

การศึกษาเปรียบเทียบประสิทธิภาพระหว่างระบบคอมพิวเตอร์ปัญญาประดิษฐ์ 2 ระบบในการการ
ช่วยตรวจจับตึงเนื้อลำไส้ใหญ่โดยใช้ค่าเกณฑ์ชี้วัดที่เหมาะสมเพื่อลดอัตราการเกิดผลบวกложงในผู้ที่มา
ส่องกล้องคัดกรองมะเร็งลำไส้ใหญ่



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาอายุรศาสตร์ ภาควิชาอายุรศาสตร์
คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2564
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Comparative Analysis of Diagnostic Performance Between Two Artificial Intelligence-
Assisted Polyp Detection Systems Using a Selected Optimal False Positive
Thresholds.



A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Medicine
Department of Medicine
FACULTY OF MEDICINE
Chulalongkorn University
Academic Year 2021
Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การศึกษาเปรียบเทียบประสิทธิภาพระหว่างระบบคอมพิวเตอร์ปัญญาประดิษฐ์ 2 ระบบในการการช่วยตรวจจับตังเนื้อลำไส้ใหญ่โดยใช้ค่าเกณฑ์ชีวิตที่เหมาะสมเพื่อลดอัตราการเกิดผลบวกลวงในผู้ที่มีมาส่องกล้องคัดกรองมะเร็งลำไส้ใหญ่
โดย	น.ส.จุฬาลักษณ์ คฤหาสน์สุวรรณ
สาขาวิชา	อายุรศาสตร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ นายแพทย์ภาฤทธิ์ เมฆอรุณกมล
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	ศาสตราจารย์ นายแพทย์รังสรรค์ ฤกษ์นิมิตร

คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้รับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

..... คณบดีคณะแพทยศาสตร์
(รองศาสตราจารย์ นายแพทย์ฉันทชาย สิทธิพันธุ์)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ศาสตราจารย์ นายแพทย์วีรพันธุ์ โขวิฑูรกิจ)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผู้ช่วยศาสตราจารย์ นายแพทย์ภาฤทธิ์ เมฆอรุณกมล)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(ศาสตราจารย์ นายแพทย์รังสรรค์ ฤกษ์นิมิตร)

..... กรรมการ
(รองศาสตราจารย์ แพทย์หญิงปิยวรรณ กิตติสกุลนาม)

..... กรรมการภายนอกมหาวิทยาลัย
(แพทย์หญิงฉันทพร ฉันทโรจน์ศิริ)

จุฬาลักษณ์ คุฤหาสน์สุวรรณ : การศึกษาเปรียบเทียบประสิทธิภาพระหว่างระบบคอมพิวเตอร์ปัญญาประดิษฐ์ 2 ระบบในการการช่วยตรวจจับตึงเนื้อลำไส้ใหญ่โดยใช้ค่าเกณฑ์ชี้วัดที่เหมาะสมเพื่อลดอัตราการเกิดผลบวกลวงในผู้ที่มีมาส่งกล้องคัดกรองมะเร็งลำไส้ใหญ่. (Comparative Analysis of Diagnostic Performance Between Two Artificial Intelligence-Assisted Polyp Detection Systems Using a Selected Optimal False Positive Thresholds.) อ.ที่ปรึกษาหลัก : ผศ. นพ.ภาฤทธิ์ เมฆอรุณกมล, อ.ที่ปรึกษาร่วม : ศ. นพ.รังสรรค์ ฤกษ์นิมิตร

ความสำคัญและที่มา: จำนวนของผลบวกลวง (false positive) ปริมาณมากที่เกิดขึ้นในการส่องลำไส้ใหญ่โดยใช้ระบบคอมพิวเตอร์ปัญญาประดิษฐ์ช่วยในการหาตึงเนื้อลำไส้ อาจทำให้แพทย์ผู้ทำการส่องกล้องเกิดความรำคาญ เสียสมาธิในการส่องกล้อง เสียเวลามากขึ้นในการเข้าไปตรวจเช็คซ้ำบริเวณที่มีกรอบขึ้น ทำให้ใช้เวลาในการส่องกล้องลำไส้ใหญ่นานขึ้น อย่างไรก็ตาม การปรับค่าเกณฑ์ชี้วัดที่ช่วยลดผลบวกลวงทำให้ความไวและความถูกต้องของประสิทธิภาพของระบบคอมพิวเตอร์ปัญญาประดิษฐ์ในการตรวจพบตึงเนื้อลำไส้ใหญ่ลดลง

วัตถุประสงค์: เพื่อเปรียบเทียบประสิทธิภาพระหว่างระบบคอมพิวเตอร์ปัญญาประดิษฐ์ 2 ระบบคือ CAD EYE และ Deep-GI ในการการช่วยตรวจจับตึงเนื้อลำไส้ใหญ่ในผู้ที่มีมาส่งกล้องคัดกรองมะเร็งลำไส้ใหญ่ และหาค่าเกณฑ์ชี้วัดที่เหมาะสมเพื่อลดอัตราการเกิดผลบวกลวง

ระเบียบวิธีการวิจัย: ระบบ Deep-GI ได้ถูกพัฒนาขึ้น โดยใช้กระบวนการเรียนรู้เชิงลึกของคอมพิวเตอร์บน Model ทางวิศวกรรมศาสตร์ ชื่อ YOLOv5 deep learning framework ซึ่งมีการใช้ข้อมูลตัวอย่างในการพัฒนา Deep-GI ถึง 15,188 ภาพ และใช้ในการทดสอบอีก 1,520 ภาพ เมื่อเริ่มทำการศึกษาก็จะมีการบันทึกวิดีโอขณะกำลังถอยกล้องตรวจดูลำไส้ใหญ่ ทั้งวิดีโอที่มีระบบ CAD EYE (CAD EYE™; Fujifilm, Tokyo, Japan) และ วิดีโอที่นำระบบ Deep-GI มาใส่เข้าไปภายหลัง หลังจากนั้น วิดีโอที่มีระบบคอมพิวเตอร์ปัญญาประดิษฐ์ ทั้ง 2 วิดีโอ จะถูกตรวจสอบและบันทึกข้อมูลต่างๆ โดยอายุรแพทย์ทางเดินอาหารที่มีความเชี่ยวชาญด้านการส่องกล้องลำไส้ใหญ่ 2 คน ข้อมูลใดที่มีความเห็นไม่ตรงกันจะถูกตรวจสอบซ้ำโดยอายุรแพทย์ทางเดินอาหารที่มีความเชี่ยวชาญพิเศษ ผลบวกจริง (true positive) คือ การที่ปัญญาประดิษฐ์ขึ้นกรอบเตือนว่ามีตึงเนื้อในบริเวณที่มีตึงเนื้อจริง หลังจากแพทย์ตรวจสอบวิดีโอและบันทึกข้อมูลแล้ว จะนำข้อมูลต่างๆ เช่น อัตราการพบตึงเนื้อลำไส้ใหญ่ (polyp detection rate), อัตราการไม่พบตึงเนื้อลำไส้ใหญ่ (polyp miss rate) และอัตราการพบผลบวกลวง (false positive rate) มาคำนวณเปรียบเทียบประสิทธิภาพของระบบคอมพิวเตอร์ปัญญาประดิษฐ์ทั้ง 2 ระบบโดยเปรียบเทียบในหลายๆเกณฑ์ชี้วัด ทั้งเกณฑ์ชี้วัด ≥ 0.5 วินาที, ≥ 1 วินาที, ≥ 1.5 วินาที และ ≥ 2 วินาที

ผลการวิจัย: จากผู้ป่วยที่มีมาส่งกล้องเพื่อคัดกรองมะเร็งลำไส้ใหญ่ทั้งหมด 170 ราย พบว่า มีผู้ป่วย 137 รายที่ส่องกล้องพบตึงเนื้อลำไส้ใหญ่อย่างน้อย 1 ตึงเนื้อ โดยมีค่าเฉลี่ยระยะเวลาในการถอยกล้องเท่ากับ 7.8 ± 2.7 นาที จากผู้ป่วย 137 ราย พบตึงเนื้อลำไส้ใหญ่ทั้งหมด 501 ตึงเนื้อ แบ่งเป็นตึงเนื้อชนิด adenoma 262 ตึงเนื้อ คิดเป็นร้อยละ 52.3 และตึงเนื้อชนิด hyperplastic 239 ตึงเนื้อ คิดเป็นร้อยละ 47.7 มีตึงเนื้อลำไส้ใหญ่ 3 ตึงเนื้อที่ไม่ถูกตรวจพบโดยระบบ Deep-GI คิดเป็นร้อยละ 0.6 ในขณะที่มีตึงเนื้อลำไส้ใหญ่ 73 ตึงเนื้อที่ไม่ถูกตรวจพบโดยระบบ CAD EYE คิดเป็นร้อยละ 14.6 ซึ่งสูงกว่าอย่างมีนัยยะสำคัญทางสถิติ ($p < 0.01$) โดยตึงเนื้อที่ไม่ถูกพบโดยระบบปัญญาประดิษฐ์ พบเป็นตึงเนื้อชนิด hyperplastic 64 ตึงเนื้อ คิดเป็นร้อยละ 85 เมื่อเปรียบเทียบประสิทธิภาพระหว่าง Deep-GI และ CAD EYE พบว่า ระบบ Deep-GI มีอัตราการตรวจพบตึงเนื้อลำไส้ใหญ่ (polyp detection rate) ที่สูงกว่า (99.4% vs. 85.4%; $p < 0.01$), มีอัตราการไม่พบตึงเนื้อลำไส้ใหญ่ (polyp miss rate) ที่ต่ำกว่า (0.6% vs. 14.6%; $p < 0.01$) และมีจำนวนการเกิดผลบวกลวงต่อการระบบส่องกล้องลำไส้ใหญ่น้อยกว่าระบบ CAD EYE (349±169 vs. 624±468; $p < 0.01$) เมื่อดูที่อัตราการตรวจพบตึงเนื้อชนิด adenoma (adenoma detection rate) พบว่า ระบบ Deep-GI มีอัตราการตรวจพบตึงเนื้อชนิด adenoma ที่สูงกว่าระบบ CAD EYE อย่างมีนัยยะสำคัญทางสถิติ นอกจากนี้ Deep-GI ยังมีอัตราการตรวจพบตึงเนื้อลำไส้ใหญ่ ที่สูงกว่าทั้งในเกณฑ์ชี้วัดสำหรับผลบวกลวงที่ ≥ 0.5 วินาที และ ≥ 1 วินาที (99.4% vs. 85%, $p < 0.01$ และ 98% vs. 84.2%, $p < 0.01$, ตามลำดับ) และยังพบว่า อัตราการพบผลบวกลวง (False positive alarm rate) น้อยกว่า 5 ครั้ง/การส่องกล้องลำไส้ใหญ่ในทั้ง 2 ระบบและอัตราการพบผลบวกลวงในระบบ CAD EYE จะสูงกว่าระบบ Deep-GI อย่างมีนัยยะสำคัญทางสถิติเมื่อปรับเกณฑ์ชี้วัดสำหรับผลบวกลวงเป็น ≤ 1.5 วินาที แต่อัตราการไม่พบตึงเนื้อลำไส้ใหญ่จะเพิ่มสูงขึ้นเป็น 10-25% ถ้าปรับค่าเกณฑ์ชี้วัดสำหรับผลบวกลวงเป็น ≥ 1.5 วินาที

สรุปผลการวิจัย: การศึกษานี้พบว่า เมื่อเปรียบเทียบประสิทธิภาพระหว่างระบบคอมพิวเตอร์ปัญญาประดิษฐ์ 2 ระบบ ในการการช่วยตรวจจับตึงเนื้อลำไส้ใหญ่ ระบบ Deep-GI มีอัตราการตรวจพบตึงเนื้อลำไส้ใหญ่ (overall polyp detection rate) ที่สูงกว่า ในขณะที่ทำให้เกิดผลบวกลวงน้อยกว่า ที่เกณฑ์ชี้วัดสำหรับผลบวกลวง ≥ 0.5 วินาที และ ≥ 1 วินาที อย่างมีนัยยะสำคัญทางสถิติ ค่าเกณฑ์ชี้วัดสำหรับผลบวกลวงที่ ≥ 1 วินาที เป็นค่าเกณฑ์ชี้วัดที่เหมาะสมสำหรับระบบ Deep-GI และ CAD EYE เนื่องจากเป็นค่าที่ทำให้อัตราการไม่พบตึงเนื้อลำไส้ใหญ่ (polyp miss rate) และเกิดผลบวกลวงต่ำที่สุด

สาขาวิชา	อายุรศาสตร์	ลายมือชื่อนิสิต
ปีการศึกษา	2564	ลายมือชื่อ อ.ที่ปรึกษาหลัก
		ลายมือชื่อ อ.ที่ปรึกษาร่วม

6370076030 : MAJOR MEDICINE

KEYWORD: Colorectal cancer screening Artificial intelligence-assisted colonoscopy Computer-aided polyp detection (CADe) Adenoma detection Colonoscopy

Julalak Karuehardsuwan : Comparative Analysis of Diagnostic Performance Between Two Artificial Intelligence-Assisted Polyp Detection Systems Using a Selected Optimal False Positive Thresholds.. Advisor: Asst. Prof. Parit Mekaroonkamol, M.D. Co-advisor: Prof. RUNGSUN RERKNIMITR, M.D.

BACKGROUND AND AIMS: High false positive (FP) rate in artificial intelligence (AI)-assisted colonoscopy for polyp detection can induce visual disturbance, unnecessary checking on non-pathological areas, and prolonged procedural time. However, a higher FP threshold inevitably results in decreased polyp detection sensitivity. This study aims to compare diagnostic performance between our newly developed AI model and previously validated AI-based computer-aided polyp detection (CADe) system using optimal FP thresholds.

METHODS: We developed an AI software for polyp detection, "Deep-GI", using a convolutional neural network based on the YOLOv5 deep learning framework. Deep-GI was trained and tested using 15,188 and 1,520 colonoscopy images, respectively. Consecutive screening colonoscopy videos were collected prospectively at our center. Two AI models, the validated CAdE system (CAD EYE™; Fujifilm, Tokyo, Japan) and our Deep-GI system, were run and analyzed on each video. The AI-labeled videos were independently reviewed by two experienced gastroenterologists. True positive was defined as a polyp detected by AI that both gastroenterologists confirmed. Another senior gastroenterologist resolved any disagreement. Polyp detection rates, polyp miss rates, and FP rates were calculated and compared using different FP thresholds (≥ 0.5 seconds, ≥ 1 second, ≥ 1.5 seconds, and ≥ 2 seconds) for the duration of an alert box appearing on the screen.

RESULTS: A total of 170 colonoscopies from 170 patients were enrolled. Mean withdrawal time was 7.8 ± 2.7 min. A total of 501 polyps (52.3% adenomatous polyps, $n=262$) were detected in 137 colonoscopies. Three polyps (0.6%) were missed by Deep-GI, and 73 polyps (14.6%) were missed by CAdE with a statistically significant difference; $p<0.01$. Sixty-four (85%) of the missed polyps were <5 mm, while 2.7% were 5-10 mm hyperplastic polyps. When compared to the validated CAdE model, Deep-GI demonstrated higher polyp detection rate (99.4% vs. 85.4%; $p<0.01$), lower polyp miss rate (0.6% vs. 14.6%; $p<0.01$), and lower FP alarm per colonoscopy (349 ± 169 vs. 624 ± 468 ; $p<0.01$). There was a significant difference in adenoma detection between both systems in FP threshold ≥ 0.5 , and ≥ 1 second. Deep-GI has a significantly higher overall polyp detection rate with lower FP alarm than CAdE at ≥ 0.5 , and ≥ 1 second of FP threshold (99.4% vs. 85%; $p<0.01$ and 98% vs. 84.2%; $p<0.01$, respectively). The false-positive-alarm rates were lower than 5 frames/colonoscopy in both groups for all FP thresholds with significantly higher FP rate in the CAdE system when the FP threshold was ≤ 1.5 seconds. The polyp miss rate increased to 10-25% as FP thresholds were adjusted to ≥ 1.5 second.

CONCLUSION: Compared to a validated CAdE system, Deep-GI demonstrated a higher overall polyp detection rate with significantly lower false positive alarm at ≥ 0.5 and ≥ 1 second thresholds. The ≥ 1 second threshold was optimal for the Deep-GI model and CAdE system as it provided the lowest polyp miss rate and false positive alarm for each system.

Field of Study: Medicine
Academic Year: 2021

Student's Signature
Advisor's Signature
Co-advisor's Signature

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยดีด้วยความช่วยเหลือของ ผศ.นพ.ภาฤทธิ์ เมฆอรุณกมล, อ.พญ.เกศินี เขียวกานนท์, รศ.(พิเศษ) พญ.สติมัย อนิวรรณ และ ศ.นพ.รังสรรค์ ฤกษ์นิมิตร หน่วยโรคทางเดินอาหารและตับ ภาควิชาอายุรศาสตร์ คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย และศูนย์ความเป็นเลิศทางการแพทย์ด้านการส่องกล้องระบบทางเดินอาหารแห่งโรงพยาบาลจุฬาลงกรณ์ สภากาชาดไทย อาคารภูมิสิริมังคลานุสรณ์ชั้น 10 ที่ให้คำแนะนำและคำปรึกษาในการทำวิจัย

ขอขอบคุณ ผศ.ดร.พีรพล เวทีกุล และ นายภานุกร สุนทรเวชพงษ์ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่ช่วยร่วมกันคิดค้นและพัฒนาระบบคอมพิวเตอร์ปัญญาประดิษฐ์ Deep-GI และโปรแกรมที่ใช้ในงานวิจัยนี้

ขอขอบคุณ คุณนาฏอนงค์ คงทับ และ คุณชลดา ณะธง หน่วยโรคทางเดินอาหาร โรงพยาบาลจุฬาลงกรณ์สภากาชาดไทย ที่ช่วยดูแลเรื่องเอกสาร ชักประวัติเก็บข้อมูล และดูแลผู้ป่วยในงานวิจัยเป็นอย่างดี ขอขอบคุณ คุณพยาบาลและเจ้าหน้าที่ ศูนย์ความเป็นเลิศทางการแพทย์ด้านการส่องกล้องระบบทางเดินอาหารแห่งโรงพยาบาลจุฬาลงกรณ์สภากาชาดไทย ที่ช่วยในการส่องกล้องและดูแลผู้ป่วยเป็นอย่างดี ขอขอบคุณผู้ป่วยทุกท่านที่ให้ความร่วมมือทำให้งานวิจัยสำเร็จลุล่วงด้วยดี สุดท้ายนี้ขอขอบคุณ บิดา มารดา และน้องสาวที่เป็นกำลังใจสำคัญ



จุฬาลักษณ์ คฤหาสน์สุวรรณ

สารบัญ

	หน้า
.....	ค
บทคัดย่อภาษาไทย.....	ค
.....	ง
บทคัดย่อภาษาอังกฤษ	ง
กิตติกรรมประกาศ	จ
สารบัญ.....	ฉ
สารบัญตาราง.....	ช
สารบัญรูปภาพ.....	ฌ
บทที่ 1.....	1
บทนำ	1
1.1 ความสำคัญและที่มาของปัญหาทางวิจัย (Background and rationale).....	1
1.2 คำถามงานวิจัย (Research question).....	2
1.3 สมมติฐานของการวิจัย (Research hypothesis).....	3
1.4 วัตถุประสงค์ของงานวิจัย (Objectives).....	3
1.5 กรอบแนวคิดในการวิจัย (Conceptual framework)	3
1.6 วิธีดำเนินการวิจัยโดยย่อ (Brief study procedure)	4
1.7 ข้อจำกัดในการวิจัย (Limitation).....	5
1.8 ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย (Expected and application).....	6
บทที่ 2.....	7
ทบทวนวรรณกรรมที่เกี่ยวข้อง	7
บทที่ 3.....	19

