivity	Specificity	AUC
Manufacturer	Carestream Health	**0.988** (0.987–0.990)	0.878(0.862–0.892)	0.831 (0.826–0.836)	0.872 (0.862–0.883)
	AGFA	0.987 (0.985–0.989)	0.865 (0.840–0.887)	**0.836** (0.829–0.843)	**0.878** (0.862–0.893)
	Fujifilm	0.991 (0.989–0.993)	**0.894** (0.870–0.914)	0.817 (0.810–0.824)	0.868 (0.851–0.884)
Age (years)	16 and less	**0.995**(0.988–0.997)	0.901 (0.790–0.957)	0.819 (0.797–0.840)	**0.886** (0.826–0.946)
	16–45	0.993 (0.992–0.994)	0.792 (0.763–0.819)	**0.922** (0.919–0.925)	0.878 (0.862–0.893)
	45 and above	0.981 (0.978–0.983)	**0.905** (0.893–0.916)	0.694 (0.688–0.701)	0.809 (0.798–0.819)
Gender	Male	**0.990** (0.989–0.991)	**0.884** (0.869–0.897)	**0.833** (0.829–0.838)	**0.875** (0.865–0.885)
	Female	0.987 (0.985–0.989)	0.871 (0.853–0.888)	0.821 (0.815–0.826)	0.866 (0.854–0.878)

BOLD denotes best performance in the stratified category for the different metrics.

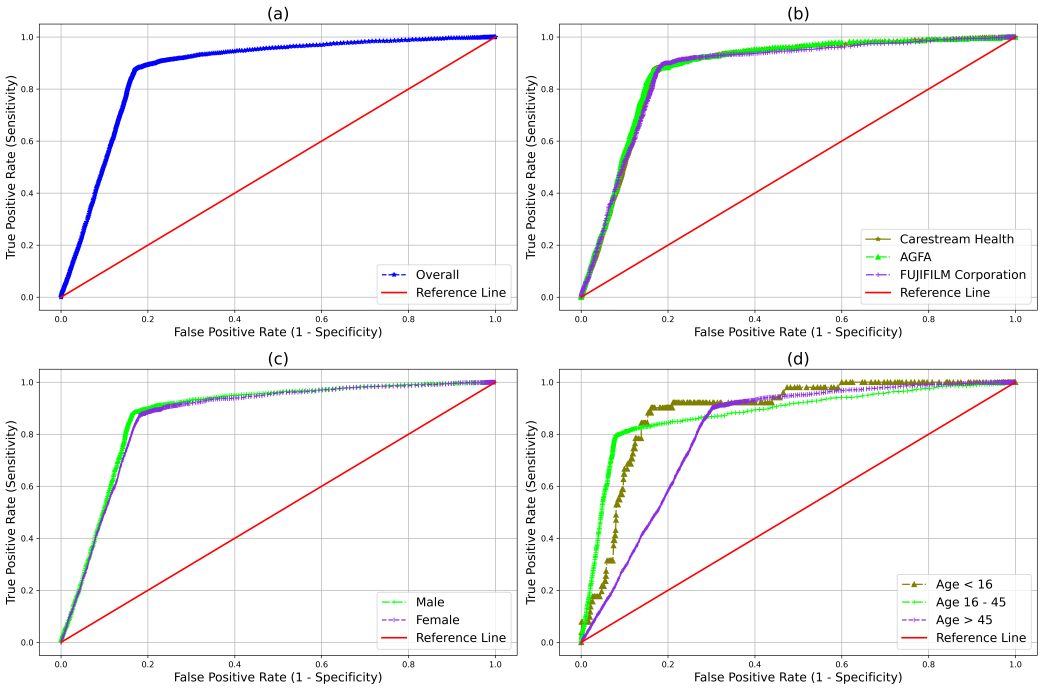

Figure 3. ROC curve. (**a**) Overall qXR performance. (**b**) Stratified across three different manufacturer subgroups. (**c**) Stratified across gender. (**d**) Stratified across 3 different age subgroups.

