Strategic Framework for Predictive Modeling of Clinical Readmission Risk

Leveraging Machine Learning to Shift Diabetic Care from Reactive to Proactive.

FINAL PROJECT
STUDENT: ALFONSO GARCIA ORTIZ
DATE: 30/01/2026

A Data-Driven Strategy to Mitigate Clinical Readmission Risk

Strategic Impact: Implementing an XGBoost predictive model identified 471 high-risk patients in testing, projecting a Net ROI of $273,000 by preventing costly readmissions.

The Challenge

Diabetic readmissions drive penalties and poor health outcomes.

The Engine

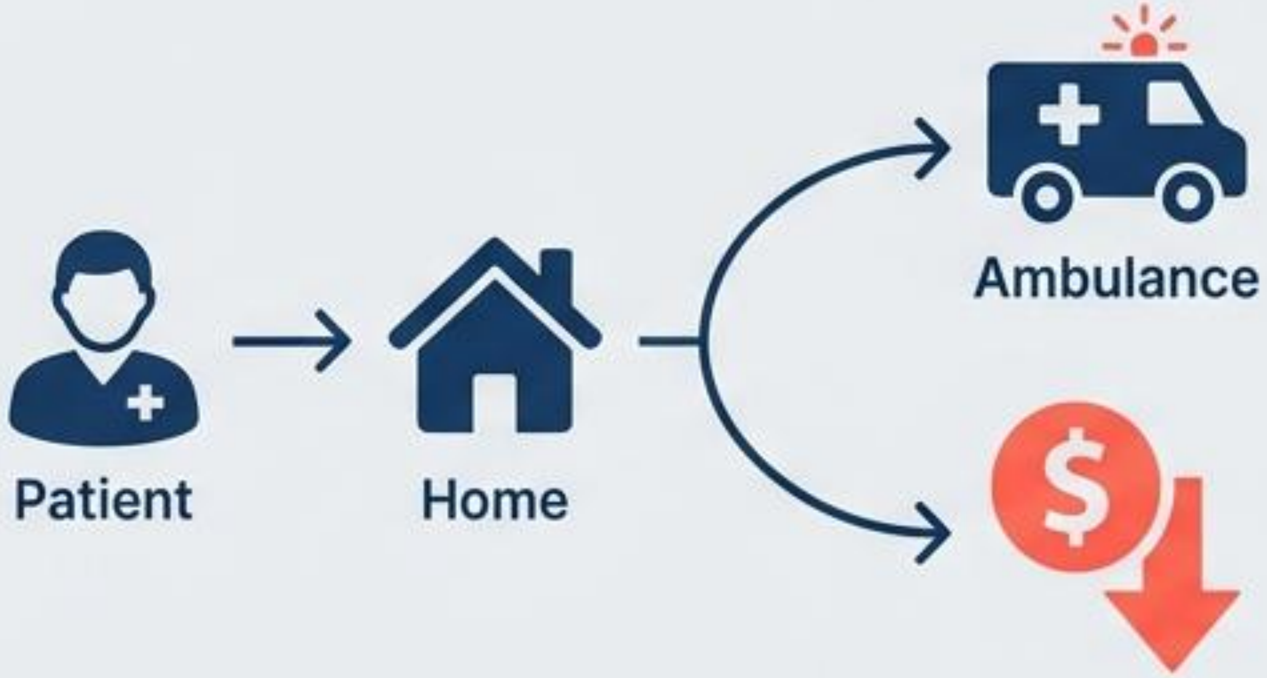
An end-to-end Python pipeline using SMOTE and Gradient Boosting (XGBoost), optimized for Recall to minimize false negatives.

The Output

A deployed Streamlit Web Application providing real-time decision support for clinicians.

Shifting the Paradigm from Reactive Penalties to Proactive Care

Current State: Reactive



Unplanned returns erode margins through HRRP penalties and signal gaps in chronic disease management.

