The Union

Paris, November 15-18

WORLD CONFERENCE ON LUNG HEALTH 2023

TRANSFORMING EVIDENCE INTO PRACTICE

CONFLICT OF INTEREST DISCLOSURE FORM

I have no Conflict of Interest to report.

Computer-Aided Detection of TB from Chest Radiographs in TB Prevalence Survey: Evaluation of Commercially Available AI Software

Zhi Zhen Qin, Stop TB Partnership

Background

- Computer-aided detection (CAD) products use AI to read CXR images and detect TB-associated abnormalities and have been recommended by the World Health Organization (WHO) for interpreting chest X-ray to triage and screen for TB in those aged 15 years and older, as an alternative to human readers
- However, this recommendation does not stipulate specific products, and notes more research is needed on performance in sub-populations.
- As the CAD market rapidly expands, many new products and software updates await independent scrutiny.

Study objectives

Using data from a recent TB prevalence study in South Africa, one of the high TB burden countries, this study aimed to:

- Evaluate the performance of 11 commercially available CAD products in detecting TB
- Analyse CAD performance in key sub-populations using data from the national TB prevalence survey in South Africa, a high TB/HIV-burden country.
- Identify the best thresholds for each product, depending on programmatic aims including target sensitivities of 90% and 80%, specificity of 70%, and a 10% CXR abnormality rate within sub-populations.

Methods

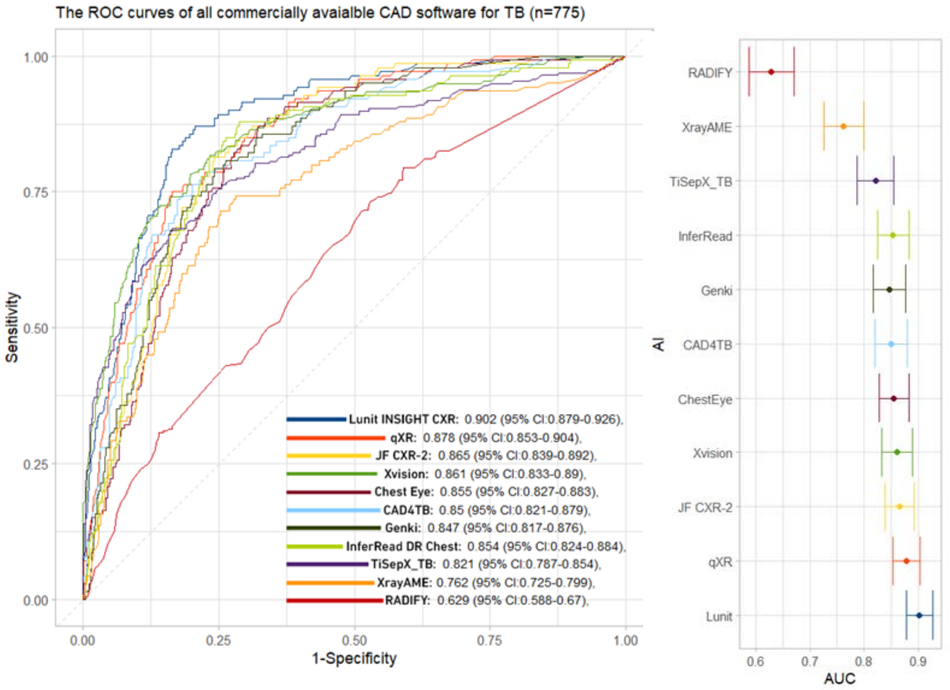
- This was a case-control evaluation used digital CXR images and metadata from individuals 15 years and older who participated in the South African national TB prevalence survey between 15 August 2017 and 28 July 2019.
- Only those with microbiological test results were eligible, with 516 bacteriologically negative and 258 bacteriologically positive.
- We used Xpert and culture results to form a composite microbiology reference standard (MRS); cases were people considered bacteriologically positive (Bac+) if they tested positive for TB on either or both tests.
- The chest X-ray images of the participants were stored on the Stop TB Partnership's server using Secure File Transfer Protocol (SFTP).
- Products were installed on the server to analyze the X-rays without any data transfer to developers. CAD readings were aggregated and exported.

Methods - analysis

- To compare overall performance of products, we plotted the Receiver Operating Characteristic (ROC) curve for each product and compared the AUCs.
- To evaluate the products as TB screening and triage tests, we analyzed the performance of each product against the target 90% sensitivity and 70% specificity values stipulated in the WHO's TPP.
- To provide insights on threshold selection, we modelled the programmatic effect of threshold selection and applied the sensitivity and specificity calculated in the case-control evaluation at all thresholds to calculate the X-ray abnormality rate, and confirmation test positivity rate.
- Then we defined criteria based on common programmatic targets, including target sensitivity, target specificity, and target test referral rate to evaluate how this affected threshold selection.

