



Test Report

MAI-CXRv4-Object Version 1.0.0

โดย บริษัท แมเนจ เอไอ โซลูชั่น จำกัด

รายงานผลการทดสอบ

โดยราชวิทยาลัยรังสีแพทย์แห่งประเทศไทย

ทดสอบใช้กับภาพรังสีทรวงอก ในกรณี



- คัดกรอง (screening) วัณโรคปอด

รายงานผลการทดสอบปัญญาประดิษฐ์ในภาพถ่ายรังสีทรวงอก

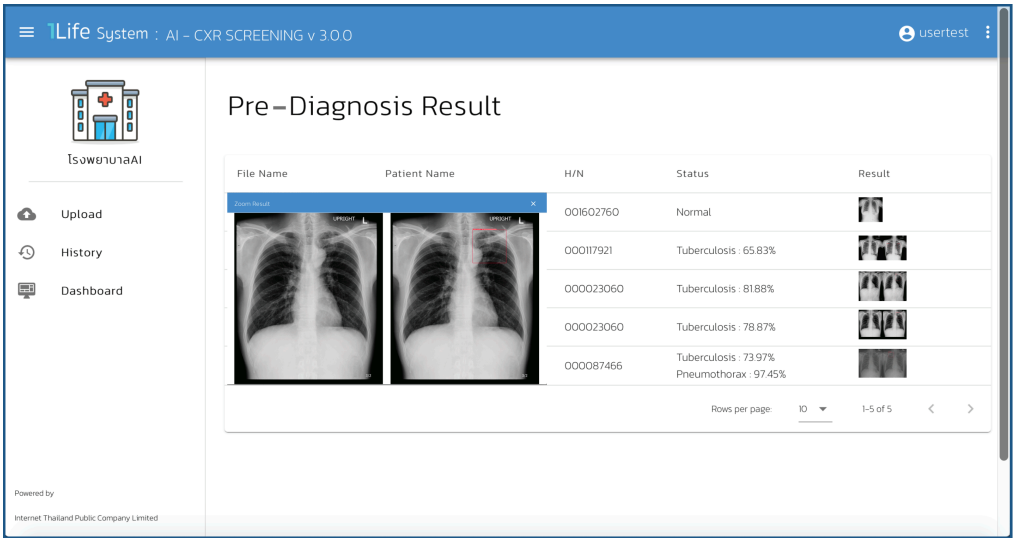
บริษัทที่ยื่นทดสอบ

หน่วยงาน	บริษัท แมเนจ เอไอ โซลูชัน จำกัด	 
ที่อยู่	1768 อาคารไทยซัมมิท ทาวเวอร์ ชั้น 16 ถนนเพชรบุรีตัดใหม่ แขวงบางกะปิ เขตห้วยขวาง กรุงเทพมหานคร 10310	
ชื่อผู้ติดต่อ	Mr.Sarun Angamnuaysiri Tel.065-521-4419 Mr.Suphanat Boonbundransuk Tel. 065-527-2073	

บริษัทผู้พัฒนา

บริษัท	บริษัท แมเนจ เอไอ โซลูชัน จำกัด	 
ที่อยู่	1768 อาคารไทยซัมมิท ทาวเวอร์ ชั้น 16 ถนนเพชรบุรีตัดใหม่ แขวงบางกะปิ เขตห้วยขวาง กรุงเทพมหานคร 10310	
ประเทศ	ไทย	
เว็บไซต์	https://manageai.co.th	

ระบบที่ยื่นทดสอบ

ชื่อระบบ	MAI-CXRv4-Object (Tuberculosis, Cardiomegaly, Pneumothorax)
เวอร์ชัน	V1.0.0 (Build 032023)
ข้อมูลระบบ	 <p>Figure 1 Output form Model Contain bounding boxes and confidence</p> <p>ระบบคัดกรองผู้ป่วยจากภาพถ่ายรังสีทรวงอก สำหรับวิเคราะห์และตรวจสอบความเสี่ยงในการเกิดโรคต่าง ๆ เช่น วัณโรค โดยนำเข้าข้อมูลภาพถ่ายรังสีทรวงอก</p>

