Early Prediction of Sepsis and ICU Length of Stay

A Critical Analysis using MIMIC-IV Data

Giuseppe Budano
Department of Computer Science and Engineering, University of Bologna
Student ID: 0001117305

The Clinical Motivation: Time is Survival

20% of global deaths are Sepsis-related.

Survival drops 8% for every hour of delay.

The Problem:

- Traditional scores (SIRS/qSOFA) trade sensitivity for specificity, missing early signs.

The Goal:

- Build an automated "Early Warning System" (EWS) to analyze physiological signals immediately upon admission.

Project Objectives

Task A: Risk Stratification

Type: Classification

Question: Is the patient at high risk of developing sepsis?

Goal: Immediate intervention.

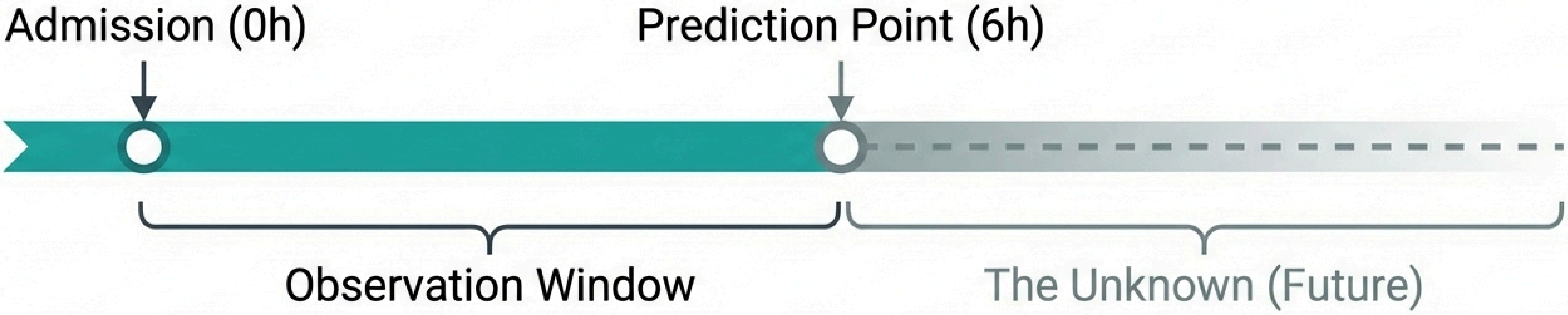
Task B: Resource Planning

Type: Regression

Question: How many days will the the patient stay in the ICU?

Goal: Optimize bed management and allocation.

The Constraint: The "6-Hour Window"



