

Original Article

Identifying malignant nodules on chest X-rays: A validation study of radiologist versus artificial intelligence diagnostic accuracy

Bassam Mahboub¹, Manoj Tadepalli², Tarun Raj², Rajalakshmi Santhanakrishnan², Mahmood Yaseen Hachim³, Usama Bastaki⁴, Rifat Hamoudi¹, Ehsan Haider⁵, Abdullah Alabousi⁵

¹Clinical Sciences Department, College of Medicine, University of Sharjah, Sharjah, United Arab Emirates, ²Artificial Intelligence, Qure.ai, Mumbai, Maharashtra, India, ³Molecular Medicine, MBRU, Dubai, United Arab Emirates, ⁴Radiology, Dubai Health Authority, Dubai, United Arab Emirates, ⁵Radiology, McMaster University, Hamilton, ON, Canada

ABSTRACT

Background: Three and half million anonymous X-rays were gathered from 45 locations worldwide (in-hospital and outpatient settings). qXR was initially trained on this massive dataset. We used an independent dataset of 13,426 chest X-rays from radiologists' reports. The test data set included 213,459 X-rays chosen at random from a pool of 3.5 million X-rays. The dataset (development) was developed using the remaining X-rays received from the remaining patients.

Methods: qXR is a deep learning algorithm-enabled software that is used to study nodules and malignant nodules on X-rays. We observed moderate to a substantial agreement even when observations were made with normal X-rays.

Results: qXR presented a high area under the curve (AUC) of 0.99 with a 95% confidence interval calculated with the Clopper–Pearson method. The specificity obtained with qXR was 0.90, and the sensitivity was 1 at the operating threshold. The sensitivity value of qXR in detecting nodules was 0.99, and the specificity ranged from 0.87 to 0.92, with AUC ranging between 0.98 and 0.99. The malignant nodules were detected with a sensitivity ranging from 0.95 to 1.00, specificity between 0.96 and 0.99, and AUC from 0.99 to 1. The sensitivity of radiologists 1 and 2 was between 0.74 and 0.76, with a specificity ranging from 0.98 to 0.99. In detecting the malignant nodules, specificity ranged between 0.98 and 0.99, and sensitivity fell between 0.88 and 0.94.

Conclusion: Machine learning model can be used as a passive tool to find incidental cases of lung cancer or as a triaging tool, which accelerate the patient journey through standard care pipeline for lung cancer.

Keywords: Artificial intelligence, chest X-rays, convolutional neural network, deep learning, malignant nodules

BACKGROUND

Deep learning (DL) algorithms help us detect various lung abnormalities in chest X-rays [1]. DL algorithms were used to detect diabetic retinopathy by using photographs of the retinal fundus [2]. While investigating skin cancer, the fine-grained object categories were analyzed using deep convolutional neural networks (CNNs) [3]. The availability of original and viable datasets representing various lung abnormalities is

seldom available [4]. DL-based algorithms outperformed experienced physicians in determining several thoracic diseases [5]. Although larger datasets are available, most of the X-rays were certified normal, limiting us in feeding the machine tools with accurate training data. CheXpert is one of the largest chest radiograph datasets that can provide uncertainty labels and supply comparisons between expert

Submitted: 08-03-2022 Revised: 16-04-2022
Accepted: 19-04-2022 Published: 27-07-2022

Address for correspondence: Dr. Bassam Mahboub,
Clinical Sciences Department, College of Medicine, University of Sharjah, Sharjah,
United Arab Emirates.
E-mail: drbmahboub@gmail.com

Access this article online	
Website: www.abhsjournal.net	Quick Response Code 
DOI: 10.4103/abhs.abhs_17_22	

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

How to cite this article: Mahboub B, Tadepalli M, Raj T, Santhanakrishnan R, Hachim MY, Bastaki U, *et al.* Identifying malignant nodules on chest X-rays: A validation study of radiologist versus artificial intelligence diagnostic accuracy. *Adv Biomed Health Sci* 2022;1:137-43.

