MORTALITY AND DEATH TIME PREDICTIONS ON ICU DATA

JUN CHEN
SHUAI LIU

CREATING THE NEXT®

OVERVIEW

Georgia Tech

- **Introduction**
- **Objectives**
- **Methods**
  - Database Summary
  - Experimental Setup
- **Results**
  - Model comparisons
  - Model prediction
- **Discussion**
- **Conclusion**

INTRODUCTION (LITERATURE REVIEW)

- In the United States, each year over 30 million patients visit hospitals, 83% of which use an electronic health record (EHR) system.
- Early and accurate identification of patients with high risk of in-hospital death can help physicians in intensive care units (ICUs) make optimal clinical decisions.
- With the accumulation of big data and the development of techniques for data storage, machine learning methods have attracted considerable research attention.
- eXtreme Gradient Boosting (XGBoost) is a machine learning technique with the remarkable features of processing the missing data efficiently and flexibly and assembling weak prediction models to build an accurate one. XGBoost has been widely recognized in a number of machine learning and data mining challenges.
- Another promising big data analytics tools for data science professionals that have emerged with those qualities are Hadoop and Spark.
- Using all those tools, computing resource and models are making the analysis and process for predicting early mortality in ICU much faster and accuracy.

OBJECTIVE

The goal of this study was twofold:

- Firstly, we attempted to compare the performance of machine learning (XGBoost) model with traditional prediction models in the prediction of the 1_day mortality and death time using MIMIC-III.
- Secondly, we planned to use the XGBoost to predict the mortality and death time in the early stage of ICU stay using 1_day ,2_day and 3_day timeframe.

DATABASE SUMMARY

- Medical Information Mart for Intensive Care III database version 1.4 (MIMIC III v1.4)
- MIMIC-III contains data associated with 46,520 distinct patients admitted to critical care units between 2001 and 2012.
- Removing data due to unusual short stays (less than one hour) and only consider adult patients with age between 16 and 89.
- The final study population covered 49,632 ICU stays of 36,343 patients.
- The database contains charted events; documents International (ICD-9) codes; records hourly physiologic data from bedside monitors validated by ICU nurses; and stores written evaluations of radiologic films by specialists covering in the corresponding time period.
- After passing a training course on the website. We were approved to extract data from this database for research purpose

EXPLORATORY DATA ANALYSIS

Table 1: Summary statistics of the study population

Variables	**Statistics** (mean ± std)
Age	62.61 ± 16.93
Gender	Male: 57.8% Female: 42.2%
Ethnicity	White: 70.4% Black: 7.1%
Admission Type	Emergency: 82.31%
Length of ICU stays	1.36 ± 1.06 days
In-hospital mortality ratio	11.6%

CREATING THE NEXT®

DATA EXTRACTION

- For each ICU stay, we have extracted data from the 1_day, 2_day and 3_day since ICU admission.
- Extracted patients' features until they were discharged from the ICU.
- Aggregated the feature values in the specified timeframe (1_day, 2_day and 3_day). There are altogether 123 extracted features, 6 categories.
- The Categories included Demographic and static features, Vital signs, Glasgow coma scale, Blood gases and chemistry values, Lab results and Urine output. In each Categories, there are also many physiological variables.
- All these variables were aggregated and processed by maximum, minimum, and average on the specified timeframe (1_day, 2_day or 3_day).

TWO-PHASE MODEL EVALUATION

- Phase 1 mortality prediction. A binary classifier was trained on the extracted features to predict in-hospital mortality
- Phase 2 death time prediction, we label each data to one of the three classes specified. The distribution of data is following. Class 0 is death in one day. Class 1 is death between 1 to 7 days. Class 2 is death great than 7 days. A multiclass classifier was trained on the same set of extracted features to predict death time

MODEL EVALUATION

Georgia Tech

- The dataset will be split into 80% training set and 20% test set. Hyperparameter tuning was done on 5-fold CV of the training set and the final evaluation of model performance was done on the test set.
- Eight models were trained on the extracted features of the study population for the specified timeframe.
- Use the XGBboost to predict the mortality in the early stage of ICU stay using 1_day ,2_day and 3_day timeframe.
- XGBoost multiclass classifiers were then trained on the training set to predict the death time label using 1_day, 2_day and 3_day data respectively
- AUROC, accuracy, RMSE were also used during the model testing stage to provide a full picture of model performance. And features importance from machine learning will be also reported.

