Test Report

MOPH AI CXR Version 1

โดย บริษัท แคริว่า (ประเทศไทย) จำกัด

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.

MOPH AI CXR

ส่งทดสอบโดย บริษัท แคริว่า (ประเทศไทย) จำกัด

304 อาคารวานิชเพลซ อารีย์ (อาคาร เอ)
ชั้นที่ 25 ยูนิต 2501 ถนนพหลโยธิน แขวงสามเสนใน
เขตพญาไท กรุงเทพมหานคร 10400

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	March 29, 2026	**Revision Number**	2
Latest Report Date	April 8, 2026	**Approval Date**	April 9, 2026

1. Executive Summary

MOPH AI CXR is an artificial intelligence (AI)-based software for chest radiograph analysis developed by Cariva (Thailand) Co., Ltd. The software is intended to function as an adjunctive software tool for tuberculosis (TB) screening, triage, and radiographic decision support on chest X-ray images.

Clinical performance evaluation was conducted using a Thai multi-centre reader-reference dataset (Ref. 123AAA) comprising 1,012 posterior–anterior digital radiography chest images from individuals aged 15 years and older. Each image was independently interpreted by three NIOSH-certified B Readers selected from a pool of six, and majority consensus was used as the reference standard. Inter-rater reliability for key target findings was high, with ICC > 0.90 and Cohen's κ > 0.75 for principal findings including abnormalities, opacity, and tuberculosis, supporting the robustness of the reference standard. For TB, B Reader vs B Reader pairwise agreement was 0.9282, with Cohen's κ of 0.8532.

MOPH AI CXR demonstrated high discriminative performance for TB detection, with an area under the receiver operating characteristic curve (AUC) of 0.9836. At the manufacturer-recommended operating point (threshold = 0.2290), TB performance was sensitivity 0.9504, specificity 0.9087, positive predictive value (PPV) 0.9345, and negative predictive value (NPV) 0.9305. Agreement with B Readers at this operating point was 0.9200, with Cohen's κ of 0.8357, indicating performance approaching expert-reader consensus.

Based on the threshold analysis presented in this report, MOPH AI CXR can be configured to align with different WHO TB screening target product profile (TPP) categories across multiple intended use scenarios, including high-sensitivity/high-specificity screening, high-sensitivity screening, and high-specificity screening. At the manufacturer-recommended operating point, the software is most consistent with a high-sensitivity screening application, supporting its use in screening and triage workflows where minimizing missed TB cases is a priority.

2. AI Software Summary

2.1 Applicant

Applicant	บริษัท แคริว่า (ประเทศไทย) จำกัด
Address	304 อาคารวานิชเพลซ อารีย์ (อาคาร เอ) ชั้นที่ 25 ยูนิต 2501 ถนนพหลโยธิน แขวงสามเสนใน เขตพญาไท กรุงเทพมหานคร 10400
Country	Thailand

2.2 Manufacturer / Developer

Manufacturer / Developer	บริษัท แคริว่า (ประเทศไทย) จำกัด
Address	304 อาคารวานิชเพลซ อารีย์ (อาคาร เอ) ชั้นที่ 25 ยูนิต 2501 ถนนพหลโยธิน แขวงสามเสนใน เขตพญาไท กรุงเทพมหานคร 10400
Country	Thailand
Website	https://www.cariva.co.th/

2.3 Software Identification

Proprietary Name	MOPH AI CXR
Common/Usual Name	MOPH AI CXR
Model/Version	1.0
Software Category	Software as a Medical Device (SaMD) – Computer-Aided Detection (CADe) for Tuberculosis Screening from Chest X-ray Images

2.4 Software Description

MOPH AI CXR เป็นโมเดลปัญญาประดิษฐ์ (AI) ที่ถูกพัฒนาขึ้นเพื่อวิเคราะห์และคัดกรองผู้ป่วยวัณโรคจากภาพถ่ายรังสีทรวงอก (CXR) โดยระบบสามารถประเมินความเสี่ยง และแสดง heatmap เพื่อช่วยบุคลากรทางการแพทย์ในการระบุรอยโรคเบื้องต้นอย่างแม่นยำ เพื่อใช้เป็นเครื่องมือสนับสนุนการตัดสินใจในการคัดกรอง (Screening) ก่อนเข้าสู่กระบวนการวินิจฉัยและรักษาลำดับต่อไป

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 Images	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	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 Summary	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 model 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 Rule	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 MOPH AI CXR V1.0 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

MOPH AI CXR V1.0 outputs a TB score ranging from 0.00–1.00 (see Table 7). **Manufacturer-recommended threshold** is the standard threshold suggested by CARIVA at 0.2290.

