Heinz 95-845: Forecasting Postoperative Mortality after General Surgery Based on MIMIC III Data on Machine Learning for Health Care

Xinmi Li XINMIL@ANDREW.CMU.EDU
Heinz College
Carnegie Mellon University
Pittsburgh, PA, United States

Abstract

This study aims to use machine learning models to predict postoperative mortality for patients in intensive care units (ICU) after general surgery. Multiple machine learning algorithms are trained and tested with the data collected in Multiparameter Intelligent Monitoring in Intensive Care III (MIMIC III) database. Through these models, items considered to be significant in post-surgery care are examined and their contribution to post-surgery death are disccussed in this study. In this study, the performance of each algorithm is also compared with each other, to suggest on the most useful and reliable model that could be used in future postoperative mortality prediction.

1. Introduction

In medical field, the traditional method to analyzing the effect of an intervention is to conduct a randomized clinical trial and analyze the outcomes. Though high levels of recommendation could derive from them, with many limitations such as the size of population involved in the trial and ethical concerns, randomized clinical trials may not be able to be launched or completed, or the conclusions could be not broadly applicable. Besides, the complicated recruiting stage, the compliance issue of participants, and the long period of the trial and follow-ups make the randomized clinical trial unefficient and costly.

Recent advances in machine learning and the transfer from paper health records to eletronic health records have provided huge opportunities to health care research. With large volume of patient data and the high-speed data processing, we are now able to research on the topics that are used to be impposible for randomized clinical trial.

MIMIC III database is a large and freely available database that consists of deidentified health-related data on over 40,000 patients collected from ICU of the Beth Israel Deaconess Medical Center between 2001 and 2012 cit (2016). Some amount of progress on mortality prediction has been made by studying MIMIC III data. One research uses the abundance of ICU data to analyze the relationship between using selective serotonin reuptake inhibitors and the increase of mortality (Ghassemi M and LA, 2014), but it only does statistical tests on the outcome instead of building machine learning models. And another reaserch makes mortality predictions among patients with sepsis and hypotension by using dynamic data during hypotensive episode(Mayaud L and D, 2013). Although it only uses logistic regression to build the model, it has the movelty that using multiple perform measures to compare this model with traditional medical protocols. And another study builds a targeted real-time early warning score for septic shock by using lab data (Katharine E. Henry and Saria, 2015), which outperforms other widely-used protocols.

In this work, we use the MIMIC III data to predict the post-surgery mortality for ICU patients who just had genral surgery. Unlike to other studies using MIMIC data, this work makes the prediction not related to a specific condition. The purpose of it is to have a more general forecast that can be applied to all ICU patients with general surgery, and to provide a guide for nurses to provide better post-surgery care through the examination of indicators of mortality.

In Section 2, we provide background on the postoperative mortality. In section 3, we have a basic elaboration of the models used in this study. Then in section 4, experimental setup is provided to allow for replication of the study. Section 5 gives the results of the model. And more discussions on this study and related work are in section 6.

2. Background

With more and more emphases on heath care quality and patient safety, lowering the post-operative risk is extremely significant. Although the national death rate from complications of medical and surgical care is decreasing over years cit (2012), post-surgery death is still a big issue especially for the acute or unplanned surgery. Therefore, good postoperative care is important. To have a guidance on indicators of postoperative risk of death means not only better patient safety outcomes but also a relief from alert fatigue for nurses.

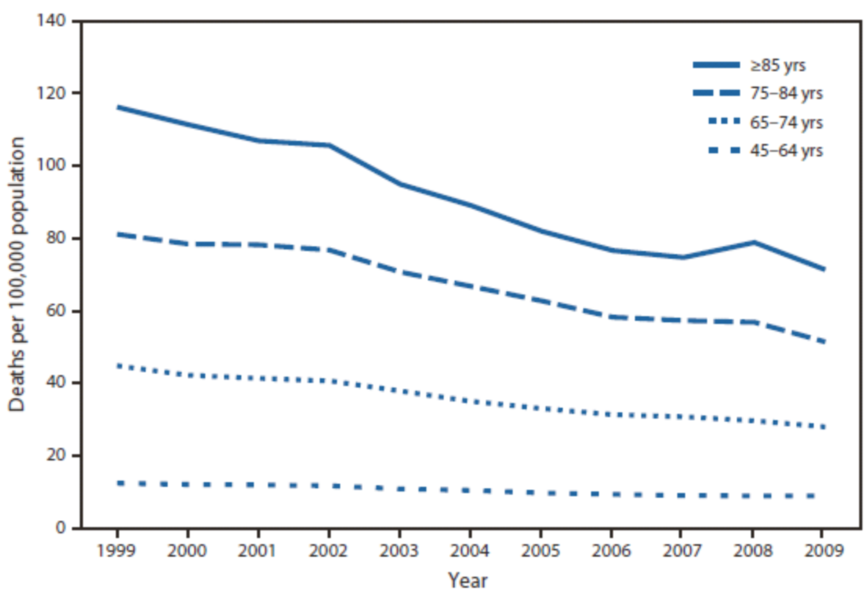


Figure 1: Death Rate From Complications of Medical and Surgical Care Among Adults Aged 45 Years, by Age Group United States, 1999 2009

