

การเปรียบเทียบการแปลผลภาพถ่ายรังสีทรวงอก: กรณีการแปลผลโดยรังสีแพทย์และเทคโนโลยีปัญญาประดิษฐ์ โรงพยาบาลกาฬสินธุ์

รุ่งนภา อนุพงศ์พิพัฒน์

บทคัดย่อ ปัจจุบันปริมาณของภาพถ่ายรังสีทรวงอก (CXR) เพิ่มขึ้นอย่างต่อเนื่อง ส่งผลให้ภาระงานของรังสีแพทย์เพิ่มมากขึ้นจึงได้มีการนำเทคโนโลยีปัญญาประดิษฐ์ (Artificial intelligence: AI) มาช่วยในการแปลผล เพื่อลดขั้นตอนและเพิ่มประสิทธิภาพในการทำงานให้ดีขึ้น อย่างไรก็ตามเทคโนโลยี AI อยู่ในระหว่างการพัฒนาอย่างต่อเนื่อง เพื่อให้มีความถูกต้องแม่นยำที่สูงขึ้น การศึกษานี้มีวัตถุประสงค์เพื่อเปรียบเทียบการแปลผลภาพถ่ายรังสีทรวงอกโดยเทคโนโลยี AI และการแปลผลโดยรังสีแพทย์ ทำการศึกษาแบบย้อนหลังเชิงพรรณนา (retrospective descriptive study) กลุ่มตัวอย่างจาก CXR ในโรงพยาบาลกาฬสินธุ์ที่มี AI ในการแปลผลภาพและมีรายงานผลโดยรังสีแพทย์อย่างเป็นทางการ จำนวน 2,207 ภาพ วิเคราะห์เปรียบเทียบความสอดคล้องของการแปลผล การแปลผลภาพถ่ายรังสีทรวงอกโดย AI และโดยรังสีแพทย์ ส่วนใหญ่มีความสอดคล้องกันในระดับที่สูงถึงสูงมาก Kappa เท่ากับ 0.67 ถึง 1.0 และมีค่าความไวและความจำเพาะใกล้เคียงกัน แต่จุดขนาดเล็กในปอด (nodule) และจุดหินปูน (calcified nodule) มีค่าความไวและความสอดคล้องที่ลดลง ส่วนรอยโรคกระดูกหัก (fracture bone) ไม่พบการบ่งชี้จากการแปลผล AI ซึ่งแตกต่างจากการแปลผลโดยรังสีแพทย์ และพบมีความไม่สอดคล้องกันบ้างในบางรอยโรค เช่น รอยโรคในปอดที่ขาวขึ้น หรือจุดขนาดเล็กและบางตำแหน่ง เช่น ขั้วปอดทั้งสองข้าง ขั้วปอด และซ่อนทับเงาหัวใจ โดยสรุปการแปลผลภาพถ่ายรังสีทรวงอกโดยเทคโนโลยี AI และรังสีแพทย์ มีความสอดคล้องกันสูงถึงสูงมาก แม้ว่าส่วนของรอยโรคกระดูกหัก (fracture bone) ไม่พบการบ่งชี้จากการแปลผลโดย AI ซึ่งแตกต่างจากการแปลผลโดยรังสีแพทย์ การแปลผลโดยเทคโนโลยี AI จะมีความไวต่ำกว่าในจุดขนาดเล็กในปอดหรือจุดหินปูน และบางตำแหน่ง เช่น ขั้วปอด และ ซ่อนทับเงาหัวใจ จะต้องมีความระมัดระวังในการแปลผล อย่างไรก็ตามสามารถนำ AI ไปประยุกต์ใช้ในการแปลผลภาพถ่ายรังสีทรวงอก เพื่อลดขั้นตอนในการทำงานและลดระยะเวลาการรอคอย (วารสารโรคมะเร็ง 2567;44:173-183)

คำสำคัญ: การเปรียบเทียบ เทคโนโลยีปัญญาประดิษฐ์ ภาพถ่ายรังสีทรวงอก

แผนกรังสีวิทยา โรงพยาบาลกาฬสินธุ์

Comparison of Interpretations of Chest Radiographs: Case of Interpretation by Radiologist and Artificial Intelligence Technology, Kalasin Hospital

By **Roongnapa Anupongpipat**

Radiologist Department of Radiology, Kalasin Hospital

Abstract The amount of chest radiograph (CXR) is continuously increasing, resulting in increase the workload of radiologist. Artificial intelligence (AI) technology has been to assist in the interpretation of CXR for reduce steps and increase work efficiency. AI technology is continuously being develop to achieve higher accuracy. Objective: To study the comparison of CXR interpretation by AI technology and radiologist. Methods: Retrospective descriptive study, a samples of 2,207 CXR images in Kalasin hospital with interpreting by AI technology and by radiologist. Comparative analysis of the interpreting of CXR by AI technology and radiologists. Results: The interpreting by AI technology and radiologists is mostly highly consistent with a Kappa of 0.67 to 1.0, and the sensitivity and specificity are similar. Some pathologies such as nodules and calcified nodules have lower sensitivity and consistency. As for fracture bone, it is not indicated by AI technology, but indicate by radiologist. There were inconsistencies in some lesions such as infiltration, nodule and some locations such as both lower lungs, hilar and retrocardiac regions. Conclusions: The interpretation of CXR by AI technology and radiologists had a high to very high level of consistency. Although fracture bone was not indicated by AI interpretation which was different from the interpretation by radiologists. The interpretation by AI technology will have lower sensitivity such as nodule or calcified nodule and some locations such as hilar and retrocardiac regions that require caution in interpretation. However AI technology can be applied to the interpretation of CXR images to reduces work steps and waiting times (*Thai Cancer J* 2024;44:173-183)

