การประเมินความสามารถในการจำแนกและความสอดคล้องของ ChatGPT ร่วมกับคำสั่งเสียงในการคัดกรองผู้ป่วยฉุกเฉินตามเกณฑ์ ESI ในโรงพยาบาลวารินชำราบ

พัทนียา บุตรอ่อน*
ห้องฉุกเฉิน โรงพยาบาลวารินชำราบ

*ผู้ประพันธ์บรรณกิจ
พัทนียา บุตรอ่อน
ห้องฉุกเฉิน โรงพยาบาลวารินชำราบ ถนนวาริน–ศรีสะเกษ ตำบลคำน้ำแซบ อำเภอวารินชำราบ
จังหวัดอุบลราชธานี 34190
อีเมล: leklunla@gmail.com
โทรศัพท์ที่ทำงาน: 045 424250
โทรศัพท์มือถือ: -

DOI: 10.14456/tjem.2025.3

วันที่รับบทความ: 2025-08-15
วันที่แก้ไขบทความ: 2025-10-02
วันที่ตอบรับบทความ: 2025-11-04

บทคัดย่อ

บทนำ

การคัดกรองผู้ป่วยฉุกเฉิน (triage) เป็นขั้นตอนสำคัญต่อความปลอดภัยของผู้ป่วยและการจัดสรรทรัพยากร นอกจากนี้ความแม่นยำในการคัดกรองยังสะท้อนถึงประสิทธิภาพของระบบบริการสาธารณสุข

วัตถุประสงค์

เพื่อประเมินความสามารถในการจำแนกและความสอดคล้องของ ChatGPT ที่รองรับคำสั่งเสียงภาษาไทย ในการจำแนกระดับ triage ตามหลักเกณฑ์ Emergency Severity Index (ESI) เมื่อเทียบกับแพทย์ฉุกเฉินและพยาบาลแรกรับ

วิธีการศึกษา

การศึกษาเชิงพรรณนาแบบภาคตัดขวาง ณ แผนกฉุกเฉิน โรงพยาบาลวารินชำราบ ระหว่าง 1 เมษายน–31 พฤษภาคม 2568 ผู้ป่วย 387 ราย ได้รับการให้ระดับ triage โดยแพทย์ฉุกเฉิน พยาบาลแรกรับ และ ChatGPT-4o ซึ่งรับข้อมูลจากแบบฟอร์มคัดกรองผ่านเสียงภาษาไทย วิเคราะห์ประสิทธิภาพการจำแนกด้วย Sensitivity, Specificity, PPV, NPV, Area under the Receiver Operating Characteristic curve (AuROC) ความสอดคล้องด้วย weighted kappa และ ทดสอบความคลาดเคลื่อนด้วย McNemar test

ผลการศึกษา

สำหรับการจำแนกกลุ่มผู้ป่วยฉุกเฉิน (ESI 1–2) ChatGPT มีค่าความไว 97.6%, ความจำเพาะ 95.9%, ค่าพยากรณ์ผลบวก (PPV) 94.8%, ค่าพยากรณ์ผลลบ (NPV) 98.1% และ AuROC 97% (95%CI 95.0–99.0) ขณะที่พยาบาลมีความไว 92.3%, ความจำเพาะ 75.7%, PPV 74.6%, NPV 92.7% และ AuROC 84% (95%CI 81.0–87.0) ChatGPT มีค่าความสอดคล้องกับแพทย์อยู่ที่ weighted kappa = 0.915 (95%CI 0.844–0.986) ในระดับ almost perfect agreement ขณะที่พยาบาลมีค่าอยู่ที่ 0.607 (95%CI 0.536–0.678) ซึ่งจัดอยู่ในระดับ substantial agreement ความคลาดเคลื่อนของ ChatGPT พบ over-triage ร้อยละ 2.6 (10 ราย) และ under-triage ร้อยละ 3.4 (13 ราย) รวม ร้อยละ 5.9 (23 ราย) McNemar p=0.270 ขณะที่พยาบาลมี over-triage ร้อยละ 23.3 (90 ราย) และ under-triage ร้อยละ 5.2 (20 ราย) รวม ร้อยละ 28.4 (110 ราย) McNemar = p<0.001

สรุปผลการศึกษา

ChatGPT ร่วมคำสั่งเสียงภาษาไทย แสดงความสามารถในการจำแนกร้อยละ 97 ความสอดคล้องกับแพทย์ในระดับสูงมาก ดังนั้นศักยภาพของ ChatGPT จึงควรถูกนำไปใช้เป็นเครื่องมือสนับสนุนการตัดสินใจ (decision-support) ลดความผิดพลาดในการคัดกรอง และเป็นประโยชน์ในสถานการณ์ที่มีข้อจำกัดด้านบุคลากร

คำสำคัญ

การคัดกรองผู้ป่วยฉุกเฉิน, การประเมินความสามารถในการจำแนก, โมเดลภาษาปัญญาประดิษฐ์

Assessing the accuracy and agreement of ChatGPT Integrated with Voice Commands in Emergency Severity Index (ESI) Triage for Emergency Patients at Warin Chamrap Hospital

Pathaneeya Boodon*
Emergency Department, Warinchamrap Hospital

*coresponding author
Pathaneeya Boodon
Emergency Department, Warinchamrap Hospital,
Warin–Sisaket Road, Kham Nam Saep Subdistrict,
Warin Chamrap District, Ubon Ratchathani 34190, Thailand
Email: Chanikan.cwl@gmail.com
Tel 045 424250
Mobile: -

DOI: 10.14456/tjem.2025.3

Receive date: 2025-08-15
Revised date: 2025-10-02
Accepted date: 2025-11-04

Abstract

Background

Emergency department (ED) triage is critical for patient safety and optimal resource use. This study assessed the accuracy and agreement of Thai voice–enabled ChatGPT in Emergency Severity Index (ESI) triage compared with emergency physicians and triage nurses.

Methods

Cross-sectional study conducted in the ED of Warinchamrap Hospital, Thailand, from 1 April to 31 May 2025. Consecutive ED patients (n = 387) underwent triage by an expert emergency physician panel (reference standard), by triage nurses, and by ChatGPT-4o using data from the standardized triage form entered via Thai voice commands and/or typing. Agreement across the five ESI levels was assessed with weighted κ (kappa). Binary classification performance for critical (ESI 1–2) versus non-critical (ESI 3–5) was reported as sensitivity, specificity, PPV, NPV, and Area under the Receiver Operating Characteristic curve (AuROC). Directional misclassification (over-/under-triage) was tested with the McNemar test.

Results

For classifying emergent cases (ESI 1–2), ChatGPT showed sensitivity 97.6%, specificity 95.9%, PPV 94.8%, NPV 98.1%, and AuROC 97% (95%CI 95.0–99.0), while the nurse showed sensitivity 92.3%, specificity 75.7%, PPV 74.6%, NPV 92.7%, and AuROC 84% (95%CI 81.0–87.0). Agreement with physicians was weighted kappa =0.915 (95%CI 0.844–0.986) (almost perfect agreement) for ChatGPT and 0.607 (95%CI 0.536–0.678) (substantial agreement) for the nurse. ChatGPT showed over-triage 2.6% (10 cases) and under-triage 3.4% (13 cases), total 5.9% (23 cases), McNemar p = 0.270 the nurse showed over-triage 23.3% (90 cases) and under-triage 5.2% (20 cases), total 28.4% (110 cases), McNemar $p < 0.001$.