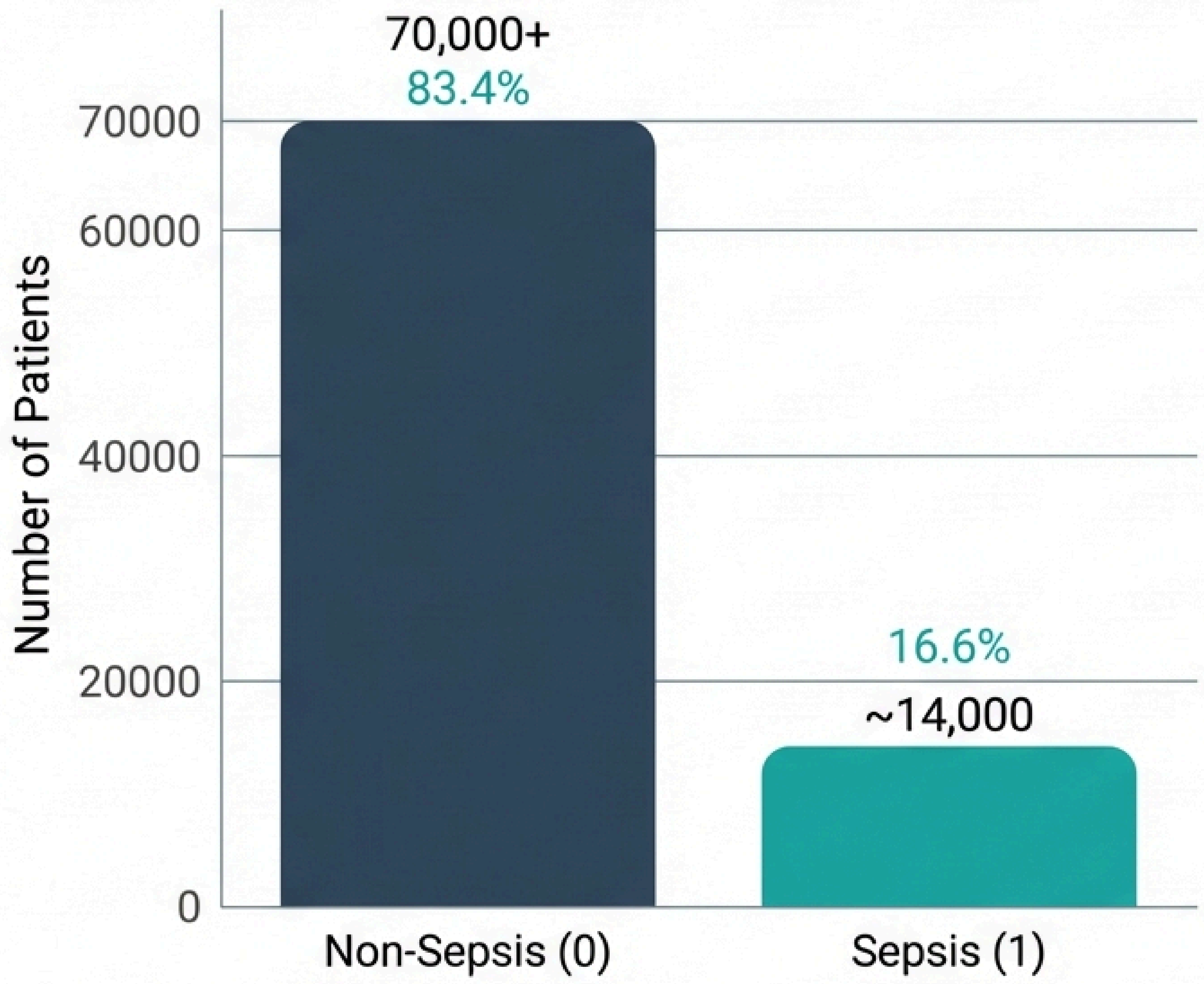
Rationale: Simulates real-time deployment unlike retrospective studies.

Trade-off: Prioritizes operational realism over theoretical maximum accuracy.

The Dataset: MIMIC-IV (v2.2)

Source: Beth Israel Deaconess Medical Center

The Data Challenge: Class Imbalance



- Implication: Models are biased toward the majority class.
- Standard Accuracy metrics become misleading.

Engineering the Pipeline

Temporal Windowing

Filtering events strictly within [t_in, t_in + 6h]

Statistical Aggregation

Flattening time-series to feature vectors.

- Mean, Median, Std Dev, Min, Max

Imputation

Median Imputation for skewed distributions.

Feature Selection Strategy

Physiological Signals

- Heart Rate
- BP (Sys/Dia)
- Respiratory Rate
- Temperature
- SpO2

Lab Markers

- Lactate
- White Blood Cell count (if available in first 6h)

Comorbidities

- Extracted via Charlson Index
- Diabetes, Hypertension, Cancer

Outlier Strategy: Conservative removal. Extreme values are often **signals, not noise**.

Modeling Strategy: Manual vs. Automated

The Baseline: Random Forest

- Ensemble nature robust to overfitting.
- Manual hyperparameter selection.
- Stratified Cross-Validation.

VS

The Challenger: AutoML (FLAML)

- Cost-Frugal Search algorithm.
- Explores **Gradient Boosting** (**XGBoost**, **LightGBM**, **CatBoost**).
- Objective: Maximize **ROC-AUC** (Task A) & Minimize **MAE** (Task B).