Keyword: comparison, artificial intelligence technology, chest radiograph

บทนำ

ภาพถ่ายรังสีทรวงอก (CXR) เป็นเครื่องมือการตรวจเพื่อช่วยในการวินิจฉัยโรคทางทรวงอก อีกทั้งยังช่วยในการคัดกรองโรคทางทรวงอก เช่น วัณโรคปอด เป็นต้น ทำให้ปัจจุบันปริมาณของภาพถ่ายรังสีทรวงอกมากขึ้นอย่างต่อเนื่อง ส่งผลให้ภาระงานของรังสีแพทย์เพิ่มมากขึ้น ซึ่งในปัจจุบันได้มีการนำเทคโนโลยีปัญญาประดิษฐ์ (Artificial intelligence: AI) มาใช้ เพื่อช่วยในการแปลผลภาพถ่ายรังสีทรวงอก¹ อย่างไรก็ตามเทคโนโลยีปัญญาประดิษฐ์อยู่ในระหว่างการพัฒนาอย่างต่อเนื่อง เพื่อให้มีความไวและความจำเพาะในการแปลผลที่สูงขึ้น² จึงได้เกิดการศึกษาขึ้นเพื่อให้เกิดการพัฒนาเทคโนโลยีปัญญาประดิษฐ์ให้มีประสิทธิภาพมากยิ่งขึ้นในลำดับต่อไป เพื่อนำไปประยุกต์ใช้ โดยเฉพาะในโรงพยาบาลที่ไม่มีรังสีแพทย์ หรือมีรังสีแพทย์ไม่เพียงพอ

วัสดุและวิธีการ

เป็นการศึกษาแบบย้อนหลัง (retrospective descriptive study) กลุ่มตัวอย่าง คือ ภาพถ่ายรังสีทรวงอกของผู้มารับบริการโรงพยาบาลกาฬสินธุ์ที่มีปัญญาประดิษฐ์ในการแปลผลภาพ และมีรายงานผลโดยรังสีแพทย์อย่างเป็นทางการ ในช่วงระหว่างวันที่ 1 ม.ค. 2566 ถึง 30 มิ.ย. 2566 จำนวนทั้งสิ้น 2,207 ภาพ ศึกษาถึงพยาธิสภาพในทรวงอก และพยาธิสภาพของกระดูกทรวงอกเป็นหลัก โดยแบ่งพยาธิสภาพในทรวงอกเป็น 6 กลุ่มดังนี้ รอยกระดูกหัก (fracture bone) รอยโรคในปอดที่ขาวขึ้น (infiltration) น้ำในช่องเยื่อหุ้มปอด (pleural effusion) ก้อน (mass) จุดหินปูน (calcified nodule) จุดเล็กในปอด (nodule) ศึกษาความสอดคล้องการแปลผลภาพถ่ายรังสีทรวงอกโดยเทคโนโลยีปัญญาประดิษฐ์ และการแปลผลโดยรังสีแพทย์ที่ไม่เห็นภาพจากเทคโนโลยีปัญญาประดิษฐ์ ใช้สถิติเชิงพรรณนาในการอธิบายลักษณะประชากร ได้แก่ ตัวแปรเชิงคุณภาพ เช่น เพศ จะนำเสนอเป็นความถี่ และร้อยละ ตัวแปรเชิงปริมาณ เช่น อายุ จะนำเสนอด้วยค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน ข้อมูลปัจจัยที่มีลักษณะข้อมูลเชิงคุณภาพ เช่น เพศ จะใช้สถิติ Chi-square หรือ Fisher's exact test ใช้สถิติเชิงอนุมานในการศึกษาประเมินความตรงกันของการแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์และการแปลผลโดยรังสีแพทย์ที่ไม่เห็นภาพจากเทคโนโลยีปัญญาประดิษฐ์ นำเสนอด้วยค่าความไว (sensitivity) ความจำเพาะ (specificity) ค่าการทำนายโรคเมื่อผลการทดสอบเป็นบวก (Positive predictive value; PPV) ค่าการทำนายโรคเมื่อผลการทดสอบเป็นลบ (Negative predictive value; NPV) และประเมินความสอดคล้องของการแปลผลใช้สถิติ Kappa และ 95%CI ของ Kappa และร้อยละของความสอดคล้อง (Percent agreement)

ผลการศึกษา

จากการวิเคราะห์ลักษณะทางประชากรของการศึกษานี้ ส่วนใหญ่เป็นภาพถ่ายรังสีทรวงอกเพศชาย จำนวน 1,327 ราย คิดเป็นร้อยละ 60.13 และมีภาพถ่ายรังสีทรวงอกเพศหญิงจำนวน 880 ราย คิดเป็นร้อยละ 39.87 อายุเฉลี่ยโดยรวมเท่ากับ 49.60 ปี ถ้าเทียบอายุระหว่างเพศชายกับเพศหญิง อายุเฉลี่ยของเพศชายเท่ากับ 45.09 อายุเฉลี่ยของเพศหญิง เท่ากับ 56.40 เพศหญิงมีอายุเฉลี่ยที่มากกว่าอย่างมีนัยสำคัญ (ตารางที่ 1)