Conclusions

ChatGPT with Thai voice commands demonstrated 97% discriminative ability and very high agreement with physicians. Its potential should be applied as a decision-support tool to reduce triage errors and be useful where staffing is limited.

Keywords

Triage for Emergency, Assessment of discriminative ability, ChatGPT

บทนำ

ปัจจุบันระบบบริการทางการแพทย์ฉุกเฉินมีบทบาทสำคัญต่อการดูแลรักษาผู้ป่วยที่ต้องการความช่วยเหลือเร่งด่วน ระบบการคัดกรองผู้ป่วย (Triage) เป็นกระบวนการสำคัญที่ใช้ในการประเมินลำดับความเร่งด่วนของผู้ป่วย เพื่อให้สามารถจัดสรรทรัพยากรทางการแพทย์ได้อย่างมีประสิทธิภาพและช่วยลดอัตราการเสียชีวิตของผู้ป่วยในภาวะฉุกเฉิน ระบบการคัดกรองที่มีประสิทธิภาพต้องสามารถจำแนกผู้ป่วยตามระดับความรุนแรงของอาการได้อย่างถูกต้องและรวดเร็ว เพื่อให้ได้รับการรักษาอย่างเหมาะสมและทันท่วงที[1,2]

หนึ่งในระบบคัดกรองที่ใช้กันอย่างแพร่หลายในแผนกฉุกเฉินของโรงพยาบาล คือ ดัชนีความรุนแรงของเหตุฉุกเฉิน (Emergency Severity Index - ESI)[3] ซึ่งแบ่งระดับความรุนแรงผู้ป่วยออกเป็น 5 ระดับ[4,5] ระบบ ESI ช่วยให้บุคลากรทางการแพทย์สามารถบริหารจัดการผู้ป่วยได้อย่างมีประสิทธิภาพ ลดภาระงาน และเพิ่มความแม่นยำในการจัดลำดับความเร่งด่วนของผู้ป่วย อย่างไรก็ตาม ในทางปฏิบัติยังมีข้อจำกัดบางประการ เช่น ภาระงานที่มากเกินไปของบุคลากรทางการแพทย์ที่ทำหน้าที่คัดกรอง การตัดสินใจที่ขึ้นอยู่กับประสบการณ์ของพยาบาลแต่ละคน และความล่าช้าในการคัดแยกผู้ป่วยในช่วงเวลาที่มีผู้ป่วยจำนวนมาก ปัจจัยเหล่านี้ส่งผลให้การบริหารจัดการในห้องฉุกเฉินมีความซับซ้อนและต้องการแนวทางที่สามารถช่วยสนับสนุนการทำงานของบุคลากรได้ดียิ่งขึ้น[6]

เทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence: AI) ได้รับความสนใจเพิ่มขึ้นในวงการแพทย์และสาธารณสุข โดยเฉพาะการนำ AI มาใช้ในกระบวนการคัดกรองผู้ป่วยแผนกฉุกเฉิน เพื่อลดภาระงานของบุคลากรและเพิ่มความแม่นยำในการคัดแยกผู้ป่วย ChatGPT ซึ่งเป็นโมเดลปัญญาประดิษฐ์ที่พัฒนาโดย OpenAI มีศักยภาพในการช่วยวิเคราะห์ข้อมูลผู้ป่วยผ่านการสนทนาแบบอัตโนมัติ และสามารถใช้เป็นเครื่องมือช่วยเหลือในการคัดกรองผู้ป่วยโดยอ้างอิงจากระบบ ESI การนำ ChatGPT มาใช้ในกระบวนการคัดกรองมีข้อดีหลายประการ ได้แก่ 1) การช่วยซักถามและรวบรวมข้อมูลอาการเบื้องต้นจากผู้ป่วย 2) การวิเคราะห์ข้อมูลอาการเพื่อช่วยแนะนำระดับความรุนแรงของผู้ป่วยตามดัชนี ESI 3) การลดภาระงานของพยาบาลที่ทำหน้าที่คัดกรอง และ 4) การช่วยฝึกอบรมบุคลากรทางการแพทย์ในการคัดแยกผู้ป่วยฉุกเฉิน อย่างไรก็ตาม การใช้ AI ในการคัดกรองยังมีข้อจำกัด เช่น ความแม่นยำของข้อมูลที่ได้รับจากผู้ป่วย การทำงานร่วมกับระบบเวชระเบียนอิเล็กทรอนิกส์ และข้อกังวลด้านจริยธรรมและความปลอดภัยของข้อมูล[(7)] ปัจจุบันมีการศึกษาความแม่นยำของ ChatGPT ในการคัดแยกผู้ป่วยตามดัชนีความรุนแรงฉุกเฉิน (ESI) โดยทีมแพทย์ผู้เชี่ยวชาญ 5 คนได้ประเมินระดับความเร่งด่วนของผู้ป่วยตามเกณฑ์ ESI และเปรียบเทียบกับผลของ ChatGPT โดยใช้ค่ามัธยฐานของทีมแพทย์เป็นมาตรฐาน พบว่ามีความสอดคล้องในระดับปานกลาง (Cohen's Kappa = 0.659, $p < 0.001$) และมีความแม่นยำ 76.6% สำหรับผู้ป่วยที่อยู่ในกลุ่มฉุกเฉินรุนแรง (ESI-1 และ 2) ChatGPT มีความแม่นยำสูงมาก (Cohen's Kappa = 0.828) โดยมีความจำเพาะ 95.63%, ค่าพยากรณ์เชิงลบ 98.17% และความแม่นยำรวม 94.90% ผลการศึกษานี้ชี้ให้เห็นว่า ChatGPT

สามารถช่วยคัดแยกผู้ป่วยฉุกเฉินรุนแรงได้อย่างมีประสิทธิภาพ ซึ่งเป็นแนวทางที่น่าสนใจสำหรับการนำ AI มาใช้ในกระบวนการคัดแยกผู้ป่วยใน ED[8]

อย่างไรก็ตามการศึกษาครั้งนี้มีความสำคัญต่อระดับประสิทธิภาพของระบบคัดกรองผู้ป่วยฉุกเฉิน ทั้งในด้านการเพิ่มความถูกต้องและความรวดเร็วของการจำแนกผู้ป่วย การลดภาระงานของบุคลากรทางการแพทย์ในภาวะห้องฉุกเฉินแออัด อีกทั้งยังเป็นการประเมินความเป็นไปได้ในการบูรณาการเทคโนโลยีปัญญาประดิษฐ์ (AI) โดยเฉพาะ ChatGPT เข้าสู่ระบบสาธารณสุขของประเทศไทย ผลการศึกษาไม่เพียงช่วยให้เข้าใจถึงศักยภาพของ ChatGPT ในบริบทของโรงพยาบาลระดับภูมิภาคเท่านั้น แต่ยังสามารถพัฒนาและปรับปรุงระบบคัดกรองผู้ป่วยในอนาคต อันอาจขยายไปสู่การประยุกต์ใช้ AI ในกระบวนการดูแลสุขภาพด้านอื่น เพื่อยกระดับคุณภาพบริการทางการแพทย์และเสริมสร้างความปลอดภัยของผู้ป่วยในระยะยาว

วัตถุประสงค์

1. เพื่อประเมินความสามารถในการจำแนกของการใช้โมเดล ChatGPT-4o ที่รองรับการป้อนข้อมูลภาษาไทย ในการคัดกรองผู้ป่วยฉุกเฉินในบริบทห้องฉุกเฉิน