EXPERIMENTAL SETUP

- Three stages
  1. Data summary and processing (ETL process) using Hadoop, Hive, Pig and on Local Docker environment (1 terabyte space, 16 GB RAM, and 8 processors, 6 GB GPU).
  2. Modeling and Cross Validation (Modeling) on a local cluster (1 terabyte space, 16 GB RAM, and 8 processors, 6 GB GPUs).
  3. Prediction mortality and death time using XGBboost on a local cluster (1 terabyte space, 16 GB RAM, and 8 processors, 6 GB GPUs). The decompressed dataset of MIMIC-III requires around 50 GB of space. of implementation:
- Apache Hadoop to perform data preprocessing and feature engineering since the dataset is quite large. Data output in the first stage is then used as feature input for the model training in the second stage. We also used Python and packages such as Pandas and Scikit-learn for efficient model testing, hyperparameter tuning and model evaluation.

RESULTS: MODEL COMPARISONS

Georgia Tech

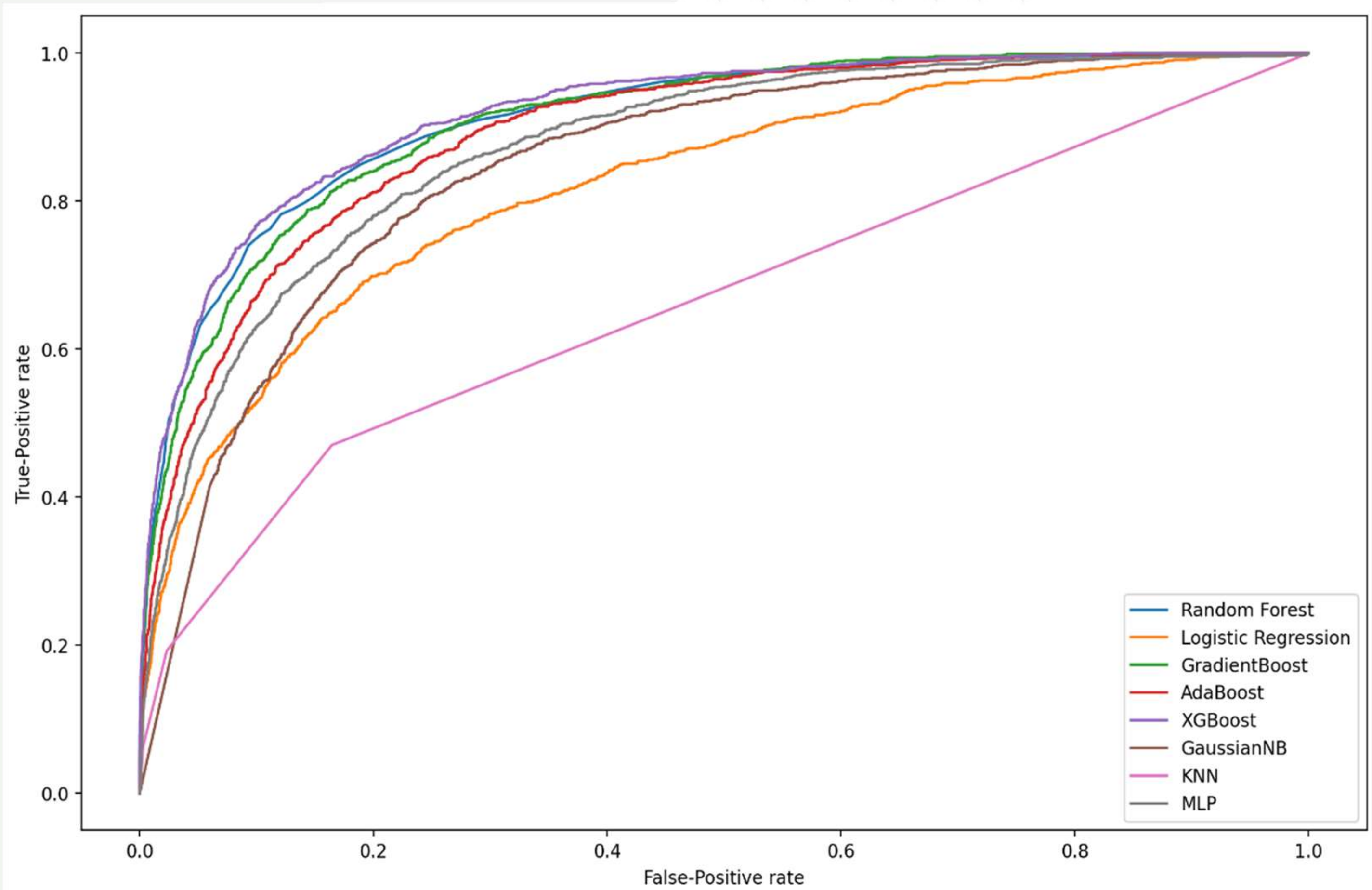


