Who Comes Back?: Machine Learning for ICU Readmission Prediction

Master of Science in Applied Data Science

Joseph Lattanzi

ICU discharge is not the end of the story. For 20% of patients, it's the beginning of a dangerous and costly cycle. Despite $26 billion in annual readmission costs and penalties for hospitals, predicting which patients will return remains unsolved. We developed a machine learning model to predict 30-day ICU readmissions using the MIMIC-IV dataset (545,316 admissions from 2008-2019). We engineered 57 clinical features spanning comorbidity indices, healthcare utilization, medication risk, and clinical complexity from 150+ million raw data points across six interconnected tables. Using temporal validation (training 2008-2017, testing 2018-2019), we compared logistic regression, random forest, and XGBoost models. XGBoost achieved 0.683 AUC (area under the ROC curve, measuring discrimination ability) on the held-out test set, correctly identifying 7 of 10 readmissions, 49% better than baseline. The model's top predictors were medication count, age, length of stay, and hospital mortality flags, validating multi-dimensional complexity over single-condition flags. Risk stratification showed the highest-risk 33% of patients contained 52.5% of all readmissions, enabling efficient resource allocation. With 50% greater efficiency than random intervention, the model requires screening only 13 patients to prevent one readmission, delivering substantial financial impact: $17.25M annual net benefit for large academic hospitals with 280% ROI. However, critical limitations exist: single-center data limits generalizability, missing post-discharge variables (housing, medication adherence, caregiver support) create a performance ceiling, and fairness analysis revealed an 11 percentage point sensitivity disparity between Black and White patient groups requiring group-specific calibration. Prospective validation and randomized pilot testing with ongoing equity monitoring are required before clinical deployment to ensure real-world effectiveness and fair implementation across all patient populations.

---

Word count: 250 words