



Test Report

CAD4TB Version 7

โดย กองวัณโรค กรมควบคุมโรค กระทรวงสาธารณสุข

รายงานผลการทดสอบ

โดยราชวิทยาลัยรังสีแพทย์แห่งประเทศไทย

ทดสอบใช้กับภาพรังสีทรวงอก ในกรณี

- คัดกรอง (screening) วัณโรคปอด
- อ่านผลซ้ำ (double reading) ให้กับรังสีแพทย์ เพื่อเพิ่มคุณภาพการวินิจฉัย
- เพิ่มความแม่นยำในการค้นหาพยาธิสภาพให้กับรังสีแพทย์
- ประมาณความยาก-ง่ายในการแปลผล
- จัดลำดับความเร่งด่วน (triage) ในการแปลผลให้แก่รังสีแพทย์



Performance of Artificial Intelligence for Tuberculosis Screening in Chest X-Ray Images of the Thai Population.



CAD4TB 7

Developed by DELFT Imaging

Submitted by Division of Tuberculosis, Department Disease Control, Ministry of Public Health

ส่งทดสอบโดย กองวัณโรค กรมควบคุมโรค กระทรวงสาธารณสุข

This report evaluates an artificial intelligence system developed for tuberculosis screening on chest radiographs in the Thai population. The primary objective is to assess screening performance for pulmonary tuberculosis; the secondary objective is to assess detection of other clinically relevant chest abnormalities. The evaluation was conducted under the project “Development of a Dataset for Testing Artificial Intelligence for Tuberculosis Screening in Chest X-ray Images of the Thai Population,” funded by the Thailand Centre of Excellence for Life Sciences (TCELS).

Dataset Code	123AAA	Number of Images	1,012
Test Completion Date	September 11, 2025	Revision Number	1
Latest Report Date	September 26, 2025	Approval Date	November 12, 2025

1. Executive Summary

CAD4TB v7 is a software CAD that assigns a 0–100 tuberculosis (TB) score to adult PA chest radiographs. CAD4TB v7 is intended as an adjunct to clinical assessment and microbiological testing, not a stand-alone diagnostic.

Clinical evaluation used a Thai multi-centre dataset (Ref. 123AAA) of 1,012 PA CXRs (≥ 15 years). Three NIOSH-certified B Readers independently interpreted each image; majority consensus served as the reference standard. Inter-rater reliability for key targets was strong (ICC > 0.90 ; Cohen's $\kappa > 0.75$ for overall abnormalities, small/large opacities, and TB), supporting the robustness of the ground truth. B Reader vs B Reader pairwise agreement was 0.9282 (κ 0.8532).

CAD4TB demonstrated high discrimination (AUC 0.9770). At a dataset-optimized operating point, performance was sensitivity 0.9402, specificity 0.9110, PPV 0.9354, NPV 0.9175; pairwise agreement with B Readers 0.9190 (κ 0.8340).

CAD4TB v7 can be tuned to meet WHO screening targets across multiple use cases for high-sensitivity/high-specificity screening (minimal accuracy), high-sensitivity screening (minimal accuracy and optimal accuracy), high-specificity screening (minimal accuracy and optimal accuracy).

2. AI Software Summary

2.1 Applicant

Applicant	Division of Tuberculosis, Department Disease Control, Ministry of Public Health กองวัณโรค กรมควบคุมโรค กระทรวงสาธารณสุข
Address	116 ถนนสุดประเสริฐ (ฝั่งขวา) แขวงบางโคล่ เขตบางคอแหลม กรุงเทพมหานคร 10120
Country	Thailand

2.2 Manufacturer / Developer

Manufacturer / Developer	Delft Imaging
Address	Utrechtseweg 48, 3704 HE Zeist
Country	The Netherlands
Website	https://www.delft.care

2.3 Software Identification

Proprietary Name	CAD4TB
Common/Usual Name	Computer-Aided Detection for Tuberculosis
Model/Version	7
Software Category	Computer-Aided Detection (CAD) Software

2.4 Software Description

CAD4TB (Computer-Aided Detection for Tuberculosis) version 7 is a software solution developed by Delft Imaging for the automated analysis of digital chest radiographs. The system uses artificial intelligence algorithms to detect radiological abnormalities suggestive of pulmonary tuberculosis (TB).