Results - overall comparison

The ROC curves of 11 commercially available CAD products against the composite MRS overall (n=775)



Results - TPP targets

<table>
<tr><th>AI</th><th>Sensitivity</th><th>Specificity</th></tr>
<tr><td>Lunit</td><td rowspan="5">90%</td><td>73.3% (69.2-77.0%)</td></tr>
<tr><td>ChestEye, qXR, InferRead, JF CXR-2</td><td>60-70%</td></tr>
<tr><td>XVision, Genki, CAD4TB</td><td>50-60%</td></tr>
<tr><td>TiSepX TB</td><td>40-50%</td></tr>
<tr><td>XrayAME, RADIFY</td><td>30-40%</td></tr>
<tr><th>AI</th><th>Sensitivity</th><th>Specificity</th></tr>
<tr><td>Lunit</td><td>91.4% (85.5-95.5%)</td><td rowspan="4">70%</td></tr>
<tr><td>CAD4TB, Genki, ChestEye, qXR, JF CXR-2, XVision, InferRead</td><td>80-90%</td></tr>
<tr><td>XrayAME, TiSepX TB</td><td>70-80%</td></tr>
<tr><td>RADIFY</td><td>45.7% (39.5-52%)</td></tr>
</table>

Results - threshold selection

Thresholds to match 10% test referral rate, i.e., X-ray abnormal rate (sorted by sensitivity) = best and worst performers

Product	Threshold	Sensitivity	Specificity
Lunit	0.7	82.1% (74.8-88.1%)	83.7% (80.2-86.8%)
JF CXR-2	0.8	80.7% (73.2-86.9%)	75.8% (71.8-79.4%)
Xvision	0.58	42.6% (36.5-48.9%)	90.7% (87.9-93.1%)
RADIFY	0.8	30.2% (24.7-36.2%)	86.1% (82.8-88.9%)

Results -subgroup analysis

Sensitivity and Specificity for qXR in each subgroup at a threshold of 0.5

Sensitivity		Specificity	
New cases	**History of TB**	**New cases***	**History of TB***
84.8% (76.2-91.3%)	85.4% (70.8-94.4%)	84.9% (80.9-88.3%)	27.3% (19.9-35.7%)
HIV-	**HIV+**	**HIV- ***	**HIV+ ***
89.5% (81.1-95.1%)	80.6% (62.5-92.5%)	74.6% (69.6-79.1%)	54.7% (43.5-65.4%)
Symptomatic	**Asymptomatic**	**Symptomatic***	**Asymptomatic***
80.0% (68.2-88.9%)	89.3% (80.1-95.3%)	82.3% (77.4-86.6%)	55.4% (48.7-61.9%)

Sensitivity			Specificity		
Young (15 to <35 years)	**Middle age (35 to <55)**	**Older (55 years +)**	**Young (15 to <35 years)**	**Middle age (35 to <55)**	**Older (55 years +)**
92.0% (80.8-97.8%)	88.7% (77-95.7%)	70.3% (53-84.1%)	78.3% (70.7-84.8%)	60.5% (52.5-68.1%)	72.0% (65.5-78%)

Results -subgroup analysis

qXR specificity and threshold at 90% and 80% sensitivity and disaggregated by subgroups.

Sensitivity		New Case	With TB History	HIV-	HIV+	Young 15<35 years	Middle Age 35<55 years	Older 55 years +
90%	***Threshold***	*0.30*	*0.35*	*0.41*	*0.3*	*0.58*	*0.52*	*0.18*
	Specificity	75.3% (70.6-79.5%)	25.0% (17.9-33.3%)	72.5% (67.4-77.2%)	50.0% (39-61%)	78.3% (70.7-84.8%)	61.7% (53.8-69.2%)	49.8% (42.8- 56.7%)
80%	***Threshold***	*0.59*	*0.73*	*0.81*	*0.52*	*0.85*	*0.84*	*0.30*
	Specificity	86.7% (82.9-89.9%)	46.2% (37.5-55.1%)	83.1% (78.7-87%)	58.1% (47-68.7%)	87.4% (80.8-92.4%)	77.2% (69.9-83.4%)	58.3% (51.3-65%)

Conclusion

- Many CAD products performed at a similarly high level overall in a high HIV/TB-burden population, but there are important differences in algorithms.
- No significant difference was noted in PLHIV compared to HIV-naive populations in products' AUC, thus further evaluations in this and other populations are necessary.
- Several context-specific factors should be considered when deciding on which product and threshold to use, including population screened, diagnostic test availability and co-morbidities.
- A CAD evaluation platform is urgently needed to aid in threshold selection attuned to specific contexts.

Acknowledgements

Martie Van der Walt, South African Medical Research Council

Sizulu Moyo, Human Sciences Research Council

Farzana Ismail, National Institute for Communicable diseases, South Africa

Phaleng Maribe, Human Sciences Research Council

Claudia M. Denkinger, Division of Infectious Disease and Tropical Medicine, University Hospital Heidelberg

Rachael Barrett, Stop TB Partnership

Lindiwe Mvusi, South African National Department of Health

Sarah Zaidi, Stop TB Partnership

Nkateko Mkhondo, World Health Organization Country Office South Africa

Khangelani Zuma, Human Sciences Research Council

Samuel Manda, South African Medical Research Council

Thuli Mthiyane, South African Medical Research Council

Jacob Creswell, Stop TB Partnership