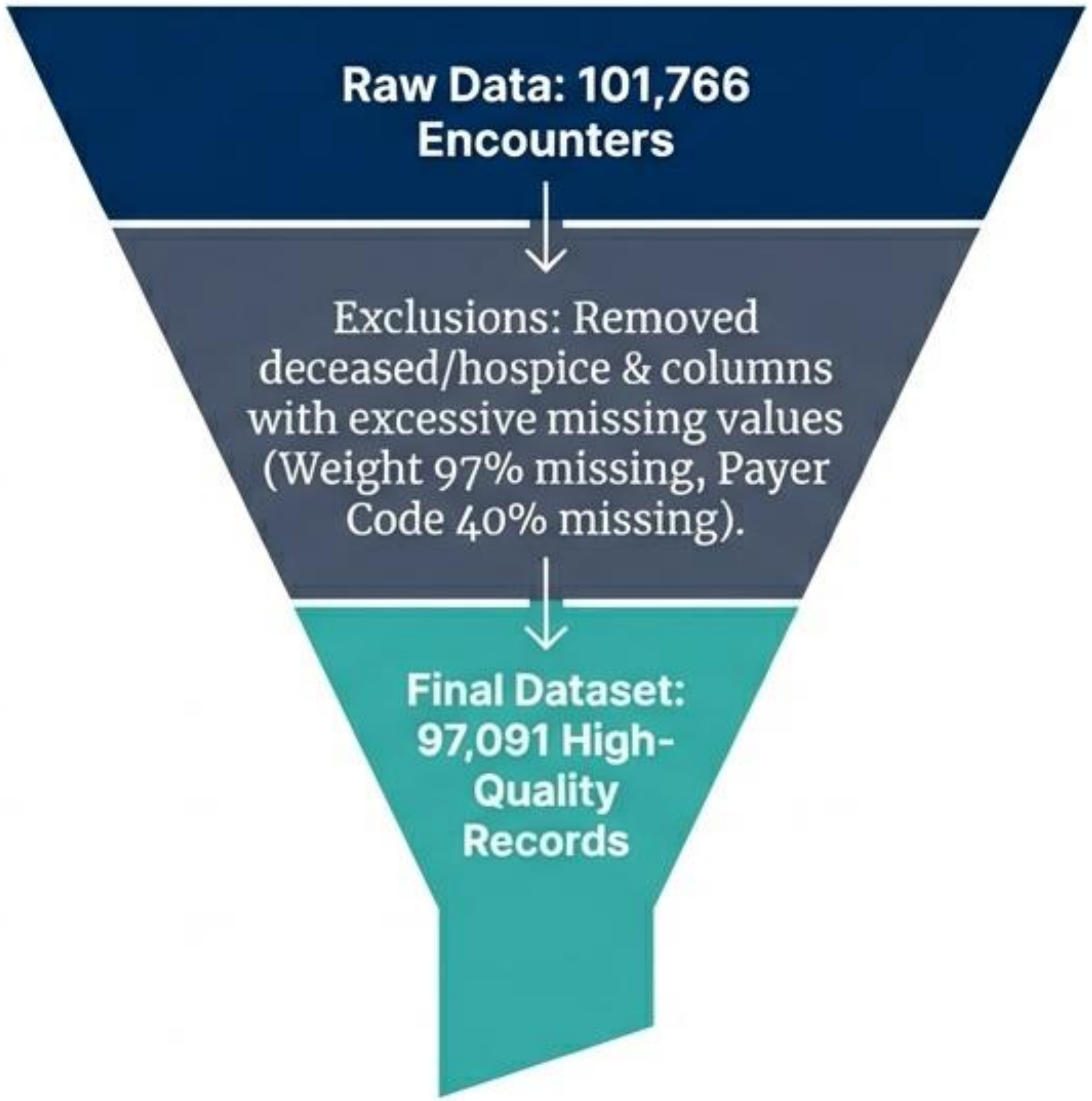
Future State: Proactive

Identifying high-risk profiles before discharge allows for targeted intervention and improved quality of life.

Strategic Goal: Move from treating patients *after* failure to identifying risk *before* discharge.

