Daniel Akama Nyamweya
Group 13

Explainable AI for Real-Time Clinical Decision Support in Predicting Readmission in Diabetic Patients

Introduction

Hospital readmissions among diabetic patients continue to present a serious clinical and financial burden for healthcare systems worldwide. According to the Medical Centers, readmission rates for chronic conditions such as diabetes contribute significantly to preventable healthcare costs and hospital penalties under the Hospital Readmissions Reduction Program (HRRP). Despite the development of predictive models, their clinical adoption remains limited due to barriers in explainability, real-time usability, and integration with existing Electronic Health Records (EHRs).

This research project seeks to bridge that gap by developing an explainable, real-time clinical decision support (CDS) model that predicts the 30-day readmission risk for diabetic patients. The goal is not only to achieve predictive accuracy but also to enhance model transparency and practical usability in hospital workflows.

Recent developments in the field have shown a shift from black-box AI systems toward explainable and human-centered CDS frameworks. Studies highlight the importance of incorporating interpretability tools such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) to make AI predictions clinically trustworthy (Lundberg & Lee, 2017; Shortliffe & Sepúlveda, 2018). Moreover, the integration of AI models with real-time data streams such as EHRs and hospital information systems has become a major research focus, promoting timely interventions rather than retrospective analysis (Rajkomar et al., 2019).

Literature Review Updates

Recent literature emphasizes that the success of AI-based CDS tools depends not only on accuracy but also on **interpretability, workflow fit, and clinical validation**. For instance, contemporary work by (Chaddad, A., et al. 2023) observed that interpretable machine learning models are better than opaque neural networks in clinician trust and adoption rates when predicting.

Similarly, explainable CDS systems have begun integrating **user-centric design principles**, allowing healthcare staff to visualize patient-specific risk factors in real time. Furthermore, there has been a surge in **hybrid CDS frameworks** that combine AI predictions with clinical guidelines. This ensures that model outputs align with existing hospital decision pathways, improving usability and ethical compliance. These insights strengthen the rationale for this research's focus on real-time, explainable, and easily integrable CDS solutions.

Large studies such as (Strack et al., 2014b) analyzed a large clinical database to assess diabetes care during hospitalization and identify potential improvements. It found that HbA1c measurements were rarely conducted in hospitalized diabetic patients. The analysis, using multivariable regression, showed that the likelihood of early readmission was influenced by whether HbA1c was measured, with variations depending on the primary diagnosis. The findings suggest that better monitoring of HbA1c in hospital settings could improve patient outcomes and reduce healthcare costs.

Methodology

This research follows a **supervised machine learning approach**, structured as follows:

1. **Data Preprocessing:**
    - Clean missing and inconsistent entries.
    - Deal with outliers.
    - Encode categorical variables (e.g., diagnosis codes, medications) using Label-encoding.
    - Normalize numerical features for algorithmic stability.
    - Ensured Patient Level Splitting to prevent data leakage.
    - Feature Engineering.
    - Managing class imbalance.

2. **Exploratory Data Analysis:**
    - Visualize distributions: Use plots to see the distribution of individual variables (e.g., histograms for numerical data).
    - Perform univariate analysis: Analyze each variable on its own to understand its distribution
    - Perform Bi-variate analysis - to look at relationships between predictors in relation to each other and the target.
3. **Model Development:**
    - Train baseline models such as Logistic Regression and Random Forest.
    - Experiment with advanced models like Gradient Boosting (XGBoost) and Neural Networks for performance comparison.
4. **Explainability & CDS Integration:**
    - Apply SHAP values to interpret feature contributions for each prediction.
    - Develop real-time decision support outputs that could integrate into EHR dashboards (e.g., patient risk scores with highlighted contributing factors).(Mienye et al., 2024).
5. **Evaluation Metrics:**
    - Assess models using AUC-ROC, precision-recall, and F1-score.
    - Evaluate interpretability using clinician feedback simulations and SHAP summary plots.

Dataset

Identified Dataset : UCI Diabetes 130-US Hospitals for Years 1999–2008

Over 100k hospital admissions for diabetic patients across 130 hospitals in the United States. It Includes demographic information, admission/discharge details, diagnoses, medications, and readmission status.

Population: Adult patients over 18 years with a primary or secondary diagnosis of diabetes.

Some of the features include:

- Demographics: Age, gender, race.
- Hospital & Admission Info: Admission type (emergency, urgent, elective), discharge disposition, length of stay.
- Medical Diagnoses: Primary and secondary ICD-9 codes.
- Treatments & Medications: Insulin and oral diabetes medication changes (metformin, sulfonylureas, etc.)
- Laboratory Results: HbA1c measurements, glucose serum test results.

- Comorbidities: Hypertension, cardiovascular disease, obesity indicators via ICD-9.

