Test Report

Inspectra CXR Version 4

โดย บริษัท เพอเซ็ปทรา 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.

PERCEPTRA

Inspectra CXR

ส่งทดสอบโดย บริษัท เพอเซ็ปทรา จำกัด

98/155 หมู่ 9 ต.บางแก้ว อ.บางพลี จ.สมุทรปราการ 10540

This report evaluates an artificial intelligence system developed for tuberculosis screening on PA 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	February 20, 2026	**Revision Number**	1
Latest Report Date	March 3, 2026	**Approval Date**	March 5, 2026

1. Executive Summary

Inspectra CXR is an artificial intelligence (AI)-based software intended for the analysis of chest radiographs and developed by Perceptra Co., Ltd. Inspectra CXR is intended to function as an adjunctive software tool for TB screening, triage, and radiographic decision support.

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 the principal findings was high, with ICC > 0.90 and Cohen's κ > 0.75 for key targets including abnormalities, opacity, and TB, 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.

Inspectra CXR demonstrated high discriminative performance for TB detection, with an area under the receiver operating characteristic curve (AUC) of 0.9785. At the manufacturer-recommended operating point (threshold=0.433334), TB performance was sensitivity 0.9590, specificity 0.8782, positive predictive value (PPV) 0.9151, and negative predictive value (NPV) 0.9399. Agreement with B Readers at this operating point was 0.9173, with Cohen's κ of 0.8293, indicating performance approaching expert-reader consensus.

Based on the threshold analysis presented in this report, Inspectra 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 เพอเซ็ปทรา จำกัด
Address	98/155 หมู่ 9 ต.บางแก้ว อ.บางพลี จ.สมุทรปราการ 10540
Country	Thailand

2.2 Manufacturer / Developer

Manufacturer / Developer	บริษัท เพอเซ็ปทรา จำกัด
Address	98/155 หมู่ 9 ต.บางแก้ว อ.บางพลี จ.สมุทรปราการ 10540
Country	Thailand
Website	https://perceptra.tech

2.3 Software Identification

Proprietary Name	Inspectra CXR
Common/Usual Name	Inspectra CXR
Model/Version	Version 4
Software Category	AI-based Software as a Medical Device (SaMD)

2.4 Software Description

Inspectra CXR is an AI-based Software as a Medical Device (SaMD) intended for use in the computerized analysis of PA chest radiographs. The software provides algorithm-generated outputs to assist qualified healthcare professionals in the review of medical images by highlighting findings that may be associated with abnormalities visible on PA chest radiographs.

Inspectra CXR is intended to support clinical workflow, including image review and prioritization, and shall be used only as an aid to clinical assessment. Final interpretation, diagnosis, and patient management decisions remain the responsibility of the qualified physician or radiologist. Inspectra CXR may be deployed as a standalone software solution or integrated with relevant healthcare information systems, including radiology workflow systems, subject to the configured use environment and product specifications.

Inspectra CXR has received Thai FDA approval and Singapore HSA approval. Additional certifications include ISO 13485:2016, IEC 62304:2006, and IEC 60601-1:2005.

3. Chest X-Ray Images Representative of the Thai Population

A dataset of digital PA chest radiographs was curated from multiple hospitals across Thailand. This section describes the population characteristics, reader labelling protocol, distribution of radiographic findings, inter-rater reliability, and key limitations of the dataset. The dataset was curated as a reader-reference standard evaluation dataset to assess the performance of the software on chest radiographs representative of the intended use population in Thailand. None of the images included in this dataset were used for model training.

3.1 Dataset Population

The dataset, referred to as 123AAA, included 1,012 posterior–anterior (PA) digital radiography PA chest radiographs 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 because of 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 PA chest radiographs used for TB screening while controlling for confounding pathologies that may interfere with CAD interpretation.

Table 1. Dataset 123AAA overview and criteria

Dataset Reference	123AAA
Total Images	1,012
Age	≥15 years
Modality	Digital Radiography (DR) PA chest radiographs
Projection	Posterior–anterior (PA)
Key Exclusion Criteria	Images were excluded if the patient had positive HIV serology, opportunistic pulmonary infections, mixed pulmonary infections that could confound radiographic interpretation, or image quality below the predefined minimum threshold.

