The utility of artificial intelligence in identifying radiological evidence of lung cancer and pulmonary tuberculosis in a high-burden tuberculosis setting

Z Z Nxumalo,[1] MBChB; **E M Irusen,**[1] FCP (SA), PhD; **B W Allwood,**[1] FCP (SA), PhD;
M Tadepalli,[2] BTech CS; **J Bassi,**[2] BE (Hons) Chemical; **C F N Koegelenberg,**[1] FCP (SA), PhD

[1] *Division of Pulmonology, Department of Medicine, Faculty of Medicine and Health Sciences, Stellenbosch University and Tygerberg Hospital, Cape Town, South Africa*
[2] *Qure.ai, Mumbai, India*

Corresponding author: *Z Z Nxumalo (durbantrees@gmail.com)*

Background. Artificial intelligence (AI), using deep learning (DL) systems, can be utilised to detect radiological changes of various pulmonary diseases. Settings with a high burden of tuberculosis (TB) and people living with HIV can potentially benefit from the use of AI to augment resource-constrained healthcare systems.
Objective. To assess the utility of qXR software (AI) in detecting radiological changes compatible with lung cancer or pulmonary TB (PTB).
Methods. We performed an observational study in a tertiary institution that serves a population with a high burden of lung cancer and PTB. In total, 382 chest radiographs that had a confirmed diagnosis were assessed: 127 with lung cancer, 144 with PTB and 111 normal. These chest radiographs were de-identified and randomly uploaded by a blinded investigator into qXR software. The output was generated as probability scores from predefined threshold values.
Results. The overall sensitivity of the qXR in detecting lung cancer was 84% (95% confidence interval (CI) 80 - 87%), specificity 91% (95% CI 84 - 96%) and positive predictive value of 97% (95% CI 95 - 99%). For PTB, it had a sensitivity of 90% (95% CI 87 - 93%) and specificity of 79% (95% CI 73 - 84%) and negative predictive value of 85% (95% CI 79 - 91%).
Conclusion. The qXR software was sensitive and specific in categorising chest radiographs as consistent with lung cancer or TB, and can potentially aid in the earlier detection and management of these diseases.

S Afr Med J 2024;114(6):e1846. https://doi.org/10.7196/SAMJ.2024.v114i6.1846

Medical imaging, particularly chest radiography, is well positioned for the application and adoption of artificial intelligence (AI) using deep learning (DL) systems. Radiological AI algorithms are now well established and serve several narrow image-analysis functions to aid clinicians and radiologists with the quantification, triage and enhancement of images.[1,2] There is increasing evidence for the use of AI in pulmonary medicine, and numerous studies have suggested that it has the potential to expedite and improve the interpretation of chest radiographs.[1,2]

Sub-Saharan Africa, and particularly southern Africa, falls victim to intersections of multiple colliding epidemics, namely tobacco smoking, pulmonary tuberculosis (PTB) and HIV.[3] South Africa (SA) has the world's largest burden of people living with HIV and one of the world's highest incidences of PTB. Moreover, lung cancer remains the most common cause of cancer-related death in the country.[4] The estimated age-standardised incidence of lung cancer in SA is 18.3/100 000.[5]

Plain chest radiography is available at practically all levels of healthcare in SA, yet the vast majority of chest radiographs performed on individuals who do not have access to medical funders are not reported by a specialist radiologist.

Given the paucity of data of the use of AI in a population with both a high burden of PTB and HIV as well as smoking-related diseases, we aimed to assess the utility of AI in detecting radiological changes compatible with lung cancer or PTB.

Methods

Study design and participants

We conducted an observational study at Tygerberg Hospital in Cape Town, SA. The institution is a 1 380-bed referral hospital that provides tertiary service to a population of approximately 4 million people with an incidence of PTB of >500/100 000 and incidence of lung cancer of ~18/100 000.[5,6]

Chest radiographs

Chest radiographs of participants with known diagnoses who were previously recruited for prospective registries were exported in a digital imaging and communications in medicine (DICOM) format, and de-identified. The registries included a registry containing normal chest radiographs (as reported by a thoracic radiologist) that were collected for another study on AI, radiographs from patients with histologically confirmed lung cancer (a lung cancer registry) and microbiologically confirmed PTB (from a registry of known cases). Of note is that no clinical information or laboratory data were collected.