วิธีการดำเนินงาน	19
3.1 รูปแบบการวิจัย (Research design).....	19
3.2 ระเบียบวิธีวิจัย (Research methodology).....	19
3.3 การให้คำนิยามเชิงปฏิบัติที่ใช้ในการวิจัย (Operational definition).....	19
3.4 เกณฑ์การคัดเลือกเข้าร่วมโครงการวิจัย (Inclusion criteria).....	20
3.5 เกณฑ์การคัดออกจากโครงการวิจัย (Exclusion criteria).....	20
3.6 การคำนวณขนาดตัวอย่าง (Sample size determination)	21
3.7 ขั้นตอนในการดำเนินการวิจัย (Study processing).....	21
3.8 การรวบรวมข้อมูล (Data collection).....	25
3.9 การวิเคราะห์ข้อมูลและสถิติที่ใช้วิเคราะห์ (Data Analysis and Statistics).....	25
3.10 ข้อพิจารณาด้านจริยธรรม (Ethical considerations)	25
3.11 ข้อจำกัดในการวิจัย (Limitation)	26
3.12 ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย (Expected and application).....	27
3.13 สถานที่ทำวิจัย (Venue of the Study)	27
3.14 การบริหารงานวิจัยและตารางการปฏิบัติงาน (Administration and time schedule) ...	27
บทที่ 4.....	28
ผลการวิจัย	28
บทที่ 5.....	38
อภิปรายผลและข้อเสนอแนะ	38
บรรณานุกรม.....	43
ประวัติผู้เขียน.....	47

สารบัญตาราง

	หน้า
ตารางที่ 1 สรุปการศึกษาของ CADe ที่นำมาใช้ทางคลินิก	12
ตารางที่ 2 Deep-GI dataset และประสิทธิภาพในการตรวจหาติ่งเนื้อลำไส้ใหญ่ของโปรแกรม Deep-GI	29
ตารางที่ 3 แสดงข้อมูลพื้นฐานของผู้ป่วยที่มาส่งกล้องคัดกรองมะเร็งลำไส้ใหญ่และติ่งเนื้อลำไส้ใหญ่ ที่ตรวจพบ (n=170)	31
ตารางที่ 4 แสดงประสิทธิภาพของระบบ Deep-GI เทียบกับระบบ CAD EYE ในการตรวจพบติ่งเนื้อ ลำไส้ใหญ่ทั้งหมด (Diagnostic performance of Deep-GI compared to CAD EYE for overall polyp detection)	32
ตารางที่ 5 แสดงประสิทธิภาพของระบบ Deep-GI เทียบกับระบบ CAD EYE ในการตรวจพบติ่งเนื้อ ชนิด adenoma (Diagnostic performance of Deep-GI compared to CAD EYE for adenoma detection)	34
ตารางที่ 6 แสดงข้อมูลพื้นฐานของติ่งเนื้อลำไส้ใหญ่ที่ตรวจไม่พบ (Baseline characteristics of missed polyps)	35
ตารางที่ 7 แสดงประสิทธิภาพของระบบ Deep-GI เทียบกับระบบ CAD EYE ในการตรวจพบผลบวก ลวงเมื่อใช้ค่าเกณฑ์ชี้วัดสำหรับผลบวกลวงที่ระยะเวลาต่างๆ (Diagnostic performance of Deep-GI system compared to CAD EYE system using different thresholds for false positive alerts)	36

สารบัญรูปภาพ

	หน้า
รูปที่ 1 แสดงกรอบความคิดในการวิจัย (Conceptual framework).....	3
รูปที่ 2 คุณภาพการเตรียมลำไส้ตาม Boston Bowel Preparation Scale (BBPS) 0, 1, 2 และ 3 คะแนนตามลำดับ (14).....	10
รูปที่ 3 การพัฒนา CAD EYE เพื่อนำมาใช้ในการตรวจหาและวินิจฉัยติ่งเนื้อลำไส้ใหญ่.....	13
รูปที่ 4 การต่อกล่อง processor ของโปรแกรม CAD EYE (กล่องสีขาวล่างสุด) เข้ากับ station ของ กล้องปกติเพื่อใช้งาน	13
รูปที่ 5 การปรับ mode ของ CAD EYE โดยใช้ปุ่มเปิด - ปิด ที่บริเวณด้านบนกล้องส่องลำไส้ปกติ และ หน้าจอแสดงผลภาพลำไส้.....	14
รูปที่ 6 ภาพหน้าจอเหมือนการส่องลำไส้ปกติ เพียงแต่มีกรอบสีเหลี่ยมบริเวณที่คาดว่าจะเป็นที่ติ่งเนื้อ จากโปรแกรม CAD EYE	15
รูปที่ 7 Framework การพัฒนาโปรแกรม Deep-GI.....	15
รูปที่ 8 แผนผังการต่อกล่อง processor ของโปรแกรม Deep-GI (กล่องสีขาวในกรอบ) เข้ากับ station ของกล้องปกติเพื่อใช้งาน.....	16
รูปที่ 9 ตัวอย่างผลบวกหลง (false positive).....	17
รูปที่ 10 แสดงขั้นตอนในการดำเนินการวิจัย.....	24
รูปที่ 11 ขั้นตอนในการคัดเลือกและดำเนินการวิจัย.....	30

บทที่ 1

บทนำ

1.1 ความสำคัญและที่มาของปัญหางานวิจัย (Background and rationale)

อุบัติการณ์ของการเกิดโรคมะเร็งลำไส้ใหญ่ในประเทศไทยพบบ่อยเป็นอันดับที่ 3 (1) โดยอุบัติการณ์ในการเกิดมะเร็งลำไส้ใหญ่นั้นจะเพิ่มมากขึ้นในกลุ่มผู้ที่มีอายุตั้งแต่ 50 ปีขึ้นไป ดังนั้นในประเทศไทยการตรวจคัดกรองมะเร็งลำไส้ใหญ่ในประชากรทั่วไปที่ไม่มีความเสี่ยงจึงแนะนำให้เริ่มทำตั้งแต่อายุ 50 ปีขึ้นไป (2) เช่นเดียวกับของสหรัฐอเมริกาและยุโรป ซึ่งการตรวจคัดกรองทำได้หลายวิธี การส่องกล้องลำไส้ใหญ่ (colonoscopy) เป็นวิธีที่ได้รับความนิยม เนื่องจากประสิทธิภาพสูงที่สุดในการตรวจหาติ่งเนื้อ adenoma ในลำไส้ใหญ่ โดยเฉพาะติ่งเนื้อ adenoma ขนาดเล็ก และหากเห็นติ่งเนื้อแล้วสามารถตัดออกได้ในคราวเดียวกัน (3) อย่างไรก็ตามการส่องกล้องลำไส้ใหญ่ยังมีข้อจำกัดเนื่องจากคุณภาพในการตรวจขึ้นกับแพทย์ผู้ทำการส่องกล้องลำไส้ใหญ่ (operator-dependent) รวมถึงกล้องส่องลำไส้ใหญ่ จากงานวิจัยก่อนหน้ามีรายงานการตรวจพบมะเร็งลำไส้ใหญ่ได้ แม้จะเข้ารับการตรวจคัดกรองแล้วก็ตาม (4)

อัตราการตรวจพบติ่งเนื้อลำไส้ชนิด adenoma (adenoma detection rate; ADR) ถือเป็นตัวชี้วัดที่สำคัญที่สุดในการประเมินคุณภาพของการตรวจคัดกรองมะเร็งลำไส้ใหญ่ด้วยการส่องกล้อง จากการศึกษาก่อนหน้านี้พบว่า ADR ที่เพิ่มขึ้นร้อยละ 1 จะทำให้อัตราการเกิดมะเร็งลำไส้ใหญ่หลังการส่องกล้อง (post-colonoscopy colorectal cancer) ลดลงถึงร้อยละ 3 (5) ทางการแพทย์จึงมีความพยายามที่จะเพิ่มอัตราการตรวจพบติ่งเนื้อลำไส้ (Polyp detection rate) เพื่อหวังว่าจะสามารถเพิ่มขีดความสามารถในการป้องกันการเกิดมะเร็งลำไส้ใหญ่และลดอัตราการเสียชีวิตจากมะเร็งลำไส้ใหญ่มาตลอด

ปัจจุบันปัญญาประดิษฐ์ได้เข้ามามีส่วนร่วมทางการแพทย์มากขึ้น โดยมีการพัฒนา computer-aided detection (CADe) และ computer-aided diagnosis (CADx) ขึ้นมาเพื่อช่วยในการตรวจหาและวินิจฉัยชนิดของติ่งเนื้อลำไส้ใหญ่ให้มีประสิทธิภาพมากขึ้น ซึ่งจากการศึกษาก่อนหน้านี้พบว่า CAD สามารถช่วยเพิ่มอัตราการตรวจพบติ่งเนื้อ adenoma ได้จริง (6) โดยล่าสุดทางบริษัท Fujifilm ได้มี software ที่ชื่อ CAD EYE ออกมา แต่อุปกรณ์ดังกล่าวมีราคาแพงและมีข้อจำกัดที่ใช้ได้กับเฉพาะกล้อง colonoscope ของบริษัท Fujifilm เท่านั้น จึงได้มีความร่วมมือกันระหว่างคณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยและหน่วยโรคระบบทางเดินอาหาร

โรงพยาบาลจุฬาลงกรณ์ ในการสร้างและพัฒนาระบบ neural networks ชื่อ Deep-GI ขึ้นสำหรับตรวจจับตึงเนื้อลำไส้

นอกจากประสิทธิภาพของปัญญาประดิษฐ์ในแง่การช่วยเพิ่มอัตราการตรวจพบตึงเนื้อลำไส้ใหญ่แล้ว สิ่งที่สำคัญอีกอย่างคือ จำนวนของผลบวกหลง (false positive) ที่เกิดขึ้นในการส่องลำไส้ใหญ่แต่ละครั้ง โดย false positive คือ การที่ปัญญาประดิษฐ์ขึ้นกรอบเตือนว่ามีตึงเนื้อ ทั้งๆที่ในความเป็นจริงไม่มีตึงเนื้อบริเวณนั้น การที่เกิด false positive alarm อาจทำให้แพทย์ผู้ทำการส่องกล้องเกิดความรำคาญ เสียสมาธิในการส่องกล้อง เสียเวลามากขึ้นในการเข้าไปตรวจเช็คซ้ำบริเวณที่มีกรอบขึ้น ทำให้ใช้เวลาในการส่องกล้องลำไส้ใหญ่นานขึ้น

อย่างไรก็ตาม การปรับลดเกณฑ์ที่ใช้ในการแสดง false positive alarm ของระบบคอมพิวเตอร์ปัญญาประดิษฐ์มักทำให้อัตราการตรวจพบตึงเนื้อลำไส้ใหญ่ลดลง ดังนั้นสิ่งสำคัญสำหรับการสร้าง พัฒนาและทดสอบระบบคอมพิวเตอร์ปัญญาประดิษฐ์คือ การหาค่าเกณฑ์ที่เหมาะสมที่ทำให้แสดง false positive alarm น้อยที่สุดโดยไม่ทำให้อัตราการตรวจพบตึงเนื้อลำไส้ใหญ่ลดลง

ทางกลุ่มผู้วิจัยจึงได้ทำการศึกษาเพื่อเปรียบเทียบประสิทธิภาพระหว่างระบบคอมพิวเตอร์ปัญญาประดิษฐ์ 2 ระบบคือ CAD EYE และ Deep-GI ในการการช่วยตรวจจับตึงเนื้อลำไส้ใหญ่ในผู้ที่มาส่องกล้องคัดกรองมะเร็งลำไส้ใหญ่ และหาค่าเกณฑ์ชี้วัดที่เหมาะสมเพื่อลดอัตราการเกิดผลบวก

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

1.2 คำถามงานวิจัย (Research question)

ประสิทธิภาพของระบบคอมพิวเตอร์ปัญญาประดิษฐ์ 2 ระบบคือ CAD EYE และ Deep-GI ในการการช่วยตรวจจับตึงเนื้อลำไส้ใหญ่ในผู้ที่มาส่องกล้องคัดกรองมะเร็งลำไส้ใหญ่มีความแตกต่างกันหรือไม่ และค่าเกณฑ์ชี้วัดที่เหมาะสมที่จะช่วยลดอัตราการเกิดผลบวกหลงคือเท่าไร

1.3 สมมติฐานของการวิจัย (Research hypothesis)

ประสิทธิภาพของระบบคอมพิวเตอร์ปัญญาประดิษฐ์ 2 ระบบคือ CAD EYE และ Deep-GI ในการการช่วยตรวจจับตึงเนื้อลำไส้ใหญ่ในผู้ที่มาส่องกล้องคัดกรองมะเร็งลำไส้ใหญ่มีประสิทธิภาพดีไม่แตกต่างกัน

1.4 วัตถุประสงค์ของงานวิจัย (Objectives)

วัตถุประสงค์หลัก (primary objective)

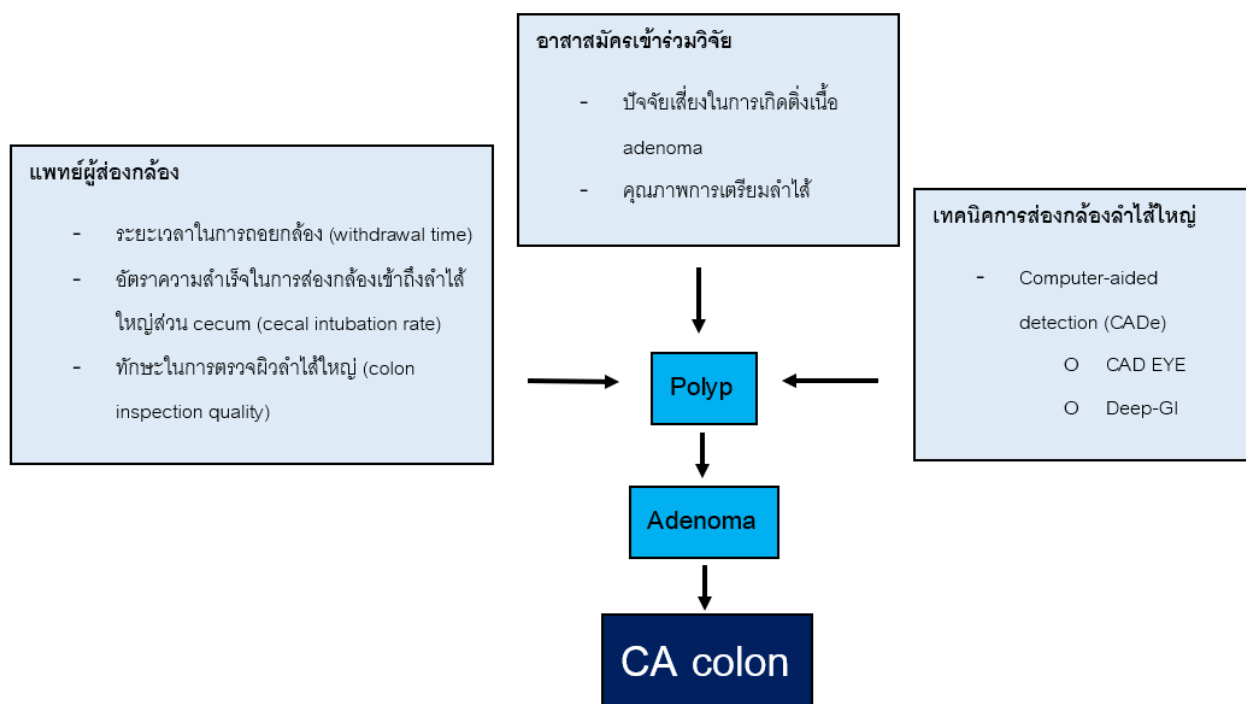
เพื่อศึกษาประสิทธิภาพของระบบคอมพิวเตอร์ปัญญาประดิษฐ์ 2 ระบบคือ CAD EYE และ Deep-GI ในการการช่วยตรวจจับตึงเนื้อลำไส้ใหญ่ในผู้ที่มาส่องกล้องคัดกรองมะเร็งลำไส้ใหญ่ และหาค่าเกณฑ์ชี้วัดที่เหมาะสมที่จะช่วยลดอัตราการเกิดผลบวกลวง

วัตถุประสงค์รอง (secondary objective)

เพื่อศึกษาและเปรียบเทียบอัตราการตรวจพบตึงเนื้อชนิด adenoma, Polyp miss rate, false positive alarm per colonoscopy ระหว่างการใช้ระบบคอมพิวเตอร์ปัญญาประดิษฐ์ 2 ระบบคือ CAD EYE และ Deep-GI ในผู้ที่มาส่องกล้องคัดกรองมะเร็งลำไส้ใหญ่

1.5 กรอบแนวคิดในการวิจัย (Conceptual framework)

รูปที่ 1 แสดงกรอบความคิดในการวิจัย (Conceptual framework)



1.6 วิธีดำเนินการวิจัยโดยย่อ (Brief study procedure)

ขั้นตอนในการดำเนินการวิจัย

- คัดเลือกอาสาสมัครจากกลุ่มผู้ที่มาตรวจคัดกรองมะเร็งลำไส้ใหญ่ที่คลินิกคัดกรองมะเร็งลำไส้ใหญ่ ศูนย์ความเป็นเลิศด้านการส่องกล้องระบบทางเดินอาหาร โรงพยาบาลจุฬาลงกรณ์ที่ตรงกับ inclusion criteria และไม่มี exclusion criteria
- แพทย์ที่ทำการส่องกล้อง จะเป็นแพทย์ที่เชี่ยวชาญการส่องกล้องตรวจลำไส้ โดยมีอัตราการตรวจพบติ่งเนื้อ adenoma ตั้งแต่ร้อยละ 35 ขึ้นไป และก่อนที่จะเข้าร่วมในงานวิจัยนี้ แพทย์ผู้ทำการส่องกล้องจะต้องมีประสบการณ์ในการใช้ CAD EYE อย่างน้อย 20 ครั้ง เพื่อให้แน่ใจว่าสามารถใช้ CAD EYE ได้อย่างมีประสิทธิภาพ
- เริ่มทำการส่องกล้องลำไส้ใหญ่โดยแพทย์ใส่กล้องจากลำไส้ส่วน rectum ขึ้นไปถึง cecum
- ขณะกำลังจะเริ่มถอยกล้องตรวจดูลำไส้ใหญ่ มีการอัด video ขณะถอยกล้องทั้ง video ที่มีระบบ CAD EYE detection และ video ที่ไม่มีระบบ CAD EYE detection (จะได้ videos 2 ไฟล์)

5. เริ่มถอยกล้องตรวจดูลำไส้โดยใช้ระบบ CAD EYE detection ช่วยในการตรวจดูลำไส้ใหญ่ โดยดูอย่างละเอียดตามมาตรฐานระยะเวลาถอยกล้องดูลำไส้อย่างน้อย 6 นาที หากตรวจพบติ่งเนื้อ ลำไส้ใหญ่จะทำการตัดออก และส่งตรวจทางพยาธิวิทยาต่อไป หลังจากนั้นถอยกล้องต่อจนกระทั่งถึง ลำไส้ส่วน rectum และนำกล้องออกจากอาสาสมัคร ผู้วิจัยทำการบันทึกตำแหน่ง จำนวน ขนาด และ รูปร่างติ่งเนื้อที่ตรวจพบรวมถึงระยะเวลาในการถอยดูลำไส้ใหญ่ โดยเวลาในการถอยดูลำไส้ใหญ่จะถูกหยุดหากมีการล้างทำความสะอาดลำไส้ ดูดน้ำ หรือตัดติ่งเนื้อ

6. บันทึกข้อมูลทั้งหมดลงใน case record form

7. นำ program Deep-GI detection มาใส่ใน video ที่ไม่มีระบบ CAD EYE detection ที่ได้บันทึกไว้ก่อนหน้านี้