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, thereby ensuring coverage across geographic regions and patient demographics.

Table 2. Source institutions and sampling summary

Source Institutions	1. Songklanagarind Hospital (Southern Thailand) 2. Chiangrai Prachanukroh Hospital (Nortern 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 PA 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 PA chest radiographs for pneumoconiosis, which ensures a high standard of expertise in detecting subtle parenchymal and pleural abnormalities. Readers were blinded to clinical information, the interpretations of the other readers, and the software output. All image assessments were performed independently on the basis of the chest radiographs alone.

As shown in Table 3, all readings were performed independently, 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 standard for evaluating Inspecta CXR 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 PA 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 presents both the **Reader Count** and the **Consensus Count** for each finding. Reader Count refers to the total number of markings recorded across all readers and images, whereas Consensus Count refers to the number of images assigned a finding on the basis of majority agreement among the three readers.

Parenchymal opacities were categorized as **small opacity** or **large opacity**, and a combined category, **"Opacity (small/large)"**, was used to facilitate CAD evaluation. Other key findings, including **mass/nodule, cavity, fibrosis, calcification, pleural effusion, pleural thickening, pneumothorax, hilar adenopathy, mediastinal adenopathy, and tuberculosis-related changes**, are also summarized. Among these, **active tuberculosis** and **indeterminate tuberculosis** were combined to support a more 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 the inter-rater reliability measures for each radiographic finding.

Findings such as **overall abnormalities, small opacity, large opacity, combined opacity, and tuberculosis** demonstrated high agreement, with **ICC values greater than 0.9** and **Cohen's κ values greater than 0.75** for the principal findings, indicating strong consensus among readers. Other findings with lower prevalence, such as **mediastinal adenopathy** and **pneumothorax**, demonstrated lower or less stable agreement estimates, reflecting the challenges associated with detecting rare or subtle radiographic 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 **hospital-based settings**. Although some images were sourced from **community-based active case finding**, the dataset 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 from that observed in the present dataset. In addition, certain findings, such as **pneumothorax** and **mediastinal adenopathy**, had relatively few occurrences, thereby limiting the reliability and interpretability of inter-rater agreement metrics for those findings.

4. Evaluation Results

4.1 Label Mapping

To standardize comparisons between **Inspectra CXR** outputs and radiologist interpretations, the AI output classes were mapped to the corresponding radiologist-defined findings. Table 6 summarizes the mapping and aggregation rules applied. For **TB**, the AI output class **"TB"** was mapped indirectly to the radiologist-assessed categories of **active tuberculosis** and **indeterminate tuberculosis**.

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

AI Output	Findings	Mapping Rule
Lung Opacity	Small Opacity and Large Opacity	Indirect Mapping
Mass	Mass or Nodule	Indirect Mapping
Pleural Effusion	Pleural Effusion	Direct Mapping
TB	Active TB and Indeterminate TB	Indirect Mapping

4.2 Results Comparison: AI vs Radiologist

Inspectra CXR generates an output score ranging from **0.00 to 1.00** for each AI output class (Table 7). The **manufacturer-recommended thresholds** were used as the primary operating points for comparison with radiologist interpretations. For **TB**, the manufacturer-recommended threshold was **0.433334**.

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

AI Output	Score Range	Threshold
Lung Opacity	0.00–1.00	0.40
Mass	0.00–1.00	0.50
Pleural Effusion	0.00–1.00	0.55
TB	0.00–1.00	0.433334

Agreement between AI outputs and radiologist interpretations was assessed using **pairwise agreement** and **Cohen's κ**. Table 8 summarizes the agreement between **Inspectra CXR** and the B Readers at the manufacturer-recommended thresholds and provides **inter-radiologist agreement** as a reference comparator.

Overall, Inspectra CXR demonstrated strong agreement with B Reader interpretations across the evaluated output classes. For **Lung Opacity**, the pairwise agreement between AI and radiologists was **0.9384**, with a **Cohen's κ of 0.8638**, which was slightly higher than the corresponding inter-radiologist agreement. For **Mass**, the pairwise agreement between AI and radiologists was **0.7256**, with a **Cohen's κ of 0.3335**. For **Pleural Effusion**, the pairwise agreement between AI and radiologists was **0.8999**, with a **Cohen's κ of 0.6372**. For **TB**, the pairwise agreement between AI and radiologists

was **0.9173**, with a **Cohen's κ of 0.8293**. For reference, inter-radiologist agreement for TB showed a pairwise agreement of **0.9282** and a **Cohen's κ of 0.8532**.