Task A Results: Sepsis Prediction

Accuracy > 83% (Misleading due to imbalance)

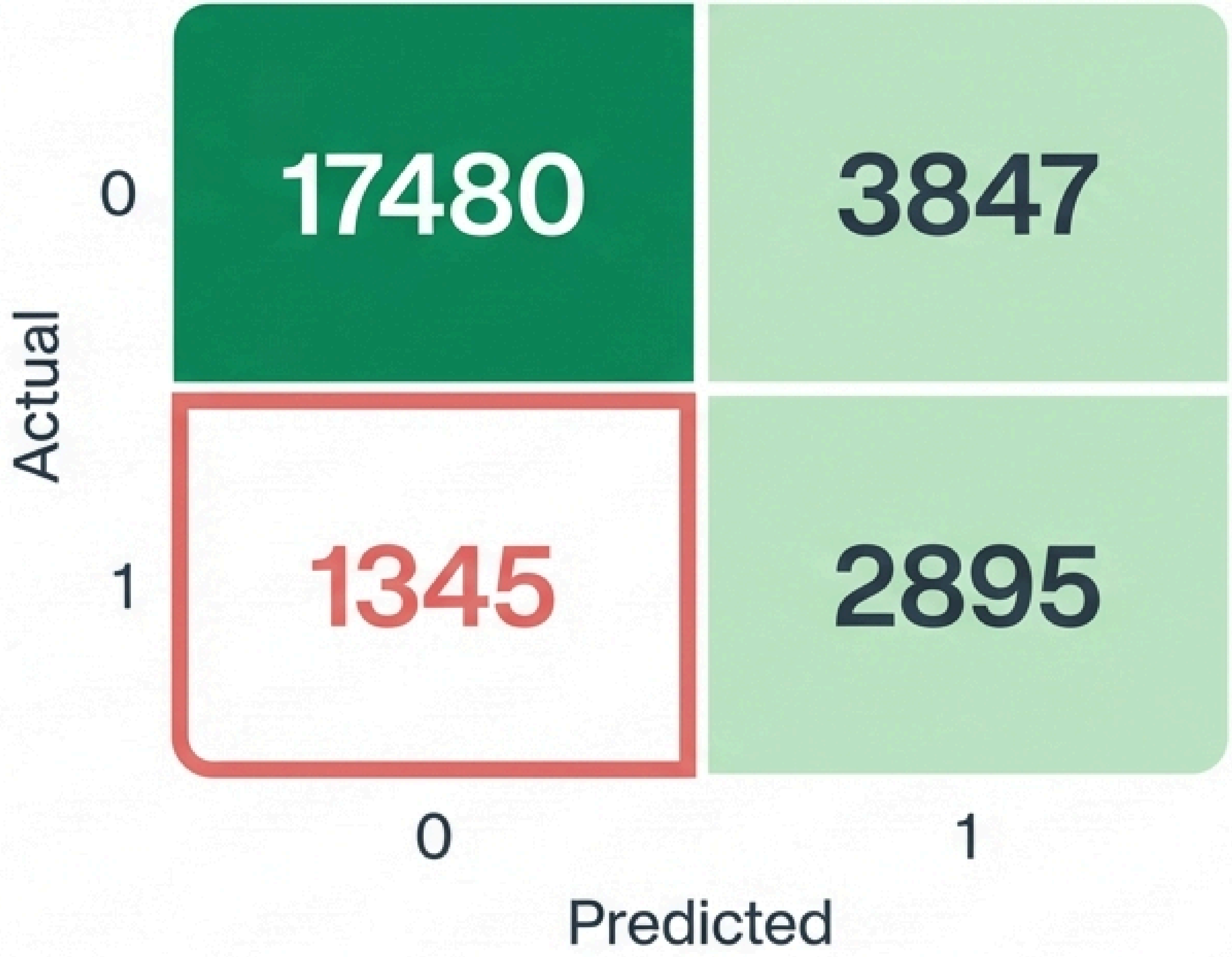
Recall (Sensitivity)

Random Forest: 0.68

AutoML: 0.48

Key Insight: AutoML optimized for global AUC, ‘playing it safe’ to reduce False Positives. Random Forest was better at catching the actual sick patients.