Figure 2. ROC curve graph for 8 models

RESULTS: MODEL COMPARISONS

Georgia Tech

Table 2: Model performance of Roc_Auc_Score, Accuracy and RMSE for phase 1 data.

Methods	Roc_Auc_Score	Accuracy score	RMSE
Random Forest	0.9117	0.9168	0.2885
Logistic Regression	0.8212	0.9011	0.3145
GradientBoost	0.907	0.9165	0.289
AdaBoost	0.8928	0.9095	0.3008
XGBoost	0.9186	0.9226	0.2781
GaussianNB	0.8473	0.8228	0.4209
KNN	0.6638	0.889	0.3332
MLP	0.8733	0.8608	0.3731

RESULTS: MODEL COMPARISONS

Table 3: Model performance of Accuracy and RMSE for phase 2 data.

Methods	Accuracy score	RMSE
Random Forest	0.9444	0.2835
Logistic Regression	0.9386	0.3092
GradientBoost	0.9469	0.2646
AdaBoost	0.9445	0.2801
XGBoost	0.949	0.2553
GaussianNB	0.8064	0.4959
KNN	0.9309	0.341
MLP	0.9416	0.2977

RESULTS: MODEL PREDICTION

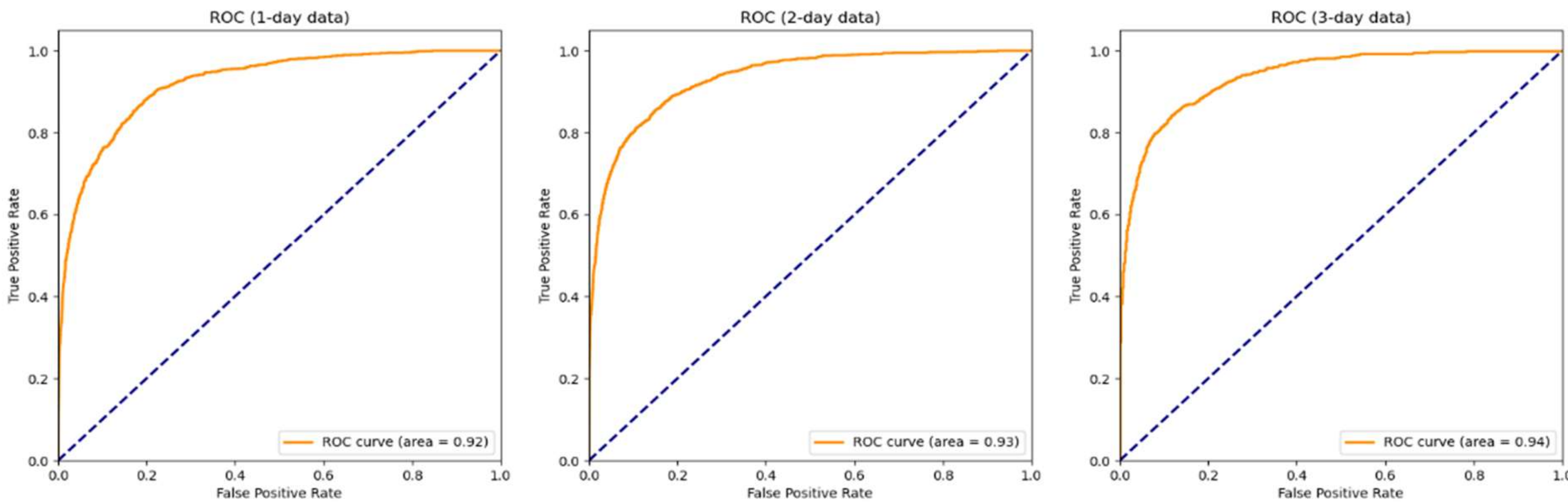
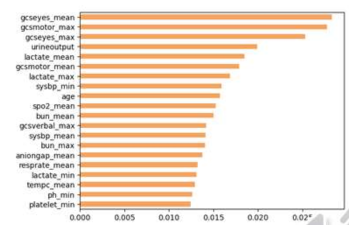


Figure 3. ROC curves of the three classifiers using XGBoost for Phase 1 data.



