Project Proposal: Predicting Hospital Readmissions with Time-Series Deep Learning Models

Chuqiao Lu (cl695) Jiaqi Wu (jw933) Josie Xing (zx142)
Yuchen Zhou (yz946) Xi Chen (xc166)

September 15, 2025

Project Category

Healthcare & Biomedicine

Problem Statement

Hospital readmissions, particularly within 30 days of discharge, are a critical concern in modern healthcare. High readmission rates contribute to increased costs, patient burden, and Medicare penalties for hospitals. Chronic disease patients, such as those with heart failure, diabetes, or COPD, are especially vulnerable to repeated hospitalizations. Traditional statistical models often fail to capture the dynamic, temporal nature of patient health records. This project aims to address this gap by building time-series deep learning models that predict 30-day readmission risk using rich clinical data. The significance lies not only in reducing healthcare costs but also in improving patient outcomes by identifying high-risk individuals who may benefit from additional follow-up care.

Challenges

We anticipate several challenges. First, medical data is often irregularly sampled and incomplete, making time-series construction and imputation non-trivial. Second, the dataset is imbalanced, since most patients are not readmitted, which may bias models toward the majority class. Third, while deep learning models may achieve high predictive accuracy, their complexity introduces challenges for interpretability—a critical factor in healthcare decision-making.

Dataset

The primary dataset will be MIMIC-IV (Medical Information Mart for Intensive Care, version 4), which contains de-identified health records of over 70,000 ICU admissions. We will focus on patients with chronic diseases and extract demographic features (age, gender, comorbidities), static clinical information (length of stay, discharge summary), and dynamic time-series variables (vital signs, lab results, medication logs). Data will be preprocessed into multivariate time-series sequences with imputation strategies applied to missing values.

Method/Algorithm

We will implement a set of baseline models, including logistic regression, random forests, and ARIMA for univariate temporal trends. Deep learning approaches will include LSTM and GRU networks to model sequential dependencies, and Transformer encoders to capture long-term relationships. Existing implementations (e.g., PyTorch, TensorFlow libraries) will be adapted with custom preprocessing pipelines and attention visualization for interpretability. Feature importance will be analyzed using SHAP values, and attention weights will be visualized to provide clinical insight.

Literature Review

Prior studies have shown that traditional regression and machine learning models can predict readmissions with moderate accuracy [1, 2]. More recent research highlights the advantages of deep sequential models, such as LSTMs and Transformers, in capturing temporal dynamics from EHR data [3, 4]. These works suggest that integrating temporal signals with static patient characteristics can significantly improve performance and clinical relevance.

Evaluation

Evaluation will involve both quantitative and qualitative analysis. Quantitative metrics include AUROC, AUPRC, F1-score, and Recall@Top-k to evaluate predictive performance under class imbalance. Qualitative assessments will include SHAP-based feature importance plots and visualizations of attention heatmaps, which reveal clinically interpretable signals. Comparative analysis across baseline and deep models will highlight trade-offs between accuracy and interpretability. Ultimately, the evaluation will not only determine model performance but also uncover actionable insights for clinicians and healthcare policymakers.

References

[1] Zhou, H., Della, P. R., Roberts, P., Goh, L., & Dhaliwal, S. (2016). *Utility of models to predict 28-day or 30-day unplanned hospital readmissions: An updated systematic review.* BMJ Open, 6(6), e011060.

[2] Futoma, J., Morris, J., & Lucas, J. (2017). *A comparison of models for predicting early hospital readmissions.* Journal of Biomedical Informatics, 56, 229–238.

[3] Rajkomar, A., et al. (2018). *Scalable and accurate deep learning with electronic health records.* npj Digital Medicine, 1(1), 18.

[4] Xu, Y., Biswal, S., Li, J., et al. (2020). *RAIM: Recurrent attentive and intensive model of multimodal patient monitoring data.* Proceedings of the AAAI Conference on Artificial Intelligence, 34(1), 1298–1305.