European Journal of Radiology Open 13 (2024) 100606

Contents lists available at ScienceDirect

European Journal of Radiology Open

journal homepage: www.elsevier.com/locate/ejro

Post-deployment performance of a deep learning algorithm for normal and abnormal chest X-ray classification: A study at visa screening centers in the United Arab Emirates

Amina Abdelqadir Mohamed AlJasmi [a], Hatem Ghonim [b], Mohyi Eldin Fahmy [b], Aswathy Nair [c,*], Shamie Kumar [c], Dennis Robert [c], Afrah Abdikarim Mohamed [a], Hany Abdou [b], Anumeha Srivastava [c], Bhargava Reddy [c]

[a] Emirates Health Services, DSO Digital Park Building A8, Dubai Silicon Oasis, Dubai, UAE
[b] Unison Capital Investment LLC, Park Heights Square, Dubai Hills Estate, UAE
[c] Qure.ai Technologies Pvt Ltd, Prestige Summit, 6, St Johns Rd, Halasuru, Bengaluru, India

A R T I C L E I N F O

Keywords:
Chest radiograph
Abnormality
Visa screening
Artificial intelligence
Workflow
Agreement

A B S T R A C T

Background: Chest radiographs (CXRs) are widely used to screen for infectious diseases like tuberculosis and COVID-19 among migrants. At such high-volume settings, manual CXR reporting is challenging and integrating artificial intelligence (AI) algorithms into the workflow help to rule out normal findings in minutes, allowing radiologists to focus on abnormal cases.
Methods: In this post-deployment study, all the CXRs acquired during the visa screening process across 33 centers in United Arab Emirates from January 2021 to June 2022 (18 months) were included. The qXR v2.1 chest X-ray interpretation software was used to classify the scans into normal and abnormal, and its agreement against radiologist was evaluated. Additionally, a digital survey was conducted among 20 healthcare professionals with prior AI experience to understand real-world implementation challenges and impact.
Results: The analysis of 1309,443 CXRs from 1309,431 patients (median age: 35 years; IQR [29–42]; 1030,071 males [78.7 %]) in this study revealed a Negative Predictive Value (NPV) of 99.92 % (95 % CI: 99.92, 99.93), Positive Predictive Value (PPV) of 5.06 % (95 % CI: 4.99, 5.13) and overall percent agreement of the AI with radiologists of 72.90 % (95 % CI: 72.82, 72.98). In the survey, majority (88.2 %) of the radiologists agreed to turnaround time reduction after AI integration, while 82 % suggested that the AI improved their diagnostic accuracy.
Discussion: In contrast with the existing studies, this research uses a substantially large data. A high NPV and satisfactory agreement with human readers indicate that AI can reliably identify normal CXRs, making it suitable for routine applications.

1. Introduction

In countries with high immigration rates and low incidence of pulmonary diseases such as tuberculosis, immigrants are screened either before departure, upon arrival, or post-departure [1]. Holding the fifth position in international migrant population, the United Arab Emirates (UAE) has 7.8 million migrants out of a total population of 9.2 million as per United Nations' estimates of 2013 [2], in UAE, the medical visa screening of expatriates and migrants above 18 years of age is a mandatory process, regulated by comprehensive guidelines for screening and confirming the presence or absence of targeted infectious diseases [3]. Chest radiography, as a primary and pre-dominant screening tool, plays an integral part in the visa screening process, systematically detecting thoracic abnormalities such as tuberculosis, COVID-19 and other abnormalities among migrants. The substantial number of CXRs generated every month in such settings presents a significant challenge for radiologists to report on time amidst subjective variability in interpretations [4,5].

With over 80 % of health systems reporting radiology department shortages and projections indicating persistent staff shortfalls and

* Correspondence to: Qure.ai, Bengaluru, Karnataka, India.
E-mail address: aswathy.nair@qure.ai (A. Nair).

https://doi.org/10.1016/j.ejro.2024.100606
Received 22 July 2024; Received in revised form 20 September 2024; Accepted 10 October 2024

ageing radiologists over the next decade, addressing these challenges and burnout is crucial. According to the Association of American Medical Colleges' analysis, the deficit of radiologists and other specialists could exceed 35,000 by 2034 [6]. In Dubai, the healthcare service utilization is influenced by the epidemiological profile, leading to disparities between demand and supply. This underscores the necessity for more efficient infrastructure utilization and expert availability [7]. UAE radiographers also emphasize on the importance of training and standard guidelines to expand their roles beyond image acquisition, potentially mitigating staff shortage challenges [8,9].

The unreported scans lead to delayed or missed diagnoses, adversely impacting patient care, further leading to outsource and insource requirements to cope with growing demands. Manual interpretation of high-volume CXRs is also a labour-intensive process directing the radiologists toward decision fatigue, missed or incorrect diagnosis and delayed reports for patients. However, precisely stratifying CXRs that appear 'normal' or without any abnormality is critical, to prevent missing or overlooking any notable finding. Consequently, radiologists end up spending a considerable amount of time in evaluating and ruling out the 90 % normal CXRs available in abundance with no evident pathologies, limiting their time and focus to assess complex abnormalities in 10 % of suspected or abnormal cases [10–12].

To mitigate these challenges of burnout, workload, potential medical errors and delays in reporting, the integration of AI-based computer aided design tools into such intricate workflows can aid in segregating normal CXRs with minimal false negatives. AI is also increasingly favoured due to its capability to manage large volumes of images or datasets and analyse a vast number of investigations daily. Recent advancements in the application of deep learning to the interpretation of medical images also target to improve the efficiency and quality of radiological interpretation in routine clinical settings [13]. The process of analysing high-volume CXRs through AI models in clinical routine would assist in reducing processing time, standardizing the reporting workflow, identifying abnormalities with minimal diagnostic errors, prioritizing patients in emergency settings, etc. [14].

Furthermore, significant advancements of AI-powered diagnostic radiology have happened in the field of tuberculosis (TB) detection, with AI assisting the interpretation of chest X-rays. In high-volume settings with limited resources and identifying incidental cases of infectious diseases, AI-powered computer-aided detection (CAD) tools have demonstrated its potential in screening and triage [15]. Recognizing AI's potential in TB screening and triage, the World Health Organization (WHO) has approved AI-based computer-aided detection (CAD) for individuals aged 15 or older and their use as an alternative to human interpretation of chest X-rays [16,17]. In low-incidence countries like Oman, policy changes by Minister of Health have enabled the screening of expatriates and immigrants for communicable diseases like TB mandatory. With the change from symptom based to radiological screening since 2018, designing operational studies can further assess the yield for future planning [18]. Deep learning algorithm have also shown potential in detecting and classifying cancer growths from radiology images [19–21].

Diagnostic accuracy of the AI software used in this study for classifying normal versus abnormal has been previously studied and reported. With a sensitivity of 96 % and specificity of 100 %, AI have identified all missed and mis-labelled findings in CXRs reducing errors among radiologist [22]. One study has reported the sensitivity of qXR in finding abnormal CXRs is about 997 % and specificity of about 67.4 % [13]. Another study reported the sensitivity and specificity of qXR in finding abnormal CXRs to be 96.88 % and 75.55 % respectively [23]. Both studies reported a high NPV of >99 % too. Both studies used a panel of expert radiologists to establish ground truth. However, these studies were limited by relatively small sample size of < 1000 CXRs and were done in controlled atmosphere of a research study evaluation thereby raising questions on real-world performance of AI. These studies also utilized CXRs from patients presenting to hospital for one or more reasons. However, this study is not designed to be a controlled validation study and hence have utilized a large number of consecutive CXRs in a VISA screening setting which is substantially different from the study populations studied earlier.

