นิพนธ์ต้นฉบับ | Original Articles

การประเมินศักยภาพของ ChatGPT ในการพยากรณ์การเสียชีวิตของผู้ป่วยติดเชื้อในกระแสเลือดในห้องฉุกเฉินโรงพยาบาลวารินชำราบ จังหวัดอุบลราชธานี

Evaluation of ChatGPT's Potential for Predicting Mortality Among Patients with Sepsis in the Emergency Department at Warinchamrab Hospital, Ubon Ratchathani Province

*ภัทรกันย์ แต้ภักดี, พ.บ., เวชกรรมสาขาเวชศาสตร์ฉุกเฉิน**
*Pathtarakun Taepakdee, M.D., Diploma of the Thai BoardEmergency Medicine**
**กลุ่มงานอุบัติเหตุและฉุกเฉิน โรงพยาบาลวารินชำราบ จังหวัดอุบลราชธานี ประเทศไทย 34190*
**Department of Emergency Medicine, Warinchamrab Hospital, Ubonratchathani Province, Thailand, 34190*
Corresponding author, E-mail address: P_kunyakarn@hotmail.com
Received: 17 May 2026. Revised: 04 Jun 2026. Accepted: 09 Aug 2026.

บทคัดย่อ

หลักการและเหตุผล : ภาวะติดเชื้อในกระแสเลือด (Sepsis) เป็นภาวะฉุกเฉินที่มีอัตราการเสียชีวิตสูง อย่างไรก็ตาม ข้อมูลเกี่ยวกับศักยภาพของ ChatGPT ในการพยากรณ์การเสียชีวิตของผู้ป่วย sepsis ในบริบทเวชปฏิบัติจริงของประเทศไทยยังมีจำกัด

วัตถุประสงค์ : เพื่อประเมินศักยภาพของ ChatGPT ในการพยากรณ์การเสียชีวิตภายใน 7 วัน และ 28 วันของผู้ป่วยภาวะติดเชื้อในกระแสเลือดที่เข้ารับการรักษาในห้องฉุกเฉิน

วิธีการศึกษา : การศึกษาครั้งนี้เป็น prospective study ในผู้ป่วยภาวะติดเชื้อในกระแสเลือดที่เข้ารับการรักษา ณ ห้องฉุกเฉิน โรงพยาบาลวารินชำราบ จังหวัดอุบลราชธานี ระหว่างวันที่ 1 มกราคม ถึง 31 พฤษภาคม พ.ศ. 2569 จำนวน 272 ราย ข้อมูลทางคลินิกวันแรกถูกนำเข้าสู่ ChatGPT (GPT-4) ผ่าน structured prompt เพื่อพยากรณ์ความเสี่ยงการเสียชีวิตภายใน 7 วัน และ 28 วัน

ผลการศึกษา : จากผู้ป่วยทั้งหมด 272 ราย พบผู้เสียชีวิตภายใน 7 วัน จำนวน 12 ราย (ร้อยละ 4.4) และเมื่อครบ 28 วัน มีผู้เสียชีวิตสะสมรวม 26 ราย (ร้อยละ 9.6) ผู้ป่วยที่เสียชีวิตมีอายุเฉลี่ยสูงกว่า มีโรคร่วมมากกว่าและมีความรุนแรงของโรคมากกว่าอย่างมีนัยสำคัญทางสถิติ โดยพบภาวะ septic shock ความดันโลหิต systolic ต่ำ อัตราการหายใจสูง ค่าออกซิเจนปลายนิ้วต่ำ การใช้ยากระตุ้นความดันโลหิต และการได้รับสารน้ำภายใน 1 ชั่วโมงแรกมากกว่ากลุ่มรอดชีวิต ($p<0.05$) ChatGPT สามารถพยากรณ์การเสียชีวิตภายใน 7 วันได้ด้วย sensitivity ร้อยละ 83.3 specificity ร้อยละ 53.5 PPV ร้อยละ 7.6 และ NPV ร้อยละ 98.6 โดยมีค่า AUROC เท่ากับ 0.684 (95% CI: 0.57-0.79) สำหรับการพยากรณ์การเสียชีวิตภายใน 28 วัน พบ sensitivity ร้อยละ 73.1 specificity ร้อยละ 56.1 PPV ร้อยละ 15.0 และ NPV ร้อยละ 95.2 โดยมีค่า AUROC เท่ากับ 0.646 (95% CI: 0.55-0.73)

สรุป : ChatGPT มีความสามารถในการจำแนกผู้ป่วยที่เสียชีวิตและรอดชีวิตได้ โดยมีค่า AUROC เท่ากับ 0.684 สำหรับการพยากรณ์การเสียชีวิตภายใน 7 วัน และ 0.646 สำหรับการเสียชีวิตภายใน 28 วัน

คำสำคัญ : ภาวะติดเชื้อในกระแสเลือด ปัญญาประดิษฐ์ การพยากรณ์การเสียชีวิต

ABSTRACT

Background : Sepsis is a medical emergency associated with high mortality. Early prognostication is important for treatment planning. However, evidence regarding the potential of ChatGPT for mortality prediction among patients with sepsis in real-world clinical practice in Thailand remains limited.

Objective : To evaluate the potential of ChatGPT in predicting 7-day and 28-day mortality among patients with sepsis presenting to the emergency department.