The software assigns each chest X-ray (CXR) an abnormality score that reflects the likelihood of TB. A user-defined threshold can be applied to decide whether individuals require further diagnostic testing, such as sputum Xpert MTB/RIF or culture. CAD4TB v7 is designed to support healthcare providers in TB screening programs, particularly in settings with high TB prevalence and shortages of trained radiologists.

CAD4TB has been widely deployed in global TB programs, particularly in low-resource settings, and is integrated with mobile X-ray units to enable point-of-care TB screening in remote areas. The software provides rapid, standardized, and reproducible results, helping reduce reliance on human readers, minimize diagnostic delays, and increase case detection rates.

3. Dataset: Chest X-Ray Images of the Thai Population

A dataset of digital chest radiographs was curated from multiple hospitals across Thailand. This section describes the population characteristics, reader labelling protocol, distribution of radiographic findings, and inter-rater reliability, along with key limitations of the dataset.

3.1 Population

The dataset, referred to as 123AAA, included 1,012 posterior–anterior (PA) digital radiography chest images from individuals aged 15 years and older (Table 1). Images were excluded if the patient had a positive HIV serology or other opportunistic pulmonary infections, or if the image quality fell below minimum thresholds due to poor positioning, exposure, motion, or artifacts.

Table 1 summarizes the dataset reference, total number of images, patient age, modality, projection, and key exclusion criteria. These criteria were applied to ensure that the dataset reflects typical chest radiographs for TB screening while controlling for confounding pathologies that may interfere with CAD interpretation.

Table 2 lists the source institutions and sampling summary. Images were randomly sampled from five institutions across Southern, Northern, Northeastern, and Central Thailand, including tertiary hospitals and the Tuberculosis Division of the Department of Disease Control. The total of 1,012 images used in the evaluation was drawn from a larger curated pool of 1,100 annotated images, ensuring coverage across geographic regions and patient demographics.

Table 1. Dataset 123AAA overview and criteria

Dataset Reference	123AAA
Total Image	1,012
Age	≥15 years
Modality	Digital Radiography (DR) Chest Radiographs
Projection	Posterior–anterior (PA)
Key Exclusion	HIV serology positive; Opportunistic or co-infections (e.g., Mycobacterium tuberculosis, Histoplasmosis, Cryptococcosis, Melioidosis, Acinetobacter baumannii); Positioning/Exposure/Motion/Artifact below minimum quality threshold

Table 2. Source institutions and sampling summary

Source Institutions	<ol style="list-style-type: none"> 1. Songklanagarind Hospital (Southern Thailand) 2. Chiangrai Pracharuk Hospital (Northern Thailand) 3. Udon Thani Hospital (Northeastern Thailand) 4. Suttawet Hospital (Northeastern Thailand) 5. Tuberculosis Division, Department of Disease Control, Ministry of Public Health (Central Thailand)
Sampling	A total of 1,012 posterior–anterior digital radiography chest radiographs randomly sampled from a pool of 1,100 images.

3.2 Reader Labelling Protocol

To establish a reference standard, each image was independently interpreted by three B Readers, randomly selected from a pool of six B reader certified radiologists. B Readers are radiologists trained and certified by the National Institute for Occupational Safety and Health (NIOSH) in the United States to classify chest radiographs for pneumoconiosis, which ensures a high standard of expertise in detecting subtle parenchymal and pleural abnormalities.

As shown in Table 3, all readings were independent, and a majority consensus rule was applied to determine the final label for each finding. This approach reduces the impact of individual reader variability and provides a robust reference for evaluating CAD4TB v7 performance.