Although AI–radiologist agreement for TB was slightly lower than radiologist–radiologist agreement, the difference was modest, indicating that the performance of Inspectra CXR approached expert-reader consensus at the manufacturer-recommended threshold.

Table 8. Pairwise agreement measures for AI vs radiologists

AI Output	Threshold	Dataset Consensus	Model Positives	Pairwise Agreement		Cohen's Kappa	
				B vs B	AI vs B	B vs B	AI vs B
Lung Opacity	0.40	645	674	0.9289	0.9384	0.8344	0.8638
Mass	0.50	189	343	0.7582	0.7256	0.3088	0.3335
Pleural Effusion	0.55	141	188	0.9282	0.8999	0.7079	0.6372
TB	0.433334	585	613	0.9282	0.9173	0.8532	0.8293

4.3 Diagnostic Performance

Diagnostic performance was assessed against the **radiologist consensus reference standard**, defined by the **majority rule of three B Readers**. Table 9 summarizes the key performance metrics at the selected thresholds, including **true positives (TP)**, **false positives (FP)**, **false negatives (FN)**, **true negatives (TN)**, and the derived measures of **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
Lung Opacity	0.40	639	35	6	332	0.9907	0.9046	0.9481	0.9822
Mass	0.50	138	205	51	618	0.7302	0.7509	0.4023	0.9238
Pleural Effusion	0.55	117	71	24	800	0.8298	0.9185	0.6223	0.9709
TB	0.433334	561	52	24	375	0.9590	0.8782	0.9151	0.9399

4.4 Sensitivity and Specificity across Thresholds

To evaluate the diagnostic performance of **Inspectra CXR** across a range of operating points, **sensitivity** and **specificity**were calculated over a wide range of thresholds for the **TB** output class. As shown in Table 10, lower thresholds were associated with very high sensitivity but lower specificity, indicating that most true TB cases were detected at the expense of a higher number of false-positive results. Conversely, higher thresholds were associated with improved specificity but reduced sensitivity, reflecting stricter criteria for the detection of TB-related abnormalities.

These findings demonstrate the expected trade-off between sensitivity and specificity and highlight the importance of selecting an appropriate threshold according to the intended clinical application, whether the priority is to **maximize case detection in screening settings** or to **reduce false-positive results in more specific decision-support workflows**.

Table 10. Sensitivity and specificity across thresholds for Tuberculosis

Threshold	Sensitivity	Specificity	Threshold	Sensitivity	Specificity
0.01	1.0000	0.0000	0.26	0.9812	0.8103
0.02	1.0000	0.0000	0.27	0.9795	0.8126
0.03	1.0000	0.2600	0.28	0.9795	0.8197
0.04	1.0000	0.3934	0.29	0.9778	0.8220
0.05	1.0000	0.5082	0.30	0.9778	0.8220
0.06	1.0000	0.5222	0.31	0.9744	0.8337
0.07	1.0000	0.5386	0.32	0.9726	0.8384
0.08	1.0000	0.5621	0.33	0.9726	0.8454
0.09	1.0000	0.5831	0.34	0.9726	0.8525
0.10	1.0000	0.6066	0.35	0.9726	0.8595
0.11	1.0000	0.6464	0.36	0.9709	0.8618
0.12	1.0000	0.6792	0.37	0.9709	0.8642
0.13	1.0000	0.7143	0.38	0.9692	0.8665
0.14	0.9983	0.7307	0.39	0.9675	0.8665
0.15	0.9949	0.7494	0.40	0.9641	0.8689
0.16	0.9932	0.7518	0.41	0.9641	0.8735
0.17	0.9915	0.7611	0.42	0.9641	0.8782
0.18	0.9915	0.7611	0.43	0.9607	0.8782
0.19	0.9897	0.7728	0.44	0.9573	0.8782
0.20	0.9880	0.7752	0.45	0.9556	0.8782
0.21	0.9846	0.7799	0.46	0.9487	0.8782
0.22	0.9846	0.7892	0.47	0.9436	0.8782
0.23	0.9846	0.7963	0.48	0.9436	0.8852
0.24	0.9829	0.8033	0.49	0.9436	0.8852
0.25	0.9812	0.8103	0.50	0.9419	0.8876