Methods : This prospective study included 272 patients with sepsis presenting to the emergency department at Warinchamrab Hospital, Ubon Ratchathani Province, between January and May 2026. First-day clinical data were entered into ChatGPT (GPT-4) using a structured prompt to predict 7-day and 28-day mortality. Prediction outputs were converted into binary outcomes and compared with observed outcomes.

Results : Among 272 patients, 12 (4.4%) died within 7 days, while a total of 26 patients (9.6%) had died by 28 days. Patients who died were significantly older, had more comorbidities, and had greater disease severity. They had a higher prevalence of septic shock, lower systolic blood pressure, higher respiratory rate, lower oxygen saturation, more frequent vasopressor use, and greater intravenous fluid administration within the first hour than survivors ($p<0.05$). ChatGPT predicted 7-day mortality with a sensitivity of 83.3%, specificity of 53.5%, PPV of 7.6%, NPV of 98.6%, and an AUROC of 0.684 (95% CI: 0.57-0.79). For 28-day mortality, the sensitivity was 73.1%, specificity was 56.1%, PPV was 15.0%, NPV was 95.2%, and the AUROC was 0.646 (95% CI: 0.55-0.73).

Conclusions : ChatGPT demonstrated the ability to discriminate between survivors and non-survivors, with an AUROC of 0.684 for predicting 7-day mortality and 0.646 for predicting 28-day mortality.

Keyword : Sepsis, Artificial Intelligence, Mortality Prediction.

หลักการและเหตุผล

ภาวะ sepsis เป็นภาวะฉุกเฉินที่มีความรุนแรงและเป็นสาเหตุสำคัญของการเสียชีวิตทั่วโลก โดยองค์การอนามัยโลก (WHO) รายงานว่ามีผู้ป่วยมากกว่า 49 ล้านรายต่อปี และเสียชีวิตมากกว่า 11 ล้านราย หรือคิดเป็นร้อยละ 20 ของการเสียชีวิตทั้งหมด[1] สำหรับประเทศไทย ภาวะติดเชื้อในกระแสเลือดยังคงมีอัตราการเสียชีวิตสูงอย่างต่อเนื่อง โดยเฉพาะ severe sepsis ที่มีอัตราการเสียชีวิตมากกว่าร้อยละ 30 ในช่วงปี พ.ศ. 2562-2565 สะท้อนถึงความสำคัญของการพัฒนาระบบคัดกรอง การประเมินความรุนแรง และการ

พยากรณ์ผลลัพธ์ของผู้ป่วย sepsis อย่างมีประสิทธิภาพ[1,2] การพยากรณ์ผลลัพธ์ของผู้ป่วย sepsis ตั้งแต่ระยะแรกมีความสำคัญในแผนกฉุกเฉิน แนวทาง Surviving Sepsis Campaign 2021 จึงเน้นการประเมินและรักษาอย่างทันท่วงที โดยแนะนำให้ยาต้านจุลชีพภายใน 1 ชั่วโมงในผู้ป่วย septic shock หรือมีความเป็นไปได้สูงที่จะเป็น sepsis และให้สารน้ำอย่างเหมาะสม[3] ปัจจุบันมีการใช้เครื่องมือประเมินความรุนแรงหลายชนิด เช่น SOFA, qSOFA, APACHE II และ SAPS II แต่เครื่องมือเหล่านี้ยังมีข้อจำกัดด้านความซับซ้อนและการต้องใช้ข้อมูลทางห้องปฏิบัติการหลายค่า ทำให้การนำไปใช้ในแผนกฉุกเฉินที่มีผู้ป่วยจำนวนมากอาจทำได้จำกัด[4,5] ในช่วงที่ผ่านมา เทคโนโลยีปัญญาประดิษฐ์ (AI) และ Machine Learning ได้รับความสนใจเพิ่มขึ้นในการพยากรณ์การเสียชีวิตของผู้ป่วย sepsis โดยสามารถเพิ่มความแม่นยำในการทำนายได้ อย่างไรก็ตาม โมเดลเหล่านี้ยังมีข้อจำกัดด้านการตีความผลลัพธ์และความโปร่งใสในการใช้งานจริง เนื่องจากเป็นลักษณะ black-box model[6-8]

เทคโนโลยี Large Language Models (LLMs) เช่น GPT-3, GPT-4, BioGPT และ Meditron ได้ปฏิวัติแนวคิดการประมวลผลข้อมูลทางการแพทย์ จากเดิมที่เน้นข้อมูลเชิงตัวเลข สู่การใช้ภาษาธรรมชาติ (Natural Language Processing: NLP) ซึ่งสามารถเข้าใจและเชื่อมโยงความสัมพันธ์เชิงเหตุผลระหว่างข้อมูลทางคลินิกที่อยู่ในรูปข้อความได้ LLMs สามารถตอบคำถาม วิเคราะห์เคส และอธิบายเหตุผลทางคลินิกได้ในรูปแบบที่แพทย์เข้าใจ ซึ่งเป็นจุดแข็งเมื่อเทียบกับโมเดล ML แบบเดิม[9] งานวิจัยที่สำคัญโดย Oh et al. จากประเทศเกาหลีใต้ ได้ศึกษาศักยภาพของ ChatGPT-3.5 และ ChatGPT-4 ในการพยากรณ์การเสียชีวิตในโรงพยาบาลของผู้ป่วย sepsis โดยใช้ฐานข้อมูล Korean Sepsis Alliance (KSA) ซึ่งรวบรวมข้อมูลผู้ป่วยกว่า 4,700 รายจาก 16 โรงพยาบาล ผลการศึกษาพบว่า GPT-4 สามารถทำนายการเสียชีวิตภายใน 7 วันหลังเข้าหอผู้ป่วยวิกฤติได้ด้วย AUROC สูงสุด 0.83 โดยใช้เพียงตัวแปร 4 ตัว ได้แก่ เพศ อายุ (Charlson Comorbidity Index) คะแนน SOFA และระดับกรดแลกติก (lactate level) เท่านั้น ซึ่งมีความแม่นยำ (AUROC 0.79) ผลลัพธ์ดังกล่าวแสดงให้เห็นถึงความสามารถของ LLM ในการเข้าใจบริบททางคลินิกและอนุมานผลลัพธ์โดยไม่ต้องพัฒนาอัลกอริธึมใหม่ ซึ่งแตกต่างจาก Machine Learning แบบเดิมที่ต้องฝึกจากข้อมูลจำนวนมาก[10]