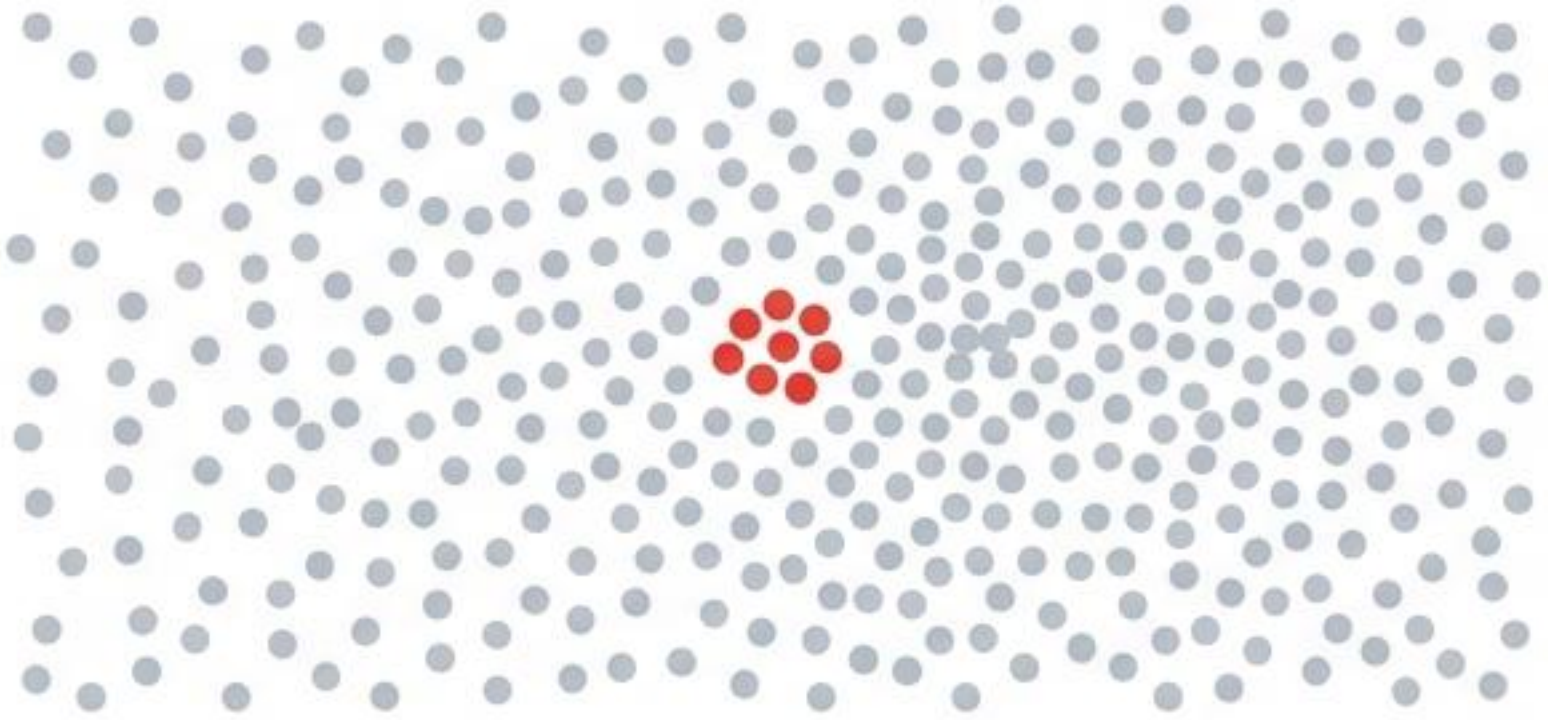
Mining a Decade of Clinical Insights Across 130 Hospitals

Source: Diabetes 130-US hospitals dataset (UCI Machine Learning Repository). Timeframe: 1999–2008.



