Predicting Hypotension in ICU Patients Receiving Vasopressor Therapy

A Deep Learning Approach using MIMIC-IV Data

Team 5 Aaryan Bammi, Pushpraj Singh, Yimeng Dong

Course BA878 – Fall 2025

The Critical Challenge: Moving Beyond Reactive Care

Hypotension in the ICU is a life-threatening event associated with organ failure and increased mortality. Current clinical practices often rely on reactive measures, treating pressure drops only after they occur.

The Gap

Traditional scores like SOFA and SAPS II are static, failing to capture rapid physiological changes. Manual monitoring often misses subtle early warning signs.

The Goal

Shift from **Reactive → Proactive** care by predicting hypotension onset within the first **24 hours** of vasopressor initiation.

Study Design & Cohort Selection

We leveraged the MIMIC-IV v2.2 database from Beth Israel Deaconess Medical Center, utilising high-resolution, minute-by-minute time-series vitals data.

1

Adult Patients

Restricted to patients aged 18 and older to ensure physiological consistency.

2

Target Definition:

Physiological Hypotension defined as Mean Arterial Pressure (MAP) < 65 mmHg sustained for ≥ 25 minutes

3

Active Treatment

Patients actively receiving vasopressors (Norepinephrine, Epinephrine, Phenylephrine, etc.).

Data Methodology: From Raw Records to Tensors

Ensuring data integrity was crucial for model stability. We constructed a robust pipeline to handle the complexities of real-world clinical data.

Extraction

Aggregated Demographics, Lab results, Vitals, and Medication administration logs. Aggregated into **10-minute bins** to reduce noise while preserving trend fidelity.

Cleaning & Imputation

Addressed missingness using **Forward Filling** to respect temporal causality, followed by zero Imputation for remaining gaps.

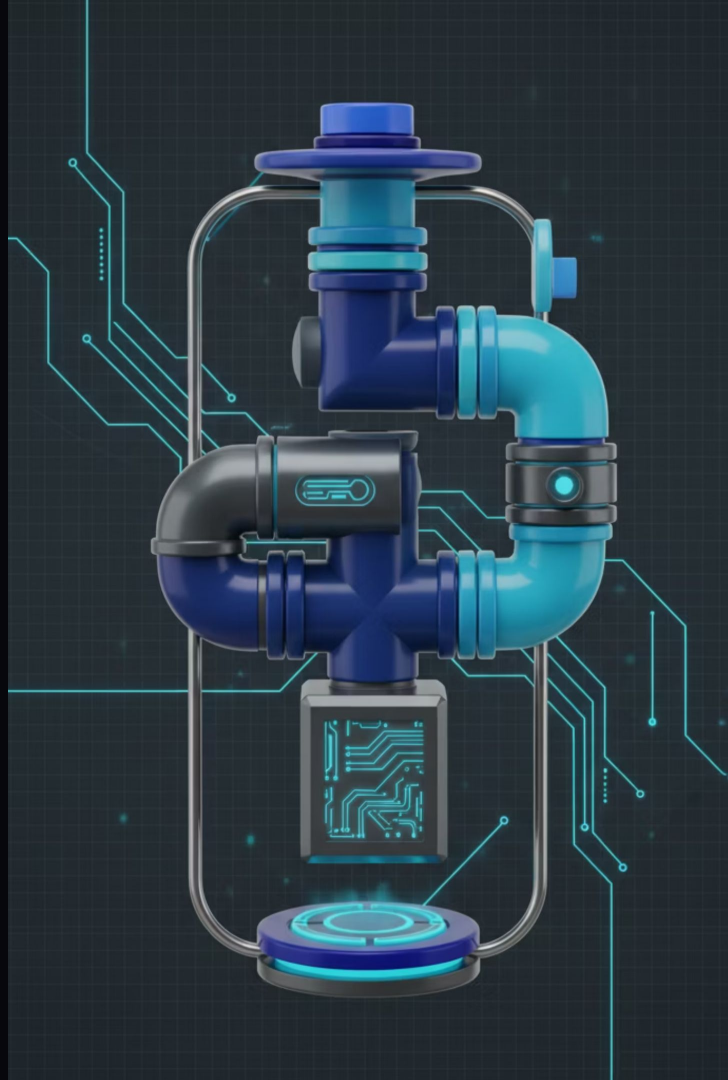
Normalization

Applied StandardScaler to normalize disparate clinical units (e.g., Blood Pressure mmHg vs. Lactate mmol/L).

Feature Engineering

Dynamic Tensor: Created **3-hour sliding windows** (T_{-180} to T_{0}) to capture physiological history.