การนำ ChatGPT มาประยุกต์ใช้ในทางเวชปฏิบัติโดยเฉพาะในห้องฉุกเฉิน มีความน่าสนใจเนื่องจากสามารถทำหน้าที่เป็น Clinical Decision Support System (CDSS) ที่ช่วยให้แพทย์ประเมินความเสี่ยงของผู้ป่วยได้รวดเร็วขึ้น งานของ Amacher et al. รายงานว่า GPT-4 สามารถทำนายผลการรอดชีวิตหลังภาวะหัวใจหยุดเต้น (cardiac arrest) ได้ด้วยค่า AUROC 0.85 ซึ่งใกล้เคียงกับคะแนน CAHP และ OHCA ที่ใช้ในการประเมินมาตรฐาน[11] นอกจากนี้ Basaran et al. พบว่า AtlasGPT และ ChatGPT 4.0 สามารถพยากรณ์การต้องทำ hemicraniectomy ในผู้ป่วย subarachnoid hemorrhage ได้แม่นยำกว่า WFNS scale (AUC 0.80 vs 0.72)[12] อย่างไรก็ตามภาวะ sepsis เป็นปัญหาสำคัญทางสาธารณสุขที่มีอัตราการเสียชีวิตสูง การพยากรณ์ผลลัพธ์ตั้งแต่ระยะแรกเป็นกุญแจสำคัญในการลดอัตราการเสียชีวิตและเพิ่มประสิทธิภาพของการใช้ทรัพยากรในห้องฉุกเฉินและ ICU การนำเทคโนโลยี Large Language Models โดยเฉพาะ ChatGPT เข้ามาช่วยประมวลผลข้อมูลทางคลินิกและให้การพยากรณ์เชิงบริบท มีศักยภาพที่จะเป็นจุดเปลี่ยนสำคัญของการแพทย์ฉุกเฉินในอนาคต อย่างไรก็ตาม ยังต้องมีการศึกษาเพิ่มเติมเพื่อทดสอบความแม่นยำ ความน่าเชื่อถือ และความปลอดภัยของการใช้งานในบริบทจริงของประเทศไทย

วัตถุประสงค์

เพื่อประเมินศักยภาพของ ChatGPT ในการช่วยพยากรณ์การเสียชีวิตของผู้ป่วยภาวะติดเชื้อในกระแสเลือด (Sepsis) ภายใน 7 วัน และ 28 วัน หลังเข้ารับการรักษาในห้องฉุกเฉิน โรงพยาบาลวารินชำราบ จังหวัดอุบลราชธานี

วิธีการศึกษา

ระเบียบวิธีวิจัย

การศึกษาครั้งนี้เป็นการศึกษาแบบไปข้างหน้า prospective study โดยมุ่งศึกษาศักยภาพของ ChatGPT ซึ่งเป็นแบบจำลองปัญญาประดิษฐ์ด้านภาษาขนาดใหญ่ (Large Language Model: LLM) ในการช่วยพยากรณ์การเสียชีวิตของผู้ป่วยภาวะติดเชื้อในกระแสเลือด (Sepsis) โดยใช้ข้อมูลทางคลินิกพื้นฐานที่ได้จากวันแรกของการเข้ารับการรักษาในโรงพยาบาล

ประชากรที่ศึกษา

ผู้ป่วยฉุกเฉินที่เข้ารับการรักษาในโรงพยาบาลวารินชำราบ ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2569 ถึง 31 พฤษภาคม พ.ศ. 2569

ขนาดตัวอย่าง

ในการศึกษาครั้งนี้ ผู้วิจัยได้คำนวณขนาดตัวอย่างจากการทบทวนวรรณกรรมของ Oh และคณะ[10] ซึ่งศึกษาความสามารถของ ChatGPT ในการพยากรณ์การเสียชีวิตของผู้ป่วย sepsis และพบว่า ChatGPT สามารถทำนายการเสียชีวิตภายใน 28 วันได้ร้อยละ 63 โดยใช้สูตรประมาณค่าสัดส่วน กำหนดค่า α = 0.05, Z = 1.96, prevalence = 0.63 และ precision = 0.063 ได้ขนาดตัวอย่างจำนวน 226 ราย และเมื่อเผื่อการสูญเสียข้อมูลร้อยละ 20 จึงกำหนดขนาดตัวอย่างทั้งหมด 272 ราย Inclusion criteria ได้แก่ ผู้ป่วยที่ได้รับการคัดแยกที่แผนกฉุกเฉินเป็นครั้งแรก ส่วน exclusion criteria ได้แก่ ผู้ป่วยที่มีเอกสารหรือข้อมูล ทางเวชระเบียนไม่ครบถ้วนสมบูรณ์

