LLM-guided Causal Bayesian Network construction for pediatric patients on ECMO

Saurabh Mathur*, Ranveer Singh*, Michael Skinner, Ethan Sanford, Neel Shah, Phillip Reeder, Lakshmi Raman, and Sriraam Natarajan

What?

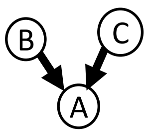
Problem Statement

Extracorporeal Membrane Oxygenation (ECMO) is a method for supporting patients with severe cardiac or respiratory failure. However, ECMO patients are at a higher risk of neurological injury. Understanding underlying causal mechanisms is critical for clinical decision-making. To this effect, we ask:

> What are the underlying causal relationships between the risk factors of neurological injury for patients on ECMO?

Background

Causal Bayesian Networks (CBN) represents causal relationships between variables using a directed acyclic graph (DAG)

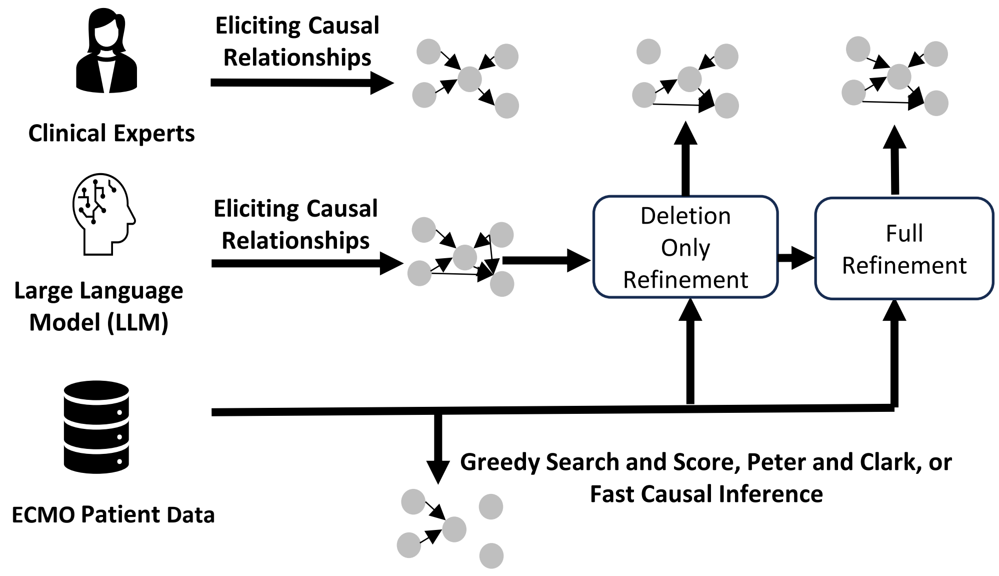


- Each edge B→A represents "B causes A"
- Allows for reasoning over interventions
- Causality cannot be inferred purely from observational data

Large Language Models capture knowledge, including causal relations, from their training data but they are stochastic and not amenable to reasoning.

How?

Methodology



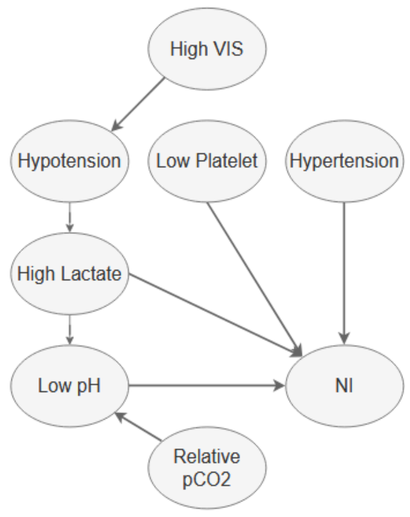
Data Description

Variable	Description	Frequency
High VIS	Vasoactive-inotropic score > 10	21.1%
Hypotension	Mean Arterial Pressure < 10th percentile for age	23.9%
Hypertension	Mean Arterial Pressure > 90th percentile for age	4.2%
Low Platelet	Platelet count < 50k/mm^3	32.4%
High Lactate	Lactate > 3 mmol/Liter	59.2%
Low pH	pH < 7.1	9.8%
Relative pCO2	Relative change in pCO2 from pre-ECMO to 24 hours on ECMO > 30%	29.6%
NI	Neurological Injury	23.9%