The CXR categorized as abnormal are further classified into two main categories and 15 major subabnormalities. The remarkable performance of qXR in terms of NPV for different abnormalities i.e., blunted CP angle (99.5%), hilar dysmorphism (99.9%), cardiomegaly (99.7%), reticulonodular pattern (99.9%), rib fracture (99.9%), scoliosis (99.9%), atelectasis (99.9%), calcification (99.7%), consolidation (99.6%), emphysema (99.9%), fibrosis (99.7%), nodule (99.8%), opacity (99.2%), pleural effusion (99.7%), and pneumothorax (99.9%) can be observed from Table 3. Furthermore, the capability of qXR in distinguishing the presence

or absence of specific abnormalities like cardiomegaly, nodule, and pneumothorax in terms of AUC is 0.965, 0.915, and 0.999 can be observed from Table 3. The ROC curve qXR performance in identifying subabnormality from abnormal CXR is given in Figures 4 and 5.

Table 3. Performance of qXR in categorizing abnormal CXR into subabnormalities by using radiologist as reference standard. Negative predictive value (NPV). Area under the curve (AUC).

	Abonrmality	NPV	Sensitivity	Specificity	AUC
Extra Pulmonary	Blunted CP angle	0.995 (0.995–0.996)	0.484 (0.435–0.534)	0.990 (0.989–0.991)	0.973 (0.709–0.766)
	Hilar Dysmorphism	0.999 (0.999–0.999)	0.216 (0.113–0.371)	0.992 (0.991–0.993)	0.864 (0.789–0.939)
	Cardiomegaly	0.997 (0.997–0.997)	0.804 (0.770–0.835)	0.962 (0.960–0.964)	0.965 (0.955–0.975)
	Reticulonodular Pattern	0.999 (0.998–0.999)	0.511 (0.407–0.614)	0.983 (0.981–0.984)	0.913 (0.872–0.954)
	Rib Fracture	0.999 (0.999–0.999)	0.840 (0.653–0.935)	0.991 (0.991–0.992)	0.984 (0.951–1.000)
	Scoliosis	0.999 (0.999–0.999)	0.698 (0.593–0.786)	0.995 (0.995–0.996)	0.981 (0.961–1.000)
Pulmonary	Atelectasis	0.999 (0.998–0.999)	0.607 (0.5108–0.697)	0.982 (0.980–0.983)	0.962 (0.936–0.987)
	Calcification	0.997 (0.997–0.997)	0.804 (0.770–0.835)	0.962(0.960–0.964)	0.965 (0.955–0.975)
	Consolidation	0.996 (0.995–0.996)	0.702 (0.663–0.737)	0.967 (0.966–0.969)	0.956 (0.944–0.967)
	Emphysema	0.999 (0.999–0.999)	0.580 (0.442–0.706)	0.988 (0.987–0.989)	0.960 (0.922–0.998)
	Fibrosis	0.997 (0.996–0.997)	0.650 (0.598–0.698)	0.977 (0.976–0.978)	0.955 (0.940–0.970)
	Nodule	0.998 (0.997–0.998)	0.719 (0.667–0.766)	0.955 (0.953–0.956)	0.915 (0.894–0.936)
	Opacity	0.992 (0.991–0.993)	0.828 (0.811–0.844)	0.921 (0.919–0.924)	0.925 (0.917–0.933)
	Pleural Effusion	0.997 (0.996–0.997)	0.667 (0.619–0.712)	0.986 (0.985–0.987)	0.972 (0.961–0.984)
	Pneumothorax	0.999 (0.999–0.999)	0.857 (0.486–0.974)	0.998 (0.998–0.998)	0.999 (0.983–1.000)