Static Vector: Engineered a parallel feature set of demographics (Age, Race) and comorbidities (Sepsis, Heart Failure) for model fusion

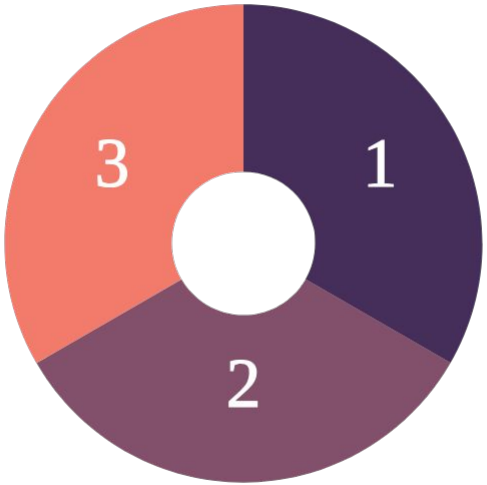


Model Architecture

GRU-Static and Transformer paths converge into a calibrated Super Learner

Super Learner Ensemble

Stacked generalization combines base outputs to optimize final risk probability and calibration.



Model A: GRU-Static Fusion

Temporal GRU with a static fusion layer integrates continuous vital signs and immutable patient attributes.

Model B: Transformer

Multi-head attention extracts long-range and cross-feature patterns across the 3-hour window and different modalities.

Model Comparison: **Top Performer** Spotlight

Quick comparison of predictive performance

Model Architecture	AUROC	Precision (PPV)	Recall (Sensitivity)	F1-Score
GRU-Fusion	0.91	0.84	0.73	0.78
Transformer	0.87	0.69	0.76	0.72
Super Learner (Ensemble)	0.91	0.82	0.75	0.78

We selected the **GRU-Fusion model** because it achieved superior **Precision (0.80 vs. 0.69)** compared to the Transformer.

Furthermore, the GRU architecture demonstrated greater training stability and computational efficiency on sparse clinical data, making it a robust and lightweight candidate for real-time bedside deployment.