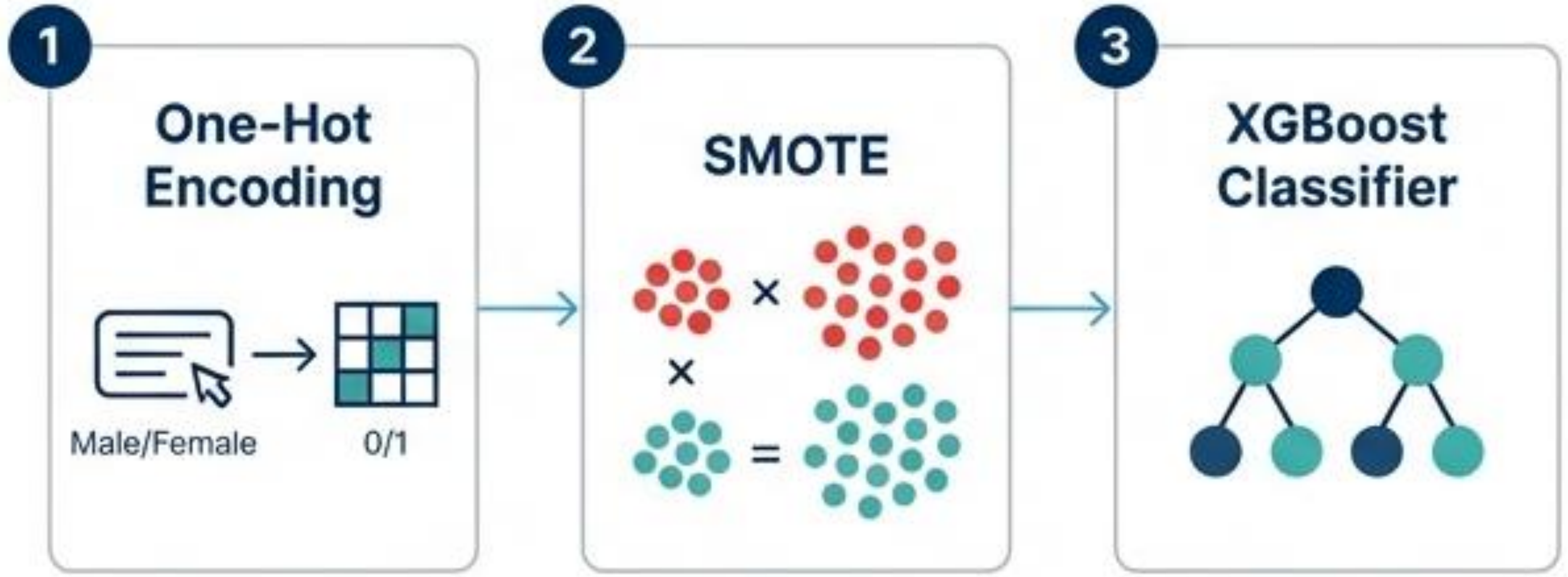
Engineering a Solution for Complex Clinical Data

The Imbalance Challenge



Only ~11% of patients were readmitted. Standard models ignore these “needle in a haystack” cases.

The Technical Solution



Synthetic Minority Over-sampling Technique (SMOTE) taught the model to recognize high-risk patterns, while XGBoost handled non-linear clinical complexity.

Prioritizing Safety: Why We Optimize for Recall

In healthcare, a False Negative (missing a sick patient) is the most dangerous error.

Metric	Baseline (Logistic Regression)	Optimized XGBoost
Recall (Sensitivity)	51%	98.7%
Ability to Detect Risk	Misses 1 in 2 patients	Identifies ~99% of patients

