Predictive Modeling of Critical Care Patient Mortality Rates Using Neural Networks: A Study on US Healthcare Data

Yuta Tsukumo

March 31, 2024

Abstract

This study aims to develop predictive models for mortality rates among critical care patients using neural networks, utilizing large-scale healthcare data from the United States. While acknowledging the limitations of generalizing findings to other healthcare systems, we emphasize the significance of accurately predicting mortality rates for informed resource allocation and quality assessment in intensive care settings.

Introduction

Predicting mortality rates for critically ill patients requiring intensive care is essential for making informed decisions regarding allocation of limited medical staff, medical equipment prioritization, assessing quality of treatment facilities, and appropriately classifying severity of illness for clinical research. However, traditional severity assessment standards in critical care, such as Apache scores, are widely used but have been reported to have low calibration for predicting mortality rates. Therefore, creating a model with higher discrimination and calibration using neural network techniques, based on large-scale patient data from the United States, holds significant importance.

Around the world, deep learning technology is being put to practical use in the medical field, for example, in image diagnosis and intraoperative image recognition systems to support surgeons. On the other hand, healthcare database research lags behind in Japan, making studies utilizing large-scale patient data challenging.

Personally, I aspire to contribute to the integration of patient data and the establishment of large-scale databases as a physician working for the Ministry of Health, Labor and Welfare in Japan after graduation, aiming to advance data science in Japanese healthcare. Through this project, I aim to analyze the advantages and disadvantages of deep learning compared to conventional methods in data analysis and evidence building, while reconfirming the significance and points to keep in mind for data sharing and database building in the Japanese healthcare system.

Related Work

In 2022, a study by Jesse et al. at MIT proposed a novel scoring system using The Global Open Source Severity of Illness Score (GOSSIS) data to predict mortality rates for patients requiring intensive care. While the specific code details were not provided in the paper, the study utilized a logistic regression model based on generalized additive mixed models (GAM) to ensure interpretability and explanatory power. In our project, there's potential to achieve higher discrimination and calibration in predicting mortality rates by leveraging neural networks.

Proposed Work

First, data cleaning will be performed, and missing values will be imputed using either multiple imputation or model-based imputation. Additionally, the dataset comprises 85 columns, and feature engineering will be conducted, utilizing exploratory data analysis (EDA) and medical knowledge, to select or discard variables suspected of multicollinearity. Afterward, model fitting will proceed, comparing the results of discrimination and calibration between the generalized additive mixed model-based logistic regression chosen by Jesse et al. and other models, including neural networks.

Datasets

The dataset consists of 91,714 rows (patients) and 85 columns, with a data size of 31.4 MB. The columns in the dataset include Hospital ID, patient-specific information such as Patient ID, age, gender, race, as well as ICU type. Additionally, Apache scores(APACHE II, APACHE III), blood pressure, heart rate, underlying medical conditions such as diabetes or immunodeficiency, and information regarding in-hospital mortality status are included. After data cleaning, the dataset will be divided into 70% training data and 30%test data. For hyperparameter tuning, the training data will be further divided into five subsets for cross-validation.

Data Cleaning

Of the total 85 columns, the variable 'Unnamed: 83' was removed because it did not contain any values.' After removing 'Unnamed: 83', the missing values accounted for 2.55% of the total data (all cells), while the sample with any missing value accounted for 37.92% of the total sample rows. This result suggests that the use of complete data, in which all samples with missing values are removed, is not desirable in model creation because it results in the loss of a large number of samples. Therefore, the missing values were imputed, since it

was deemed acceptable to consider missing values as occurring randomly, the mean of the continuous variable was used to compensate for the missing values.

For categorical variables for which the class was initially assigned an object, the class was corrected to be a categorical variable. Prior to model creation, rows with missing values for categorical variables were removed, and these data processes resulted in a sample size of 89,488 from the original sample size of 91,714.