ตารางที่ 1 ลักษณะพื้นฐานทางประชากร

Characteristic	Total (n=2,207)	Male (n=1,327)	Female (n=880)	P
Age (years), mean±SD	49.60±16.77	45.09±16.25	56.40±15.18	<0.001
15–29 years	319 (14.5)	266 (20.0)	53 (6.0)	<0.001
30–44 years	607 (27.5)	459 (34.6)	148 (16.8)	
45–59 years	594 (26.9)	314 (23.7)	280 (31.8)	
≥60 years	687 (31.1)	288 (21.7)	399 (45.3)	

ตารางที่ 2 เปรียบเทียบการแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์กับรายงานผลโดยรังสีแพทย์ที่ไม่เห็นภาพจากเทคโนโลยีปัญญาประดิษฐ์

Finding	Positive n (%)	Negative n (%)	Sensitivity (95%CI)	Specificity (95%CI)	PPV (95%CI)	NPV (95%CI)	KAPPA (95%CI)	% Agreement	P
AI of Film (Standard)									
1. Fracture bone, n (%)									
P1 positive	0 (0.0)	9 (0.4)	-	-	-	-	0.0 (NA-1.0)	99.59	-
P1 negative	0 (0.0)	2198 (99.6)							
2. Infiltration, n (%)									
P1 positive	304 (66.2)	22 (1.3)	66.2 (61.7-70.5)	98.7 (98.1-99.2)	93.3 (90.0-95.7)	91.8 (90.4-93.0)	0.73 (0.69-0.77)	91.98	<0.001
P1 negative	155 (33.8)	1726 (98.7)							
AI of Film (Standard)									
3. Pleural effusion, n (%)									
P1 positive	4 (100)	0 (0.0)	100 (39.8-100)	100 (99.8-100)	100 (39.8-100)	100 (99.8-100)	1.0 (1.0-1.0)	100	<0.001
P1 negative	0 (0.0)	2203 (100)							
4. Mass, n (%)									
P1 positive	10 (100)	9 (0.4)	100 (69.2-100)	99.6 (99.2-99.8)	52.6 (28.6-75.6)	100 (99.8-100)	0.69 (0.49-0.88)	99.59	<0.001
P1 negative	0 (0.0)	2188 (99.6)							

Finding	Positive n (%)	Negative n (%)	Sensitivity (95%CI)	Specificity (95%CI)	PPV (95%CI)	NPV (95%CI)	KAPPA (95%CI)	% Agreement	P
AI of Film (Standard)									
5. Calcified, n (%)									
P1 positive	11 (100)	6 (0.3)	100 (71.5-100)	99.7 (99.4-99.9)	64.7 (38.3-85.8)	100 (99.8-100)	0.78 (0.62-0.95)	99.73	<0.001
P1 negative	0 (0.0)	2190 (99.7)							
6. Nodule, n (%)									
P1 positive	33 (94.3)	30 (1.4)	94.3 (80.8-99.3)	98.6 (98.0-99.1)	52.4 (39.4-65.1)	99.9 (99.7-100)	0.67 (0.56-0.78)	98.55	<0.001
P1 negative	2 (0.1)	2142 (98.6)							
Report of Film (Standard)									
1. Fracture bone, n (%)									
P1 positive	9 (100)	0 (0.0)	100 (66.4-100)	100 (99.8-100)	100 (66.4-100)	100 (99.8-100)	1.0 (1.0-1.0)	100	<0.001
P1 negative	0 (0.0)	2198 (100)							
AI positive	0 (0.0)	0 (0.0)	0.0 (0.0-33.6)	100 (99.8-100)	-	99.6 (99.2-99.8)	0.0 (NA-1.0)	99.59	-
AI negative	9 (100)	2198 (100)							
2. Infiltration, n (%)									
P1 positive	322 (99.4)	4 (0.2)	99.4 (97.8-99.9)	99.8 (99.5-99.9)	98.8 (96.9-99.7)	99.9 (99.6-100)	0.99 (0.98-1.0)	99.73	<0.001
P1 negative	2 (0.6)	1879 (99.8)							
AI positive	305 (66.4)	154 (8.2)	94.1 (91.0-96.4)	91.8 (90.5-93.0)	66.4 (61.9-70.8)	98.9 (98.3-99.3)	0.73 (0.70-0.77)	92.2	<0.001
AI negative	19 (1.1)	1729 (91.8)							
3. Pleural effusion, n (%)									
P1 positive	4 (100)	0 (0.0)	100 (39.8-100)	100 (99.8-100)	100 (39.8-100)	100 (99.8-100)	1.0 (1-1)	100	<0.001
P1 negative	0 (0.0)	2203 (100)							
AI positive	4 (100)	0 (0.0)	100 (39.8-100)	100 (99.8-100)	100 (39.8-100)	100 (99.8-100)	1.0 (1.0-1.0)	100	<0.001
AI negative	0 (0.0)	2203 (100)							

