



Test Report

LUNIT INSIGHT CXR Version 3.1.5.0

โดย Lunit Inc

รายงานผลการทดสอบ

โดยราชวิทยาลัยรังสีแพทย์แห่งประเทศไทย

ทดสอบใช้กับภาพรังสีทรวงอก ในกรณี

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

รายงานผลการทดสอบปัญญาประดิษฐ์ในภาพถ่ายรังสีทรวงอก

บริษัทที่ยื่นทดสอบ

หน่วยงาน	Lunit Inc	
ที่อยู่	15F, 27, Teheran-ro 2-gil, Gangnam-gu, Seoul, Republic of Korea	
ชื่อผู้ติดต่อ	Jo Kyung-sik	

บริษัทผู้พัฒนา

บริษัท	Lunit Inc	
ที่อยู่	15F, 27, Teheran-ro 2-gil, Gangnam-gu, Seoul, Republic of Korea	
ประเทศ	เกาหลีใต้	
เว็บไซต์	https://www.lunit.io	

ระบบที่ยื่นทดสอบ

ชื่อระบบ	Lunit INSIGHT CXR																											
เวอร์ชัน	3.1.5.0																											
ข้อมูลระบบ	<p>Product specifications Information source: Vendor Last updated: Feb. 19, 2023</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th colspan="2" style="background-color: #f2f2f2;">General</th> </tr> </thead> <tbody> <tr> <td>Product name</td> <td>Lunit INSIGHT CXR</td> </tr> <tr> <td>Subspeciality</td> <td>Chest</td> </tr> <tr> <td>Modality</td> <td>X-ray</td> </tr> <tr> <td>Disease targeted</td> <td>Atelectasis, calcification, cardiomegaly, consolidation, fibrosis, mediastinal widening, nodule, pleural effusion, pneumothorax, pneumoperitoneum, tuberculosis</td> </tr> <tr> <td>Key-features</td> <td>Radiologic finding detection, abnormality score, text interpretation</td> </tr> <tr> <td>Suggested use</td> <td>Before: adapting worklist order, flagging acute findings During: perception aid (prompting all abnormalities/results/heatmaps), report suggestion</td> </tr> </tbody> </table> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th colspan="2" style="background-color: #f2f2f2;">Data characteristics</th> </tr> </thead> <tbody> <tr> <td>Population</td> <td>Patients aged 14 years or older</td> </tr> <tr> <td>Input</td> <td>chest PA(posterior-anterior view), chest AP(anterior-posterior view)</td> </tr> <tr> <td>Input format</td> <td>DICOM</td> </tr> <tr> <td>Output</td> <td>Localization (color map, grayscale map, combined map, single colormap), abnormality score for each lesion/case, binary assessment of abnormality, worklist order, draft radiology report generation</td> </tr> <tr> <td>Output format</td> <td>DICOM Secondary Capture, DICOM GSPS (Grayscale Softcopy Presentation State), Prepopulated Report (API integration only), HL7</td> </tr> </tbody> </table>		General		Product name	Lunit INSIGHT CXR	Subspeciality	Chest	Modality	X-ray	Disease targeted	Atelectasis, calcification, cardiomegaly, consolidation, fibrosis, mediastinal widening, nodule, pleural effusion, pneumothorax, pneumoperitoneum, tuberculosis	Key-features	Radiologic finding detection, abnormality score, text interpretation	Suggested use	Before: adapting worklist order, flagging acute findings During: perception aid (prompting all abnormalities/results/heatmaps), report suggestion	Data characteristics		Population	Patients aged 14 years or older	Input	chest PA(posterior-anterior view), chest AP(anterior-posterior view)	Input format	DICOM	Output	Localization (color map, grayscale map, combined map, single colormap), abnormality score for each lesion/case, binary assessment of abnormality, worklist order, draft radiology report generation	Output format	DICOM Secondary Capture, DICOM GSPS (Grayscale Softcopy Presentation State), Prepopulated Report (API integration only), HL7
General																												
Product name	Lunit INSIGHT CXR																											
Subspeciality	Chest																											
Modality	X-ray																											
Disease targeted	Atelectasis, calcification, cardiomegaly, consolidation, fibrosis, mediastinal widening, nodule, pleural effusion, pneumothorax, pneumoperitoneum, tuberculosis																											
Key-features	Radiologic finding detection, abnormality score, text interpretation																											
Suggested use	Before: adapting worklist order, flagging acute findings During: perception aid (prompting all abnormalities/results/heatmaps), report suggestion																											
Data characteristics																												
Population	Patients aged 14 years or older																											
Input	chest PA(posterior-anterior view), chest AP(anterior-posterior view)																											
Input format	DICOM																											
Output	Localization (color map, grayscale map, combined map, single colormap), abnormality score for each lesion/case, binary assessment of abnormality, worklist order, draft radiology report generation																											
Output format	DICOM Secondary Capture, DICOM GSPS (Grayscale Softcopy Presentation State), Prepopulated Report (API integration only), HL7																											