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

AI Output	Score Range	Threshold
TB	0.00–1.00	0.2290

Inter-rater agreement was assessed using pairwise agreement and Cohen's κ. As shown in **Table 8.** At the manufacturer-recommended threshold of 0.2290, CARIVA demonstrated strong agreement with B Readers. The pairwise agreement between AI and radiologists was 0.9200, with a Cohen's κ of 0.8357. For reference, inter-radiologist agreement (B vs B) showed a higher pairwise agreement of 0.9282 and a Cohen's κ of 0.8532. While AI–radiologist agreement was slightly lower than radiologist–radiologist agreement, the difference was modest, suggesting that CARIVA performs at a level approaching expert consensus.

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	0.2290	585	0.9282	0.9200	0.8532	0.8357

4.3 Diagnostic Performance

Diagnostic performance was assessed against the radiologist consensus (majority rule of three B Readers). Table 9 summarizes key performance metrics a threshold of 0.2290, 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	0.2290	556	39	29	388	0.9504	0.9087	0.9345	0.9305

4.4 Sensitivity and Specificity across Thresholds

To evaluate the diagnostic performance of MOPH AI CXR V1.0, 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
0.01	1.0000	0.0047	0.26	0.9350	0.9251
0.02	1.0000	0.0515	0.27	0.9333	0.9274
0.03	1.0000	0.3513	0.28	0.9316	0.9368
0.04	0.9949	0.5785	0.29	0.9299	0.9391
0.05	0.9949	0.7237	0.30	0.9299	0.9415
0.06	0.9846	0.7681	0.31	0.9265	0.9438
0.07	0.9846	0.7939	0.32	0.9248	0.9438
0.08	0.9812	0.8056	0.33	0.9197	0.9461
0.09	0.9778	0.8267	0.34	0.9179	0.9461
0.10	0.9744	0.8361	0.35	0.9162	0.9485
0.11	0.9726	0.8454	0.36	0.9111	0.9508
0.12	0.9709	0.8548	0.37	0.9111	0.9508
0.13	0.9709	0.8618	0.38	0.9094	0.9508
0.14	0.9709	0.8642	0.39	0.9094	0.9532
0.15	0.9709	0.8689	0.40	0.9077	0.9532
0.16	0.9692	0.8712	0.41	0.9060	0.9555
0.17	0.9624	0.8782	0.42	0.9060	0.9578
0.18	0.9590	0.8899	0.43	0.9043	0.9578
0.19	0.9556	0.8970	0.44	0.9026	0.9578
0.20	0.9538	0.8993	0.45	0.9009	0.9602
0.21	0.9538	0.9016	0.46	0.9009	0.9625
0.22	0.9538	0.9063	0.47	0.8991	0.9625
0.23	0.9504	0.9110	0.48	0.8957	0.9625
0.24	0.9436	0.9133	0.49	0.8940	0.9625
0.25	0.9402	0.9227	0.50	0.8923	0.9625

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

Threshold	Sensitivity	Specificity	Threshold	Sensitivity	Specificity
0.51	0.8855	0.9625	0.76	0.8085	0.9883
0.52	0.8821	0.9649	0.77	0.8051	0.9883
0.53	0.8821	0.9649	0.78	0.8017	0.9930
0.54	0.8803	0.9649	0.79	0.8000	0.9930
0.55	0.8769	0.9649	0.80	0.7932	0.9930
0.56	0.8769	0.9649	0.81	0.7795	0.9930
0.57	0.8735	0.9672	0.82	0.7709	0.9930
0.58	0.8667	0.9672	0.83	0.7658	0.9953
0.59	0.8564	0.9672	0.84	0.7538	0.9953
0.60	0.8547	0.9696	0.85	0.7436	0.9953
0.61	0.8496	0.9696	0.86	0.7316	0.9953
0.62	0.8496	0.9696	0.87	0.7197	0.9953
0.63	0.8496	0.9696	0.88	0.7077	0.9953
0.64	0.8496	0.9696	0.89	0.6940	0.9953
0.65	0.8479	0.9696	0.90	0.6786	1.0000
0.66	0.8444	0.9696	0.91	0.6701	1.0000
0.67	0.8427	0.9719	0.92	0.6479	1.0000
0.68	0.8410	0.9742	0.93	0.6222	1.0000
0.69	0.8376	0.9742	0.94	0.6051	1.0000
0.70	0.8359	0.9813	0.95	0.5761	1.0000
0.71	0.8359	0.9836	0.96	0.5248	1.0000
0.72	0.8222	0.9859	0.97	0.4513	1.0000
0.73	0.8222	0.9859	0.98	0.3453	1.0000
0.74	0.8205	0.9859	0.99	0.1214	1.0000
0.75	0.8171	0.9883	1.00	0.0000	1.0000

Note: The recommended threshold for MOPH AI CXR V1.0 is 0.2290.