The primary objective of this multi-centre retrospective study is to assess the agreement of AI (qXR version 2.1) in classifying normal versus abnormal CXRs with reporting radiologists. Further, digital survey assessments of clinical and technical users provide insights into the perceived implementation challenges and benefits of integrating AI into clinical workflows.

2. Methods

2.1. Device description

Developed by Qure.ai, a Software as a Medical Device organization, qXR version 2.1 is an AI computer-aided detection (CAD) device intended for detecting presence or absence of various radiological abnormalities in CXRs. It consists of multiple deep-learning algorithms for multiple radiological abnormalities and was trained on a dataset of over 2 million CXR images using expert radiologist labels for supervised classification task [24]. qXR is CE marked for classifying CXRs into normal and abnormal ones which was the focus of this study. A description of the algorithm development has been previously reported and the same article also reported a high area-under-the-receiver-operating-characteristics-curve (AUC) of 0.92 (CI 0.91–0.94) for detection of abnormal CXRs [25]. The AI software classifies the CXRs into normal when no radiological abnormalities are detected in the CXR and abnormal if abnormalities (e.g. lung nodule, blunted costophrenic angle, cardiomegaly, cavity, consolidation, fibrosis, hilar enlargement, opacities, pleural effusion, pneumothorax, radiological signs of tuberculosis, rib fracture) are detected. In this study, we focused on the normal versus abnormal classification feature of qXR.

2.2. Study design

For agreement analysis: This is a retrospective multicentre post-deployment study in which all the consecutive frontal (PA/AP) CXRs between January 2021 and June 2022 (18 months) acquired from patients aged 18 years or older as a part of the visa screening process from 33 centres in UAE were considered for analysis. All the CXRs with patient demographics like age and gender, quarter or year of X-ray acquisition, and normal versus abnormal classification by AI and radiologist were obtained for analysis. Any CXRs with a non-available AI result and CXRs where the radiologist suggested a repeat CXR due to lack of image quality were excluded from analysis.

For survey analysis: The digital survey based on forms were electronically sent to the radiologists and IT professionals working in the study sites. The survey questionnaire included 20 questions (S1 supplementary), and the corresponding responses were collected between 14th and 29th of February 2024. The email responses from 17 radiologists and 3 health professionals (PACS/ IT managers) who had prior experience in using the AI software were collected for analysis post receiving their consent, all within a period of 1-month.

2.3. Dataset

The retrospectively collected CXR dataset included variables like anonymous patient identifier, AI and radiologist interpretation as normal or abnormal), age of the subject at the time of CXR acquisition in years, gender of the subject, and quarter of the year (e.g.: 2022 Q1) in which the CXR was done. Definitions of classification metrics for evaluating the performance of AI models are shown in Table 1. All the data collected as per the pre-defined criteria during the course of this study was stripped of any patient identifying information in compliance with

Table 1
Basic classification metrics and its definition for statistical estimation of model performance.

Metric	Definition
True Positive (TP)	Both AI and radiologist classified a CXR as abnormal.
True Negative (TN)	Both AI and radiologist classified a CXR as normal.
False Positive (FP)	AI classified a CXR as abnormal, but the radiologist deemed it as normal.
False Negative (FN)	AI classified a CXR as normal, but the radiologist deemed it as abnormal.

HIPAA and GDPR guidelines.

2.4. Statistical analysis

For a normal versus abnormal classification, the NPV (normal predictive correlation), PPV (abnormal predictive correlation), positive percent agreement (PPA), and negative percent agreement (NPA), are reported along with their 95 % confidence interval (95 % CI) constructed using the exact binomial method. We used PPA and NPA instead of sensitivity and specificity guided by recommendations from the Federal Drug Agency (FDA) for computer-aided detection (CAD) evaluations using non-reference standards [26]. Statistically speaking, PPA and NPA are similar in mathematical forms to sensitivity and specificity, respectively. Like sensitivity, positive percent agreement indicates the proportion of true positives (considered positive by both AI and radiologist) to the total number of CXRs confirmed as positive, i.e., abnormal by radiologist (this includes the true positives and false negatives). Similarly, like specificity, negative percent agreement indicates the proportion of true negatives (considered negative by both AI and radiologist) to the total number of CXRs confirmed as negative, i.e., normal by radiologist (this includes the true negatives and false positives).

The calculations of these four metrics are shown below.

$$NPV = \frac{TN}{\mathrm{TN} + \mathrm{FN}}$$

$$PPA = \frac{TP}{\mathrm{TP} + \mathrm{FN}}$$

$$NPA = \frac{TN}{\mathrm{TN} + \mathrm{FP}}$$

$$PPV = \frac{TP}{\mathrm{TP} + \mathrm{FP}}$$

Subgroup analysis stratified by gender, age group, and time period (quarter) are also reported. Agreement statistics quantifying the overall percentage agreement (PA) along with 95 % CI are also reported.

2.5. Deployment at site

With 33 radiology centres among 136 facilities in the UAE, the site faces the operational challenges of processing millions of CXRs involving timely detection and reporting of abnormalities for making prompt clinical decisions, ensuring an efficient visa process. Balancing the workload of radiologists and healthcare professionals also enhances clinical and administrative productivity, leading to faster decisions and improved healthcare quality. To tackle these issues, the team of radiologists from the site opts for the central deployment of an evidence-based, data-driven, chest X-ray interpretation software across their 33 visa screening centres from January 2021 to streamline the visa screening processes with minimal clinical and operational risks. AI classify all the CXRs into normal and abnormal (include flagging of cases with tuberculosis and COVID among immigrants) within minutes. Further investigation on the reported flags and discordant cases are carried out by the radiologists.

3. Results

3.1. Diagnostic accuracy estimation

Based on the Standards for Reporting of Diagnostic Accuracy (STARD) guidelines, the data flow diagram showing the selection process of CXRs are represented in Figure After excluding 706 CXRs with low image quality and 93,370 without AI processed outputs, a total of 1309,443 CXRs from 1309,431 patients (median age: 35 years; IQR [29–42]; 1030,071 males [78.7 %]) were included in the analysis. The baseline characteristics of the selected 1309,443 CXRs with respect of age, gender, and time period of CXR acquisition quarter wise are shown in Table 2. The contingency table (Table 3) shows the samples of TP, TN, FP, and FN used for analysis from the selected dataset as per the defined test criteria against reference standard.

The analysis of 1309,443 CXRs in this study revealed a NPV of 99.92 % (95 % CI: 99.92, 99.93) and a PPV of 5.06 % (95 % CI: 4.99, 5.13). The point estimates and 95 % CI of NPV, PPA, NPA, and PPV are shown in Table 4. As a comparison, a hypothetical naïve classifier classifying all CXRs as normal would have attained an NPV of 98.50 % due to high observed prevalence of about normal CXRs in the sample. This means that the margin of improvement for any other classifier models is 1.5 % suggesting that AI crossed 94.67 % (1.42 % marginal increase observed using the AI software in this study) of that margin. The overall percent agreement between the AI and reporting radiologist was found to be 72.90 % (95 % CI: 72.82, 72.98).