Table 3. Reader labelling protocol and consensus rules

Readers	Three B Readers per image from a pool of six B Readers
B Reader Definition	Radiologists certified by NIOSH to interpret and classify chest radiographs for pneumoconiosis
Labelling Method	Independent reads
Consensus Finding	Majority consensus

3.3 Characteristics of Findings

The distribution of radiographic findings across individual readers and consensus labels is summarized in Table 4. The table shows both the Reader Count (sum of markings across all readers) and Consensus Count (majority agreement among the three readers) for each finding.

Parenchymal opacities were categorized as small or large, and a combined category (“Opacity (small/large)”) was used to facilitate CAD evaluation. Other key findings, including masses/nodules, cavities, fibrosis, calcifications, pleural effusions, pleural thickening, pneumothorax, hilar and mediastinal adenopathy, and tuberculosis-related changes, are also detailed. Among these, active tuberculosis and indeterminate tuberculosis were combined to provide a comprehensive assessment of TB-related abnormalities.

Table 4. Findings counts by radiologists and consensus (1,012 images)

Findings	Reader Count	Consensus Count
Abnormalities	2,106	692
Opacity (small/large) [#]	1,955	645
• Small opacity	1,692	572
• Large opacity	1,628	561
Mass/nodule	686	189
Cavity	1,134	380
Fibrosis	1,033	339
Calcification	415	86
Pleural effusion	436	141
Pleural thickening	740	238
Pneumothorax	18	5
Hilar adenopathy	420	105
Mediastinal adenopathy	123	21
Tuberculosis (active/indeterminate) [#]	1,742	585
• Active Tuberculosis	1,559	538
• Indeterminate tuberculosis	183	37

Note: Reader Count (RC) = sum of radiologists marks over all readers and images; Consensus Count (CC) = images marked by majority rule (≥ 2 of 3 radiologists). Opacity (small/large) refers to combined parenchymal opacity encompassing both small and large opacities.

3.4 Inter-rater reliability

The consistency of reader annotations was assessed using multiple metrics, including pairwise agreement, intraclass correlation coefficient (ICC), Cohen's κ , and Fleiss's κ . Table 5 presents inter-rater reliability measures for each radiographic finding.

Findings such as overall abnormalities, small and large opacities, and tuberculosis showed high agreement (ICC > 0.9; Cohen's κ > 0.75), indicating robust consensus among readers. Other findings with lower prevalence, such as mediastinal adenopathy or pneumothorax, demonstrated lower agreement, reflecting the challenges of detecting rare or subtle features.

Table 5. Inter-rater reliability measures for each finding (Internal Validation)

Findings	Agreement	ICC	Cohen's κ	Fleiss's κ
Abnormalities	0.9289	0.9372	0.8326	0.8326
Small opacity	0.9236	0.9376	0.8334	0.8334
Large opacity	0.8432	0.8657	0.6823	0.6823
Opacity (small/large) [#]	0.8794	0.9037	0.7577	0.7576
Mass/nodule	0.7582	0.5733	0.3088	0.3088
Cavity	0.8498	0.8641	0.6794	0.6791
Fibrosis	0.7510	0.7072	0.4458	0.4454
Calcification	0.8030	0.3733	0.1658	0.1654
Pleural effusion	0.9282	0.8793	0.7079	0.7081
Pleural thickening	0.8314	0.7808	0.5428	0.5426
Pneumothorax	0.9967	0.8856	0.7206	0.7206
Hilar adenopathy	0.8386	0.5896	0.3233	0.323
Mediastinal adenopathy	0.9381	0.4356	0.1993	0.2035
Tuberculosis	0.9282	0.9458	0.8532	0.8532
• Active Tuberculosis	0.8900	0.9141	0.7798	0.7798
• Indeterminate tuberculosis	0.9196	0.5517	0.2911	0.2906

Notes: Opacity (small/large) refers to combined parenchymal opacity encompassing both small and large opacities. Tuberculosis refers to combined active tuberculosis and indeterminate tuberculosis.