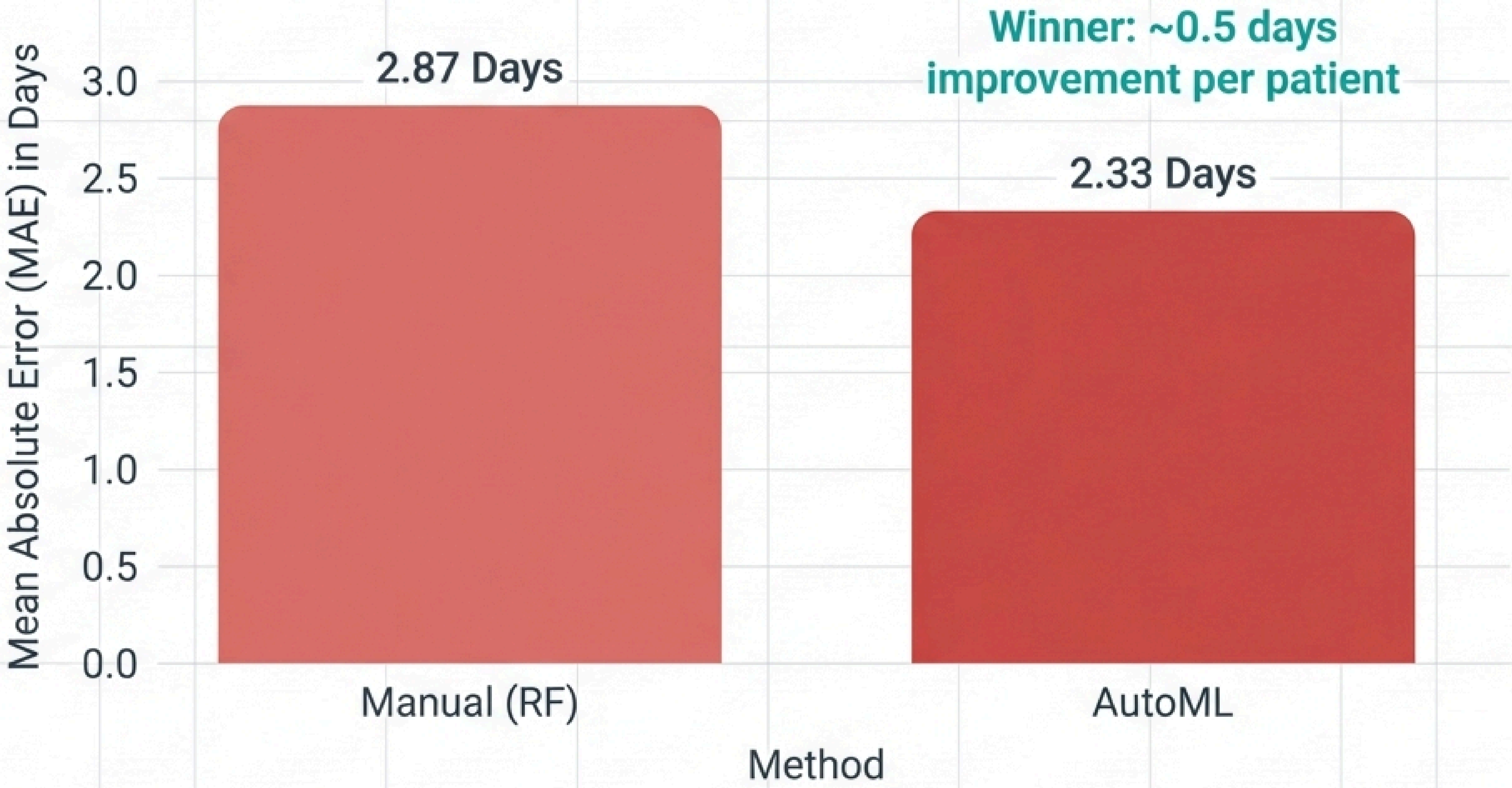
Visualizing Performance: The Confusion Matrix



Focus on False Negatives (1345).

