# Retrospective Prediction on ICU Patients' Mortality and Death Time using MIMIC III
# [A Machine Learning Approach]

Group No.: 103 | Department of CS : Second Year
Project II [PROJ-CS481]
University of Engineering & Management, Kolkata

# CREATED BY,


- Digbijoy Dasgupta [CSE 2nd year, Section : G, Roll no. 21]
- Shreya Bose [CSE 2nd year, Section : G, Roll no. 56]
- Raktim Karmakar [CSE 2nd year, Section : A, Roll no. 41]
- Ishita Pahari [CSE 2nd year, Section : F, Roll no. 25]
- Udayan Misra [CSE 2nd year, Section : H, Roll no. 75]
- Abhishek Sharma [CSE 2nd year, Section : I, Roll no. 01]

## Under the Guidance of,

Prof. (Dr.) Sudipta Basu Pal

# Literature Review

- **Over the past few decades, several ICU scoring systems have been developed using rule-based method / data mining approach**, e.g. APACHE, SAPS, SOFA.

- **Recently, better mortality prediction models trained on larger input features have been developed using machine learning approach.** E.g.
    1. Applying Latent Dirichlet Allocation to free-text hospital notes
    2. Using Recurrent Neural Network to build benchmark models
    3. Applying ensemble methods to improve mortality prediction

- **However, most mortality models in the literature were designed for at least 24 hours or 48 hours after ICU admission.**

- **Could we predict in-hospital mortality in the early stage of ICU, say 6 hours since ICU admission?**

- **Also can we predict the death time for the critical patients?**

# OBJECTIVE

- The goal is to build a two-phase model framework to predict

  **(1) in-hospital mortality and**

  **(2) death hours since ICU admission**

during the early stage of ICU stay, i.e. first 6-hour since ICU admission.

- The model would be useful to promptly identify high-risk patients who might be dead within hours or days since ICU admission, so that resources can be efficiently allocated during the early stage of ICU stay.

# Exploratory Data Analysis

- The original dataset consists of 61,532 distinct ICU stays of 46,520 unique Patients.

- To form our study population, we filtered out patients with
  - ICU stays < one hour
  - age less < 16 or age > 89

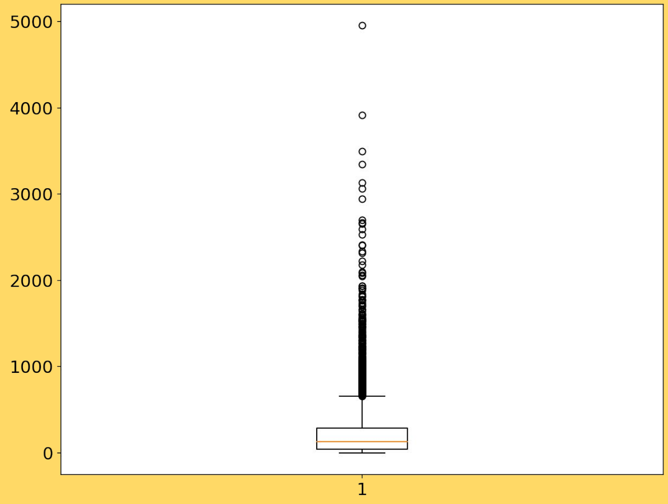
- The final study population consists of 49,632 ICU stays of 36,343 patients.