3.5 Limitations

Images were primarily collected from hospitals. Although some images were sourced from community-based active case findings, this may not fully represent the community-level distribution of TB abnormalities. This may reduce generalizability in high HIV prevalence settings, where TB presentation can differ. Certain findings (e.g., pneumothorax, mediastinal adenopathy) had few occurrences, limiting the reliability of inter-rater agreement metrics.

4. Evaluation Results

4.1 Label Mapping

To standardize comparisons between CAD4TB v7 outputs and radiologist interpretations, AI output classes were mapped to radiologist findings. Table 6 summarizes the mapping and aggregation rules applied. For TB, the AI output class “TB” was mapped indirectly to radiologist-assessed active and indeterminate TB findings.

Table 6. Radiologist's finding and AI Label mapping and aggregation rules

AI Output	Findings	Mapping Rule
TB	Tuberculosis	Indirect Mapping (Active TB and Indeterminate TB)

4.2 Results Comparison: AI vs Radiologist

CAD4TB v7 outputs a TB score ranging from 0–100 (see Table 7). Two types of thresholds were evaluated:

1. **Manufacturer-recommended threshold** which is the standard threshold suggested by Delft Imaging at 60.
2. **Dataset-optimized threshold** which was selected based on an ROC analysis to maximize **Youden's index** for the dataset. The dataset-derived thresholds were selected on the full evaluation dataset. They are presented for comparison only and must not be interpreted as validated cutoffs for clinical deployment without calibration.

Table 7. Manufacturer's recommended thresholds for AI output classes

AI Output	Score Range	Recommended Threshold	Dataset-optimized threshold [#]
TB	0–100	60	46

Inter-rater agreement was assessed using pairwise agreement and Cohen's κ . As shown in **Table 8**, CAD4TB v7 achieved an agreement with radiologists at both thresholds. Agreement was slightly higher at the dataset-optimized threshold (pairwise agreement 0.9190, κ 0.8340) than at the manufacturer-recommended threshold (pairwise agreement 0.8834, κ 0.7664). For comparison, radiologist–radiologist agreement (“B vs B”) was higher overall (pairwise agreement 0.9282, κ 0.8532).

Table 8. Pairwise agreement measures for AI vs radiologists

AI Output	Threshold	N	Pairwise Agreement		Cohen's Kappa	
			B vs B	AI vs B	B vs B	AI vs B
TB	46	588	0.9282	0.9190	0.8532	0.8340
TB	60	592	0.9282	0.8834	0.8532	0.7664

4.3 Diagnostic Performance

Diagnostic performance was assessed against the radiologist consensus (majority rule of three B Readers). Table 9 summarizes key performance metrics at the two thresholds, including true positives (TP), false positives (FP), false negatives (FN), true negatives (TN), and derived measures: sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

Table 9. Diagnostic performance metrics AI vs radiologist consensus

AI Output	Threshold	TP	FP	FN	TN	Sensitivity	Specificity	PPV	NPV
TB	46	550	38	35	389	0.9402	0.9110	0.9354	0.9175
TB	60	494	18	91	409	0.8444	0.9578	0.9648	0.8180

4.4 Sensitivity and Specificity across Thresholds

To evaluate the diagnostic performance of CAD4TB, sensitivity and specificity were calculated for a wide range of thresholds (Table 10). As expected, lower thresholds resulted in very high sensitivity but low specificity, suggesting that most true TB cases were detected at the expense of a large number of false positives. Conversely, higher thresholds resulted in improved specificity but lower sensitivity, reflecting stricter criteria for the detection of abnormalities.

This highlights the inherent trade-off between sensitivity and specificity and underlines the importance of selecting an appropriate threshold for deployment depending on the intended use case, whether maximizing case detection (screening) or minimizing false positives (confirmatory testing).