8. video จากผู้ป่วยรายเดียวกันที่มี CAD EYE detection และ Deep-GI detection รวม 2 videos จะถูกตรวจสอบและบันทึกข้อมูลต่างๆ โดยอายุรแพทย์ทางเดินอาหารที่มีความเชี่ยวชาญ ด้านการส่องกล้องลำไส้ใหญ่ 2 คน ข้อมูลใดที่มีความเห็นไม่ตรงกันจะถูกตรวจสอบซ้ำโดยอายุรแพทย์ ทางเดินอาหารที่มีความเชี่ยวชาญพิเศษ

9. นำทั้ง 2 videos (CAD EYE detection และ Deep-GI detection) ไปนับจำนวน false positive และ duration โดยใช้โปรแกรมที่ทางคณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยจัดทำ ขึ้น

10. คำนวณหาค่า true positive, false negative, false positive rate, polyp detection rates และ polyp miss rates ของทั้ง 2 โปรแกรมปัญญาประดิษฐ์ โดยเปรียบเทียบใน หลายๆเกณฑ์ชี้วัด ทั้งเกณฑ์ชี้วัด ≥ 0.5 วินาที, เกณฑ์ชี้วัด ≥ 1 วินาที, เกณฑ์ชี้วัด ≥ 1.5 วินาที และ เกณฑ์ชี้วัด ≥ 2 วินาที

1.7 ข้อจำกัดในการวิจัย (Limitation)

การส่องกล้องลำไส้ใหญ่โดยใช้ปัญญาประดิษฐ์ทำให้เกิดภาวะแทรกซ้อนได้เท่ากับการส่อง กล้องด้วยวิธีการปกติ แต่หากเกิดภาวะแทรกซ้อนขึ้น ทางคณะผู้วิจัยก็ได้เตรียมแนวทางการแก้ไขไว้ แล้ว เช่น หากมีภาวะลำไส้ทะลุก็สามารถการเย็บปิดลำไส้ใหญ่โดยการส่องกล้องได้ทันที และให้นอน โรงพยาบาลเพื่อสังเกตอาการ หากมีภาวะออกซิเจนในเลือดต่ำกว่าปกติหรือภาวะความดันโลหิตต่ำใน

ระหว่างการให้ยานอนหลับก็สามารถแก้ไขปัญหาได้ทันที เนื่องจากทางทีมผู้วิจัยมีการเฝ้าวัดสัญญาณชีพผู้ป่วยตลอดเวลา (extensive monitoring) และมีอุปกรณ์ช่วยชีวิตขั้นพื้นฐานในห้องส่องกล้อง

ในกรณีที่ความเห็นของอายุรแพทย์ทางเดินอาหารที่มีความเชี่ยวชาญด้านการส่องกล้องลำไส้ใหญ่ 2 คนไม่ตรงกัน video จะถูกตรวจสอบซ้ำโดยอายุรแพทย์ทางเดินอาหารที่มีความเชี่ยวชาญพิเศษ

1.8 ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย (Expected and application)

ได้ทราบถึงประสิทธิภาพของระบบคอมพิวเตอร์ปัญญาประดิษฐ์ 2 ระบบคือ CAD EYE และ Deep-GI ในการช่วยตรวจจับตึงเนื้อลำไส้ใหญ่ในผู้ที่มีมาส่องกล้องคัดกรองมะเร็งลำไส้ใหญ่ และหาค่าเกณฑ์ชีวิตที่เหมาะสมที่จะช่วยลดอัตราการเกิดผลบวกลวง

บทที่ 2

ทบทวนวรรณกรรมที่เกี่ยวข้อง

อุบัติการณ์และความสำคัญของโรคมะเร็งลำไส้ใหญ่

มะเร็งลำไส้ใหญ่และทวารหนักเป็นมะเร็งที่พบได้บ่อยเป็นอันดับที่ 3 และมีอัตราการตายเป็นอันดับที่ 4 ของโรคมะเร็งทั่วโลก โดยมีอุบัติการณ์ 1.4 ล้านรายและเสียชีวิต 700,000 รายในปีค.ศ. 2012 (7) ในประชากรไทยอุบัติการณ์ของการเกิดโรคมะเร็งลำไส้ใหญ่เป็นอันดับที่ 3 ของโรคมะเร็งในเพศชาย และอันดับที่ 5 ของโรคมะเร็งในเพศหญิง (1, 8) โดยจากหลายการศึกษาพบว่าอุบัติการณ์ของการเกิดโรคมะเร็งลำไส้ใหญ่นั้นพบมากกว่าในประเทศที่มีเศรษฐกิจดี แต่ในปัจจุบันอุบัติการณ์ของโรคมะเร็งลำไส้ใหญ่นั้นเพิ่มมากขึ้นทั่วโลก (9) โดยชนิดของมะเร็งที่พบเป็นส่วนมากคือ adenocarcinoma ซึ่งพบได้ร้อยละ 96 ของมะเร็งลำไส้ใหญ่ทั้งหมด (10)

พยาธิกำเนิดมะเร็งลำไส้

การเกิดโรคมะเร็งลำไส้ใหญ่ เซลล์ที่ผิดปกติ (dysplasia) หรือเซลล์มะเร็งมักเกิดที่บริเวณเยื่อบุผิวของลำไส้ใหญ่แล้วจึงลุกลามลงไปในพื้นที่ลึกขึ้น (10) โดยระยะเวลาที่ตั้งเนื้องอกจะกลายเป็นมะเร็งลำไส้ใหญ่ใช้เวลาประมาณ 10 ปี แนวทางการตรวจคัดกรองมะเร็งลำไส้ใหญ่ทั่วโลกจึงแนะนำให้ทำการตรวจคัดกรองอีก 10 ปีถัดไป หากไม่พบตั้งเนื้องอกจากการส่องกล้องลำไส้ใหญ่ (5) ยกเว้นในคนที่มีความเสี่ยงสูงเช่น มีความผิดปกติทางพันธุกรรมที่เสี่ยงต่อการเกิดมะเร็งลำไส้ใหญ่อาจต้องทำทุก 1-2 ปี เช่นในกลุ่มที่เป็น Lynch syndrome (11)

การแบ่งประเภทของตั้งเนื้อในลำไส้ใหญ่ตามลักษณะพยาธิวิทยา (histopathology)

1. ตั้งเนื้องอกโดยทั่วไป (conventional adenomas) เป็นตั้งเนื้องอกที่มีเซลล์ผิดปกติและมีโอกาสเกิดมะเร็งลำไส้ใหญ่ในอนาคต โดยชนิดของเนื้องอกพิจารณาตามหลัก 2 ข้อต่อไปนี้

1.1 ลักษณะของเซลล์ที่ผิดปกติ (dysplasia)

- High grade dysplasia
- Low grade dysplasia

1.2 ส่วนประกอบของ villous ในตั้งเนื้องอก (villousity)

- Tubular adenoma คือ มีส่วนของ villous น้อยกว่าร้อยละ 25
- Tubulovillous adenoma คือ มีส่วนของ villous มากกว่าร้อยละ 25
- Villous adenoma คือ มีส่วนของ villous มากกว่าร้อยละ 75

2. ดั้งเนื้อชนิดแบนราบ (serrated polyps) มีหลายชนิด โดยมีทั้งชนิดที่มีและไม่มีโอกาสกลายเป็นมะเร็งลำไส้ใหญ่ในอนาคต

2.1 Hyperplastic polyps (ดั้งเนื้อที่ไม่ใช่กลายเป็นมะเร็งลำไส้ใหญ่)

2.2 Sessile serrated polyp (SSP) หรือ sessile serrated adenoma (SSA)

2.3 Traditional serrated adenoma

โดยทั้งดั้งเนื้อชนิด sessile serrated adenoma และ traditional serrated adenoma เป็นดั้งเนื้ออกที่สามารถกลายเป็นมะเร็งลำไส้ใหญ่ได้ในอนาคต

เนื้องอกลำไส้ใหญ่ชนิด advanced (advanced adenoma) คือดั้งเนื้ออกขนาดมากกว่าหรือเท่ากับ 1 เซนติเมตร หรือเป็น tubulovillous หรือ villous adenoma หรือมี high grade dysplasia ซึ่งลักษณะเหล่านี้เป็นลักษณะของดั้งเนื้ออกที่มีโอกาสกลายเป็นมะเร็งลำไส้ใหญ่สูง โดยดั้งเนื้ออกที่มีขนาด ≤ 5 มิลลิเมตรพบว่ามีโอกาสเป็นมะเร็งน้อยมาก และดั้งเนื้ออกที่มีขนาด 6-9 มิลลิเมตรพบว่ามีโอกาสเป็นมะเร็งน้อยกว่าร้อยละ 1 (12)

การตรวจคัดกรองมะเร็งลำไส้ใหญ่ด้วยการส่องกล้องลำไส้ใหญ่ (Colonoscopy)

อุบัติการณ์ในการเกิดมะเร็งลำไส้ใหญ่นั้นจะเพิ่มมากขึ้นในกลุ่มผู้ที่มีอายุตั้งแต่ 50 ปีขึ้นไป ดังนั้นการตรวจคัดกรองมะเร็งลำไส้ใหญ่ในประชากรทั่วไปที่ไม่มีความเสี่ยงจึงแนะนำให้เริ่มทำในอายุ 50-75 ปี (2) โดยมีเป้าหมายคือ เพื่อให้พบมะเร็งลำไส้ใหญ่ตั้งแต่ในระยะเริ่มแรก หรือพบดั้งเนื้ออกที่อาจจะกลายเป็นมะเร็งลำไส้ใหญ่ได้ในอนาคตแล้วให้รักษา เช่น การตัดดั้งเนื้ออกออก หรือผ่าตัดลำไส้ใหญ่ในรายที่มีข้อบ่งชี้ เพราะการตรวจพบมะเร็งลำไส้ใหญ่ในระยะหลังนั้นทำให้อัตราการเสียชีวิตเพิ่มขึ้นและสูญเสียค่าใช้จ่ายในการรักษาค่อนข้างมาก (9)

การตรวจคัดกรองมะเร็งลำไส้ใหญ่โดยการส่องกล้องลำไส้ใหญ่มีข้อดีเมื่อเทียบกับวิธีการตรวจคัดกรองแบบอื่นๆ คือมีประสิทธิภาพสูงที่สุดในการตรวจหาดั้งเนื้ออก โดยเฉพาะดั้งเนื้อขนาดเล็ก และหากเห็นดั้งเนื้ออกแล้วสามารถตัดออกได้ในคราวเดียวกัน ดังนั้นหากการส่องกล้องลำไส้ใหญ่ไม่

พบติ่งเนื้องอกก็สามารถตรวจติดตามได้อีก 10 ปีถัดไป อย่างไรก็ตามการส่องกล้องลำไส้ใหญ่ก็ยังมีข้อด้อยเมื่อเทียบกับวิธีอื่นๆ คือต้องเตรียมลำไส้ใหญ่โดยการรับประทานยาระบายปริมาณมาก รวมทั้งยังมีค่าใช้จ่ายที่สูง ต้องใช้เครื่องมือพิเศษ และแพทย์ผู้ชำนาญในการส่องกล้องเนื่องจากเป็นวิธีที่มีความเสี่ยงสูง (3, 5) นอกจากนี้ยังมีข้อจำกัดในเรื่องของแพทย์ที่ทำการส่องกล้องแต่ละรายอาจมีความสามารถต่างกัน (operator-dependent) ทำให้อาจเกิดความผิดพลาดได้ เช่น ส่องกล้องลำไส้ใหญ่ได้ไม่ครบ หรือตรวจไม่พบติ่งเนื้องอกในบางตำแหน่ง (3, 5)

ตัวชี้วัดคุณภาพในการส่องกล้องลำไส้ใหญ่ (3, 13, 14)

1. แพทย์ผู้ทำการส่องกล้องลำไส้ใหญ่ควรมีอัตราการตรวจพบติ่งเนื้อชนิด adenoma (Adenoma detection rate, ADR) โดยรวมอย่างน้อยร้อยละ 25 หรือร้อยละ 30 ในเพศชาย และร้อยละ 20 ในเพศหญิง โดยคำนวณ ADR จากจำนวนผู้เข้ารับการส่องกล้องลำไส้ใหญ่ที่มีติ่งเนื้องอก (adenoma)หารด้วยจำนวนผู้เข้ารับการส่องกล้องลำไส้ใหญ่ทั้งหมด

อัตราการตรวจพบติ่งเนื้อลำไส้ชนิด adenoma (adenoma detection rate; ADR) ถือเป็นตัวชี้วัดที่สำคัญที่สุดในการประเมินคุณภาพของการตรวจคัดกรองมะเร็งลำไส้ใหญ่ด้วยการส่องกล้อง จากการศึกษาที่ผ่านมาพบว่า ADR rates ระดับสูง (ADR: 33.51-52.51%) จะลดอัตราการเกิดมะเร็งลง 50% เมื่อเทียบกับกลุ่มที่ ADR ต่ำ (ADR 7.35-19.05%) และ ADR ที่เพิ่มขึ้นร้อยละ 1 จะทำให้อัตราการเกิดมะเร็งลำไส้ใหญ่หลังการส่องกล้อง (post-colonoscopy colorectal cancer) ลดลงถึงร้อยละ 3 (5)

ADR คำนวณจาก จำนวนผู้ป่วยที่ตรวจพบติ่งเนื้อในลำไส้ใหญ่ (Adenomatous polyp) อย่างน้อย 1 ตี่งเนื้อ หารด้วยจำนวนผู้ป่วยทั้งหมดที่เข้ารับการตรวจ

2. แพทย์ผู้ทำควรมีอัตราความสำเร็จในการส่องกล้องเข้าไปถึงลำไส้ใหญ่ cecum (cecal intubation rate) โดยรวมร้อยละ 90 หรือร้อยละ 95 ในกลุ่มประชากรที่มาตรฐานคัดกรองโรคมะเร็งลำไส้ใหญ่

3. แพทย์ควรมีการบันทึกผลการตรวจทวารหนักที่ตรวจด้วยวิธี digital rectal examination

4. การเตรียมลำไส้ ควรให้มีการรับประทานยาระบายในวันก่อนที่จะเข้ารับการส่องกล้องและเข้าวันที่มาเข้ารับการส่องกล้อง ร่วมกับการจดบันทึกคุณภาพของการเตรียมลำไส้ของผู้เข้ารับการ

ส่องกล้องในแต่ละราย โดยคุณภาพของการเตรียมลำไส้มีเกณฑ์หลายแบบ โดยที่นิยมคือ Boston Bowel Preparation Scale (BBPS) โดยแบ่งเป็นคะแนน 0-3 คะแนน โดยดูจากลำไส้ 3 ส่วนคือ ลำไส้ใหญ่ด้านขวา (Right colon), ตรงกลาง (Transverse colon) และด้านซ้าย (Left colon)

0 คือ ไม่สามารถมองเห็นเยื่อบุผิวลำไส้ได้ เนื่องจากอุจจาระในลำไส้มากไม่สามารถล้างออกได้ระหว่างการส่องกล้อง (unprepared colon segment with mucosa not seen due to solid stool that cannot be cleared)

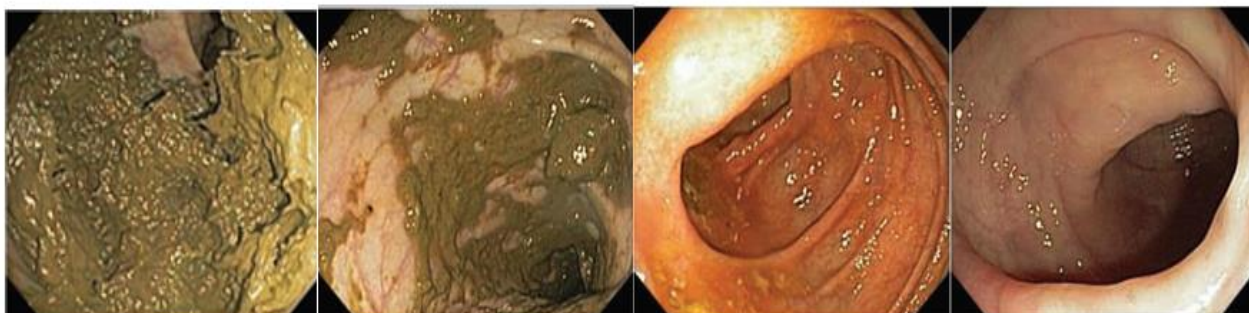
1 คือ สามารถเห็นเยื่อบุผิวลำไส้ในบริเวณนั้นได้เพียงบางส่วน เนื่องจากยังมีอุจจาระค้างอยู่ (portion of mucosa of the colon segment seen, but other areas of the colon segment not well seen due to staining, residual stool and/or opaque liquid)

2 คือ สามารถมองเห็นเยื่อบุผิวลำไส้ในบริเวณนั้นได้เกือบทั้งหมด แต่ยังมีอุจจาระค้างอยู่เล็กน้อย (minor amount of residual staining, small fragments of stool and/or opaque liquid, but mucosa of colon segment seen well)

3 คือ สามารถมองเห็นเยื่อบุผิวลำไส้ได้ชัดเจนทั้งหมด ไม่มีเศษอุจจาระอยู่เลย (entire mucosa of colon segment seen well with no residual staining, small fragments of stool and/or opaque liquid)

โดยคะแนนรวมคือ 0-9 คะแนน ซึ่งคะแนนในการเตรียมลำไส้ที่ยอมรับได้คือตั้งแต่ 5 คะแนน ขึ้นไป (14)

รูปที่ 2 คุณภาพการเตรียมลำไส้ตาม Boston Bowel Preparation Scale (BBPS) 0, 1, 2 และ 3 คะแนนตามลำดับ (14)



5. ระยะเวลาในการถอยกล้องออก (Withdrawal time) ควรนานตั้งแต่ 6 นาทีขึ้นไป รวมถึงต้องมีการถ่ายรูปให้ครบทุกส่วนของลำไส้ใหญ่ เพื่อแสดงให้เห็นว่าตรวจได้ครบทุกจุด

ปัญญาประดิษฐ์ (Artificial intelligent)

การส่องกล้องเพื่อตรวจลำไส้ใหญ่โดยแพทย์นั้น แม้แพทย์ผู้ส่องกล้องจะมีอัตราการตรวจพบติ่งเนื้อ adenoma มากกว่าร้อยละ 40 ก็ยังมีความผิดพลาดได้ค่อนข้างมาก (15)

ปัจจุบันมีการนำปัญญาประดิษฐ์เข้ามาช่วยในการส่องกล้องลำไส้ใหญ่ (computer-aided diagnosis, CAD) ทั้งในแง่ของการตรวจหาและวินิจฉัยชนิดของติ่งเนื้อลำไส้ใหญ่ โดยหลักการประดิษฐ์ส่วนใหญ่ใช้การเรียนรู้แบบ Deep Learning ซึ่งเป็นการเรียนรู้แบบอัตโนมัติเลียนแบบการทำงานของโครงข่ายประสาทของมนุษย์ โดยนำระบบโครงข่ายประสาท (Neural Network) หลายชั้นมาซ้อนกัน และทำการเรียนรู้ข้อมูลตัวอย่าง ซึ่งข้อมูลตัวอย่างจะถูกนำไปใช้ในการตรวจจับรูปแบบหรือจัดหมวดหมู่ข้อมูลเพื่อใช้ประมวลผลในอนาคตต่อไป (16)

ที่ผ่านมาทีมงานวิจัยที่อาศัยการเก็บข้อมูลภาพจากการส่องกล้อง เพื่อทำฐานข้อมูลสำหรับการตรวจจับติ่งเนื้อลำไส้ใหญ่หลายงานวิจัยโดยใช้ Model ทางวิศวกรรมศาสตร์ หลาย model เช่น