2. เพื่อประเมินความสอดคล้องของ ChatGPT-4o ร่วมกับคำสั่งเสียงในการคัดกรองผู้ป่วยฉุกเฉินตามเกณฑ์ ESI ในโรงพยาบาลวารินชำราบ

3. เพื่อเปรียบเทียบประสิทธิภาพการตัดสินของ ChatGPT กับพยาบาลแรกรับและแพทย์ฉุกเฉิน

รูปแบบการศึกษา การศึกษาครั้งนี้เป็นการศึกษาแบบตัดขวาง (Cross-sectional Study)

ประชากรที่ศึกษา เก็บข้อมูลแบบ consecutive case ของผู้ป่วยที่มารับบริการ ณ แผนกฉุกเฉิน ช่วงวันที่ 1 เม.ย. – 31 พ.ค. 2568 โดย ไม่เลือกคัดตามระดับ ESI (no enrichment by ESI level) และใช้เฉพาะการคัดกรองครั้งแรกต่อหนึ่งผู้ป่วย

ขนาดตัวอย่าง ในการศึกษาครั้งนี้ ผู้วิจัยได้คำนวณขนาดตัวอย่าง ซึ่งจากการทบทวนวรรณกรรม ของ Colakca และคณะ[8] ที่ได้ศึกษาเรื่อง Emergency department triaging using ChatGPT based on emergency severity index principles: a cross-sectional study พบ accuracy was 94.90% คำนวณขนาดตัวอย่างโดยใช้สูตร ประมาณค่าสัดส่วน Alpha (α) = 0.05, Standard normal value(Z) = 1.96, Prevalence(P) = 0.94, Absolute Presicion (d) = 0.025, 95% confidence interval Sample size (n) = 347 ราย ดังนั้นจะใช้กลุ่มตัวอย่างทั้งสิ้น 387 ราย

Inclusion Criteria

- ผู้ป่วยที่ได้รับการคัดแยกที่แผนกฉุกเฉินเป็นครั้งแรก

Exclusion Criteria

- เอกสารไม่ครบถ้วนสมบูรณ์ ได้แก่ ESI triage level จากแพทย์ฉุกเฉิน (reference standard) ESI triage level จากพยาบาลคัดกรอง ESI triage level จาก ChatGPT-4o

เครื่องมือที่ใช้ในการวิจัย

ในการศึกษาครั้งนี้ ผู้วิจัยได้ทำการเก็บข้อมูลด้วยแบบบันทึกข้อมูล ได้แก่ เพศ อายุ (ปี) โรคประจำตัว สิทธิ์การรักษา เวร วันที่มา

โรงพยาบาล วิธีการมาโรงพยาบาล เวลาที่มาถึงโรงพยาบาล สัญญาณชีพ Trauma 25 กลุ่มอาการ Criteria Based Dispatch (CBD) Critical outcome Triage level ได้แก่ พยาบาลคัดกรอง แพทย์ฉุกเฉิน และ ChatGPT

การกำหนดระดับ Emergency Severity Index (ESI) gold standard กระทำโดยผู้วิจัยหลักที่เป็นแพทย์ฉุกเฉิน จำนวน 1 คน ซึ่งมีประสบการณ์ทำงานในห้องฉุกเฉินเฉลี่ย 7 ปี พยาบาลคัดกรอง ผู้กรอกข้อมูล ChatGPT ถูกปกปิดระหว่างกันทั้งหมด โดยผู้วิจัยทำหน้าที่กรอกข้อมูลจากแบบฟอร์มคัดกรองเข้าสู่ระบบ ChatGPT โดยไม่มีการเปิดเผยผลการประเมินของแพทย์ต่อพยาบาลหรือระบบ AI และในทำนองเดียวกัน พยาบาลและระบบ AI ก็ไม่ทราบผลการประเมินของแพทย์เช่นกัน ทั้งนี้เพื่อป้องกันอคติ

ขั้นตอนการเก็บข้อมูล

1. การรับผู้ป่วยเข้าสู่ระบบ (Patient Arrival and Registration) ผู้ป่วยที่เข้ารับบริการในแผนกฉุกเฉินอาจมาด้วยตนเอง (walk-in) หรือโดยรถพยาบาลฉุกเฉิน (Emergency Medical Service, EMS) หลังจากมาถึง เจ้าหน้าที่จะดำเนินการลงทะเบียนข้อมูลพื้นฐาน ได้แก่ ชื่อ–สกุล อายุ เพศ และวิธีการมาถึง พร้อมบันทึกลงในระบบโรงพยาบาล

2. การวัดสัญญาณชีพและบันทึกข้อมูล (Vital Signs Measurement and Data Recording) พยาบาลคัดกรอง (triage nurse) ทำการวัดและบันทึกสัญญาณชีพ ได้แก่ ความดันโลหิต (Blood pressure, BP) อัตราการเต้นของหัวใจ (Heart rate, HR) ค่าความอิ่มตัวของออกซิเจนในเลือด (Oxygen saturation, SpO_2) อุณหภูมิร่างกาย (Body temperature, Temp) จากนั้นทำการซักถามอาการสำคัญ (chief complaint) และโรคร่วม (comorbidities) ของผู้ป่วย ก่อนบันทึกข้อมูลทั้งหมดลงในแบบฟอร์มคัดกรอง (case record form)

3. การให้คะแนนระดับความเร่งด่วนโดยแพทย์ผู้เชี่ยวชาญ (แพทย์ฉุกเฉิน) ทำการประเมินข้อมูลของผู้ป่วยที่บันทึกไว้ และให้คะแนนระดับความเร่งด่วนตามเกณฑ์ Emergency Severity Index (ESI)

4. การให้คะแนนระดับความเร่งด่วนโดย ChatGPT-4o (รองรับคำสั่งเสียง) การศึกษาใช้โมเดลภาษาปัญญาประดิษฐ์ GPT-4o (OpenAI) ช่วงวันที่ 1 เมษายน – 31 พฤษภาคม 2568 ผ่านบัญชีสถาบัน ข้อมูลจากแบบฟอร์มคัดกรองมาตรฐาน (อาการสำคัญ สัญญาณชีพ โรคร่วม วิธีมาถึง และหมวดอาการนำ Criteria Based Dispatch, CBD) ผู้วิจัยเข้าสู่ระบบ ChatGPT ใช้ระบบรู้จำเสียงอัตโนมัติ (Automatic Speech Recognition; ASR) ภายในแอปพลิเคชัน เพื่อแปลงเสียงพูดภาษาไทยเป็นข้อความ (Speech-to-Text) โดยไม่มีการบันทึกหรือเก็บไฟล์เสียงไว้ในระบบ

การตั้งค่าไม่ได้ใช้ seed หรือปลั๊กอินเสริมใด ๆ การสั่งงาน (prompt) กำหนดชัดเจนให้โมเดลทำการประเมินระดับความเร่งด่วนตามเกณฑ์ Emergency Severity Index (ESI) โดยอ้างอิงเหตุผลจากสัญญาณชีพ อาการที่มีความเสี่ยงสูง หากผลลัพธ์ที่ได้ไม่สมบูรณ์ เช่น ไม่ปรากฏระดับ ESI หรือให้ข้อมูลไม่ครบถ้วน จะทำการป้อนข้อมูลซ้ำไม่เกิน 1 ครั้งต่อราย และบันทึกเหตุการณ์ดังกล่าวไว้เพื่อความโปร่งใส