Table 10. Sensitivity and specificity across thresholds for Tuberculosis

Threshold	Sensitivity	Specificity	Threshold	Sensitivity	Specificity
1	1.0000	0.1920	26	0.9897	0.7939
2	1.0000	0.3489	27	0.9880	0.7986
3	1.0000	0.4473	28	0.9880	0.8033
4	1.0000	0.5363	29	0.9880	0.8033
5	1.0000	0.5855	30	0.9846	0.8080
6	0.9983	0.6159	31	0.9812	0.8126
7	0.9983	0.6440	32	0.9795	0.8197
8	0.9966	0.6674	33	0.9795	0.8244
9	0.9966	0.6979	34	0.9778	0.8314
10	0.9966	0.7049	35	0.9744	0.8337
11	0.9966	0.7190	36	0.9709	0.8384
12	0.9949	0.7377	37	0.9658	0.8407
13	0.9949	0.7424	38	0.9658	0.8501
14	0.9949	0.7471	39	0.9641	0.8571
15	0.9949	0.7564	40	0.9573	0.8712
16	0.9949	0.7588	41	0.9538	0.8782
17	0.9949	0.7611	42	0.9504	0.8806
18	0.9949	0.7658	43	0.9453	0.8829
19	0.9949	0.7682	44	0.9453	0.8970
20	0.9949	0.7705	45	0.9453	0.8993
21	0.9932	0.7705	46	0.9402	0.9110
22	0.9932	0.7752	47	0.9316	0.9227
23	0.9932	0.7775	48	0.9248	0.9251
24	0.9915	0.7822	49	0.9197	0.9297
25	0.9915	0.7845	50	0.9128	0.9344

Table 10. Sensitivity and specificity across thresholds for Tuberculosis (continued)

Threshold	Sensitivity	Specificity	Threshold	Sensitivity	Specificity
51	0.9128	0.9344	76	0.5419	0.9930
52	0.9094	0.9368	77	0.5316	0.9930
53	0.9043	0.9415	78	0.5197	0.9930
54	0.8940	0.9485	79	0.5026	0.9930
55	0.8906	0.9485	80	0.4838	0.9930
56	0.8821	0.9485	81	0.4684	0.9930
57	0.8735	0.9485	82	0.4564	0.9930
58	0.8667	0.9508	83	0.4513	0.9930
59	0.8632	0.9555	84	0.4393	0.9953
60	0.8547	0.9578	85	0.4256	0.9953
61	0.8444	0.9578	86	0.3932	0.9953
62	0.8393	0.9602	87	0.3846	0.9953
63	0.8325	0.9625	88	0.3624	0.9953
64	0.8308	0.9649	89	0.3333	0.9977
65	0.8274	0.9696	90	0.3026	0.9977
66	0.8103	0.9719	91	0.2684	0.9977
67	0.7932	0.9719	92	0.2308	0.9977
68	0.7829	0.9766	93	0.1897	0.9977
69	0.7761	0.9766	94	0.1727	0.9977
70	0.7521	0.9789	95	0.1419	1.0000
71	0.7368	0.9813	96	0.0889	1.0000
72	0.7094	0.9813	97	0.0615	1.0000
73	0.6821	0.9859	98	0.0359	1.0000
74	0.6547	0.9859	99	0.0137	1.0000
75	0.6068	0.9859	100	0.0017	1.0000

Note: The recommended threshold for CAD4TB is 60

5. WHO Target Product Profile (TPP) Performance Analysis

The World Health Organization (WHO) sets performance targets for TB screening tests to help health workers choose the right tests for their population. These targets define the **minimum-acceptable performance** and the **optimal performance** for different tests.