	<p>(CXR) จากนั้นระบบ AI จะวิเคราะห์ภาพ CXR รวมถึงลักษณะทางกายภาพของการเกิดโรคต่าง ๆ เพื่อใช้เป็นเครื่องมือเสริมสำหรับแพทย์ในการคัดกรองผู้ป่วยเบื้องต้น ก่อนยื่นผลการอ่านผล CXR ของผู้ป่วยด้วยแพทย์รังสีอีกครั้ง</p> <p>In this project study, 13504 Image of Tuberculosis confirmed patients imaged, 15427 Image of Cardiomegaly confirmed patients imaged, 12324 Image of Pneumothorax confirmed patients imaged, and more than 100000 matching CXRs of normal patients imaged were collected in Thailand hospital and used to build a deep learning 3 model of YOLO object detection for detecting patients positive in 3 diseases. You Only Look Once (YOLO) proposes using an end-to-end neural network that makes predictions of bounding boxes and class probabilities all at once and is enhanced by process of generating new transformed versions of images from the given image dataset to increase its diversity and process of accurately identifying regions and boundaries of the lung field from surrounding thoracic tissue. And used confidence threshold above 0.5 to detect positive case of diseases</p>
--	--

ชุดข้อมูลที่ใช้ทดสอบ

รหัสชุดข้อมูล	1A
จำนวนภาพถ่าย	300 ภาพ

คุณลักษณะของข้อมูลที่ใช้ทดสอบ

จำนวนภาพถ่าย 300 ภาพได้ถูกสุ่มมาจากชุดข้อมูลทดสอบจำนวน 1,500 ภาพซึ่งได้รับการสนับสนุนจาก โรงพยาบาลสงขลานครินทร์ จังหวัดสงขลา, โรงพยาบาลเชียงรายประชารักษ์ จังหวัดเชียงราย, โรงพยาบาลอุดรธานี จังหวัดอุดรธานี, โรงพยาบาลสุทธาเวช จังหวัดมหาสารคาม, และกองวัณโรค กรมควบคุมโรค กระทรวงสาธารณสุข โดยทุกภาพจะถูกอ่านโดยรังสีแพทย์ B Reader จำนวน 3 ท่าน

คุณลักษณะของภาพถ่ายรังสีในโครงการ:

- ภาพถ่ายรังสีทรวงอกของผู้ป่วยที่มีอายุตั้งแต่ 15 ปี ด้วยเครื่องเอกซเรย์คอมพิวเตอร์
- ไม่มีภาพถ่ายจากผู้ป่วยที่มีสถานะ HIV Serology เป็นบวก
- ไม่มีภาพถ่ายจากผู้ป่วยที่มีการติดเชื้อที่ปอดแบบฉวยโอกาสอื่น ๆ หรือการติดเชื้อร่วม เช่น การติดเชื้อไมโครแบคทีเรีย อีสโตพลาสโมซิส ไคริปโตคอคโคซิส เมลิออยโดซิส และแอกติโนมัยโคซิส

การวัดความตรงภายในในชุดข้อมูล ใช้ตัววัดดังต่อไปนี้

- Pairwise Agreement ค่าเฉลี่ยของความเหมือนกันกันระหว่างแต่ละคู่ของรังสีแพทย์
- Intraclass Agreement (ICC) ค่าเฉลี่ยของ Pearson's correlation แบบ ICC(2,3) เมื่อมีรังสีแพทย์ 3 ท่าน อ่านภาพถ่ายรังสีแบบสุ่ม
- การวิเคราะห์ความตรงระหว่างผู้ประเมินโดยใช้สถิติ Pairwise Cohen's Kappa และ Fless' Kappa

จำนวนรอยโรคในชุดข้อมูล

ตารางที่ 1 จำนวนรอยโรคที่ระบุโดยรังสีแพทย์ ในภาพถ่ายรังสีที่ได้รับการวินิจฉัยไม่เป็นวัณโรคปอด ($N_{\text{Non-TB}}$) และในภาพถ่ายรังสีได้รับยืนยันผลวัณโรคปอดโดยผลตรวจเสมหะย้อมเชื้อหรือผลตรวจเพาะเชื้อ (N_{TB}) รวมถึงจำนวนรอยโรคที่ได้รับการยืนยันว่ามีรอยโรคจากรังสีแพทย์ส่วนใหญ่ ($N_{\text{Consensus}}$) ในชุดข้อมูล 1A จำนวน 300 ภาพ แต่ละภาพมีการอ่านโดยรังสีแพทย์ “B” Reader จำนวน 3 รายแบบสุ่ม

