Harnessing large language models for prognostic variable discovery in intraventricular hemorrhage

Tanima Arora, MD MHS & Kristyn Beam, MD MPH
Division of Neonatology, Beth Israel Deaconess Medical Center

HARVARD MEDICAL SCHOOL TEACHING HOSPITAL
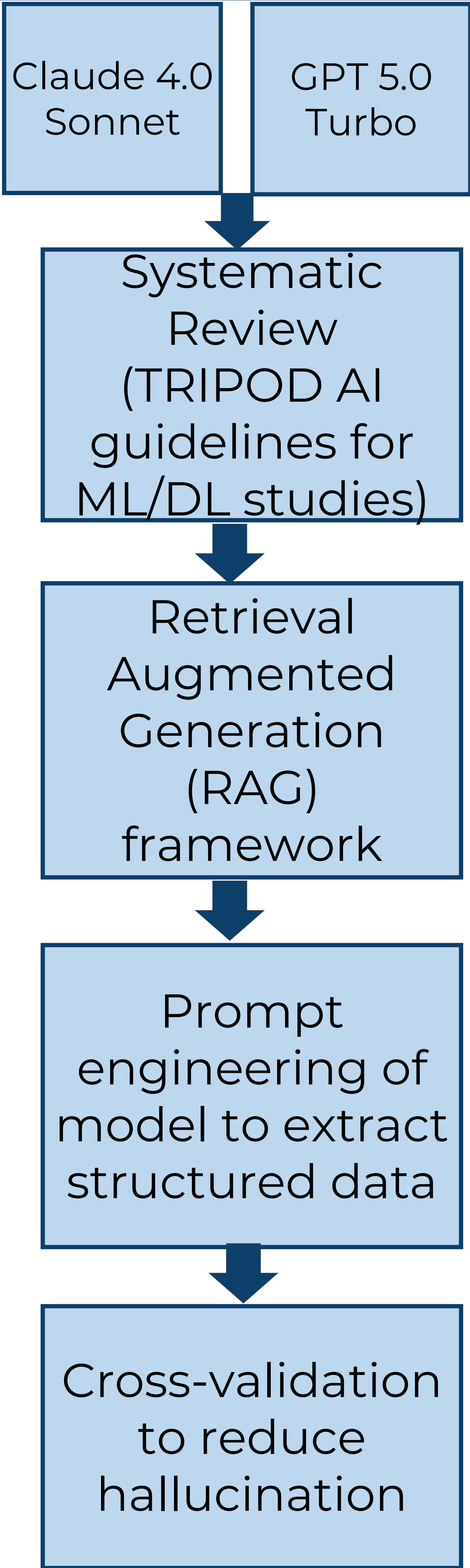Beth Israel Lahey Health

Abstract

Intraventricular hemorrhage (IVH) is a serious complication of prematurity, and current prognostication relies heavily on clinical judgement. This study used large language models (LLMs) within a retrieval-augmented framework to systematically identify predictors of IVH outcomes from existing machine learning literature. Across 31 studies, gestational age (GA), birth weight (BW), and APGAR scores emerged as the most consistent high-impact factors, with some additional outcome-specific variables identified. The findings provide a structural foundation for developing future AI-driven prognostic models.

Introduction

- IVH is a major complication of prematurity, particularly in very low birth weight infants

- Mortality rates range from 4-40% depending on severity

- Prognostication currently relies on clinical gestalt and basic demographic factors

- Generative AI models such as LLMs offer novel opportunities in the prediction of neonatal IVH outcomes

- The objective of this study was to apply LLMs for literature synthesis and prognostic variable discovery in IVH