There are three main types of TB screening tests:

1. **High-sensitivity, high-specificity test** aims at detecting almost all TB cases (high sensitivity) and has few false positives (high specificity). This is useful in general population screening, where you want to detect TB without over-referring healthy people.
2. **High-sensitivity test** aims at detecting most TB cases (high sensitivity), but may produce more false positives (lower specificity). This is useful when missing TB cases is more dangerous than over-referring people.
3. **High-specificity test** aims at detecting fewer TB cases (lower sensitivity) but almost never mislabels healthy people as TB-positive (high specificity). This is useful when confirming TB after an initial screening test.

These targets help ensure that TB screening tests detect enough cases without overwhelming the health system with false positives, guiding safe and efficient TB control programs.

Table 11. WHO Target Product Profile (TPP) for different TB screening tests

Test Type	Minimal Accuracy		Optimal Accuracy	
	Sensitivity	Specificity	Sensitivity	Specificity
High-sensitivity, High specificity screening test	90%	80%	95%	95%
	<i>Conditions Met (Thresholds: 28–53)</i>			
High-sensitivity screening test	90%	60%	95%	85%
	<i>Conditions Met (Thresholds: 6–53)</i>		<i>Conditions Met (Thresholds: 38–45)</i>	
High-specificity screening test	60%	98%	70%	98%
	<i>Conditions Met (Thresholds: 71–75)</i>		<i>Conditions Met (Thresholds: 71–72)</i>	

Notes: Minimal Accuracy is Achievable even in populations where TB is rare (~1%). Optimal Accuracy is achievable even in very low-prevalence populations (~0.25%). For high-specificity tests, minimal sensitivity may be lower (50%) because the focus is on avoiding false positives.

For tuberculosis detection, CAD4TB can meet WHO **minimal** TPP requirements for high-sensitivity/high-specificity screening over a broad range of thresholds (28–53). It can also be tuned to meet high-specificity TPP targets ($\geq 98\%$ specificity) at higher thresholds (71–75) when minimizing false positives is paramount. The vendor-recommended threshold of 60 favours specificity (95.8%) but sacrifices sensitivity relative to radiologists and therefore may be better suited to confirmatory or triage workflows than mass screening.

6. Aligning with WHO Policy Statement on the Use of CAD for TB Screening

Aligning with the WHO policy statement on "Use of computer-aided detection software for tuberculosis screening", clinical performance of CAD software with that of expert radiologists was compared.

The radiologists' sensitivity was 0.9018 (95% CI: 0.8819–0.9218) and specificity was 0.9408 (95% CI: 0.9229–0.9586), as detailed in Table 12. With these results as a comparator, we evaluate the corresponding sensitivity and specificity of CAD4TB across different thresholds as shown in the corresponding Receiver operating characteristic (ROC) curve in Figures 1 and 2. The area under the curve (AUC) was 0.9770. At a fixed sensitivity of 0.9018, the specificity of CAD4TB was 0.9438. At a fixed specificity of 0.9408, the sensitivity of CAD4TB was 0.9043.

Table 12. Radiologists' sensitivity and specificity with 95% Tango's CI

Metrics	Mean	95% CI Upper	95% CI Lower
Sensitivity	0.9018	0.8819	0.9218
Specificity	0.9408	0.9229	0.9586

