



# Machine Learning to Predict Three Types of Outcomes After Traumatic Brain Injury Using Data at Admission: A Multi-Center Study for Development and Validation

Matsuo, Kazuya ; Aihara, Hideo ; Hara, Yoshie ; Morishita, Akitsugu ; Sakagami, Yoshio ; Miyake, Shigeru ; Tatsumi, Shotaro ; Ishihara, ...

---

(Citation)

Journal of Neurotrauma, 40(15-16):1694-1706

(Issue Date)

2023-08-16

(Resource Type)

journal article

(Version)

Version of Record

(Rights)

© Kazuya Matsuo et al 2023; Published by Mary Ann Liebert, Inc.  
This Open Access article is distributed under the terms of the Creative Commons License (CC-BY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

(URL)

<https://hdl.handle.net/20.500.14094/0100483044>



Open camera or QR reader and  
scan code to access this article  
and other resources online.



ORIGINAL ARTICLE

CLINICAL STUDIES

# Machine Learning to Predict Three Types of Outcomes After Traumatic Brain Injury Using Data at Admission: A Multi-Center Study for Development and Validation

Kazuya Matsuo,<sup>1,2,\*</sup> Hideo Aihara,<sup>3,\*\*</sup> Yoshie Hara,<sup>1</sup> Akitsugu Morishita,<sup>4,\*\*</sup> Yoshio Sakagami,<sup>5</sup> Shigeru Miyake,<sup>6</sup> Shotaro Tatsumi,<sup>7,\*\*</sup> Satoshi Ishihara,<sup>8</sup> Yoshiki Tohma,<sup>9,\*\*</sup> Haruo Yamashita,<sup>1</sup> and Takashi Sasayama<sup>2</sup>

## Abstract

The difficulty of accurately identifying patients who would benefit from promising treatments makes it challenging to prove the efficacy of novel treatments for traumatic brain injury (TBI). Although machine learning is being increasingly applied to this task, existing binary outcome prediction models are insufficient for the effective stratification of TBI patients. The aim of this study was to develop an accurate 3-class outcome prediction model to enable appropriate patient stratification. To this end, retrospective balanced data of 1200 blunt TBI patients admitted to six Japanese hospitals from January 2018 onwards (200 consecutive cases at each institution) were used for model training and validation. We incorporated 21 predictors obtained in the emergency department, including age, sex, six clinical findings, four laboratory parameters, eight computed tomography findings, and an emergency craniotomy. We developed two machine learning models (XGBoost and dense neural network) and logistic regression models to predict 3-class outcomes based on the Glasgow Outcome Scale-Extended (GOSE) at discharge. The prediction models were developed using a training dataset with  $n = 1000$ , and their prediction performances were evaluated over two validation rounds on a validation dataset ( $n = 80$ ) and a test dataset ( $n = 120$ ) using the bootstrap method. Of the 1200 patients in aggregate, the median patient age was 71 years, 199 (16.7%) exhibited severe TBI, and emergency craniotomy was performed on 104 patients (8.7%). The median length of stay was 13.0 days. The 3-class outcomes were good recovery/moderate disability for 709 patients (59.1%), severe disability/vegetative state in 416 patients (34.7%), and death in 75 patients (6.2%). XGBoost model performed well with 69.5% sensitivity, 82.5% accuracy, and an area under the receiver operating characteristic curve of 0.901 in the final validation. In terms of the receiver operating characteristic curve analysis, the XGBoost outperformed the neural network-based and logistic regression models slightly. In particular, XGBoost

<sup>1</sup>Department of Neurosurgery, Hyogo Emergency Medical Center and Kobe Red Cross Hospital, Kobe, Japan.

<sup>2</sup>Department of Neurosurgery, Kobe University Graduate School of Medicine, Kobe, Japan.

<sup>3</sup>Department of Neurosurgery, Hyogo Prefectural Himeji Cardiovascular Center, Himeji, Japan.

<sup>4</sup>Department of Neurosurgery, <sup>9</sup>Acute Care Medical Center, Hyogo Prefectural Kakogawa Medical Center, Kakogawa, Japan.

<sup>5</sup>Department of Neurosurgery, Hyogo Prefectural Awaji Medical Center, Sumoto, Japan.