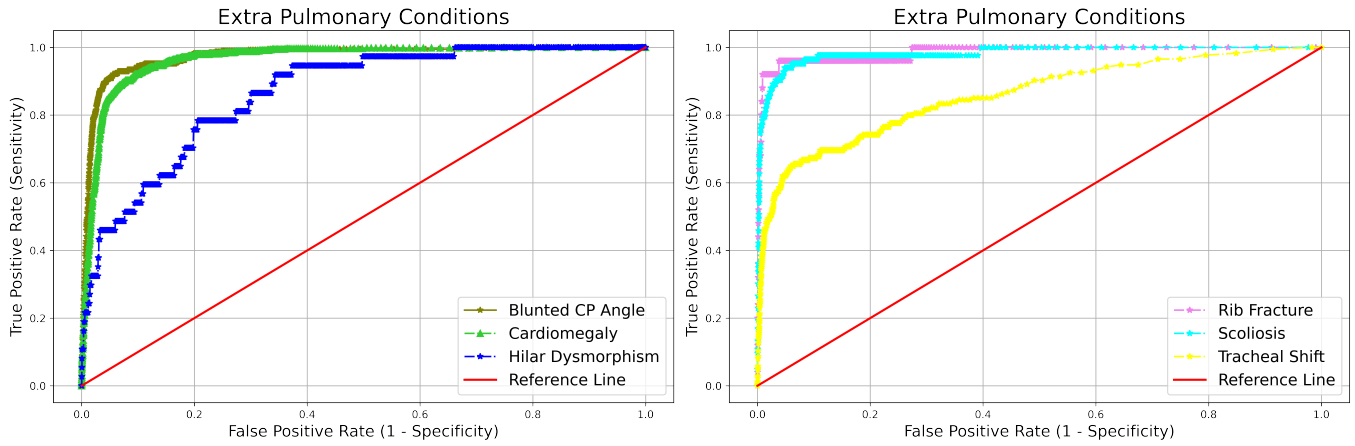

Figure 4. ROC curve qXR performance in identifying subabnormality from abnormal CXR and categorizing into extra pulmonary.

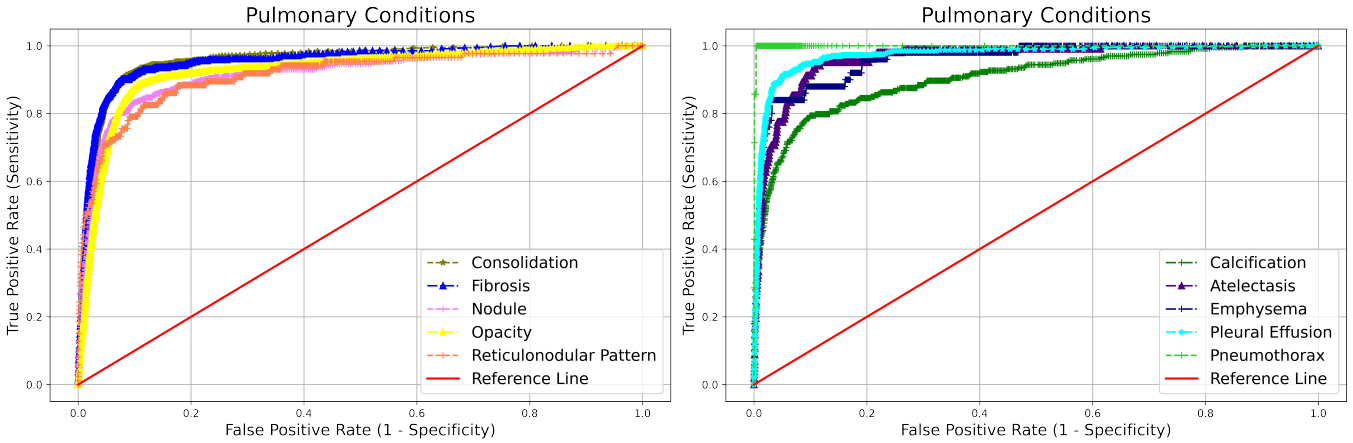

Figure 5. ROC curve qXR performance in identifying subabnormality from abnormal CXR and categorizing into pulmonary conditions.

There were total of 410 CXRs which qXR predicted as normal but were marked as abnormal by the radiologist. Representative CXR processed by the qXR, predicted as normal and abnormal along with abnormality localization is represented in Figure 6.