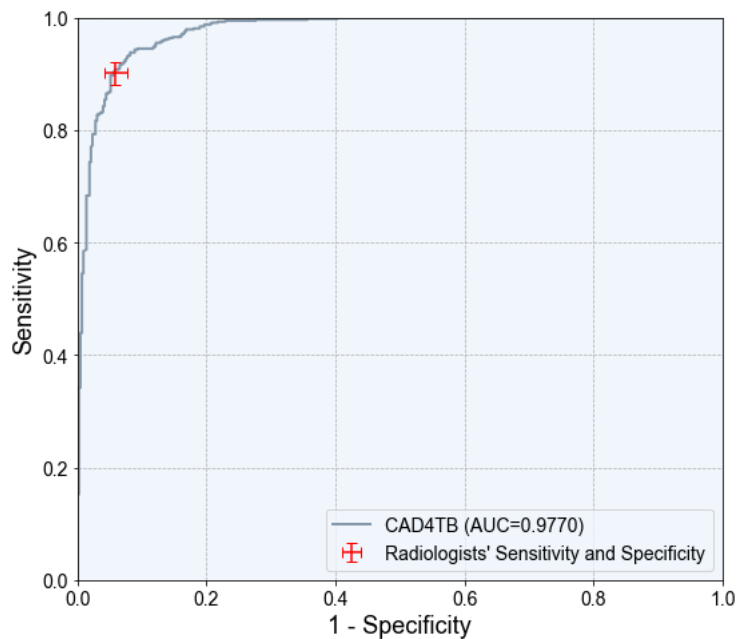


Figure 1. ROC curves showing the performance of CAD4TB, compared with radiologists.

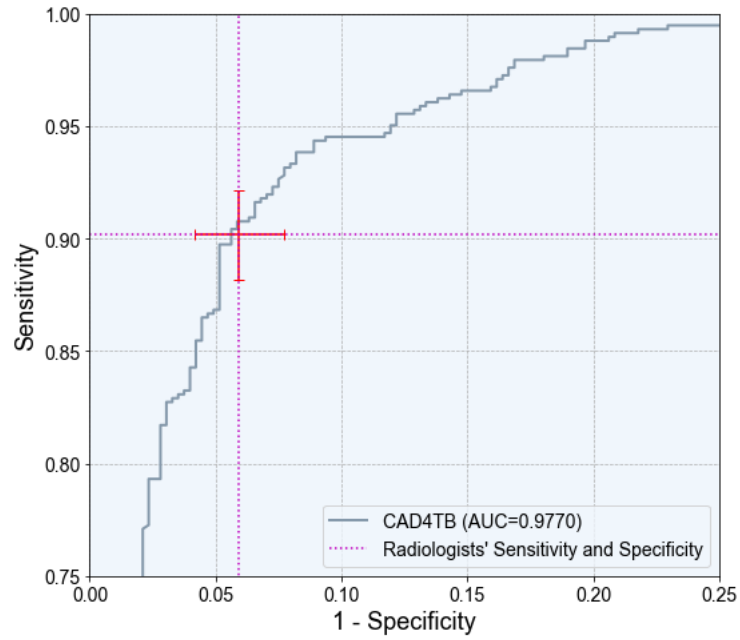


Figure 2. Partial ROC curves at 75-100% specificity showing the performance of CAD4TB, compared with radiologists.

For CAD4TB, we observed higher TB scores for people with TB, whereas the scores were on the lower end of the distribution for the non-TB cohort. Figure 3 presents a histogram showing distribution of the TB scores from CAD4TB.

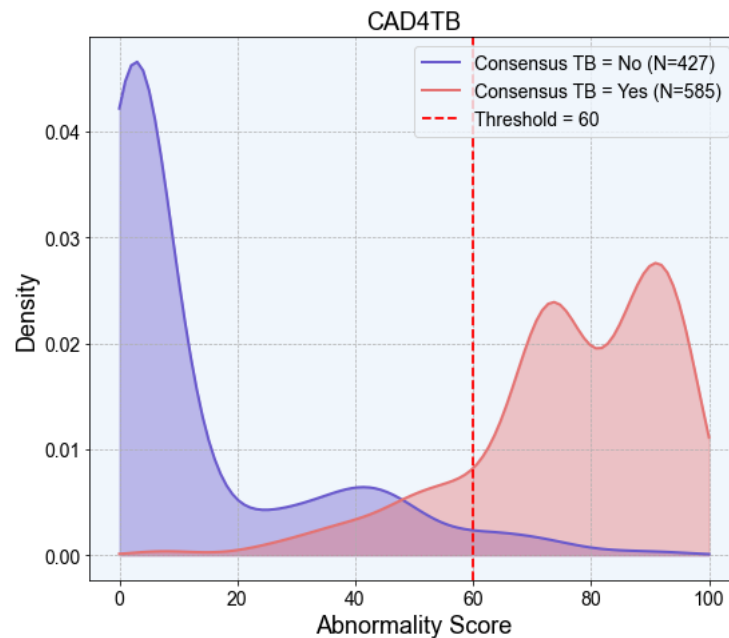


Figure 3. Density plot of TB scores for CAD4TB for TB and non-TB cases (as per the radiologist consensus)

