Mortality Predictive Analysis on Intensive Care Unit Patients of The MIMIC-IV Dataset

Group 5
Danlan Chen [1288528]
Edoardo de Duro [1497010]
Muhan Guan [1407870]
Lulu Qurotaini [1328810]

COMP90089 - Machine Learning for Health

1 Introduction

Acute Pancreatitis (AP) is a severe inflammation of the pancreas and it is one of the most common causes for hospitalization admission in the USA (Lee & Papachristoum, 2019). Despite the mortality caused by AP having recently decreased, this disease remains a cause of death in ICU departments (1%-5%; Krishna et al., 2017). Particularly, the mortality rate seems to be higher in subgroups characterized by unclear clinical and laboratory risk factors (Vege, 2022). Here, we use MIMIC-IV (Jonson et al, 2020) database to predict mortality and identify the strongest predictors of death due to AP.

2 Data Sources

2.1 Patient Cohort Extraction

To ensure robustness in our predictions, we plan to derive patient cohorts with two distinct extraction schemes and subsequently compare the predictive outcomes.

Cohort 1
Identify the patients with AP using its corresponding ICD code.

Cohort 2
ICU patients classified as AP based on digital phenotypes. They must meet at least two of three criteria (Lee & Papachristoum, 2019):

1. Abdominal Pain
2. Serum Amylase and/or Lipase levels exceeding three times the upper limit of normal (100-300 U/I for Serum Amylase and 50-160 U/I for Lipase)
3. Imaging results consistent with AP. Extraction will be done throughout NLP on free-text reports extracted from mimic-iv-notes module

Duplicates will be removed and cohort 2 will be finalized.

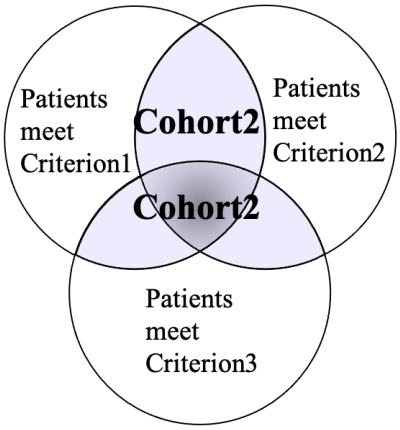

FIGURE 1: Cohort 2 extraction.

2.2 Features Preparation

Drawing inspiration from multiple scoring systems for severity in AP (Vege, 2022), we will categorize our features into two main groups:

1. Clinical predictors (categorical variables)
2. Laboratory predictors (various numerical and categorical variables derived from laboratory measurements)

Additionally, we will incorporate the patients' demographic features, with eventually the predictive variable being their survival status at discharge. For data integrity, we will address missing values based on their distributions and types and eventually exclude patients below 18, resulting in two datasets for our models.

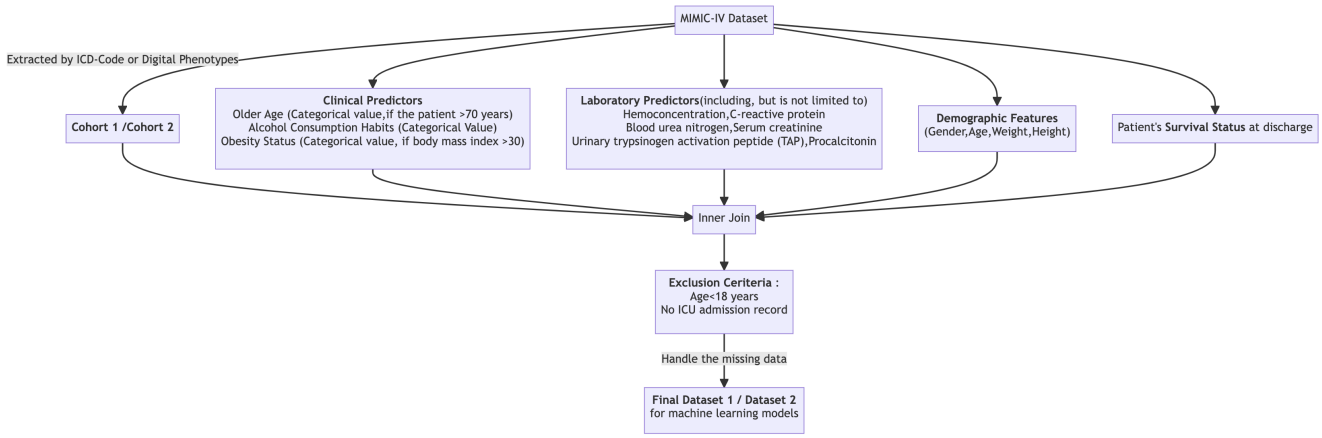

FIGURE 2: Dataset construction from MIMIC-IV database.