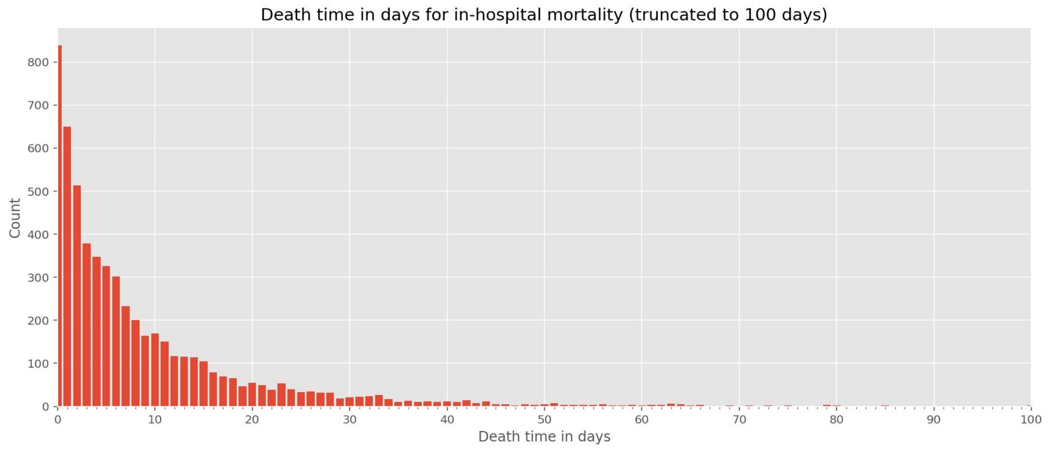
| Variables | Statistics |
|---|---|
| Age | Mean 62.61 |
| Gender | Male 57.29% |
| Ethnicity | White 71.00% |
| Admission Type | Emergency 82.31% |
| Number of ICU Stays | Mean 1.37 |
| In-hospital mortality ratio | 11.62% |

Summary Statistics of the population

# Exploratory Data Analysis (contd.)

● Next, we consider only the dead patients with non-negative death time since ICU admission.
● There are 5,718 in-hospital mortality.
● Their average death time since ICU admission is 9.57 days, maximum death time is 206.38 days and minimum death time is 0 day.



# Machine Learning Models

## Phase 1

- In Phase 1, the study population consists of 49,632 ICU stays. We split the data into **80% training set** and **20% test set**.

- We then built a custom machine learning pipeline to select features, transform data, impute missing values and train a **Random Forest Classifier using grid search on 5-fold CV** to predict the in-hospital mortality label.

- We also test the trained model (best parameter set from grid search) on the test set, and compare the model result using 6-hour, 12-hour and 24-hour data.

## Phase 2

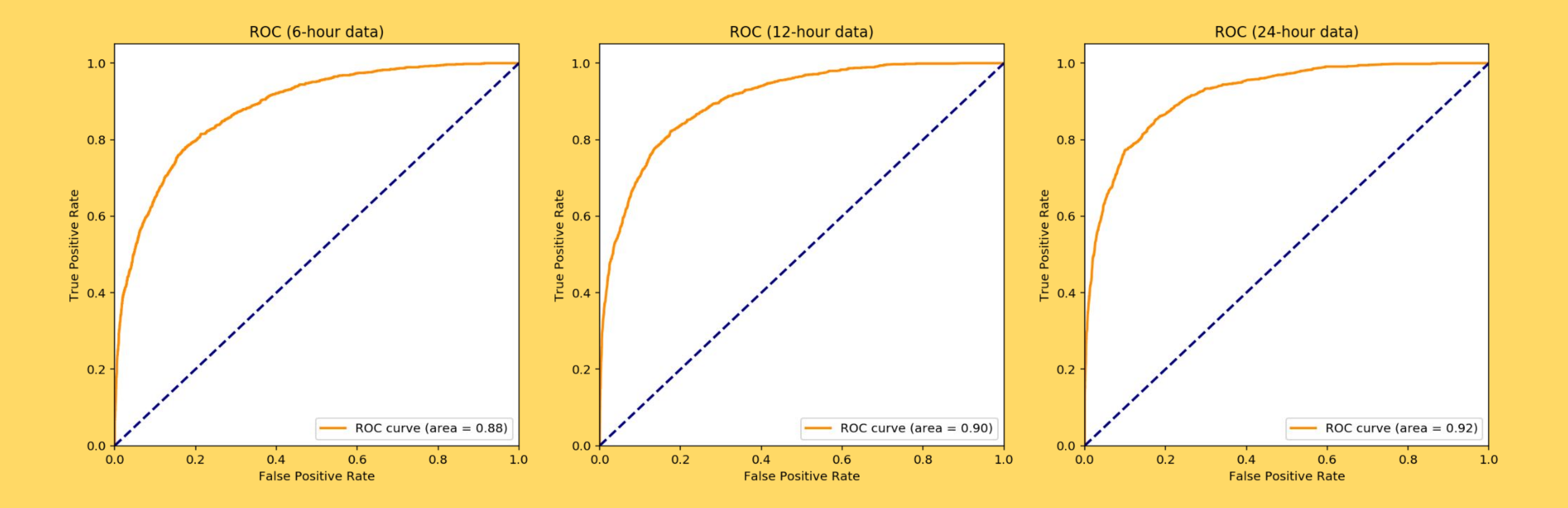
- In Phase 2, we filtered out dead patients with negative death time. A total of 5,718 ICU stays of dead patients were split into **80% training set** and **20% test set**.