Finding		$N_{\text{Non-TB}}$	N_{TB}	$N_{\text{Consensus}}$	
Abnormality		53	448	158	
Small opacity		33	374	135	
	Primary nodular	1	274	98	
	Primary reticular	32	87	21	
	Secondary nodular	9	224	76	
	Secondary reticular	24	125	30	
Large opacity		2	349	119	
Mass/nodule		2	114	29	
Cavity		0	242	81	
Fibrosis		7	196	58	
Calcification		5	61	12	
Pleural effusion		0	99	34	
Pleural thickening		7	122	39	
Pneumothorax		0	1	0	
Hilar adenopathy		2	87	18	
Mediastinal adenopathy		1	30	7	
Consistent with tuberculosis		0	422	144	
	Active Tuberculosis	0	387	136	
		Patchy infiltration	0	265	99
		Cavity with surrounding consolidation	0	210	73
		Unilateral hilar/paratracheal lymph node enlargement	0	39	5
		Pleural effusion	0	52	15
		Miliary nodules	0	74	20
		Indeterminate tuberculosis	0	35	6
	Reticulonodular infiltration	0	22	4	
	Destroyed lung or bronchiectasis	0	3	0	
Inconsistent with tuberculosis		450	28	156	

ความตรงภายในของชุดข้อมูล

ตารางที่ 2 ความตรงภายในระหว่างผู้ประเมิน (Inter-rater Reliability) ของแต่ละรอยโรคในชุดข้อมูล 1A จำนวน 300 ภาพ ของรังสีแพทย์ “B” Reader จำนวน 3 ราย ซึ่งวัดโดยค่าสถิติ Pairwise Agreement, ICC(2,3), Cohen’s kappa และ Fleiss’ kappa

Finding	Agreement	ICC	Cohen’s	Fleiss’
Abnormality	0.8956	0.9181	0.7884	0.7884
Small opacity	0.8356	0.8584	0.6681	0.6681
Primary nodular	0.8422	0.8357	0.6281	0.6282
Primary reticular	0.7911	0.2301	0.0892	0.0897
Secondary nodular	0.7222	0.5352	0.2751	0.2761
Secondary reticular	0.7578	0.3006	0.1260	0.1233
Large opacity	0.8800	0.8993	0.7477	0.7478
Mass/nodule	0.8378	0.5365	0.2777	0.2776
Cavity	0.8600	0.8449	0.6446	0.6439
Fibrosis	0.7978	0.6866	0.4212	0.4212
Calcification	0.8889	0.4027	0.1829	0.1825
Pleural effusion	0.9578	0.9163	0.7838	0.7844
Pleural thickening	0.8711	0.7319	0.4762	0.4752
Pneumothorax	0.9978	0.0000	0.0000	0.0000
Hilar adenopathy	0.8644	0.4867	0.2395	0.2394
Mediastinal adenopathy	0.9556	0.6008	0.3130	0.3318
Consistent with tuberculosis	0.9644	0.9751	0.9286	0.9286
Active Tuberculosis	0.9178	0.9373	0.8323	0.8323
Patchy infiltration	0.8178	0.7940	0.5613	0.5614
Cavity with surrounding consolidation	0.8333	0.7755	0.5333	0.5342
Unilateral hilar/paratracheal lymph node enlargement	0.9244	0.2289	0.0846	0.0887
Pleural effusion	0.9444	0.7431	0.4941	0.4897
Miliary nodules	0.8978	0.5905	0.3212	0.3227
Indeterminate tuberculosis	0.9444	0.5114	0.2588	0.2568
Reticulonodular infiltration	0.9600	0.3705	0.1947	0.1613
Destroyed lung or bronchiectasis	0.9933	0.0067	0.0015	0.0033
Inconsistent with tuberculosis	0.9644	0.9751	0.9286	0.9286

ตารางที่ 3 การตีความค่า ICC และ Kappa ตาม Landis and Koch (1977)¹

ICC/Kappa Statistic	Strength of Agreement
<0.00	Poor
0.00 – 0.20	Slight
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Substantial
0.81 – 1.00	Almost Perfect

¹Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. In *Biometrics* (Vol. 33, Issue 1, p. 159). JSTOR. <https://doi.org/10.2307/2529310>