A blinded investigator uploaded the deidentified DICOM files in a random order to the Qure.ai website (Mumbai, India) for AI reporting.

qXR usage

qXR software (Qure.ai, Mumbai, India) was used to generate a threshold score; radiographs are scored between 0 and 1 for particular abnormality. qXR is trained using deep learning wherein the neural networks are trained to infer binary outcomes using a positive and negative label. Multiple neural network algorithms were combined for each finding to generate the final probability score. The threshold scores are tested on global datasets, and configurations have been built for multiple-use cases that are optimised for sensitivity and/or specificity. The output that was generated was probability scores of an identified suspicious lesion on the chest radiograph of interest, being malignant or illustrating features in keeping with PTB. Numerical thresholds for each condition being true or false were noted as follows: mass or specific pulmonary nodule cut-off of 0.2, PTB of 0.7 and for any pulmonary abnormality (nonspecific) of 0.84. If the chest radiograph was compatible with the pathology of interest, the score was higher than threshold, hence true, and the converse false.

Statistical analysis

In order to estimate an expected sensitivity of 90% with a maximum allowed lower limit for sensitivity for the 95% confidence interval (CI) as 0.85 with a minimum of 80%, it was calculated that a minimum of 384 chest radiographs containing confirmed abnormalities at a 5% level of significance.[7,8] Analysis was conducted using statistical software for social science (SPSS) version 27 (IBM, USA). Sensitivity, specificity, positive predictive value and negative predictive value along with their 95% CIs were used as primary metrics to evaluate the performance of qXR against the gold standard. For both diagnoses (lung cancer and PTB), the same reading could be judged as true positive (TP) for lung cancer, but also false positive (FP) for TB or vice versa if both cut-off values were deemed positive.

Ethical approval

The capturing and de-identification of data were approved by the Stellenbosch University Health Research Ethics Committee (Ethics Reference No: S22/03/039).

Results

A total sample size of 382 CXRs were used, of which 127 were lung cancer (Fig. 1), 144 PTB (Fig. 2) and 111 normal controls (Fig. 3). The performance of the qXR software for the effective detection of lung cancer and PTB is summarised in Table 1. Of note is the fact that the overall sensitivity of qXR in identifying lung cancer was 84% (95% CI 80 - 87), specificity 91% (95% CI 84 - 97) and positive predictive value of 97% (95% CI 95 - 99). For PTB, it had a sensitivity of 90% (95% CI 87 - 93) and specificity of 79% (95% CI 73 - 84) and negative predictive value of 85% (95% CI 79 - 91).

Discussion

The overall sensitivity of qXR in identifying lung cancer was 84%, with a specificity of 91% and positive predictive value of 97%, and for PTB, it had a sensitivity of 90% and specificity of 79%, and a negative predictive value of 85%, making it highly efficient in categorising chest radiographs as abnormal and consistent with cases of confirmed lung cancer or TB.

In a healthcare system that faces the collision of two epidemics (PTB and smoking-related diseases), a highly sensitive screening test would result in the earlier detection of pathology and prompt referrals to facilities where patients can be further adequately managed.