การเก็บข้อมูล

1. บันทึกข้อมูลทางคลินิกวันแรก จากเวชระเบียนและระบบสารสนเทศโรงพยาบาล (HOSxP) โดยใช้แบบฟอร์มที่กำหนด ซึ่งรวมถึงข้อมูลดังนี้

- อายุ เพศ โรคร่วมสำคัญ
- อาการนำและสัญญาณชีพ (BP, HR, RR, Temp, SpO_2, GCS)
- ผลการตรวจทางห้องปฏิบัติการ (Lactate, WBC, Platelet, Band neutrophil ถ้ามี)
- ข้อมูลการรักษาเบื้องต้น เช่น การให้สารน้ำ ยาปฏิชีวนะ ยากระตุ้นความดันโลหิต

2. บันทึกผลลัพธ์การรักษา (Outcome Data)

- การรอดชีวิตหรือเสียชีวิตภายใน 7 วัน และ 28 วัน หลังเข้ารับการรักษา

3. การป้อนข้อมูลเข้าสู่ ChatGPT (GPT-4 Model)

1. เพื่อประเมินศักยภาพของ ChatGPT (GPT-4) ในการพยากรณ์ความเสี่ยงการเสียชีวิตของผู้ป่วย Sepsis ผู้วิจัยได้ออกแบบ Prompt เฉพาะทาง (Specialized Prompt) ให้ ChatGPT ทำหน้าที่เสมือน "ผู้ช่วยแพทย์เวชศาสตร์ฉุกเฉิน" (Emergency Physician Assistant AI) ที่สามารถวิเคราะห์ข้อมูลทางคลินิกและให้การพยากรณ์ผลลัพธ์เชิงเหตุผลได้การออกแบบ prompt ดังกล่าวมีพื้นฐานจาก ตัวแปรสำคัญในแบบประเมินมาตรฐานทางคลินิก ได้แก่ Quick Sequential Organ Failure Assessment (qSOFA) Acute Physiology and Chronic Health Evaluation II (APACHE II) Mortality in Emergency Department Sepsis (MEDS)

โดยได้มีการคัดเลือกเฉพาะตัวแปรที่สำคัญและตัดตัวแปรที่ซ้ำซ้อนออก เพื่อให้โมเดลได้รับข้อมูลที่จำเป็นและสะท้อนสภาพทางคลินิกของผู้ป่วยได้อย่างครบถ้วน โดยไม่เกิดการซ้ำซ้อนของข้อมูล

System Prompt (TH): ChatGPT ผู้ช่วยแพทย์ฉุกเฉินพยากรณ์การเสียชีวิตผู้ป่วย Sepsis

"คุณคือผู้ช่วยแพทย์เวชศาสตร์ฉุกเฉิน ทำหน้าที่ช่วยประเมินความรุนแรงและพยากรณ์การเสียชีวิตของผู้ป่วยภาวะติดเชื้อในกระแสเลือด (Sepsis) ให้ดำเนินการดังนี้ :

1. รับข้อมูลพื้นฐานของผู้ป่วย (เพศ อายุ โรคร่วมสำคัญ ภาวะโรคระยะสุดท้าย)
2. รับข้อมูลอาการนำและสัญญาณชีพวันแรก (HR, RR, SBP, Temp, SpO_2, GCS, Septic shock)
3. รับค่าผลแลปที่สำคัญ (Lactate, Platelet, WBC)
4. วิเคราะห์ความรุนแรงและพยากรณ์ความเสี่ยงการเสียชีวิตภายใน 7 วัน และ 28 วัน
5. ให้ผลลัพธ์ในรูปแบบ "ความเสี่ยงต่ำ / ปานกลาง / สูง" พร้อมเหตุผลทางคลินิกประกอบ เช่น ภาวะ hypotension, tachycardia, hyperlactatemia ฯลฯ
6. ให้ข้อเสนอแนะเบื้องต้นในการดูแลผู้ป่วย (initial management suggestion)"

โครงสร้างของแบบฟอร์ม Prompt

ตัวอย่างบทเขียน "ผู้ป่วยชาย อายุ 72 ปี มีโรคประจำตัวเบาหวานและความดันโลหิตสูง อาการนำ: มีไข้ หนาวสั่น หายใจเร็ว และซึมเล็กน้อย สัญญาณชีพวันแรก: BP 90/60 mmHg, HR 120 ครั้ง/นาที RR 30 ครั้ง/นาที Temp 39°C, SpO_2 91%, GCS 13 ผลตรวจทางห้องปฏิบัติการ: Lactate 3.8 mmol/L, Platelet 110,000/µL, WBC 18,000/µL มีภาวะช็อกจากการติดเชื้อ (Septic shock) โปรดประเมินความเสี่ยงการเสียชีวิตของผู้ป่วยรายนี้ภายใน 7 วันและ 28 วัน พร้อมอธิบายเหตุผลทางคลินิกประกอบ และข้อเสนอแนะเบื้องต้นในการดูแลผู้ป่วย"

