

An Exploration of Various Models for Predicting Mortality of ICU Patients

Project Proposal

October 13, 2022

Group Members:

Hannah Damico

Jackson Dial

Caitlyn Nguyen

Github Repository:

<https://github.com/jacksondial/707FinalProject>

I. Problem Description

Intensive care units (ICUs) provide critical medical care for patients with life-threatening conditions.¹ During an ICU visit, many vital measurements are taken to serve as markers for the patient's health.¹ Machine learning methods can particularly be leveraged to automate utilization of the rich ICU data and support data-driven clinical decisions. Acuity scoring systems such as Simplified Acute Physiology Score (SAPS) have been used in combination with classical statistical methods such as logistic regression to predict mortality risk for ICU patients.²⁻⁵ More recently, deep learning methods have been explored as frameworks for mortality prediction using neural networks. The machine learning-based models demonstrated greater prediction accuracy than the linear regression models, but at the expense of interpretability.⁶⁻⁸ Additionally, these methods mainly included only sequential features (e.g., laboratory tests), and not non-sequential features (e.g., patient demographics). An interpretable model based on both logistic regression and recurrent neural networks was implemented to utilize both the sequential features and non-sequential features of ICU encounters.⁹ This joint recurrent neural networks (RNN) and long short-term memory (LSTM) model enabled a degree of interpretability while improving performance from a baseline logistic regression model.⁹ A large fallback of this study is that the data was from only the medical ICU of a single institution; thus, the model may not be generalizable to other ICU categories or different regions.

This study seeks to (1) utilize both demographic and physiological data routinely collected during ICU visits for improved patient-specific predictions and (2) compare the predictive performance of models based on different architectures to determine the best model for predicting mortality of patients of various ICU categories. Our hypothesis is that by combining both demographic data and physiological data routinely collected during ICU visits, patient-specific prediction for mortality will improve. Additionally, we hypothesize that the RNN-LSTM model will be the best model in terms of interpretability and predictive performance. These claims are supported by the observance of a higher accuracy when implementing sequential data into the RNN-LSTM model for predicting medical ICU survival, as previously discussed.⁹

It is important to develop a prediction model that considers all patient information provided from an ICU encounter as a patient-specific clinical decision support tool. There is an abundance of time series data of lab measurements and basic demographic data of patients readily available from ICU visits. This data is historically underused and wasted, but predictive models could utilize this large data to predict mortality and potentially serve as a screening tool.¹⁰ Clinicians could use this tool to support their data-driven decisions on medical treatment and initiate quality of life conversations, impacting the lives of thousands of patients in critical condition.

II. Data Description

Data from this project is sourced from the 2012 PhysioNet/Computing in Cardiology Challenge formulated to develop methods for patient-specific prediction of mortality within the first 48 hours of ICU hospitalization. This data is a subset of the Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II) Clinical Database, version 2.6, which is a collection of

ICU clinical data collected between 2001 to 2008 from a single tertiary teaching hospital.¹¹ Patients involved in the study include adults admitted to including, but not limited to, cardiac, medical, surgical and trauma ICUs. Individuals were excluded if ICU stays were less than 48 hours or patient age was less than 16 years. This dataset is of interest for this study as it is rich in both sequential and non-sequential data and includes encounters of various ICU types.

The data is composed of 12,000 records from ICU stays. There are 42 predictor variables in total, five of which are general non-sequential demographic descriptors collected upon admission, and 37 of which are sequential time-stamped variables measured once, more than once, or not at all in some cases. General descriptor variables include: Age (in years, numeric); Gender (Binary, 0: Female, 1: Male); Height (cm, numeric); ICUType (Factor, 4 nominal levels, 1: Coronary Care Unit, 2: Cardiac Surgery Recovery Unit, 3: Medical ICU, or 4: Surgical ICU); and Weight (kg, numeric). Time series variables are in chronological order with time stamps denoting the time elapsed since admission to ICU in the format of hh:mm. Not all time intervals are the same and range from hourly to daily to irregular. The outcome variable is in-hospital death, a binary variable in which 0 is for survivor and 1 is for died in-hospital. Concerns when working with this data include the use of different tools to measure the same predictor variable resulting in multiple measurements at the same time point. There is also a high degree of missingness for some of the sequential lab measurements, as lab measurements could be measured once, more than once, or not at all in some cases. The final data dimension is 299,264 time checkpoints by 42 predictors.