	Technology
Integration	Integration in standard reading environment (PACS), Integration via AI marketplace or distribution platform, Stand-alone third party application, Stand-alone webbased
Deployment	Locally on dedicated hardware, Locally virtualized (virtual machine, docker), Cloud-based
Trigger for analysis	Automatically, right after the image acquisition
Processing time	3 - 10 seconds
	Certification
CE	Certified, Class IIa, MDR
FDA	510(k) cleared, Class II
	Market presence
On market since	11-2019
Distribution channels	GE Healthcare (Thoracic Care Suite), Philips Healthcare, INFINITT AI Platform, FujiFilm, Agfa, CARPL.ai, Osimis

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 ฮิสโตพลาสโมซิส ไครปโตคอคโคซิส 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

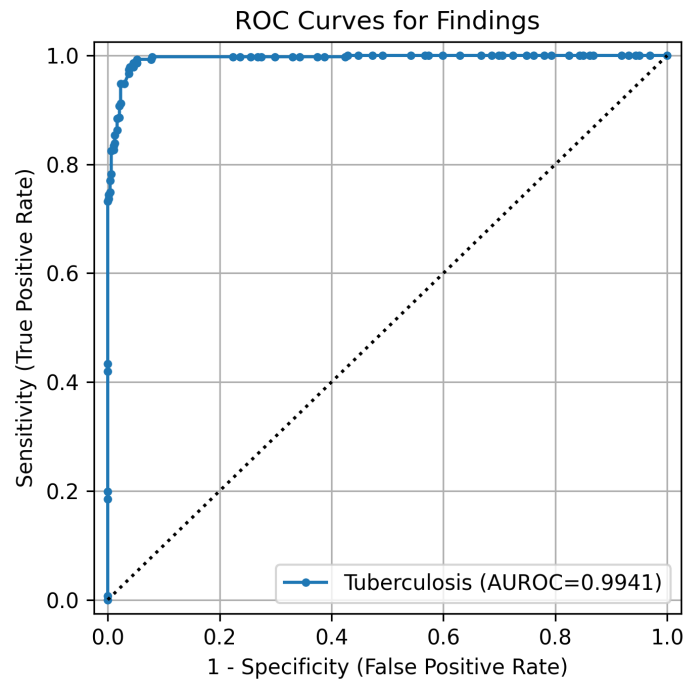
Finding	N	Threshold	Pairwise Agreement		Cohen's Kappa	
			"B" vs "B"	"B" vs AI	"B" vs "B"	"B" vs AI
Tuberculosis	422	0.15	0.9635	0.9689	0.9261	0.9377

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	0.15	0.9477	0.9929	0.9437	0.9934	0.9941

รูปภาพที่ 1 เส้นโค้ง ROC ของแต่ละรอยโรค



บทวิจารณ์

สำหรับการทดสอบ Lunit INSIGHT CXR เวอร์ชัน 3.1.5.0 ทางผู้ผลิตได้ส่งรายงานรอยโรคมา 1 รายการ คือ Tuberculosis

จากตารางที่ 6 เมื่อเทียบ Pairwise Agreement และ Cohen's Kappa ระหว่างรังสีแพทย์กับปัญญาประดิษฐ์ ("B" vs AI) และรังสีแพทย์ด้วยกัน ("B" vs "B") แล้ว Lunit INSIGHT CXR เวอร์ชัน 3.1.5.0 มีประสิทธิภาพใกล้เคียงกับรังสีแพทย์ (ความแตกต่างน้อยกว่า 5%) ในส่วนของการวินิจฉัยวัณโรค (Tuberculosis) ซึ่งปัญญาประดิษฐ์มีคะแนนมากกว่ารังสีแพทย์อยู่ 0.54% สำหรับ Pairwise Agreement และ คะแนนน้อยกว่ารังสีแพทย์อยู่ 1.16% 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	0.9689	+0.54%	0.9261	0.9377	+1.16%