System Prompt (TH) “คุณคือผู้ช่วยคัดกรองผู้ป่วยฉุกเฉินตามเกณฑ์ ESI ทำงานเป็น

ขั้นตอน: (1) รับข้อมูลจากแบบฟอร์มมาตรฐาน (gender, chief complaint, vital signs, comorbidities, arrival mode, CBD/หมวดอาการนำ, red flags) (2) ทวนข้อมูลแบบ readback อย่างเป็นระเบียบ (3) ถ้าข้อมูลไม่ครบเฉพาะส่วนสำคัญต่อ ESI ให้ขอเติม เพียงรายการที่จำเป็น (4) สรุประดับ ESI 1–5 พร้อมเหตุผลอิงเกณฑ์ ESI และข้อเสนอแนะเบื้องต้น และ (5) ส่งออกผลตาม โครงสร้างเอาต์พุตมาตรฐาน ด้านล่างโดยย่อและชัดเจน”

ตัวอย่างบทพูด “ผู้ป่วยเพศหญิง อาการนำเจ็บแน่นหน้าอกมา 30 นาที ร้าวไปแขนซ้าย เหงื่อออก สัญญาณชีพ: ชีพจร 120 ครั้งต่อนาที ความดัน 90/60 มิลลิเมตรปรอท หายใจ 28 ครั้งต่อนาที ค่าออกซิเจนปลายนิ้ว 90% ไม่มีโรคประจำตัว มาถึงโดยระบบการแพทย์ฉุกเฉิน (EMS) หมวดอาการนำ เจ็บแน่นทรวงอก Red flags ความดันต่ำ ชีพจรเร็ว เหงื่อออก ร้าวแขนซ้าย”

จริยธรรมการวิจัย

โครงการนี้ได้รับการพิจารณาและอนุมัติจากคณะกรรมการพิจารณาจริยธรรมการวิจัยในมนุษย์ สำนักงานสาธารณสุขจังหวัดอุบลราชธานี (เลขที่รับรอง SSJ.UB 2568-03.008) เมื่อวันที่ 31 มีนาคม พ.ศ. 2568 การดำเนินการวิจัยเป็นไปตามหลักเกณฑ์จริยธรรมการวิจัยในมนุษย์ที่กำหนดไว้ใน Declaration of Helsinki, Belmont Report และแนวทางปฏิบัติการวิจัยที่ดี (Good Clinical Practice, GCP) โดยผู้เข้าร่วมวิจัยทุกคนได้รับข้อมูลอย่างเพียงพอ และได้ลงนามในเอกสารแสดงความยินยอมก่อนเข้าร่วมการศึกษา

ในการเก็บข้อมูลด้วยคำสั่งเสียงภาษาไทย ระบบจะทำการประมวลผลเสียงเป็นข้อความโดยอัตโนมัติทันที โดยไม่มีการจัดเก็บไฟล์เสียงต้นฉบับ ข้อมูลเสียงทั้งหมดถูกลบหลังการประมวลผลเพื่อป้องกันการเข้าถึงซ้ำ และไม่มีการเก็บข้อมูลระบุตัวตนที่สามารถเชื่อมโยงกับผู้ป่วยได้

สถิติที่ใช้ในการวิเคราะห์ข้อมูล

ใช้โปรแกรมสำเร็จรูปสำหรับประมวลผลและวิเคราะห์ข้อมูล

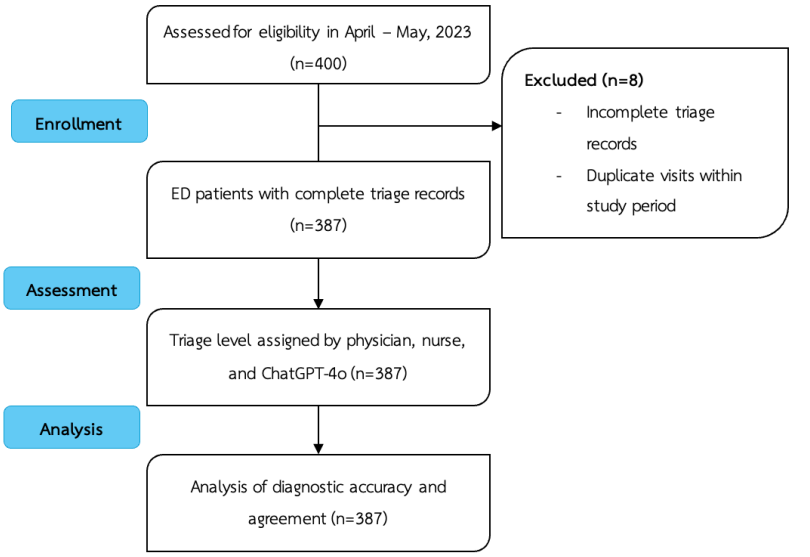
1. ใช้สถิติพรรณนาข้อมูล กรณีที่ข้อมูลมีการแจกแจงปกติ นำเสนอค่าเฉลี่ยส่วนเบี่ยงเบนมาตรฐาน และในกรณีข้อมูลแจกแจงไม่ปกติ นำเสนอค่ามัธยฐาน ค่า interquartile range ค่าสูงสุด และค่าต่ำสุด

2. ค่าทางวินิจฉัย (Diagnostic performance) สำหรับการจำแนกกรณีวิกฤต (ESI 1–2) เทียบกับ ไม่วิกฤต (ESI 3–5) คำนวณ sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) และ Area under the Receiver Operating Characteristic curve (AuROC) โดยใช้ exact 95% confidence intervals (DeLong’s method)

3. การวัดความสอดคล้อง (Agreement) ระดับความสอดคล้องระหว่างการประเมินของ ChatGPT-4o, พยาบาลคัดกรอง และแพทย์ฉุกเฉิน (มาตรฐานอ้างอิง) วิเคราะห์ด้วย quadratic-weighted κ (Fleiss–Cohen weights) พร้อมรายงานค่าประมาณและช่วงความเชื่อมั่น 95%

4. การทดสอบความแตกต่าง Comparative tests สำหรับการเปรียบเทียบ binary outcome ใช้ McNemar test วิเคราะห์จากตาราง 2×2 ของ discordant pairs

5. Confusion matrices แสดงผลในรูป confusion matrices ทั้งสำหรับการจำแนก



ภาพที่ 1 แผนผังการคัดเลือกผู้ป่วย การประเมินระดับ ESI โดยแพทย์ พยาบาล และ ChatGPT-4o (n=387)

5 ระดับ (5×5) และ (2×2) เพื่อให้เห็นทิศทางการจัดระดับที่คลาดเคลื่อน

ผลการศึกษา

กลุ่มตัวอย่างมีผู้ป่วยทั้งหมด 387 ราย อายุเฉลี่ย 46.8±23.9 โดยมีเพศชาย ร้อยละ 52.2 (202 ราย) และเพศหญิง ร้อยละ 47.8 (185 ราย) ผู้ป่วยส่วนใหญ่ ไม่มีโรคประจำตัว ร้อยละ 60.0 (231 ราย) ด้านสิทธิ์การรักษาพบว่าใช้สิทธิหลักประกันสุขภาพถ้วนหน้าเป็นส่วนใหญ่ ร้อยละ 80.6 (312 ราย) การมารับบริการเวรบ่ายมากที่สุด ร้อยละ 48.1 (186 ราย) มากกว่าเวรเช้า ร้อยละ 30.5 (118 ราย) และเวรดึก ร้อยละ 21.4 (83 ราย) โดยส่วนใหญ่มาในวันราชการ ร้อยละ 78.0 (302 ราย) และมาด้วยตนเอง ร้อยละ 74.7 (289 ราย) มีสัดส่วนผู้ป่วยอุบัติเหตุ (trauma) ร้อยละ 28.4 (110 ราย) การนำตามเกณฑ์ CBD ที่พบบ่อยที่สุด 5 อันดับ ได้แก่ ปวดท้อง/หลัง/เชิงกราน/ขาหนีบ ร้อยละ 18.9, ป่วย/อ่อนเพลีย/อัมพาตเรื้อรัง/ไม่ทราบสาเหตุจำเพาะ ร้อยละ 15.5, พลัดตกหกล้ม/อุบัติเหตุ/เจ็บปวด ร้อยละ 10.8, หายใจลำบาก ร้อยละ 10.6 และอุบัติเหตุยานยนต์ ร้อยละ 8.8 (ตารางที่ 1)