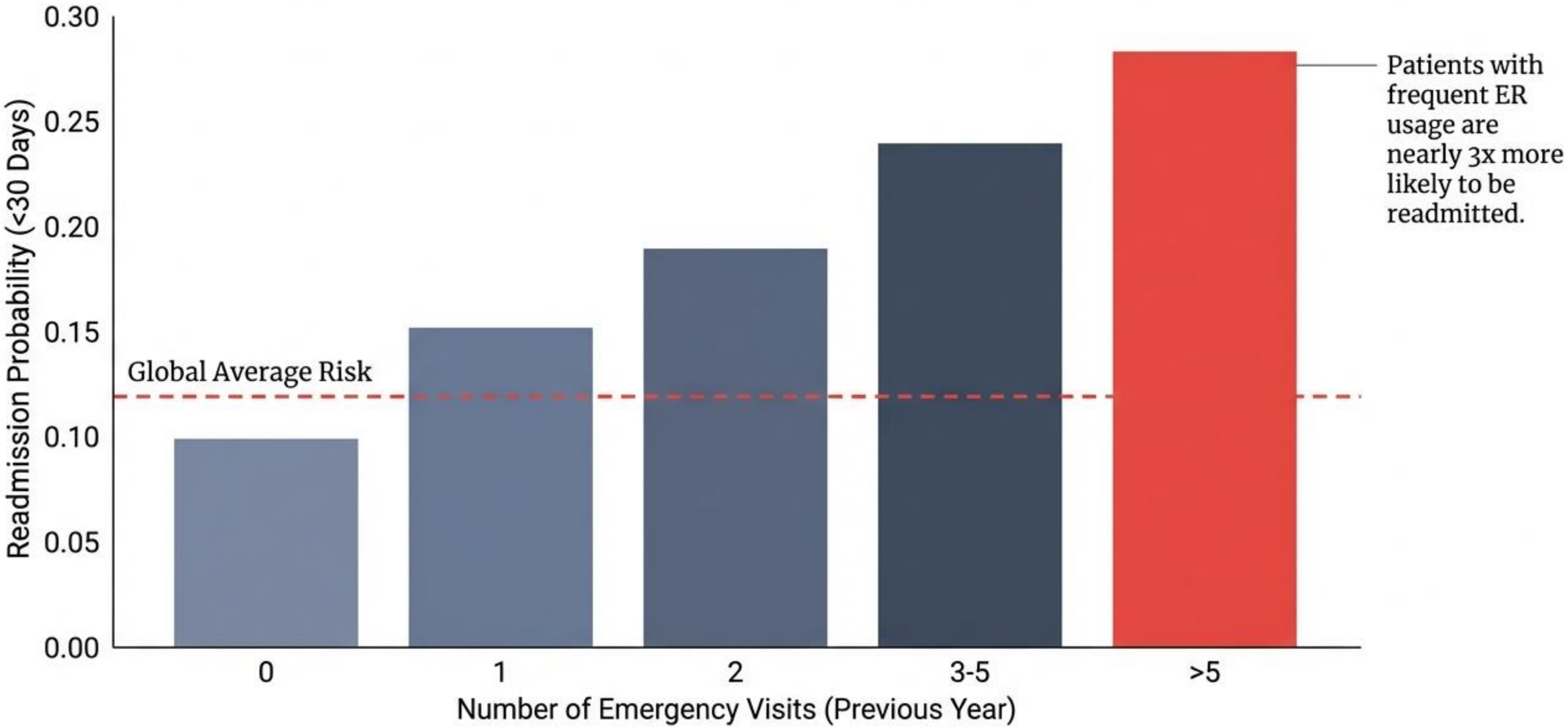
Data Evidence

Test Set Performance (Class 1 - Readmitted)

Precision:	0.12
Recall:	0.99
F1-Score:	0.21