The raw ICU data was obtained and parsed using Python. An analytic dataset was constructed to be in patient-time format, in which each row of data contains the patient record ID, time of record, and corresponding predictor values. The dataset and parsing scripts can be found at <https://github.com/jacksondial/707FinalProject>.¹²

III. Exploratory Data Analysis

An EDA has been performed and is recorded in the repository. For purposes of this document, minimal conclusions will be shared or discussed. One important aspect of the data is the missingness. The data has a dimension size of 299,264 X 43, with nearly 82% of the cells containing NA values. This is due to not all labs being measured at each timepoint, as discussed earlier. Preprocessing will be done to fill in the missing values using the last observation carried forward (LOCF) method, as that will provide the best estimate for the individual at that time. The distribution of age and correlation matrix of descriptor variables are presented in **Figure 1**. The distribution of age was largely left-skewed, meaning that many patients were of higher ages. This is expected because as age increases, the risk of health complications, hospitalization, and mortality also increases. Additionally, the correlation matrix indicated that the descriptor variables were of loose correlation to one another, except for weight and height being somewhat correlated, as expected.

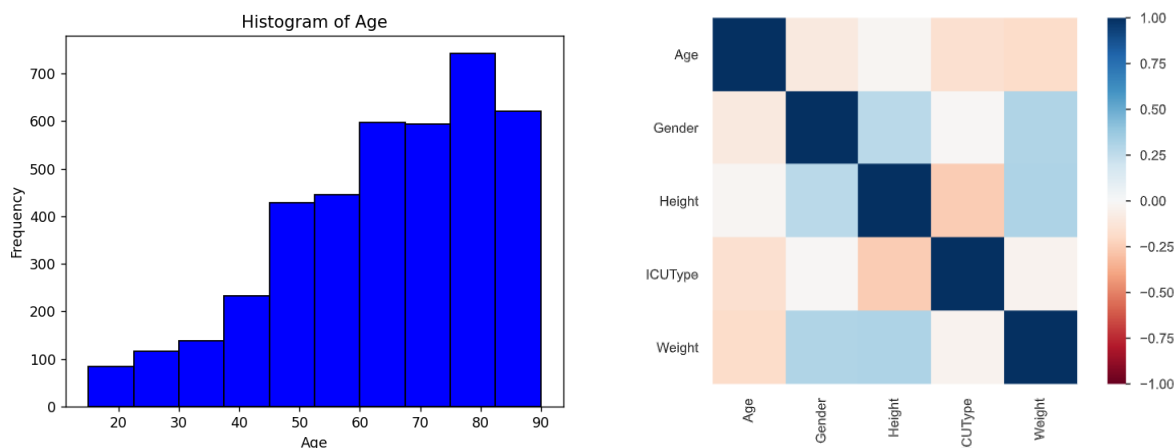


Figure 1. Histogram of age (left) and correlation matrix (right) of descriptor predictors. The left-skewed histogram (left) demonstrated that the population was mainly composed of older individuals. The correlation matrix (right) showed that the descriptor variables were weakly correlated.