judgments [6]. It has been a difficult task to collect X-rays with less defined edges, irregular margins, and orders that appear circumscribed from reliable sources. In most lung cancer patients, solitary lung nodules can appear on the X-rays, which can be considered as *prima facie* finding of the possible lung cancer in at least 20% of cases [7]. This is critical in scaling up the investigation of the possible lung abnormalities. Although several tools are in place, non-contrast computed tomography (CT) is still a reliable technique to identify pulmonary nodules [8]. Automated detection of intracranial hemorrhage using head CT scans was attempted using DL algorithms [9]. CheXNet was used to detect pneumonia from chest X-rays using DL [10]. Stage I cancers can be diagnosed with chest X-rays in the initial screening process [11]. In all, 44% of the central lung cancers and 98% of peripheral lung cancers can appear as nodules [12]. These nodules might be associated with or without adenopathy. In the related studies, it was suggested that 65% of the lung cancers appeared as nodules [13]. There are always instances that ignore nodules that would otherwise need to investigate the lung cancers. This difficulty or oversight can be minimized with the use of machine algorithms. Although not a sole tool can present accurate results, these algorithms can aid in investigating lung abnormalities. Several computer-aided diagnostic methods can be used in lung cancer to be investigated on the X-rays [14]. Several studies evaluated DL algorithms to gauge nodules on the test dataset that contained X-rays with normal and abnormal lung conditions. qXR v2.0 was used to analyze chest X-rays of tuberculosis patients [15]. qXR is a software application that can utilize DL algorithms to detect nodules on chest X-rays. This study evaluates the performance of qXR in detecting malignant nodules. In addition, some of the abnormalities such as pleural effusion, opacity, and consolidation were also investigated using this software. The model score was established for each X-ray, and the performance was evaluated accordingly. Through this study, training and validation of qXR to detect nodules on chest X-rays are explained.

MATERIALS AND METHODS

Dataset

About 3.5 million anonymous X-rays were collected from 45 centers (in-hospital and outpatient settings) spread worldwide. qXR was trained with this huge dataset initially. An independent dataset comprising 13,426 chest X-rays based on the reports generated by radiologists was used. These X-rays were acquired in posterior–anterior (PA), anterior–posterior (AP), supine, or lateral views. The lateral chest X-rays were not available; we made all efforts to get them but could not. Although lateral X-rays can be included if they are available, it is very hard to find the corresponding lateral X-rays for most

AP/PA X-rays taken in general practice, and even harder to find them in cases of suspected nodules. The biopsy report should be used as the final ground truth when available. Radiologist interpretation has been used due to the lack of biopsy reports in these cases. The diversity in the dataset is based on the quality of X-rays, resolution, and size distribution. The test data comprised 213,459 X-rays that were randomly selected from a pool of 3.5 million X-rays used. The development dataset was framed from the remaining X-rays collected from the remaining patients and was used to develop the algorithm. In all, 500 nodules were randomly selected from the nodule-positive subset and 500 from the nodule-negative subset (determined by the original radiologist report). No signs or symptoms or clinical data were considered when deciding the inclusion/exclusion of a case from the study.

Validation and training datasets were extracted from these X-rays. In this study, chest X-rays from patients younger than 15 years and those exposed in the lateral view were excluded from test and development datasets. The development dataset and test dataset were treated as independent in the development of the algorithm. The presence/absence of the nodules as reported by radiologists was extracted using natural language processing techniques [16]. At least one nodule was present in the test data comprising 10,200 X-rays, which was confirmed by radiologists. Randomized selections were made to choose 894 scans. From this collection, X-rays that exhibited nodules were classified as positive test sets, and those devoid of abnormalities and nodules were classified as negative test sets. Ground truth was established based on the annotated reports of three radiologists.

The radiologists classified these scans based on the presence/absence of nodules and if they were malignant or benign. The ground truth was confirmed based on the consensus of the opinions among radiologists. The performance of qXR with ground truth was compared with the reads of two radiologists. The accuracy of qXR was evaluated using positive predictive value (PPV), negative predictive value (NPV), specificity, sensitivity, and area under the curve (AUC) [17]. Cohen's kappa was used to assess the variability between a pair of radiologists [18]. The consistency of the agreement in the radiologist's group was tested using Fleiss' kappa [19]. The inter-rated agreement was used to determine the accuracy in interpreting chest X-rays of tuberculosis patients [20] and pneumonia patients [21].

Ground truth

In the primary stage of this research, three radiologists with 10 years of experience were recruited. Their respective identities were masked to ensure independent and unbiased

reporting of the results. The radiological and clinical information of the anonymous patients was intentionally denied to the radiologists. Each radiologist was given the yes/no option to mark the X-rays for the presence of nodules. If nodules were present, the radiologists were asked to mark X-rays as either malignant or benign. Knowledge about histopathology of the nodules was not studied, and this study purely relied on the radiologists' opinions. Radiologists employed their unbiased judgment based on the internal characteristics, calcification, contour, margin, and size distribution. In the presence of multiple nodules, and if at least one was malignant, the radiologists read it as malignant [22]. In case of discrepancy raised with reports, the majority of the three reports were classified as ground truth. The selection of the cases was entirely made using the radiological report and was blinded to the clinical data/symptoms of the patient.