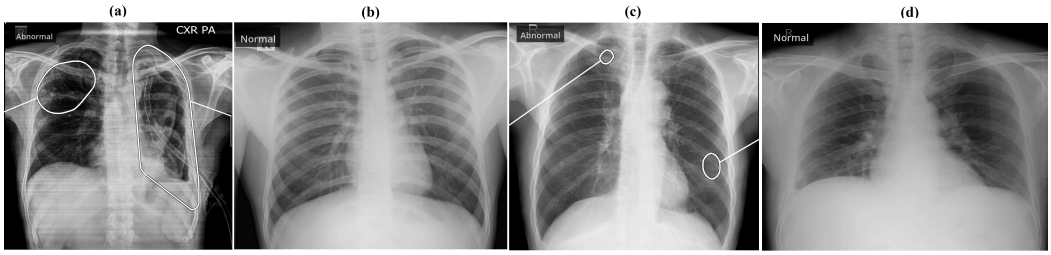

Figure 6. The CXR classified by qXR as (**a**) true positive (TP), (**b**) true negative (TN), (**c**) false positive (FP), and (**d**) false negative (FN). (**a**) Predicted as abnormal with impression of cavity, pleural effusion, pneumothorax, opacity, consolidation and fibrosis in the left lung while cavity and opacity were also predicted in the right lung. (**b**) Predicted as normal. (**c**) Predicted as abnormal with pleural effusion, pneumothorax and opacity in the left lung and right lung. (**d**) Predicted as normal.

Turnaround-Time Analysis

The baseline i.e., pre-qXR data was procured from the radiological scan centres during a three-month period. After the qXR was deployed i.e., post-qXR, we obtained the data randomly from the 48,940 CXR that was considered in final analyis for comparing difference in radiology turnaround time (TAT) between the two periods. TAT as the difference between time at which the final CXR report was reported by the radiologist and the time at which the CXR was assigned to the radiologist. The unpaired *t*-test is used to compute the statisctial signficance of the change in TAT for pre-qXR and post-qXR [27].

There were a total of 18,496 CXRs (9248 (50%) pre-qXR, and 9248 (50%) post-qXR deployment) considered for this analysis. The mean TAT decreased by about 40.63% (83.028 min in the pre-qXR period and 50.287 min in the post-qXR period). This reduction was significant based on the unpaired *t*-test (p-value < 0.0000001). The summary statistics of TAT between the groups are given in the Table 4.

Table 4. Summary statistics of TAT analysis between pre-qXR and post-qXR.

Attributes	Pre-qXR (minutes)	Post-qXR (minutes)
Minimum	11.547	**6.249**
Mean	83.028	**50.287**
Maximum	24,918.617	**14,290.85**

BOLD denotes least time in minutes.

4. Discussion

In this prospective multicenter quality-improvement study, we have analyzed CXR with the help of a commercially available deep learning-based AI algorithm (qXR). The purpose was to identify normal and abnormal findings and to demonstrate the efficacy of triaging AI in routine CXR screening. The high NPV (98.9%) in categorizing normal and abnormal CXR demonstrates the utility of qXR as a screening tool in high-volume facilities. In a prior study, deep learning algorithms were used for interpreting 420 CXRs interpretation from a publicly available database. The reported sensitivity for atelectasis, cardiomegaly, edema, and pleural effusion was 0.750, 0.617, 0.712, and 0.806, respectively [28]. In another study using three AI algorithms for CXR interpretations, 13 abnormalities were reported with a mean AUC for DenseNet121, InceptionResNetV2, ResNet152V2 was 0.793, 0.801, and 0.751, respectively [29]. Prior studies have reported the use of AI as a standalone or a second reader [30] with comparable performance, but validation of such tools in real clinical settings is still missing and represents an active area of research.

