Medical Temporal Data Analysis

AI Capstone Project

Team members:
111550196 狄豪飛
110550201 周恭麟
111550203 提姆西

National Yang Ming Chiao Tung University

April 24, 2025

Project Topic: Medical Temporal Data Analysis

Task 2: Predicting In-Hospital Mortality in ICU Patients

- Develop a machine learning model to predict in-hospital mortality for ICU patients
- Utilize data from the first 6 hours of ICU admission including vital signs, lab results, diagnoses, and demographics
- Apply advanced ML techniques to achieve high predictive performance

Significance

- Early mortality risk prediction can guide clinical decision-making
- Potential to improve resource allocation and patient outcomes
- Contributes to the development of AI-assisted critical care

Project Workflow

1. **Cohort Selection (TODO 1)**
   - Filter patients with 1 ICU stay (first stay only)
   - Ensure 6 hours of records before discharge
   - Create selection flow diagram
2. **Feature Extraction (TODO 2)**
   - Demographics: Age, Gender
   - BMI (calculation needed)
   - Lab results: BUN, Bilirubin, etc.
   - Vital signs: Heart Rate, BP, etc.
   - Prior diagnoses
3. **Data Preprocessing (TODO 3)**
   - Handle multiple temporal measurements
   - Address missing values and outliers
   - Feature normalization/scaling
4. **Model Development (TODO 4)**
   - Implement ML algorithms
   - Incorporate temporal data
   - Model evaluation and interpretation

Techniques and Dataset

Planned Techniques

- **XGBoost**: Selected based on Hou et al. (2020), who achieved superior performance (AUC 0.857) compared to traditional models for mortality prediction
- **Random Forest**: Chosen based on Iwase et al. (2022), who demonstrated exceptional performance (AUC 0.945) for ICU mortality prediction
- **SMOTE**: Implemented following Gao et al. (2024), who showed significant model robustness improvement by addressing class imbalance
- **SHAP Analysis**: Adopted from Gao et al. (2024), who successfully identified key mortality predictors, providing clinical interpretability
- **Time-window Approach**: Based on Singh et al. (2015), who effectively incorporated temporal measurements for predictive modeling

Dataset

MIMIC-IV database (30% sample): 18,995 patients – 15,347 survivors (81%) and 3,648 non-survivors (19%)

Planned Outcome and Evaluation

Evaluation Metrics:

- Area Under ROC Curve (AUROC)
- Precision, Recall, F1-score
- Confusion matrix analysis
- Feature importance ranking

Success Criteria:

- AUROC bigger than 0.85 (benchmark from literature)
- Clinically interpretable feature importance

Expected Deliverables:

- Cohort selection flow chart
- Feature descriptive analysis
- Documented preprocessing strategies
- Model architecture visualization
- Performance evaluation results
- Complete code implementation
- Temporal data incorporation analysis