เมื่อแบ่งกลุ่มตามระดับความเร่งด่วนของ triage เป็น critical (ระดับ 1–2) และ non-critical (ระดับ 3–5) โดยยึดการประเมินของแพทย์ฉุกเฉินเป็นมาตรฐานอ้างอิง พบว่า ChatGPT มีสมรรถนะในการจำแนกกลุ่ม critical อยู่ในเกณฑ์สูง โดยให้ค่าความไว (sensitivity) ร้อยละ 97.6 (95%CI 94.1–99.4) ความจำเพาะ (specificity) ร้อยละ 95.9 (95%CI 92.3–98.1) ค่าพยากรณ์ผลบวก (PPV) ร้อยละ 94.8 (95%CI 90.4–97.6) ค่าพยากรณ์ผลลบ (NPV) ร้อยละ 98.1 (95%CI 95.3–99.5) และ (AuROC) ร้อยละ 97 (95%CI 95.0–99.0) ขณะที่ พยาบาลให้ค่าความไว ร้อยละ 92.3 (95%CI 87.2–95.8) ความจำเพาะ ร้อยละ 75.7 (95%CI 69.4–81.2) PPV ร้อยละ 74.6 (95%CI 68.2–80.4) NPV

ตารางที่ 1 ข้อมูลทั่วไปของผู้ป่วยที่มารับบริการแผนกฉุกเฉิน (n=387)

ตัวแปร	จำนวน (ร้อยละ)
เพศ	
ชาย	202 (52.2)
หญิง	185 (47.8)
อายุ (ปี) (mean ± SD)	46.8±23.9
โรคประจำตัว	
ไม่มีโรคประจำตัว	231 (60.0)
มีโรคประจำตัว	156 (40.0)
สิทธิ์การรักษา	
หลักประกันสุขภาพถ้วนหน้า (บัตรทอง)	312 (80.6)
ประกันสังคม	22 (5.7)
ข้าราชการ เบิกได้	25 (6.5)
อื่นๆ	28 (7.2)
เวรที่ผู้ป่วยเข้ารับการรักษา	
เช้า	118 (30.5)
บ่าย	186 (48.1)
ดึก	83 (21.4)
วันที่มาโรงพยาบาล	
วันราชการ	302 (78.0)
วันหยุด / เสาร์ อาทิตย์	85 (22.0)
วิธีการมาโรงพยาบาล	
มาเอง	289 (74.7)
ระบบการแพทย์ฉุกเฉิน(EMS)	49 (12.7)
ระบบส่งต่อ	28 (7.2)
ส่งจากห้องผู้ป่วยนอก	21 (5.4)
สัญญาณชีพ (mean ± SD)	
Temperature (°C)	36.9±0.8
Pulse (bpm)	90.3±21.1
Respiratory rate (rpm)	22.6±5.0
Systolic blood pressure (mmHg)	146.7±34.1
Oxygen saturation (%)	98.7±7.3
ประเภทผู้ป่วย	
Non-trauma	277 (71.6)
Trauma	110 (28.4)

ตารางที่ 1 ข้อมูลทั่วไปของผู้ป่วยที่มารับบริการแผนกฉุกเฉิน (n=387) (ต่อ)

ตัวแปร	จำนวน (ร้อยละ)
Criteria Based Dispatch (CBD)	
- CBD1 ปวดท้องบริเวณหลัง เชิงกราน และขาหนีบ	73 (18.9)
- CBD2 ภูมิแพ้ แอนาฟิแล็กซิส	7 (1.8)
- CBD3 สัตว์กัด	20 (5.2)
- CBD4 เลือดออก	2 (0.5)
- CBD5 หายใจลำบาก หายใจติดขัด	41 (10.6)
- CBD6 หัวใจหยุดเต้น	2 (0.5)
- CBD7 เจ็บแน่นทรวงอก	21 (5.4)
- CBD8 สำลัก อุดกั้นทางเดินหายใจ	0
- CBD9 เบาหวาน	5 (1.3)
- CBD10 ภาวะฉุกเฉินเหตุสิ่งแวดล้อม	0
- CBD11 -	0
- CBD12 ปวดศีรษะ ภาวะผิดปกทางตา หู คอ จมูก	22 (5.7)
- CBD13 คลุ้มคลั่ง ภาวะทางจิตประสาท	9 (2.3)
- CBD14 พิษ รับยาเกินขนาด	11 (2.8)
- CBD15 มีครรภ์ คลอด นรีเวช	6 (1.5)
- CBD16 ชัก มีสัญญาณบอกเหตุการชัก	9 (2.3)
- CBD17 อ่อนเพลีย อัมพาตเรื้อรัง	60 (15.5)
- CBD18 กล้ามเนื้ออ่อนแรง สูญเสียความรู้สึก	11 (2.8)
- CBD19 ไม่รู้สติ ไม่ตอบสนอง หมดสติชั่ววูบ	9 (2.3)
- CBD20 เด็ก กุมารเวช	0
- CBD21 ถูกทำร้าย	3 (0.8)
- CBD22 ไหม้ ลวกเหตุความร้อน ไฟฟ้าช็อต	0
- CBD23 ตกน้ำ จมน้ำ บาดเจ็บทางน้ำ	0
- CBD24 พลัดตกหกล้ม	42 (10.8)
- CBD25 อุบัติเหตุยานยนต์	34 (8.8)

*คำย่อ (Abbreviations): SD = standard deviation, °C = degrees Celsius, bpm = beats per minute, rpm = respirations per minute; mmHg = millimeters of mercury, CBD = Criteria Based Dispatch.