1. การใช้ CVC-ClinicDB (17) ซึ่งเป็นชุดข้อมูลรูปที่อยู่ในรูปแบบไฟล์ TIFF ขนาด 384×288 พิกเซล ซึ่งเป็นรูปสี่ที่มี 3 ช่อง โดยมีทั้งหมด 612 รูป ซึ่งเป็นรูปที่ถ่ายโดยโรงพยาบาล Hospital Clinic จากเมือง Barcelona ประเทศสเปน
2. ETIS-LaribPolypDB (18) เป็นชุดข้อมูลรูปที่อยู่ในรูปแบบไฟล์ TIFF ขนาด 1225×966 พิกเซล ซึ่งเป็นรูปสี่ที่มี 3 ช่อง โดยมีทั้งหมด 192 รูป ซึ่งถ่ายโดยโรงพยาบาล Lariboisière Hospital-APHP ประเทศฝรั่งเศส และระบุตำแหน่งของติ่งเนื้อโดย ETIS laboratory, ENSEA, University of Cergy-Pontoise ประเทศฝรั่งเศส
3. Mask R-CNN (19) มีทั้งหมด 2,484 รูป แล้วประเมินผลบน CVC-ClinicDB และ ETIS-LaribPolypDB
4. YOLOv2 (20) ใช้รูปภาพ Polyp จากวิดีโอส่องกล้องจำนวน 8,075 รูปของ Asan Medical Center

การใช้ปัญญาประดิษฐ์ตรวจหาติ่งเนื้อลำไส้ใหญ่ (computer-aided detection, CAdE)

การใช้ CADe มีวัตถุประสงค์เพื่อเพิ่มอัตราการตรวจพบติ่งเนื้อ adenoma และลดอัตราการตรวจไม่พบติ่งเนื้อลำไส้ใหญ่ (polyp miss rate) (21) ต่อมาเริ่มมีการนำ CADe มาใช้จริงในการส่องกล้องลำไส้ใหญ่ในผู้เข้ารับการส่องกล้องจำนวน 1058 ราย โดยเปรียบเทียบระหว่างการส่องกล้องโดยใช้ CADe กับการส่องกล้องปกติ (ไม่ใช่ CADe) พบว่าอัตราการตรวจพบติ่งเนื้อ adenoma จากการส่องกล้องโดยใช้ CADe สูงกว่าการส่องกล้องปกติอย่างมีนัยสำคัญทางสถิติ คือร้อยละ 29.1 และร้อยละ 20.3 ตามลำดับ ($P < 0.001$) และจำนวนติ่งเนื้อ adenoma เฉลี่ยต่อการส่องกล้องแต่ละครั้ง (Adenomas per colonoscopy, APC) ในกลุ่มที่ใช้ CADe สูงกว่ากลุ่มที่ไม่ใช้ CADe คือ 0.53 และ 0.31 ตามลำดับ ($P < 0.001$) โดยติ่งเนื้อ adenoma ที่พบเพิ่มส่วนใหญ่เป็นที่บริเวณลำไส้ใหญ่ส่วน transverse, descending และ sigmoid และติ่งเนื้อที่พบเพิ่มมักมีขนาดไม่เกิน 5 มิลลิเมตรและมีลักษณะแบนราบ (flat หรือ sessile) (22) ต่อมามีการทำ meta-analysis ที่มีประชากรที่ศึกษารวม 2815 ราย (กลุ่มที่ใช้ CADe 1400 ราย และกลุ่มที่ส่องกล้องปกติ 1415 ราย) พบว่ากลุ่มที่ใช้ CADe มีอัตราการตรวจพบติ่งเนื้อ adenoma สูงกว่ากลุ่มที่ไม่ใช้คือร้อยละ 32.9 และร้อยละ 20.8 ตามลำดับ ($P < 0.001$) (6) ซึ่งจาก meta-analysis ดังกล่าวข้างต้นมีอัตราการตรวจพบติ่งเนื้อ adenoma ค่อนข้างน้อยซึ่งอาจเป็นผลจากผู้เข้าร่วมวิจัยที่อายุก่อนข้าน้อย โดยมีอายุเฉลี่ยคือ 50 ปี และข้อบ่งชี้ในการส่องกล้องลำไส้ใหญ่มีทั้งผู้ป่วยที่มีอาการและผู้ป่วยที่เข้ารับการตรวจคัดกรองมะเร็งลำไส้ใหญ่

ตารางที่ 1 สรุปการศึกษาของ CADe ที่นำมาใช้ทางคลินิก

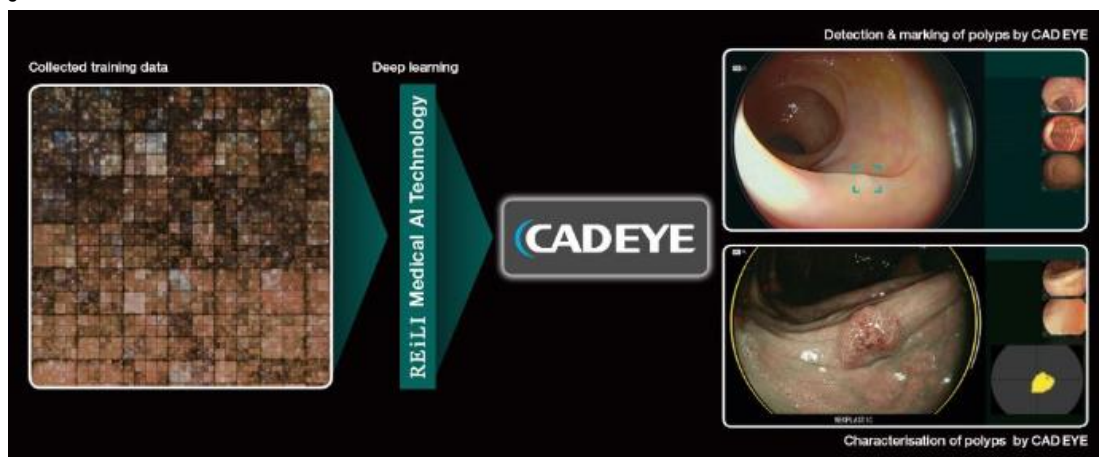
จำนวนประชากรที่ศึกษา	รูปแบบการศึกษา	กลุ่มศึกษาที่ใช้เทคนิค	กลุ่มควบคุม	ADR (ร้อยละ)	APC	เอกสารอ้างอิง
1058 คน	RCT	CADe	SC	29.1 vs. 20.3 ($P < 0.001$)	0.53 vs. 0.31 ($P < 0.001$)	(22)
1026 คน	RCT	CADe	SC	39.1 vs. 23.9 ($P < 0.001$)	0.52 vs. 0.34 ($P < 0.001$)	(23)
659 ราย	RCT	CADe	SC	28.9 vs. 16.5 ($P < 0.001$)	0.37 vs 0.18 ($P < 0.001$)	(24)

ADR คืออัตราการตรวจพบติ่งเนื้อ adenoma, APC คือจำนวนติ่งเนื้อ adenoma เฉลี่ยต่อการส่องกล้องลำไส้ใหญ่แต่ละครั้ง, CADe คือ computer-aided colonoscopy, SC คือ standard colonoscopy, NS คือ not significant

CAD EYE (25)

CAD EYE เป็นปัญญาประดิษฐ์ของบริษัท Fujifilm ที่พัฒนาออกมาเพื่อตรวจหาและวินิจฉัยชนิดของติ่งเนื้อลำไส้ใหญ่ (detection and characterization) ในเครื่องเดียวกัน โดยได้ CE mark แล้ว และเริ่มวางขายในเดือนมีนาคม ค.ศ. 2020 โดยการพัฒนาเครื่องใช้กระบวนการ deep learning เพื่อเรียนรู้ภาพทั้งหมดที่ใส่เข้าไปในระบบ

รูปที่ 3 การพัฒนา CAD EYE เพื่อนำมาใช้ในการตรวจหาและวินิจฉัยติ่งเนื้อลำไส้ใหญ่

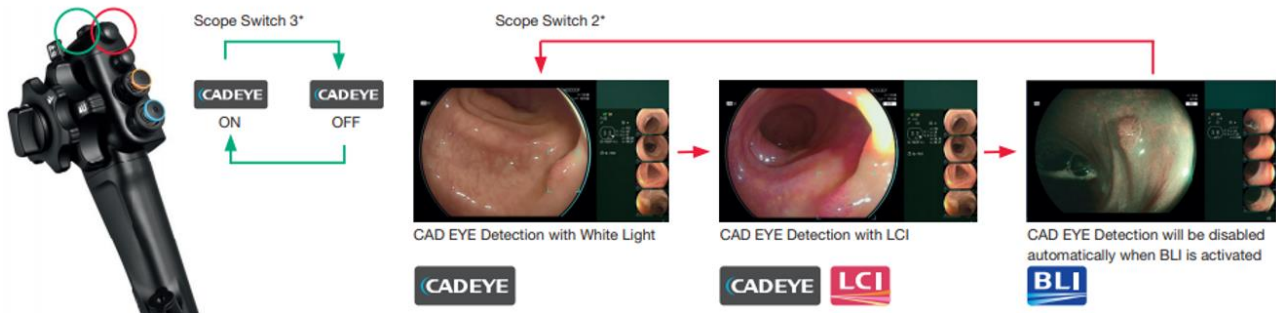


CAD EYE เป็น software ในกล่อง processor สามารถนำมาต่อใช้กับกล้องรุ่น ELUXEO 7000 ซึ่งเป็นกล้องมาตรฐานที่ใช้ในปัจจุบันของรพ. จุฬาลงกรณ์ได้เลย การใช้งานสามารถเปิดใช้งานทำงานแต่ละ mode ได้แบบ real-time ในขณะที่ส่องกล้องลำไส้ใหญ่เพียงแคกดปุ่มบริเวณด้ามกล้อง ซึ่งการทำงานของ CAD EYE จะแบ่งเป็น 2 ระบบคือ CAD EYE Detection และ CAD EYE Characterization

รูปที่ 4 การต่อกล่อง processor ของโปรแกรม CAD EYE (กล่องสีขาวล่างสุด) เข้ากับ station ของกล้องปกติเพื่อใช้งาน

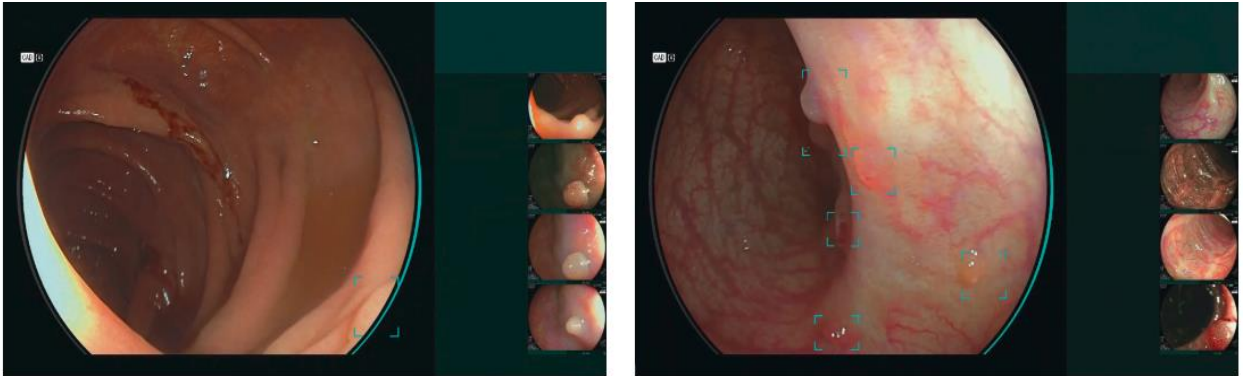


รูปที่ 5 การปรับ mode ของ CAD EYE โดยใช้ปุ่มเปิด - ปิด ที่บริเวณด้ามกล้องส่องลำไส้ปกติ และหน้าจอแสดงภาพลำไส้



CAD EYE Detection เป็นการตรวจหาติ่งเนื้อจากการส่องกล้องลำไส้ใหญ่ เมื่อ AI นี้ ตรวจพบติ่งเนื้อจะมีเสียงเตือนดังขึ้น และมีกรอบสี่เหลี่ยมบริเวณที่คาดว่าจะเป็นติ่งเนื้อ

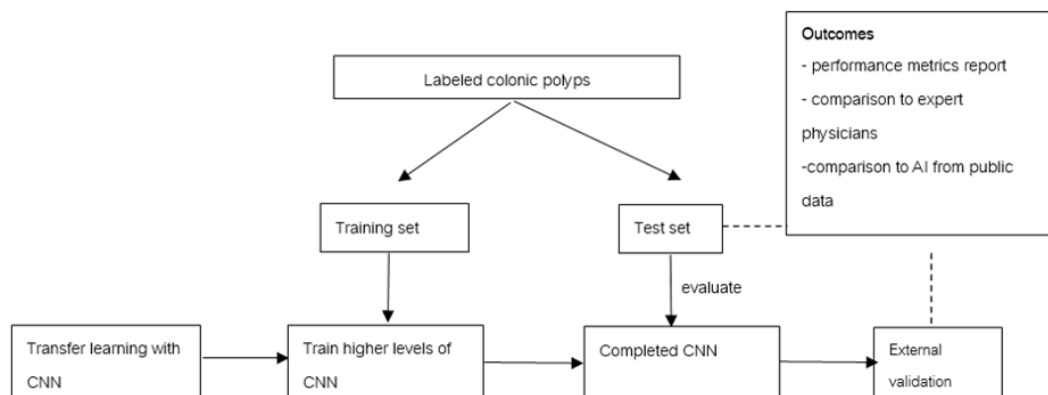
รูปที่ 6 ภาพหน้าจอเหมือนการส่องลำไส้ปกติ เพียงแต่มีกรอบสี่เหลี่ยมบริเวณที่คาดว่าจะเป็นที่ตั้งเนื้อ จากโปรแกรม CAD EYE



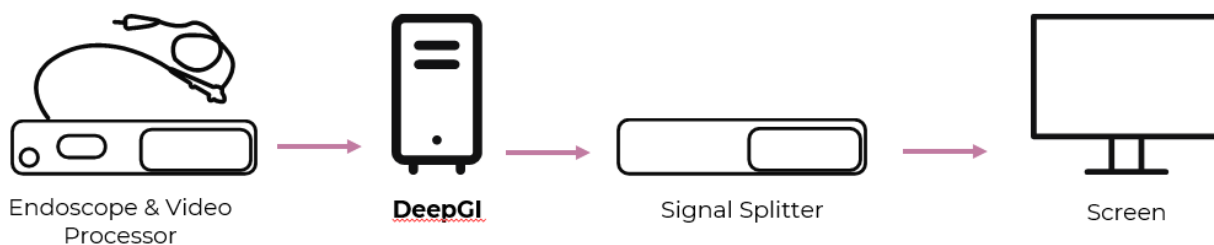
Deep-GI

Deep-GI เป็นปัญญาประดิษฐ์ที่เกิดจากความร่วมมือกันระหว่างคณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยและหน่วยโรกระบบทางเดินอาหาร โรงพยาบาลจุฬาลงกรณ์ เพื่อตรวจหาตั้งเนื้อลำไส้ใหญ่ โดยการพัฒนาเครื่องใช้กระบวนการเรียนรู้เชิงลึกของคอมพิวเตอร์ หรือที่เรียกว่า deep learning โดยใช้ Model ทางวิศวกรรมศาสตร์ ชื่อ YOLOv5 deep learning framework เพื่อเรียนรู้ภาพทั้งหมดที่ใส่เข้าไป ซึ่งเป็นการเรียนรู้แบบอัตโนมัติเลียนแบบการทำงานของโครงข่ายประสาทของมนุษย์ โดยนำระบบโครงข่ายประสาท (Neural Network) หลายชั้นมาซ้อนกัน และทำการเรียนรู้ข้อมูลตัวอย่าง ซึ่งข้อมูลตัวอย่างจะถูกนำไปใช้ในการตรวจจับรูปแบบหรือจัดหมวดหมู่ข้อมูลเพื่อใช้ประมวลผลต่อไป

รูปที่ 7 Framework การพัฒนาโปรแกรม Deep-GI



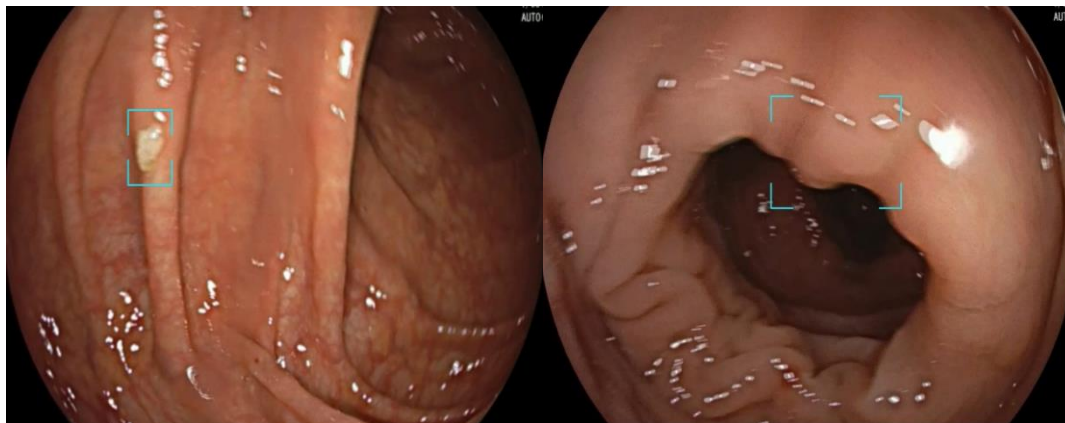
รูปที่ 8 แผนผังการต่อกล่อง processor ของโปรแกรม Deep-GI (กล่องสีขาวในกรอบ) เข้ากับ station ของกล้องปกติเพื่อใช้งาน



False positive alarm

นอกจากประสิทธิภาพของปัญญาประดิษฐ์ในแง่การช่วยเพิ่มอัตราการตรวจพบติ่งเนื้อลำไส้ใหญ่แล้ว สิ่งที่สำคัญอีกอย่างคือ จำนวนของผลบวกลวง (false positive) ที่เกิดขึ้นในการส่องลำไส้ใหญ่แต่ละครั้ง โดย false positive คือ การที่ปัญญาประดิษฐ์ขึ้นกรอบเตือนว่ามีติ่งเนื้อ ทั้งๆที่ในความเป็นจริงไม่มีติ่งเนื้อบริเวณนั้น โดยผลบวกลวงอาจเกิดจากฟอง เศษอุจจาระ หรือรอยพับของลำไส้ใหญ่ การที่เกิด false positive alarm อาจทำให้แพทย์ผู้ทำการส่องกล้องเกิดความรำคาญ เสียสมาธิในการส่องกล้อง เสียเวลามากขึ้นในการเข้าไปตรวจเช็คซ้ำบริเวณที่มีกรอบขึ้น ทำให้ใช้เวลาในการส่องกล้องลำไส้ใหญ่นานขึ้น

รูปที่ 9 ตัวอย่างผลบวกกลาง (false positive)



อย่างไรก็ตาม การปรับลดเกณฑ์ที่ใช้ในการแสดง false positive alarm ของระบบคอมพิวเตอร์ปัญญาประดิษฐ์มักทำให้อัตราการตรวจพบติ่งเนื้อลำไส้ใหญ่ลดลง ดังนั้นสิ่งที่สำคัญสำหรับการสร้าง พัฒนาและทดสอบระบบคอมพิวเตอร์ปัญญาประดิษฐ์คือ การหาค่าเกณฑ์ที่เหมาะสมที่ทำให้แสดง false positive alarm น้อยที่สุดโดยที่ไม่ทำให้อัตราการตรวจพบติ่งเนื้อลำไส้ใหญ่ลดลง