RESULTS: MODEL PREDICTION

Georgia Tech

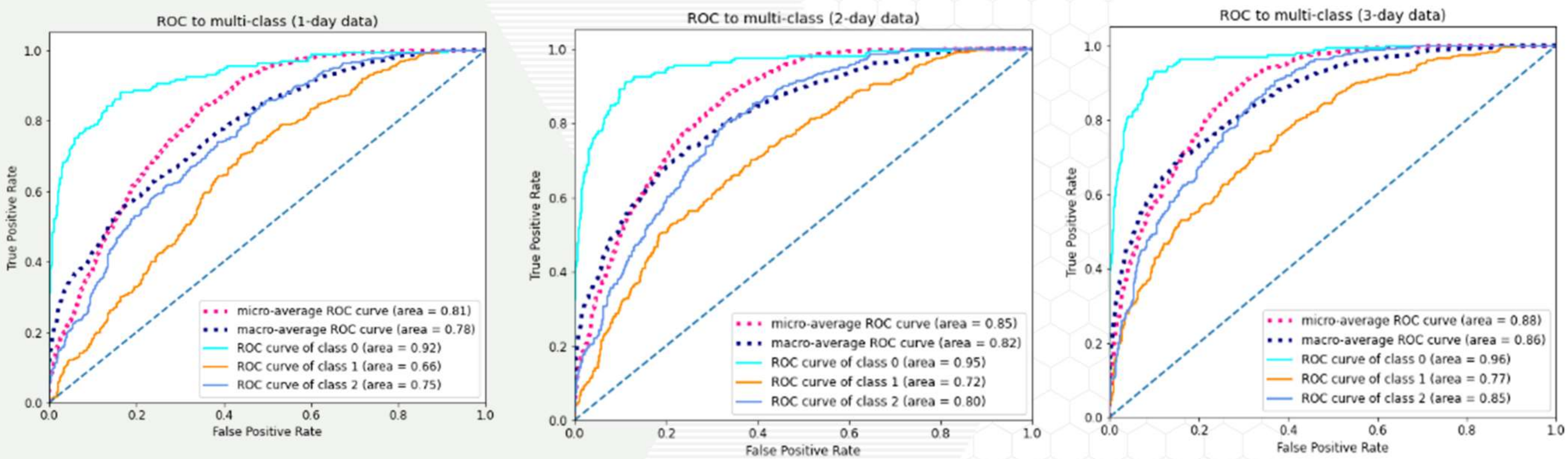
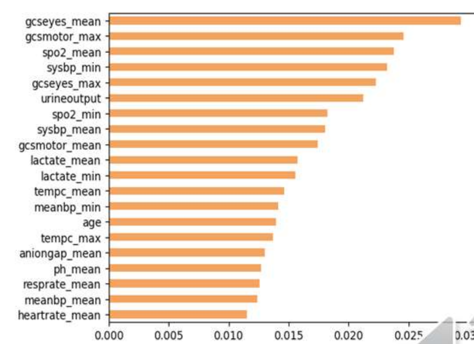


Figure 4. ROC curves of the three classifiers using XGBoost for Phase 2 data.



DISCUSSION

- Our results have showed that XGBoost has outperform the other methods based on all these three criterions. But there are several limits in XGBoost methods. For instance, the features selected were according to clinical experience but not algorithm.
- Some other limitations of our study and it may provide a base for potential improvement.
  - The data come from only one database and the majority of patients were white, potential bias may occur;
  - Further exploration for the database was not performed, which may lead to the abandonment of some key variables;
  - The proposed model was not designed to be validated by developing set from the database or our clinical data.

CONCLUSION

Georgia Tech

We have evaluated eight models to predict the mortality and death time categories. this study demonstrated that the machine learning based on XGboost algorithm does outperform conventional machine learning methods on both phases. Finally, this XGboost model may prove clinically useful and better predict the early mortality in ICU.