COMPUTER-AIDED DETECTION OR ARTIFICIAL INTELLIGENCE FOR CHEST X-RAY IN EARLY DETECTION OF TUBERCULOSIS

EXECUTIVE SUMMARY

(Adapted from the report by SYFUL AZLIE MD FUZI)

Authors (HTA Experts):
Syful Azlie Md Fuzi
Dr. Aidatul Azura Abdul Rani
Lee Sit Wai
Dr. Roza Sarimin
Dr. Izzuna Mudla Mohamed Ghazali

Information Specialist:
Subhiyah Ariffin

Expert Committee:
Dato' Dr. Mat Zuki Mat Jaeb
Dr. Nurhayati Mohd Marzuki
Dr. Suzana Mohd Hashim
Dr. Zuhanis Abdul Hamid
Dr. Nik Mazlina Mohammad
Dr. Asmah Razali
Dr. Noor Haslinda Ismail
Dr. Fun Weng Hong
Nor Hasnida Mat Desa
Maznah Mohamad
Prof Dr Sanjay Rampal A/L Lekhraj Rampal

External Reviewer:
Dr. Norzaini Rose Mohd Zain
Prof. Dr. Sharifa Ezat Wan Puteh

For further information please contact:

Malaysian Health Technology
Assessment Section (MaHTAS)
Medical Development Division
Ministry of Health Malaysia
Level 4, Block E1, Precinct 1
Government Office Complex
62590 Putrajaya.

Tel: 603 8883 1229

Available at the following website:
http://www.moh.gov.my

2024

Background

For many years, tuberculosis (TB) has been the leading infectious killer until arise of COVID-19. Approximately 10.6 million new TB cases were diagnosed worldwide (2021) whereby 1.6 million people died from TB. The number of TB cases in Malaysia on the other hand has remained high for the past 30 years despite high cure rates that are achievable with a timely diagnosis and an appropriate antibiotic treatment (recorded a total of 25,391 cases in 2022 compared with 21,727 cases reported a year earlier). Malaysia also recorded 2,572 TB deaths in 2022, an increase of 284 deaths (12%) from 2021. While improvements in digital radiography technology have improved the chest radiography (CXR) image quality, lack of access to these facilities and skilled/trained personnel to interpret CXR images continues to be a problem, especially in underdeveloped regions with a high TB prevalence. Advances in artificial intelligence (AI) technology for radiological image analysis, also called computer-aided detection (CAD) have led to the recent development and commercial availability of software that automate the interpretation of CXR for detecting TB. Most CAD products generate radiology reports, as well as a continuous numeric abnormality score to represent the likelihood that TB-related abnormalities are present, along with a heatmap indicating the location of identified abnormalities. Nevertheless, the performance of CAD software in clinical settings is limited by the small number of published studies. As CAD software and software versions are rapidly emerging, so is evidence-based reporting on their accuracy. There is, therefore, a need to evaluate the latest evidence on diagnostic performance of various CAD software for automated interpretation of CXRs for early TB detection to ensure an accurate screening programme, and whether it can be introduced in the management of TB in Malaysia.

This assessment was prepared in response to the request made by the Head of TB/ Leprosy Sector, Disease Control Division, Ministry of Health Malaysia.

Technical Features

The concept of AI, which was based on *deep learning (DL)* and *deep convolutional neural network (DCNNs)* is inspired by the cognitive function of the human mind in terms of learning, problem-solving, and decision making. Recently, CAD software products have been increasingly used to enhance the feasibility and accuracy of CXR interpretation. It uses AI to analyse digital CXR images for abnormalities and provides a quantitative scoring for TB-related abnormalities. Most CAD software express their reading of an abnormality as a numerical score (between 0 - 100 or 0 - 1). A score above a determined threshold can trigger a referral for further diagnostic evaluation for TB. It also generates a second image with a heatmap or bounding boxes, indicating the detected abnormalities' location.

Policy Question

Should CAD or AI for CXR interpretation be used for early TB detection in Malaysia?

Objective:

i. To assess the diagnostic accuracy of CAD or AI for CXR interpretation for early TB detection.
ii. To assess the effectiveness of CAD or AI for CXR interpretation in patient outcomes - facilitate early diagnosis of TB.
iii. To assess the safety of CAD or AI for CXR interpretation for early TB detection - adverse events or complications.
iv. To assess the economic implication, social, ethical, and organisational aspects related to CAD or AI for CXR interpretation for early TB detection.

