Bigger is Not Always Better: Computational Efficiency in Lexical Prehospital Triage Modeling

Aaron C. Weidman, PhD, Chase Zikmund, Leonard S. Weiss, MD, Francis X. Guyette, MD, Ronald K. Poropatich, MD, David D. Salcido, PhD

University of Pittsburgh School of Medicine, Department of Emergency Medicine

Introduction

- Future LSCOs may strain military medics' capabilities
- AI analysis of natural language may support triage at scale (Fig. 1)
- Many assume that sophisticated large language models (LLMs) hold the most promise, but simpler machine learning models (MLMs) offer edge implementation advantages due to computational efficiency

Methods

- Sample: *N* = 12,913 patients treated by STAT MedEvac air medical transport from 2012-2021 (mean age = 52 years; 63% men)
- Medic Impressions: Avg length = 7.43 words (*SD* = 4.23)
- Lifesaving interventions (LSI): 28.2% sample-wide rate (*n* = 3,643)
  - Airway (e.g., endotracheal intubation, bag-valve-mask)
  - Bleeding control (e.g., tourniquet; tranexamic acid)
  - Cardiac (e.g., CPR)
  - Thoracic (e.g., chest tubes, needle decompression)
  - Vasopressors (e.g., dopamine, epinephrine)
  - Volume resuscitation (e.g., packed red blood cells; crystalloid)

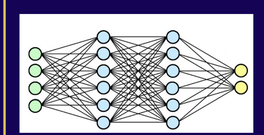
AI Modeling

Example text: *"Gunshot wound to lower leg with lots of blood loss."*

Machine Learning Models (MLMs)

1. Bag-of-words natural language processing on unique words and 2-word phrases
2. Histogram gradient boosting algorithm implemented via Scikit-Learn

Large Language Models (LLMs)

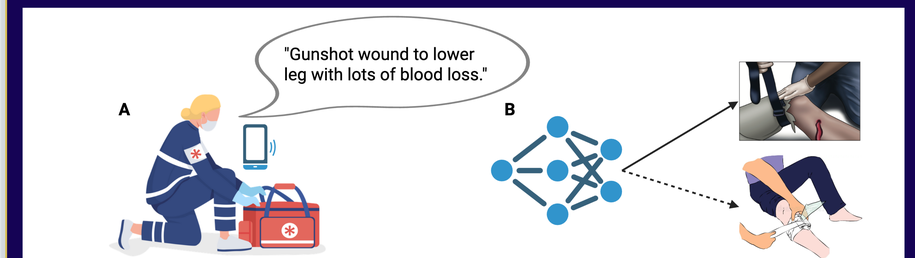


1. Impressions tokenized with embeddings
2. Two LLMs implemented via HuggingFace
   a) BioBERT (trained on medical text)
   b) DistilBERT (trained on general language)

All modeling performed on an iMac with M3 chip and 24GM memory

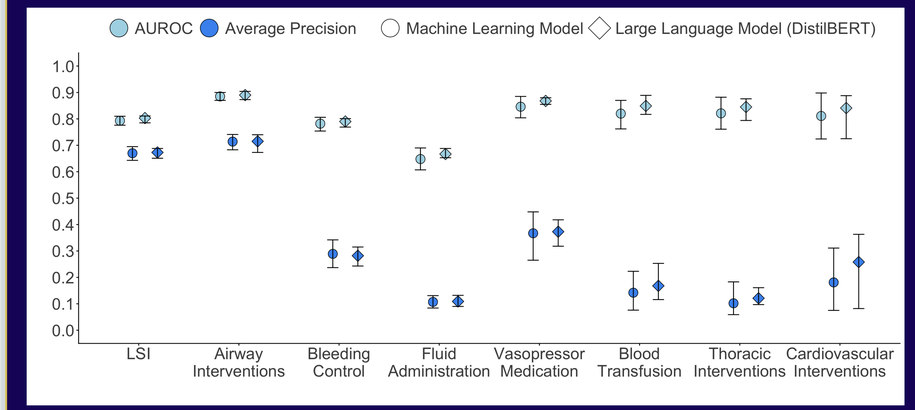
Results

Figure 1: Notional schematic of decision support via AI and natural language



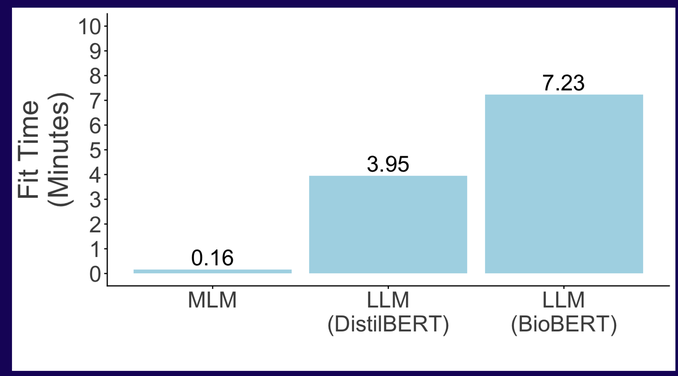
(A) On-scene clinician voices first impression to edge computing device. (B) Model recommends two interventions based on lexical input (solid line indicates best option if resources permit).

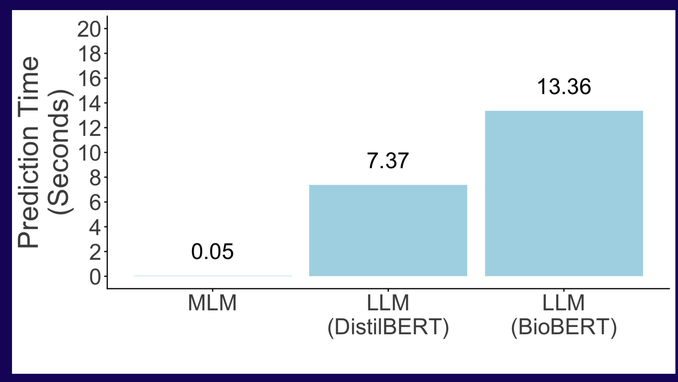
Figure 2: Results predicting LSI receipt from MLM and LLM



AUROC: Area under receiver-operator curve
Average Precision: Akin to area under the precision-recall curve
Note: Results for BioBERT LLM were nearly identical to DistilBERT LLM

Figure 3: Computational resources for MLMs and LLMs





Note: Fit time represents the time required to fully train the model on 80% of cases. Prediction time represents the time requires for the model to predict LSI receipt from the remaining 20% of unseen cases.

Limitations

- Retrospective study based on EHR data
- Use of only two encoder-based LLMs
  - Decoder-based LLMs (e.g., ChatGPT) could be examined
- Deployment of a local computing solution for MLMs and LLMs
  - LLMs could run faster on cloud computing devices

Conclusion

- Computational efficiency is a critical benchmark for military medical AI applications given the unique constraints of combat environments
- Computationally light MLMs performed equivalent to computationally expensive LLMs on a task analogous to battlefield casualty triage
- Future work should examine feasibility of AI assisted triage in field contexts via analysis of natural language as well as ambient acoustical data