มีการศึกษาโดยใช้ระบบคอมพิวเตอร์ปัญญาประดิษฐ์ที่พัฒนาขึ้นเพื่อช่วยในการตรวจจับติ่งเนื้อลำไส้ใหญ่ในคนไข้ทั้งหมด 62 คน โดยมีการปรับเกณฑ์ชีวิตเป็นระยะเวลาต่างๆ เพื่อหาเกณฑ์ที่เหมาะสมที่ทำให้เกิดผลบวกกลางน้อยที่สุด โดยพบว่า ที่เกณฑ์ชีวิต ≥ 0.5 วินาที มีความจำเพาะ (specificity) คือร้อยละ 93.2 และความถูกต้อง (accuracy) คือร้อยละ 97.8 ที่เกณฑ์ชีวิต ≥ 1 วินาที มีความจำเพาะคือร้อยละ 98.6 และความถูกต้องคือร้อยละ 99.5 ในขณะที่ ที่เกณฑ์ชีวิต ≥ 2 วินาที มีความจำเพาะคือร้อยละ 99.8 และความถูกต้องคือร้อยละ 99.9 ทางผู้วิจัยพบว่า เกณฑ์ชีวิต ≥ 2 วินาทีที่น่าจะเป็นเกณฑ์ที่เหมาะสมที่สุดที่จะนำมาใช้ในเวชปฏิบัติ เนื่องจากเป็นเวลาที่มากพอที่จะใช้ในการตรวจสอบเรื่องผลบวกกลาง และ ไม่ได้เกิดผลบวกกลางมากจนเกินไป (26)

ทางกลุ่มผู้วิจัยจึงได้ทำการศึกษาเพื่อเปรียบเทียบประสิทธิภาพระหว่างระบบคอมพิวเตอร์ปัญญาประดิษฐ์ 2 ระบบคือ CAD EYE และ Deep-GI ในการการช่วยตรวจจับติ่งเนื้อลำไส้ใหญ่ในผู้ที่มาส่งกล้องคัดกรองมะเร็งลำไส้ใหญ่ และหาค่าเกณฑ์ชีวิตที่เหมาะสมเพื่อลดอัตราการเกิดผลบวกกลาง โดยเหตุผลที่เลือกเปรียบเทียบ Deep GI กับ CAD eye ของบริษัท Fujifilm เนื่องจากเป็นระบบคอมพิวเตอร์ปัญญาประดิษฐ์ที่ได้รับการพัฒนามาเป็นเวลานาน มีการศึกษาเกี่ยวกับประสิทธิภาพใน

การตรวจหาและวินิจฉัยชนิดของต้งเนื้อลำไส้ใหญ่ (detection and characterization) มากมายว่า
สามารถใช้ได้จริง (sensitivity และ accuracy 95-98%)



บทที่ 3 วิธีการดำเนินงาน

การวิจัยนี้เป็นการศึกษาในผู้มาตรวจส่องกล้องคัดกรองมะเร็งลำไส้ใหญ่ ซึ่งเข้ารับการตรวจรักษาที่คลินิกโรคทางเดินอาหาร โรงพยาบาลจุฬาลงกรณ์

3.1 รูปแบบการวิจัย (Research design)

เป็นการศึกษาแบบ Prospective single center study

3.2 ระเบียบวิธีวิจัย (Research methodology)

ประชากรเป้าหมาย (Target population)

ผู้มาตรวจส่องกล้องคัดกรองมะเร็งลำไส้ใหญ่อายุ 50-75 ปี ในโรงพยาบาลจุฬาลงกรณ์

3.3 การให้คำนิยามเชิงปฏิบัติที่ใช้ในการวิจัย (Operational definition)

1. การส่องกล้องลำไส้ใหญ่คือ การส่องกล้องจากทวารหนักเข้าไปจนถึงลำไส้ใหญ่ส่วน cecum
2. เวลาที่ใช้ในการส่องกล้องลำไส้ใหญ่ขาเข้า เริ่มตั้งแต่กล้องเข้าทวารหนักจนถึงลำไส้ใหญ่ส่วน cecum
3. เวลาที่ใช้ในการถอยกล้อง เริ่มจับเวลาตั้งแต่ถอยกล้องออกจากลำไส้ใหญ่ส่วน cecum จนถึงออกมานอกร่างกายผู้เข้าร่วมวิจัย โดยจะหยุดเวลาเมื่อมีการฉีดยาล้างผนังลำไส้ ดูดน้ำ หรือมีการตัดติ่งเนื้อลำไส้ใหญ่
4. การตรวจคัดกรองมะเร็งลำไส้ใหญ่ในผู้ที่มีความเสี่ยงในการเกิดมะเร็งลำไส้ใหญ่เทียบเท่ากับประชากรทั่วไป เริ่มทำตั้งแต่อายุ 50 ปีขึ้นไป
5. Adenoma detection rate (ADR) คือ อัตราการตรวจพบติ่งเนื้อลำไส้ชนิด adenoma ADR คำนวณจาก จำนวนผู้ป่วยที่ตรวจพบติ่งเนื้อในลำไส้ใหญ่ (Adenomatous polyp) อย่างน้อย 1 ติ่งเนื้อหารด้วยจำนวนผู้ป่วยทั้งหมดที่เข้ารับการตรวจ

6. Adenoma per colonoscopy (APC) คือ จำนวนติ่งเนื้อ adenoma เฉลี่ยต่อการส่องกล้องลำไส้ใหญ่แต่ละครั้ง

APC คำนวณจาก จำนวนติ่งเนื้อในลำไส้ใหญ่ (Adenomatous polyp) ทั้งหมดที่ตรวจพบหารด้วยจำนวนการส่องกล้องลำไส้ใหญ่ในผู้ป่วยทั้งหมดที่เข้ารับการตรวจ

7. ผลบวกหลง (false positive) คือ การที่ปัญญาประดิษฐ์ขึ้นกรอบเตือนว่ามีติ่งเนื้อ ทั้งๆ ที่ในความเป็นจริงไม่มีติ่งเนื้อบริเวณนั้น

8. ผลบวกจริง (true positive) คือ การที่ปัญญาประดิษฐ์ขึ้นกรอบเตือนว่ามีติ่งเนื้อในบริเวณที่มีติ่งเนื้อจริง

9. ผลลบหลง (false negative) คือ การที่ปัญญาประดิษฐ์ไม่ขึ้นกรอบเตือนว่ามีติ่งเนื้อในบริเวณที่มีติ่งเนื้อจริง

3.4 เกณฑ์การคัดเลือกเข้าร่วมโครงการวิจัย (Inclusion criteria)

ผู้ที่มีอายุ 50-75 ปี ที่เข้ารับการตรวจคัดกรองมะเร็งลำไส้ใหญ่และเป็นผู้สมัครใจเข้าร่วมโครงการด้วยตนเอง

3.5 เกณฑ์การคัดออกจากโครงการวิจัย (Exclusion criteria)

1. ผู้ที่มีโรคทางพันธุกรรมที่ส่งผลให้มีโอกาสเกิดเนื้องอกหรือมะเร็งลำไส้ใหญ่เพิ่มมากขึ้น เช่น familial adenomatous polyposis, Lynch syndrome หรือเคยเป็นมะเร็งลำไส้ใหญ่มาก่อน
2. ผู้ที่มีโรคลำไส้อักเสบ (Inflammatory bowel disease)
3. ผู้ที่เคยได้รับการผ่าตัดลำไส้ใหญ่
4. มีอาการผิดปกติเช่น ถ่ายเป็นเลือด ถ่ายท้องผูกสลับท้องเสีย โลหิตจางโดยไม่ทราบสาเหตุ หรือ น้ำหนักลดผิดปกติ
5. ผู้ที่แพทย์ไม่สามารถส่องกล้องเข้าไปถึงลำไส้ใหญ่ cecum ได้สำเร็จ

3.6 การคำนวณขนาดตัวอย่าง (Sample size determination)

เนื่องจากงานวิจัยนี้เป็นการเปรียบเทียบประสิทธิภาพของระบบคอมพิวเตอร์ปัญญาประดิษฐ์ 2 ระบบคือ CAD EYE และ Deep-GI ในการช่วยตรวจจับตึงเนื้อลำไส้ใหญ่ จึงคำนวณ sample size โดยใช้สูตรสำหรับ non-inferiority or superiority trial for binary data

$$n_2 = \frac{(z_{1-\alpha} + z_{1-\beta})^2}{(\epsilon - \delta)^2} \left[\frac{p_1(1-p_1)}{k} + p_2(1-p_2) \right]$$

$$\epsilon = p_1 - p_2$$

$$k = \frac{n_1}{n_2}$$

$$n_1 = kn_2$$

Proportion in group1 (p_1) = 0.970 (ความถูกต้องในการตรวจพบตึงเนื้อลำไส้ใหญ่ จากงานวิจัย CADe อื่นคือร้อยละ 97) (26-27)

Proportion in group2 (p_2) = 0.990 (สมมติให้ความถูกต้องในการตรวจพบตึงเนื้อลำไส้ใหญ่โดยการใช้โปรแกรม Deep-GI คือร้อยละ 99)

Non-inferiority or superiority margin (δ) = 0.060

Ratio between 2 groups (k) = 1.0

Alpha (α) = 0.05, $Z(0.950) = 1.96$

Beta (β) = 0.20, $Z(0.800) = 0.841621$

เมื่อแทนค่าในสูตรข้างต้นจะคิด Sample size Group1 (n_1) = 151 โดยกำหนดอำนาจการทดสอบ (power) ที่ 80% และความแตกต่างอย่างมีนัยสำคัญ (2-sided significant level) ที่ 5% และเมื่อคิดเผื่อ drop out ประมาณ 10% จะได้ $n = 170$ ต่อกลุ่ม

3.7 ขั้นตอนในการดำเนินการวิจัย (Study processing)

1. คัดเลือกอาสาสมัครจากกลุ่มผู้ที่มาตรวจคัดกรองมะเร็งลำไส้ใหญ่ที่คลินิกคัดกรองมะเร็งลำไส้ใหญ่ ศูนย์ความเป็นเลิศด้านการส่องกล้องระบบทางเดินอาหาร โรงพยาบาลจุฬาลงกรณ์ที่ตรงกับ inclusion criteria และไม่มี exclusion criteria

2. ผู้วิจัยอธิบายขั้นตอน ผลดี ผลเสีย และวิธีการส่องกล้องลำไส้ใหญ่ให้ผู้เข้าร่วมการวิจัยรับทราบ และให้ลงชื่อยินยอมเข้าร่วมงานวิจัยโดยสมัครใจ
3. ผู้ช่วยวิจัยทำการนัดหมายวันส่องกล้อง และให้คำแนะนำขั้นตอนการปฏิบัติตัวก่อนมาส่องกล้องลำไส้ ได้แก่ การเตรียมอาหาร และการรับประทานยาระบายเตรียมลำไส้ก่อนตรวจ
4. วันส่องกล้องตรวจลำไส้ ผู้ช่วยวิจัยจะตรวจสอบและประเมินคุณภาพการเตรียมลำไส้ก่อนส่องกล้อง หากคุณภาพการเตรียมลำไส้ยังไม่สะอาดเพียงพอจะให้รับประทานยาระบายเพิ่มที่ห้องเตรียมตัวก่อนส่องกล้อง ตามขั้นตอนปฏิบัติเช่นเดียวกับผู้ป่วยอื่นทั่วไป
5. ผู้เข้าร่วมวิจัยที่จะได้รับยานอนหลับ (conscious sedation) ระหว่างการส่องกล้องลำไส้ใหญ่ โดยขณะให้ยานอนหลับจะมีพยาบาลเฝ้าระวังสัญญาณชีพของผู้เข้าร่วมวิจัยตลอดเวลา ตามขั้นตอนปฏิบัติเช่นเดียวกับผู้ป่วยอื่นทั่วไป
6. แพทย์ที่ทำการส่องกล้อง จะเป็นแพทย์ที่ความเชี่ยวชาญการส่องกล้องตรวจลำไส้ โดยมีอัตราการตรวจพบติ่งเนื้อ adenoma ตั้งแต่ร้อยละ 35 ขึ้นไป และก่อนที่จะเข้าร่วมในงานวิจัยนี้ แพทย์ผู้ทำการส่องกล้องจะต้องมีประสบการณ์ในการใช้ CAD EYE อย่างน้อย 20 ครั้ง เพื่อให้แน่ใจว่าสามารถใช้ CAD EYE ได้อย่างมีประสิทธิภาพ
7. เริ่มทำการส่องกล้องลำไส้ใหญ่โดยแพทย์ใส่กล้องจากลำไส้ส่วน rectum ขึ้นไปถึง cecum
8. ขณะกำลังจะเริ่มถอยกล้องตรวจดูลำไส้ใหญ่ มีการอัด video ขณะที่ถอยกล้องทั้ง video ที่มีระบบ CAD EYE detection และ video ที่ไม่มีระบบ CAD EYE detection (จะได้ videos 2 ไฟล์)
9. เริ่มถอยกล้องตรวจดูลำไส้โดยใช้ระบบ CAD EYE detection ช่วยในการตรวจดูลำไส้ใหญ่ โดยดูอย่างละเอียดตามมาตรฐานระยะเวลาถอยกล้องดูลำไส้อย่างน้อย 6 นาทีและไม่เกิน 10 นาที หากเป็นไปได้ (เพื่อป้องกันการตรวจที่นานเกินปกติ) หากตรวจพบติ่งเนื้อลำไส้ใหญ่จะทำการตัดออกและส่งตรวจทางพยาธิวิทยาต่อไป หลังจากนั้นถอยกล้องต่อจนกระทั่งถึงลำไส้ส่วน rectum และนำกล้องออกจากอาสาสมัคร ผู้วิจัยทำการบันทึกตำแหน่ง จำนวน ขนาด และรูปร่างติ่งเนื้อที่ตรวจพบ รวมถึงระยะเวลาในการถอยดูลำไส้ใหญ่ โดยเวลาในการถอยดูลำไส้ใหญ่จะถูกหยุดหากมีการล้างทำความสะอาดลำไส้ ดูดน้ำ หรือตัดติ่งเนื้อ

10. หลังจากส่องกล้องเสร็จ ผู้เข้าร่วมวิจัยจะถูกย้ายไปในห้องพักพื้นที่มีการตรวจวัดสัญญาณชีพเป็นระยะจนผู้เข้าร่วมวิจัยตื่นระดับความรู้สึกตัวดี ผู้วิจัยแจ้งผลการส่องกล้องเบื้องต้น และในผู้ป่วยที่มีตัดชิ้นเนื้อจะการนัดตรวจติดตาม เพื่อฟังผลการตรวจและผลชิ้นเนื้อที่แผนกผู้ป่วยนอก

11. จะพิจารณาคัดผู้เข้าร่วมงานวิจัยออกจากการศึกษา (drop out) ในกรณี

- คุณภาพของการเตรียมลำไส้แย่ (poor) คือ Boston Bowel Preparation Scale (BBPS) < 5 คะแนน
- ไม่สามารถทำการส่องกล้องลำไส้ใหญ่ได้จนถึง cecum (incomplete colonoscopy)
- สัญญาณชีพไม่คงที่ (hemodynamic instability) หรือมีผลแทรกซ้อนในระหว่างการส่องกล้อง

12. บันทึกข้อมูลทั้งหมดลงใน case record form

13. นำ video ที่บันทึกไว้มาทำการตัดต่อ โดยหลังจากที่เจอตังเนื้อลำไส้ใหญ่และแพทย์ผู้ทำการส่องกล้องได้ทำจากพิจารณา ถ่ายรูปและตัดสินใจเรื่องการตัดตังเนื้อแล้ว ช่วงเวลาที่แพทย์ผู้ทำการส่องกล้องทำการตัดชิ้นเนื้อจะถูกตัดออกไป เพื่อดูระยะเวลาในการถอยดูลำไส้ใหญ่รวมทั้งหมด (เวลาในการถอยดูลำไส้ใหญ่จะถูกหยุดหากมีการล้างทำความสะอาดลำไส้ ดูดน้ำ หรือตัดตังเนื้อ)

14. นำ program Deep-GI detection มาใส่ใน video ที่ไม่มีระบบ CAD EYE detection ที่ได้บันทึกและตัดต่อไว้ไว้ก่อนหน้า

15. video จากผู้ป่วยรายเดียวกันที่มี CAD EYE detection และ Deep-GI detection รวม 2 videos ที่ถูกตัดต่อเหมือนกัน จะถูกตรวจสอบและบันทึกข้อมูลต่างๆ โดยอายุรแพทย์ทางเดินอาหารที่มีความเชี่ยวชาญด้านการส่องกล้องลำไส้ใหญ่ 2 คนตลอดการศึกษาวินิจฉัย ข้อมูลใดที่มีความเห็นไม่ตรงกันจะถูกตรวจสอบซ้ำโดยอายุรแพทย์ทางเดินอาหารที่มีความเชี่ยวชาญพิเศษคนที่ 3 และนับความเห็นที่ผู้เชี่ยวชาญเห็นตรงกันอย่างน้อย 2 ใน 3 เป็นความเห็นที่ถูกต้อง

16. ข้อมูลที่ตรวจสอบและบันทึกจาก video

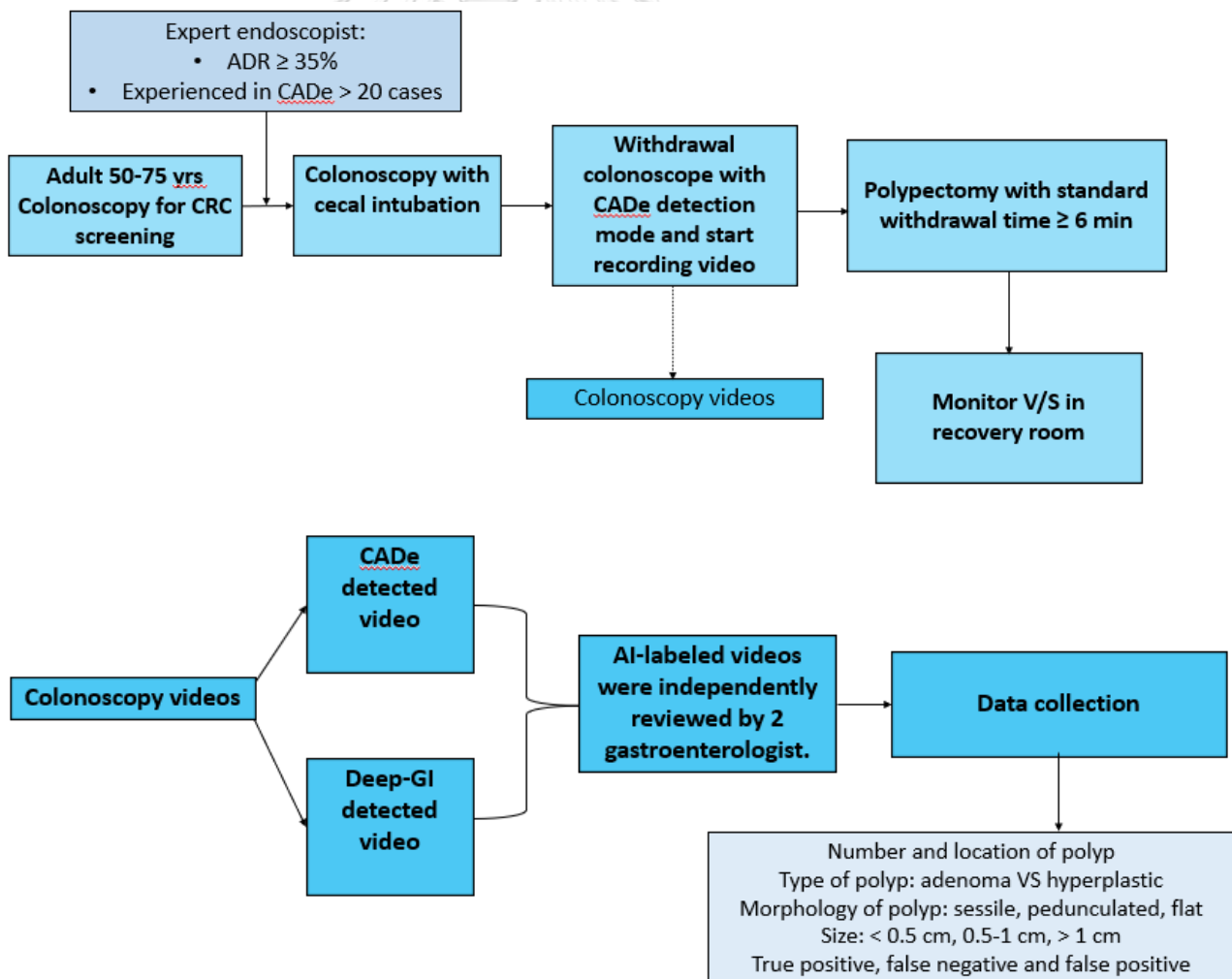
- จำนวนตังเนื้อทั้งหมด ตำแหน่ง
- รูปร่าง (sessile, pedunculated, flat)
- ขนาด (< 0.5 cm, 0.5-1 cm, > 1 cm)