Algorithm

DL was used to train the CNNs to detect nodules on the X-rays [23]. Specific architectures that constitute the basic blocks of this system are versions of residual networks aided with squeeze–excitation modules [24]. We used modified vanilla versions of these architectures to process information with appropriate resolution. The quality, resolution, size distribution, and diversity were observed in the selected development dataset before feeding it into CNNs. X-rays were subjected to downsampling and image normalization to minimize source-dependent variation. The classification networks that built nodule detection systems were pre-trained to demarcate chest X-rays from others [25]. Super-set was utilized in this process, and it constituted all the X-rays. The network output score ranged between 0 and 1, which indicated nodule occurrence [26]. A pixel map was produced specifying the location of the nodule [27]. An algorithm was developed to automatically detect malignancy using nodules on X-rays [28]. To train the algorithm, 3000 patches isolated from nodule-bearing X-rays were used. Every patch was extracted by manually demarcating the nodule with a bounding box and resized appropriately. Radiologists gave labels stating malignant/benign nodules. Each chest X-ray was passed through a nodule detecting algorithm. A pixel map was produced according to a specific lung area. The obtained pixel map was cropped, and the resultant patch could pass onto the malignancy detection algorithm. This algorithm assigned a malignancy score to every patch. Each X-ray got a malignancy score after aggregating the scores of all the nodules present. Multiple networks of nodule and malignancy detection algorithms were trained in this process. These networks were classified based on the dataset (training) distribution, model initialization conditions, and type of architecture. A major ensembling scheme and a subset of these models were selected based on heuristics. This

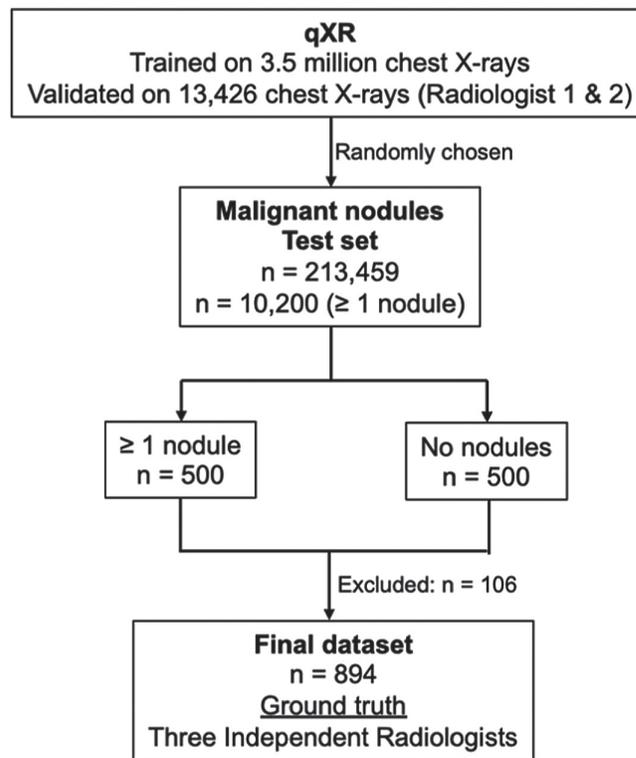


Figure 1: Datasets and algorithm training.

subset was used to integrate prediction and to produce final decisions utilizing each algorithm. The schema of datasets and algorithm training is represented in Figure 1.

Statistical analysis

qXR was evaluated opposite ground truth of detecting nodules and even malignant nodules by utilizing area under the receiver operating characteristic (ROC) curve at 95% confidence interval (CI) [29]. We have used PPV, NPV, specificity, and sensitivity in this process. Beta distribution was examined using the Clopper–Pearson method to calculate CI at 95% [30]. This analysis was done using Python with packages such as NumPy, Pandas, PyMongo, and scikit-learn [31]. The data used in this study were stored in the Mongo database [32].

RESULTS

It was observed that 47% of the X-rays (patients of age greater than 30 years), 75% of X-rays have gender information, of which 58% of them are taken from male patients. In all, 14% of the X-rays reflected malignant nodules, whereas 45% exhibited nodules. The clinical characteristics of the test set are shown in Table 1.

qXR detected nodules with a high AUC of 0.99 with a 95% CI. The specificity obtained was 0.90 with a sensitivity of 1

during the operating point [Table 2]. Agreement among the radiologists was moderate to substantial in the detection of nodules. Cohen's κ value was used to describe the agreement between a pair of radiologists, and Fleiss' κ was used to know agreement among all radiologists.