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

Threshold	Sensitivity	Specificity	Threshold	Sensitivity	Specificity
0.51	0.9402	0.8899	0.76	0.8650	0.9602
0.52	0.9402	0.8923	0.77	0.8496	0.9649
0.53	0.9402	0.8923	0.78	0.8393	0.9719
0.54	0.9385	0.8923	0.79	0.8222	0.9742
0.55	0.9316	0.8970	0.80	0.8120	0.9789
0.56	0.9265	0.8970	0.81	0.8017	0.9813
0.57	0.9248	0.9016	0.82	0.7880	0.9813
0.58	0.9248	0.9040	0.83	0.7709	0.9883
0.59	0.9179	0.9110	0.84	0.7573	0.9906
0.60	0.9162	0.9157	0.85	0.7316	0.9906
0.61	0.9145	0.9157	0.86	0.7026	0.9930
0.62	0.9145	0.9204	0.87	0.6752	0.9953
0.63	0.9128	0.9227	0.88	0.6239	0.9953
0.64	0.9128	0.9274	0.89	0.5590	0.9953
0.65	0.9128	0.9274	0.90	0.4581	0.9953
0.66	0.9060	0.9297	0.91	0.3504	0.9953
0.67	0.9043	0.9344	0.92	0.1590	0.9977
0.68	0.9026	0.9415	0.93	0.0017	1.0000
0.69	0.8957	0.9415	0.94	0.0000	1.0000
0.70	0.8940	0.9508	0.95	0.0000	1.0000
0.71	0.8872	0.9508	0.96	0.0000	1.0000
0.72	0.8838	0.9532	0.97	0.0000	1.0000
0.73	0.8769	0.9555	0.98	0.0000	1.0000
0.74	0.8769	0.9578	0.99	0.0000	1.0000
0.75	0.8684	0.9578	0.76	0.8650	0.9602

Note: The recommended threshold for Inspectra CXR is 0.433334.

5. WHO Target Product Profile (TPP) Performance Analysis

The **World Health Organization (WHO)** defines target product profiles (TPPs) for TB screening tests to support the selection of tests appropriate for different populations and use settings. These targets specify both **minimum acceptable** and **optimal** performance criteria for different types of screening tests.

There are three main categories of TB screening tests:

1. **High-sensitivity, high-specificity screening test**
   This type of test is intended to detect most TB cases while also maintaining a relatively low false-positive rate. Such a performance profile is particularly useful in general population screening, where the objective is to identify TB cases without over-referring healthy individuals.
2. **High-sensitivity screening test**
   This type of test prioritizes the detection of most TB cases by maintaining high sensitivity, although this may occur at the expense of lower specificity and a higher false-positive rate. This profile is useful in settings where missing TB cases is considered more harmful than over-referral.
3. **High-specificity screening test**
   This type of test prioritizes specificity in order to minimize false-positive classifications, although sensitivity may be lower. This profile is useful in settings where limiting unnecessary referral or confirmatory testing is particularly important.

These WHO targets help ensure that TB screening tests detect a sufficient proportion of TB cases without overwhelming the health system with excessive false-positive results, thereby supporting 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: 0.24–0.68)		*Conditions Not Met*	
High-sensitivity screening test	90%	60%	95%	85%
	Conditions Met (Thresholds: 0.10–0.68)		*Conditions Met (Thresholds: 0.34–0.45)*	
High-specificity screening test	60%	98%	70%	98%
	Conditions Met (Thresholds: 0.81–0.88)		*Conditions Met (Thresholds: 0.81–0.86)*	

Notes: Minimal accuracy refers to the minimum acceptable performance target, including in populations where TB prevalence is low (approximately 1%). Optimal accuracy refers to the preferred performance target, including in very low-prevalence populations (approximately 0.25%). For high-specificity tests, the minimum acceptable sensitivity may be lower because the primary objective is to minimize false-positive results.