<sup>6</sup>Department of Neurosurgery, Kita-harima Medical Center, Ono, Japan.

<sup>7</sup>Department of Neurosurgery, Hirohata Steel Memorial Hospital, Himeji, Japan.

<sup>8</sup>Department of Emergency and Critical Care Medicine, Hyogo Emergency Medical Center, Kobe, Japan.

**\*\*Current affiliation of Hideo Aihara, Akitsugu Morishita, and Shotaro Tatsumi: Department of Neurosurgery, Hyogo Prefectural Harima-Himeji General Medical Center, Himeji, Japan. Current affiliation of Yoshiki Tohma: Hyogo Prefectural Harima-Himeji General Medical Center Acute Care Medical Center, Himeji, Japan.**

\*Address correspondence to: Kazuya Matsuo, MD, Department of Neurosurgery, Hyogo Emergency Medical Center and Kobe Red Cross Hospital, 1-3-1 Wakinohamakaigandori, Chuo-ku, Kobe, Hyogo 651-0073, Japan E-mail: kkmatsuo@outlook.jp

© Kazuya Matsuo et al 2023; Published by Mary Ann Liebert, Inc. This Open Access article is distributed under the terms of the Creative Commons License (CC-BY) (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

outperformed the logistic regression model significantly in predicting severe disability/vegetative state. Although each model predicted favorable outcomes accurately, they tended to miss the mortality prediction. The proposed machine learning model was demonstrated to be capable of accurate prediction of in-hospital outcomes following TBI, even with the three GOSE-based categories. As a result, it is expected to be more impactful in the development of appropriate patient stratification methods in future TBI studies than conventional binary prognostic models. Further, outcomes were predicted based on only clinical data obtained from the emergency department. However, developing a robust model with consistent performance in diverse scenarios remains challenging, and further efforts are needed to improve generalization performance.

**Keywords:** artificial intelligence; Glasgow Outcome Scale-Extended; machine learning; neural network; traumatic brain injury

## Introduction

There have been no significant advances in the treatment of traumatic brain injury (TBI) in recent decades.<sup>1</sup> Although mortality for patients with severe TBI has decreased slightly, which is primarily attributable to improvements in the emergency medical system, the prevalence of severe disability and vegetative state has increased. Moreover, the rate of favorable outcomes has not been improved.<sup>2,3</sup> As a result, TBI remains the leading cause of death among young people and the leading cause of death and disability in all age groups globally.<sup>1,4</sup> Further, the incidence of TBI is expected to continue to increase.<sup>5</sup> As such, various clinical trials have been conducted to improve the prognosis of TBI.<sup>6</sup>

However, despite promising pre-clinical results, most randomized clinical trials on medical and surgical treatments have failed to demonstrate effectiveness.<sup>1,6,7</sup> The failure of prospective trials to demonstrate statistical superiority for such treatments may be attributed to the inability to identify patients who would benefit from novel therapies accurately. Existing screening methods, such as the Glasgow Coma Scale (GCS), should be insufficient on their own for successful stratification in clinical trials involving TBI patients with multiple underlying pathophysiological mechanisms. The pathogenesis of TBI involves not only primary brain injury from hematoma, cerebral contusion, diffuse axonal injury, and diffuse brain swelling, but also ischemia-reperfusion injury, inflammatory reaction, brain herniation, hypoxia, and hypotension caused by extracranial injuries, and secondary brain swelling resulting from these processes.<sup>3,6</sup>

Machine learning techniques have been applied to classify TBI patients appropriately in several studies, including our previous study.<sup>8</sup> However, most of these models only provide binary predictions, such as in-hospital mortality, and are insufficient to describe patient severity and thus contribute to effective stratification.<sup>9–15</sup> Currently, it remains unclear whether machine learning models can predict more specific outcome categories to stratify TBI patients more precisely.

The objective of this study was to develop an accurate 3-class outcome prediction model that can serve as the basis for appropriate patient stratification in future TBI studies. We only used clinical data obtained from the emergency department (ED) to train machine learning models based on multi-institutional retrospective data.

## Methods

### Ethical approval and data acquisition

The ethics committee of the Japanese Red Cross Kobe Hospital approved this study (No. 247) and waived the requirement for informed consent, as this was a retrospective observational study. In compliance with the Ethical Guidelines for Medical and Health Research Involving Human Subjects in Japan, participants in the study were given the option to withdraw from the study at any time through an opt-out method on institutional websites. The data supporting this study's findings are available from the corresponding author upon reasonable request. This study followed the transparent reporting of a multi-variable prediction model for individual prognosis or diagnosis guidelines.<sup>16</sup>

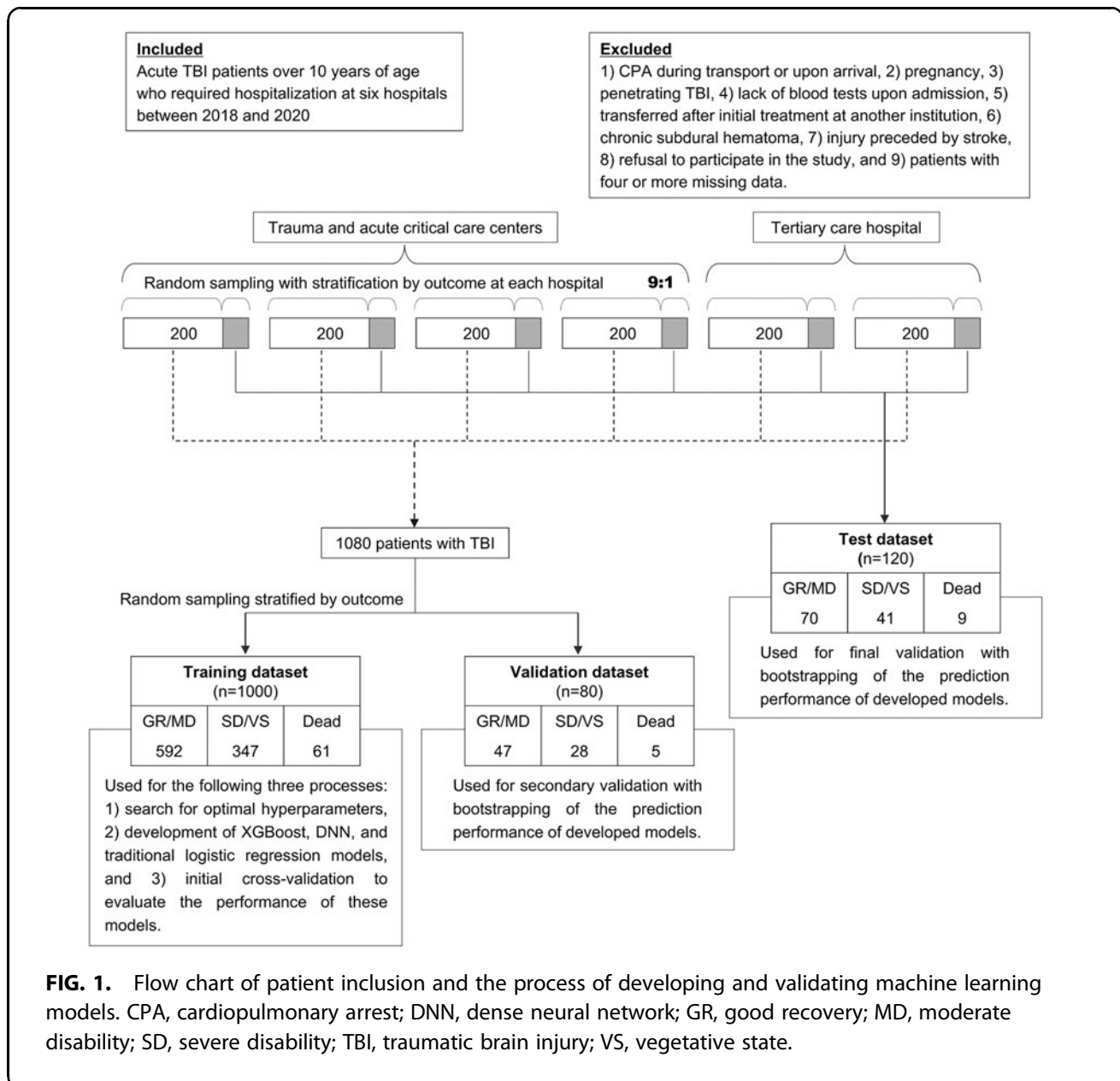
### Study design and participants

Data for 200 consecutive patients admitted for acute TBI treatments to each of six hospitals in Hyogo, Japan, from January 2018 onwards were collected retrospectively. To enroll 200 patients from each institution, patient data were collected until November 2021 at hospitals with fewer TBI patients, and until May 2019 at hospitals with more patients. Four of the six facilities are trauma and acute critical care centers, while two are tertiary care hospitals. All participating institutions offer 24-h neurosurgical services and acute neurological care to TBI patients, with board-certified neurosurgeons providing standard management according to the guidelines and consensus.<sup>17,18</sup> All information was obtained from electronic health record systems or institutional trauma registries.

Inclusion criteria were: 1) male or female participants (> 10 years of age); and 2) diagnosed with TBI requiring hospitalization. Exclusion criteria were: 1) cardiopulmonary arrest during transport or upon arrival; 2) pregnancy; 3) penetrating TBI; 4) lack of blood tests upon admission; 5) transferred after initial treatment at another institution; 6) chronic subdural hematoma; 7) injury preceded by stroke; 8) refusal to participate in the study; and 9) patients with four or more missing data. Patients who met the exclusion criteria were pre-screened, and a total of 1200 TBI patients were included in the study, with 200 consecutively admitted patients from each of the six hospitals. Setting a target period and including all consecutive cases within that period would have increased the variability in the number of patients at each institution,

resulting in potential heterogeneity in the training data and, consequently, significant bias during the development and evaluation of the prediction model. Therefore, we decided to use a balanced dataset in this study by including a fixed number of consecutive TBI cases from each institution, instead of opting for a universal data collection period.

We randomly selected 10% of the 200 patients at each hospital to construct the test dataset (comprising 120 patients in aggregate). The remaining 1080 patients were randomly divided into validation (80 patients) and training (1000 patients) datasets (Fig. 1). The difference between the validation and test datasets was that the latter consisted of equal numbers of TBI patients from all hospitals, whereas the validation dataset was not stratified by



hospital. The sample size was determined based on our previous findings.<sup>8</sup> In the case of binary prediction, our previous model showed that approximately 180 patients from a single institution are sufficient to ensure good prediction performance with an appropriate selection of clinical factors. As a result, we assumed that 90% of the obtained data should be used for training, and set the number of patients collected from each institution to be 200. The six emergency hospitals were rigorously selected based on their ability to provide high-quality TBI care during the study period consistently, yielding a total sample size of 1200 patients.

### Predictive parameters

We developed prediction models incorporating the following 21 predictors—age; sex; six clinical findings, including systolic blood pressure, GCS total score, GCS motor score, the presence of spinal cord injury, the presence of major extracranial injuries and pupillary abnormalities; four laboratory parameters including hemoglobin, glucose, C-reactive protein, and D-dimer levels on blood tests; eight computed tomography (CT) findings including acute subdural hematoma (ASDH), acute epidural hematoma, traumatic subarachnoid hemorrhage (tSAH), cerebral contusion, mass lesion, midline shift, basal cistern compression, and modified Rotterdam CT score (Supplementary Table S1)<sup>19</sup>; and emergency craniotomy. These clinical parameters were obtained upon initial presentation to the ED. The definition of each clinical factor is presented in the Supplementary Methods.

### Outcome classification

Patient outcomes were evaluated based on the Glasgow Outcome Scale (GOS) and GOS-extended (GOSE) at discharge.<sup>20,21</sup> The evaluating neurosurgeon determined and rated the assessment items, such as social activities, ability to work as before, and maximum time for which the patients could take care of themselves at home, based on their in-hospital activities of daily living and need for care. We defined three outcome categories for a predictive target: 1) GOSE 5 (lower moderate disability [MD])–8 (upper good recovery [GR]) as good; 2) GOSE 2 (vegetative state [VS])–4 (upper severe disability [SD]) as poor; and 3) GOSE 1 as death.