- ชนิด (adenoma, hyperplastic)
- true positive, false negative และ false positive

17. นำทั้ง 2 videos (CAD EYE detection และ Deep-GI detection) ไปนับจำนวน false positive และ duration โดยใช้โปรแกรมที่เขียนขึ้นมาเพื่อนับกรอบปัญหาประดิษฐ์โดยเฉพาะที่ทางคณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยจัดทำขึ้น

18. คำนวณหาค่า true positive, false negative, false positive rate, polyp detection rates และ polyp miss rates ของทั้ง 2 โปรแกรมปัญญาประดิษฐ์ โดยเปรียบเทียบในหลายๆเกณฑ์ชี้วัด ทั้งเกณฑ์ชี้วัด ≥ 0.5 วินาที, เกณฑ์ชี้วัด ≥ 1 วินาที, เกณฑ์ชี้วัด ≥ 1.5 วินาที และ เกณฑ์ชี้วัด ≥ 2 วินาที

รูปที่ 10 แสดงขั้นตอนในการดำเนินการวิจัย



3.8 การรวบรวมข้อมูล (Data collection)

- ข้อมูลพื้นฐาน : อายุ เพศ ประวัติการสูบบุหรี่ ประวัติการดื่มแอลกอฮอล์ ส่วนสูง น้ำหนัก BMI
- โรคประจำตัว และยาที่ใช้ประจำ
- ประวัติโรคมะเร็งลำไส้ใหญ่ในครอบครัว
- คุณภาพของการเตรียมลำไส้ใหญ่ก่อนเริ่มส่องกล้อง ระยะเวลาในการส่องกล้องลำไส้ใหญ่
- ลักษณะของติ่งเนื้อ ได้แก่ จำนวน ตำแหน่ง รูปร่าง ขนาด ชนิด
- ภาวะแทรกซ้อนหลังจากการส่องกล้องลำไส้ใหญ่
- true positive, false negative และ false positive

3.9 การวิเคราะห์ข้อมูลและสถิติที่ใช้วิเคราะห์ (Data Analysis and Statistics)

วิเคราะห์ข้อมูลโดยใช้โปรแกรม SPSS Statistics ในการวิเคราะห์ข้อมูล โดยสถิติที่ใช้มีดังนี้

- ข้อมูลประเภทกลุ่ม (categorical data) ได้แก่ เพศ โรคประจำตัว ประวัติโรคมะเร็งในครอบครัว ประวัติสูบบุหรี่ ประวัติดื่มสุรา จำนวน ตำแหน่ง ตำแหน่ง รูปร่าง ขนาด ชนิด คำนวณเป็นร้อยละ (percentage)
- ข้อมูลต่อเนื่อง (continuous data) ได้แก่ อายุของอาสาสมัคร คุณภาพของการเตรียมลำไส้ใหญ่ก่อนเริ่มส่องกล้อง ระยะเวลาในการส่องกล้องลำไส้ใหญ่ คำนวณเป็นค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน (mean and standard deviation)
- ใช้ค่า true positive, false negative คำนวณ sensitivity
- เปรียบเทียบค่า polyp detection rates ระหว่างระบบคอมพิวเตอร์ปัญญาประดิษฐ์ 2 ระบบ
- P values จะมีนัยสำคัญทางสถิติเมื่อค่า p value < 0.05

3.10 ข้อพิจารณาด้านจริยธรรม (Ethical considerations)

หลักความเคารพในบุคคล (Respect for person)

คือผู้ป่วยที่เข้าร่วมในการวิจัยจะได้รับข้อมูล เกี่ยวกับรายละเอียดของงานวิจัยอย่างครบถ้วน และตัดสินใจอย่างอิสระในการให้ความยินยอมจากผู้เข้าร่วมวิจัย รวมทั้งข้อมูลที่ได้จากการศึกษาและประวัติของผู้ป่วยจะถูกเก็บเป็นความลับ โดยคำนึงถึงสิทธิของผู้ป่วยเป็นสำคัญ และการบันทึกข้อมูลจะบันทึกโดยไม่มีข้อมูลที่จะระบุถึงตัวผู้ป่วย รวมไปถึงการนำเสนอผลการศึกษาก็จะเป็นการนำเสนอภาพรวมของการศึกษาทั้งหมด และไม่นำเสนอข้อมูลรายบุคคล อนึ่งตามพระราชบัญญัติสุขภาพแห่งชาติ พ.ศ.2550 ที่ผู้วิจัยต้องขอความยินยอมจากผู้ป่วย

หลักการให้ประโยชน์ไม่ก่อให้เกิดอันตราย (Beneficence/Non-maleficence)

ในการศึกษาวิจัยนี้ ทางทีมผู้วิจัยมีเกณฑ์การคัดเข้าและคัดออกที่ชัดเจน ผู้ที่มีความเสี่ยงสูงไม่เหมาะสมจะเข้ารับการส่องกล้องลำไส้ใหญ่จะถูกคัดออกจากงานวิจัยตั้งแต่วินิจฉัย การเข้าร่วมวิจัยมีข้อดีคือ ได้รับการตรวจคัดกรองลำไส้ใหญ่โดยใช้ปัญญาประดิษฐ์ที่มีข้อมูลชัดเจนแล้วว่าทำให้ตรวจพบติ่งเนื้อ adenoma มากขึ้น โดยภาวะแทรกซ้อนเทียบเท่ากับการส่องกล้องแบบปกติ เช่น ลำไส้ใหญ่ทะลุพบได้ร้อยละ 0.01 เลือดออกในทางเดินอาหารพบได้ร้อยละ 0.1 รวมถึงระยะเวลาที่ใช้ในการส่องกล้องไม่ต่างจากเดิม นอกจากนี้ทางทีมผู้วิจัยมีการเฝ้าวัดสัญญาณชีพผู้ป่วยตลอดเวลา (extensive monitoring) และจะหยุดการส่องกล้องและช่วย resuscitation ทันทีหากมีภาวะแทรกซ้อนเกิดขึ้น

หลักความยุติธรรม (Justice)

คือการศึกษานี้มีเกณฑ์การคัดเลือกผู้ป่วยเข้าและคัดออกชัดเจน

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

3.11 ข้อจำกัดในการวิจัย (Limitation)

การส่องกล้องลำไส้ใหญ่โดยใช้ปัญญาประดิษฐ์ทำให้เกิดภาวะแทรกซ้อนได้เท่ากับการส่องกล้องด้วยวิธีการปกติ แต่หากเกิดภาวะแทรกซ้อนขึ้น ทางคณะผู้วิจัยก็ได้เตรียมแนวทางการแก้ไขไว้แล้ว เช่น หากมีภาวะลำไส้ทะลุก็สามารถการเย็บปิดลำไส้ใหญ่โดยการส่องกล้องได้ทันที และให้นอนโรงพยาบาลเพื่อสังเกตอาการ หากมีภาวะออกซิเจนในเลือดต่ำกว่าปกติหรือภาวะความดันโลหิตต่ำในระหว่างการให้ยานอนหลับก็สามารถแก้ไขได้ทันที เนื่องจากทางทีมผู้วิจัยมีการเฝ้าวัดสัญญาณชีพผู้ป่วยตลอดเวลา (extensive monitoring) และมีอุปกรณ์ช่วยชีวิตขั้นพื้นฐานในห้องส่องกล้อง

3.12 ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย (Expected and application)

ได้ทราบถึงประสิทธิภาพของระบบคอมพิวเตอร์ปัญญาประดิษฐ์ 2 ระบบคือ CAD EYE และ Deep-GI ในการการช่วยตรวจจับตึงเนื้อลำไส้ใหญ่ในผู้ที่มีมาส่องกล้องคัดกรองมะเร็งลำไส้ใหญ่ และหาค่าเกณฑ์ชี้วัดที่เหมาะสมที่จะช่วยลดอัตราการเกิดผลบวกลวง

3.13 สถานที่ทำวิจัย (Venue of the Study)

หน่วยโรคระบบทางเดินอาหารและตับ ภาควิชาอายุรศาสตร์ คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย (ศูนย์ความเป็นเลิศทางการแพทย์ด้านการส่องกล้องระบบทางเดินอาหารแห่งโรงพยาบาลจุฬาลงกรณ์อาคารภูมิ สิริมังคลานุสรณ์ชั้น 10) เลขที่ 1873 ถนน พระรามที่ 4 แขวง ปทุมวัน เขตปทุมวัน กรุงเทพมหานคร 10330 โทร. 02-256-4356

3.14 การบริหารงานวิจัยและตารางการปฏิบัติงาน (Administration and time schedule)

กิจกรรม/เดือน	12/64	1/65	2/65	3/65	4/65	5/65	6/65
เขียนโครงร่างงานวิจัย/ค้นคว้ารวบรวมเอกสาร ที่เกี่ยวข้อง/ขอจริยธรรมการวิจัย							
คัดเลือกผู้ป่วยเข้าร่วมการศึกษาและรวบรวมข้อมูล							
การวิเคราะห์ข้อมูลและสรุปผล							
การเขียนต้นฉบับและส่งตีพิมพ์							

บทที่ 4

ผลการวิจัย

งานวิจัยนี้เป็นการศึกษาในผู้ป่วยที่มาส่งกล้องคัดกรองมะเร็งลำไส้ใหญ่ ณ หน่วยโรคระบบทางเดินอาหารและตับ ภาควิชาอายุรศาสตร์ คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย (ศูนย์ความเป็นเลิศทางการแพทย์ด้านการส่องกล้องระบบทางเดินอาหารแห่งโรงพยาบาลจุฬาลงกรณ์อาคารภูมิสิริมังคลานุสรณ์ชั้น 10) มีผู้ป่วยเข้าร่วมการศึกษาทั้งสิ้น 170 ราย โดยผู้ป่วยทั้งหมด 170 รายสามารถทำการศึกษาได้จนจบเสร็จสิ้นตลอดการวิจัย

ประสิทธิภาพในการตรวจหาติ่งเนื้อลำไส้ใหญ่ของโปรแกรม Deep-GI (Deep-GI model Performance)

การคิดค้นและพัฒนา Deep-GI ใช้กระบวนการเรียนรู้เชิงลึกของคอมพิวเตอร์ หรือที่เรียกว่า deep learning โดยใช้ Model ทางวิศวกรรมศาสตร์ ชื่อ YOLOv5 deep learning framework เพื่อเรียนรู้ภาพทั้งหมดที่ใส่เข้าไป ซึ่งเป็นการเรียนรู้แบบอัตโนมัติเลียนแบบการทำงานของโครงข่ายประสาทของมนุษย์ โดยนำระบบโครงข่ายประสาท (Neural Network) หลายชั้นมาซ้อนกัน และทำการเรียนรู้ข้อมูลตัวอย่าง ข้อมูลตัวอย่างจะถูกนำไปใช้ในการตรวจจับรูปแบบหรือจัดหมวดหมู่ข้อมูลเพื่อใช้ประมวลผลต่อไป โดยมีการใช้ข้อมูลตัวอย่างในการพัฒนา Deep-GI ถึง 15,188 ภาพ แบ่งเป็นการใส่ข้อมูลเพื่อให้ Deep-GI เรียนรู้ 12,148 ภาพ และใช้ในการทดสอบ Deep-GI อีก 3,040 ภาพ หลังจากที่ได้ทดสอบประสิทธิภาพของ Deep-GI จากชุดข้อมูลตัวอย่าง พบว่า Deep-GI มีความไว (sensitivity), ความจำเพาะ (specificity), ค่าทำนายผลบวก (positive predictive value, PPV), ค่าทำนายผลลบ (negative predictive value, NPV) และค่าความแม่นยำ (accuracy) เท่ากับ ร้อยละ 95, ร้อยละ 92, ร้อยละ 86, ร้อยละ 97, และร้อยละ 91 ตามลำดับ (ตารางที่ 2)

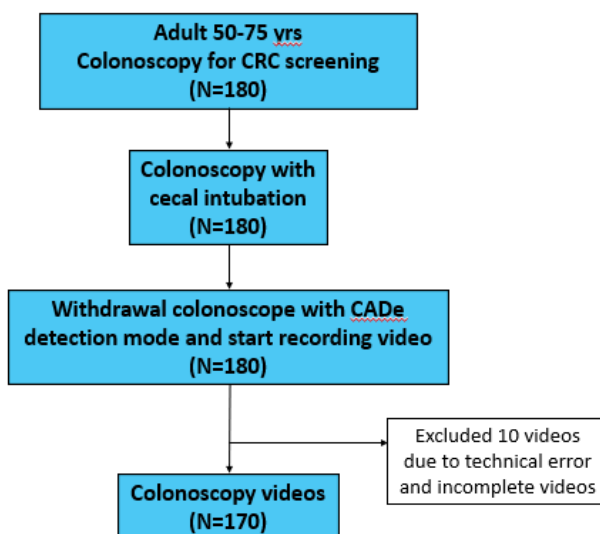
ตารางที่ 2 Deep-GI dataset และประสิทธิภาพในการตรวจหาตั้งเนื้อลำไส้ใหญ่ของโปรแกรม Deep-GI

Dataset	Total images	Polyp images	Non-polyp images
Training dataset	12,148	4,609	7,539
Validating dataset	1,520	577	943
Testing dataset	1,520	577	943
Total	15,188	5,763	9425
Performance status on still images			
Sensitivity			94.96 %
Specificity			91.73%
Positive predictive value (PPV)			86.42%
Negative predictive value (NPV)			97.05%
Accuracy			90.69%

มีผู้ป่วยเข้าร่วมการศึกษาทั้งสิ้น 180 ราย ผู้ป่วยทั้งหมด 180 รายนี้ เป็นผู้ป่วยที่มาส่งกล้องเพื่อคัดกรองมะเร็งลำไส้ใหญ่ โดยที่ไม่มีโรคหรืออาการผิดปกติอื่นๆ เช่น โรคทางพันธุกรรมที่ส่งผลให้มีโอกาสเกิดเนื้องอกหรือมะเร็งลำไส้ใหญ่เพิ่มมากขึ้น, โรคลำไส้อักเสบ (Inflammatory bowel disease), ผู้ที่เคยได้รับการผ่าตัดลำไส้ใหญ่, มีอาการถ่ายเป็นเลือด, ถ่ายท้องผูกสลับท้องเสีย, โลหิตจางโดยไม่ทราบสาเหตุ หรือ น้ำหนักลดผิดปกติ เป็นต้น

ผู้ป่วยทั้ง 180 รายได้รับการส่องกล้องลำไส้ใหญ่สมบูรณ์จากลำไส้ส่วน rectum ขึ้นไปถึง cecum และถอยกล้องตรวจดูลำไส้ใหญ่โดยเปิดระบบ CAD EYE detection พร้อมทั้งอัด video ขณะถอยกล้องทั้ง video ที่มีระบบ CAD EYE detection และ video ที่ไม่มีระบบ CAD EYE หลังจากที่ได้ตรวจสอบ video หลังส่องกล้องเสร็จ พบว่า มี video ที่มีปัญหา เช่น อัด video มาได้ไม่ครบทั้งการส่องกล้อง หรือ ไม่สามารถเปิด video ได้ เป็นต้น รวมทั้ง 10 videos ดังนั้นจึงได้ผู้ป่วยรวมทั้งหมดในการศึกษา 170 ราย ดังแสดงในรูปที่ 11 จากการศึกษาในผู้ป่วยทั้งหมด 170 ราย ผู้ป่วยมีอายุเฉลี่ย 62.7 ± 8.4 ปี มีผู้ป่วยเพศหญิง 102 รายคิดเป็นร้อยละ 60 เพศชาย 68 ราย คิดเป็นร้อยละ 40 ค่าเฉลี่ยคุณภาพการเตรียมลำไส้ตาม Boston Bowel Preparation Scale (BBPS) เท่ากับ 8.6 ± 0.63 และค่าเฉลี่ยระยะเวลาในการถอยกล้องเท่ากับ 7.8 ± 2.7 นาที

รูปที่ 11 ขั้นตอนในการคัดเลือกและดำเนินการวิจัย



จากผู้ป่วยที่มาส่งกล้องเพื่อคัดกรองมะเร็งลำไส้ใหญ่ทั้งหมด 170 ราย พบว่า มีผู้ป่วย 137 รายที่ส่งกล้องพบติ่งเนื้อลำไส้ใหญ่อย่างน้อย 1 ตี่งเนื้อ และจาก 137 รายที่พบติ่งเนื้อลำไส้ใหญ่นั้น พบว่าเป็นติ่งเนื้อชนิด adenoma ถึง 105 ราย คิดเป็น adenoma detection rate (ADR) เท่ากับ ร้อยละ 61.76 โดยจากผู้ป่วยทั้งหมด พบติ่งเนื้อลำไส้ใหญ่ทั้งหมด 501 ตี่งเนื้อ แบ่งเป็นติ่งเนื้อชนิด adenoma 262 ตี่งเนื้อ คิดเป็นร้อยละ 52.3 และติ่งเนื้อชนิด hyperplastic 239 ตี่งเนื้อ คิดเป็นร้อยละ 47.7 ค่าเฉลี่ยของจำนวนติ่งเนื้อลำไส้ใหญ่ที่พบในการส่องกล้อง 1 ครั้ง 2.95 ตี่งเนื้อ/การส่องกล้องลำไส้ใหญ่ โดยตลอดการศึกษานี้ ไม่พบว่ามีข้อมูลใดที่มึความเห็นไม่ตรงกันระหว่างอายุรแพทย์ทางเดินอาหารที่มีความเชี่ยวชาญด้านการส่องกล้องลำไส้ใหญ่ 2 คน

สำหรับติ่งเนื้อชนิด adenoma พบว่า ร้อยละ 67.9 เป็นติ่งเนื้อที่มีขนาดน้อยกว่า 5 มม. และร้อยละ 80 เป็นติ่งเนื้อชนิดไม่มีก้าน (sessile) สำหรับติ่งเนื้อชนิด hyperplastic พบว่า ร้อยละ 96.7 เป็นติ่งเนื้อที่มีขนาดน้อยกว่า 5 มม. และประมาณร้อยละ 50 เป็นติ่งเนื้อชนิดไม่มีก้าน (sessile) ในขณะที่อีกครึ่งหนึ่ง เป็นติ่งเนื้อแบน (flat) โดยที่ไม่พบติ่งเนื้อแบบมีก้าน (pedunculated) ในกลุ่มติ่งเนื้อชนิด hyperplastic เลย ข้อมูลพื้นฐานของผู้ป่วยที่มาส่งกล้องและลักษณะติ่งเนื้อลำไส้ใหญ่ที่พบทั้งหมด ดังแสดงในตารางที่ 3

ตารางที่ 3 แสดงข้อมูลพื้นฐานของผู้ป่วยที่มาส่งกล้องคัดกรองมะเร็งลำไส้ใหญ่และติ่งเนื้อลำไส้ใหญ่ที่ตรวจพบ (n=170)

Demographic data (n=170)		
Age	62.7 ± 8.4	
Sex (male)	68 (40%)	
BBPS	8.6 ± 0.63	
Withdrawal time (min)	7.8 ± 2.7	
Total polyps	501	
	Adenoma = 262 (52.3%)	Hyperplastic = 239 (47.7%)
Size		
<0.5 cm	178 (67.9%)	231 (96.7%)
0.5-1 cm	63 (24.1%)	8 (3.3%)
>1 cm	21 (8%)	0
Morphology		
Sessile	211 (80.6%)	124 (51.9%)
Pedunculate	25 (9.5%)	0
Flat	26 (9.9%)	115 (48.1%)