For nodules, Cohen's κ value was 0.37–0.66 with Fleiss' κ value of 0.49. For malignant nodules, Cohen's κ value was 0.59–0.77, with Fleiss' κ value of 0.67 [Table 3]. There was a slight agreement observed in the case of other abnormalities and non-nodular opacities. We observed a moderate to substantial

agreement for even observations made with normal X-rays. The ROC curve of qXR versus radiologists is presented in Figures 2 and 3, respectively. The AUC of malignant nodules was 1, and for nodules, it was 0.99. Owing to the appropriate AUC, we stratified the test set accordingly using the gender and age of the patients at the operating point. Gender was classified as men, women, and other classes, and age was stratified as ≤ 30 , >30 – ≤ 60 , and >60 . We analyzed NPV, PPV, specificity, sensitivity, and accuracy. The performance of qXR remained the same when tested across all the subgroups [Tables 4 and 5].

The sensitivity value of qXR in detecting nodules was 0.99, and the specificity ranged from 0.87 to 0.92, with AUC ranging between 0.98 and 0.99. The malignant nodules were detected with a sensitivity ranging from 0.95 to 1.00, specificity between 0.96 and 0.99, and AUC from 0.99 to 1. The sensitivity of radiologists 1 and 2 was between 0.74 and 0.76, with a specificity ranging from 0.98 to 0.99. In detecting the malignant nodules, specificity ranged between 0.98 and 0.99, and sensitivity fell between 0.88 and 0.94.

Table 1: Clinical characteristics of the validation dataset.

Characteristics	n (894)	%
Age		
≤ 30	74	8.2
>30 and ≤ 60	236	26.4
>60	180	20.1
Not available	404	45.2
Gender		
Women	281	31.4
Men	389	43.5
Other	2	0.2
Not available	222	24.8
Abnormalities*		
Nodule	402	45.0
Malignant nodules	121	13.5
Non-nodular opacities	69	7.7
Others	37	4.1
None	387	43.3

*More than one abnormality was present in certain X-rays

DISCUSSION

qXR is a robust model trained with 3.5 million chest X-rays, and it is being used to screen tuberculosis with high accuracy. This study used the expertise of three radiologists to establish ground truth. Compared with the ground truth, qXR exhibited high accuracy in detecting nodules and even labeled malignant nodules in the lungs. This accuracy was estimated utilizing specificity, sensitivity, and AUC of qXR in the detection of nodules. This model succeeded in detecting malignant nodules in all the subgroups of gender and age.

Table 2: Performance of qXR and radiologists in detecting nodules and malignant nodules versus ground truth.

Metric	Nodule		Malignant nodule	
	qXR (95% CI)	Ground truth	qXR (95% CI)	Ground truth
AUC	0.99 (0.98–0.99)	-	1.00 (0.99–1.00)	-
Sensitivity	1.00 (0.98–1.00)	0.76–0.74	0.99 (0.95–1.00)	0.94–0.88
Specificity	0.90 (0.87–0.92)	0.99–0.98	0.98 (0.96–0.99)	0.97–0.98
NPV	1.00 (0.98–1.00)	0.83–0.82	1.00 (0.99–1.00)	0.99–0.98
PPV	0.89 (0.86–0.92)	0.99–0.96	0.83 (0.76–0.92)	0.83–0.90

AUC: area under the curve, CI: confidence interval, NPV: negative predictive value, PPV: positive predictive value

Table 3: Inter-rater agreement between three independent radiologists (I, II, and III).

Abnormality	Pairs (Cohen's κ)			All (Fleiss' κ)
	I, II	II, III	I, III	
Nodules	0.66	0.43	0.37	0.49
Malignant nodules	0.77	0.63	0.59	0.67
Non-nodular opacities	0.21	0.26	0.17	0.22
Other abnormalities	0.45	0.09	0.12	0.25
None	0.70	0.51	0.46	0.56

Cohen's κ was used to assess agreement between radiologist pairs, and Fleiss' κ was used to assess agreement among all three radiologists