- Outcomes: mortality, disease progression, complications (such as post-hemorrhagic hydrocephalus) and resolution
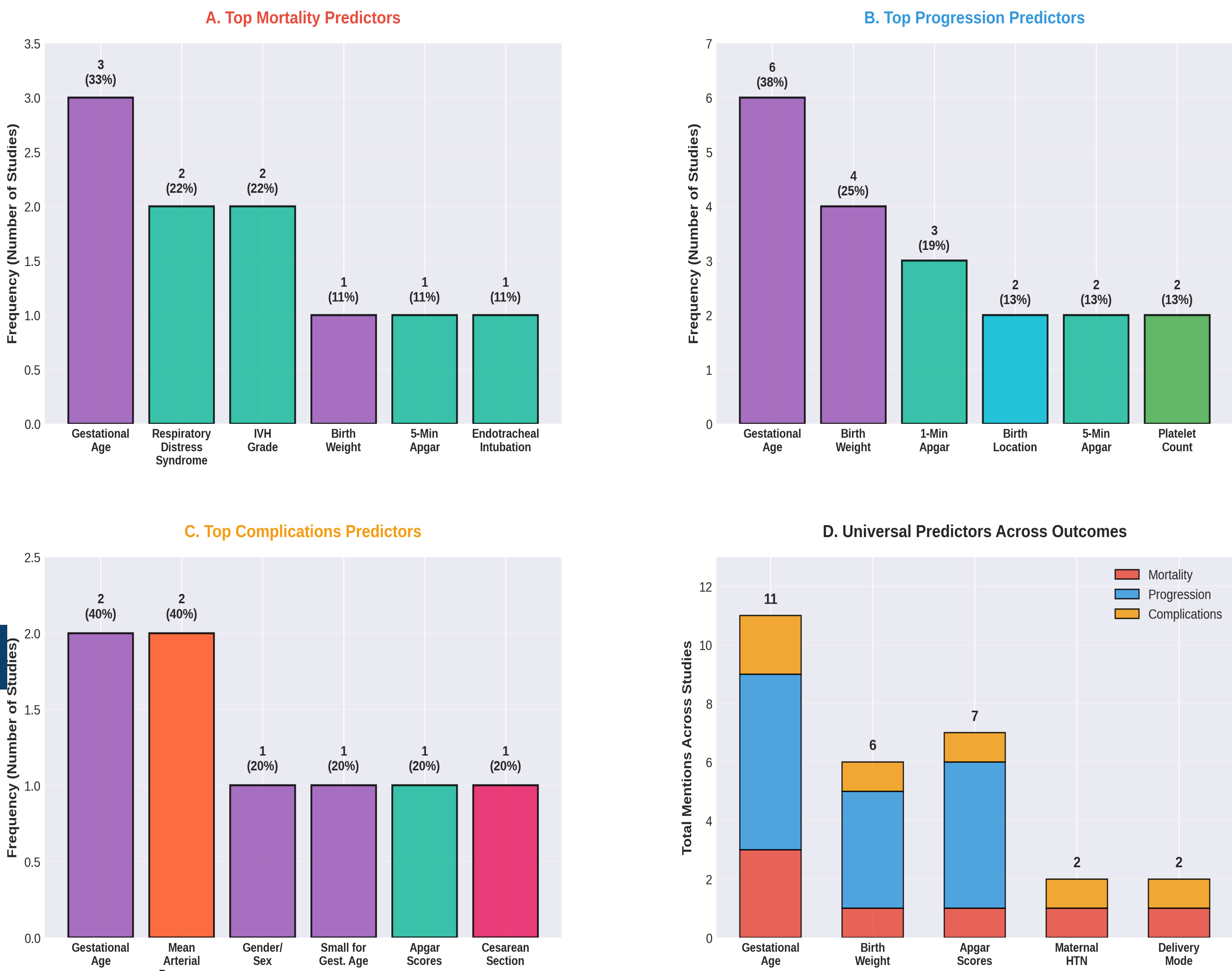
Methods

Claude 4.0 Sonnet | GPT 5.0 Turbo
↓
Systematic Review (TRIPOD AI guidelines for ML/DL studies)
↓
Retrieval Augmented Generation (RAG) framework
↓
Prompt engineering of model to extract structured data
↓
Cross-validation to reduce hallucination

Results

Figure 1. LLM extracted predictors for IVH outcomes

A: Top mortality predictors included GA, RDS, and Grade of IVH

B: An interesting predictor of the progression of IVH includes birth location

C: One of the top predictive variables for post-IVH complications includes mean arterial pressure

D: Across all outcome, GA, BW and APGAR score were the most frequent predictors

Discussion

- First systematic application of LLMs to neonatal IVH prognostication

- Demonstrates the feasibility and utility of using LLMs to consolidate heterogeneous findings in the IVH prognostication literature

- The absence of studies on resolution prediction highlights an unmet clinical need

- Discovery of outcome-specific modifiers, such as respiratory distress syndrome and IVH grade for mortality, platelet count and birth location for progression, and mean arterial pressure for complications, suggests nuances and potentially targetable physiologic pathways in IVH management

- A well-defined variable hierarchy provides a solid foundation for future AI-driven prognostic models

- Such models could empower neonatologists and families with clearer prognostic expectations, guide timing for neurosurgical consultation, and support decisions around escalation or redirection of care

- Future directions: training and validating predictions of text-based and image-based generative models (using head ultrasound data) in comparison to neonatologist prognostication

References

*References available on request

Let's connect: tanima.arora@childrens.harvard.edu