Most AI models suffer from poor generalization due to target domain divergence. Previously, the qXR has been evaluated retrospectively in more than 2.3 million CXR, where it categorized CXR into normal and different abnormalities [20]. To determine whether an AI algorithm can be utilized in real clinical settings, a validation of its performance in the real world is required. The added advantage of our study from previously published literature [20,31–33] is that the AI (qXR) has been evaluated prospectively in clinical settings. The algorithm showed high efficacy in classifying CXR into clinically relevant abnormalities, demonstrating its usability in real-time routine CXR screening in large clinical facilities. The overall sensitivity and specificity obtained is 87.9% and 82.9%, respectively. In general clinical practice, CXR serves as a preliminary screening examination for different thoracic and pulmonary abnormalities, where the sensitivity metric is more important, especially in high volume and low-resource facilities. Additionally, the minimum number of false negatives of 0.83% demonstrates its potential as a computer-aided diagnosis tool in emergency departments to improve CXR interpretations. Among the different types of abnormalities, the NPV obtained was 99.9% in almost all abnormalities. Additionally, the high AUC score obtained for different abnormalities shows the capability of the qXR in separating and categorizing different abnormalities. However, among all the abnormalities, the hilar dysmorphism, and reticulonodular pattern achieved a least AUC score of 0.864, 0.913 respectively. This may be attributed to the limited number of true positive samples, i.e., 37, for hilar dysmorphism, and 86 for reticulonodular pattern. In the subgroup analysis, the qXR showed consistent performance in terms of NPV for different subgroups in manufacturer, age, and sex, demonstrating the AI algorithm's robustness.

In the conventional clinical approach, there is minimal use of technology. However, there are many advantages to using AI in radiology departments. With the usage of AI, the delivery of healthcare services to the patient can become more efficient and timely. AI can reduce wait times for patients to get the final report, especially for normal cases, as we have demonstrated in this study. In a conventional system, the normal and abnormal CXR are in the same worklist, and there is no way to segregate/triage normal CXR without opening the CXR. AI as a secondary reader assists in reducing errors in the reports and missed diagnoses [22]. Beyond reduction in reporting time and improvement in report quality, the use of AI will lead to more appropriate treatments for the patients in a timely manner. This is also expected to be reflected in better clinical outcomes downstream in

care management. Adoption of AI in radiology can improve quality metrics and volume of reports significantly over time. This will benefit the radiologists, different department levels of the healthcare facilities, and also the end-to-end patient care.

Although there are many advantages to using technology, some costs/efforts are associated with introducing any innovation. Every innovation has upfront costs, and the benefits are realized over time. Therefore, some physicians and departments might want to wait before they choose to spend to adopt this innovation [34]. Overall, the advantages are much longer than of inconveniences; this study also demonstrates robust evidence to make a case for quicker adoption of AI in radiology. AI not only adds to the efficiency for the clinicians, but a faster adoption is ethical so that the benefits of innovation are passed on to the patients who can benefit from precise, timely, and better care.

There were certain limitations in our study. First, our study was based on a single demographic and racial category; thus, whether the performance of the qXR is reproducible in another demographic and racial category needs to be validated. Secondly, the number of true positives samples was only 6.9% of the final study population. A larger number of true positives would provide more exhaustive testing and further analysis of the qXR. Finally, we want to improve the qXR algorithm and fine tune the threshold to increase the normal CXR prediction, which is truly normal and thus decrease the radiologist's workload further. In the future, it would be beneficial for us to conduct a study with a larger sample and more findings. Another helpful step would be further developing qXR at larger hospitals in emergency settings to measure its performance and compare the TAT in reducing the radiologist's workload.

5. Conclusions

This study has prospectively demonstrated that using AI as an assistance tool can be beneficial in high-workload healthcare facilities. In this multicenter prospective study, the high NPV obtained for overall and multiple abnormalities indicates the use of AI in finding and localizing the abnormalities on CXR. The remarkable AUC score obtained for different clinically relevant abnormalities shows the capability of the AI tool in categorizing the CXR with multiple abnormalities. The triaging of AI in routine CXR screening becomes more critical for developing and underdeveloped countries with a shortage of skilled radiologists. AI tools with high NPV like qXR can be utilized for screening purposes to screen out normal patients, thus allowing clinicians to focus more on patients with abnormalities and their treatment pathways. Additionally, the AI as a second reader enables radiologists to decide rapidly with higher confidence and thus reduce the interrater variability and workload in high CXR volume facilities.