Finding	Positive n (%)	Negative n (%)	Sensitivity (95%CI)	Specificity (95%CI)	PPV (95%CI)	NPV (95%CI)	KAPPA (95%CI)	% Agreement	P
Report of Film (Standard)									
4. Mass, n (%)									
P1 positive	17 (100)	2 (0.1)	100 (80.5-100)	99.9 (99.7-100)	89.5 (66.9-98.7)	100 (99.8-100)	0.94 (0.87-1.0)	99.91	<0.001
P1 negative	0 (0.0)	2188 (99.9)							
AI positive	10 (58.8)	0 (0.0)	58.8 (32.9-81.6)	100 (99.8-100)	100 (69.2-100)	99.7 (99.3-99.9)	0.74 (0.55-0.93)	99.68	<0.001
AI negative	7 (41.2)	2190 (99.7)							
5. Calcified, n (%)									
P1 positive	16 (100)	1 (0.0)	100 (79.4-100)	100 (99.7-100)	94.1 (71.3-99.9)	100 (99.8-100)	0.97 (0.91-1.00)	99.95	<0.001
P1 negative	0 (0.0)	2190 (100)							
AI positive	11 (68.8)	0 (0.0)	68.8 (41.3-89.0)	100 (99.8-100)	100 (71.5-100)	99.8 (99.5-99.9)	0.81 (0.65-0.97)	99.77	<0.001
AI negative	5 (31.2)	2191 (100)							
6. Nodule, n (%)									
P1 positive	59 (100)	4 (0.2)	100 (93.9-100)	99.8 (99.5-99.9)	93.7 (84.5-98.2)	100 (99.8-100)	0.97 (0.93-1.0)	99.82	<0.001
P1 negative	0 (0.0)	2144 (99.8)							
AI positive	33 (55.9)	2 (0.1)	55.9 (42.4-68.8)	99.9 (99.7-100)	94.3 (80.8-99.3)	98.8 (98.3-99.2)	0.70 (0.59-0.80)	98.73	<0.001
AI negative	26 (44.1)	2146 (99.9)							

หมายเหตุ: AI คือ การแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์ P1 คือการแปลผลจากรังสีแพทย์โดยไม่เห็นภาพจากเทคโนโลยีปัญญาประดิษฐ์ Report คือ รายงานผลอย่างเป็นทางการของรังสีแพทย์ในระบบ PACs

จากตารางที่ 2 พบว่าพยาธิสภาพที่มีน้ำในช่องเยื่อหุ้มปอด (pleural effusion) การแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์เทียบกับการแปลผลโดยรังสีแพทย์ มีความสอดคล้องกันในระดับที่สูงมาก Kappa เท่ากับ 1.0 (95%CI: 1.0-1.0) และมีความสอดคล้องกันอย่างมีนัยสำคัญทางสถิติ ($P<0.001$) มีความไวและจำเพาะสูง โดยมีค่าความไวและความจำเพาะเท่ากับร้อยละ 100

กลุ่มรอยโรคในปอดที่ขาวขึ้น (infiltration) พบว่าการแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์เทียบกับการแปลผลโดยรังสีแพทย์ที่ไม่เห็นภาพเทคโนโลยีปัญญาประดิษฐ์ มีความสอดคล้องกันในระดับสูง Kappa

เท่ากับ 0.73 (95%CI: 0.69-0.77) และมีความสอดคล้องกันอย่างมีนัยสำคัญทางสถิติ ($P<0.001$) แต่การแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์มีค่าความไวสูงกว่า โดยมีค่าความไวเท่ากับร้อยละ 94.1 ในขณะที่การแปลผลโดยรังสีแพทย์ที่ไม่เห็นภาพจากเทคโนโลยีปัญญาประดิษฐ์มีค่าความไวเท่ากับร้อยละ 66.2

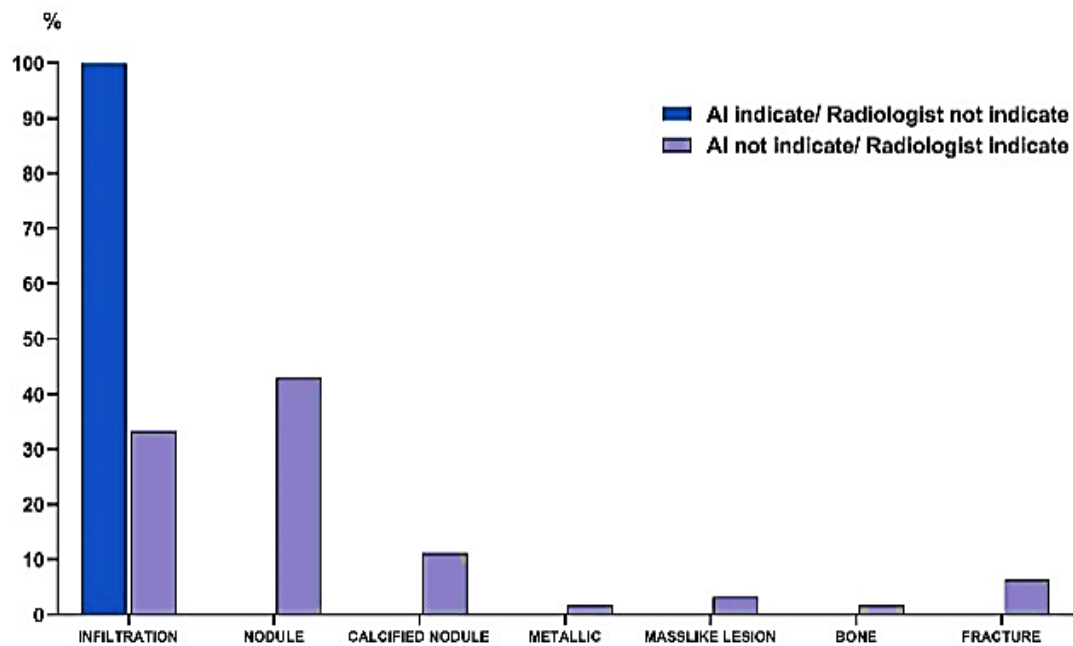
ลักษณะจุดหินปูน (calcified nodule) พบว่าการแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์เทียบกับการแปลผลโดยรังสีแพทย์ที่ไม่เห็นภาพเทคโนโลยีปัญญาประดิษฐ์ มีความสอดคล้องกันในระดับสูง Kappa เท่ากับ 0.78 (95%CI: 0.62-0.95) และมีความสอดคล้องกันอย่างมีนัยสำคัญทางสถิติ ($P<0.001$) แต่การแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์มีความไวต่ำกว่า โดยมีค่าความไวเท่ากับร้อยละ 68.8 แต่มีค่าความจำเพาะ ความน่าจะเป็นที่เป็นโรค ความน่าจะเป็นที่จะไม่เป็นโรคที่สูง คือ มีค่าเท่ากับร้อยละ 100 ร้อยละ 100 และ ร้อยละ 99.8 ตามลำดับ