การแปลผล ผลการพยากรณ์จาก ChatGPT ถูกแปลผลและจัดกลุ่มให้อยู่ในรูปแบบผลลัพธ์ทวิภาค (binary outcome) เพื่อใช้ในการวิเคราะห์ทางสถิติ โดยผู้วิจัยกำหนดให้คำตอบที่ ChatGPT ประเมินว่า "ความเสี่ยงต่ำ" และ "ความเสี่ยงปานกลาง" ถูกจัดให้อยู่ในกลุ่ม "รอดชีวิต" และกำหนดรหัสเป็น 0 ขณะที่คำตอบที่ประเมินว่า "ความเสี่ยงสูง" ถูกจัดให้อยู่ในกลุ่ม "เสียชีวิต" และกำหนดรหัสเป็น 1 ทั้งนี้การแปลงผลลัพธ์ให้อยู่ในรูปแบบทวิภาคดังกล่าวมีวัตถุประสงค์เพื่อให้สามารถนำไปวิเคราะห์ประสิทธิภาพของแบบจำลองในการพยากรณ์การเสียชีวิตได้

สถิติที่ใช้ในการวิเคราะห์ข้อมูล

ใช้โปรแกรม STATA 10.1 สำหรับประมวลผลและวิเคราะห์ข้อมูล

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

2. โดยใช้สถิติอนุมาน chi-square test หรือ fisher exact ในตัวแปร categorical data และสถิติ independent t-test และ Mann-Whitney-U test ตัวแปรเชิงปริมาณที่มีการกระจายตัวปกติและไม่เป็นปกติ ตามลำดับ

วิเคราะห์ประสิทธิภาพด้วย sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive and negative likelihood ratios (LR+ และ LR–), odds ratio (OR) และ area under the receiver operating characteristic curve (AUROC) พร้อมช่วงความเชื่อมั่นร้อยละ 95 (95% CI)

จริยธรรมการวิจัย

ในการวิจัยครั้งนี้ ได้ผ่านการรับรองโดยคณะกรรมการพิจารณาจริยธรรมการวิจัยในมนุษย์ สำนักงานสาธารณสุขจังหวัดอุบลราชธานี เอกสารรับรองเลขที่ SSJ.UB 2568-11.027 วันที่รับรอง 28 พฤศจิกายน พ.ศ. 2568

ผลการศึกษา

จากผู้ป่วยทั้งหมด 272 ราย พบผู้เสียชีวิตภายใน 7 วัน จำนวน 12 ราย (ร้อยละ 4.4) และเมื่อครบ 28 วัน มีผู้เสียชีวิตสะสมรวม 26 ราย (ร้อยละ 9.6) ผู้ป่วยที่เสียชีวิตภายใน 28 วัน มีอายุสูงกว่า มีโรคร่วม ภาวะโรคระยะสุดท้าย และภาวะ septic shock มากกว่ากลุ่มรอดชีวิต นอกจากนี้ยังมีอาการหอบเหนื่อยเป็นอาการสำคัญแรกเริ่มบ่อยกว่า มีค่าความดันโลหิต systolic และ diastolic ต่ำกว่า อัตราการหายใจสูงกว่า ค่าออกซิเจนปลายนิ้วต่ำกว่า ระดับความรู้สึกตัว (GCS) ต่ำกว่า และมีการใช้ยากระตุ้นความดันโลหิตรวมทั้งการได้รับสารน้ำทางหลอดเลือดดำภายใน 1 ชั่วโมงแรกมากกว่า อีกทั้งมีจำนวนเกล็ดเลือดต่ำกว่า และมีคะแนน qSOFA, APACHE II และ MEDS สูงกว่ากลุ่มรอดชีวิตอย่างมีนัยสำคัญทางสถิติ ($p<0.05$) ขณะที่เพศ อัตราการเต้นของหัวใจ อุณหภูมิร่างกาย ระดับ lactate และจำนวนเม็ดเลือดขาวไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ (ตารางที่ 1)

ตารางที่ 1 เปรียบเทียบลักษณะพื้นฐานและข้อมูลทางคลินิกของผู้ป่วยระหว่างกลุ่มรอดชีวิตและเสียชีวิตภายใน 28 วัน

Variable	Survived within 28 days (n=246)	Died within 28 days (n=26)	p-value
Age (years), mean ± SD	61.2 ± 18.1	71.4 ± 15.4	<0.01
Male sex, n (%)	115 (46.7%)	16 (61.5%)	0.15
Presence of comorbidity, n (%)	159 (64.6%)	22 (84.6%)	0.04
Terminal illness, n (%)	0 (0.0%)	2 (7.7%)	0.01
Septic shock, n (%)	10 (4.1%)	8 (30.8%)	<0.01
First symptom: dyspnea, n (%)	76 (30.9%)	13 (50.0%)	0.04
SBP (mmHg), mean ± SD	145.1 ± 32.3	116.5 ± 34.4	<0.01
DBP (mmHg), mean ± SD	82.2 ± 17.8	70.4 ± 19.7	<0.01
HR (beats/min), mean ± SD	101.9 ± 20.5	100.9 ± 28.3	0.81
RR (breaths/min), mean ± SD	25.9 ± 6.5	36.4 ± 9.2	<0.01
Temperature (°C), mean ± SD	37.7 ± 1.3	37.5 ± 1.2	0.32
SpO_2 (%), mean ± SD	97.1 ± 3.4	95.1 ± 4.5	<0.01
GCS score, median (IQR)	15 (15-15)	15 (8-15)	<0.01
Vasopressor use, n (%)	9 (3.7%)	7 (26.9%)	<0.01
IV fluid within 1 hour, n (%)	14 (5.7%)	10 (38.5%)	<0.01
Lactate (mmol/L), median (IQR)	2.1 (1.4-2.8)	1.8 (1.4-3.7)	0.48
WBC (cells/μL), median (IQR)	11,525 (8,240-14,460)	13,170 (9,580-15,300)	0.20
Platelet count (cells/μL), median (IQR)	250,000 (187,000-336,000)	158,000 (101,000-280,000)	0.01
qSOFA score, mean ± SD	0.6 ± 0.6	1.6 ± 0.9	<0.01
APACHE II score, mean ± SD	0.5 ± 3.2	5.2 ± 10.3	<0.01
MEDS score, mean ± SD	4.1 ± 3.0	10.2 ± 3.3	<0.01