Sub-group analysis based on age and gender is presented in Table 5, while Fig. 2 illustrates the correlation of the increasing trend in NPV values following the threshold adjustments made in 2021 Q2 for the classification of normal versus abnormal cases. NPV was found to be comparable in males 99.92 (95 % CI: 99.92–99.93) and females 99.91 (95 % CI: 99.90–99.92) while a slightly lower NPV was observed in the participants aged > 65 years (NPV: 99.70, 95 % CI: 99.45–99.86) compared to younger participants aged 18–35 years (NPV: 99.94, 95 % CI:99.93–99.94).

Table 2
Baseline demographic characteristics of the CXR data included in the analysis. *Nine out of the 1309,431 patients were having age more than 100 years. The possibility of a data entry issue can not be ruled out. We did not exclude these patients because this group constituted only a very miniscule proportion and the CXRs from these patients were all valid CXRs.

Age	
Metric	**Value**
N	1309,431
Mean	3613
Median	35.00
Inter-quartile range	29.00 – 42.00
Minimum	18.00
Maximum*	134.00
Standard Deviation	9.91
Age Group	
Group	**N (%)**
18 – 35	691540 (52.81 %)
36 – 65	608252 (46.45 %)
> 65	9639 (0.74 %)
Gender	
Group	**N (%)**
Male	1030,071 (78.66 %)
Female	279360 (21.34 %)
Number of CXRs by Quarter (Total N = 1309,443)	
Group	**N (%)**
2021 Q1	212071 (16.19 %)
2021 Q2	234537 (17.91 %)
2021 Q3	197305 (15.07 %)
2021 Q4	195310 (14.91 %)
2022 Q1	276206 (21.09 %)
2022 Q2	194014 (14.83 %)

Table 3
Contingency table showing true positive (TP), false positive (FP), false negative (FN) and true negative (TN) numbers of AI in comparison to reporting radiologist for normal versus abnormal classification of CXRs.

	Rad Abnormal	Rad Normal	Total
AI Abnormal	18,865 (TP)	354,100 (FP)	372,965
AI Normal	737 (FN)	935,741 (TN)	936,478
Total	19,602	1289,841	1309,443

Table 4
Overall results showing NPV, PPA, NPA and PPV point estimates and 95 % confidence intervals.

Metric	Point Estimate	Lower 95 % CI Limit	Upper 95 % CI Limit
NPV	99.92	99.92	99.93
PPA	96.24	95.96	96.50
NPA	72.55	72.47	72.62
PPV	5.06	4.99	5.13

Table 5
Diagnostic accuracy estimates stratified by the sub-groups like age, gender, and quarter of CXR data acquisition.

Age Group	NPV (95 % CI)	PPA (95 % CI)	NPA (95 % CI)	PPV (95 % CI)
18–35	99.94 (99.93–99.94)	95.58 (95.09–96.04)	78.13 (78.03–78.23)	4.53 (4.43–4.63)
36–65	99.90 (99.89–99.91)	96.56 (96.21–96.88)	66.69 (66.57–66.81)	5.33 (5.24–5.43
> 65	99.70 (99.45–99.86)	98.26 (96.83–99.16)	37.06 36.07–38.07	9.03 (8.33–9.76)
Gender	**NPV (95 % CI)**	**PPA (95 % CI)**	**NPA (95 % CI)**	**PPV (95 % CI)**
Male	99.92 (99.92–99.93)	96.52 (96.22–96.80)	71.83 (71.74–71.92)	5.11 (5.03–5.180
Female	99.91 (99.90–99.92)	95.02 (94.27–95.70)	75.19 (75.03–75.35)	4.86 (4.71–5.02)
Quarter	**NPV (95 % CI)**	**PPA (95 % CI)**	**NPA (95 % CI)**	**PPV (95 % CI)**
2021Q1	99.88 (99.86–99.90)	94.09 (93.21–94.88)	75.54 (75.35–75.72)	5.56 (5.37–5.75)
2021Q2	99.88 (99.86–99.90)	93.87 (93.01–94.64)	76.46 (76.28–76.63)	5.63 (5.44–5.82)
2021Q3	99.95 (99.94–99.96)	97.65 (97.01–98.19)	70.77 (70.57–70.97)	4.40 (4.24–4.57)
2021Q4	99.95 (99.94–99.96)	97.64 (96.99–98.18)	70.31 (70.1–70.51)	4.35 (4.19–4.52)
2022Q1	99.94 (99.92–99.95)	97.01 (96.44–97.51)	70.92 (70.75–71.09)	4.84 (4.69–4.99)
2022Q2	99.94 (99.92–99.95)	97.51 (96.94–98)	70.93 (70.72–71.13)	5.73 (5.55–5.93)

3.2. Survey analysis

The digital survey was planned to understand the real-life challenges and benefits of the deployed AI software from its actual users and the questionnaire (S1 supplementary) covers the aspects of trust factor, apprehensions, awareness, and opinions related to the evolving use of AI used in radiology and its adoption within the clinical workflow. The distribution of survey participants is shown in Fig. 3 and out of the 17 radiologists participated in the survey, 14 (82.3 %) routinely report CXRs for visa screening investigations and though the remaining 3 PACS or IT Managers do not report CXRs, they do validate the images on arrival. Approximately 65 % of the respondents have experience in using AI before, leaving out 35 % with no prior experience. Further, 16 (80 %) suggested they had to put in minimal effort to implement the AI technology at work and 17 (85 %) did have a good understanding of this technology they use and know how it works.

In the survey, the responses from participated radiologists toward the different questions were measured in Likert scale (Fig. 4) and 71 % of them were familiar and at ease to use AI in their practice, with majority acknowledging its use positively. 82 % of radiologists also agreed that the AI aiding the CXR interpretation improved their diagnostic accuracy, reducing the likelihood of missed diagnoses. 88.2 % of them suggested that there is a reduction in turnaround time after adopting AI into the workflow with a positive impact and no disagreement was reported among radiologists in AI assisting routine CXR interpretation. Based on their previous experience in using AI, 70.5 % positively acknowledged that their reporting times are faster when compared to manual analysis, while 29.4 % remained unbiased in their opinion. Other suggestions for improvement provided by the radiologists are shown in Fig. 5.

About 29.4 % of participating radiologists also consider it realistic that AI software could eventually automate certain aspects of their role, such as routine checks of the Preventative Medicine Department, generating directed reports for normal radiographs, and serving as an adjuvant tool. While 41.1 % hold the opposite view on the same, the remaining 29.4 % expressed their responses in a neutral way. Among the perceived benefits of AI in the clinical process or radiologist workflow, majority (53 %) of the radiologists suggested reduced reporting time, followed by workload reduction, improved sensitivity and productivity, and workflow enhancement. While 24 % of radiologists had neutral opinions on the AI performance in diagnostic workflows, positive suggestions for AI integration included its use for classifying and segmenting lesions, assisting in routine screening, detecting X-ray fractures, and more. The need for more AI accuracy studies in real-world settings was also highlighted. The radiologists also suggested the general need for improvement in areas of lesion detection, customization required as per area/ region of interest, regular update of variable normal appearances, etc. 24 % of the radiologists also raised concerns related to AI in terms of overreading, low accuracy, over-reliance, and flagging insignificant things. The PACS/IT managers also highlighted the use of AI in routine CXR applications to reduce the workload of the radiologists and improve their reporting time and overall productivity.