จุดขนาดเล็กในปอด (nodule) พบว่ามีความสอดคล้องกันในระดับสูง ในการแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์เทียบกับการแปลผลโดยรังสีแพทย์ที่ไม่เห็นภาพเทคโนโลยีปัญญาประดิษฐ์ Kappa เท่ากับ 0.67 (95%CI: 0.56-0.78) และมีความสอดคล้องกันอย่างมีนัยสำคัญทางสถิติ ($P<0.001$) แต่การแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์ มีความไวต่ำกว่าการแปลผลโดยรังสีแพทย์ โดยมีค่าความไวเท่ากับร้อยละ 55.9 มีค่าความจำเพาะ ความน่าจะเป็นที่เป็นโรค ความน่าจะเป็นที่จะไม่เป็นโรคที่สูง คือ มีค่าเท่ากับร้อยละ 99.9 ร้อยละ 94.3 และร้อยละ 98.8 ตามลำดับ

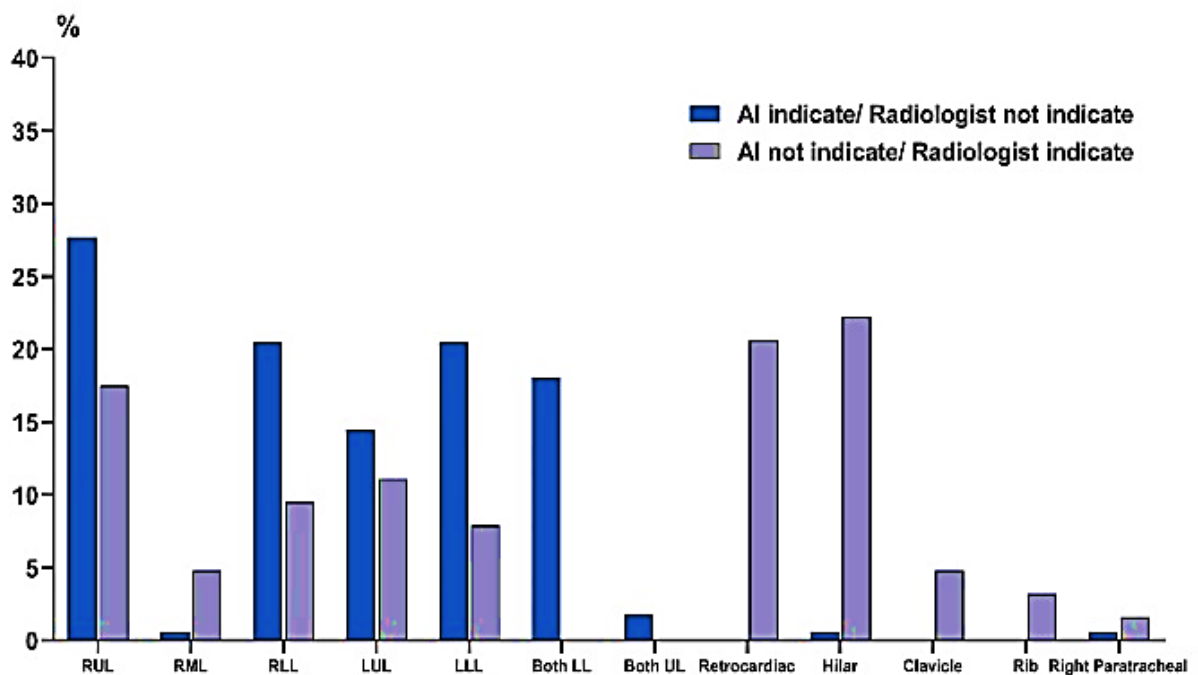
พยาธิสภาพที่เป็นก้อน (mass) มีความสอดคล้องกันในระดับสูงในการแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์เทียบกับการแปลผลโดยรังสีแพทย์โดยไม่เห็นภาพเทคโนโลยีปัญญาประดิษฐ์ Kappa เท่ากับ 0.69 (95%CI: 0.49-0.88) และมีความสอดคล้องกันอย่างมีนัยสำคัญทางสถิติ ($P<0.001$) แต่การแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์ มีความไวต่ำกว่าการแปลผลโดยรังสีแพทย์ โดยมีค่าความไวเท่ากับร้อยละ 58.8 มีค่าความจำเพาะ ความน่าจะเป็นที่เป็นโรค ความน่าจะเป็นที่จะไม่เป็นโรคที่สูง คือ มีค่าเท่ากับร้อยละ 100 ร้อยละ 100 และร้อยละ 99.7 ตามลำดับ

