

Prediction of Hospital Mortality Rate for Intensive Care Unit Patients

1 Problem

The objective of this project is to construct a model which can predict a patient's mortality rate in an ICU, given their biological information.

In Intensive Care Units (ICU), patients require continuous monitoring and support. Yet, resources tend to be insufficient, often with not enough personnel and equipment for all the patients. Given this constraint, it is very important to understand what patients need to be prioritized and what level of care is required. One natural way of prioritizing patients is by determining the severity of their illness or physiological health. Logically, patients who are predicted to die should obtain more immediate attention and care. On the other hand, in a very utilitarian (and controversial) sense, patients who are predicted to die during their stay in the ICU should be given lower priority since resources that can be used on other patients are wasted on them. Whatever the case, our duty as data scientists is to generate the prediction of mortality. How our results are used is under the discretion of the hospital. In this project, we will predict a patient's mortality given their biological information.

To predict this we have focused on the first 48 hours of ICU patients and we have used the correspondent label hospital mortality:

- 0: the patient will not die
- 1: the patient will die

2 Related Works

Research on prior machine learning projects related to the ICU yielded mainly studies on developing algorithms to detect the possibility of the onset of sepsis for patients entering the ICU.

One study used 65 features of ICU patients for the model and had reasonable success predicting the possibility of sepsis for patients 4 - 12 hours before clinical recognition¹. Another used a multivariate combination of easily obtainable patient data that included vitals, peripheral capillary oxygen saturation, Glasgow Coma Score, and age and an algorithm that used the 4 stages of data partitioning, feature construction, and classifier training and testing to detect sepsis in 21,000 visits². The last study used a random forest model constructed with over 500 clinical variables and proved its machine learning model to be more effective than Clinical Decision Rules³.

These studies show the feasibility of using clinical data of ICU patients to predict health complications. Simultaneously, they show a paucity in research on predicting other labels beyond sepsis for patients in the ICU, differentiating our project from other related works.

3 Data

We have used data from 12,000 ICU stays of patients who were admitted for a variety of reasons to develop the algorithm. The biological information used involve 42 values that were collected at least once during the first 48 hours of admission into the ICU. These features include 5 general descriptors. The features selected are:

¹Nemati et al., *An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU*.

²Desautels et al., *Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data*.

³Taylor et al., *Prediction of In-hospital Mortality in Emergency Department Patients With Sepsis*.

1. Record ID
2. Age
3. Gender
4. Height
5. ICU type

The remaining 37 values are time series variables observed either once, several times or not at all in some cases. Some examples of these variables are Albumin, creatinine and pH levels.

4 Approach

We first normalized the data with the exception of static variables that have no benefit if normalized. These static variables are RecordID, Age, Gender, Height and ICUType.

We then used a support vector classifier through a 4-way cross-validation to predict the probability of death while being in the ICU.

5 Results and Conclusion

The cross-validation results obtained an accuracy of about 70 - 75% in its predictions. This is a good baseline to show that machine learning algorithms can fairly accurately predict mortality rates in the Intensive Care Unit. Thus, it could assist doctors in prioritising patients to tend to while leaving the actual ordering up to the doctors' own discretion.

With improvements to the algorithm and greater quantity of data along with more features (health signs), we believe the algorithm could potentially be a great aid in helping the healthcare system to more intelligently invest their limited resources.

6 References

- T Desautels, J Calvert, J Hoffman, M Jay, Y Kerem, L Shieh, D Shimabukuro, U Chettipally, Feldman MD, C Barton, DJ Wales, R Das. 2016. "Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach." *JMIR Med Inform* 2016; 4(3):e28.
- Shamim Nemati, Andre Holder, Fereshteh Razmi, Mathew D Stanley, Gary D Clifford, and Timothy G Buchman. 2018. "An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU." *Crit Care Med.* 2018 April 46(4): 547–553.
- R. Andrew Taylor MD, MHS, Joseph R. Pare MD Arjun K. Venkatesh MD, MBA, MHS Hani Mowafi MD, MPH Edward R. Melnick MD, MHS William Fleischman MD M. Kennedy Hall MD, MHS. "Prediction of In-hospital Mortality in Emergency Department Patients With Sepsis: A Local Big Data–Driven, Machine Learning Approach." *Academic Emergency Medicine* 2016; 23: 269– 278