4. Discussion

In our study, we observed a high NPV of AI (99.92 %, 95 % CI: 99.92, 99.93) indicating that the algorithm can accurately classify the normal CXRs with minimal false negative rate. It can be argued that false negative (wrongly flagging an abnormal CXR as normal) result is more harmful than false positive result, especially in the context of a hypothetical auto normal reporting scenario because in such a scenario a false positive CXR will still be reviewed a radiologist. In this context, a high NPV is desired because it indicates high confidence that a CXR classified as normal by AI is truly normal. The observed PPV of 5.06 % (95 % CI: 4.99–5.13) is apparently low and the low prevalence of radiologist-deemed abnormal CXRs could have contributed to this. Both NPV and PPV are dependent not only on the device accuracy, but also on the prevalence of the target condition. The observed PPA (96.52, 95 % CI: 96.22–96.80) and NPA (71.83, 95 % CI: 71.74–71.92) suggests that AI has very good agreement with radiologists for abnormal CXRs and moderate agreement with radiologists for normal CXRs. The relatively lower NPA may be due to AI flagging even non-clinically relevant findings such as old rib fractures, small and long-standing lung nodules, etc [23]. Post-deployment measures taken to improve the NPV through threshold adjustments are necessary for optimising the performance of AI. The increasing NPV trend following the threshold adjustments made in 2021 Q2 for the classification of normal versus abnormal cases further supports it.

The need for automated interpretation of normal CXRs to reduce the radiologist workload is clear and AI system has reported abnormalities from 28 % of normal CXRs with 99 % sensitivity [27]. In health-screening settings, AI triage can efficiently identify normal CXRs, reducing the workload of radiologist by 40 % and improving reader specificity [28]. AI has also demonstrated consistently high performance

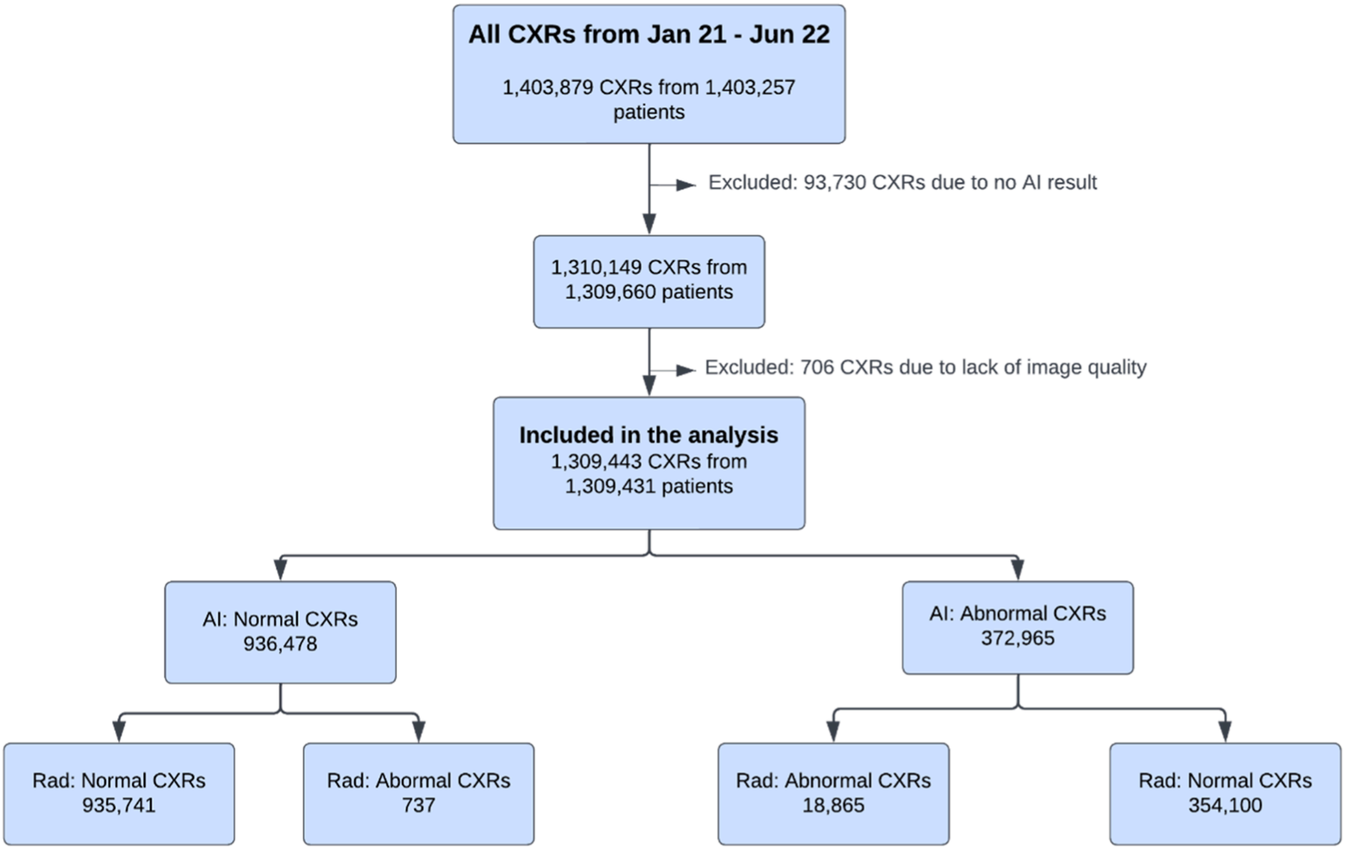

Fig. 1. Data flow diagram illustrating the selection process of the CXRs used in the study analysis.