ส่วนพยาธิสภาพของรอยกระดูกหัก (fracture bone) จำนวน 9 ราย การแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์ไม่ได้บ่งชี้ว่ามีพยาธิสภาพทั้ง 9 ราย ในขณะที่การแปลผลภาพโดยรังสีแพทย์ที่ไม่เห็นภาพจากเทคโนโลยีปัญญาประดิษฐ์ และรายงานผลอย่างเป็นทางการของรังสีแพทย์ในระบบ PACs บ่งชี้ว่ามีพยาธิสภาพของรอยกระดูกหักทั้ง 9 ราย อย่างไรก็ตามพบมีความไม่สอดคล้องกันอยู่บ้างลักษณะที่เทคโนโลยีปัญญาประดิษฐ์แปลผลบ่งชี้ว่ามีพยาธิสภาพ แต่รังสีแพทย์ไม่ได้บ่งชี้ว่ามีพยาธิสภาพ ส่วนใหญ่คือรอยโรคในปอดที่ขาขึ้น (infiltration) (แผนภูมิที่ 1) ในขณะที่การแปลผลโดยรังสีแพทย์บ่งชี้ว่ามีพยาธิสภาพ แต่การแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์ไม่ได้บ่งชี้ว่ามีพยาธิสภาพ ส่วนใหญ่คือจุดขนาดเล็กในปอด (nodule) (แผนภูมิที่ 1) หากพิจารณาโดยตำแหน่ง พบว่าตำแหน่งที่การแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์บ่งชี้ว่ามีพยาธิสภาพ แต่การแปลผลโดยรังสีแพทย์ไม่ได้บ่งชี้ว่ามีพยาธิสภาพ ได้แก่ ชายปอดทั้งสองข้าง (both LL) มีร้อยละ 18.1 ปอดกลีบขวาบน (RUL) มีร้อยละ 27.7 และปอดกลีบขวาล่าง (RLL) มีร้อยละ 20.5 (แผนภูมิที่ 2) ตำแหน่งที่การแปลผลโดยรังสีแพทย์บ่งชี้ว่ามีพยาธิสภาพแต่การแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์ไม่บ่งชี้

ว่ามีพยาธิสภาพ มักเป็นตำแหน่งซั้วปอด (hilar) และตำแหน่งซ้อนทับเงาหัวใจ (retrocardiac) (แผนภูมิที่ 2)



แผนภูมิที่ 1 แสดงจำนวนร้อยละของการแปลผลโดยรังสีแพทย์และการแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์
เห็นไม่สอดคล้องกันตามลักษณะตามลักษณะพยาธิสภาพ



แผนภูมิที่ 2 แสดงจำนวนร้อยละของการแปลผลโดยรังสีแพทย์และการแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์
เห็นไม่สอดคล้องกันตามตำแหน่งปอด

วิจารณ์

การแปลผลภาพถ่ายรังสีทรวงอกของพยาธิสภาพในปอด โดยใช้เทคโนโลยีปัญญาประดิษฐ์ มีความสอดคล้องกันกับการแปลผลโดยรังสีแพทย์ อยู่ในระดับที่สูงถึงสูงมากโดยมีค่า Kappa อยู่ในช่วง 0.67-1.00 ร้อยละของความสอดคล้องอยู่ในช่วงร้อยละ 89.35-100 ซึ่งสอดคล้องกับการศึกษาก่อนหน้านี้ของ Joy T. Wu และคณะ³ แต่มีบางพยาธิสภาพ เช่น ลักษณะเป็นก้อน (mass) จุดขนาดเล็กในปอด (nodule) หรือจุดหินปูน (calcified nodule) การแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์จะมีความไวลดลง แต่ยังมีค่าความจำเพาะที่สูง ซึ่งเป็นไปในทิศทางเดียวกันกับการศึกษาก่อนหน้านี้ของ วีรยา น้อยศิริ และคณะ⁴ และการศึกษาของ Pattaranan และคณะ⁵ ที่พบว่าการแปลผลภาพถ่ายรังสีทรวงอกด้วยเทคโนโลยีปัญญาประดิษฐ์ ในมะเร็งปอด มีความไวที่ต่ำกว่ารังสีแพทย์ แต่มีความจำเพาะสูงใกล้เคียงกัน การศึกษาของ Khemipa Sanklaa⁶ ที่พบว่าการแปลผลภาพถ่ายรังสีทรวงอกด้วยเทคโนโลยีปัญญาประดิษฐ์ในการคัดกรองวัณโรคปอด มีความจำเพาะสูงแต่มีความไวต่ำ และการศึกษาของ Parisa Kaviani และคณะ⁷ ที่พบว่าการแปลผลภาพถ่ายรังสีทรวงอกด้วยเทคโนโลยีปัญญาประดิษฐ์มักจะไม่พบพยาธิสภาพที่เป็นจุดขนาดเล็กในปอด (nodule)