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%
	Conditions Met (Thresholds: 0.08–0.46)		*Conditions Not Met*	
High-sensitivity screening test	90%	60%	95%	85%
	Conditions Met (Thresholds: 0.05–0.46)		*Conditions Met (Threshold: 0.12–0.23)*	
High-specificity screening test	60%	98%	70%	98%
	Conditions Met (Thresholds: 0.71–0.84)		*Conditions Met (Thresholds: 0.71–0.88)*	

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.

Notably, the manufacturer-recommended threshold (0.2290) favours sensitivity (95.04%) but does not reach the WHO ≥95% specificity criterion for the high-sensitivity, high-specificity TPP.

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 MOPH AI CXR V1.0 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 0.9836.

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

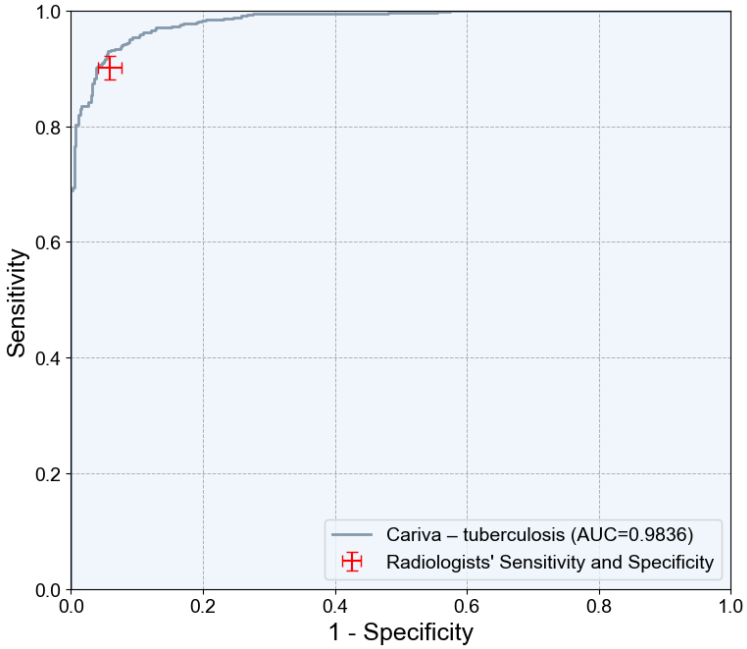

Figure 1. ROC curves showing the performance of MOPH AI CXR V1.0, compared with radiologists.

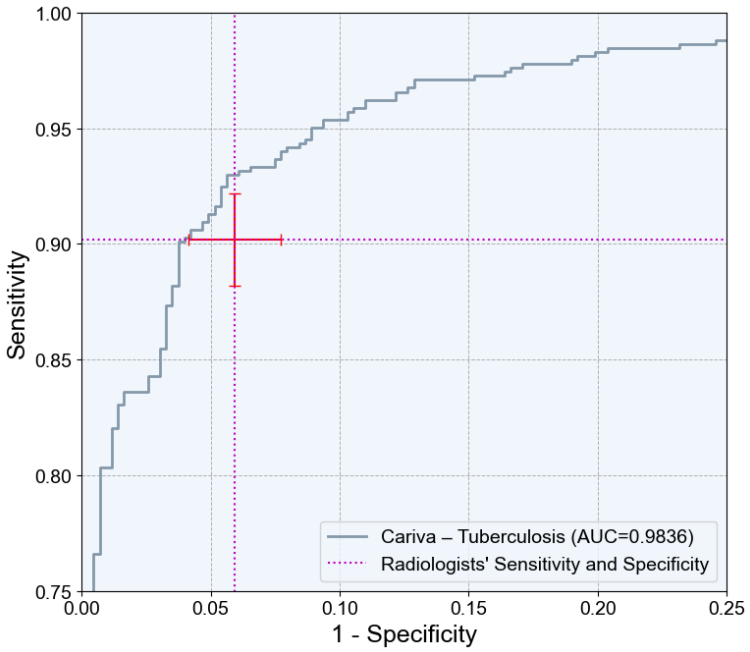

Figure 2. Partial ROC curves at 75-100% specificity showing the performance of MOPH AI CXR V1.0, compared with radiologists.