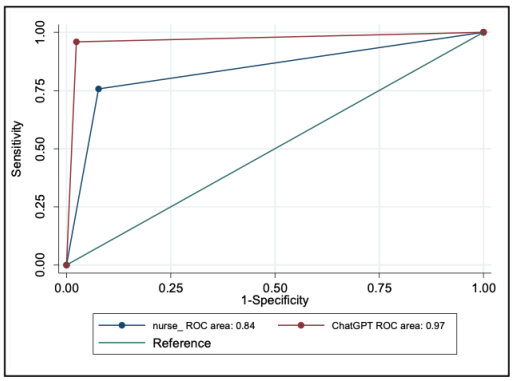
ร้อยละ 92.7 (95%CI 87.8–96.1) และ (AuROC) ร้อยละ 84 (95%CI 81.0–87.0) โดย ChatGPT แสดงค่าความไวและความจำเพาะสูงกว่าพยาบาล (ตารางที่ 2 และภาพที่ 2-3)

เมื่อเปรียบเทียบการจัดระดับความเร่งด่วนตาม triage ระหว่างแพทย์ พยาบาล และ ChatGPT พบว่า แพทย์จำแนกอยู่ในระดับ triage 2 และ 3 มากที่สุด (ร้อยละ 39.0 และ 45.5 ตามลำดับ) และไม่มีผู้ป่วยที่แพทย์จัดเป็นระดับ triage 5 ขณะที่พยาบาลมีแนวโน้มจัดอยู่ในระดับสูง (รุนแรงกว่า) มากกว่าแพทย์ โดยสัดส่วนระดับ triage 2 สูงถึง ร้อยละ 48.8 และระดับ triage 4 ต่ำกว่า

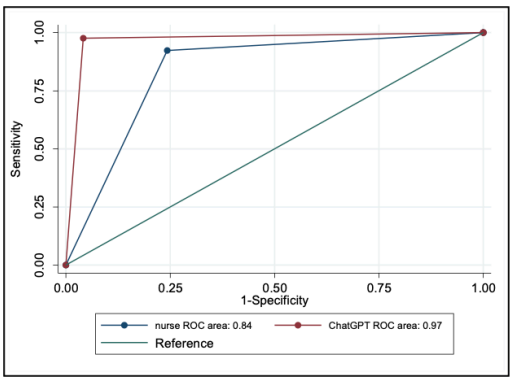
ตารางที่ 2 เปรียบเทียบความสามารถในการจำแนกของการใช้โมเดล ChatGPT-4o

Indicators % (95%CI)	Critical cases (ESI 1, 2)		Non-critical cases (ESI 3–5)	
	ChatGPT	พยาบาล	ChatGPT	พยาบาล
Sensitivity	97.6 (94.1–99.4)	92.3 (87.2–95.8)	95.9 (92.3–98.1)	75.7 (69.4–81.2)
Specificity	95.9 (92.3–98.1)	75.7 (69.4–81.2)	97.6 (94.1–99.4)	92.3 (87.2–95.8)
PPV	94.8 (90.4–97.6)	74.6 (68.2–80.4)	98.1 (95.3–99.5)	92.7 (87.8–96.1)
NPV	98.1 (95.3–99.5)	92.7 (87.8–96.1)	94.8 (90.4–97.6)	74.6 (68.2–80.4)
AuROC	97.0 (95.0–99.0)	84.0 (81.0–87.0)	97.0 (95.0–99.0)	84.0 (81.0–87.0)

* Triage 1–2 (critical / high acuity), Triage 3–5 (non-critical / lower acuity), ESI = Emergency Severity Index, PPV = positive predictive value, NPV = negai ve predictive value, AuROC = area under the ROC curve, CI = confidence interval.



ภาพที่ 2 เปรียบเทียบความสามารถในการจำแนกของการใช้โมเดล ChatGPT-4o ของกลุ่ม Critical cases (ESI 1, 2)



ภาพที่ 3 เปรียบเทียบความสามารถในการจำแนกของการใช้โมเดล ChatGPT-4o ของกลุ่ม Critical cases (ESI 3-5)

แพทย์ชัดเจน (ร้อยละ 2.6 เทียบกับ ร้อยละ 10.8) ส่วน ChatGPT ให้สัดส่วนใกล้เคียงแพทย์ในระดับ triage 1–3 และมีสัดส่วนระดับ triage 4 สูงกว่าแพทย์เล็กน้อย (ร้อยละ 12.7) เมื่อวิเคราะห์ด้วยดัชนีความสอดคล้อง (ordinal agreement) weighted kappa ระหว่างแพทย์กับ ChatGPT อยู่ที่ 0.915 (95%CI 0.844-0.986) สะท้อนความสอดคล้องระดับ almost perfect ขณะที่ระหว่างแพทย์กับพยาบาลอยู่ที่ 0.607 (95%CI 0.536-0.678) แสดงความสอดคล้องระดับ (substantial) (ตารางที่ 3)

การวิเคราะห์ความคลาดเคลื่อนของการ

ตารางที่ 3 การเปรียบเทียบความสอดคล้องของระดับ ChatGPT พยาบาลคัดกรองและแพทย์

ระดับ ESI	ChatGPT จำนวน (ร้อยละ)	พยาบาล จำนวน (ร้อยละ)	แพทย์ จำนวน (ร้อยละ)
1	18 (4.6)	20 (5.2)	18 (4.6)
2	156 (40.3)	189 (48.8)	151 (39.0)
3	163 (42.1)	166 (42.9)	176 (45.5)
4	49 (12.7)	10 (2.6)	42 (10.8)
5	1 (0.3)	2 (0.5)	0 (0.0)
รวม	387 (100.0)	387 (100.0)	387 (100.0)
Weighted kappa	0.915 (0.844-0.986)	0.607 (0.536-0.678)	Reference

คัดกรองพบว่า ChatGPT มี over-triage ร้อยละ 2.6 (10 ราย) และ under-triage ร้อยละ 3.4 (13 ราย) รวมความคลาดเคลื่อน ร้อยละ 5.9 (23 ราย) (p=0.270) ในขณะที่ พยาบาล มี over-triage ร้อยละ 23.3 (90 ราย) และ under-triage ร้อยละ 5.2 (20 ราย) รวมความคลาดเคลื่อน ร้อยละ 28.4 (110 ราย) (p<0.001) (ตารางที่ 4)

อภิปรายผล

จากผลการศึกษาครั้งนี้แสดงให้เห็นถึงความเป็นไปได้ของ ChatGPT-4o ที่รองรับคำสั่งเสียงภาษาไทย ในการคัดกรองผู้ป่วยฉุกเฉินตามเกณฑ์ Emergency Severity Index (ESI) ได้อย่างมีประสิทธิภาพเทียบเคียงกับแพทย์ฉุกเฉิน และสูงกว่าการคัดกรองโดยพยาบาลคัดกรอง ChatGPT-4o แสดงผลการจำแนกกลุ่มผู้ป่วย critical (ESI 1–2) ได้อย่างแม่นยำสูง โดยมีค่าความไว (sen-

ตารางที่ 4 ความคลาดเคลื่อนการคัดกรอง (Over-triage / Under-triage)

Parameters	ChatGPT จำนวน (ร้อยละ)	พยาบาล จำนวน (ร้อยละ)	p-value
Under-triage	13 (3.4)	20 (5.2)	0.270
Over-triage	10 (2.6)	90 (23.3)	<0.001
Overall mis-triage	23 (5.9)	110 (28.4)	<0.001