พยาธิสภาพที่มีลักษณะกลุ่มรอยโรคในปอดที่ขาขึ้น (infiltration) การแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์เทียบกับรายงานผลอย่างเป็นทางการของรังสีแพทย์ในระบบ PACs พบว่ามีความไวสูงใกล้เคียงกัน แต่หากเปรียบเทียบการแปลผลโดยรังสีแพทย์ที่ไม่เห็นภาพเทคโนโลยีปัญญาประดิษฐ์ พบว่ามีความไวลดลง เนื่องจากรายงานผลอย่างเป็นทางการของรังสีแพทย์ในระบบ PACs นั้นเห็นภาพเทคโนโลยีปัญญาประดิษฐ์ขณะแปลผล ทำให้มีการแปลผลคล้อยตามกับเทคโนโลยีปัญญาประดิษฐ์ได้สอดคล้องกับการศึกษาของ Michael H. Bernstein และคณะ⁸ หากพิจารณาในส่วนที่ไม่สอดคล้องกันอยู่บ้าง พบว่าการแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์บ่งชี้ว่ามีพยาธิสภาพ แต่การแปลผลโดยรังสีแพทย์ผู้เชี่ยวชาญไม่ได้บ่งชี้ว่ามีพยาธิสภาพนั้น ส่วนใหญ่เป็นรอยโรคในปอดที่ขาขึ้น (infiltration) และหากพิจารณาตำแหน่งการแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์บ่งชี้ว่ามีพยาธิสภาพ แต่การแปลผลโดยรังสีแพทย์ไม่ได้บ่งชี้ว่ามีพยาธิสภาพ มักจะเป็นบริเวณยอดปอดและชายปอดทั้งสองข้าง น่าจะมีผลจากปัจจัยของการซ้อนทับของเงา เช่น การซ้อนทับของเงาเนื้อเยื่อ เต้านม การซ้อนทับของเงากระดูก เป็นต้น ซึ่งเทคโนโลยีปัญญาประดิษฐ์อาจจะยังไม่สามารถแยกแยะได้อย่างชัดเจน ตำแหน่งที่การแปลผลของรังสีแพทย์บ่งชี้ว่ามีพยาธิสภาพ แต่การแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์ไม่บ่งชี้ว่ามีพยาธิสภาพ มักพบในตำแหน่งซั้วปอด (hilum) ร้อยละ 22.2 และตำแหน่งซ้อนทับเงาหัวใจ (retrocardiac) ร้อยละ 20.6 ซึ่งสอดคล้องกับการศึกษาก่อนหน้านี้ของ Sicular และคณะ⁹ ที่พบว่าการแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์พลาดการพบรอยโรคในตำแหน่งซั้วปอดด้านซ้าย และการศึกษาของ วีรยา น้อยศิริ และคณะ² ที่พบว่าตำแหน่งที่การแปลผลภาพถ่ายรังสีทรวงอกด้วยเทคโนโลยีปัญญาประดิษฐ์ให้ผลลบลง ได้แก่ ตำแหน่งซั้วปอดทั้งสองข้าง ด้านหลังเงาหัวใจ ยอดปอดด้านบนและฐานปอดด้านล่าง ข้อจำกัดของการศึกษานี้คือการแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์ในโรงพยาบาลกาฬสินธุ์ มีการนำมาใช้เฉพาะภาพถ่ายรังสีทรวงอกในรถเอกซเรย์เคลื่อนที่เท่านั้น ยังมิได้มีการนำมาใช้กับผู้ป่วยที่มารับบริการในโรงพยาบาล และเป็นการศึกษาแบบย้อนหลัง อาจมีการขาดความครอบคลุมครบถ้วนในบางข้อมูล

สรุป

จากผลการศึกษาวิเคราะห์พบว่า การแปลผลภาพถ่ายรังสีทรวงอกเทคโนโลยีปัญญาประดิษฐ์ และการแปลผลโดยรังสีแพทย์มีความสอดคล้องกันในระดับที่สูงถึงสูงมาก โดยมีค่าความสอดคล้อง Kappa อยู่ในช่วง 0.67-1.00 แม้ว่าจะมีบางพยาธิสภาพ เช่น จุดขนาดเล็กในปอด (nodule) หรือจุดหินปูน (calcified nodule) จะมีความไวลดลงบ้าง แต่ยังมีค่าความจำเพาะที่สูงใกล้เคียงกับการแปลผลโดยรังสีแพทย์ ส่วนพยาธิสภาพที่เป็นรอยกระดูกหัก (fracture bone) จำนวน 9 ราย การแปลผลโดยเทคโนโลยีปัญญาประดิษฐ์ไม่ได้บ่งชี้ว่ามีพยาธิสภาพ¹⁰ ทั้ง 9 ราย ในขณะที่การแปลผลภาพโดยรังสีแพทย์ที่ไม่เห็นภาพ จากเทคโนโลยีปัญญาประดิษฐ์ และรายงานผลอย่างเป็นทางการของรังสีแพทย์ในระบบ PACs บ่งชี้ว่ามีพยาธิสภาพของรอยกระดูกหักทั้ง 9 ราย และมีบางตำแหน่งเช่นตำแหน่งซั้วปอด (hilum) และตำแหน่งซ่อนทับเงาหัวใจ (retrocardiac) ที่ควรจะต้องมีระมัดระวังในการแปลผลมากขึ้น อย่างไรก็ตามเทคโนโลยีปัญญาประดิษฐ์สามารถนำไปประยุกต์ใช้เป็นเครื่องมือทางเลือกสำหรับการแปลผลภาพถ่ายรังสีทรวงอก เพื่อช่วยลดขั้นตอนในการทำงานของรังสีแพทย์เพิ่มประสิทธิภาพในการแปลผล และลดระยะเวลาในการรอคอย^{11,12} รวมถึงสามารถนำไปประยุกต์ใช้ในโรงพยาบาลที่ไม่มีรังสีแพทย์หรือโรงพยาบาลชุมชนเพื่อช่วยในการวินิจฉัยที่ถูกต้องแม่นยำ