ผลการทดสอบ

การวัดความตรงภายนอกนั้นวัดโดยใช้ Pairwise Agreement หรือค่าเฉลี่ยของความเหมือนกันระหว่างคู่ของแต่ละรังสีแพทย์และปัญญาประดิษฐ์และ Pairwise Cohen's Kappa หรือค่าเฉลี่ยของค่าสถิติ Cohen's Kappa ระหว่างคู่ของแต่ละรังสีแพทย์และปัญญาประดิษฐ์ เพื่อเปรียบเทียบระหว่างรังสีแพทย์กับปัญญาประดิษฐ์ ("B" vs AI) และรังสีแพทย์ด้วยกัน ("B" vs "B")

ตารางที่ 4 การวิเคราะห์ความตรงกันภายนอกระหว่างรังสีแพทย์และปัญญาประดิษฐ์

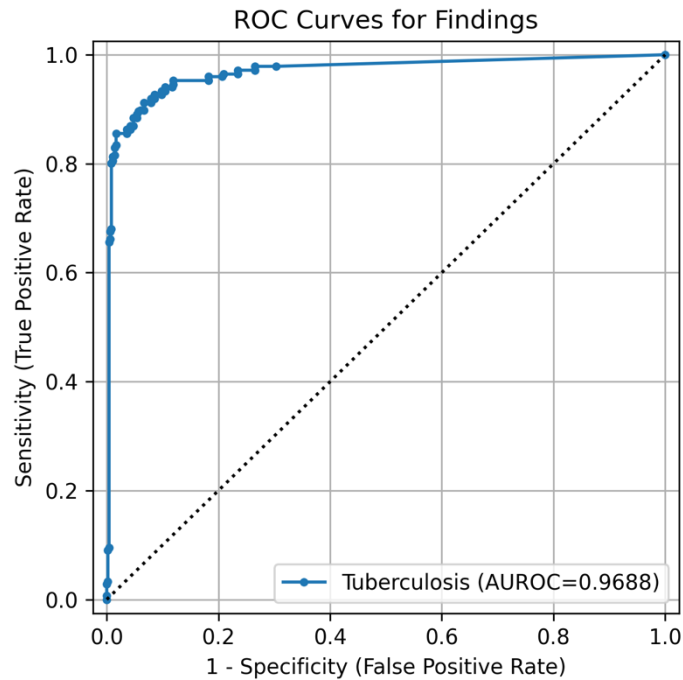
Finding	N	Threshold	Pairwise Agreement		Cohen's Kappa	
			"B" vs "B"	"B" vs AI	"B" vs "B"	"B" vs AI
Tuberculosis	422	0.50	0.9635	0.9167	0.9261	0.8324

ในการวัดประสิทธิภาพในการวินิจฉัยของแต่ละรอยโรคนั้นใช้เกณฑ์ ความไว (Sensitivity), ความจำเพาะ (Specificity), อัตราส่วนการทำนายผู้ป่วยที่เป็นโรคที่ถูกต้อง (Positive Prediction Rate, PPR) และอัตราส่วนการทำนายผู้ป่วยที่ไม่เป็นโรคที่ถูกต้อง (Negative Prediction Rate, NPR) ด้วยเกณฑ์การวินิจฉัย (Diagnostic Threshold) ที่ระบุโดยผู้ผลิต รวมถึงพื้นที่ใต้กราฟ ROC เช่นกัน

ตารางที่ 5 ประสิทธิภาพในการวินิจฉัยของแต่ละรอยโรคของปัญญาประดิษฐ์เปรียบเทียบกับรังสีแพทย์

Finding	N	Threshold	Specificity	Sensitivity	PPV	NPV	AUROC
Tuberculosis	422	0.50	0.9331	0.8981	0.9221	0.9121	0.9688

รูปภาพที่ 1 เส้นโค้ง ROC ของแต่ละรอยโรค



บทวิจารณ์

สำหรับการทดสอบ MAI-CXRv4-Object (Build 032023) เวอร์ชัน 1.0.0 ทางผู้ผลิตได้ส่งรายงานรอยโรคมา 1 รายการ คือ Tuberculosis