There has been promising but limited research done locally, highlighting the need for computer-aided detection (CAD) in conditions such as PTB. Melendez *et al.*,[9] in a single-centre study using a similar sample size of 330 chest radiographs, assessed four CAD software programs' ability to accurately differentiate between PTB, silicosis and normal images. Three out of the four CAD softwares had sensitivities surpassing 90%, which was in agreement with the results of TB detection in our study.[9] Once again, similar results were reported in a multicentre retrospective study evaluating five commercial AI products, yielding sensitivities of >90% in assessing the ability of AI to detect PTB on chest radiographs in

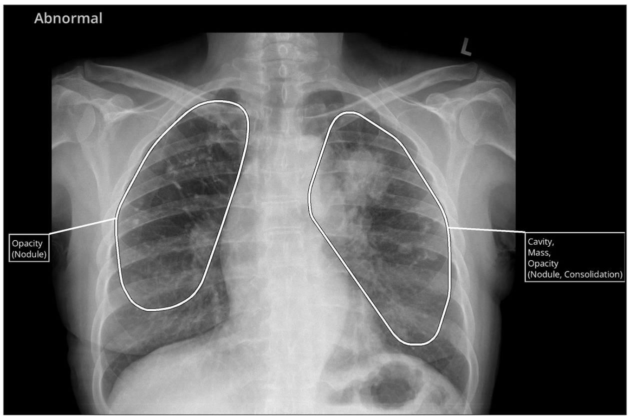

Fig. 1. An example of a chest radiograph of a patient with lung cancer correctly identified as abnormal and lung cancer (true positive).

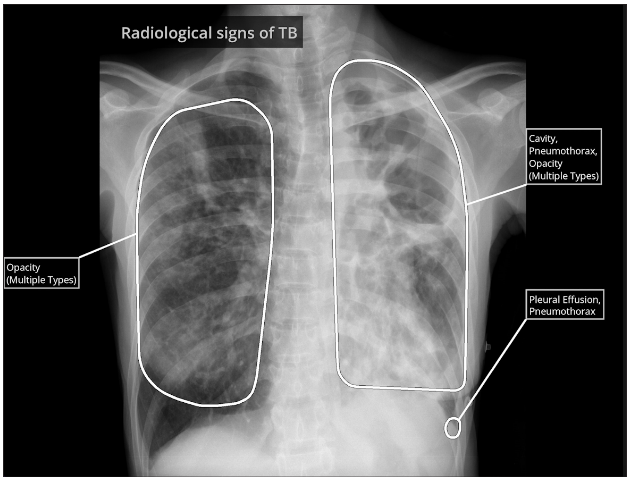

Fig. 2. An example of a chest radiograph of a patient with confirmed pulmonary tuberculosis correctly identified on the left (true positive), but the reading also suggested a lung mass on the right (false positive).

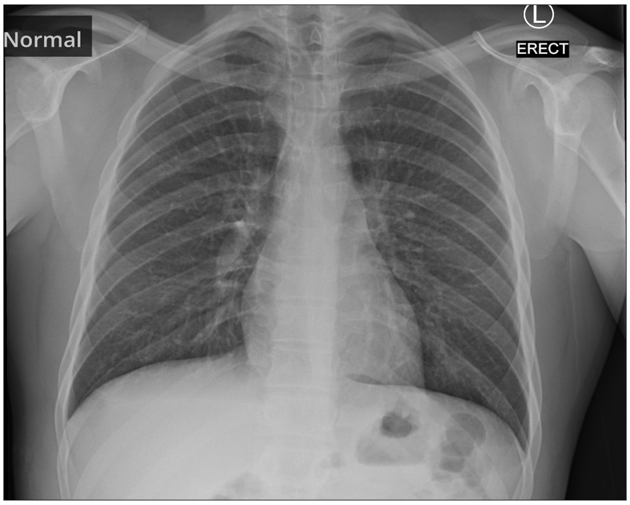

Fig. 3. An example of a normal chest radiograph correctly identified as normal (true negative).