Some of the features we engineered:

- age_group - converts numeric age to categorical risk buckets.
- comorbidity_score - counts the number of diagnoses.
- total_meds - aggregates medications or drug columns.
- prev_readmit - captures patient's readmission history.

These engineered features can later be mapped back to clinical reasoning in your explainable AI phase.

Target Variable: Whether the patient was **readmitted within 30 days, re-admitted after 30 days, or not readmitted**.

Process of validating dataset as 'good data'.

To ensure that the dataset used in this study is of high quality and suitable for AI modeling within a clinical decision support framework, several validation steps were conducted.

1. Source Validation - The dataset originates from a trusted and peer-recognized repository (UCI Machine Learning Repository). It includes real-world clinical information such as demographics, lab tests, diagnoses (ICD codes), medications, and readmission outcomes. The dataset has been used in multiple peer-reviewed studies for developing predictive models, confirming its acceptance and reproducibility in research *(Strack et al., 2014, Journal of Biomedical Informatics).*
2. Data Integrity and Consistency - All patient IDs (patient_nbr) were unique identifiers with multiple hospital encounters correctly linked to the same patient. Date fields (e.g., admission, discharge) showed chronological consistency and no negative time gaps. Outlier inspection and removal were conducted for key numeric variables such as *time_in_hospital*, *num_lab_procedures*, and *num_medications*, ensuring realistic distributions.
3. Data Completeness and Representativeness **-** The dataset contains over numerous records covering a broad age range and multiple comorbidities, ensuring that it captures a diverse population typical of diabetic patients. Class imbalance (readmitted vs. not readmitted) will be analyzed and corrected during preprocessing using SMOTE resampling and class weights to ensure fairness in modeling. Most features (age, lab procedures, comorbidities, medications, admission types, etc.) are clinically relevant predictors of readmission.

4. Data Validation Through Reproducibility - The dataset was compared to values reported in previous publications using the same data source. Descriptive statistics (mean hospital stay, readmission rates, medication counts) aligned closely with published benchmarks, indicating no corruption or modification.
5. Ethical and Privacy Considerations - The data is fully de-identified, containing no PHI (Protected Health Information), aligning with HIPAA compliance standards. All data were obtained from publicly available and ethically cleared sources.

The dataset meets the criteria of “good data” for AI-based clinical decision support. These checks collectively ensure that the data provides a robust, fair, and clinically valid foundation for developing explainable real-time decision support models.

Future Work Plan

The next stages of this project will include:

1. Exploratory Data Analysis: Univariate and Bi-variate analysis.
2. Model Building and evaluation: Fine-tune models using hyperparameter tuning (e.g., grid search) to enhance predictive accuracy.
3. Explainability Evaluation: Implement SHAP visualizations and test their interpretability with a small group of healthcare professionals.
4. Workflow Simulation: Design a prototype dashboard to simulate how real-time readmission alerts could appear in an EHR interface.
5. Reporting and Validation: Compare model outcomes against existing literature benchmarks; prepare a manuscript detailing methods, performance, and usability outcomes.

References

1. Strack B, DeShazo JP, Gennings C, Olmo JL, Ventura S, Cios KJ, Clore JN. Impact of HbA1c measurement on hospital readmission rates: analysis of 70,000 clinical database patient records. Biomed Res Int. 2014;2014:781670. doi: 10.1155/2014/781670. Epub 2014 Apr 3. PMID: 24804245; PMCID: PMC3996476. https://pmc.ncbi.nlm.nih.gov/articles/PMC3996476/
2. Chaddad, A., Peng, J., Xu, J., & Bouridane, A. (2023). Survey of Explainable AI Techniques in Healthcare. Sensors, 23(2), 634. https://doi.org/10.3390/s2302063
3. Lundberg, S., & Lee, S. (2017, May 22). A unified approach to interpreting model predictions. arXiv.org. https://arxiv.org/abs/1705.07874
4. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347–1358. https://doi.org/10.1056/nejmra1814259
5. UCI Machine Learning Repository. (n.d.). https://archive.ics.uci.edu/dataset/296/diabetes+130-us+hospitals+for+years+1999-2008
6. Mienye, I. D., Obaido, G., Jere, N., Mienye, E., Aruleba, K., Emmanuel, I. D., & Ogbuokiri, B. (2024). A survey of explainable artificial intelligence in healthcare: Concepts, applications, and challenges. Informatics in Medicine Unlocked, 51, 101587. https://doi.org/10.1016/j.imu.2024.101587
7. Shortliffe, E. H., & Sepúlveda, M. J. (2018). Clinical decision support in the era of artificial intelligence. JAMA, 320(21), 2199. https://doi.org/10.1001/jama.2018.17163