จากตารางที่ 6 เมื่อเทียบ Pairwise Agreement และ Cohen's Kappa ระหว่างรังสีแพทย์กับปัญญาประดิษฐ์ (“B” vs AI) และรังสีแพทย์ด้วยกัน (“B” vs “B”) แล้ว MAI-CXRv4-Object เวอร์ชัน 1.0.0 (Build 032023) ในส่วนของการวินิจฉัยวัณโรค (Tuberculosis) ซึ่งปัญญาประดิษฐ์มีคะแนนน้อยกว่ารังสีแพทย์อยู่ 4.68% สำหรับ Pairwise Agreement และ คะแนนน้อยกว่ารังสีแพทย์อยู่ 9.37% สำหรับ Cohen's Kappa (N = 422)

ตารางที่ 6 ความแตกต่างระหว่าง Pairwise Agreement และ Cohen's Kappa

Finding	Pairwise Agreement			Cohen's Kappa		
	B vs “B”	“B” vs AI	Diff	“B” vs “B”	“B” vs AI	Diff
Tuberculosis	0.9635	0.9167	-4.68%	0.9261	0.8324	-9.37%

ในส่วนของการคัดกรองวัณโรคปอด MAI-CXRv4-Object เวอร์ชัน 1.0.0 (Build 032023) เมื่อวิเคราะห์บนชุดข้อมูล 1A จำนวน 300 ภาพ พบว่าประสิทธิภาพในการวินิจฉัยวัณโรคใกล้เคียงกับรังสีแพทย์ และมีค่าพื้นที่ใต้กราฟตัวรับ (AUROC) อยู่ที่ 0.9688 ค่าความไว (Sensitivity) อยู่ที่ 0.8981 และค่าความจำเพาะ (Specificity) อยู่ที่ 0.9331 ที่ Threshold เป็น 0.50

อ้างอิงถึงเกณฑ์ [The Target Product Profiles \(TPPs\) for a rapid non-sputum-based biomarker test for tuberculosis detection](#) ของ World Health Organisation (WHO) ตามตารางที่ 7 จะเห็นได้ว่าแต่ละเกณฑ์การวัดมีค่า Sensitivity และ Specificity ที่แตกต่างกัน

ตารางที่ 7 TPP for a rapid non-sputum-based biomarker test for tuberculosis detection

	Minimal Requirements		Optimal Requirements	
	Sensitivity	Specificity	Sensitivity	Specificity
Smear-replacement test	Overall >80%	98%	Overall >95%	98%
	Positive >99%		Positive >99%	
	Negative >60%		Negative >68%	
Non-sputum based biomarker test	Overall >65%	98%	Positive >98%	98%
	Positive >98%		Negative >68%	
Triage test	90%	70%	95%	80%

เอกสารอ้างอิง: https://academic.oup.com/jid/article/211/suppl_2/S29/2490781

ผลลัพธ์ของการคัดกรองวัณโรคปอด MAI-CXRv4-Object เวอร์ชัน 1.0.0 (Build 032023) ใน Threshold ต่าง ๆ เมื่อเทียบกับเกณฑ์ WHO TPP โดยกำหนดให้เป็นค่า Threshold ที่สูงที่สุดที่จะทำให้ค่า Specificity นั้นมีความใกล้เคียงกับ WHO TPP มากที่สุด ผลลัพธ์ที่ได้เป็นดังตารางที่ 8

ตารางที่ 8 ค่า Sensitivity และ Specificity ที่ Threshold ต่าง ๆ ตามเกณฑ์ WHO TPP

Threshold	Sensitivity	Specificity
0.6257	0.8554	0.9833
0.5000	0.8981	0.9331
0.1264	0.9597	0.8180
0.0628	0.9787	0.7343

ทั้งนี้เมื่อนำผลลัพธ์ที่ได้เปรียบเทียบกับเกณฑ์ WHO TPP พบว่า MAI-CXRv4-Object นั้นผ่านเกณฑ์ Triage test ทั้งใน Minimal Requirements และ Optimal Requirements และผ่านเกณฑ์ Smear-replacement test และ Non-sputum based biomarker test สำหรับเกณฑ์ Minimal Requirements ซึ่งผลการทดสอบ ได้ถูกสรุปในตารางที่ 9

ตารางที่ 9 ผลการทดสอบการคัดกรองวัณโรคปอด MAI-CXRv4-Object ตามเกณฑ์ WHO TPP

	Minimal Requirements	Optimal Requirements
Smear-replacement test	ผ่าน	ไม่ผ่าน
Non-sputum based biomarker test	ผ่าน	ไม่ผ่าน
Triage test	ผ่าน	ผ่าน