Key results

Key predictive performance and reliability indicators for clinical deployment

0.914

AUC

Strong discrimination between outcomes

80 %

Precision

Minimizes false alarms

0.1

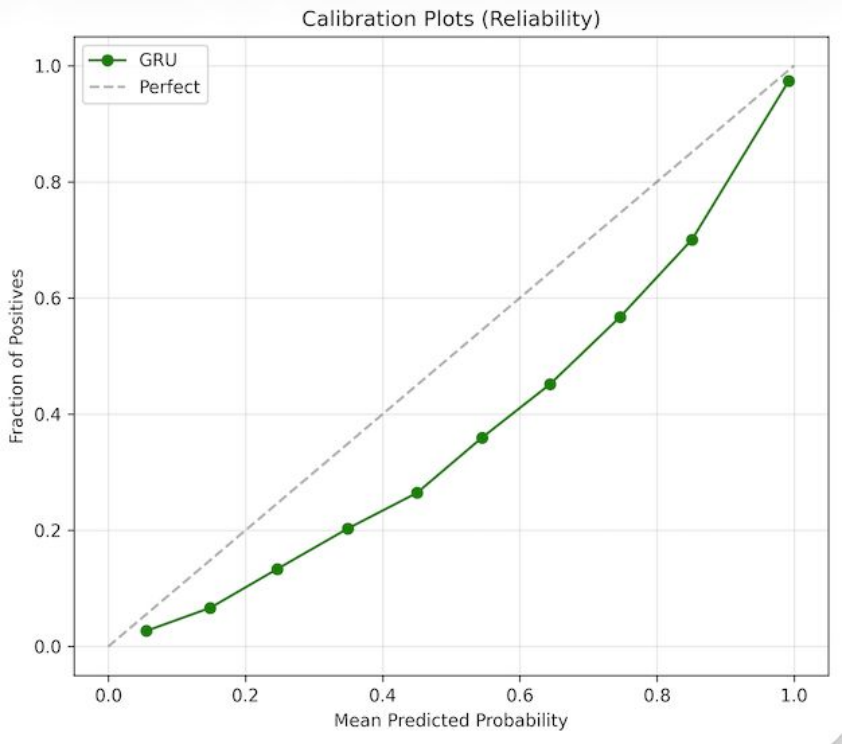
Brier Score

High reliability in probability estimates

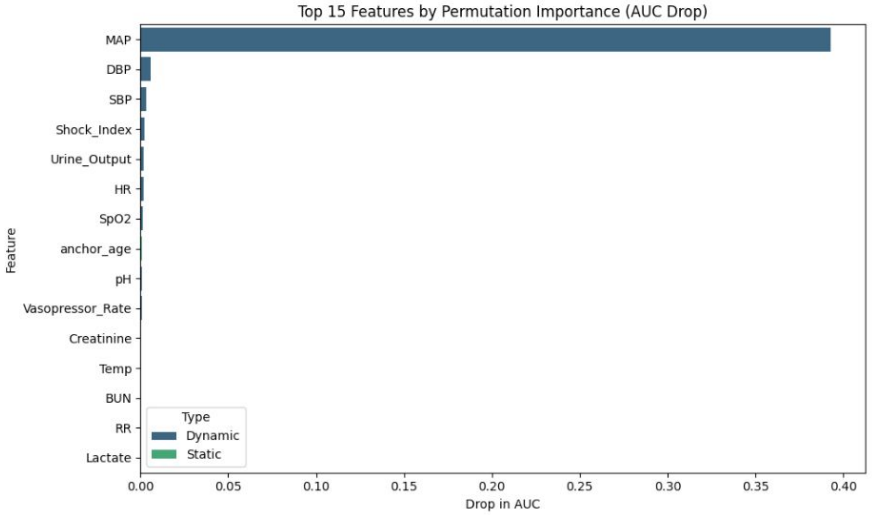
76 %

Precision

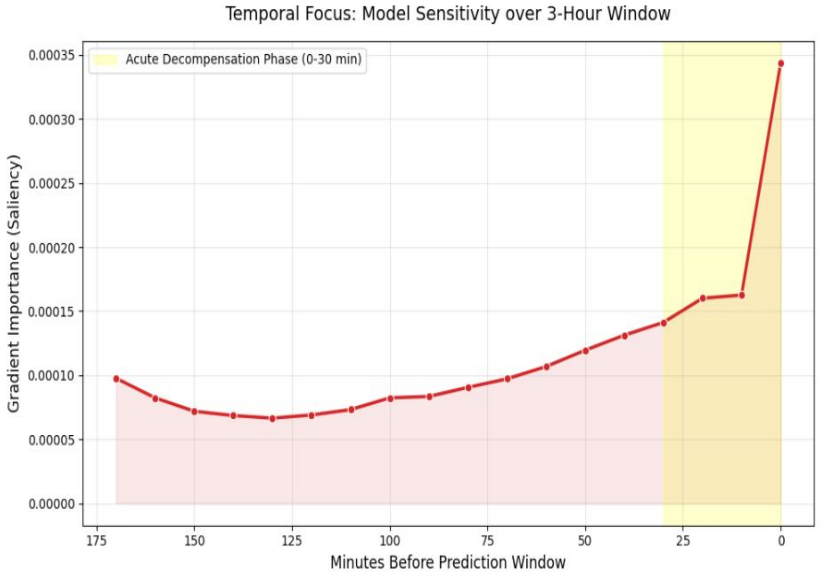
Minimizes false alarms



Feature Importance



Primary Driver: MAP Trends (Mean Arterial Pressure). The model prioritizes blood pressure dynamics over static values.



Saliency mapping reveals a massive spike in model attention in the **0–30 minute window** before prediction

Algorithmic Fairness Audit

Ensure the model does not propagate systemic bias or fail on minority subpopulations. We perform Bootstrapped hypothesis testing with **Bonferroni Correction** ($\alpha = 0.005$) to control for multiple comparisons and prevent false discoveries.

Category	Comparison Group	Reference Group	AUC (Group)	AUC (Ref)	Difference	P-Value	Sig.
Gender	Female	Male	0.916	0.912	+0.004	0.006	
Race	Black	White	0.921	0.909	+0.012	< 0.001	*
	Hispanic	White	0.930	0.909	+0.021	< 0.001	*
	Asian	White	0.920	0.909	+0.011	0.018	
Age	Age 65+	Age <65	0.906	0.919	-0.013	< 0.001	*
Comorbidity	Sepsis	No Sepsis	0.908	0.916	-0.007	< 0.001	*
	Heart Failure	No Heart Failure	0.915	0.913	+0.002	0.080	
	Renal Failure	No Renal Failure	0.915	0.913	+0.002	0.132	
ICU Unit	Surgical ICU (SICU)	Medical ICU (MICU)	0.934	0.916	+0.019	< 0.001	*
	Cardiac Vascular ICU (CVICU)	Medical ICU (MICU)	0.902	0.916	-0.013	< 0.001	*

- **Universal Safety (Gender Parity):** The model achieved strict equity across gender lines, with virtually identical discrimination scores for **Male (0.912)** and **Female (0.916)** patients
- **Minority Performance :** The model demonstrated statistically superior performance for **Hispanic (0.930)** and **Black (0.921)** subpopulations compared to the White baseline (0.909).

Clinical Implications & Impact

Early Intervention

Clinicians can adjust vasopressor titration *before* the crash occurs, smoothing patient hemodynamics.

Resource Allocation

High-risk patients can be flagged for closer nursing attention or 1:1 monitoring ratios during critical windows.

Patient Outcomes

Potential to reduce the duration of hypoperfusion, thereby mitigating risks of kidney injury and reducing ICU length of stay.

Future Directions & Conclusion

Limitations

Our study is retrospective and based on a single centre (BIDMC). External validation across different hospital systems is required to ensure generalisability.

Future Work: Explainability

We aim to incorporate **SHAP** (SHapley Additive exPlanations) to tell doctors *why* the model predicted risk (e.g., "Risk High because Lactate is rising").

Summary

We successfully built an end-to-end pipeline from raw EHR data to a high-performing predictive model.

Final Thought

AI can serve as a "silent partner" in the ICU, monitoring patients 24/7 to support life-saving decisions.

Thank You | Q&A