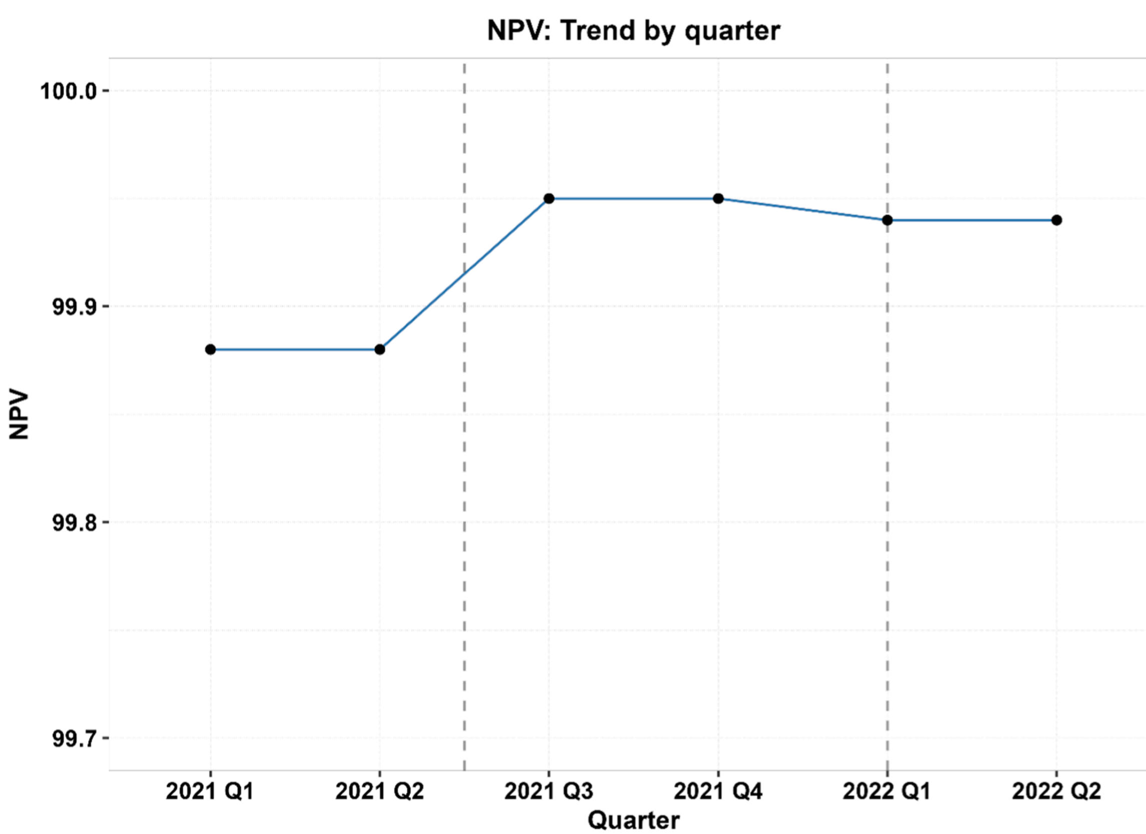

Fig. 2. Trend of NPV stratified by quarter during the study period (January 2021 to June 2022). The two dashed vertical lines indicate the two threshold adjustments done for improved normal versus abnormal classification in August 2021 and March 2022.

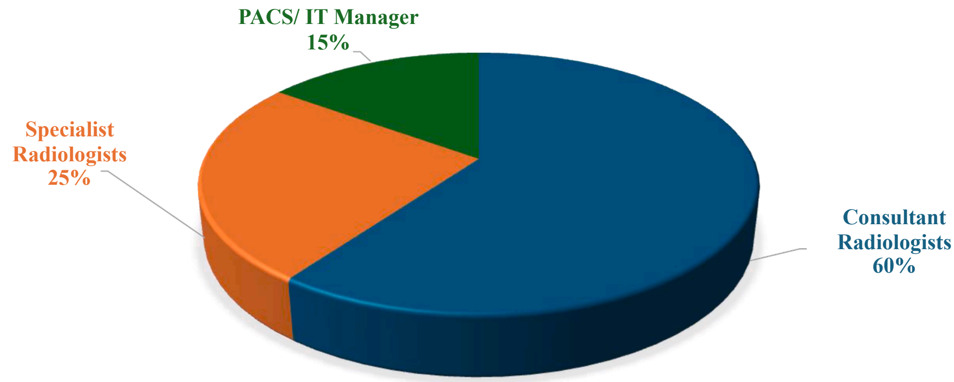

Fig. 3. Percentage distribution of total participants (n=20) responded to the digital survey across various health categories of the study setting.

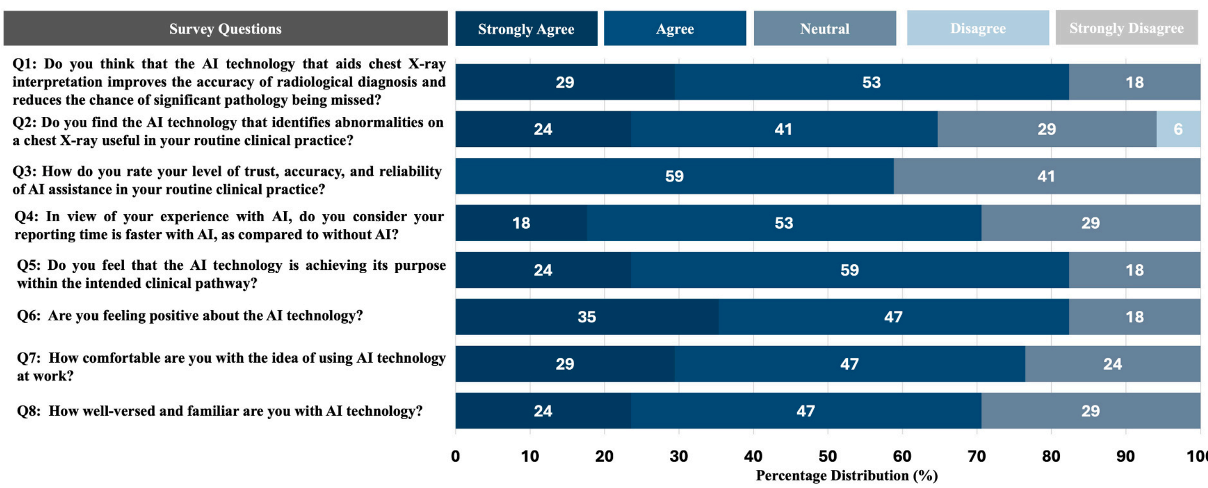

Fig. 4. Responses of participated radiologists (n=17) expressed in Likert scale toward the prime survey questions regarding the impact of artificial intelligence technology integration in the study setting.

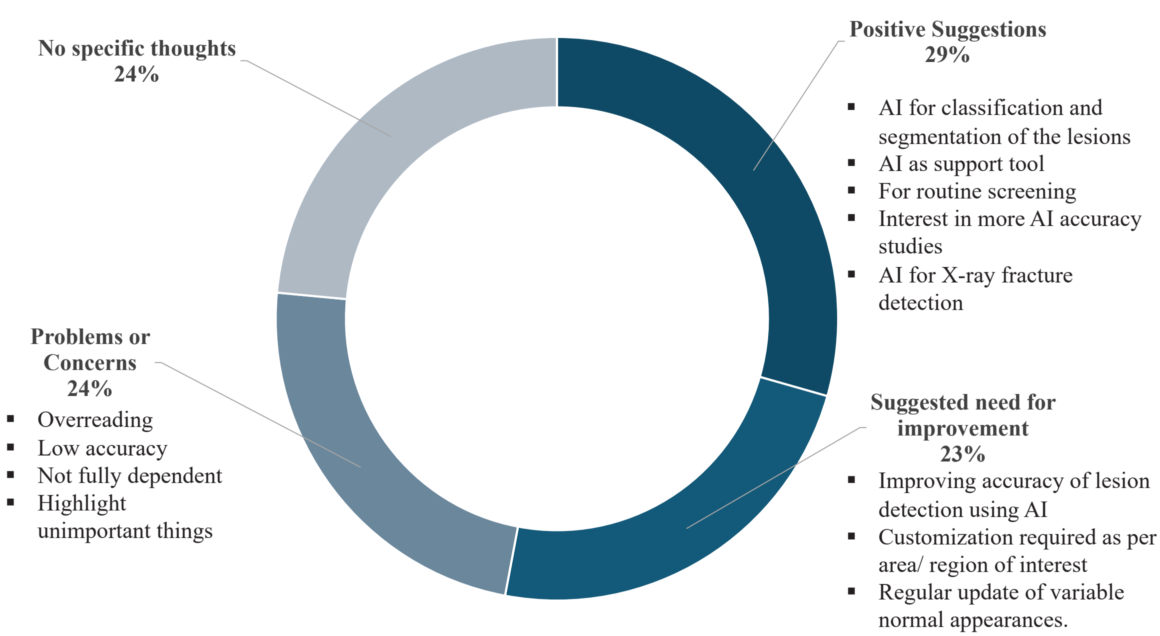

Fig. 5. Critical inputs expressed by the radiologists (n=17) related to using artificial intelligence for interpreting chest X-rays, highlighting the areas of improvement and concerns.

