ICU
MANAGEMENT & PRACTICE

LLM Capabilities in Perioperative Risk Prediction, Prognostication

General-domain large language models may perform risk stratification and predict postoperative outcome measures using procedure descriptions and a patient's electronic health record notes.

A recent study examined the predictive performance of a model on eight tasks: prediction of American Society of Anesthesiologists Physical Status (ASA-PS), hospital admission, ICU admission, unplanned admission, hospital mortality, PACU phase 1 duration, hospital duration, and ICU duration.

This prognostic study utilised task-specific datasets from two years of retrospective electronic health records gathered during routine clinical care. Cases and clinical notes were formatted into prompts and inputted into the large language model GPT-4 Turbo (OpenAI) to generate predictions and explanations.

The study was conducted at a quaternary care centre consisting of three academic hospitals and affiliated clinics within a single metropolitan area. Participants included patients who underwent surgery or procedures with anaesthesia and had at least one clinician-written note in the electronic health record before surgery. Data analysis was conducted between November and December 2023.

The study evaluated results on task-specific datasets, each comprising 1000 cases, except for unplanned admission (949 cases) and hospital mortality prediction (576 cases). Key findings included F1 scores of 0.50 for ASA-PS, 0.64 for hospital admission, 0.81 for ICU admission, 0.61 for unplanned admission, and 0.86 for hospital mortality prediction. However, performance on duration prediction tasks was uniformly poor across all prompt strategies, with mean absolute errors of 49 minutes for PACU phase 1 duration, 4.5 days for hospital duration, and 1.1 days for ICU duration prediction.

Current general-domain large language models show potential to aid clinicians in perioperative risk stratification for classification tasks but fall short in accurate numerical duration predictions. However, their capability to provide high-quality natural language explanations for predictions could prove valuable in clinical workflows and may complement traditional risk prediction models.

Source: JAMA

Image Credit: iStock

Published on : Tue, 11 Jun 2024

© For personal and private use only. Reproduction must be permitted by the copyright holder. Email to copyright@mindbyte.eu.