Nodules larger than 2 cm can be translated as malignant lung cancer with 75% accuracy. Moreover, several artificial intelligence (AI) algorithms can process X-rays images in seconds enabling researchers to perform large-scale screening efficiently. qXR is an AI model used to screen patients with

coronavirus, tuberculosis, and several lung diseases. Sim *et al.* studied 150 normal X-rays and 450 abnormal X-rays (cancerous as defined by CT and pathology) [33]. CNN model read these scans. Although the addition of CNN to the radiologist reports enhanced the instrument’s sensitivity to detect malignant nodules, it did not outperform the judgment of only radiologists. In their study, the radiologists’ sensitivity was 0.54–0.84 and carried a false positive rate between 0.1 and 0.3. Finally, Nam *et al.* examined nodules using a large set of normal and abnormal X-rays with deep learning-based automatic detection (DLAD) algorithm with an AUC of 0.92–0.99 [34]. qXR succeeded in attaining the same level of performance as DLAD. qXR can detect calcification of nodules; the final malignancy score assigned by qXR to each nodule depends on the calcification status of that nodule and is inversely proportional (i.e., the more calcified the nodule, the less score qXR assigns for malignancy).

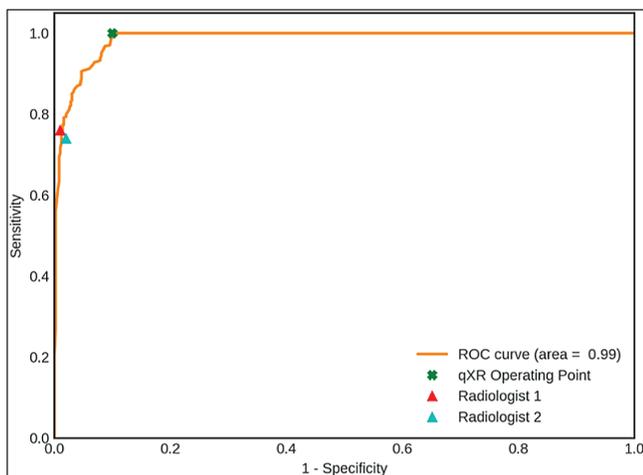


Figure 2: ROC curve to detect nodules.

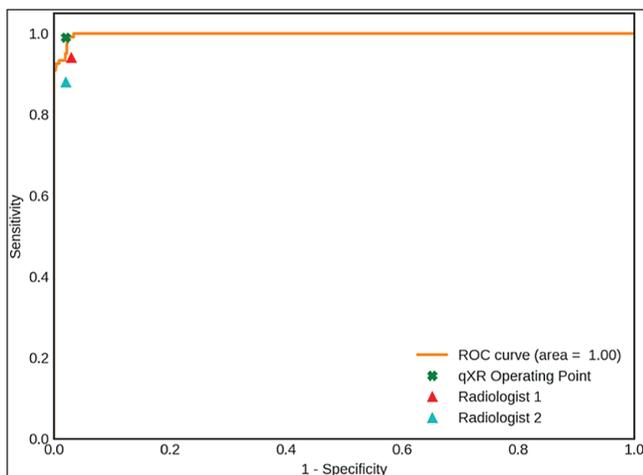


Figure 3: ROC curve to detect malignant nodules.

Justification

We compared the reports generated from three radiologists with reports generated from qXR. In the detection of malignant nodules, radiologists and qXR exhibited similar sensitivity and specificity. In the detection of nodules, qXR presented high sensitivity than radiologists. This study hypothesizes that the high performance exhibited by qXR can be attributed to the size of the nodules as large-sized nodules reflect malignancy. Nodule size can define the solitary pulmonary nodule present in the chest CT scan, that is, <3 cm diameter (if greater than 3 cm, it can be considered mass). The nodules were marked as malignant nodules by the qXR and radiologists due to their larger size. Cautious interpretation is necessary in this case as slight size variation among nodules can slip the judgment and result in false positives/false negatives. In detecting malignant nodules, low sensitivity was observed due to their small size, which remained the same even when the CNN model was used. This study further hypothesized that calcification, contour, and margins could affect the sensitivity. Calcifications can be observed using CT, and low-

Table 4: Performance of qXR in detecting nodules stratified by age and gender.