sitivity) ร้อยละ 97.6 และค่าความจำเพาะ (specificity) ร้อยละ 95.9 ซึ่งสูงกว่าเมื่อเทียบกับพยาบาลที่มีความไวเพียงร้อยละ 92.3 และความจำเพาะเพียงร้อยละ 75.7 สะท้อนให้เห็นถึงความสามารถของ AI ในการประมวลผลข้อมูลทางคลินิกและตัดสินใจในเชิงระบบได้แม่นยำและคงที่กว่าการพึ่งพาการตัดสินใจของบุคลากรแต่ละคนค่าพยากรณ์ผลบวก (PPV) และค่าพยากรณ์ผลลบ (NPV) ของ ChatGPT-4o ยังอยู่ในระดับสูงถึงร้อยละ 94.8 และ 98.1 ตามลำดับ เมื่อเปรียบเทียบกับพยาบาลที่มีค่าเพียงร้อยละ 74.6 และ 92.7 ตามลำดับ นอกจากนี้ ค่า Area under the Receiver Operating Characteristic curve (AuROC) ของระบบ ChatGPT อยู่ที่ร้อยละ 97 เทียบกับเพียงร้อยละ 84 ของพยาบาล ซึ่งสอดคล้องกับการศึกษา Kaboudi และคณะ[9] ดำเนินการทบทวนวรรณกรรมอย่างเป็นระบบและวิเคราะห์อภิมาน ครอบคลุมผู้ป่วยหรือสถานการณ์จำลองรวม 1,412 ราย ChatGPT 4.0 แสดงความแม่นยำรวมที่ 0.86 (ช่วงความเชื่อมั่น 95%: 0.64-0.98) ความแม่นยำที่สูงขึ้นของ ChatGPT-4o ในการศึกษาครั้งนี้ เกิดจากการประยุกต์ใช้โมเดลภาษาเวอร์ชันใหม่ที่สามารถรับข้อมูลผ่านคำสั่งเสียงภาษาไทยได้โดยตรง ประกอบกับการป้อนข้อมูลที่เป็นระบบและครบถ้วนจากแบบฟอร์มคัดกรองที่ผ่านการออกแบบอย่างเหมาะสม อีกทั้งยังสามารถประมวลผลข้อมูลได้อย่างคงที่ในทุกกรณี ไม่ได้รับผลกระทบจากอารมณ์ ประสบการณ์ หรือภาระงานของผู้ประเมิน ซึ่งอาจเป็นปัจจัยที่ทำให้ผลลัพธ์ของพยาบาลมีความแปรปรวนมากกว่า

ค่าความสอดคล้องของ ChatGPT กับแพทย์โดยใช้ weighted kappa อยู่ที่ 0.915 ซึ่งจัดอยู่ในระดับ “almost perfect agreement” ตามเกณฑ์ของ Landis และ Koch ขณะที่พยาบาลมีค่าความสอดคล้องอยู่ที่ 0.607 หรือ “substantial agreement” ความแตกต่างนี้ชี้ให้เห็นว่า ChatGPT-4o สามารถเรียนรู้และประมวลผลข้อมูลได้อย่างมีประสิทธิภาพจนสามารถตัดสินใจได้ใกล้เคียงกับผู้เชี่ยวชาญมากกว่าการประเมินของบุคลากรทั่วไป เมื่อพิจารณาเปรียบเทียบกับการศึกษาของ Paslı และคณะ[10] รายงานว่าค่าความสอดคล้อง (Cohen’s Kappa) ระหว่าง GPT-4 กับ gold standard = 0.899 ทีม triage กับ gold standard = 0.893 อย่างไรก็ตาม การศึกษาครั้งนี้ได้ศึกษาขั้นด้วยการใช้ GPT-4o ซึ่งสามารถประมวลผลจากคำสั่งเสียงภาษาไทยร่วมกับข้อมูลจากแบบฟอร์มคัดกรอง ESI ที่เป็นข้อมูลจริงจากหน้างาน (real-world triage data) โดยไม่ต้องผ่านการแปลงเป็นข้อความด้วยตนเอง ซึ่งสะท้อนบริบทการใช้งานจริงในแผนกฉุกเฉินที่มีข้อจำกัดด้านเวลาและทรัพยากรบุคคล

การศึกษาครั้งนี้พบว่าความคลาดเคลื่อน ChatGPT มี over-triage เพียงร้อยละ 2.6 และ under-triage ร้อยละ 3.4 รวมความคลาดเคลื่อนอยู่ที่ร้อยละ 5.9 และไม่แตกต่างจากแพทย์อย่างมีนัยสำคัญทางสถิติ (p=0.270) ซึ่งมีศักยภาพสูงกว่าการศึกษาของ Colakca และคณะ[8] ที่พบว่าผลการเปรียบเทียบระหว่างคณะผู้เชี่ยวชาญ (EC) และ ChatGPT พบว่า ChatGPT ให้การคัดกรองแบบมากเกินไป (over-triage) ร้อยละ 10.7 (80 ราย) และแบบน้อยเกินไป (under-triage) ร้อยละ 12.6 (94 ราย) เป็นไปได้ว่า ความแตกต่างของผลการวิจัยนี้อาจอธิบายได้จากหลายปัจจัยสำคัญ ประการแรก การศึกษาที่โรงพยาบาลวารินชำราบใช้ ChatGPT-4o ซึ่งมีความสามารถในการรองรับภาษาไทยได้อย่าง

แม่นยำ รวมถึงรองรับการรับข้อมูลผ่านคำสั่งเสียงโดยตรง ซึ่งสะท้อนถึงบริบทจริงของการใช้งานในระบบสุขภาพไทย อีกทั้งการศึกษาครั้งนี้ใช้ข้อมูลจริงจากผู้ป่วยในแผนกฉุกเฉิน (real-world data) ที่มีความหลากหลายทั้งในด้านอาการ วิธีการมาโรงพยาบาล และบริบทของโรงพยาบาลระดับภูมิภาคที่มีข้อจำกัดด้านทรัพยากร จึงสะท้อนถึงศักยภาพของ ChatGPT ในการใช้งานจริงได้ดีกว่าการจำลองข้อมูลในห้องทดลองหรือสถานการณ์สมมติ

การศึกษาครั้งนี้พบว่าพยาบาลมีอัตรา over-triage สูงถึงร้อยละ 23.3 และ under-triage ร้อยละ 5.2 รวมถึงความคลาดเคลื่อนสูงถึงร้อยละ 28.4 ซึ่งมีนัยสำคัญทางสถิติ ($p<0.001$) ซึ่งสูงกว่าการศึกษาของ Hinson และคณะ[11] ที่ทำการวิเคราะห์ย้อนหลังข้อมูลผู้ป่วยกว่า 96,071 ราย ที่เข้ารับการประเมิน triage ด้วยระบบ ESI โดยพยาบาลในแผนกฉุกเฉินประเทศบราซิล พบว่า มีการคัดกรองผิดพลาด (mistriage) ถึง 16,426 ราย หรือคิดเป็นร้อยละ 17.1 ของผู้ป่วยทั้งหมด เป็นไปได้ว่าความแตกต่างด้านโรคที่พบบ่อย ความเร่งด่วนในการเข้าถึงบริการหรือภาระงานของพยาบาลในแต่ละระบบ อาจส่งผลต่อความแม่นยำในการคัดกรอง ดังนั้นการพัฒนาคุณภาพต่อเนื่องควรมุ่งลดการคัดกรองคลาดเคลื่อน (mistriage) ด้วย 2 ด้านหลัก ได้แก่ (1) ด้านบุคลากร จัดอบรมระยะสั้นสำหรับพยาบาลคัดกรอง ใช้เคสจริงและจำลอง และทำการทบทวนผลงานพร้อมแจ้งผลกลับทุกเดือน และ (2) ด้านระบบ ให้มีการตรวจทานโดยเจ้าหน้าที่ก่อนยืนยันผลจาก ChatGPT-4o พร้อมทบทวน ถ้อยคำมาตรฐานตัวอย่างคำสั่งภาษาไทยทุกไตรมาสร่วมกับแพทย์ฉุกเฉิน