7. Conclusion

Software & Sponsor:

CAD4TB v7 by Delft Imaging is an AI CAD tool for automated analysis of digital chest radiographs to support TB screening. The Thai Division of Tuberculosis (Department of Disease Control, MOPH) served as the applicant and test submitter.

Dataset:

A curated Thai CXR dataset (Ref. 123AAA) comprised 1,012 PA DR images (≥ 15 years). Each image was read independently by three NIOSH-certified B Readers from a pool of six; majority rule provided the consensus label. Inter-rater reliability was high for core targets (e.g., abnormalities, small/large opacities, TB), with ICC > 0.90 and Cohen's $\kappa > 0.75$ for these findings.

AI Output & Thresholds:

CAD4TB v7 yields a TB score 0–100. Two operating points were emphasized:

- Dataset-optimized threshold: 46 (chosen by ROC/Youden)
- Manufacturer-recommended threshold: 60

AI vs Radiologists:

Pairwise agreement and Cohen's κ showed strong concordance:

- Baseline (B-reader vs B-reader): Agreement 0.9282, κ 0.8532
- Threshold 46: Agreement 0.9190, κ 0.8340
- Threshold 60: agreement 0.8834, κ 0.7664

Diagnostic Performance on consensus agreement:

- Threshold 46: Sensitivity 0.9402, Specificity 0.9110, PPV 0.9354, NPV 0.9175
- Threshold 60: Sensitivity 0.8444, Specificity 0.9578, PPV 0.9648, NPV 0.8180

WHO TPP Alignment

CAD4TB v7 can be tuned to meet WHO screening TPPs:

- High-sensitivity/High-specificity screening met requirements at thresholds 28–53 (minimal accuracy).
- High-sensitivity screening met requirements at 6–53 (minimal accuracy) and 38–45 (optimal accuracy).
- High-specificity screening met requirements at 71–75 (minimal accuracy) and 71–72 (optimal accuracy).

Use-Case Guidance:

- Community or mass screening (maximize case detection, tolerate more positives): favour ~40–50 to achieve ~94% sensitivity with $\geq 91\%$ specificity.
- Triage or confirmatory workflows (limit secondary testing): consider ~60 for ~96% specificity and very high PPV.
- Very low-prevalence or constrained capacity: consider ≥ 71 to meet $\geq 98\%$ specificity per WHO high-specificity TPP, acknowledging reduced sensitivity.

Summary:

CAD4TB v7 demonstrates substantial agreement with expert readers and high discriminative performance (AUC 0.977). Threshold tuning enables alignment with WHO TPPs across diverse programs from high-throughput screening to specificity-focused triage.

8. References

1. World Health Organization. (2021). *Determining the local calibration of computer-assisted detection (CAD) thresholds and other parameters: A toolkit to support the effective use of CAD for TB screening* (ISBN 978-92-4-002861-6). Geneva: WHO. <https://iris.who.int/bitstream/handle/10665/345925/9789240028616-eng.pdf>
2. World Health Organization. (2025). *Use of computer-aided detection software for tuberculosis screening: WHO policy statement* (ISBN 978-92-4-011037-3). Geneva: WHO. <https://www.who.int/publications/i/item/9789240110373>
3. World Health Organization. (2025). *Target product profiles for tuberculosis screening tests* (ISBN 978-92-4-011357-2). Geneva: WHO. <https://www.who.int/publications/i/item/9789240113572>