IV. Tentative Predictive Model Building Approach

We plan to evaluate the performance of various model architectures for predicting survivorship of ICU patients within their 48-hour ICU period. The health outcome of interest is mortality; thus, predictive models will be framed to classify individuals as non-survivor or survivor. Three different supervised learning methods will be built and evaluated: (1) linear-based models, (2) tree-based approaches, (3) neural networks. Specifically, logistic regression models regularized via L1 penalty (LASSO) and L2 penalty (ridge) will be assessed as the linear-based models. The ensembles of XGBoost and random forest will be used as benchmarks for tree-based approaches. A multi-layer perceptron (MLP) will be constructed to evaluate the performance of a neural network. A joint RNN-LSTM will be used as a neural network which incorporates both interpretability and complexity. This approach is feasible as the different model types will cover a range of generalizability, interpretability, and complexity. There is an increase in complexity and decrease in generalizability and interpretability from linear-based models to tree-based approaches, then neural networks. However, a joint RNN-LSTM may provide a degree of interpretability while also improving accuracy, due to the interpretable weights and complexity of the model. For right now, Hannah will do the linear-based models, Caitlyn will do the tree-based models, and Jackson will do the neural networks. Once the area under the receiver-operator curve is assessed for each model, we will decide on the best architecture and optimize it together.

References

1. Marshall JC, Bosco L, Adhikari NK, et al. What is an intensive care unit? A report of the task force of the World Federation of Societies of Intensive and Critical Care Medicine. *Journal of Critical Care*. 2017;37:270-276. doi:10.1016/j.jcrc.2016.07.015
2. Sharma A, Shukla A, Tiwari R, Mishra A. Mortality Prediction of ICU patients using Machine Learning. In: *Proceedings of the International Conference on Compute and Data Analysis - ICCDA '17*. ACM Press; 2017. Accessed October 13, 2022. <http://dx.doi.org/10.1145/3093241.3093267>
3. Vincent J-L, Moreno R. Clinical review: Scoring systems in the critically ill. *Critical Care*. 2010;14(2):207. doi:10.1186/cc8204
4. Le Gall JR. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA: The Journal of the American Medical Association*. 1993;270(24):2957-2963. doi:10.1001/jama.270.24.2957
5. Moody GB, Mark RG, Goldberger AL. PhysioNet: a research resource for studies of complex physiologic and biomedical signals. In: *Computers in Cardiology 2000. Vol.27 (Cat. 00CH37163)*. IEEE. Accessed October 13, 2022. <http://dx.doi.org/10.1109/cic.2000.898485>
6. Harutyunyan H, Khachatryan H, Kale DC, Ver Steeg G, Galstyan A. Multitask learning and benchmarking with clinical time series data. *Scientific Data*. 2019;6(1). doi:10.1038/s41597-019-0103-9
7. Aczon M, Ledbetter D, Ho L, et al. Dynamic Mortality Risk Predictions in Pediatric Critical Care Using Recurrent Neural Networks. arXiv.org. Published January 23, 2017. <https://arxiv.org/abs/1701.06675>
8. Jo Y, Lee L, Palaskar S. Combining LSTM and Latent Topic Modeling for Mortality Prediction. arXiv.org. Published September 8, 2017. <https://arxiv.org/abs/1709.02842>
9. Ge W, Huh JW, Park YR, Lee JH, Kim YH, Turchin A. An Interpretable ICU Mortality Prediction Model Based on Logistic Regression and Recurrent Neural Networks with LSTM units. *AMIA Annu Symp Proc*. 2018;2018:460-469. Published 2018 Dec 5.
10. Anthony Celi L, Mark RG, Stone DJ, Montgomery RA. "Big Data" in the Intensive Care Unit. Closing the Data Loop. *American Journal of Respiratory and Critical Care Medicine*. 2013;187(11):1157-1160. doi:10.1164/rccm.201212-2311ed
11. Saeed M, Villarroel M, Reisner AT, et al. Multiparameter Intelligent Monitoring in Intensive Care II: A public-access intensive care unit database*. *Critical Care Medicine*. 2011;39(5):952-960. doi:10.1097/ccm.0b013e31820a92c
12. Damico H, Dial J, Nguyen C. Predictive Modeling for Mortality of ICU Patients. GitHub repository. <https://github.com/jacksondial/707FinalProject>