For MOPH AI CXR V1.0, 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 MOPH AI CXR V1.0.

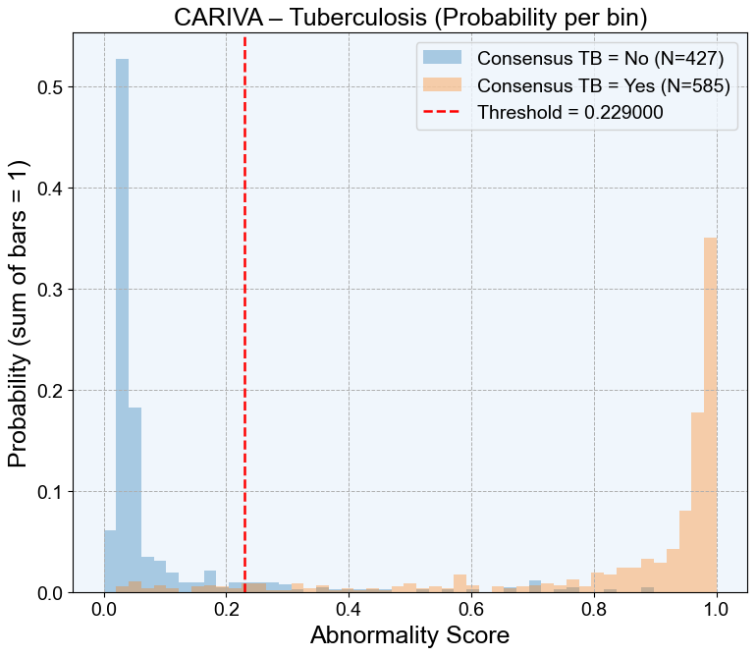

Figure 3. Histogram plot of TB scores for MOPH AI CXR V1.0 for TB and non-TB cases (as per the radiologist consensus)

7. Conclusion

Software & Sponsor

MOPH AI CXR is an artificial intelligence (AI)-based software for chest radiograph analysis developed by Cariva (Thailand) Co., Ltd. The software is intended to support healthcare professionals in the screening of pulmonary tuberculosis from chest X-ray images.

Dataset

Performance was evaluated using a curated Thai chest radiograph dataset (Ref. 123AAA) comprising 1,012 posterior–anterior digital radiography chest images from individuals aged 15 years and older. The dataset was collected from five institutions across Thailand and was used as a reader-reference standard dataset for AI performance evaluation. Each image was independently interpreted by three NIOSH-certified B Readers from a pool of six, and the final label was assigned by majority consensus.

AI vs Radiologists

Pairwise agreement and Cohen's κ showed strong concordance between MOPH AI CXR and expert readers for tuberculosis detection.

- Baseline inter-radiologist agreement (B vs B) for TB:
    - Agreement: 0.9282
    - Cohen's κ: 0.8532
- AI vs radiologists at the manufacturer-recommended threshold (0.2290) for TB:
    - Agreement: 0.9200
    - Cohen's κ: 0.8357

For TB, AI–radiologist agreement was slightly lower than radiologist–radiologist agreement, but remained close to expert-reader consensus.

Diagnostic Performance Against Radiologist Consensus

For **TB**, at the manufacturer-recommended threshold (**0.2290**), **MOPH AI CXR** achieved the following diagnostic performance against the radiologist consensus reference standard:

- **Sensitivity:** 0.9504
- **Specificity:** 0.9087
- **PPV:** 0.9345
- **NPV:** 0.9305

WHO TPP Alignment

MOPH AI CXR can be tuned to meet different **WHO TB screening TPP** categories depending on the selected threshold:

- **High-sensitivity / high-specificity screening:** minimal criteria met at thresholds 0.08–0.46
- **High-sensitivity screening:** minimal criteria met at thresholds 0.05–0.46; optimal criteria met at thresholds 0.12–0.23
- **High-specificity screening:** minimal criteria met at thresholds 0.71–0.84; optimal criteria met at thresholds 0.71–0.88

Summary

MOPH AI CXR demonstrated strong agreement with expert readers, high sensitivity for TB detection, and excellent overall discrimination.

For TB, at the manufacturer-recommended threshold, the software achieved:

- **Sensitivity:** 0.9504
- **Specificity:** 0.9087
- **AUC:** 0.9785

At the recommended threshold, the software favors high sensitivity, which is desirable for screening and case-finding programs where missed TB cases are a major concern. Accordingly, the performance data support the use of MOPH AI CXR as an adjunctive software tool for TB screening, triage, and radiographic decision support, subject to professional clinical oversight and appropriate confirmatory diagnostic follow-up.

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