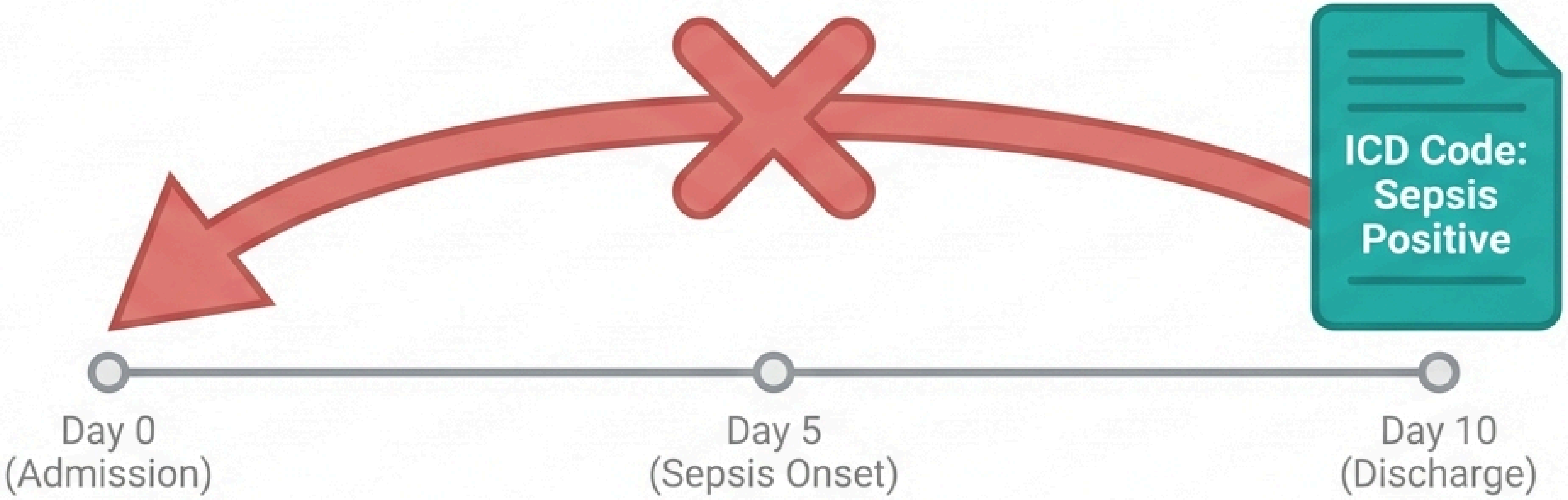
- These are missed diagnoses—the most costly error in medicine.
- Confirms the model is conservative.

Task B Results: Length of Stay



AutoML successfully modeled complex dependencies for continuous variables.

The Critical Flaw: The ICD Trap



- **Issue:** Ground truth relies on Discharge ICD Codes.
- **Result:** The model is trained on ‘future’ knowledge.
- **Data Leakage:** Comorbidities often derived from discharge summaries, not admission history.

Future Directions

1. Dynamic Definitions

- Move to Sepsis-3 criteria.
- Define onset by real-time SOFA score changes, not static codes.

2. Advanced Modeling

- Replace statistical aggregation with Recurrent Neural Networks (RNNs/LSTMs).

Conclusion: From Algorithm to Bedside

Feasibility: Robust baselines established using only the first 6 hours of data.

The Verdict:

AutoML excels at numerical optimization (Length of Stay).

Manual oversight is needed for clinical definitions (Recall).

Final Thought:

These models are ‘Triage Support Systems’, not autonomous diagnostic agents.

Next Step:

Prospective validation required.