	AUC	Accuracy	Sensitivity	Specificity	NPV	PPV
Age						
≤30	0.99 (0.97–1.00)	0.96 (0.89–0.99)	1.00 (0.80–1.00)	0.95 (0.86–0.99)	1.00 (0.94–1.00)	0.85 (0.62–0.97)
>30 and ≤60	0.99 (0.98–1.00)	0.97 (0.93–0.99)	1.00 (0.97–1.00)	0.93 (0.87–0.97)	1.00 (0.97–1.00)	0.93 (0.86–0.97)
>60	0.99 (0.97–1.00)	0.94 (0.90–0.97)	0.99 (0.95–1.00)	0.86 (0.76–0.94)	0.98 (0.91–1.00)	0.93 (0.87–0.97)
Not available	0.98 (0.96–0.99)	0.92 (0.89–0.95)	0.99 (0.97–1.00)	0.88 (0.83–0.92)	1.00 (0.97–1.00)	0.84 (0.78–0.89)
Gender						
Women	0.99 (0.975–1.0)	0.95 (0.92–0.97)	0.99 (0.96–1.00)	0.90 (0.84–0.95)	0.99 (0.95–1.00)	0.92 (0.87–0.96)
Men	0.99 (0.97–1.00)	0.95 (0.92–0.97)	0.99 (0.97–1.00)	0.90 (0.85–0.94)	0.99 (0.97–1.00)	0.90 (0.85–0.94)
Other	1.00 (1.00–1.00)	1.00 (0.16–1.00)	1.00 (0.03–1.00)	1.00 (0.03–1.00)	1.00 (0.03–1.00)	1.00 (0.03–1.00)
Not available	0.98 (0.96–1.00)	0.92 (0.88–0.95)	1.00 (0.95–1.00)	0.89 (0.83–0.94)	1.00 (0.97–1.00)	0.80 (0.70–0.88)

AUC: area under the curve, CI: confidence interval, NPV: negative predictive value, PPV: positive predictive value
 Values are displayed with 95% CIs

Table 5: Performance of qXR in detecting malignant nodules stratified by age and gender.

	AUC	Accuracy	Sensitivity	Specificity	NPV	PPV
Age						
≤30	1.00 (0.97–1.00)	0.99 (0.93–1.00)	1.00 (0.59–1.00)	0.99 (0.92–1.00)	1.00 (0.95–1.00)	0.88 (0.47–1.00)
>30 and ≤60	1.00 (0.99–1.00)	0.98 (0.95–0.99)	0.98 (0.90–1.00)	0.97 (0.94–0.99)	0.99 (0.97–1.00)	0.91 (0.81–0.97)
>60	1.00 (0.99–1.00)	0.96 (0.92–0.98)	1.00 (0.91–1.00)	0.94 (0.89–0.98)	1.00 (0.97–1.00)	0.83 (0.69–0.92)
Not available	1.00 (0.98–1.00)	0.98 (0.96–0.99)	1.00 (0.85–1.00)	0.97 (0.95–0.99)	1.00 (0.99–1.00)	0.69 (0.50–0.84)
Gender						
Women	1.00 (0.99–1.00)	0.98 (0.95–0.99)	0.98 (0.90–1.00)	0.97 (0.94–0.99)	1.00 (0.98–1.00)	0.90 (0.80–0.96)
Men	1.00 (0.99–1.00)	0.97 (0.95–0.99)	1.00 (0.94–1.00)	0.97 (0.94–0.98)	1.00 (0.99–1.00)	0.85 (0.74–0.92)
Other	-	-	-	-	-	-
Not available	1.00 (0.96–1.00)	0.97 (0.94–0.99)	1.00 (0.48–1.00)	0.97 (0.94–0.99)	1.00 (0.98–1.00)	0.42 (0.15–0.72)

AUC: area under the curve, CI: confidence interval, NPV: negative predictive value, PPV: positive predictive value

Values are displayed with 95% CIs

kilovolt radiography can allow us to see calcifications present inside the nodules. All these parameters might have affected the judgment of the radiologists. We assume that mass size might have affected the low agreement established while interpreting nodules compared to malignant nodules.

Limitations

There was no clinical context of patients supplied before the radiologists to establish ground truth. Providing clinical history might have affected the reports given by radiologists resulting in the biased ground truth. The room for error was significantly lowered by establishing ground truth, considering the majority of the opinions offered by the radiologists. There was no histology reports or follow-up CT scan to confirm malignancy. At least one-fourth of lung cancer cases can be diagnosed by radiography. Although chest X-rays offer low sensitivity, nodules detected can be investigated, and malignancy was confirmed in most cases. Guidelines recommend not to consider chest X-rays alone to screen for malignancy in the lungs. AI algorithms fed with X-ray images can provide results in few seconds, and hence, there is a possibility of low-cost screening within a limited time frame.

CONCLUSION

We propose using the machine learning model either as a passive tool, monitoring all the X-rays processed at an institution, thereby helping to find incidental cases of lung cancer, or as a triaging tool, which helps accelerate the patient journey through the standard care pipeline for lung cancer. The machine learning model can be embedded in its totality, wherein it takes the complete digital imaging and communications in medicine as the input and produces a lung malignancy risk score, which can be made use of as is appropriate in the given setting. This study used qXR (AI algorithm) to detect the nodules and malignant nodules in the chest X-rays with high specificity, sensitivity, and accuracy. Advanced studies can be commissioned to validate this model by the addition of clinical parameters. These studies

can also reduce costs and reduce the time needed to determine lung cancer patients' malignancy.