Author Contributions: Conceptualization, S.T. and B.R.; formal analysis, S.C.; investigation, A.G. (Arunkumar Govindarajan) and A.G. (Aarthi Govindarajan); writing—original draft preparation, S.C., R.A., and S.T.; writing—review and editing, A.G. (Arunkumar Govindarajan), A.G. (Aarthi Govindarajan), S.T., B.R., and R.A.; software, P.K. and P.P.; supervision, A.G. (Arunkumar Govindarajan) and A.G. (Aarthi Govindarajan); project administration, D.I. and A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki. This is a quality improvement (workflow optimization) study using a device that is certified by the appropriate authority for use in the region where the study was conducted. There are no safety concerns as a qualified radiologist made the final clinical report with no change in standard of operational procedure (SOP). Such quality improvement studies do not require IRB (Institutional Review Board) approval.

Informed Consent Statement: Patient consent was waived due to nature of our study. The display board for patients awareness was setup displaying the ongoing quality improvement study as per guidelines.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Raoof, S.; Feigin, D.S.; Sung, A.W.; Raoof, S.; Irugulpati, L.; Rosenow, E.C. Interpretation of plain chest roentgenogram. *Chest* **2012**, *141*, 545–558. [CrossRef] [PubMed]
2. Mathers, C.D.; Loncar, D. Projections of Global Mortality and Burden of Disease from 2002 to 2030. *PLoS Med.* **2006**, *3*, e442. [CrossRef]
3. Corne, J.; Pointon, K. *Comprar Chest X-ray Made Easy*, 3rd ed.; Corne, J., Ed.; Churchill Livingstone: London, UK, 2009; ISBN 9780443069222.
4. Singh, R.; Kalra, M.K.; Nitiwarangkul, C.; Patti, J.A.; Homayounieh, F.; Padole, A.; Rao, P.; Putha, P.; Muse, V.; Sharma, A.; et al. Deep learning in chest radiography: Detection of findings and presence of change. *PLoS ONE* **2018**, *13*, e0204155. [CrossRef] [PubMed]
5. Qin, C.; Yao, D.; Shi, Y.; Song, Z. Computer-aided detection in chest radiography based on artificial intelligence: A survey. *Biomed. Eng. Online* **2018**, *17*, 113. [CrossRef] [PubMed]
6. Majkowska, A.; Mittal, S.; Steiner, D.F.; Reicher, J.J.; McKinney, S.M.; Duggan, G.E.; Eswaran, K.; Chen, P.H.C.; Liu, Y.; Kalidindi, S.R.; et al. Chest Radiograph Interpretation with Deep Learning Models: Assessment with Radiologist-adjudicated Reference Standards and Population-adjusted Evaluation. *Radiology* **2019**, *294*, 421–431. [CrossRef]
7. Li, X.; Shen, L.; Xie, X.; Huang, S.; Xie, Z.; Hong, X.; Yu, J. Multi-resolution convolutional networks for chest X-ray radiograph based lung nodule detection. *Artif. Intell. Med.* **2020**, *103*, 101744. [CrossRef]
8. Wang, X.; Zhao, Y.; Pourpanah, F. Recent advances in deep learning. *Int. J. Mach. Learn. Cybern.* **2020**, *11*, 747–750. [CrossRef]
9. Sogancioglu, E.; Çalli, E.; van Ginneken, B.; van Leeuwen, K.G.; Murphy, K. Deep Learning for Chest X-ray Analysis: A Survey. *Med Image Anal.* **2021**, *72*, 102125.
10. Lee, J.H.; Sun, H.Y.; Park, S.; Kim, H.; Hwang, E.J.; Goo, J.M.; Park, C.M. Performance of a Deep Learning Algorithm Compared with Radiologic Interpretation for Lung Cancer Detection on Chest Radiographs in a Health Screening Population. *Radiology* **2020**, *297*, 687–696. [CrossRef]