ผลการประเมินประสิทธิภาพของ ChatGPT ในการพยากรณ์การเสียชีวิต พบว่า ChatGPT มีความไวและค่าพยากรณ์ผลลบค่อนข้างสูงทั้งในการพยากรณ์การเสียชีวิตภายใน 7 วัน และ 28 วัน ขณะที่ความจำเพาะและค่าพยากรณ์ผลบวกอยู่ในระดับค่อนข้างต่ำ โดยประสิทธิภาพในการจำแนกผู้ป่วยที่เสียชีวิตและรอดชีวิตภายใน 7 วันสูงกว่าการพยากรณ์ภายใน 28 วัน (ตารางที่ 2)

ตารางที่ 2 การประเมินประสิทธิภาพของ ChatGPT ในการพยากรณ์การเสียชีวิตของผู้ป่วยภาวะติดเชื้อในกระแสเลือดภายใน 7 วัน และ 28 วัน (n=272 ราย)

Performance measure	7-day mortality (95%CI)	28-day mortality (95%CI)
Sensitivity, % (95% CI)	83.3 (51.6-97.9)	73.1 (52.2-88.4)
Specificity, % (95% CI)	53.5 (47.2-59.6)	56.1 (49.7-62.4)
Positive predictive value (PPV), % (95% CI)	7.6 (3.7-13.6)	15.0 (9.3-22.4)
Negative predictive value (NPV), % (95% CI)	98.6 (95.0-99.8)	95.2 (90.3-98.0)
Positive likelihood ratio (LR+) (95% CI)	1.79 (1.35-2.38)	1.66 (1.27-2.19)
Negative likelihood ratio (LR–) (95% CI)	0.31 (0.09-1.11)	0.48 (0.25-0.91)
odds ratio, (95% CI)	5.74 (1.38-26.72)	3.47 (1.44-8.34)
AUROC (95% CI)	0.684 (0.570-0.798)	0.646 (0.554-0.738)
p-value	0.02	<0.01

อภิปรายผล

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อประเมินศักยภาพของ ChatGPT ในการพยากรณ์การเสียชีวิตภายใน 7 วัน และ 28 วันของผู้ป่วยภาวะติดเชื้อในกระแสเลือดที่เข้ารับการรักษาในห้องฉุกเฉิน โรงพยาบาลวารินชำราบ โดยอ้างอิงนิยามภาวะ sepsis ตาม Sepsis-3 criteria[13] ผลการศึกษาพบว่า ChatGPT มีความสามารถในการพยากรณ์การเสียชีวิตในระดับปานกลาง โดยมีค่า AUROC สำหรับการพยากรณ์การเสียชีวิตภายใน 7 วัน เท่ากับ 0.684 และสำหรับการเสียชีวิตภายใน 28 วัน เท่ากับ 0.646 แม้ว่าค่าความแม่นยำโดยรวมจะยังไม่สูงเพียงพอสำหรับการใช้เป็นเครื่องมือหลักในการตัดสินใจทางคลินิก แต่ผลการศึกษานี้สะท้อนให้เห็นว่าโมเดลภาษาขนาดใหญ่ (large language model; LLM) อย่าง ChatGPT มีศักยภาพในการนำข้อมูลทางคลินิกเบื้องต้นมาประมวลผลเพื่อช่วยคัดกรองผู้ป่วยกลุ่มเสี่ยงต่ำได้ โดยเฉพาะจากค่า negative predictive value (NPV) ที่อยู่ในระดับสูงทั้งการพยากรณ์การเสียชีวิตภายใน 7 วัน และ 28 วัน

ผลการศึกษาครั้งนี้สอดคล้องกับแนวโน้มของงานวิจัยในต่างประเทศที่พบว่า artificial intelligence และ machine learning สามารถช่วยในการพยากรณ์ผลลัพธ์ของผู้ป่วย sepsis ได้[14] โดยเฉพาะในด้านการประเมินความเสี่ยงระยะต้นและการคัดกรองผู้ป่วยที่มีแนวโน้มทรุดลงอย่างรวดเร็ว งานศึกษาของ Liu และคณะ[15] ซึ่งพัฒนา interpretable machine learning model สำหรับการทำนาย sepsis risk ในผู้ป่วยห้องฉุกเฉิน พบว่าโมเดลที่ใช้ข้อมูล triage อย่างครอบคลุม ได้แก่ vital signs, demographic characteristics, medical history และ chief complaints มีประสิทธิภาพในการพยากรณ์ sepsis สูงกว่าการใช้เพียงสัญญาณชีพ โดย Gradient Boosting model ให้ค่า AUC สูงถึง 0.83 และยังสามารถอธิบายผลการทำนายได้ผ่าน SHAP และ LIME methods ซึ่งช่วยเพิ่ม transparency ของ AI model ในการใช้งานจริงทางคลินิก อย่างไรก็ตาม งานวิจัยส่วนใหญ่ก่อนหน้านี้มักใช้โมเดลที่ถูกพัฒนาขึ้นเฉพาะทางจากฐานข้อมูลขนาดใหญ่และอาศัยตัวแปรจำนวนมาก แตกต่างจากการศึกษานี้ที่ใช้ ChatGPT ซึ่งเป็น general-purpose LLM ที่ไม่ได้ถูกออกแบบมาเฉพาะสำหรับงานด้าน critical care โดยตรง แต่ยังสามารถแสดงศักยภาพในการประเมินความเสี่ยงจากข้อมูลทางคลินิกพื้นฐานได้ในระดับหนึ่ง