Owing to its accuracy, this algorithm can be integrated with other machine learning tools to enable doctors to arrive at appropriate judgments. Furthermore, this technology can be simplified and supplied to countries with a limited technical workforce, saving patients.

Acknowledgements

The authors would like to thank the management of Qure.ai for providing funds and technical infrastructure for this study.

Authors' contributions

All authors had significant and equal contributions to the conceptualization, data collection and analysis and the write up of the first and final draft. All authors are responsible for the scientific integrity of the manuscript.

Ethical statement

This study was accepted by the ethical committee and review board of the institution where this research is done.

Declaration of patient consent

The patients were informed and they accepted to participate in this research with no restrictions.

Financial support and sponsorship

This work is funded by Qure.ai.

Conflict of interest

There are no conflicts of interest.

Data availability statement

The data set used in the current study is available (tick the appropriate option and fill the information).

Repository name—None.

Name of the public domain resources—None.

Data availability within the article or its supplementary materials

Data will be shared upon request available on request from Dr. Bassam (drbmahboub@gmail.com). Dataset can be made available after the embargo period due to commercial restrictions.

REFERENCES

- Tang YX, Tang YB, Peng Y, Yan K, Bagheri M, Redd BA, *et al.* Automated abnormality classification of chest radiographs using deep convolutional neural networks. *NPJ Digit Med* 2020;3:70.
- Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, *et al.* Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016;316:2402-10.
- Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, *et al.* Corrigendum: Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;546:686.
- Bustos A, Pertusa A, Salinas JM, de la Iglesia-Vayá M. PadChest: A large chest x-ray image dataset with multi-label annotated reports. *Med Image Anal* 2020;66:101797.
- Hwang EJ, Park S, Jin KN, Kim JI, Choi SY, Lee JH, *et al.*; DLAD Development and Evaluation Group. Development and validation of a deep learning-based automated detection algorithm for major thoracic diseases on chest radiographs. *JAMA Netw Open* 2019;2:e191095.
- Irvin J, Rajpurkar P, Ko M, Yu Y, Ciurea-Illcus S, Chute C, *et al.* CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. Proceedings of the AAAI Conference on Artificial Intelligence. 2019;33:590-7.
- Erasmus JJ, Connolly JE, McAdams HP, Roggli VL. Solitary pulmonary nodules: Part I. Morphologic evaluation for differentiation of benign and malignant lesions. *Radiographics* 2000;20:43-58.
- Bhalla AS, Das A, Naranje P, Irodi A, Raj V, Goyal A. Imaging protocols for CT chest: A recommendation. *Indian J Radiol Imaging* 2019;29:236-46.
- Chilamkurthy S, Ghosh R, Tanamala S, Biviji M, Campeau NG, Venugopal VK, *et al.* Deep learning algorithms for detection of critical findings in head CT scans: A retrospective study. *Lancet* 2018;392:2388-96.
- Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, *et al.* CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. *arXiv:1711.05225 [cs, stat]*; 25 Dec 2017. Available from: <http://arxiv.org/abs/1711.05225>. [Last accessed on 9 Sep 2021].
- Purandare NC, Rangarajan V. Imaging of lung cancer: Implications on staging and management. *Indian J Radiol Imaging* 2015;25:109-20.
- Larici AR, Farchione A, Franchi P, Ciliberto M, Cicchetti G, Calandriello L, *et al.* Lung nodules: Size still matters. *European Respiratory Review* 2017;26:1-10. Available from: <https://err.ersjournals.com/content/26/146/170025>.
- Gould MK, Donington J, Lynch WR, Mazzone PJ, Midthun DE, Naidich DP, *et al.* Evaluation of individuals with pulmonary nodules: When is it lung cancer? *Chest* 2013;143:e935-1205.
- El-Baz A, Beache GM, Gimel'farb G, Suzuki K, Okada K, Elnakib A, *et al.* Computer-aided diagnosis systems for lung cancer: Challenges and methodologies. *Int J Biomed Imaging* 2013;2013:942353.
