Development of machine learning models to process Electronic Health Records – Explainable Models

Contents

[1. Introduction 3](#_Toc87564197)

[1.1 MIMIC-III dataset 3](#_Toc87564198)

[2. Preprocessing Data 5](#_Toc87564199)

[Cohort Selection 5](#_Toc87564200)

[3. Machine Learning 13](#_Toc87564201)

[4. Deep Learning 14](#_Toc87564202)

# Introduction

Publicly available large scale Electronic Health Records (EHR) with detailed lab tests, hospitalisations and prescriptions are important in research community in order to develop robust clinical decision systems. This data include detailed information on patients history but they are complex and irregularly sampled. In this project, we aim to:

• Use current state-of-the-art machine learning models to train a model to predict risk of mortality

• Develop mechanisms based on global and local interpretability to explain the outcome of the models and identify risk factors

Programming skills: python, sklearn, pytorch

Machine learning: clustering, dimensionality reduction, deep neural networks, inter/intra subject validation protocols, predictive models

--The student needs to complete an ethics course online to be able to download the MIMIC-III EHR data

## MIMIC-III dataset

Use sigmoid at last layer to get probabilities instead

# Preprocessing Data

## Extracting Static features

The static features is the cohert selection of the patients.

Graphical user interface, text

Description automatically generatedpatients\_data

## Extraction of vital data and mapping to variables

A black screen with white text

Description automatically generated with low confidenceEvents\_data

# Machine Learning

## V1 (error with data)

Chart

Description automatically generated

Chart

Description automatically generated

Baseline Model Logistic regression model with Balanced weight

Chart

Description automatically generated

Chart, treemap chart

Description automatically generated

## Version 2.0 (with better data)

numeric\_vars = ['subject\_id', 'hadm\_id', 'hours\_in', 'Diastolic blood pressure',

       'Fraction inspired oxygen', 'Glucose', 'Heart Rate', 'Height',

       'Mean blood pressure', 'Oxygen saturation', 'Respiratory rate',

       'Systolic blood pressure', 'Temperature', 'Weight', 'pH']

## Graphical user interface, application Description automatically generated Version 2.1 (ALL DATA)

numeric\_vars = ['subject\_id', 'hadm\_id', 'hours\_in', 'Diastolic blood pressure',

       'Fraction inspired oxygen', 'Glucose', 'Heart Rate', 'Height',

       'Mean blood pressure', 'Oxygen saturation', 'Respiratory rate',

       'Systolic blood pressure', 'Temperature', 'Weight', 'pH',

       'Diastolic blood pressure1', 'Fraction inspired oxygen1', 'Glucose1',

       'Heart Rate1', 'Height1', 'Mean blood pressure1', 'Oxygen saturation1',

       'Respiratory rate1', 'Systolic blood pressure1', 'Temperature1',

       'Weight1', 'pH1', 'Diastolic blood pressure2',

       'Fraction inspired oxygen2', 'Glucose2', 'Heart Rate2', 'Height2',

       'Mean blood pressure2', 'Oxygen saturation2', 'Respiratory rate2',

       'Systolic blood pressure2', 'Temperature2', 'Weight2', 'pH2',

       "('Glascow coma scale eye opening', 'mean')\_1.0",

       "('Glascow coma scale eye opening', 'mean')\_2.0",

       "('Glascow coma scale eye opening', 'mean')\_3.0",

       "('Glascow coma scale eye opening', 'mean')\_4.0",

       "('Glascow coma scale motor response', 'mean')\_1.0",

       "('Glascow coma scale motor response', 'mean')\_2.0",

       "('Glascow coma scale motor response', 'mean')\_3.0",

       "('Glascow coma scale motor response', 'mean')\_4.0",

       "('Glascow coma scale motor response', 'mean')\_5.0",

       "('Glascow coma scale motor response', 'mean')\_6.0",

       "('Glascow coma scale total', 'mean')\_3.0",

       "('Glascow coma scale total', 'mean')\_4.0",

       "('Glascow coma scale total', 'mean')\_5.0",

       "('Glascow coma scale total', 'mean')\_6.0",

       "('Glascow coma scale total', 'mean')\_7.0",

       "('Glascow coma scale total', 'mean')\_8.0",

       "('Glascow coma scale total', 'mean')\_9.0",

       "('Glascow coma scale total', 'mean')\_10.0",

       "('Glascow coma scale total', 'mean')\_11.0",

       "('Glascow coma scale total', 'mean')\_12.0",

       "('Glascow coma scale total', 'mean')\_13.0",

       "('Glascow coma scale total', 'mean')\_14.0",

       "('Glascow coma scale total', 'mean')\_15.0",

       "('Glascow coma scale verbal response', 'mean')\_1.0",

       "('Glascow coma scale verbal response', 'mean')\_2.0",

       "('Glascow coma scale verbal response', 'mean')\_3.0",

       "('Glascow coma scale verbal response', 'mean')\_4.0",

       "('Glascow coma scale verbal response', 'mean')\_5.0",]

## Graphical user interface, chart Description automatically generated

# Deep Learning

Univariate and multivariate

MICE is a multivariate problem