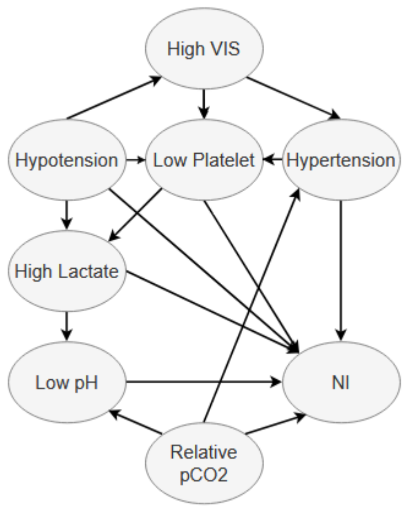
Data source: Children's Medical Center Dallas
Scope: 71 patients on ECMO from 0 to 24 hours.
Note: all variables except Relative pCO2 and NI are considered true if the corresponding condition holds at any point.

What we get

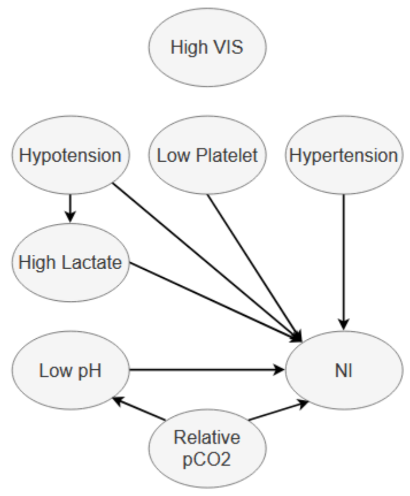
Empirical evaluation



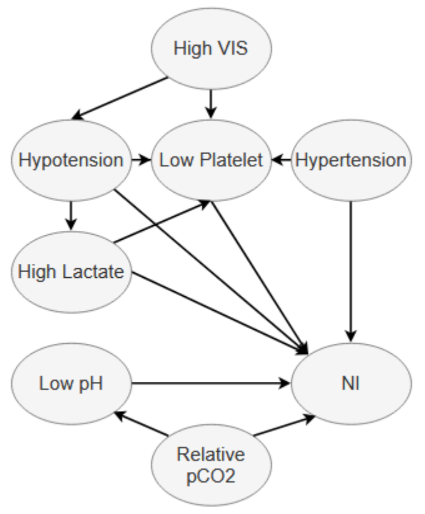
Expert Defined CBN



LLM-elicited CBN



LLM-elicited CBN after deletion only refinement



LLM-elicited CBN after full refinement

Data Driven vs LLM guided methods

Method	SE	SHD	SID
Greedy Search and Score (SS)	4.3 ± 2	9.7 ± 2	17.9 ± 2
Peter and Clark (PC)	0.8 ± 0.9	8.1 ± 1.4	16.5 ± 4
Fast Causal Inference (FCI)	0.1 ± 0.3	8 ± 0.4	14.9 ± 0.3
LLM	9	9	5
LLM + Deletion Only Refinement	3 ± 0.9	5.4 ± 1.1	5.1 ± 1.1
LLM + Full Refinement	4.8±1.1	6.6±1.3	6.2 ± 3.5

Number of Spurious Edges (SE), Structural Hamming Distance (SHD), and Structural Intervention Distance (SID) comparing the graphs constructed by data driven approaches (SS, PC and FCI), and the LLM-guided approaches (pooled output of all the LLMs, deletion-only refinement, and full refinement) over 10 bootstrap samples

Results

- LLM Constructed Graphs are closer to expert given graph compared to data-driven methods
- Deletion-only refinement correctly eliminates spurious edges
- Full refinement discovers novel edges albeit at the cost of more spurious edges

Future work

- Extending to multi-center data
- Scaling to larger variable spaces via relational models.
- Extending analysis to explicitly model relationships across time, e.g., pre-ECMO to on-ECMO.

Acknowledgements

We acknowledge the support from NIH R01 award R01NS133142.

StARLinG LAB

https://starling.utdallas.edu/

@STARLing_lab

*Equal contributors

Koller and Friedman 2009; Friedman and Goldszmidt 1996; Sprites et al. 2000; Mathur et al. 2024; Zečević et al. 2023; Shah et al. 2023