ผลการประเมินประสิทธิภาพของ ChatGPT ในการพยากรณ์การเสียชีวิต (ตารางที่ 2) พบว่า ChatGPT สามารถพยากรณ์การเสียชีวิตภายใน 7 วันได้ด้วย sensitivity ร้อยละ 83.3 specificity ร้อยละ 53.5 ค่า positive predictive value (PPV) ร้อยละ 7.6 และ negative predictive value (NPV) ร้อยละ 98.6 โดยมีค่า AUROC เท่ากับ 0.684 (95% CI: 0.57-0.79)

สำหรับการพยากรณ์การเสียชีวิตภายใน 28 วัน พบว่า ChatGPT มี sensitivity ร้อยละ 73.1 specificity ร้อยละ 56.1 ค่า PPV ร้อยละ 15.0 และค่า NPV ร้อยละ 95.2 โดยมีค่า AUROC เท่ากับ 0.646 (95% CI: 0.55-0.73) นอกจากนี้ ค่า positive likelihood ratio (LR+) สำหรับการพยากรณ์การเสียชีวิตภายใน 7 วันและ 28 วัน เท่ากับ 1.79 และ 1.66 ตามลำดับ ขณะที่ค่า negative likelihood ratio (LR–) เท่ากับ 0.31 และ 0.48 ตามลำดับ ผลการศึกษาครั้งนี้สอดคล้องกับการศึกษาของ Oh และคณะ[10] ซึ่งประเมินความสามารถของ ChatGPT ในการพยากรณ์การเสียชีวิตของผู้ป่วย sepsis จาก Korean Sepsis Alliance database โดยพบว่า GPT-4 สามารถพยากรณ์การเสียชีวิตภายใน 7 วัน ได้ด้วยค่า AUROC อยู่ระหว่าง 0.70-0.83 และการเสียชีวิตภายใน 30 วัน อยู่ในช่วงประมาณ 0.51-0.75 ขึ้นกับรูปแบบของ in-context learning และจำนวนตัวอย่างที่ใช้ในการ prompt ขณะที่ GPT-3.5 มีประสิทธิภาพต่ำกว่าอย่างชัดเจน โดยผู้วิจัยสรุปว่า GPT-4 มีศักยภาพในการพยากรณ์ short-term mortality ของผู้ป่วย sepsis แม้ว่าประสิทธิภาพจะยังต่ำกว่า machine learning models แบบเฉพาะทาง เช่น gradient boosting model อย่างไรก็ตาม specificity และ PPV ของ ChatGPT ในการศึกษานี้ยังอยู่ในระดับต่ำ สะท้อนว่าโมเดลมีแนวโน้ม overpredict ความเสี่ยงการเสียชีวิต ทำให้เกิด false positive ค่อนข้างมาก ซึ่งอาจนำไปสู่การใช้ทรัพยากรทางการแพทย์เกินความจำเป็นหากนำไปใช้จริงโดยไม่มีการประเมินร่วมกับแพทย์

ข้อจำกัด

เป็นการศึกษาแบบ single-center study และมีจำนวนผู้เสียชีวิตค่อนข้างต่ำ ซึ่งอาจส่งผลต่อความแม่นยำของผลการวิเคราะห์ นอกจากนี้ผลลัพธ์ของ ChatGPT อาจขึ้นกับรูปแบบ prompt ที่ใช้ และตัวโมเดลยังมีข้อจำกัดด้าน interpretability เนื่องจากเป็น black-box model อีกทั้งการศึกษานี้ใช้ข้อมูลเฉพาะช่วงแรกของการมารับบริการ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. ควรพิจารณานำ ChatGPT มาประยุกต์ใช้เป็นเครื่องมือสนับสนุนการตัดสินใจทางคลินิก (Clinical Decision Support Tool) ร่วมกับการประเมินของแพทย์และเครื่องมือประเมินมาตรฐาน เช่น qSOFA, SOFA และ APACHE II เพื่อสนับสนุนการคัดกรองผู้ป่วยและการจัดลำดับความเร่งด่วนในการดูแลผู้ป่วยภาวะติดเชื้อในกระแสเลือดในห้องฉุกเฉิน

2. ไม่ควรใช้ผลการพยากรณ์จาก ChatGPT เพียงอย่างเดียวในการตัดสินใจรักษา แต่ควรใช้ประกอบการประเมินทางคลินิก ผลการตรวจทางห้องปฏิบัติการ และดุลยพินิจของแพทย์ผู้ดูแล เพื่อเพิ่มความเหมาะสมและความปลอดภัยในการดูแลผู้ป่วย

3. ควรมีการศึกษาวิจัยในหลายสถาบัน (multicenter study) โดยใช้กลุ่มตัวอย่างที่มีขนาดใหญ่ขึ้น รวมทั้งเปรียบเทียบประสิทธิภาพของ ChatGPT กับเครื่องมือพยากรณ์มาตรฐานและแบบจำลองปัญญาประดิษฐ์เฉพาะทาง เพื่อยืนยันความแม่นยำ ความน่าเชื่อถือ และความเหมาะสมก่อนนำไปประยุกต์ใช้ในเวชปฏิบัติอย่างแพร่หลาย