Abbreviations: BBPS= Boston Bowel Preparation Scale

CHULALONGKORN UNIVERSITY

ประสิทธิภาพในการตรวจพบติ่งเนื้อลำไส้ใหญ่ (Polyp detection performance)

การตรวจพบติ่งเนื้อทั้งหมด (Overall polyp detection)

เมื่อเปรียบเทียบประสิทธิภาพระหว่างระบบคอมพิวเตอร์ปัญญาประดิษฐ์ 2 ระบบคือ CAD EYE และ Deep-GI ในการการช่วยตรวจจับติ่งเนื้อลำไส้ใหญ่ในผู้ที่มาส่งกล้องคัดกรองมะเร็งลำไส้ใหญ่ พบว่า ระบบ Deep-GI ตรวจพบติ่งเนื้อลำไส้ใหญ่ทั้งหมด 498 ติ่งเนื้อ คิดเป็นร้อยละ 99.4 ในขณะที่ระบบ CAD EYE ตรวจพบติ่งเนื้อลำไส้ใหญ่ทั้งหมด 428 ติ่งเนื้อ คิดเป็นร้อยละ 85.4 มีติ่งเนื้อลำไส้ใหญ่ 3 ติ่งเนื้อที่ไม่ถูกตรวจพบโดยระบบ Deep-GI คิดเป็นร้อยละ 0.6 ในขณะที่มีติ่งเนื้อลำไส้ใหญ่ 73 ติ่งเนื้อที่ไม่ถูกตรวจพบโดยระบบ CAD EYE คิดเป็นร้อยละ 14.6 ดังแสดงในตารางที่ 4

เมื่อเปรียบเทียบประสิทธิภาพระหว่าง Deep-GI และ CAD EYE พบว่า ระบบ Deep-GI มีอัตราการตรวจพบติ่งเนื้อลำไส้ใหญ่ (polyp detection rate) ที่สูงกว่า (99.4% vs. 85.4%; $p < 0.01$) และมีอัตราการไม่พบติ่งเนื้อลำไส้ใหญ่ (polyp miss rate) ที่ต่ำกว่า (0.6% vs. 14.6%; $p < 0.01$) ระบบ CAD EYE

ตารางที่ 4 แสดงประสิทธิภาพของระบบ Deep-GI เทียบกับระบบ CAD EYE ในการตรวจพบติ่งเนื้อลำไส้ใหญ่ทั้งหมด (Diagnostic performance of Deep-GI compared to CAD EYE for overall polyp detection)

Overall polyp detection	Deep-GI			CAD EYE		
Total true positive	498			428		
Total false positive	59,350			106,042		
Total false negative	3			73		
False positive/colonoscopy	349 ± 169			624 ± 468		
Threshold	TP	FP	FN	TP	FP	FN
≥0.5 sec	498	2056	3	426	3807	75
≥1 sec	492	744	9	422	1149	79
≥1.5 sec	453	334	48	380	404	121
≥2 sec	449	178	52	376	182	125
Sensitivity	Deep-GI			CAD EYE		P-value
No threshold	99.4%			85.4%		<0.01
≥0.5 sec	99.4%			85%		<0.01
≥1 sec	98%			84.2%		<0.01
≥1.5 sec	90.4%			75.6%		<0.01
≥2 sec	89.6%			75%		<0.01

Abbreviations: TP=true positive, FP=false positive, FN= false negative

การตรวจพบติ่งเนื้อชนิด adenoma (Adenoma detection)

เมื่อเปรียบเทียบประสิทธิภาพระหว่างระบบคอมพิวเตอร์ปัญญาประดิษฐ์ 2 ระบบคือ CAD EYE และ Deep-GI ในการการช่วยตรวจจับติ่งเนื้อลำไส้ใหญ่ในผู้ที่มีมาส่องกล้องคัดกรองมะเร็งลำไส้

ใหญ่ พบว่า ระบบ Deep-GI ตรวจพบติ่งเนื้อชนิด adenoma 261 ติ่งเนื้อ คิดเป็นร้อยละ 99.6 ในขณะที่ระบบ CAD EYE ตรวจพบติ่งเนื้อชนิด adenoma 253 ติ่งเนื้อ คิดเป็นร้อยละ 96.6 มีติ่งเนื้อชนิด adenoma 1 ติ่งเนื้อที่ไม่ถูกตรวจพบโดยระบบ Deep-GI คิดเป็นร้อยละ 0.4 ในขณะที่มีติ่งเนื้อชนิด adenoma 9 ติ่งเนื้อที่ไม่ถูกตรวจพบโดยระบบ CAD EYE คิดเป็นร้อยละ 3.4 ดังแสดงในตารางที่ 5

เมื่อเปรียบเทียบประสิทธิภาพระหว่าง Deep-GI และ CAD EYE พบว่า ระบบ Deep-GI มีอัตราการตรวจพบติ่งเนื้อชนิด adenoma (adenoma detection rate) ที่สูงกว่า (99.6% vs. 96.6%; $p=0.039$), มีอัตราการไม่พบติ่งเนื้อชนิด adenoma (adenoma miss rate) ที่ต่ำกว่า (0.4% vs. 3.4%; $p=0.043$) และมีจำนวนการเกิดผลบวกกลางต่อการระบบส่องกล้องลำไส้ใหญ่ น้อยกว่า CAD EYE (349 ± 169 vs. 624 ± 468 ; $p < 0.01$) ดังแสดงในตารางที่ 7

เมื่อปรับค่าเกณฑ์ชี้วัดสำหรับผลบวกกลาง (false positive threshold, FP threshold) เป็นค่าต่างๆ ตั้งแต่ ≥ 0.5 วินาที, ≥ 1 วินาที, ≥ 1.5 วินาที, and ≥ 2 วินาที พบว่า ระบบ Deep-GI ยังคงมีอัตราการตรวจพบติ่งเนื้อชนิด adenoma (adenoma detection rate) ที่สูงกว่าระบบ CAD EYE อย่างมีนัยสำคัญทางสถิติ ที่เกณฑ์ชี้วัดสำหรับผลบวกกลาง ≥ 0.5 วินาที (99.6% vs. 96.2%; $p=0.027$) ในขณะที่ ที่เกณฑ์ชี้วัดสำหรับผลบวกกลางอื่นๆ ระบบ Deep-GI ก็มีอัตราการตรวจพบติ่งเนื้อชนิด adenoma (adenoma detection rate) ที่สูงกว่าระบบ CAD EYE เช่นกัน แต่ไม่มีนัยสำคัญทางสถิติ ดังแสดงในตารางที่ 5

ตารางที่ 5 แสดงประสิทธิภาพของระบบ Deep-GI เทียบกับระบบ CAD EYE ในการตรวจพบติ่งเนื้อชนิด adenoma (Diagnostic performance of Deep-GI compared to CAD EYE for adenoma detection)

Adenoma detection	Deep-GI			CAD EYE		
Total true positive	261			253		
Total false positive	59,350			106,042		
Total false negative	1			9		
False positive/colonoscopy	349 ± 169			624 ± 468		
Threshold	TP	FP	FN	TP	TP	FN
≥0.5 sec	261	2056	1	252	3807	10
≥1 sec	260	744	2	252	1149	10
≥1.5 sec	254	334	8	248	404	14
≥2 sec	254	178	8	247	182	15
Sensitivity	Deep-GI		CAD EYE		P-value	
No threshold	99.6%		96.6%		0.039	
≥0.5 sec	99.6%		96.2%		0.027	
≥1 sec	99.2%		96.2%		0.059	
≥1.5 sec	96.9%		94.7%		0.293	
≥2 sec	96.9%		94.2%		0.229	

Abbreviations: TP=true positive, FP=false positive, FN= false negative

ข้อมูลพื้นฐานของติ่งเนื้อลำไส้ใหญ่ที่ตรวจไม่พบโดยระบบปัญญาประดิษฐ์ (Characteristics of missed polyps by CAD EYE)

จากติ่งเนื้อลำไส้ใหญ่ทั้งหมดที่พบโดยอายุรแพทย์ทางเดินอาหารที่มีความเชี่ยวชาญด้านการส่องกล้องลำไส้ใหญ่จำนวน 501 ติ่งเนื้อ พบว่า มีติ่งเนื้อลำไส้ใหญ่ทั้งหมด 75 ติ่งเนื้อ ที่ตรวจไม่พบโดยระบบปัญญาประดิษฐ์ ทั้ง 2 ระบบ แบ่งเป็นติ่งเนื้อชนิด adenoma 9 ติ่งเนื้อ คิดเป็นร้อยละ 12 และติ่งเนื้อชนิด hyperplastic 66 ติ่งเนื้อ คิดเป็นร้อยละ 88

โดยระบบ Deep-GI ตรวจไม่พบติ่งเนื้อทั้งหมด 3 ตี่งเนื้อ แบ่งเป็นติ่งเนื้อชนิด adenoma 1 ตี่งเนื้อ และติ่งเนื้อชนิด hyperplastic 2 ตี่งเนื้อ ในขณะที่ระบบ CAD EYE ตรวจไม่พบติ่งเนื้อทั้งหมด 73 ตี่งเนื้อ แบ่งเป็นติ่งเนื้อชนิด adenoma 9 ตี่งเนื้อ และติ่งเนื้อชนิด hyperplastic 64 ตี่งเนื้อ เกือบทั้งหมดของติ่งเนื้อที่ตรวจไม่พบโดยระบบปัญญาประดิษฐ์ ทั้งติ่งเนื้อชนิด adenoma และติ่งเนื้อชนิด hyperplastic เป็นติ่งเนื้อที่มีขนาดน้อยกว่า 5 มม. มีเพียงติ่งเนื้อชนิด adenoma ที่ขนาดมากกว่า 10 มม. เพียงติ่งเนื้อเดียวที่ตรวจไม่พบโดยระบบ CAD EYE ข้อมูลพื้นฐานของติ่งเนื้อลำไส้ใหญ่ที่ตรวจไม่พบ ดังแสดงในตารางที่ 6

ตารางที่ 6 แสดงข้อมูลพื้นฐานของติ่งเนื้อลำไส้ใหญ่ที่ตรวจไม่พบ (Baseline characteristics of missed polyps)

Missed polyps (False negative)	75	
	Adenoma = 9 (12%)	Hyperplastic = 66 (88%)
Deep-GI	1 (11.1%)	2 (3%)
CAD EYE	9 (100%)	64 (97%)
Size		
< 0.5 cm	8 (88.9%)	64 (97%)
0.5-1 cm	0	2 (3%)
> 1 cm	1 (11.1%)	0
Morphology		
Sessile	8 (88.9%)	39 (59.1%)
Pedunculate	0	0
flat	1 (11.1%)	27 (40.9%)

ค่าเกณฑ์ชี้วัดสำหรับผลบวกหลง (False positive threshold, FP threshold)

เมื่อเปรียบเทียบประสิทธิภาพระหว่างระบบคอมพิวเตอร์ปัญญาประดิษฐ์ 2 ระบบ ในการช่วยตรวจจับติ่งเนื้อลำไส้ใหญ่ในเรื่องของจำนวนการเกิดผลบวกหลงต่อการระบบส่องกล้องลำไส้ใหญ่ พบว่า ระบบ Deep-GI เกิดผลบวกหลงน้อยกว่าระบบ CAD EYE เมื่อค่าเกณฑ์ชี้วัดสำหรับผลบวกหลงอยู่ที่ ≥ 0.5 วินาที และ ≥ 1 วินาที (12.1 ± 10.3 vs. 22.4 ± 23.5 ; $p < 0.01$ และ 4.4 ± 4.8 vs. 6.8 ± 7.6 ; $p < 0.01$ ตามลำดับ) ดังแสดงในตารางที่ 7

พบว่า อัตราการพบผลบวกปลอม (False positive alarm rate) น้อยกว่า 5 ครั้ง/การส่องกล้องลำไส้ใหญ่ ในระบบคอมพิวเตอร์ปัญญาประดิษฐ์ทั้ง 2 ระบบ เมื่อค่าเกณฑ์ชี้วัดสำหรับผลบวกปลอมอยู่ที่ ≥ 1 วินาที และพบว่า อัตราการพบผลบวกปลอมในระบบ CAD EYE จะสูงกว่าระบบ Deep-GI อย่างมีนัยสำคัญทางสถิติ เมื่อปรับค่าเกณฑ์ชี้วัดสำหรับผลบวกปลอมเป็น < 1.5 วินาที

ตารางที่ 7 แสดงประสิทธิภาพของระบบ Deep-GI เทียบกับระบบ CAD EYE ในการตรวจพบผลบวกปลอมเมื่อใช้ค่าเกณฑ์ชี้วัดสำหรับผลบวกปลอมที่ระยะเวลาต่างๆ (Diagnostic performance of Deep-GI system compared to CAD EYE system using different thresholds for false positive alerts)

Parameters	Deep-GI (N=170)	CAD EYE (N=170)	P-value			
Overall polyp detection rate (%)	498 (99.4%)	428 (85.43%)	<0.01			
Adenoma detection rate (%)	261 (99.62%)	253 (96.56%)	0.039			
Overall polyp miss rate (%)	3 (0.6%)	73 (14.6%)	<0.01			
Adenoma miss rate (%)	1 (0.4%)	9 (3.4%)	0.043			
Total number of false positive alarms	59,350	106,042	-			
False positives per colonoscopy (mean \pm SD)	349 \pm 169	624 \pm 468	<0.01			
Comparative analysis using different thresholds for false positive alerts						
Threshold	True positive (Polyp detection rate)			False positive alarm/colonoscopy (mean \pm SD)		
	Deep-GI	CAD EYE	P value	Deep-GI	CAD EYE	P value
For all polyps (N=501)						
≥ 0.5 sec	498 (99.4%)	426 (85%)	<0.01	12.1 \pm 10.3	22.4 \pm 23.5	<0.01
≥ 1 sec	492 (98%)	422 (84.2%)	<0.01	4.4 \pm 4.8	6.8 \pm 7.6	<0.01
≥ 1.5 sec	453 (90.4%)	380 (75.9%)	<0.01	2 \pm 2.9	2.4 \pm 3.8	0.276
≥ 2 sec	449 (89.6%)	376 (75.1%)	<0.01	1 \pm 1.9	1.1 \pm 2.2	0.654
For adenomas (N=262)						
≥ 0.5 sec	261 (99.6%)	252 (96.2%)	0.027	12.1 \pm 10.3	22.4 \pm 23.5	<0.01
≥ 1 sec	260 (99.2%)	252 (96.2%)	0.059	4.4 \pm 4.8	6.8 \pm 7.6	<0.01

≥ 1.5 sec	254 (96.9%)	248 (94.7%)	0.293	2 ± 2.9	2.4 ± 3.8	0.276
≥ 2 sec	254 (96.9%)	247 (94.2%)	0.228	1 ± 1.9	1.1 ± 2.2	0.654



บทที่ 5

อภิปรายผลและข้อเสนอแนะ

ในปัจจุบัน ได้มีการพัฒนาและนำระบบปัญญาประดิษฐ์มาใช้ในทางการแพทย์มากขึ้น ในระบบทางเดินอาหารเองก็ได้มีการคิดค้นและพัฒนาระบบปัญญาประดิษฐ์เพื่อช่วยในการส่องกล้องลำไส้ใหญ่ (computer-aided diagnosis, CADe) ทั้งในแง่ของการตรวจหาและวินิจฉัยชนิดของติ่งเนื้อลำไส้ใหญ่ มีการศึกษามากมายที่ได้นำ CADe มาใช้จริงในการส่องกล้องลำไส้ใหญ่ในผู้เข้ารับการส่องกล้องคัดกรองมะเร็งลำไส้ใหญ่ พบว่าอัตราการตรวจพบติ่งเนื้อชนิด adenoma จากการส่องกล้องโดยใช้ CADe สูงกว่าการส่องกล้องปกติอย่างมีนัยสำคัญทางสถิติ (22, 27, 28) สมาคมส่องกล้องทางเดินอาหารทั้งทางยุโรปและอเมริกา มีการพูดถึงและสนับสนุนการนำระบบปัญญาประดิษฐ์มาใช้ในการส่องกล้องลำไส้ใหญ่ในผู้เข้ารับการส่องกล้องคัดกรองมะเร็งลำไส้ใหญ่มากขึ้น (29, 30) แต่อย่างไรก็ดี ยังคงต้องมีการศึกษาวิจัยต่อเนื่องก่อนที่จะนำมาใช้เป็นแนวทางปฏิบัติระดับประเทศได้

จากการศึกษานี้พบว่า การส่องกล้องโดยใช้ CADe มีอัตราการตรวจพบติ่งเนื้อชนิด adenoma หรือ adenoma detection rate (ADR) สูงถึงประมาณร้อยละ 60 โดยระบบปัญญาประดิษฐ์ที่ดีควรมีค่าความไว (sensitivity) ที่สูงในการตรวจพบติ่งเนื้อลำไส้ใหญ่ในขณะที่มีอัตราการพบผลบวกปลอม (False positive alarm rate) ต่ำ (31, 32)

การศึกษานี้เป็นการศึกษาแรกที่เปรียบเทียบประสิทธิภาพระหว่างระบบคอมพิวเตอร์ปัญญาประดิษฐ์ 2 ระบบคือ CAD EYE และ Deep-GI ในการช่วยตรวจจับติ่งเนื้อลำไส้ใหญ่ในผู้ที่มาส่องกล้องคัดกรองมะเร็งลำไส้ใหญ่ และหาค่าเกณฑ์ชี้วัดที่เหมาะสมเพื่อลดอัตราการเกิดผลบวกปลอม โดยจากการศึกษานี้พบว่า Deep-GI มีอัตราการตรวจพบติ่งเนื้อลำไส้ใหญ่ (polyp detection rate, PDR) ที่สูงกว่าทั้งในเกณฑ์ชี้วัดสำหรับผลบวกปลอมที่ ≥ 0.5 วินาที และ ≥ 1 วินาที (99.4% vs. 85.4%,

$p < 0.01$ และ 97.8 % vs. 84.2%, $p < 0.01$ ตามลำดับ) เมื่อดูแค่เฉพาะอัตราการตรวจพบติ่งเนื้อชนิด adenoma (adenoma detection rate, ADR) พบว่า ระบบ Deep-GI ยังคงมีอัตราการตรวจพบติ่งเนื้อชนิด adenoma (adenoma detection rate) ที่สูงกว่าระบบ CAD EYE อย่างมีนัยสำคัญทางสถิติ ที่เกณฑ์ชี้วัดสำหรับผลบวก ≥ 0.5 วินาที (99.6% vs. 96.2%; $p = 0.027$) ผลของการศึกษานี้ทำให้สรุปได้ว่า ระบบ Deep-GI มีอัตราการตรวจพบติ่งเนื้อลำไส้ใหญ่ (polyp detection rate) ที่สูงกว่าระบบ CAD EYE ในติ่งเนื้อทุกชนิด ไม่ว่าจะเป็นติ่งเนื้อชนิด adenoma หรือติ่งเนื้อชนิด hyperplastic ส่วนในแง่ของความสามารถในการวินิจฉัยแยกชนิดของติ่งเนื้อลำไส้ใหญ่ของระบบปัญญาประดิษฐ์ทั้ง 2 ระบบนั้น การศึกษานี้ไม่สามารถบอกได้ เนื่องจากอยู่นอกเหนือจากคำถามงานวิจัยและไม่ได้ออกแบบมาเพื่อศึกษาในเรื่องนี้ แต่เป็นคำถามที่น่าสนใจและกำลังจะทำการศึกษาต่อไปในอนาคต