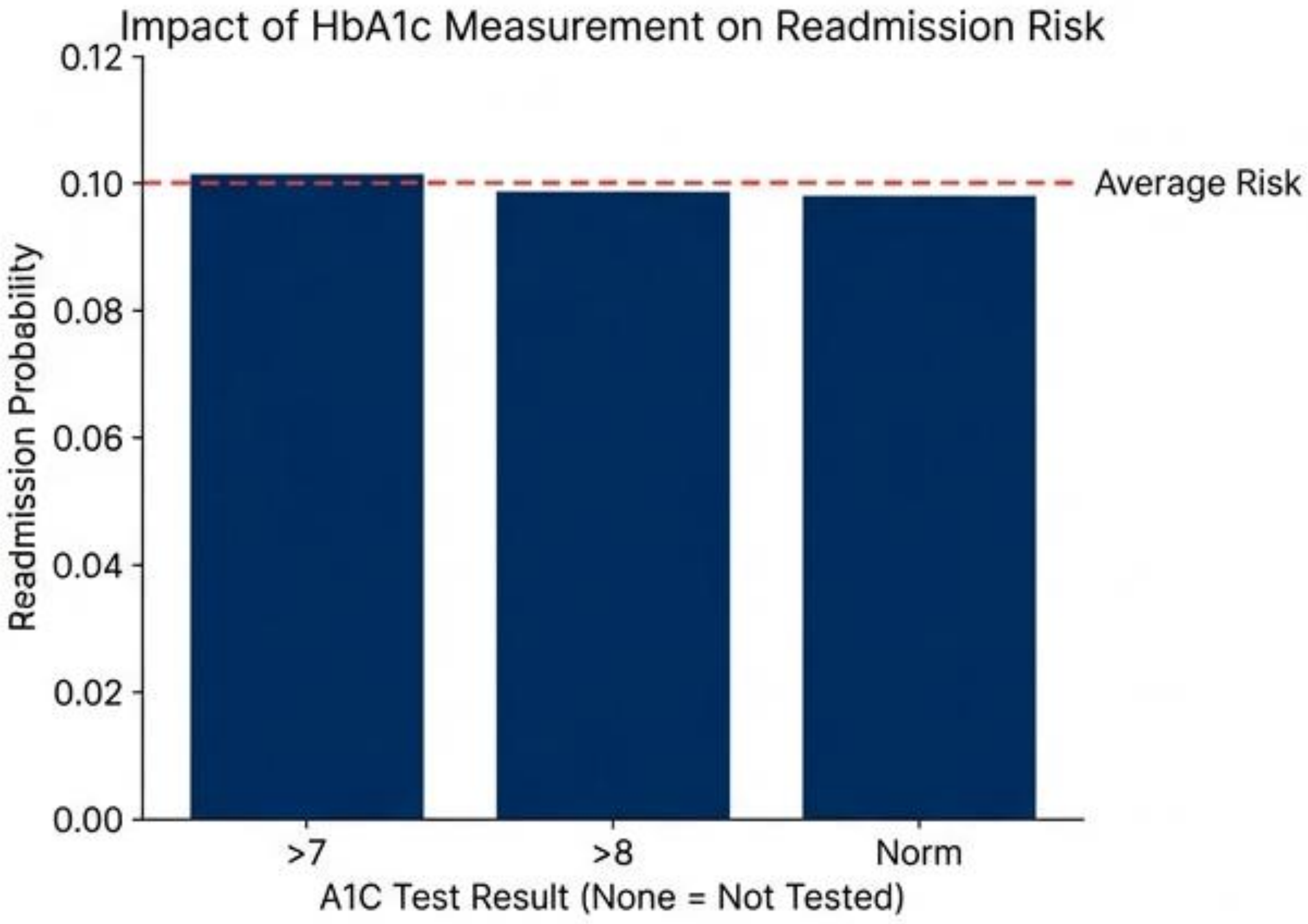
Model tuned to aggressively flag potential risk, accepting lower precision to ensure safety.