Research questions:

i. What is the diagnostic performance of CAD or AI for CXR interpretation for early TB detection?
ii. Does screening with CAD or AI for CXR improve TB mortality rate?
iii. Is CAD or AI for CXR interpretation for early TB cost-effective?
iv. What is the social, ethical, and organisational implication/impact related to CAD or AI for CXR interpretation for early TB detection?

Methods

Part A: Systematic Review of Literature

Literature search was developed by the main author and *Information Specialist* who searched for published articles pertaining to CAD or AI for CXR interpretation for early TB detection. The following electronic databases were searched through the Ovid interface: Ovid MEDLINE(R) ALL 1946 to February 2024, EBM Reviews - Health Technology Assessment (4th Quarter 2016), EBM Reviews - Cochrane Database of Systematic Review (2005 to February 2024), EBM Reviews - Cochrane Central Register of Controlled Trials (January 2024), and EBM Reviews - NHS Economic Evaluation Database (4th Quarter 2016). Parallel searches were run in PubMed, US FDA and INAHTA database. Search was limited to articles in English and in human. The detailed search strategy is in **Appendix 2**. The last search was performed on 7th February 2024. Additional articles were identified by reviewing the references of retrieved articles.

Part B: Economic Evaluation

An economic evaluation was conducted to assess the cost-effectiveness and estimate the incremental cost-effectiveness ratio (ICER) of CAD or AI for CXR compared to radiologist interpretation for early TB detection in Malaysia using decision analytic modelling. The data on sensitivity and specificity of both CAD or AI and conventional CXR are human-captured through the published literature. The cost inputs are collected based on the perspectives of MOH, Malaysia.

Results:

Part A: Systematic Review of Literature

A total of 184 records were identified through the Ovid interface and PubMed while 14 were identified from other sources (references of retrieved articles). After removing one duplicate, 197 titles were found to be potentially relevant and abstracts were screened using the inclusion and exclusion criteria. Of these, 34 relevant abstracts were retrieved in full text. After reading, appraising and applying the inclusion and exclusion criteria to the 34 full-text articles, nine were included. The nine full-text articles selected in this review comprised three systematic reviews with/out meta-analysis, one pragmatic randomised trial, three diagnostic studies, and one each for prospective cohort and economic evaluation. All studies included were published in English language between 2016 and 2023 and were mostly conducted in Canada, United Kingdom, Palestine, South Africa, Zambia, Tanzania, Pakistan, Malawi, Bangladesh, and China.

Diagnostic accuracy and effectiveness

From the systematic review, evidence suggested that CAD or AI is a useful tool that can assist in rapid and consistent CXR interpretation for TB. The findings in general demonstrated that CAD could achieve an equivalent diagnostic accuracy or sometimes even superior to experienced, certified physician radiologist readings for detecting bacteriologically (sputum Xpert and/or culture) confirmed TB on CXR, either in a screening use-case (non-facility-based testing) or in a triage use-case (facility-based testing), with the **AUC** around **0.80 to 0.91, sensitivity** ranged from **0.88 to 0.91,** and **specificity** ranged from **0.61 to 0.76**. An updated software versions of CAD significantly improved upon their predecessor's ability to detect TB in terms of **AUC**: CAD4TB v7 **0.903** (95% CI: 0.897, 0.908) versus CAD4TB v6 **0.823** (95% CI: 0.816, 0.830), and qXR v3 **0.906** (95% CI: 0.901, 0.911) versus qXR v2 **0.872** (95% CI: 0.866, 0.878), outperforming human radiologists with specificity of **76.0**% (95% CI: 75.1, 76.9%) and **64.1**% (95% CI: 63.1, 65.2%) for both updated versions, and meeting the standard set in the WHO's Target Product Profile (TPP) for a TB triage test of sensitivity ≥90% and specificity ≥70%. Overall, the AUCs of all CAD software versions were significantly higher in new cases compared to people with a history of TB: ranging from **0.846 to 0.918** for new cases and **0.706 to 0.841** for those who had TB previously. Despite comparable sensitivity, human readers also have poorer performance with this group with specificity of **37.62%** (95% CI: 34.95, 40.34%) compared to **67.2%** (95% CI: 66.1, 68.3%) where there was no TB history. All product versions also performed significantly worse in older populations, as did human readers (AUC **0.889 to 0.931** for young age and **0.762 to 0.858** for old age). In the context of systematic HIV-TB screening, digital CXRs interpreted using CAD software significantly increased the timeliness and completeness of HIV and TB diagnosis and treatment compared to current standard approaches (health worker-directed TB and HIV screening). Median time to TB treatment initiation was shorter (**1-day versus 11-day**; hazard ratio [HR] 2.86, 95% CI: 1.04, 7.87) while HIV screening reduced the proportion with undiagnosed or untreated HIV from **2.7% to 0.2%** (risk ratio [RR] 0.09, 95% CI: 0.01, 0.71, p=0.02).