หลังจากที่แพทย์ได้มีการนำระบบปัญญาประดิษฐ์มาใช้ในทางคลินิกเพื่อช่วยในการหาติ่งเนื้อลำไส้ใหญ่ในการส่องกล้องคัดกรองมะเร็งลำไส้ใหญ่ มีประเด็นที่น่าสนใจคือการเกิดผลบวกปลอม โดยจากการศึกษาหลายๆ การศึกษาที่ผ่านมา พบว่า ผลบวกปลอมเป็นหนึ่งในตัวชี้วัดที่สำคัญที่มีการทำการศึกษากัน แต่ในการศึกษาวิจัยที่ผ่านมา มีการให้คำจำกัดความของผลบวกปลอมที่แตกต่างกัน ทั้งในแง่ของระยะเวลาของผลบวกปลอม ที่ปัญญาประดิษฐ์ขึ้นกรอบเตือนว่ามีติ่งเนื้อ ทั้งๆที่ในความเป็นจริงไม่มีติ่งเนื้อบริเวณนั้น และคำจำกัดความของตัวผลบวกปลอมเอง ในขณะที่บางการศึกษาไม่ได้มีกล่าวถึงคำจำกัดความของผลบวกปลอมเลย (33-37) ปัจจุบันยังไม่มีข้อสรุปเกี่ยวกับคำจำกัดความของผลบวกปลอมที่จะนำมาใช้ในการศึกษาเพื่อประเมินประสิทธิภาพของระบบคอมพิวเตอร์ปัญญาประดิษฐ์ในการการช่วยตรวจส่องกล้องลำไส้ใหญ่ การมีคำจำกัดความของผลบวกปลอมที่เป็นมาตรฐานและมีเกณฑ์ชี้วัดสำหรับผลบวกปลอมที่ชัดเจน จะช่วยให้สามารถศึกษาประสิทธิภาพของระบบคอมพิวเตอร์ปัญญาประดิษฐ์ได้ถูกต้อง แม่นยำและมีความจำเพาะมากขึ้น ในขณะที่เดียวกันผลบวกปลอมยังสามารถ

ใช้เป็นหนึ่งในตัวชี้วัดคุณภาพในการส่องกล้องลำไส้ใหญ่ได้ เพื่อนำมาใช้เปรียบเทียบประสิทธิภาพระหว่างระบบคอมพิวเตอร์ปัญญาประดิษฐ์แต่ละแบบได้ง่ายขึ้น โดยในการศึกษานี้ ได้นำทั้ง 2 videos (CAD EYE detection และ Deep-GI detection) ไปนับจำนวน false positive และ duration โดยใช้โปรแกรมที่เขียนขึ้นมาเพื่อนับกรอบปัญญาประดิษฐ์โดยเฉพาะ ที่ทางคณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยจัดทำขึ้น เพื่อลดโอกาสการเกิด bias ในในการศึกษานี้

จากการศึกษานี้พบว่า ค่าเกณฑ์ชี้วัดสำหรับผลบวกหลง (false positive threshold, FP threshold) ที่แตกต่างกัน ตั้งแต่ ≥ 0.5 วินาที, ≥ 1 วินาที, ≥ 1.5 วินาที, and ≥ 2 วินาที มีผลต่อทั้งจำนวนผลบวกหลงที่เกิดขึ้น รวมถึงความไวและความถูกต้องของประสิทธิภาพของระบบคอมพิวเตอร์ปัญญาประดิษฐ์ในการตรวจพบติ่งเนื้อลำไส้ใหญ่ เมื่อดูในระบบ Deep-GI พบว่า ถ้ากำหนดค่าเกณฑ์ชี้วัดสำหรับผลบวกหลงเป็น ≥ 0.5 วินาที จะพบอัตราการเกิดผลบวกหลง 12.1 ± 10.3 ครั้ง/การส่องกล้องลำไส้ใหญ่ ในขณะที่ ถ้าปรับค่าเกณฑ์ชี้วัดสำหรับผลบวกหลงเป็น ≥ 1 วินาที จะพบอัตราการเกิดผลบวกหลงลดลง เหลือเพียง 4.4 ± 4.8 ครั้ง/การส่องกล้องลำไส้ใหญ่ แต่อย่างไรก็ตาม การปรับค่าเกณฑ์ชี้วัดที่ช่วยลดผลบวกหลงทำให้ความไวและความถูกต้องของประสิทธิภาพของระบบคอมพิวเตอร์ปัญญาประดิษฐ์ในการตรวจพบติ่งเนื้อลำไส้ใหญ่ลดลงด้วย ดังนั้นสิ่งที่สำคัญในการออกแบบและพัฒนาาระบบคอมพิวเตอร์ปัญญาประดิษฐ์ในการช่วยตรวจส่องกล้องลำไส้ใหญ่ คือ การหาค่าเกณฑ์ชี้วัดที่เหมาะสมสำหรับผลบวกหลง

มีการศึกษาจาก Holzwaner และคณะ (26) กล่าวว่า เกณฑ์ชี้วัดที่เหมาะสมสำหรับผลบวกหลงคือ ≥ 2 วินาที เนื่องจากเวลา ≥ 2 วินาทีนั้น เป็นเวลาที่เพียงพอที่จะใช้แยกผลบวกหลงต่างๆ ไม่ว่าจะป็นฟองหรือเศษอุจจาระ และเป็นเวลาที่เพียงพอในการเป่าลมเพื่อให้รอยพับของลำไส้ใหญ่หายไป

จากการศึกษานี้พบว่า ค่าเกณฑ์ชี้วัดที่ ≥ 2 วินาทีทำให้ความไวและความถูกต้องของประสิทธิภาพของระบบคอมพิวเตอร์ปัญญาประดิษฐ์ในการตรวจพบติ่งเนื้อลำไส้ใหญ่ลดลงเนื่องจากมี

อัตราการไม่พบติ่งเนื้อลำไส้ใหญ่เพิ่มมากขึ้น. ในทางกลับกัน ค่าเกณฑ์ชี้วัดที่ ≥ 1 วินาที เป็นค่าเกณฑ์ชี้วัดที่เหมาะสมสำหรับผลบวกลวงสำหรับระบบ Deep-GI เนื่องจาก เป็นค่าที่ทำให้อัตราการไม่พบติ่งเนื้อลำไส้ใหญ่ (polyp miss rate) ต่ำที่สุด ในขณะที่ยังคงทำให้เกิดผลบวกลวงค่อนข้างน้อย จากผลการศึกษา สรุปลงได้ว่า ค่าเกณฑ์ชี้วัดที่เหมาะสมสำหรับผลบวกลวงสำหรับระบบคอมพิวเตอร์

ปัญญาประดิษฐ์แต่ละระบบอาจจะแตกต่างกัน

ข้อดีประการแรกของการศึกษานี้คือ เรามีวิดีโอการส่องกล้องตรวจลำไส้ใหญ่จำนวนมาก และพบติ่งเนื้อลำไส้ใหญ่ปริมาณมาก ทั้งติ่งเนื้อชนิด adenoma และติ่งเนื้อชนิด hyperplastic ซึ่งทำให้มีอำนาจเพียงพอที่จะสนับสนุนความไวและความถูกต้องของประสิทธิภาพของ Deep-GI ในการตรวจพบติ่งเนื้อลำไส้ใหญ่ ข้อดีประการที่สองคือ การศึกษานี้เป็นการศึกษาแรกที่เปรียบเทียบประสิทธิภาพระหว่างระบบคอมพิวเตอร์ปัญญาประดิษฐ์ 2 ระบบ ในแง่ของการความไวและความถูกต้องในการตรวจพบติ่งเนื้อลำไส้ใหญ่ และผลของการปรับค่าเกณฑ์ชี้วัดที่ระยะเวลาต่างๆ

นอกจากนี้ การศึกษานี้ได้ประเมินความไวและความถูกต้องของ Deep-GI ในการตรวจพบติ่งเนื้อลำไส้ใหญ่และการเกิดผลบวกลวง โดยวัดเป็นระยะเวลาที่พบที่ชัดเจน ทำให้สามารถคำนวณความถูกต้องได้อย่างแม่นยำมากขึ้น

อย่างไรก็ตาม เนื่องจากระบบ Deep-GI ได้ถูกนำมาใช้ในวิดีโอที่บันทึกและตัดต่อไว้ไว้ก่อนหน้าแล้ว ไม่ได้ถูกนำไปใช้ตั้งแต่ตอนที่ทำการส่องกล้องลำไส้ใหญ่ ทำให้ยังไม่สามารถประเมินประสิทธิภาพที่ชัดเจนในการช่วยเพิ่มอัตราการตรวจพบติ่งเนื้อชนิด adenoma (adenoma detection rate) ได้ ในอนาคตทางคณะผู้จัดทำวางแผนที่จะนำระบบ Deep-GI มาใช้ในคลินิกเพื่อศึกษาประสิทธิภาพในการเพิ่มอัตราการตรวจพบติ่งเนื้อชนิด adenoma (adenoma detection rate) และประสิทธิภาพในการวินิจฉัยแยกชนิดของติ่งเนื้อลำไส้ใหญ่ต่อไป

จากการศึกษานี้พบว่า Deep-GI สามารถช่วยตรวจจับตึงเนื้อลำไส้ใหญ่ในผู้ป่วยที่เข้ารับการส่องกล้องได้ดีกว่า ในขณะที่ทำให้เกิดผลบวกลวงน้อยกว่า CAD EYE เนื่องจาก Deep-GI ถูกพัฒนาขึ้นโดยใช้ภาพถึง 15,188 ภาพในการเรียนรู้และการทดสอบ โดยภาพที่ใช้มีทั้งภาพตึงเนื้อชนิดต่างๆ และภาพผลบวกลวงต่างๆ ทำให้ Deep-GI มีค่าความแม่นยำ (accuracy) สูงถึงร้อยละ 91

โดยสรุป การศึกษานี้พบว่า เมื่อเปรียบเทียบประสิทธิภาพระหว่างระบบคอมพิวเตอร์ปัญญาประดิษฐ์ 2 ระบบ ในการช่วยตรวจจับตึงเนื้อลำไส้ใหญ่ในผู้ป่วยที่เข้ารับการส่องกล้องคัดกรองมะเร็งลำไส้ใหญ่ ระบบ Deep-GI มีอัตราการตรวจพบตึงเนื้อลำไส้ใหญ่ (overall polyp detection rate) ที่สูงกว่า ในขณะที่ทำให้เกิดผลบวกลวงน้อยกว่า ที่เกณฑ์ชี้วัดสำหรับผลบวกลวง ≥ 0.5 วินาที และ ≥ 1 วินาที อย่างมีนัยสำคัญทางสถิติ ค่าเกณฑ์ชี้วัดสำหรับผลบวกลวงที่ ≥ 1 วินาที เป็นค่าเกณฑ์ชี้วัดที่เหมาะสมสำหรับระบบ Deep-GI และ CAD EYE เนื่องจากเป็นค่าที่ทำให้อัตราการไม่พบตึงเนื้อลำไส้ใหญ่ (polyp miss rate) และเกิดผลบวกลวงต่ำที่สุด ในอนาคต จำเป็นต้องมีการศึกษาเพิ่มเติมเพื่อศึกษาประโยชน์และประสิทธิภาพของระบบ Deep-GI ในการเพิ่มอัตราการตรวจพบตึงเนื้อชนิด adenoma (adenoma detection rate) ในคลินิก ทั้งนี้เพื่อประโยชน์ในการพัฒนาองค์ความรู้ด้านการแพทย์ในการดูแลผู้ป่วยต่อไป

บรรณานุกรม

1. Virani S, Bilheem S, Chansaard W, Chitapanarux I, Daoprasert K, Khuanchana S, et al. National and subnational population-based incidence of cancer in Thailand: assessing cancers with the highest burdens. *Cancers*. 2017;9(8):108.
2. Khuaprema T, Srivatanakul P. Colon and rectum cancer in Thailand: an overview. *Japanese journal of clinical oncology*. 2008;38(4):237-43.
3. Rex DK, Boland CR, Dominitz JA, Giardiello FM, Johnson DA, Kaltenbach T, et al. Colorectal cancer screening: recommendations for physicians and patients from the US Multi-Society Task Force on Colorectal Cancer. *Gastroenterology*. 2017;153(1):307-23.
4. Rutter MD, Beintaris I, Valori R, Chiu HM, Corley DA, Cuatrecasas M, et al. World Endoscopy Organization consensus statements on post-colonoscopy and post-imaging colorectal cancer. *Gastroenterology*. 2018;155(3):909-25. e3.
5. Corley DA, Jensen CD, Marks AR, Zhao WK, Lee JK, Doubeni CA, et al. Adenoma detection rate and risk of colorectal cancer and death. *New england journal of medicine*. 2014;370(14):1298-306.
6. Aziz M, Fatima R, Dong C, Lee-Smith W, Nawras A. The impact of deep convolutional neural network-based artificial intelligence on colonoscopy outcomes: a systematic review with meta-analysis. *Journal of Gastroenterology and Hepatology*. 2020;35(10):1676-83.
7. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer*. 2015;136(5):E359-E86.
8. Khuaprema T, Attasara P, Sriplung H, Wiangnon S, Sangrajrang S. *Cancer in Thailand vol VII, 2007-2009*. Bangkok: National Cancer Institute. 2013.
9. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut*. 2017;66(4):683-91.
10. Stewart SL, Wike JM, Kato I, Lewis DR, Michaud F. A population-based study of colorectal cancer histology in the United States, 1998–2001. *Cancer*. 2006;107(S5):1128-

41.

11. Giardiello FM, Allen JI, Axilbund JE, Boland CR, Burke CA, Burt RW, et al. Guidelines on genetic evaluation and management of Lynch syndrome: a consensus statement by the US Multi-society Task Force on colorectal cancer. *Gastroenterology*. 2014;147(2):502-26.

12. Ponugoti PL, Cummings OW, Rex DK. Risk of cancer in small and diminutive colorectal polyps. *Digestive and Liver Disease*. 2017;49(1):34-7.

13. Rees CJ, Gibson ST, Rutter MD, Baragwanath P, Pullan R, Feeney M, et al. UK key performance indicators and quality assurance standards for colonoscopy. *Gut*. 2016;65(12):1923-9.

14. Lai EJ, Calderwood AH, Doros G, Fix OK, Jacobson BC. The Boston bowel preparation scale: a valid and reliable instrument for colonoscopy-oriented research. *Gastrointestinal endoscopy*. 2009;69(3):620-5.

15. Aniwat S, Orkoonawat P, Viriyautsahakul V, Angsuwatcharakon P, Pittayanon R, Wisedopas N, et al. The secondary quality indicator to improve prediction of adenoma miss rate apart from adenoma detection rate. *Official journal of the American College of Gastroenterology | ACG*. 2016;111(5):723-9.

16. Le Berre C, Sandborn WJ, Aridhi S, Devignes M-D, Fournier L, Smail-Tabbone M, et al. Application of artificial intelligence to gastroenterology and hepatology. *Gastroenterology*. 2020;158(1):76-94. e2.

17. Bernal J, Sánchez FJ, Fernández-Esparrach G, Gil D, Rodríguez C, Vilariño F. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized Medical Imaging and Graphics*. 2015;43:99-111.

18. Silva J, Histace A, Romain O, Dray X, Granado B. Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. *International journal of computer assisted radiology and surgery*. 2014;9(2):283-93.

19. He K, Gkioxari G, Dollár P, Girshick R, editors. Mask r-cnn. *Proceedings of the IEEE international conference on computer vision*; 2017.

20. Redmon J, Farhadi A, editors. YOLO9000: better, faster, stronger. *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2017.

21. Kudo Se, Mori Y, Misawa M, Takeda K, Kudo T, Itoh H, et al. Artificial intelligence

and colonoscopy: Current status and future perspectives. *Digestive Endoscopy*. 2019;31(4):363-71.

22. Wang P, Berzin TM, Brown JRG, Bharadwaj S, Becq A, Xiao X, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut*. 2019;68(10):1813-9.

23. Liu WN, Zhang YY, Bian XQ, Wang LJ, Yang Q, Zhang XD, et al. Study on detection rate of polyps and adenomas in artificial-intelligence-aided colonoscopy. *Saudi journal of gastroenterology : official journal of the Saudi Gastroenterology Association*. 2020;26(1):13-9.

24. Su JR, Li Z, Shao XJ, Ji CR, Ji R, Zhou RC, et al. Impact of a real-time automatic quality control system on colorectal polyp and adenoma detection: a prospective randomized controlled study (with videos). *Gastrointest Endosc*. 2020;91(2):415-24.e4.

25. Fujifilm. Fujifilm acquires CE mark and launches CAD EYE afocpduAt, in Europe. Available: www.fujifilm.eu/eu/news/article/fujifilm-acquires-ce-mark-and-launches-cad-eyea-function-of-colonic-polyp-detection-utilizing-ai-t. [

26. Holzwanger EA, Bilal M, Brown JRG, Singh S, Becq A, Ernest-Suarez K, et al. Benchmarking definitions of false-positive alerts during computer-aided polyp detection in colonoscopy. *Endoscopy*. 2021;53(09):937-40.

27. Urban G, Tripathi P, Alkayali T, Mittal M, Jalali F, Karnes W, et al. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. *Gastroenterology*. 2018;155(4):1069-78. e8.

28. Repici A, Badalamenti M, Maselli R, Correale L, Radaelli F, Rondonotti E, et al. Efficacy of real-time computer-aided detection of colorectal neoplasia in a randomized trial. *Gastroenterology*. 2020;159(2):512-20. e7.

29. Bisschops R, East JE, Hassan C, Hazewinkel Y, Kamiński MF, Neumann H, et al. Advanced imaging for detection and differentiation of colorectal neoplasia: European Society of Gastrointestinal Endoscopy (ESGE) Guideline–Update 2019. *Endoscopy*. 2019;51(12):1155-79.

30. Berzin TM, Parasa S, Wallace MB, Gross SA, Repici A, Sharma P. Position statement on priorities for artificial intelligence in GI endoscopy: a report by the ASGE

Task Force. *Gastrointestinal endoscopy*. 2020;92(4):951-9.

31. Alagappan M, Brown JRG, Mori Y, Berzin TM. Artificial intelligence in gastrointestinal endoscopy: The future is almost here. *World journal of gastrointestinal endoscopy*. 2018;10(10):239.
32. Vinsard DG, Mori Y, Misawa M, Kudo S-e, Rastogi A, Bagci U, et al. Quality assurance of computer-aided detection and diagnosis in colonoscopy. *Gastrointestinal endoscopy*. 2019;90(1):55-63.
33. Wang Z, Liang Z, Li L, Li X, Li B, Anderson J, et al. Reduction of false positives by internal features for polyp detection in CT-based virtual colonoscopy. *Medical Physics*. 2005;32(12):3602-16.
34. Misawa M, Kudo S-e, Mori Y, Cho T, Kataoka S, Yamauchi A, et al. Artificial intelligence-assisted polyp detection for colonoscopy: initial experience. *Gastroenterology*. 2018;154(8):2027-9. e3.
35. Wang P, Xiao X, Glissen Brown JR, Berzin TM, Tu M, Xiong F, et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. *Nature biomedical engineering*. 2018;2(10):741-8.
36. Wang P, Liu X, Berzin TM, Brown JRG, Liu P, Zhou C, et al. Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADE-DB trial): a double-blind randomised study. *The Lancet Gastroenterology & Hepatology*. 2020;5(4):343-51.
37. Weigt J, Repici A, Antonelli G, Afifi A, Kliegis L, Correale L, et al. Performance of a new integrated computer-assisted system (CADE/CADx) for detection and characterization of colorectal neoplasia. *Endoscopy*. 2022;54(02):180-4.

ประวัติผู้เขียน

ชื่อ-สกุล	จุฬาลักษณ์ คฤหาสน์สุวรรณ
วัน เดือน ปี เกิด	14 ตุลาคม 2531
สถานที่เกิด	กรุงเทพมหานคร
วุฒิการศึกษา	จุฬาลงกรณ์มหาวิทยาลัย
ที่อยู่ปัจจุบัน	29 ซอยรามอินทรา 31 แยก 2 ถนนรามอินทรา แขวงอนุสาวรีย์ เขต บางเขน กทม 10220



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY