MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231115

Development of Machine Learning Prediction Models for Self-Extubation After Delirium Using Emergency Department Data

Koutarou MATSUMOTO[a,b1], Yasunobu NOHARA[b,c], Mikako SAKAGUCHI[d], Yohei TAKAYAMA[b,d], Takanori YAMASHITA[e], Hidehisa SOEJIMA[b] and Naoki NAKASHIMA[e]

[a] *Biostatistics Center, Kurume University, Japan*
[b] *Institute for Medical Information Research and Analysis, Saiseikai Kumamoto Hospital, Japan*
[c] *Big Data Science and Technology, Faculty of Advanced Science and Technology, Kumamoto University, Japan*
[d] *Department of Nursing, Saiseikai Kumamoto Hospital, Japan*
[e] *Medical Information Center, Kyushu University Hospital, Japan*
ORCiD ID: Koutarou Matsumoto https://orcid.org/0000-0003-0769-7763

Abstract. Delirium is common in the emergency department, and once it develops, there is a risk of self-extubation of drains and tubes, so it is critical to predict delirium before it occurs. Machine learning was used to create two prediction models in this study: one for predicting the occurrence of delirium and one for predicting self-extubation after delirium. Each model showed high discriminative performance, indicating the possibility of selecting high-risk cases. Visualization of predictors using Shapley additive explanation (SHAP), a machine learning interpretability method, showed that the predictors of delirium were different from those of self-extubation after delirium. Data-driven decisions, rather than empirical decisions, on whether or not to use physical restraints or other actions that cause patient suffering will result in improved value in medical care.

Keywords. delirium, self-extubation after delirium, machine learning

1. Introduction

Delirium is common in hospitalized patients, with a prevalence of 8%–17% in emergency department patients [1]. Delirium causes incidents such as self-extubations, so it is critical to predict the occurrence of delirium and take preventive measures before it occurs. Although many machine learning models for predicting delirium have been developed, they are not always applicable to the Japanese population, which is rapidly aging. There is a tool for predicting self-extubation incidents, but it is not specific to delirium [2]. As a result, the goal of our research is to create a model that predicts the

[1] Corresponding Author: Koutarou Matsumoto, E-Mail: matsumoto_koutarou@kurume-u.ac.jp.

onset of delirium in the early stages of hospitalization in the emergency department, where delirium is more likely to occur, as well as a machine learning prediction model that predicts the occurrence of self-extubation incidents of drains and tubes after delirium occurs.

2. Methods

2.1. Study Patients

We examined data from patients who were admitted to Saiseikai Kumamoto Hospital in Kumamoto City, Japan. This hospital is one of the altitude emergency and critical care medical centers in southern Japan. Clinical data from routine clinical practice were accumulated in the hospital database, from which clinical variables were extracted. The institutional review board approved the present study (Approval No. 1072). We included 33277 consecutive emergency patients who were hospitalized between December 2017 and February 2022 in this study. This study excluded 1100 patients (563 patients could not be traced, 356 patients died within 24 h after admission, and 181 patients were less than 18 years old). Finally, 32177 patients were included in this study (Figure 1).

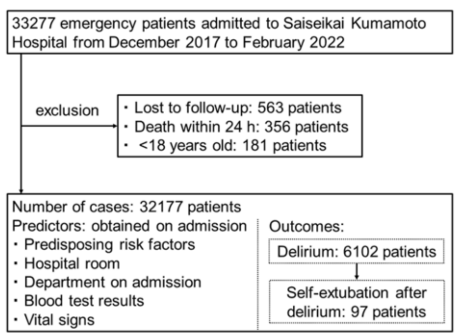

Figure 1. Flow chart for selection of study patients.

2.2. Clinical Outcomes

Incident delirium was defined as a positive confusion assessment method after hospitalization. Cases of delirium-related self-extubation among those who developed delirium and were reported in the incident report were defined as self-extubation after delirium. Patients were assessed daily by nurses trained in a specialized delirium subcommittee that includes a psychiatrist.

2.3. Predictors

Predictive models were developed using data available at the emergency department at the time of admission. A total of 45 predictors, including blood test values and vital signs, were used in addition to basic information such as age and gender.

2.4. Algorithm and Predictive Performance Evaluation

The XGBoost algorithm, a type of decision tree ensemble learning, was used to develop the prediction models. This algorithm was chosen due to its high prediction accuracy and compatibility with the later described interpretive technique. The discriminating ability of each model was assessed by the area under the receiver operating characteristic curve (AUROC) as well as sensitivity and specificity. To determine the optimal cutoff value that maximizes sensitivity and specificity, the Youden index method was used. The calibration performance was assessed by calibration slope. A stratified five-fold cross-validation was performed as internal validation, and the mean and standard deviation of each metric were calculated.

2.5. Handling of Imbalanced Data

In this study, the frequency of occurrence of the outcomes is low and extremely unbalanced. When predicting imbalanced events, special handling, such as oversampling, is required when predicting imbalanced events. We used “cost-sensitive learning,” in this study, which adjusts the loss function of the XGBoost algorithm to minimize the loss function by attaching weights to minority classes to make them more strongly identified. The weight was calculated as the inverse of the percentage of instances of the minority class in the data set.

2.6. Machine Learning Interpretability Method

To visualize the relationship between predictors and outcomes, we used Shapley additive explanation (SHAP), a model-agnostic machine learning method. SHAP is a method for visualizing predictors' contributions by using Shapley values, which are guaranteed to be distributed fairly in cooperative game theory. The R package “xgboost” was used to develop the XGBoost algorithm. To calculate and visualize SHAP values, the R package “shapviz” was used. All statistical analyses were R statistical packages (http://www.r-project.org/, version 4.2.0).

3. Results

Delirium after admission was observed in 6102 (8.0%) patients among the 32177 patients, and self-extubation after delirium was observed in 97 (0.3%). Each predictive model generally showed good discrimination (AUROC: 0.782–0.818).

Calibration performance, on the other hand, was good for the model predicting delirium but poor for the model predicting self-extubation after delirium (Table 1). The SHAP summary plots each predicting model’s top 10 predictive contributions. (Figure 2). A SHAP dependence plot depicted the relationship between age, albumin level, and body mass index, which are continuous values among the top predictors of the delirium prediction model, and delirium (Figure 3).

Table 1. Prediction accuracy of each model after stratified 5-fold cross-validation.

Prediction models	AUROC	Sensitivity	Specificity	Calibration slope
Delirium	0.818 ± 0.006	0.828 ± 0.030	0.672 ± 0.035	1.067 ± 0.044
Self-extubation after delirium	0.782 ± 0.050	0.872 ±0.046	0.631 ± 0.125	1.309± 0.270

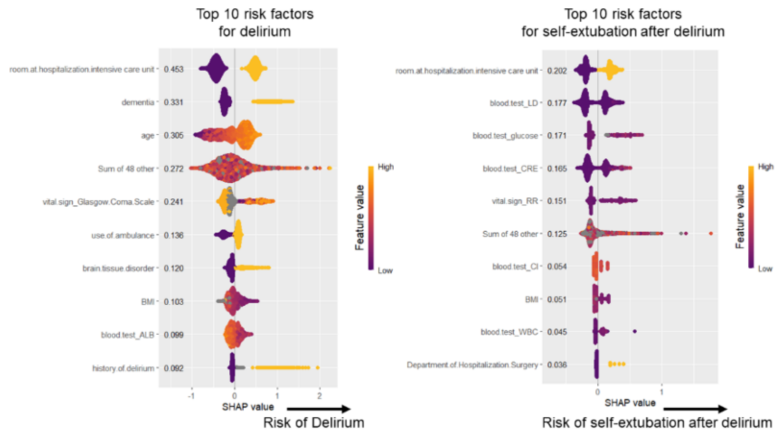

Figure 2. SHAP summary plot for each prediction model.

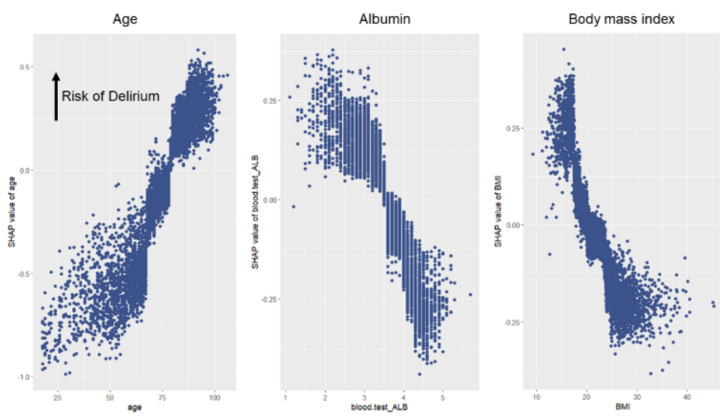

Figure 3. SHAP dependence plot in a delirium prediction model.

4. Discussion

The major findings of this study were as follows. Despite predictions based on imbalanced data, the discriminative performance of the two developed prediction models was good. By confirming the SHAP summary plots, it was visualized that there are different predictors of delirium and self-extubation after delirium.

Although predictive models using machine learning in emergency departments have been developed, high prediction accuracy is not always possible in Japan due to its super-aging population. The SHAP dependence plot demonstrates that low albumin and BMI values are associated with risk, which may reflect the nutritional frailty of elderly hospitalized patients. Given these findings, it may be preferable to develop a Japanese-specific model for predicting the occurrence of delirium rather than using a model developed in Europe or the United States.

Physical restraints are sometimes necessary when dealing with delirium, and they cause patients physical and emotional distress. Physical restraints are a strong precipitating factor for delirium, with the risk of developing delirium being approximately 2.9 times higher than in the absence of restraints [3]. On the other hand, when agitated delirium develops, the patient may self-extubate the drain or tube, resulting in death. If the risk of delirium and self-extubation after delirium can be calculated using machine learning and data from the emergency department at an early stage of hospitalization, it will help support decisions about whether or not to use physical restraints (Figure 4). If predictive performance, which is one of machine learning's strengths, can be successfully applied in the medical field, it will be possible to reduce the burden on medical personnel and patients while also preventing serious medical accidents. However, while the model for predicting self-extubation after delirium performed well in terms of discrimination, there are issues with calibration performance among the machine learning prediction models developed in this study. Continued research on machine learning prediction models using imbalanced data with extremely few events is needed.

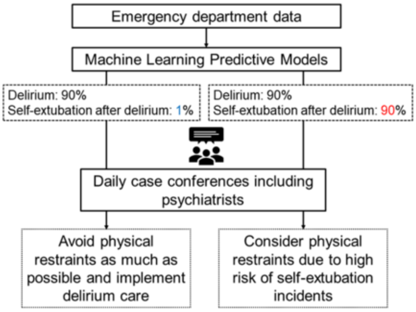

Figure 4. Conceptual diagram of decision support using machine learning predictive models.

5. Conclusions

We demonstrated that it is possible to create a delirium prediction model as well as a machine learning model to predict self-extubation following delirium. The application of these techniques to medical decisions like physical restraints would increase the value of medical care.

References

[1] Inouye SK, Westendorp RG, Saczynski JS. Delirium in elderly people. Lancet. 2014 Mar;383(9920):911-22, doi: 10.1016/S0140-6736(14)60994-6.

[2] Moons P, Boriau M, Ferdinande P. Self-extubation risk assessment tool: predictive validity in a real-life setting. Nurs Crit Care. 2008 Nov-Dec;13(6):310-4, doi: 10.1111/j.1478-5153.2008.00305.x

[3] McPherson JA, Wagner CE, Boehm LM, Hall JD, Johnson DC, Miller LR, Burns KM, Thompson JL, Shintani AK, Ely EW, Pandharipande PP. Delirium in the cardiovascular ICU: exploring modifiable risk factors. Crit Care Med. 2013 Feb;41(2):405-13, doi: 10.1097/CCM.0b013e31826ab49b.