Exploratory Data Analysis (EDA)

Next, for each continuous variable, we created a histogram identified by the binary outcome variable, in-hospital mortality, as well as a bar chart identified by the in-hospital mortality variable for each categorical variable. These distributions showed a trend toward older age in the group of in-hospital deaths and a trend toward fewer in-hospital deaths in the group of elective surgeries. The graphs did not clearly show any significant differences between the deaths or survivors with respect to gender, race, or ICU type.

The APACHE II and APACHE III scores have already been given for the present data and will be used as variables when creating the model of in-hospital mortality. Therefore, the evaluation of the relationship between individual variables, such as vital signs and other physiological indicators of the patient on ICU admission, which are included in the calculation of the Apache score are omitted from our model.

The APACHE II and APACHE III scores are theoretically highly interrelated, and incorporating both variables into the model raises concerns about col-linearity issues. However, the scatter evaluation of the relationship showed no clear and strong association between the two variable, and the correlation coefficient was not that high at 0.39. Therefore, we decided to include both variables as predictors in our model.

Evaluation

The multiple models will be evaluated using the test dataset, employing the Area Under the Receiver Operating Characteristic Curve (AUROC) for differentiation evaluation, a calibration plot for assessing calibration, and considering AIC, BIC, and the confusion matrix for additional performance metrics.

Creation Trial Model

Based on the EDA results, variables were selected as follows.

- age

- bmi
- elective_surgery
- ethnicity
- gender
- icu_admit_source
- icu_stay_type
- icu_type
- pre_icu_los_days
- apache_2_diagnosis
- apache_3j_diagnosis
- aids
- cirrhosis
- diabetes_mellitus
- immunosuppression
- leukemia
- lymphoma
- solid_tumor_with_metastasis

$$\begin{aligned}\text{logit}(p) = {} & b_0 + b_1 \times \text{age} + b_2 \times \text{bmi} + b_3 \times \text{elective\_surgery} \\ & + b_4 \times \text{ethnicity} + b_5 \times \text{gender} + b_6 \times \text{icu\_admit\_source} \\ & + b_7 \times \text{icu\_stay\_type} + b_8 \times \text{icu\_type} + b_9 \times \text{pre\_icu\_los\_days} \\ & + b_{10} \times \text{apache\_2\_diagnosis} + b_{11} \times \text{apache\_3j\_diagnosis} \\ & + b_{12} \times \text{aids} + b_{13} \times \text{cirrhosis} + b_{14} \times \text{diabetes\_mellitus} \\ & + b_{15} \times \text{hepatic\_failure} + b_{16} \times \text{immunosuppression} \\ & + b_{17} \times \text{leukemia} + b_{18} \times \text{lymphoma} + b_{19} \times \text{solid\_tumor\_with\_metastasis}\end{aligned} \tag{1}$$

Result

Data were evaluated by dividing 80% of the data into training data and 20% into test data.The results of this simple logistic regression are as follows.

```
Accuracy: 0.9131548819713242
Confusion Matrix:
 [[16744    12]
 [ 1581     6]]
Classification Report:
              precision    recall  f1-score   support

           0       0.91      1.00      0.95     16756
           1       0.33      0.00      0.01      1587

    accuracy                           0.91     18343
   macro avg       0.62      0.50      0.48     18343
weighted avg       0.86      0.91      0.87     18343
```

Figure 1:

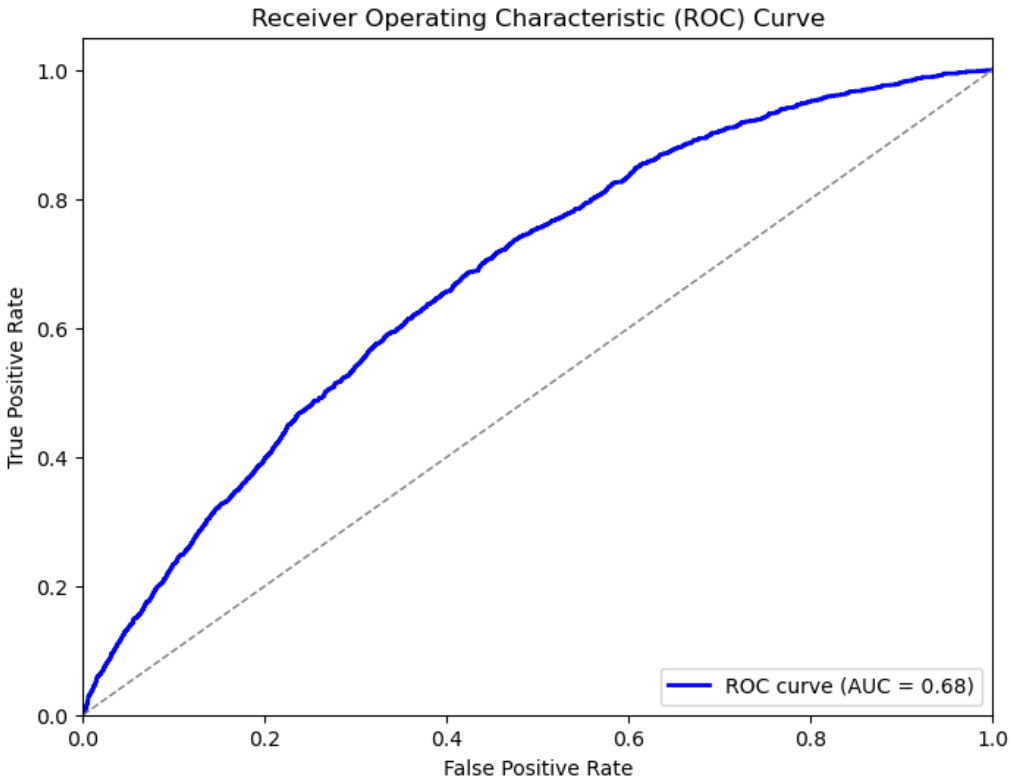

Figure 2: ROC

The Accuracy was 0.91, AUC 0.68, which is good for a simple model.

Timeline

- 2/16: Project Proposal Due
- 2/19 - 2/25: Data Cleaning
- 2/26 - 3/10: Exploratory Data Analysis (EDA)
- 3/11 - 3/17: Spring Break
- 3/18 - 3/24: Variable Selection
- 3/25 - 3/31: Creation Trial Model
- 4/1 - 4/14: Model Fitting
- 4/15 - 4/21: Wrapping Up

- 4/25, 27: Presentation
- 4/30: Submit Due Date

Conclusion

In this study, we will divide the large-scale dataset of patients receiving critical care in the United States (GOSSIS) into training and testing sets. Subsequently, we will create models using neural network methods and other traditional machine learning techniques to predict mortality rates among patients receiving critical care. Developing a model capable of accurately predicting mortality rates in critical care is considered significant as it contributes to the allocation strategies of limited medical resources in intensive care settings. Furthermore, it aids in the quality assessment of healthcare institutions and facilitates appropriate categorization of patients in research related to intensive care. On the other hand, it's important to note that the data used in this study are from the United States, and there are certain limitations to generalizing findings to other countries with different intensive care systems and healthcare resources.

Github repo link

https://github.com/2023DS598/Project

References

[1] Raffa, J. D., Johnson, A. E. W., O'Brien, Z., Pollard, T. J., Mark, R. G., Celi, L. A., Pilcher, D., & Badawi, O. (2022). The Global Open Source Severity of Illness Score (GOSSIS). Critical care medicine, 50(7), 1040–1050. https://doi.org/10.1097/CCM.0000000000005518

[2] Nassar, A. P., Jr, Mocelin, A. O., Nunes, A. L., Giannini, F. P., Brauer, L., Andrade, F. M., & Dias, C. A. (2012). Caution when using prognostic models: a prospective comparison of 3 recent prognostic models. Journal of critical care, 27(4), 423.e1–423.e4237. https://doi.org/10.1016/j.jcrc.2011.08.016