Table 1 summarises studies related to diagnostic accuracy of CAD or AI for CXR interpretation in early TB detection.

Safety

There was no retrievable evidence of the adverse events or complications related to using CAD or AI for CXR interpretation for early TB detection. However, only one study reported all-cause mortality (not related to CAD) from systematic HIV-TB screening with no significant differences between standard of care 0.7% (3/420), oral HIV testing and linkage to treatment 0.7% (3/450), and oral HIV testing combined with computer-aided digital chest X-ray plus subsequent sputum Xpert confirmation 0.9% (4/450). Currently, several AI algorithms approved by the United States Food and Drug Administration (USFDA) and CAD software was registered as CE-mark (Class IIa) medical device software category that has been validated in clinical studies.

Economic implication

Economic evaluation of CAD or AI for CXR interpretation for early TB detection has been very limited and to date, two cost-effectiveness analyses have been undertaken. The first revealed that oral HIV testing and linkage to care among adults presenting with cough in Malawi were likely to be cost-effective as compared to oral HIV testing in addition to CAD-based CXR plus subsequent sputum Xpert confirmation with an incremental cost-effectiveness ratio (ICER) of USD 901.29 versus USD 4,620.47 per QALY gained. If implemented at scale, these interventions have the potential to rapidly and efficiently improve TB and HIV diagnosis and treatment. Another study suggested that using CAD software to achieve 80% targeted diagnostic coverage in Pakistan will cost an additional investment between USD 2.65 and USD 19.23 million on top of the current infrastructure, depending on the CAD software used. Using human readers, however, would cost an additional USD 23.97 million over the course of four years.

Organisational

There was no retrievable evidence regarding procedural time points and training or learning curves related to CAD or AI for CXR interpretation for early TB detection. In recent years, several independent evaluations of CE-certified commercially available CAD solutions have demonstrated that the accuracy of these products was comparable or sometimes even superior to experienced, certified physician radiologist readings. However, the evaluations did show some variation in the diagnostic accuracy of the CAD products between use-cases and geographies, which suggests that local calibration of the threshold may be required before implementation of CAD. Most CAD products do not have a manufacturer-recommended threshold for triaging individuals needing further confirmatory testing. Also, given the variation across different contexts, the same threshold score does not necessarily provide the same sensitivity and specificity. Therefore, the WHO recommends carrying out a calibration study of the CAD product in the intended use setting according to a defined protocol. Based on the evidence, WHO released new guidelines which stipulate that CAD may be used as an alternative to human reader interpretation of digital CXR for TB screening and triage, but that its use should be limited to interpreting CXR for pulmonary TB in individuals aged 15 years or more.

Social, ethical and legal

No evidence was retrieved on social, ethical and legal issues related to CAD or AI for CXR interpretation for early TB detection. Although WHO has recently endorsed the use of CAD as a screening tool for TB in high-prevalence settings, national TB programmes (NTPs) in several countries might not have clear guidelines in TB diagnostic algorithms for using CAD as an alternative to physician radiologists. Close collaboration with NTPs regarding the choice of CAD, and the setting of threshold scores − preferably based on clinical studies done in the same or similar contexts − should be considered before deciding to procure or implement CAD.

Part B: Economic Evaluation
From the decision analytic modelling, the base-case analysis indicated that the cost per tuberculosis case detected for CXR with human reader was MYR 1,823.32 compared to MYR 1,483.84 for CXR with CAD or AI for tuberculosis screening purpose; yielding an incremental cost-effectiveness ratio (ICER) of MYR 1,293.86 per tuberculosis case detected.