Based on the threshold analysis presented in this report, the performance of **Inspectra CXR** may satisfy different WHO TPP categories at different operating points. The manufacturer-recommended threshold should therefore be interpreted in the context of the intended use. Lower thresholds tend to favor sensitivity and may be more appropriate for screening-oriented applications, whereas higher thresholds favor specificity and may be more appropriate for more selective referral or confirmatory decision-support use.

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

In alignment with the WHO policy statement on the use of computer-aided detection (CAD) software for tuberculosis screening, the clinical performance of Inspectra CXR was compared with that of expert radiologists.

The radiologists' sensitivity was 0.9018 (95% CI: 0.8819–0.9218) and specificity was 0.9408 (95% CI: 0.9229–0.9586), as summarized in Table 12. Using these results as the comparator, the corresponding sensitivity and specificity of Inspectra CXR were evaluated across different thresholds, as illustrated by the receiver operating characteristic (ROC) curves shown in Figures 1 and 2. The area under the curve (AUC) was 0.9785, indicating excellent overall discrimination between TB and non-TB cases based on the radiologist consensus reference standard.

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

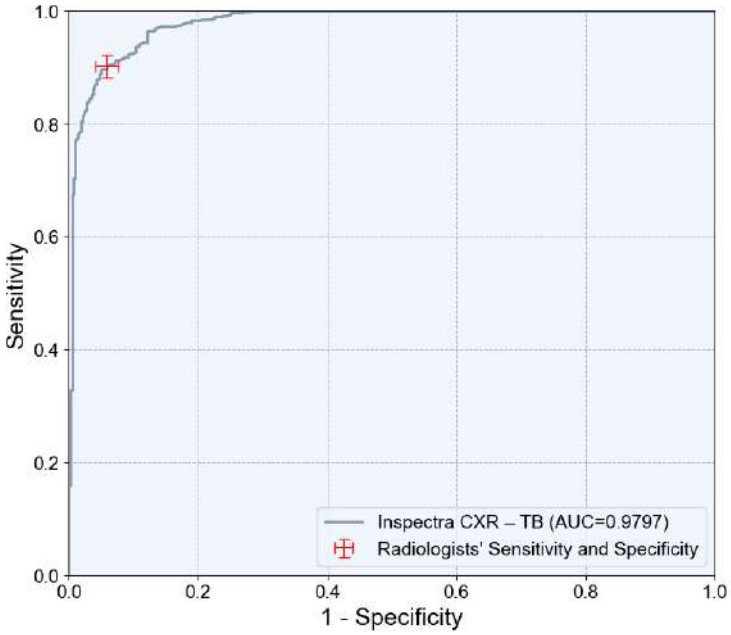

Figure 1. ROC curves showing the performance of Inspectra CXR , compared with radiologists.

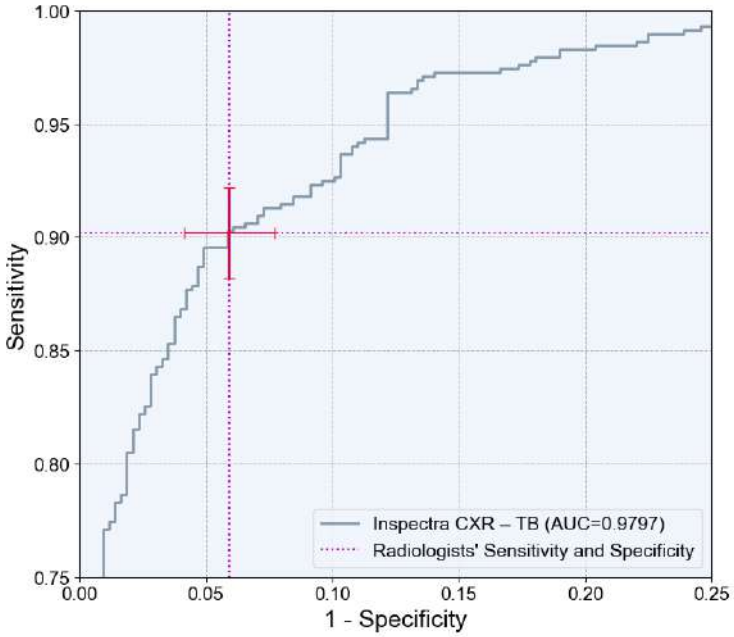

Figure 2. Partial ROC curves at 75-100% specificity showing the performance of Inspectra CXR , compared with radiologists.