Prior Emergency Utilization is the Strongest Indicator of Fragility



The HbA1c Care Gap: Lack of Monitoring Signals Risk

Absence of data is a predictor in itself.

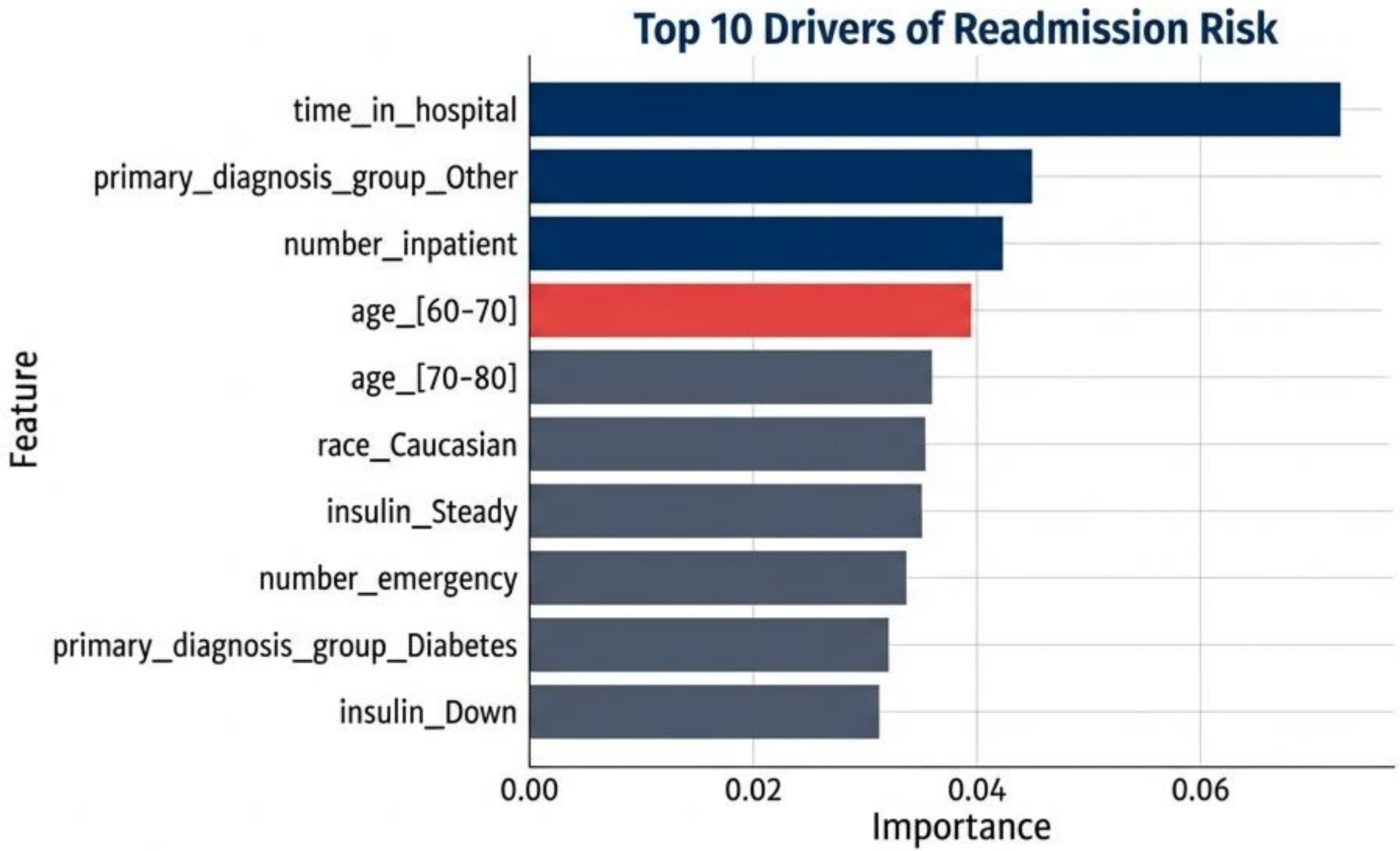


Key Insight

The Missing Variable: While the chart shows that patients with test results (even high ones) have managed risks, the literature and model data indicate that the *absence* of a test (*No Test* Performed) is a **significant risk factor**.

Interpretation: Lack of monitoring during a hospital stay implies a failure in disease management protocols, **elevating readmission probability**.

Anatomy of a Readmission: Key Risk Drivers



1. Severity Proxy: Length of stay (Time in Hospital) is the dominant predictor.

2. Utilization History: Prior inpatient and emergency visits signal fragility.

3. Demographic Focus: Patients in the 60-70 age bracket are specifically vulnerable.

Proactive Intervention Drives Tangible Economic Value

INPUT DATA	ROI CALCULATION FLOW	
Total Test Patients: **19,419**	Gross Savings (Penalties & Care Avoided):	**$508,500**
Patients Flagged for Intervention: **471**	Less Intervention Costs (Staff & Resources):	**- $235,500**
Projected Readmissions Prevented: **33**		