Conclusion:

Part A: Systematic Review of Literature
An overview of the latest evidence on CAD software for automated interpretation of CXR for early TB detection suggested that CAD can achieve an equivalent diagnostic accuracy or sometimes even superior to experienced, certified physician radiologist readings for detecting bacteriologically confirmed TB either in a screening use-case (non-facility-based testing) or in a triage use-case (facility-based testing). Because the abnormality scores produced are continuous, the sensitivity and specificity can vary from 0 to 100%, depending on where the threshold is set. Additionally, an updated software version of CAD significantly improved upon their predecessor's ability to detect TB and meet the standard set in the WHO guideline. Indeed, digital CXR interpreted using CAD software with universal HIV testing increased TB diagnoses, shortened time to TB treatment, and reduced undiagnosed or untreated HIV infection. The biggest advantage of CAD is its superior safety profile with no severe adverse events and mortality directly related to the software. Given the existing evidence, economic evaluations conducted in countries that implemented oral HIV testing and linkage to care were likely to be cost-effective, whereas using digital CXR with computer-aided interpretation for TB plus universal HIV screening was not. By comparison, CAD software could enable large-scale screening programmes in high TB-burden countries and be less costly than radiologists.

Part B: Economic Evaluation
Based on the modelling approach and willingness to pay of 1 to 3 GDP per capita, CAD AI was found to be cost-effective for detecting TB cases through screening the targeted population. However, the limitations of this study needed to be taken into consideration. A sensitivity analysis suggested that the cost, sensitivity of human readers and CAD software might be the major factors that influenced the cost-effectiveness ratio.

Recommendation
Based on the above review, CAD or AI for CXR interpretation can be used to assist in early TB detection among high-risk groups in Malaysia. Product technical evaluations (including technical specification, compatibility, inter-operability, and regulatory requirements) should be done before making procurement decisions.

Table 1: Accuracy information of clinical studies included

First Author/ Year/ Study	Country	CAD software & version	Threshold abnormality score	Comparator	Reference standard	Sensitivity	Specificity	AUC
Zhan Y et al. 2023 SR & MA	China	CAD4TB v1.08 to v7 qXR v2 to v3 Lunit INSIGHT CXR v3.1.0.0 to v4.9.0 InferRead DR v1.0.0.02 to v2 JF CXR-1 v2 to v3.0	17 to 74 0.29 to 44.1 0.076 to 0.92 0.34 to 53.8 0.6 to 83.4	Various reference standard	AFB smear, MTB culture, Xpert MTB/RIF, radiologist readings	Ranged: 0.49 to 1.00 Pooled: 0.91	Ranged: 0.063 to 0.997 Pooled: 0.65	0.91
Harris M et al. 2019 SR	Canada	CAD4TB v1.08 to v5	NR	Various reference standard	AFB smear, MTB/ liquid culture, Xpert MTB/RIF, NAAT	Triage: 0.86 to 1.00 Screening: 0.53 to 0.89	Triage: 0.23 to 0.69 Screening: 0.56 to 0.98	0.75
Pande T et al. 2016 SR	Canada	CAD4TB v1.08 to v3.07	23 to 95	Various reference standard	AFB smear, MTB/ liquid culture, Xpert MTB/RIF	0.47 to 0.91	0.52 to 0.94	0.71 to 0.87
Abuzerr S et al. 2023 Diagnostic	Palistine	CAD4TB v6	>40	Radiologist readings	Gene-Xpert	0.802	0.94	NR
Nishtar T et al. 2022 Diagnostic	Pakistan	CAD4TB, version not reported	≥70	Gene-Xpert	NR	0.832	0.127	NR
Qin ZZ et al. 2022 Diagnostic	Bangladesh			**Radiologist readings performance:**	Xpert MTB/RIF	0.882	0.628	-
		CAD4TB v6	64	CAD4TB v6		0.883	0.641	0.823
			62	WHO TPP*		0.899	0.614	
		CAD4TB v7	58	CAD4TB v7		0.882	0.760	0.903
			50	WHO TPP*		0.899	0.728	
		qQR v2	67	qXR v2		0.882	0.702	0.872
			61	WHO TPP*		0.903	0.668	
		qXR v3	65	qXR v3		0.884	0.766	0.906
			60	WHO TPP*		0.900	0.742	

Abbreviation: CAD, computer-aided detection; SR & MA, systematic review & meta-analysis; AUC, area under the receiver operating curve; NR, not reported; WHO TPP*, WHO's Target Product Profile - fixing sensitivity at 90%.