in classifying CXRs with normal and abnormal findings, including major thoracic diseases, surpassing radiologists in some cases and potentially enhancing the overall quality and efficiency of clinical workflow [29]. Setting a confidence threshold can help to classify the CXRs into highly probable normal and abnormal and aid diagnosis [30]. Studies determining the usability of AI to rule-out normal cases show a precision of 97.7 % and further emphasize on enabling radiologists to focus more on complex abnormalities [31]. It is also reported that the average time taken to report critical imaging findings decreased significantly from 11.2–2.7 days [32]. The multiple AI models showed a reported sensitivity over 95 % for classifying normal versus abnormal chest radiographs [13,33]. Another retrospective study revealed that AI could identify normal chest radiographs without pathologies in 24.5 % to 52.7 % of cases, with a sensitivity of 98 %. Also, only 1.1 % of critical errors are observed above 95 % sensitivity and it is comparable or less than that observed with the radiologist standard [34].

Prior assessments of the AI software used in this study have focused on its ability to differentiate between normal and abnormal CXRs, revealing a negative predictive value of 98.9 % [13]. Implementation of AI has resulted in a significant decrease in turnaround time, reducing it by around 40.63 % [35]. In another study, AI demonstrated high sensitivity compared to radiologists, making it particularly beneficial in resource-constrained clinical settings where there's a shortage of trained radiologists [36]. Specifically tailored for primary chest X-ray interpretation, AI holds promise in reducing both the time and cost associated with reporting normal studies, thereby allowing radiologists to dedicate more attention to abnormal cases. Using the radiologist as the reference standard, AI also demonstrated an overall sensitivity of approximately 87.9 % (95 % CI: 86.7–88.9) in detecting pulmonary abnormalities using dataset from UAE hospital. The level of agreement is observed to be consistent across various subgroups, including age and gender [37]. Globally, many cases may not be reported within the specified timeframe and patient pathway, yet the substantial negative predictive value (NPV) values hold great potential for improving clinical support. In a different service evaluation conducted using a randomly sampled retrospective CXR dataset, the AI software has revealed 97 % sensitivity and 99 % NPV in classifying a chest radiograph as normal and abnormal with 92 % AUC [23].

As this study focusses on improving the radiological efficiency in high-volume settings like visa screening centres, we believe that AI could be instrumental in reducing both the time and cost associated with reporting normal scans, thereby allowing radiologists to focus more onto abnormal cases. In clinical workflows, further categorizing the abnormal scans is also crucial to triage the migrant population and AI can be used for secondary reads to identify discrepancies in CXR findings. The model performance and generalizability of AI findings may vary with respect to age-related variation, image acquisition parameters with respect to radiation dose, sub-populations, etc. Hence, training and validating the algorithms over an expansive dataset is important [38]. Further, recent studies also highlight the role of AI in redefining traditional workflows and optimizing the radiologist efficiency with more uniform reporting, correlating diagnoses with other clinical reports and treatment plans to flag discrepancies on time, along with patient tracking and care coordination between healthcare provider and patient, optimizing resources, etc [39]. The major strength of this study is the inclusion of a huge volume dataset used for performance evaluation and incorporating data from multiple centres. The multi-centre approach also supports the generalizability of the AI findings, its impact on the routine clinical workflow, and allows for a comprehensive understanding of radiological patterns and trends across patient types and settings.

One of the limitations of this study is using a single radiologist ground truth as reference standard without clinical or microbiological confirmation, as it may lack confidence due to the possible subjective interpretations and inter-reader variability. And as this is not a validation study, with large volume of chest X-rays analysed routinely, setting the ground truth post-deployment can be challenging or operationally impractical. Also, having the single radiologist ground truth to analyze the agreement between the AI algorithm and radiologists in classifying chest X-rays as normal or abnormal contributed to meet the study objective. Apart from this, another limitation of this study is the exclusion of 93,730 CXRs from analysis (Fig. 1) as they did not have AI results due to suboptimal image quality and resolution (1440*1440 pixels) required for processing. However, this affected only 6.7 % (93730 out of 1403,879) of the CXRs. Further, discordant cases were not reviewed by another radiologist due to the retrospective nature of the analysis. Filtering these discordant cases from such extensive data volumes before analysis was not feasible, and it was beyond the scope of this study. Additionally, as a potential limitation, the current study has not quantitatively analysed the turnaround time as accessing the required datapoints in a large volume study was difficult and hence the reduction in turnaround time was assessed only through the radiologist opinions from digital survey.

So, in future, prospective study to assess the real time impact of the AI integration, prospective data related to estimating turnaround time, cost effectiveness and patient outcomes can be designed. A failure analysis, including the estimation of false positives and false negatives, along with the underlying factors influencing such results, can help identify potential overreading or underreading errors by the AI algorithm. Periodic post-marketing surveillance would help build confidence in the safety and effectiveness of AI used in routine clinical workflows. The studies focussing on the performance of AI to identify and categorize the abnormalities after eliminating the normal scans can also assist the radiologists in specific disease diagnosis and triaging among migrants.

5. Conclusion

For validated, hardware agnostic and compatible algorithms, with regulatory and data compliance (HIPAA and GDPR) in place, the adoption of AI in clinical workflows are rapidly expanding. In this study, the analysis of over 1.3 million CXRs from a visa screening centre demonstrated a NPV of 99.92 % and an overall percent agreement of 72.90 % between radiologists and AI. The digital survey indicated interest of 82 % radiologists in AI assisted CXR interpretation and 88.2 % suggested that there is a reduction in turnaround time after AI integration. In conclusion, standardization within clinical workflows is pivotal and this also includes routine checks within the department, producing computerized reports for normal radiographs, and serving as an adjuvant tool to clinicians as suggested by the radiologists in the survey. Once there is alignment between the AI interpretations to that of radiologists, integrating AI into the workflow reduce fatigue, minimize the risk of incorrect diagnoses associated with labour-intensive procedures, and enhance operational efficiency.

Ethics approval

The approval from the Research Committee of Emirates Health Services (MOHAP/DXB-REC/S.S. N/No.127 /2023) was obtained before the study initiation.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Funding source

There is no external funding obtained for this study.

Author contribution

The authors of the manuscript Amina Abdelqadir Mohamed AlJasmi, Hatem Ghonim, Mohyi Eldin Fahmy, and Shamie Kumar contributed to the conception, design, and data extraction for this study. Dennis Robert contributed to the data analysis, interpretation, and review. The corresponding author of this manuscript Aswathy Nair contributed to literature search, methodology, survey analysis, and manuscript writing. Shamie Kumar, Amina Abdelqadir Mohamed AlJasmi, Afrah Abdikarim Mohamed, Hatem Ghonim, Mohyi Eldin Fahmy, Hany Abdou, Anumeha Srivastava, Bhargava Reddy also contributed to the critical revision of the manuscript and final approval. All the authors approve the manuscript and guarantee the integrity of the work.