ข้อเสนอแนะ

เพื่อเพิ่มประสิทธิภาพความไว และความจำเพาะในการใช้เทคโนโลยีปัญญาประดิษฐ์ช่วยในการแปลผลภาพถ่ายรังสีทรวงอกให้สูงขึ้นนั้น อาจต้องพิจารณาเพิ่มข้อมูลภาพถ่ายทางรังสีทรวงอกให้เทคโนโลยีปัญญาประดิษฐ์ได้เรียนรู้เพิ่มเติมมากขึ้นในบางพยาธิสภาพ เช่น รอยกระดูกหัก (fracture bone) จุดขนาดเล็กในปอด (nodule) จุดหินปูน (calcified nodule) และบางตำแหน่ง เช่น ตำแหน่ง ซั้วปอด (hilum) หรือตำแหน่งซ่อนทับเงาหัวใจ (retrocardiac) เป็นต้น เพื่อให้มีการพัฒนาอย่างต่อเนื่องและมีประสิทธิภาพมากขึ้น

กิตติกรรมประกาศ

ขอขอบพระคุณโรงพยาบาลกาฬสินธุ์ และเจ้าหน้าที่ผู้ดูแลระบบการจัดเก็บภาพและการแปลผล (PACs) ที่อนุเคราะห์ข้อมูลสำหรับการศึกษานี้

References

1. Calandriello L, Walsh SLF. Artificial intelligence for thoracic radiology: from research tool to clinical practice. Eur Respir J 2021;57:2100625.
2. Ridder K, Preuhs A, Mertins A, Joerger C. Routine usage of AI-based chest X-ray reading support in a multi-site medical Supply Center. arXiv 2022.
3. Wu JT, Wong KCL, Gur Y, Ansari N, Karagyris A, Sharma A, et al. Comparison of chest radiograph interpretations by Artificial Intelligence Algorithm vs Radiology Residents. JAMA Netw Open 2020; 3:e2022779.

4. Noisiri W, Vijitrsaguan C, Lertrojpanya S, Jiamjit K, Chayjaroon J, Tantibundhit C. Sensitivity and specificity of artificial intelligence for chest diagnostic radiology in lung cancer. *J DMS* 2021;45:55-61.
5. Munpolsri P, Sarakarn P, Munpolsri N. Screening of lung cancer using chest radiographs with application AI chest for all (DMS TU) in the context of a Regional Cancer Hospital. *J DMS [Internet]*. 2021;46:138-44.
6. Sanklaa K. Evaluation of efficiency of artificial intelligence for chest radiograph interpretation for pulmonary tuberculosis screening in mobile x-ray vehicle. *J Assoc Med Sci* 2021;54:43-47
7. Kaviani P, Kalra MK, Digumarthy SR, Gupta RV, Dasegowda G, Jagirdar A, et al. Frequency of missed findings on Chest Radiographs (CXRs) in an international, multicenter study: application of AI to reduce missed findings. *Diagnostics (Basel)* 2022;12:2382.
8. Bernstein MH, Atalay MK, Dibble EH, Maxwell AWP, Karam AR, Agarwal S, et al. Can incorrect artificial intelligence (AI) results impact radiologists, and if so, what can we do about it? A multi-reader pilot study of lung cancer detection with chest radiography. *Eur Radiol* 2023;33:8263-9.
9. Sicular S, Alpaslan M, Ortega FA, Keathley N, Venkatesh S, Jones RM, et al. Reevaluation of missed lung cancer with artificial intelligence. *Respir Med Case Rep* 2022;39:101733.
10. Osatavanichvong K, Nakano E de G, Dessí G, Saksirinukul T. Evaluation Artificial Intelligent (AI) assists radiologist in radiographic chest interpretation. *BKK Med J* 2019;14:59-65.
11. Harris M, Qi A, Jeagal L, Torabi N, Menzies D, Korobitsyn A, et al. A systematic review of the diagnostic accuracy of artificial intelligence-based computer programs to analyze chest x-rays for pulmonary tuberculosis. *PLoS One* 2019;14:e0221339.
12. Shin HJ, Han K, Ryu L, Kim EK. The impact of artificial intelligence on the reading times of radiologists for chest radiographs. *NPJ Digit Med* 2023;6:82