11. Mehrotra, R.; Agrawal, R.; Ansari, M.A.M.H. Diagnosis of hypercritical chronic pulmonary disorders using dense convolutional network through chest radiography. *Multimed. Tools Appl.* **2022**, *81*, 7625–7649. [CrossRef]
12. Avni, U.; Greenspan, H.; Konen, E.; Sharon, M.; Goldberger, J. X-ray Categorization and Retrieval on the Organ and Pathology Level, Using Patch-Based Visual Words. *IEEE Trans. Med. Imaging* **2011**, *30*, 733–746. [CrossRef] [PubMed]
13. Noor, N.M.; Rijal, O.M.; Yunus, A.; Mahayiddin, A.A.; Gan, C.P.; Ong, E.L.; Bakar, S.A.R.S.A. Texture-Based Statistical Detection and Discrimination of Some Respiratory Diseases Using Chest Radiograph. In *Advances in Medical Diagnostic Technology*; Springer: Singapore, 2014.
14. Cicero, M.D.; Bilbily, A.; Colak, E.; Dowdell, T.; Gray, B.G.; Perampaladas, K.; Barfett, J. Training and Validating a Deep Convolutional Neural Network for Computer-Aided Detection and Classification of Abnormalities on Frontal Chest Radiographs. *Investig. Radiol.* **2017**, *52*, 281–287. [CrossRef]
15. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.P.; Yuille, A.L. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. *arXiv* **2016**, arXiv:1412.7062.
16. Candemir, S.; Jaeger, S.; Lin, W.; Xue, Z.; Antani, S.K.; Thoma, G.R. Automatic heart localization and radiographic index computation in chest x-rays. In Proceedings of the SPIE Medical Imaging 2016: Computer-Aided Diagnosis, San Diego, CA, USA, 27 February–3 March 2016.
17. Guan, Q.; Huang, Y.; Zhong, Z.; Zheng, Z.; Zheng, L.; Yang, Y. Diagnose like a Radiologist: Attention Guided Convolutional Neural Network for Thorax Disease Classification. *arXiv* **2018**, arXiv:1801.09927.
18. Wu, J.T.; Wong, K.C.L.; Gur, Y.; Ansari, N.; Karargyris, A.; Sharma, A.; Morris, M.; Saboury, B.; Ahmad, H.M.; Boyko, O.B.; et al. Comparison of Chest Radiograph Interpretations by Artificial Intelligence Algorithm vs Radiology Residents. *JAMA Netw. Open* **2020**, *3*, e2022779. [CrossRef]
19. Obuchowski, N.A.; Zhou, X.H. Prospective studies of diagnostic test accuracy when disease prevalence is low. *Biostatistics* **2002**, *3*, 477–492. [CrossRef]
20. Putha, P.; Tadepalli, M.; Reddy, B.; Raj, T.; Chiramal, J.A.; Govil, S.; Sinha, N.; Manjunath, K.; Reddivari, S.; Rao, P.; et al. Can Artificial Intelligence Reliably Report Chest X-Rays?: Radiologist Validation of an Algorithm trained on 1.2 Million X-Rays. *arXiv* **2018**, arXiv:1807.07455.
21. Qin, Z.Z.; Ahmed, S.; Sarker, M.S.; Paul, K.K.; Adel, A.S.S.; Naheyan, T.; Barrett, R.; Banu, S.; Creswell, J. Tuberculosis detection from chest x-rays for triaging in a high tuberculosis-burden setting: An evaluation of five artificial intelligence algorithms. *Lancet Digit. Health* **2021**, *3*, e543–e554. [CrossRef]
22. Kaviani, P.; Digumarthy, S.R.; Bizzo, B.C.; Reddy, B.; Tadepalli, M.; Putha, P.; Jagirdar, A.; Ebrahimian, S.; Kalra, M.K.; Dreyer, K.J. Performance of a Chest Radiography AI Algorithm for Detection of Missed or Mislabeled Findings: A Multicenter Study. *Diagnostics* **2022**, *12*, 2086. [CrossRef]
23. Ebrahimian, S.; Homayounieh, F.; Rockenbach, M.A.B.C.; Putha, P.; Raj, T.; Dayan, I.; Bizzo, B.C.; Buch, V.; Wu, D.; Kim, K.; et al. Artificial intelligence matches subjective severity assessment of pneumonia for prediction of patient outcome and need for mechanical ventilation: A cohort study. *Sci. Rep.* **2021**, *11*, 858. [CrossRef]