Ethical statement

The ethics approval from the Research Committee of Emirates Health Services (MOHAP/DXB-REC/S.S. N/No.127 /2023) was obtained before the study initiation and data collection. The research work mentioned in this article has not been published previously and is not under consideration for publication elsewhere. This article is also approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out.

CRediT authorship contribution statement

Mohyi Eldin Fahmy: Project administration, Data curation, Conceptualization. **Aswathy Nair:** Writing – original draft, Methodology. **Shamie Kumar:** Writing – review & editing, Validation, Methodology, Data curation, Conceptualization. **Dennis Robert:** Writing – review & editing, Methodology, Formal analysis. **Amina Abdelqadir Mohamed AlJasmi:** Supervision, Project administration, Data curation, Conceptualization. **Hatem Ghonim:** Project administration, Data curation, Conceptualization. **Afrah Abdikarim Mohamed:** Writing – review & editing, Project administration. **Hany Abdou:** Writing – review & editing, Project administration. **Anumeha Srivastava:** Writing – review & editing. **Bhargava Reddy:** Writing – review & editing.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Dr. Aswathy Nair reports a relationship with Qure.ai Technologies Private Limited that includes: employment. Dr. Dennis Robert, Shamie Kumar, Anumeha Srivastava, Bhargava Reddy reports a relationship with Qure.ai Technologies Private Limited that includes: employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Acknowledgement

The authors would like to thank Dr. Asma Khammas, Mr. Mohammed Salah, and Ms. Reem Matar from Emirate Health Services and Mr. Ahmed Seada from Unison IT team for their continuous support and guidance throughout the study specifically in the course of ethics committee approval, technology integration, data collection, and digital survey. The authors also acknowledge the contributions of Manoj Tadepalli and the qXR R&D team, whose support in algorithm development, training, and validation was instrumental to this work.

Consent

Informed consent was obtained from the radiologists and healthcare professionals prior to their participation in the digital survey.

Data sharing

De-identified patient data is contained within the article for analysis.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejro.2024.100606.

References

[1] G.G. Alvarez, B. Gushulak, K.A. Rumman, E. Altpeter, D. Chemtob, P. Douglas, et al., A comparative examination of tuberculosis immigration medical screening programs from selected countries with high immigration and low tuberculosis incidence rates, BMC Infect. Dis. 11 (1) (2011 Jan 4), https://doi.org/10.1186/1471-2334-11-3.

[2] Migration Policy Institute. Labor Migration in the United Arab Emirates: Challenges and Responses. [Accessed on 28th March 2024] Available from: 〈https://www.migrationpolicy.org/article/labor-migration-united-arab-emirates-challenges-and-responses〉.

[3] F.I. Al Hosani, G.A. Yahia, Prevalence of pulmonary tuberculosis among expatriates subjected to medical visa screening in Abu Dhabi, United Arab Emirates, J. Epidemiol. Glob. Health 3 (1) (2013) 23.

[4] A.M. Ajlan, Technical quality and diagnostic impact of chest x-rays in tuberculosis screening: insights from a saudi teleradiology cohort, Curēus 16 (2) (2024 Feb 3) e53509, https://doi.org/10.7759/cureus.53509.

[5] J. Singh, S. Al-Abri, E. Petersen, F. Al Yaqoubi, K. Al Rahbi, L. Al Balushi, et al., Importance of tuberculosis screening of resident visa applicants in Low TB incidence settings: experience from Oman, J. Epidemiol. Glob. Health 12 (3) (2022 Sep) 281–291, https://doi.org/10.1007/s44197-022-00040-w.

[6] Radiology Business. Minding the Gap: Strategies to Address the Growing Radiology Shortage [Accessed on 28th March 2024] Available from: 〈https://www.itnonline.com/article/minding-gap-strategies-address-growing-radiology-shortage〉.

[7] N. Monsef, E. Suliman, E. Ashkar, H.Y. Hussain, Healthcare services gap analysis: a supply capture and demand forecast modelling, Dubai 2018–2030, BMC Health Serv. Res. 23 (1) (2023 May 10), https://doi.org/10.1186/s12913-023-09401-y.

[8] M.M. Abuzaid, W. Elshami, J. McConnell, M. Baird, Changing the model of radiography practice in the UAE: a snapshot of a profession in transition, Radiography 27 (1) (2020 Jun), https://doi.org/10.1016/j.radi.2020.05.014.

[9] M.M. Abuzaid, W. Elshami, M. Kadhom, J. McConnell, S. Mc Fadden, The changing concept of radiographer's role in UAE: an analysis of radiologists' opinions and acceptance, Radiogr. (Lond.) 28 (4) (2022 Nov) 1042–1049, https://doi.org/10.1016/j.radi.2022.07.010.

[10] L.I.T. Lee, S. Kanthasamy, R.S. Ayyalaraju, R. Ganatra, The current state of artificial intelligence in medical imaging and nuclear medicine, BJR|Open 1 (1) (2019 Jul) 20190037, https://doi.org/10.1259/bjro.20190037.

[11] S.J. Adams, R.D.E. Henderson, X. Yi, P. Babyn, Artificial intelligence solutions for analysis of X-ray images, Can. Assoc. Radio. J. 72 (1) (2021) 60–72, https://doi.org/10.1177/0846537120941671.

[12] Gulf News. DHA to use Artificial Intelligence to sort medical fitness X-rays. [Accessed on 23 March 2024]. Available from: 〈https://gulfnews.com/going-out/society/dha-to-use-artificial-intelligence-to-sort-medical-fitness-x-rays-1.2166166〉.

[13] S.R. Blake, N. Das, M. Tadepalli, B. Reddy, A. Singh, R. Agrawal, S. Chattoraj, D. Shah, P. Putha, Using artificial intelligence to stratify normal versus abnormal chest x-rays: external validation of a deep learning algorithm at east kent hospitals university NHS foundation trust, Diagnostics 13 (22) (2023) 3408, https://doi.org/10.3390/diagnostics13223408.

[14] G. Irmici, M. Cè, E. Caloro, N. Khenkina, G. Della Pepa, V. Ascenti, et al., Chest X-ray in emergency radiology: what Artificial Intelligence applications are available? Diagnostics 13 (2) (2023) 216. 〈https://www.mdpi.com/2075-4418/13/2/216〉 (Available from).

[15] C. Geric, Z.Z. Qin, C.M. Denkinger, S.V. Kik, B. Marais, A. Anjos, P.M. David, F. Ahmad Khan, A. Trajman, The rise of artificial intelligence reading of chest X-rays for enhanced TB diagnosis and elimination, Int J. Tube Lung Dis. 27 (5) (2023) 367–372, 10.5588/ijtld.22.0687. 37143227; 10171486.

[16] G. Joshi, A. Jain, S.R. Araveeti, S. Adhikari, H. Garg, M. Bhandari, FDA-approved artificial intelligence and machine learning (AI/ML)-enabled medical devices: an updated landscape, Electronics 13 (3) (2024) 498, https://doi.org/10.3390/electronics13030498.

[17] WHO. WHO consolidated guidelines on tuberculosis: module 2: screening: systematic screening for tuberculosis disease. World Health Organization; 2021. Available: 〈https://www.who.int/publications/i/item/9789240022676〉.