Table 1. Diagnostic efficiency (with 95% confidence intervals) of qXR for lung cancer and PTB (*N*=382)				
Condition	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Lung cancer	84% (80 - 87)	91% (85 - 97)	97% (95 - 99)	62% (54 - 71)
PTB	90% (87 - 93)	79% (73 - 84)	86% (82 - 91)	85% (79 - 91)

PTB = pulmonary tuberculosis; PPV = positive predictive value; NPV = negative predictive value.

high-burden regions; this further highlights the potential of AI to be pragmatically utilised as a tool to rule out disease.[10] The qXR showed non-inferiority to the World Health Organization present recommendation for a PTB triage test, which scores above 0.90 for sensitivity and 0.70 for specificity.[11]

Sub-Saharan Africa shows a lower incidence of lung cancer but a higher mortality burden; proposed reasons for this include the late detection of lung cancer in the absence of official lung-screening programmes.[12,13] Low-dose computed tomography (LDCT) as a screening tool has been shown in a large, randomised control trial to reduce mortality in high-risk populations; however, due to resource constraints this may not be a feasible option in our setting. Even though AI detection for lung cancer was reported as lower than that of LDCT, it did, however, show similar sensitivities to those of reporting radiologists.[14,15] This emphasises the benefit that AI has in low-resource settings and environments that do not have personnel with the skillset to interpret chest radiographs. In these settings, chest radiograph facilities are already in use, and the integration of AI in already established workflow systems can aid in potential screening and detection of pulmonary nodules, which are known to be precursors of lung cancer.

AI does not replace clinicians, but is beneficial in facilitating the diagnostic processes; deep learning algorithms trained on a larger scale can detect multiple abnormalities close to a radiologist level of accuracy.[14,15] A consistent finding comparing radiologists to AI in the detection of lung nodules on CXRs is that AI has a higher sensitivity, while radiologists show higher specificities; this signifies that AI can identify more cases of nodules and have more false positives than radiologists.[8,16]

Radiologists having higher specificities could be accounted for by their abilities to interpret specific findings such as lines, tubes, pneumothorax, fibrosis and masses. With AI assistance, the mean performance of radiologists improved compared with unaided assistance.[17] This demonstrates the value of using AI to assist clinicians in rapid diagnosis by streamlining screening tools with AI. Once software is deployed and implemented in healthcare facilities, one needs to consider that the performance over time can diminish, as pathology often evolves when the disease prevalence changes. These models should be continuously checked and upgraded; it is unclear what the frequency will entail when used in real time.[2]

There is a paucity of research in low-resource countries such as SA evaluating the diagnostic utility of AI in settings of a high burden of PTB and lung cancer; this study is the first of its kind particularly focusing on nodule detection. Our results add to a growing interest and body of evidence that aligns to current literature.

The major strength of our study is that the gold standard diagnoses of lung cancer or PTB were confirmed by appropriate means, and not a radiological diagnosis. Another strength was that investigators were blinded to the gold standard diagnosis and all chest radiographs were de-identified; this ensured that observer bias was minimised.

Limitations which we encountered were that the dataset was from a single site, and the HIV status of all patients was not known; this may limit generalisability of the results. We also did not compare this AI algorithm to others, as this was not the aim of the study.

Conclusion

In conclusion, qXR software was sensitive and specific in categorising chest radiographs as consistent with lung cancer or TB, and can potentially aid in the earlier detection and management of these diseases. Resource-constrained healthcare systems can potentially benefit from the roll-out of this technology.

Declaration. This study was completed as part of ZZN's MMed (Int).

Acknowledgements. We would like to acknowledge Mr Daniel Mashishi from the Division of Epidemiology and Biostatistics at Stellenbosch University for his contribution to the statistical analysis.

Author contributions. ZZN and CFNK contributed to the design and data analysis. ZZN collected the chest radiographs. MT and JB performed the qXR analysis. ZZN and CFNK composed the first draft of the manuscript, which was edited all co-authors.

Funding. AstraZeneca (SA) funded the use of qXR, but had no input into the study design, data collection, analysis, or preparation of the manuscript.

Conflicts of interest. MT and JB are employed by Qure.ai, Mumbai, India. The other authors have no conflicts of interest to declare.