3 Methodology

The mortality risk prediction is expressed as **supervised learning** (probabilistic classification), assigning binary label to individual; high mortality rate ($y = 1$) or not ($y = 0$). We will employ Decision Tree model as the baseline and compare it with ensemble learning techniques (XGBoost, Random Forest, CatBoost) and neural network (Deep Neural Network). These models are chosen for (1) ability to predict probability, (2) well performance for particular task, (3) advanced methods to handle bias, variance, and non-linearity, and (4) interpretability, except DNN. Additionally, we will perform **unsupervised learning** (clustering) using K-means clustering and DBScan algorithm to find underlying characteristics among patients.

We will evaluate both datasets by phenotyping and diagnoses by **hold-out validation**, dividing each into two subsets: training and test set (80/20). Prior to training, models' hyperparameters are tuned using **grid-search cross-validation**, dividing training set into training and validation. The training set is used to train the models using appropriate hyperparameters. We will assess the models' performance on test set using MCC (Matthews Correlation Coefficient) and ROC-AUC (Receiver Operating Characteristic-Area Under Curve). Eventually, we compare **ten models**, five models for two datasets (see Figure 3).

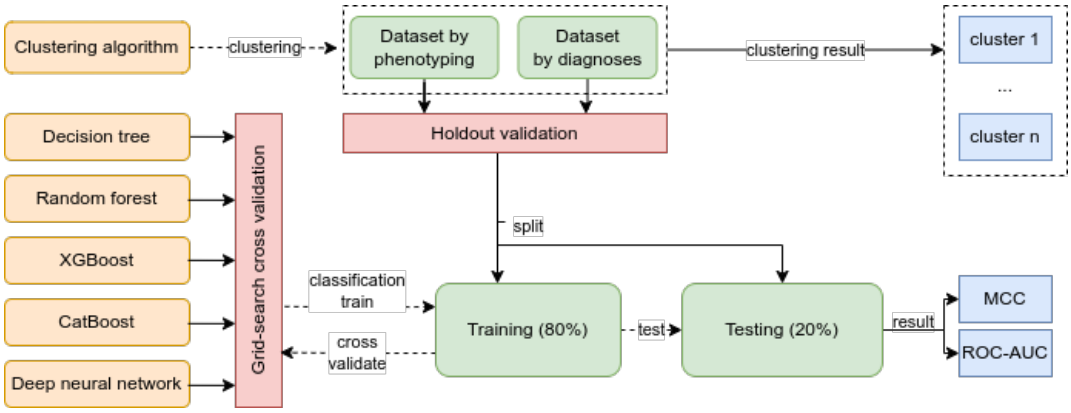

FIGURE 3: Methodology diagram.

We expect advanced models will perform better than the baseline with DNN outperform due to outstanding ability in high-dimensional and complex dataset. Nevertheless, finding the right architecture, such model not overfit, will be challenging. Ensemble learning and baseline models will be more interpretable than DNN. We will extract decision tree produced by models and feature importance from XGBoost and Random Forest. Comparing two datasets, we expect to get better performance on dataset by phenotyping due to less noise (values are already filtered based on phenotyping features). Additionally, we expect to observe at least three clusters in the dataset describing severity of symptoms. But there is possibility that clusters reveal other patterns in the dataset.

References

- Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, L. A., & Mark, R. (2020). Mimic-iv. PhysioNet. Available online at: https://physionet. org/content/mimiciv/1.0/(accessed August 23, 2021).
- Krishna, S. G., Kamboj, A. K., Hart, P. A., Hinton, A., & Conwell, D. L. (2017). The changing epidemiology of acute pancreatitis hospitalizations: a decade of trends and the impact of chronic pancreatitis. Pancreas, 46(4), 482.
- Lee, P. J., & Papachristou, G. I. (2019). New insights into acute pancreatitis. Nature reviews Gastroenterology hepatology, 16(8), 479-496.
- Vege, SS. (2022). Predicting the severity of acute pancreatitis. In S. Grover (Ed.), UpToDate, Retrieved September 14, 2023, from https://www.uptodate.com/contents/predicting-the-severity-of-acute-pancreatitis.