- Twabi HH, Semphere R, Mukoka M, Chiume L, Nzawa R, Feasey HRA, *et al.* Pattern of abnormalities amongst chest X-rays of adults undergoing computer-assisted digital chest X-ray screening for tuberculosis in Peri-Urban Blantyre, Malawi: A cross-sectional study. *Trop Med Int Health* 2021;26:1427-37.
- Pons E, Braun LM, Hunink MG, Kors JA. Natural language processing in radiology: A systematic review. *Radiology* 2016;279:329-43.
- Bradley SH, Hatton NLF, Aslam R, Bhartia B, Callister ME, Kennedy MP, *et al.* Estimating lung cancer risk from chest X-ray and symptoms: A prospective cohort study. *Br J Gen Pract* 2021;71:e280-6.
- Luo L, Luo X, Chen W, Liang C, Yao S, Huang W, *et al.* Consistency analysis of programmed death-ligand 1 expression between primary and metastatic non-small cell lung cancer: A retrospective study. *J Cancer* 2020;11:974-82.
- Endo C, Nakashima R, Taguchi A, Yahata K, Kawahara E, Shimagaki N, *et al.* Inter-rater agreement of sputum cytology for lung cancer screening in Japan. *Diagn Cytopathol* 2015;43:545-50.
- Sakurada S, Hang NT, Ishizuka N, Toyota E, Hung Le D, Chuc PT, *et al.* Inter-rater agreement in the assessment of abnormal chest X-ray findings for tuberculosis between two Asian countries. *BMC Infect Dis* 2012;12:31.
- Hopstaken RM, Witbraad T, van Engelsehoven JM, Dinant GJ. Inter-observer variation in the interpretation of chest radiographs for pneumonia in community-acquired lower respiratory tract infections. *Clin Radiol* 2004;59:743-52.
- Schultheiss M, Schmette P, Bodden J, Aichele J, Müller-Leisse C, Gassert FG, *et al.* Lung nodule detection in chest X-rays using synthetic ground-truth data comparing CNN-based diagnosis to human performance. *Sci Rep* 2021;11:15857.
- Schultheiss M, Schober SA, Lodde M, Bodden J, Aichele J, Müller-Leisse C, *et al.* A robust convolutional neural network for lung nodule detection in the presence of foreign bodies. *Sci Rep* 2020;10:12987.
- Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-excitation networks. *IEEE Trans Pattern Anal Mach Intell* 2020;42:2011-23.
- Ibrahim DM, Elshennawy NM, Sarhan AM. Deep-chest: Multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases. *Comput Biol Med* 2021;132:104348.
- Farhat H, Sakr GE, Kilany R. Deep learning applications in pulmonary medical imaging: Recent updates and insights on COVID-19. *Mach Vis Appl* 2020;31:53.
- Hollings N, Shaw P. Diagnostic imaging of lung cancer. *Eur Respir J* 2002;19:722-42.
- Del Ciello A, Franchi P, Contegiacomo A, Cicchetti G, Bonomo L, Larici AR. Missed lung cancer: When, where, and why? *Diagn Interv Radiol* 2017;23:118-26.
- Nash M, Kadavigere R, Andrade J, Sukumar CA, Chawla K, Shenoy VP, *et al.* Deep learning, computer-aided radiography reading for tuberculosis: A diagnostic accuracy study from a tertiary hospital in India. *Sci Rep* 2020;10:210.
- Jabbour SK, Lee KH, Frost N, Breder V, Kowalski DM, Pollock T, *et al.* Pembrolizumab plus concurrent chemoradiation therapy in patients with unresectable, locally advanced, stage III non-small cell lung cancer: The phase 2 KEYNOTE-799 nonrandomized trial. *JAMA Oncol* 2021;7:1351-9.
- Mathur P. Key technological advancements in healthcare. In: Mathur P, editor. *Machine Learning Applications Using Python: Cases Studies from Healthcare, Retail, and Finance*. Berkeley, California: Apress; 2019:13-35. Available from: https://doi.org/10.1007/978-1-4842-3787-8_2. [Last accessed on 9 Sep 2021].
- Ferreira Junior JR, Oliveira MC, de Azevedo-Marques PM. Cloud-based NoSQL open database of pulmonary nodules for computer-aided lung cancer diagnosis and reproducible research. *J Digit Imaging* 2016;29:716-29.
- Sim Y, Chung MJ, Kotter E, Yune S, Kim M, Do S, *et al.* Deep convolutional neural network-based software improves radiologist detection of malignant lung nodules on chest radiographs. *Radiology* 2020;294:199-209.
- Chilamkurthy S, Ghosh R, Tanamala S, Biviji M, Campeau NG, Venugopal VK, *et al.* Deep learning algorithms for detection of critical findings in head CT scans: A retrospective study. *Lancet* 2018;392:2388-96.