1. Nam JG, Hwang EJ, Kim J, et al. AI improves nodule detection on chest radiographs in a health screening population: A randomised controlled trial. Radiology 2023;307(2):e221894. https://doi.org/10.1148/radiol.221894
2. Rajpurkar P, Lungren MP. The current and future state of AI interpretation of medical images. N Engl J Med 2023;388(21):1981-1990. https://doi.org/10.1056/NEJMra2301725
3. Van Zyl Smit RN, Pai M, et al. Global lung health: The colliding epidemics of tuberculosis, tobacco smoking, HIV and COPD. Eur Respir J 2010;35(1):27-33. https://doi.org/10.1183/09031936.00072909
4. Barta JA, Powell CA, Wisnivesky JP. Global epidemiology of lung cancer. Ann Glob Health 2019;85(1):8. https://doi.org/10.5334/aogh.2419
5. Al Lehebe A, Alomair A, Mahboub B, et al. Recommended approaches for screening and early detection of lung cancer in the Middle East and Africa (MEA) region: A consensus statement. J Thorac Dis 2024;16(3):2142-2158 https://doi: 10.21037/jtd-23-1568
6. World Health Organization.Global tuberculosis report WHO 2022. Geneva: WHO, 2022. https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022/tb-disease-burden/2-1-tb-incidence (accessed 20 February 2023).
7. Putha P, Tadepalli M, Reddy B, et al. Can artificial intelligence reliably report chest X-rays?: Radiologist validation of an algorithm trained on 2.3 million X-rays. arxiv:1807.07455, 2018. https://doi.org/10.48550/arXiv.1807.07455
8. Mahboub B, Tadepalli M, Raj T, et al. Identifying malignant nodules on chest X-rays: A validation study of radiologist versus artificial intelligence diagnostic accuracy. Adv Biomed Health Sci 2022;1:137-143. https://doi.org/10.4103/abhs.abhs_17_22
9. Melendez J, Sánchez CI, Philipsen RH, et al. An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information. Sci Rep 2016;6:25265. https://doi.org/10.1038/srep25265
10. Qin ZZ, Ahmed S, Sarker MS, et al. 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(9):e543-e554. https://doi.org/10.1016/S2589-7500(21)00116-3
11. World Health Organization. High priority target product profiles for new tuberculosis diagnostics: Report of a consensus meeting, 28 -29 April 2014. Geneva: WHO, 2014. https://www.who.int/publications/i/item/WHO-HTM-TB-2014.18 (accessed 15 January 2022).
12. Vos T, Allen C, Arora M, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990 - 2015: A systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016;388(10053):1545-1602. https://doi.org/10.1016/S0140-6736(16)31678-6
13. Tandon YK, Bartholmai BJ, Koo CW. Putting artificial intelligence (AI) on the spot: Machine learning evaluation of pulmonary nodules. J Thorac Dis 2020;12(11):6954-6965. https://doi.org/10.21037/jtd-2019-cptn-03
14. Berle DR, DeMello S, Berg CD, et al; National Lung Screening Trial Research Team. Results of the two incidence screenings in the National Lung Screening Trial. N Engl J Med 2013;369(10):920-931. https://doi.org/10.1056/NEJMoa1208962
15. Yoo H, Kim KH, Singh R, Digumarthy SR, Kalra MK. Validation of a deep learning algorithm for the detection of malignant pulmonary nodules in chest radiographs. JAMA Netw Open 2020;3(9):e2017135. https://doi.org/10.1001/jamanetworkopen.2020.17135
16. Singh R, Kalra MK, Nitiwarangkul C, et al. Deep learning in chest radiography: Detection of findings and presence of change. PLoS ONE 2018;13(10):e0204155. https://doi.org/10.1371/journal.pone.0204155
17. Homayounieh F, Digumarthy S, Ebrahimian S, et al. An artificial intelligence-based chest X-ray model on human nodule detection accuracy from a multicenter study. JAMA Netw Open 2021;4(12):e2141096. https://doi.org/10.1001/jamanetworkopen.2021.41096

Accepted 4 March 2024.