3. Logistic Regression, Naive Bayes, Tree Augmented Naive Bayes, Decision Tree, and Random Forest Models

In this study, multiple machine learning models are used in R programming with packages: Logistic Regression, Naive Bayes, Tree Augmented Naive Bayes, Decision Tree, and Random Forest.

After training and testing all models, they are evaluated based on the accuracy of classification the outcome, receiver operating characteristic curve (ROC curve), and precision-recall curve (PR curve).

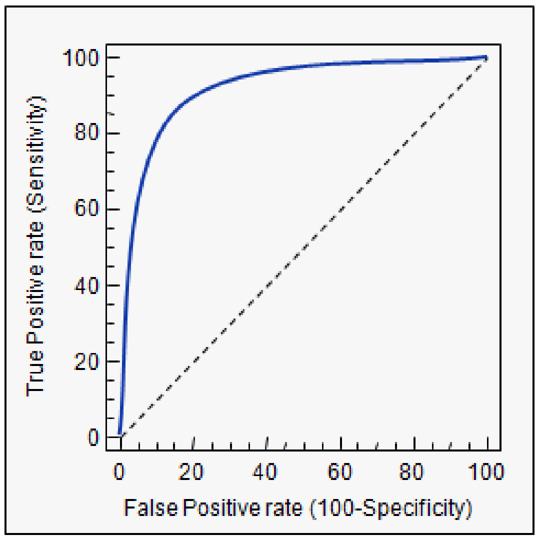
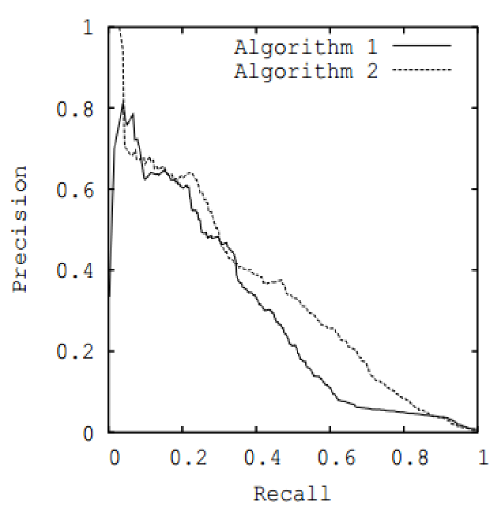


Figure 2: ROC Curve; source: https://www.medcalc.org/manual/roc-curves.php

Figure 3: PR Curve; source: https://www.quora.com/What-is-Precision-Recall-PR-curve



$$Accuracy = \frac{TruePositive + TrueNegative}{TruePositive + FalsePositive + TrueNegative + FalseNegative}$$

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

Code is available at `https://github.com/xml93/MLforHC_FinalProject`.

4. Experimental Setup

4.1 Cohort Selection

The MIMIC III database contains over 40,000 patients' records and related health information. Since this study focuses on the patients who have general surgery and stay in ICU at the time of recording, only patients who have "ICU" as the cost center and "SURG" (which represents general surgery, other specific categories of surgery are marked respectively) as the current service. After selection, only 3580 patients are left.

Gender	Outcome (%)	RR	O2 Saturation(%)	Temp(F)	Systolic BP	HR
Female	Survival: 954 (61.5%)	18.72	89.07	98.33	115.81	87.81
1551	Death: 597 (38.5%)	19.56	90.93	97.86	116.80	87.95
Male	Survival: 1164 (57.4%)	18.31	88.92	98.50	116.80	87.95
2029	Death: 865 (42.6%)	19.59	91.69	98.11	107.46	88.06

4.2 Data Extraction

All datasets are downloaded from MIMIC III database and are in different spreadsheets of topics. Basic sql queries are used to link them. Since each patient has multiple admission records and each is related to general surgery, in order to predict the mortality after surgery, only the most recent admission is kept. And all other patient information is only kept for this most recent admission record. Each patient's admission has an ICD-9 Diagnosis code indicating the diagnosis. They are transformed into 19 groups using the first 3 digitd of ICD-9 code based on the ICD-9 diagnositic groups, making such data more informative.