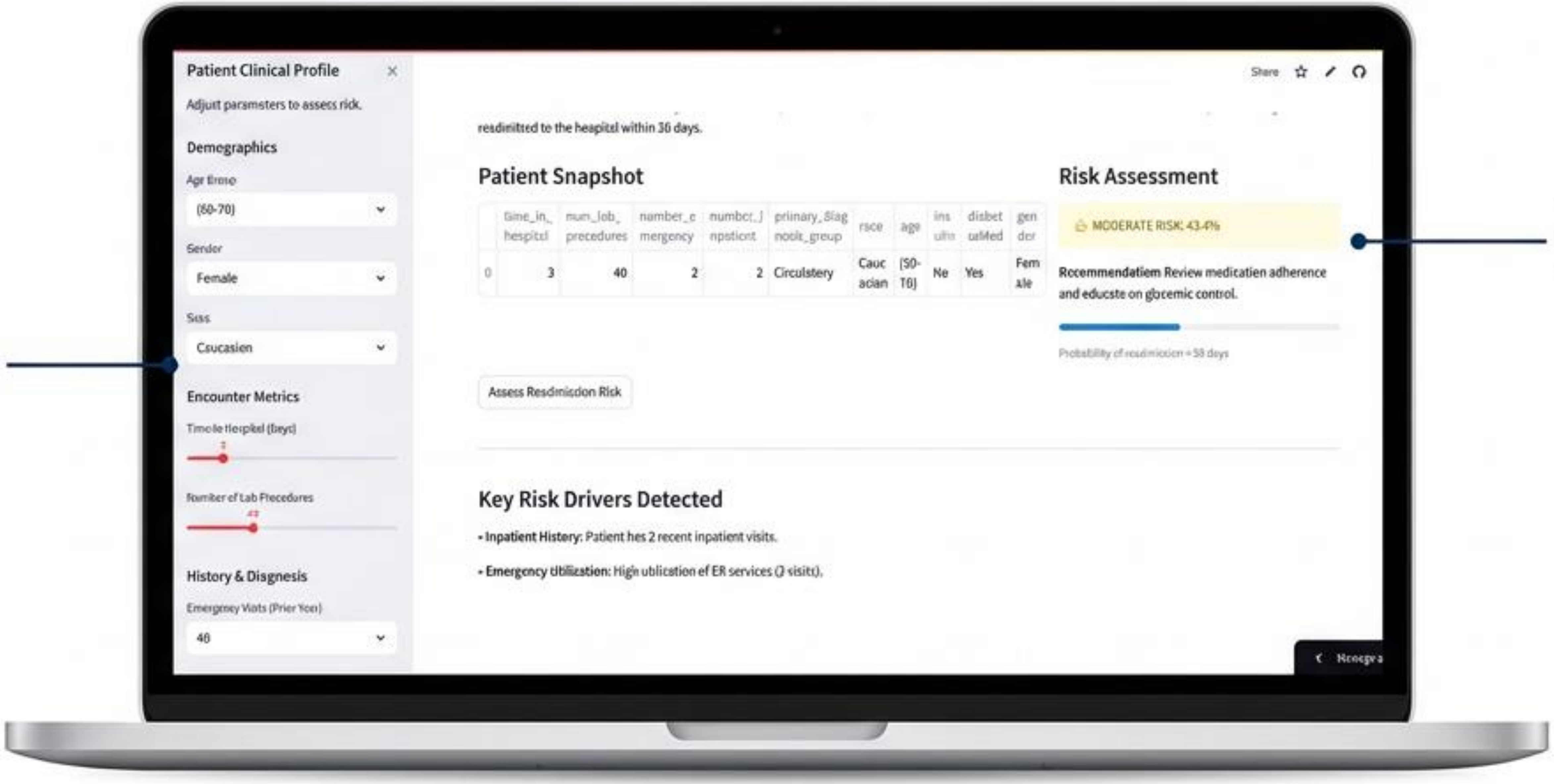
Net ROI (Savings): $273,000

Based on Test Set Analysis (n=19,419).

Empowering Clinicians with the 'Diabetic Readmission Predictor'

Bridging the gap between complex data science and bedside care.

Simple Clinical Inputs



Real-time Risk Scoring

Interpretable AI: Explaining the 'Why' Behind the Risk

Bridging the gap between complex data science and clinical understanding.

Risk Assessment

MODERATE RISK: 43.4%

Recommendation: Review medication adherence and educate on glycemic control.

Key Risk Drivers Detected

- **Inpatient History:** Patient has 2 recent inpatient visits.
- **Emergency Utilization:** High utilization of ER services (2 visits).

- **Transparency:** Moves beyond 'Black Box' AI by clearly stating the contributing factors.
- **Trust:** Clinicians can validate the model's logic against their own medical judgment.
- **Actionability:** Provides immediate, specific recommendations for care teams.

Navigating Data Limitations and Context

Temporal Constraints

Dataset spans 1999–2008.

Clinical workflows and Electronic Health Record (EHR) systems have evolved significantly post-Affordable Care Act (ACA).

Pilot deployment requires validation against modern data streams.

Bias & Fairness

Demographic representation must be monitored to ensure algorithmic fairness.

The model is designed as a Decision Support Tool—clinical judgment remains remains the final authority.

The Path Forward: Enhancing Precision & Scope

Deep Learning

Explore neural networks to capture more subtle, non-linear patterns in patient history.

Social Determinants (SDOH)

Incorporate income, neighborhood, and food access data to understand the 'whole patient' beyond clinical metrics.

EHR Integration

Move from standalone web app to full embedding within the hospital's Electronic Health Record system.

Clinical Authority

A Framework for Smarter Spending and Better Care

Success: Demonstrated 98.7% Recall in identifying high-risk diabetic patients.

Value: Projected $273,000 Net ROI per test cohort.

Tooling: Operationalized Machine Learning via an interpretable Streamlit App.

Status: Ready to move from Theory to Practice.

Recommendation: Approve Pilot Program for Clinical Integration.