Leveraging Large Language Models to Optimize Clinical Text Analysis for In-Hospital Cardiac Arrest Identification

The Center for Resuscitation Science
Penn Medicine

Penn Institute for Biomedical Informatics

Aarthi Kaviyarasu, BS[1]; Ugurcan Vurgun, MA[3]; Sy Hwang, MS[3]; Ana Acevedo, BA[1]; Danielle L. Mowery, PhD, MS, MS[3]; Benjamin S. Abella, MD, MPhil[1]; Oscar J. L. Mitchell, MD, MSCE[1,2]

[1] Center for Resuscitation Science, Department of Emergency Medicine, University of Pennsylvania, Philadelphia, PA, USA; [2] Division of Pulmonary, Allergy, and Critical Care, Department of Medicine, University of Pennsylvania, Philadelphia, PA, USA; [3] Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA, USA

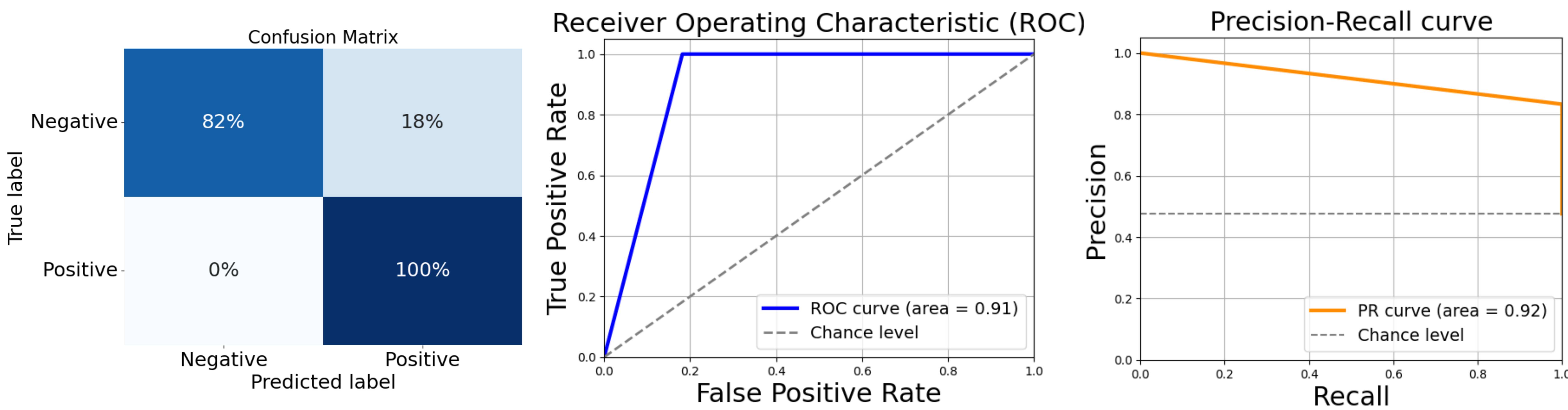
BACKGROUND

- In-hospital cardiac arrest (IHCA) is experienced by approximately 300,000 patients annually in the United States
- While individual care teams can readily identify IHCA at the bedside, post-event identification and reporting of IHCA often lacks consistency and poses a major problem at hospital across the U.S.
- Accurate and timely reporting of IHCA events is crucial for facilitating QI initiatives, such as cardiopulmonary resuscitation (CPR) quality review, optimizing clinical team performance, and benchmarking IHCA outcomes
- Large language models (LLMs) such as GPT-4 have previously been used to suggest diagnoses from free-text contained within clinical notes

OBJECTIVES & METHODS

Objective: To determine if using large language models (LLMs) for analysis of electronic health record (EHR) notes will facilitate accurate assertion of the presence of IHCA at the patient-level **Population:** Adult (≥ 18 years) inpatient encounters at the Hospital of the University of Pennsylvania from 06/2018 to 03/2022 with a reported clinical emergency **Data collection:** Patient encounters were identified using a QI database. All discharge summaries associated with the encounters of interest were pulled from the EHR and deidentified using PHIlter. Notes were reviewed by research staff to ascertain true IHCA labels (positive vs. negative). Positive IHCA was defined as the loss of pulses followed by the delivery of CPR **Large language model/Environment:** GPT4 v. 32K-Chat / Penn Medicine Microsoft Azure Databricks **Prompting methods and specifications:** **Zero-shot prompting:** Provided the LLM with only the prompt, no examples **Four-shot prompting:** Provided the LLM with 2 positive, 2 negative examples **Performance measures:** F1 Score**,** accuracy, precision, recall

FIGURES 1-3. Four-shot learning results



RESULTS

- Manual chart review determined 48% and 52% of cases to be IHCA+ and IHCA-, respectively, based on their discharge summaries
- Using **zero-shot learning**, the LLM performed with an accuracy of 81%, precision of 71%, recall of 100%, and F1-score of 83%
- Using **four-shot learning**, the LLM performed with an accuracy of 90%, precision of 83%, recall of 100%, and F1-score of 91%

To learn more about how to collaborate with PennAI, scan this QR code accessible via the PSOM VPN or UPHS network

DISCUSSION

Findings and Implications of Using LLMs to Identify IHCA

- This study demonstrates the potential efficacy of leveraging LLMs to automatically classify IHCA from discharge summaries and improved precision with n-shot learning
- These findings suggest that such technologies can be effective in real-world clinical settings, providing a scalable solution to improve patient outcomes and quality improvement efforts

Future Directions

- Refinement of model parameters (e.g., temperature, maximum output token size, selection of positive and negative examples)
- Implementation of temporal and strategic sampling of various note types to develop a more informed and clinically relevant model
- External validation of model using cases from other Penn Medicine hospitals (i.e., Penn Presbyterian Medical Center, Pennsylvania Hospital, HUP Cedar, Chester County Hospital)

Department of Medicine Research Day
Philadelphia, PA
June 11, 2024

Penn_CRS
OscarJMitchell
oscar.mitchell@pennmedicine.upenn.edu