สรุป

ChatGPT มีความสามารถในการจำแนกผู้ป่วยที่เสียชีวิตและรอดชีวิตได้ โดยมีค่า AUROC เท่ากับ 0.684 สำหรับการพยากรณ์การเสียชีวิตภายใน 7 วัน และ 0.646 สำหรับการเสียชีวิตภายใน 28 วัน อย่างไรก็ตาม ค่า negative predictive value ที่สูงสะท้อนถึงศักยภาพในการช่วยคัดกรองผู้ป่วยกลุ่มความเสี่ยงต่ำ แม้ว่าประสิทธิภาพโดยรวมยังไม่เพียงพอสำหรับการใช้แทนการประเมินทางคลินิกหรือเครื่องมือมาตรฐาน แต่ผลการศึกษานี้สะท้อนให้เห็นถึงแนวโน้มของการประยุกต์ใช้ large language model ในเวชศาสตร์ฉุกเฉินและเวชบำบัดวิกฤต ซึ่งอาจมีบทบาทสำคัญในอนาคตในฐานะ clinical decision support tool ร่วมกับการประเมินโดยแพทย์

เอกสารอ้างอิง

1. วิมลรัตน์ เสนาะเสียง. การพัฒนาระบบการดูแลผู้ป่วยติดเชื้อในกระแสเลือด (SEPSIS) โรงพยาบาลสมเด็จ. วารสารการศึกษาสุขภาพและการจัดการสิ่งแวดล้อม 2566;8(3):661-7.
2. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, et al. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study. Lancet 2020;395(10219):200-11. doi: 10.1016/S0140-6736(19)32989-7.
3. Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. Intensive Care Med 2021;47(11):1181-247. doi: 10.1007/s00134-021-06506-y.
4. Kądziołka I, Świstek R, Borowska K, Tyszecki P, Serednicki W. Validation of APACHE II and SAPS II scales at the intensive care unit along with assessment of SOFA scale at the admission as an isolated risk of death predictor. Anaesthesiol Intensive Ther 2019;51(2):107-11. doi: 10.5114/ait.2019.86275.
5. Tekin B, Kiliç J, Taşkin G, Solmaz İ, Tezel O, Başgöz BB. The Comparison of scoring systems: SOFA, APACHE-II, LODS, MODS, and SAPS-II in critically ill elderly sepsis patients. J Infect Dev Ctries 2024;18(1):122-30. doi: 10.3855/jidc.18526.
6. Bao C, Deng F, Zhao S. Machine-learning models for prediction of sepsis patients mortality. Med Intensiva (Engl Ed) 2023;47(6):315-25. doi: 10.1016/j.medine.2022.06.024.
7. Zhang G, Shao F, Yuan W, Wu J, Qi X, Gao J, et al. Predicting sepsis in-hospital mortality with machine learning: a multi-center study using clinical and inflammatory biomarkers. Eur J Med Res 2024;29(1):156. doi: 10.1186/s40001-024-01756-0.
8. Lundberg S, Lee Su-In. A Unified Approach to Interpreting Model Predictions. [Internet]. [cited 2025 Nov 1]. Available from:URL: https://www.researchgate.net/publication/317062430_A_Unified_Approach_to_Interpreting_Model_Predictions.
9. Chung P, Fong CT, Walters AM, Aghaeepour N, Yetisgen M, O'Reilly-Shah VN. Large Language Model Capabilities in Perioperative Risk Prediction and Prognostication. JAMA Surg 2024;159(8):928-37. doi: 10.1001/jamasurg.2024.1621.
10. Oh N, Cha WC, Seo JH, Choi SG, Kim JM, Chung CR, et al. ChatGPT Predicts In-Hospital All-Cause Mortality for Sepsis: In-Context Learning with the Korean Sepsis Alliance Database. Healthc Inform Res 2024;30(3):266-76. doi: 10.4258/hir.2024.30.3.266.
11. Amacher SA, Arpagaus A, Sahmer C, Becker C, Gross S, Urben T, et al. Prediction of outcomes after cardiac arrest by a generative artificial intelligence model. Resusc Plus 2024;18:100587. doi: 10.1016/j.resplu.2024.100587.
12. Basaran AE, Güresir A, Knoch H, Vychopen M, Güresir E, Wach J. Beyond traditional prognostics: integrating RAG-enhanced AtlasGPT and ChatGPT 4.0 into aneurysmal subarachnoid hemorrhage outcome prediction. Neurosurg Rev 2025;48(1):40. doi: 10.1007/s10143-025-03194-w.

13. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 2016;315(8):801-10. doi: 10.1001/jama.2016.0287.
14. Haas R, McGill SC. Artificial Intelligence for the Prediction of Sepsis in Adults: CADTH Horizon Scan [Internet]. Ottawa (ON): Canadian Agency for Drugs and Technologies in Health; 2022 [cited 2026 May 15]. (CADTH Horizon Scans). Available from:URL https://www.ncbi.nlm.nih.gov/books/NBK596676/.
15. Liu Z, Shu W, Li T, Zhang X, Chong W. Interpretable machine learning for predicting sepsis risk in emergency triage patients. Sci Rep 2025;15(1):887. doi: 10.1038/s41598-025-85121-z.