COMP90089 Machine Learning Applications for Health Group Project Proposal

Lachlan Mcalpine, 1475293 Meghana Ganapa, 1436378
Rongcong Lin, 1520403 Yueyang Wu, 1345511

1 Research Objective

Acute kidney injury (AKI) is a common, severe syndrome in critically ill patients, strongly associated with increased morbidity and mortality (Hu et al., 2022). Early detection and timely intervention can significantly improve outcomes, making AKI well-suited for predictive modeling using electronic health records (EHRs) (Martinze et al., 2020). Moreover, hypertension is the second leading cause of kidney disease (CDC, 2023). Therefore, this study aims to investigate how machine learning models using EHRs can enhance early detection and prevention of AKI in hypertensive ICU patients, leading to better clinical outcomes and more effective patient care.

2 Data source & Phenotyping

Initially we will process the data to identify and build our cohort from patients who match our digital phenotype from the Medical Information Mart for Intensive Care IV (MIMIC-IV) (those that match the clinical definition of hypertension and AKI on UpToDate).

Diagnosing hypertension is often difficult due to the lack of longitudinal measurements (of blood pressure). There is no universal standard that defines the number of abnormal readings required for a positive diagnosis. Therefore, we identify our hypertensive patients with the following criteria (Basile et al., 2024):

1. A Systolic blood pressure $\geq$ 140 mmHg or diastolic blood pressure $\geq$ 90 mmHg,
and
2. Patients who are taking medication for hypertension, such as antihypertensives.

Table 1: Data Elements for finding hypertensive patients

Hospital (hosp) Module	ICU module
1. Admissions table 2. Online Medical Record (OMR) table 3. The Electronic Medicine Administartion Record (eMAR) table 4. Pharmacy table 5. Prescriptions table	1. Dimension table describing itemid (d_items) 2. Inputevents table

To facilitate the modeling process, it is important to identify hypertensive patients who develop AKI. The diagnosis of AKI in hospitalized patients typically involves specific lab tests, with the commonly accepted criteria including (Fatehi & Hsu, 2024):

1. An increase in serum creatinine by ≥ 0.3 mg/dL (27 micromol/L) within 48 hours,

or

2. A decrease in urine volume to < 0.5 mL/kg/hour for more than six hours.

Table 2: Data Elements for finding acute kidney injury in hypertensive patients

Hospital (hosp) Module	ICU module
0. data elements from Table 1 1. Dimension table for labevents (d_labitems) 2. Labevents table	0. data elements from Table 1 1. Outputevents table

3 Methodology

Our goal is to create a machine learning model that can predict the probability of acute kidney damage (AKI) in hypertension patients admitted to the ICU. The specific methods are as follows:

1. **Data Pre-processing:** Cleaning data, managing outliers, handling missing data, normalizing numerical variables for consistency.
2. **Exploratory Data Analysis:** Exploring key variable distributions (such as vital signs, creatinine levels, and medication usage), and identifying potential correlations.
3. **Feature Selection and Engineering:** Extracting AKI risk features, including blood pressure readings, nephrotoxic drug usage, and renal function indicators, and applying PCA to reduce dimensionality. Creating new features like blood pressure change rate to enhance prediction.
4. **Data Splitting:** Separating the dataset into three parts, a training set for model development, a validation set for hyper-parameter optimization, and a set for assessing the model's performance.
5. **Model selection:** Trying algorithms such as Regression (Ridge and Lasso), Support Vector Machine, Stochastic Gradient Descent and Random Forest, and optimizing the model hyper-parameters through grid search or cross-validation.
6. **Visualization:** Visualizing the distribution of key features and the performance of models such as AUC-ROC curves.

3.1 Performance Metrics

We will evaluate the performance of our model using the following metrics:

- **Precision:** Helping the model achieves reliable accuracy in predicting AKI to reduce misjudgment and unnecessary clinical intervention.
- **Recall:** Ensuring that the model effectively identifies most patients who actually develop AKI, thereby reducing the risk of missing those with the disease.
- **F1-score:** Balancing precision and recall to dealing with the imbalanced cases.
- **AUC-ROC:** Evaluating the model's ability to distinguish between AKI and non-AKI patients at different thresholds.

3.2 Expected Outcomes

We expect the following outcomes from this project:

- Identifying AKI in ICU hypertensive patients early with an optimized machine learning model, enabling timely preventive measures and reducing the long-term impact of acute kidney injury.
- Use model predictions to help medical institutions allocate resources more effectively, and enhance overall treatment outcomes.

References

1. Basile, J., Bloch, M, J. (2024). *Overview of hypertension in adults.* UpToDate. https://www.uptodate.com/contents/overview-of-hypertension-in-adults

2. Centers for Disease Control and Prevention. (2024). *Chronic Kidney Disease in the United States, 2023.* https://www.cdc.gov/kidney-disease/media/pdfs/CKD-Factsheet-H.pdf

3. Fatehi, P., Hsu, C. -Y. (2024). *Evaluation of acute kidney injury among hospitalized adult patients.* UpToDate. https://www.uptodate.com/contents/evaluation-of-acute-kidney-injury-among-hospitalized-adult-patients

4. Hu, C., Tan, Q., Zhang, Q., Li, Y., Wang, F., Zou, X., Peng, Z. (2022). Application of Interpretable Machine Learning for Early Prediction of Prognosis in Acute Kidney Injury. *Computational and Structural Biotechnology Journal*, 20, 2861–2870. https://doi.org/10.1016/j.csbj.2022.06.003

5. Martinez, D. A., Levin, S. R., Klein, E. Y., Parikh C. R., Menez, S., Taylor, R. A., Hinson, J. S. (2020). Early Prediction of Acute Kidney Injury in the Emergency Department with Machine-Learning Methods Applied to Electronic Health Record Data. *Annals of Emergency Medicine*, 76(4), 501–514. https://doi.org/10.1016/j.annemergmed.2020.05.026