[18] J. Singh, S. Al-Abri, E. Petersen, F. Al Yaqoubi, K. Al Rahbi, L. Al Balushi, F. Al Fahdi, A. Al Balushi, F.M.M. Jawad, P.J. Kurup, Importance of tuberculosis screening of resident visa applicants in low TB incidence settings: experience from oman, J. Epidemiol. Glob. Health 12 (3) (2022) 281–291, https://doi.org/10.1007/s44197-022-00040-w.

[19] Diogo P., Morais M., Calisto F.M., Santiago C., Aleluia C., Nascimento J.C. Weakly-Supervised Diagnosis and Detection of Breast Cancer Using Deep Multiple Instance Learning. In: Proceedings of the 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI), Cartagena, Colombia, 18–21 April 2023; pp. 1–4.

[20] Abrantes J., Bento e Silva MJN, Meneses JP, Oliveira C, Calisto FMGF, Filice RW. External validation of a deep learning model for breast density classification ECR 2023 / C-16014. 10.26044/ecr2023/C-16014.

[21] Morais M., Calisto F.M., Santiago C., Aleluia C., Nascimento J.C. Classification of Breast Cancer in Mri with Multimodal Fusion. 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI), Cartagena, Colombia, 2023; pp. 1-4. 10.1109/ISBI53787.2023.10230686.

[22] P. Kaviani, S.R. Digumarthy, B.C. Bizzo, B. Reddy, M. Tadepalli, P. Putha, A. Jagirdar, S. Ebrahimian, M.K. Kalra, K.J. Dreyer, Performance of a chest radiography AI algorithm for detection of missed or mislabeled findings: a multicenter study, Diagnostics 12 (9) (2022) 2086, https://doi.org/10.3390/diagnostics12092086.

[23] A. Kumar, P. Patel, D. Robert, S. Kumar, A. Khetani, B. Reddy, A. Srivastava, Accuracy of an artificial intelligence-enabled diagnostic assistance device in recognising normal chest radiographs: a service evaluation, published online ahead of print]:tzae029, BJR Open (2024), https://doi.org/10.1093/bjro/tzae029.

[24] Putha, P., Tadepalli, M., Reddy, B., et al., (2018). Can Artificial Intelligence Reliably Report Chest X-Rays?: Radiologist Validation of an Algorithm trained on 2.3 Million X-Rays, CoRR abs/1807.0. http://arxiv.org/abs/1807.07455.

[25] R. Singh, M.K. Kalra, C. Nitiwarangkul, J.A. Patti, F. Homayounieh, A. Padole, P. Rao, P. Putha, V.V. Muse, A. Sharma, S.R. Digumarthy, Deep learning in chest radiography: detection of findings and presence of change, PLoS One 13 (10) (2018) e0204155, https://doi.org/10.1371/journal.pone.0204155, 30286097; 6171827.

[26] U.S. Food and Drug Administration (FDA). Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests. [Internet]. Silver Spring (MD): FDA; 2017. Available from: ⟨https://www.fda.gov/files/medical%20devices/published/Guidance-for-Industry-and-FDA-Staff—Statistical-Guidance-on-Reporting-Results-from-Studies-Evaluating-Diagnostic-Tests-%28PDF-Version%29.pdf⟩.

[27] L.L. Plesner, F.C. Müller, J.D. Nybing, L.C. Laustrup, F. Rasmussen, O.W. Nielsen, et al., Autonomous chest radiograph reporting using AI: estimation of clinical impact (Available from:), Radiology 307 (3) (2023) e222268, https://doi.org/10.1148/radiol.222268.

[28] H. Yoo, E.Y. Kim, H. Kim, Y.R. Choi, M.Y. Kim, S.H. Hwang, et al., Artificial Intelligence-based identification of normal chest radiographs: a simulation study in a multicenter health screening cohort, Korean J. Radio. 23 (10) (2022 Oct) 1009–1018.

[29] E.J. Hwang, S. Park, K.N. Jin, J.I. Kim, S.Y. Choi, J.H. Lee, et al., Development and validation of a deep learning-based automated detection algorithm for major thoracic diseases on chest radiographs, JAMA Netw. Open 2 (3) (2019 Mar) e191095.

[30] L. Vogado, F. Araújo, P.S. Neto, J. Almeida, J.M.R.S. Tavares, R. Veras, A ensemble methodology for automatic classification of chest X-rays using deep learning, Comput. Biol. Med 145 (2022 Jun) 105442.

[31] T. Dyer, L. Dillard, M. Harrison, T.N. Morgan, R. Tappouni, Q. Malik, et al., Diagnosis of normal chest radiographs using an autonomous deep-learning algorithm, Clin. Radio. 76 (6) (2021) 473. ⟨https://www.sciencedirect.com/science/article/pii/S0009926021000763⟩. e9-473.e15. Available from.

[32] M. Annarumma, S.J. Withey, R.J. Bakewell, E. Pesce, V. Goh, G. Montana, Automated triaging of adult chest radiographs with deep artificial neural networks (Available from:), Radiology 291 (1) (2019) 196–202, https://doi.org/10.1148/radiol.2018180921.

[33] Y.X. Tang, Y.B. Tang, Y. Peng, K. Yan, M. Bagheri, B.A. Redd, et al., Automated abnormality classification of chest radiographs using deep convolutional neural networks (Available from:), npj Digit Med 3 (1) (2020) 70, https://doi.org/10.1038/s41746-020-0273-z.

[34] L.L. Plesner, F.C. Müller, M.W. Brejnebøl, C.H. Krag, L.C. Laustrup, F. Rasmussen, O.W. Nielsen, M. Boesen, M.B. Andersen, Using AI to identify unremarkable chest radiographs for automatic reporting, Radiology 312 (2) (2024) e240272, https://doi.org/10.1148/radiol.240272.

[35] Z.Z. Qin, M.S. Sander, B. Rai, et al., Using artificial intelligence to read chest radiographs for tuberculosis detection: a multi-site evaluation of the diagnostic accuracy of three deep learning systems, Sci. Rep. 9 (2019) 15000, https://doi.org/10.1038/s41598-019-51503-3.

[36] Automated classification of chest X-rays as normal/abnormal using a high sensitivity deep learning algorithm. Presented at: European Congress of Radiology, Vienna; 2019.

[37] Albastaki U., Al Hashemi H., Seyhoglu S., Mahmoud K., Al Hashmi A., Jerome C.P. Comparative Assessment of Chest X-ray Interpretations by AI Model and Radiologist Vs Pulmonologist in Predicting the Clinical Status of Covid-19 Pneumonia Patients. J Adv Radiol Med Imaging. Year;7(1). ISSN: 2456-5504.

[38] M.B.K. Sammer, Y.S. Akbari, R.A. Barth, S.L. Blumer, J.R. Dillman, S.G. Farmakis, D.P. Frush, A. Gokli, S.S. Halabi, R. Iyer, A. Joshi, J.K. Kwon, H.J. Otero, A.C. Sher, S.T. Sotardi, B.H. Taragin, A.J. Towbin, C. Wald, Use of artificial intelligence in radiology: impact on pediatric patients, a white paper from the ACR pediatric ai workgroup, J. Am. Coll. Radio. 20 (8) (2023) 730–737, https://doi.org/10.1016/j.jacr.2023.06.003, 37498259. 15187.02084.

[39] R. Najjar, Redefining radiology: a review of artificial intelligence integration in medical imaging, Diagnostics 13 (17) (2023) 2760, https://doi.org/10.3390/diagnostics13172760.