For Inspectra CXR, higher TB scores were generally observed among TB cases, whereas scores for non-TB cases were concentrated toward the lower end of the distribution. Figure 3 presents a histogram showing the distribution of the Inspectra CXR TB scores for TB and non-TB cases, as defined by the radiologist consensus reference standard.

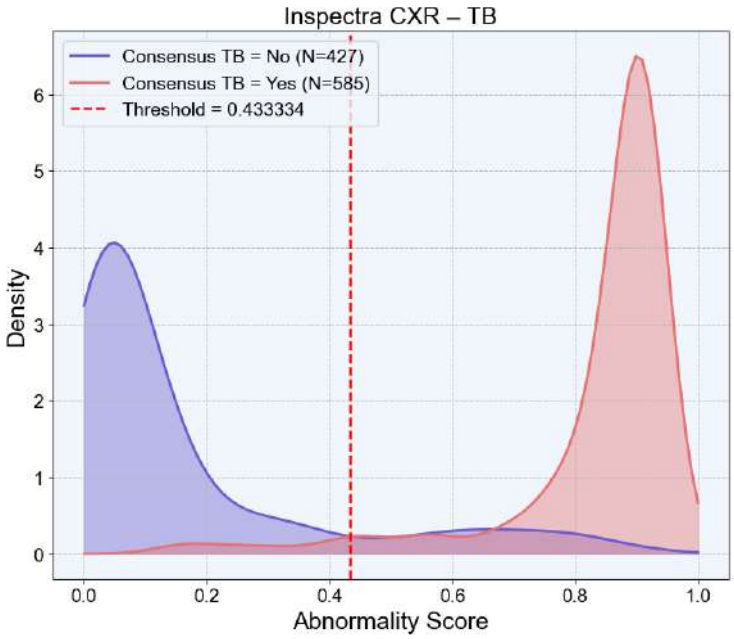

Figure 3. Histogram plot of TB scores for Inspectra CXR for TB and non-TB cases (as per the radiologist consensus)

7. Conclusion

Software & Sponsor

- **Inspectra CXR** is an artificial intelligence (AI)-based software for chest radiograph analysis developed by **Perceptra Co., Ltd.**

Dataset

- Performance was evaluated using a curated Thai chest radiograph dataset (Ref. 123AAA) comprising 1,012 PA 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 Inspectra CXR and expert readers.
- **Baseline inter-radiologist agreement (B vs B):**
  - **Lung Opacity:** agreement **0.9289**, κ **0.8344**
  - **Mass:** agreement **0.7582**, κ **0.3088**
  - **Pleural Effusion:** agreement 0**.9282**, κ **0.7079**
  - **TB:** agreement **0.9282**, κ **0.8532**
- **AI vs radiologists at the manufacturer-recommended threshold:**
  - **Lung Opacity:** agreement **0.9384**, κ **0.8638**
  - **Mass:** agreement **0.7256**, κ **0.3335**
  - **Pleural Effusion:** agreement **0.8999**, κ **0.6372**
  - **TB:** agreement **0.9173**, κ **0.8293**
- For TB, AI–radiologist agreement was slightly lower than radiologist–radiologist agreement, but remained close to expert-reader consensus.

Diagnostic Performance Against Radiologist Consensus

- **Lung Opacity:** Sensitivity **0.9907**, Specificity **0.9046**, PPV **0.9481**, NPV **0.9822**
- **Mass:** Sensitivity **0.7302**, Specificity **0.7509**, PPV **0.4023**, NPV **0.9238**
- **Pleural Effusion:** Sensitivity **0.8298**, Specificity **0.9185**, PPV **0.6223**, NPV **0.9709**
- **TB :** Sensitivity **0.9590**, Specificity **0.8782**, PPV **0.9151**, NPV **0.9399**

WHO TPP Alignment

- Inspectra 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 **0.24–0.68**
  - **High-sensitivity screening:** minimal criteria met at thresholds **0.10–0.68**; optimal criteria met at **0.34–0.45**
  - **High-specificity screening:** minimal criteria met at **0.81–0.88**; optimal criteria met at **0.81–0.86**

Summary

- Inspectra 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:** 0.9590
  - **Specificity:** 0.8782
  - **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.

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