ข้อจำกัด

การศึกษาครั้งนี้มีข้อจำกัดหลายประการ ประการแรก เป็นการศึกษาในโรงพยาบาลระดับทั่วไปในภูมิภาคเพียงแห่งเดียว ซึ่งอาจจำกัดความสามารถในการทั่วไป (generalizability) ไปยังบริบทอื่นที่มีลักษณะผู้ป่วยหรือทรัพยากรแตกต่างกัน ประการที่สอง การประเมินของระบบ ChatGPT อาศัยข้อมูลจากแบบฟอร์มคัดกรองที่จัดทำโดยพยาบาลคัดกรอง ดังนั้นความครบถ้วนและความถูกต้องของข้อมูลที่บันทึกอาจส่งผลโดยตรงต่อผลการวิเคราะห์ ประการที่สามการประเมินไม่ได้ดำเนินการแบบสุ่มลำดับ (non-random order) ซึ่งอาจทำให้เกิดอคติบางประการต่อผลลัพธ์ การตีความผลลัพธ์ควรคำนึงถึงประเด็นด้าน automation bias ซึ่งอาจทำให้บุคลากรมีแนวโน้มปรับการตัดสินใจให้สอดคล้องกับระบบ AI

นอกจากนี้การศึกษาครั้งนี้ใช้ความเห็นของแพทย์ฉุกเฉินเพียง 1 ท่านเป็นมาตรฐานอ้างอิง (gold standard) ในการตัดสินระดับความเร่งด่วนของผู้ป่วย ซึ่งอาจก่อให้เกิดอคติจากผู้ประเมิน (adjudicator bias) และไม่สามารถประเมินความแปรปรวนระหว่างผู้ประเมิน (inter-rater variability) ได้ การใช้คณะผู้เชี่ยวชาญหลายท่านหรือการตัดสินแบบฉันทามติอาจให้ค่าความสอดคล้อง (κ) ที่แตกต่างไปจากผลการศึกษาครั้งนี้ และสุดท้ายการศึกษาครั้งนี้ไม่ได้ประเมินผลลัพธ์เชิงปฏิบัติการ (operational outcomes) เช่น เวลาในการคัดกรอง ความแออัดในห้องฉุกเฉิน (ED crowding) อัตราผู้ป่วยออกจากห้องฉุกเฉินก่อนการตรวจ (leave without being seen, LWBS) หรือ throughput ของผู้ป่วย ซึ่งล้วนมีความสำคัญต่อการประเมินผลกระทบในทาง

ปฏิบัติจริง งานวิจัยในอนาคตจึงควรออกแบบให้ครอบคลุมปัจจัยเหล่านี้เพื่อสะท้อนประโยชน์และข้อจำกัดของระบบได้อย่างคลอบคลุมมากขึ้น

สรุป

การศึกษาครั้งนี้พบว่า ChatGPT-4o ซึ่งรองรับคำสั่งเสียงภาษาไทย มีความเป็นไปได้ในการช่วยคัดกรองผู้ป่วยฉุกเฉินตามเกณฑ์ ESI ได้อย่างแม่นยำใกล้เคียงแพทย์ และสูงกว่าการคัดกรองโดยพยาบาลในบางตัวชี้วัด อย่างไรก็ตาม การตีความควรคำนึงถึงข้อจำกัดด้านบริบทโรงพยาบาลเดียว ข้อมูลจากการบันทึกของพยาบาล และการไม่ประเมินผลลัพธ์เชิงปฏิบัติการ เช่น เวลาในการคัดกรองหรือ throughput ของผู้ป่วย ในระยะต่อไป ควรมีการพัฒนากลไกการแลกเปลี่ยนข้อมูล (knowledge sharing) ระหว่างโรงพยาบาลเพื่อสร้างฐานข้อมูลที่สามารถเปรียบเทียบและทดสอบซ้ำได้ (reproducibility) ตลอดจนประเมินความสามารถของโมเดลในบริบทที่หลากหลาย (generalizability) เพื่อสนับสนุนการบูรณาการ AI เข้าสู่ระบบบริการฉุกเฉินของประเทศไทยอย่างมีหลักฐานเชิงประจักษ์และอยู่ภายใต้การกำกับดูแลทางคลินิกที่เหมาะสม

กิตติกรรมประกาศ

ผู้วิจัยขอขอบพระคุณคณะผู้บริหาร แพทย์ พยาบาล และเจ้าหน้าที่แผนกฉุกเฉิน โรงพยาบาลวารินชำราบ ที่ให้ความร่วมมือและสนับสนุนการดำเนินงานวิจัยครั้งนี้ รวมทั้งผู้ป่วยทุกท่านที่เข้าร่วมการศึกษา

ผลประโยชน์ทับซ้อน (Conflict of interest)

-

ทุนวิจัย (Funding)

ไม่มีทุนสนับสนุน

เอกสารอ้างอิง

1. Sudaparn Thanyajira, Wanida Aopprasertsak. *Emergency and Mass Casualty Nursing*. 3rd ed. Bangkok: Samcharoen Panich; 2003.
2. FitzGerald G, Jelinek G, Scott D, Gerdtz M. Emergency department triage revisited. Emergency Medicine Journal. 2010 Feb 1;27(2):86–92. doi:10.1136/emj.2009.077081
3. Gilboy N, Tanabe T, Travers DA, Rosenau AM. Emergency Severity Index (ESI): A triage tool for emergency department care. J Emerg Nurs 2011.
4. Christ M, Grossmann E, Winter D, Bingisser R, Platz E. Modern triage in the emergency department. Dtsch Arztebl Int 2010;107(50): 892-8. Doi:10.3238/arztebl.2010.0892
5. Ratthapong Buriwong, editor. *MOPH ED. Triage*. Nonthaburi: Bureau of Medical Academics, Department of Medical Services, Ministry of Public Health; 2018.
6. Kanistha Suksamnan. A study of urgent patient care models in the medical outpatient department at HRH Princess Maha Chakri Sirindhorn Medical Center [Master's thesis in Public Health]. Bangkok: Thammasat University; 2019.
7. Singer RF, Infante AA, Oppenheimer CC, West CA, Siegel B. The use of and satisfaction with the Emergency Severity Index. J Emerg Nurs 2012;38(2):120-6. Doi:10.1016/j.jen.2010.07.004

8. Natcholphan Homkaew. Concept for developing an “AI” patient triage system to make the “emergency room” truly an “emergency room” [Internet]. 2024 [cited 2025 Feb 20]. Available from: https://www.thecoverage.info/news/content/6688
9. Colakca C, Ergin M, Ozensoy HS, Sener A, Guru S, Ozhasenekler A. Emergency department triaging using ChatGPT based on emergency severity index principles: a cross-sectional study. Sci Rep 2024;14:22106. Doi:10.1016/j.jen.2010.07.004
10. Kaboudi N, Firouzbakht S, Eftekhar MS, Fayazbakhsh F, Joharivarnoosfaderani N, Ghaderi S, et al. Diagnostic accuracy of ChatGPT for patients’ triage: a systematic review and meta-analysis. Arch Acad Emerg Med 2024;12(1):e60. Doi:10.1101/2024.05.20.24307543
11. Paslı S, Şahin AS, Beşer MF, Topçuoğlu H, Yadigaroğlu M, İmamoğlu M. Assessing the precision of artificial intelligence in ED triage decisions: Insights from a study with ChatGPT. Am J Emerg Med 202;78:170-5. Doi:10.1016/j.ajem.2024.01.037.
12. Hinson JS, Martinez DA, Schmitz PSK, Toerper M, Radu D, Scheulen J, et al. Accuracy of emergency department triage using the Emergency Severity Index and independent predictors of under-triage and over-triage in Brazil: a retrospective cohort analysis. Int J Emerg Med 2018;11(1):3. Doi:10.1186/s12245-017-0161-8.