ในส่วนของการคัดกรองวัณโรคปอด Lunit INSIGHT CXR เวอร์ชัน 3.1.5.0 เมื่อวิเคราะห์บนชุดข้อมูล 1A จำนวน 300 ภาพ พบว่าประสิทธิภาพในการวินิจฉัยวัณโรคใกล้เคียงกับรังสีแพทย์ และมีค่าพื้นที่ใต้กราฟตัวรับ (AUROC) อยู่ที่ 0.9941 ค่าความไว (Sensitivity) อยู่ที่ 0.9929 และค่าความจำเพาะ (Specificity) อยู่ที่ 0.9477 ที่ Threshold เป็น 0.15

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% Positive >99% Negative >60%	98%	Overall >95% Positive >99% Negative >68%	98%
Non-sputum based biomarker test	Overall >65% Positive >98%	98%	Positive >98% Negative >68%	98%
Triage test	90%	70%	95%	80%

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

ผลลัพธ์ของการคัดกรองวัณโรคปอด Lunit INSIGHT CXR เวอร์ชัน 3.1.5.0 ใน Threshold ต่าง ๆ เมื่อเทียบกับเกณฑ์ WHO TPP โดยกำหนดให้เป็นค่า Threshold ที่สูงที่สุดที่จะทำให้ค่า Specificity นั้นมีความใกล้เคียงกับ WHO TPP มากที่สุด ผลลัพธ์ที่ได้นั้นเป็นดังตารางที่ 8

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

Threshold	Sensitivity	Specificity
0.6073	0.8839	0.9833
0.1500	0.9299	0.9477
0.0923	0.9976	0.9205
0.0212	0.9976	0.7008

ทั้งนี้เมื่อนำผลลัพธ์ที่ได้เปรียบเทียบกับเกณฑ์ WHO TPP พบว่า Lunit CXR Analyser 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 Lunit INSIGHT CXR ตามเกณฑ์ WHO TPP

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

ตารางเสริม

ในรูปที่ 1 เส้นโค้ง ROC แสดงประสิทธิภาพของปัญญาประดิษฐ์ที่ระดับค่าเกณฑ์ (Threshold) ที่แตกต่างกัน โดยการปรับค่าเกณฑ์ เราสามารถเปลี่ยนแปลงความสัมพันธ์ระหว่างค่าความไว (Sensitivity) และค่าความเจาะจง (Specificity) ตาราง S1 แสดงค่าความไวและความเจาะจงที่ระดับค่าเกณฑ์ที่แตกต่างกันของ Finding ประเภท Tuberculosis

Table S1 ค่า Sensitivity และ Specificity ในระดับ Threshold ที่แตกต่างกันของ Finding ประเภท Tuberculosis (Threshold ที่ผู้ผลิตแนะนำคือ 0.15)

Threshold	Sensitivity	Specificity
0.95	0.42	1.00
0.80	0.75	0.99
0.77	0.77	0.99
0.75	0.78	0.99
0.72	0.83	0.99
0.67	0.84	0.99
0.66	0.85	0.98
0.64	0.86	0.98
0.60	0.90	0.98
0.48	0.95	0.98
0.45	0.95	0.97
0.42	0.97	0.96
0.31	0.98	0.96
0.24	0.99	0.95
0.20	0.99	0.95
0.10	0.99	0.92
0.05	0.88	0.77

โดยทั่วไปแล้ว ที่ Threshold ที่ต่ำ โมเดลจะมีค่า Sensitivity ที่สูง แต่ค่า Specificity จะต่ำ นั่นหมายความว่า โมเดลสามารถระบุคนที่ป่วยโรคได้ถูกต้องส่วนใหญ่ แต่ยังมีผลการผลบวกที่เท็จ เมื่อค่า Threshold อยู่ในระดับกลาง Sensitivity และ Specificity อยู่ในระดับที่สมดุล ซึ่งอาจเป็นทางเลือกที่ดีขึ้นในบางบริบท แต่เมื่อมี Threshold ที่สูง โมเดลจะมีค่า Specificity ที่สูงแต่ค่า Sensitivity จะต่ำ ซึ่งอาจจะเป็นทางเลือกที่เหมาะสมเมื่อต้องการความแน่นอนสูงเกี่ยวกับผลบวก

การเลือก Threshold เหมาะสมขึ้นอยู่กับความต้องการเฉพาะของบริบท ในบางครั้ง Sensitivity ที่สูงอาจจะสำคัญกว่า ในขณะที่บางครั้ง Specificity ที่สูงอาจจะสำคัญกว่า