- We label each data to one of the three specified classes, then built a custom machine learning pipeline to train a **multiclass Random Forest Classifier using grid search on 5-fold CV** to predict the death time label.

- We also test the trained model (best parameter set from grid search) on the test set, and compare the model result using 6-hour, 12-hour and 24-hour data.

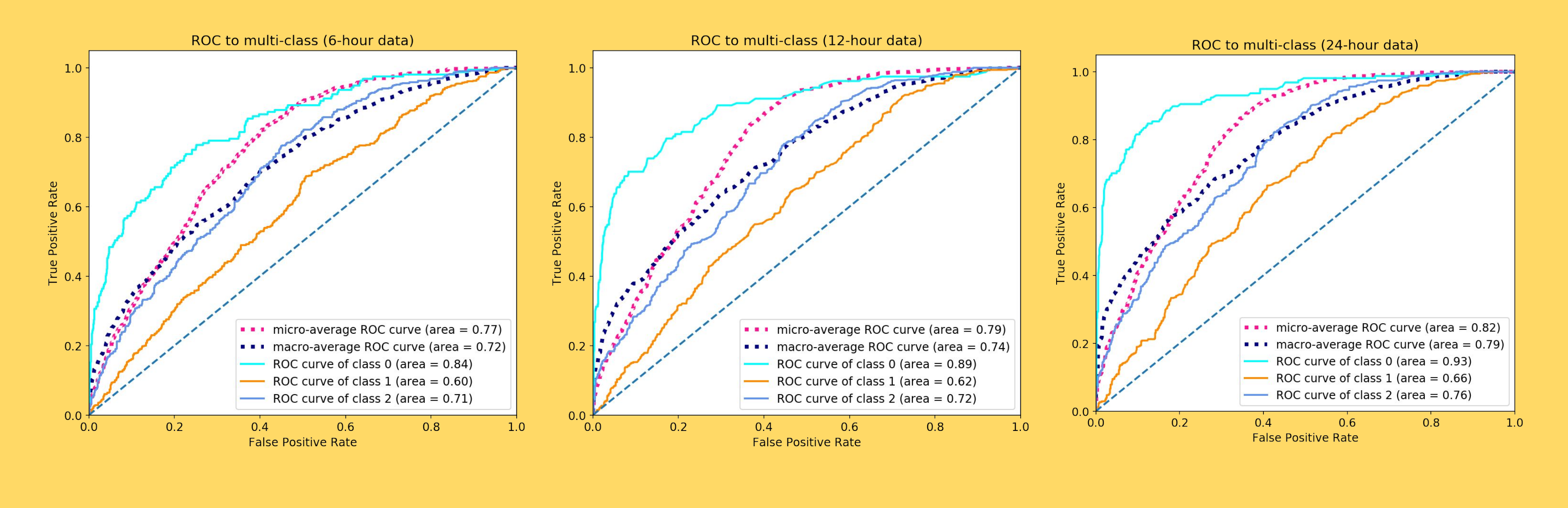
# Result Analysis and Evaluation of Phase 1 Model

The predictive performance of the mortality model trained on 6-hour data has competitive performance (AUROC 0.88) with the same model trained on 12-hour and 24-hour data (AUROC 0.90 and 0.92 respectively).



# Result Analysis and Evaluation of Phase 2 Model

The death time multiclass classifier trained on 6-hour data also provides an effective base (micro-average AUROC 0.77) to give a rough estimate of death hours since ICU admission. The result is competitive to the models trained on 12-hour or 24-hour data (micro-average AUROC 0.79 and 0.82 respectively).



# Conclusion

- Although the models trained on the data in the first 24-hour since ICU admission give better performance, the first 6 hours of ICU data provides enough information for mortality prediction and a rough estimate of death hours since ICU admission.

- The proposed framework provides a base to promptly identify high-risk patients who might be dead within hours or days since ICU admission in the early stage of ICU stay, and there are potential avenues for improvement.

Thank You! Stay Safe!