4.3 Feature Choices

Features used in the model are mainly based on protocols used in the postoperative care. We want to know which medication and what physiology of patient are useful indicators of post-surgery death. Therefore, top 10 medications with highest prescription frequency are selected. They are Lactated Ringers ("LR1000"), Insulin ("INSULIN"), Furosemide ("FURO40I"), Magnesium Sulfate ("MAG2PM"), Sodium Chloride 0.9% Flush ("NACLFLUSH"), Metoprolol ("METO5I"), Depakote 500 mg ("NS500"), 250 cc of 5 % dextrose solution ("D5W250"), Depakote 250 mg ("NS250"), and Depakote 1000 mg ("NS1000"). And the medications are trandsormed into a set of vectors with binary value indicating whether each patient has had such drug. And the patient's physiologies that matter in postoperative care are: respiratory rate ("RR"), oxygen saturation ("SPO2"), temperature ("T"), systolic blood pressure ("BP"), pulse rate, and level of consciousness, according to the National Early Warning Score (NEWS Score) cit (2015). Due to the few record size of pulse rate and the difficulty to categorize consciousness level from the raw data, the last two physiologies are not included in this study. The other 4 physiologies have abundant data in the chartevent table, which means they are recoreded several times within each admission. The mean value of each physiology is calculated as the feature value.

After checking the missing pattern of each colomn, colomns with over 50% missing values and patients with over 5 missing items are dropped. And the remaining missing values are all the averaged physiologies in the previous feature selection step instead of direct records in raw data, so they are missing at random. R package "amelia" is used to impute the missing value for the remaining dataset, with 5 folds. However, after imputation, there are still 61 patients with less than 2 missing values unable to be imputed, so they are dropped.

4.4 Evaluation Criteria and Method Comparison

Since the existing models used to predict the mortality each focuses on a specific disease or condition and the models built in this study are more general and relatively novel in providing general guidance for postoperative care, models in this study are not compared to existing models. Instead, the models themselves are compared against each other to evaluate the performance, by using accuracy, ROC curve, and PR curve.

5. Results

Present the results here. Do not describe how the results were obtained. Those descriptions belong in Section 4.

Typically there are multiple parts and subparts of your study. Use subsections to report the results.

5.1 Results on Application A

Give us some numbers about how well your method works, especially in comparison to some baselines. You should provide a summary of the results in the text, as well as in tables (such as table 5.1) and figures (such as figure 4).

You may use subfigures/wrapfigures (LaTeX packages) so that figures don't have to span the whole page or multiple figures are side by side.

Method	Outcome (%)
Us	20.1
Baseline	18.2

Table 1: Outcome by method used. These are our results.

Figure 4: Example smile graphic.

5.2 Results on Application B

Did more than one experiment type?

6. Discussion and Related Work

This is where you characterize the outcomes of your method and draw conclusions from you experiment. The discussion will build upon the Introduction and the Results sections to synthesize where your contribution brings the field. Discuss any implications of your work. Discuss limitations of your work. Are there situations where you should and should not use your method. What implications are there on policy making, clinical decision making, or future research activities? Remember to contextualize your work with respect to related work and provide references.

7. Conclusion

Summarize your work one more time, this time assuming the reader has read your paper. Build suspense for what your next extension to this method would be.

References

Quickstats: Death rate* from complications of medical and surgical care among adults aged 45 years, by age group — united states, 1999–2009, Sep 2012. URL `https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6137a6.htm`.

National early warning score (news), May 2015. URL `https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news`.

Mimic-iii v1.4 documentation, 2016. URL `https://mimic.physionet.org/about/mimic/`.

Singh N Stone DJ Ghassemi M, Marshall J and Celi LA. Leveraging a critical care database: Selective serotonin reuptake inhibitor use prior to icu admission is associated with increased hospital mortality. *Chest*, 145(4):745–752, Apr 2014.

Peter J. Pronovost Katharine E. Henry, David N. Hager and Suchi Saria. A targeted real-time early warning score (trewscore) for septic shock. *Sci Transl Med*, 7(299), Aug 2015.

Clifford GD Tarassenko L Celi LAG Mayaud L, Lai PS and Annane D. Dynamic data during hypotensive episode improves mortality predictions among patients with sepsis and hypotension. *Critical care medicine*, 41(4):954–962, Apr 2013.

Appendix A.

Some more details about those methods, so we can actually reproduce them.