24. Mahboub, B.M.; Tadepalli, M.; Raj, T.; Santhanakrishnan, R.; Hachim, M.Y.; Bastaki, U.M.A.; Hamoudi, R.A.; Haider, E.A.; Alabousi, A. Identifying malignant nodules on chest X-rays: A validation study of radiologist versus artificial intelligence diagnostic accuracy. *Adv. Biomed. Health Sci.* **2022**, *1*, 137. [CrossRef]
25. Machin, D.; Bryant, T.; Altman, D.; Gardner, M. *Statistics with Confidence: Confidence Intervals and Statistical Guidelines*; John Wiley & Sons: Hoboen, NJ, USA, 2013.
26. Wilson, E.B. Probable Inference, the Law of Succession, and Statistical Inference. *J. Am. Stat. Assoc.* **1927**, *22*, 209–212. [CrossRef]
27. Ruxton, G.D. The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. *Behav. Ecol.* **2006**, *17*, 688–690. [CrossRef]
28. Rajpurkar, P.; Joshi, A.; Pareek, A.; Chen, P.; Kiani, A.; Irvin, J.A.; Ng, A.; Lungren, M.P. CheXpedition: Investigating Generalization Challenges for Translation of Chest X-ray Algorithms to the Clinical Setting. *arXiv* **2020**, arXiv:2002.11379.
29. Albahli, S.; Rauf, H.T.; Algosaibi, A.A.; Balas, V.E. AI-driven deep CNN approach for multi-label pathology classification using chest X-Rays. *PeerJ Comput. Sci.* **2021**, *7*, e495. [CrossRef] [PubMed]
30. Engle, E.; Gabrielian, A.E.; Long, A.; Hurt, D.E.; Rosenthal, A. Performance of Qure.ai automatic classifiers against a large annotated database of patients with diverse forms of tuberculosis. *PLoS ONE* **2020**, *15*, e0224445. [CrossRef] [PubMed]
31. Bejnordi, B.E.; Veta, M.; van Diest, P.J.; van Ginneken, B.; Karssemeijer, N.; Litjens, G.J.S.; van der Laak, J.A.; Hermsen, M.; Manson, Q.F.; Balkenhol, M.C.A.; et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA* **2017**, *318*, 2199–2210. [CrossRef]
32. Hwang, E.J.; Park, S.; Jin, K.N.; Kim, J.I.; Choi, S.Y.; Lee, J.H.; Goo, J.M.; Aum, J.; Yim, J.J.; Cohen, J.G.; et al. Development and Validation of a Deep Learning–Based Automated Detection Algorithm for Major Thoracic Diseases on Chest Radiographs. *JAMA Netw. Open* **2019**, *2*, e191095. [CrossRef]
33. Hwang, E.J.; Park, S.; Jin, K.N.; Kim, J.I.; Choi, S.Y.; Lee, J.H.; Goo, J.M.; Aum, J.; Yim, J.J.; Park, C.M.; et al. Development and Validation of a Deep Learning–based Automatic Detection Algorithm for Active Pulmonary Tuberculosis on Chest Radiographs. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **2018**, *69*, 739–747. [CrossRef]
34. Coppola, F.; Faggioni, L.; Gabelloni, M.; Vietro, F.D.; Mendola, V.; Cattabriga, A.; Cocozza, M.A.; Vara, G.; Piccinino, A.; Monaco, S.L.; et al. Human, All Too Human? An All-Around Appraisal of the "Artificial Intelligence Revolution" in Medical Imaging. *Front. Psychol.* **2021**, *12*, 710982. [CrossRef]