### Machine learning model development

Missing variables were imputed using the k-nearest neighbor imputation method.<sup>22</sup> In the k-nearest neighbor imputation for missing variables, the admission hospital, the presence of pneumocephalus, and the length of hospitalization were introduced in addition to the 21 learning parameters. All 21 parameters were then standardized. We used XGBoost and a dense neural network (DNN)

as machine learning approaches.<sup>23,24</sup> The optimal hyperparameters of the models were determined using automated Bayesian optimization implemented using the Hyperopt library.<sup>25</sup> During hyperparameter optimization, we repeated the search for hyperparameters that maximize accuracy, sensitivity, F1 score, area under the receiver operating characteristic curve (AUROC), and area under the precision-recall curve (AUC-PR). The prediction model was designed to determine the prediction probability for each of the three classifications, with the classification with the highest probability serving as the model's response. Each evaluation index was calculated using a confusion matrix obtained based on the predicted and true classifications, and for indices other than accuracy, the macro-averages of the values of each outcome class were used for evaluation. Thus, the construction of this model did not require a probability threshold to be set.<sup>26</sup> Both the XGBoost and DNN models were trained using an early stopping function, which terminated training when the specified metrics reached the optimal values.

The following 10 hyperparameters of XGBoost were optimized: “max\_depth,” “min\_child\_weight,” “gamma,” “subsample,” “colsample\_bytree,” “lambda,” “alpha,” “max\_delta\_step,” “eval\_metric,” and learning rate. The number of decision trees was fixed to 100,000, and the value of “early\_stopping\_rounds” was fixed to 100. The following 11 hyperparameters of the DNN model were optimized: number of hidden layers, activation method for hidden layers, number of units of hidden layers, dropout rate on input, dropout rate on hidden layers, type of optimizer, learning rate of the optimizer, usage of batch normalization on a hidden layer, batch size, L1 regularization weight, and L2 regularization weight. Regularization is a machine learning technique that adds a regularization term to loss functions to reduce overfitting. L1 regularization adds an L1 penalty equal to the absolute value of the coefficient magnitude. L2 regularization adds an L2 penalty equal to the square of the coefficient magnitude.

The training epoch was fixed to 1000 rounds, the patience for early stopping was fixed to 100, and the loss metric was fixed to mean squared error. We adopted the hyperparameter set that produced the highest and second-highest values for the five predefined statistical indicators during hyperparameter optimization for model training on the training dataset, yielding 10 XGBoost and 10 DNN models in aggregate. Python version 3.9, scikit-learn version 1.1.3, and XGBoost version 1.7.1 were used to develop and validate the machine learning models. The DNN model was implemented using Tensorflow version 2.6.0. A subset of the program code generated for this study is available at GitHub and can be accessed on <https://gist.github.com/kkmatsuo/75b8e1bcbe93c59492d24f9af690ae46>.

## Analysis of model performance

First, as initial validation of the developed models, we performed three-repeated 10-fold stratified cross-validation on the training dataset. For secondary validation using the validation dataset, we selected the XGBoost and DNN models with the top three AUC-PR scores as measured during the initial validation. Subsequently, we selected the XGBoost and DNN models with the highest AUC-PR scores on the validation dataset for further evaluation on the test dataset as final validation (i.e., test). We also performed multinomial logistic regression to compare prediction performances on the training, validation, and test datasets.<sup>27</sup> This logistic regression model was designed as a statistical method with no regularization or penalization.<sup>28,29</sup> We performed bootstrap analysis 1000 times on the validation and test datasets to generate a value distribution for each evaluation index.

Sensitivity, positive predictive value (PPV), accuracy, F1 score, AUROC, and AUC-PR were used to evaluate prediction performance. Except for accuracy, the averages of the respective prediction results for the three outcome classes were employed as the evaluation indices (i.e., macro-average). We compared the prediction performances of representative XGBoost, DNN, and logistic regression models using 3-class confusion matrices and 3-class ROC curves. No model updates (e.g., recalibration) were conducted throughout the modeling process in response to the validation results. In addition, we assessed the contribution of each predictor to the predicted outcome using SHapley Additive exPlanations (SHAP) values.

## Statistical analysis

Categorical variables were reported as frequencies and percentages, whereas continuous variables were reported as medians and interquartile ranges (IQRs). For comparison between datasets, the chi-squared test and Kruskal-Wallis one-way analysis of variance were used to compare categorical and continuous variables. In comparisons of patient characteristics between the six hospitals, *p* values were adjusted using Bonferroni correction. On the test dataset, we compared the AUROC corresponding to each outcome class of the best performing XGBoost and logistic regression models using the DeLong test and calculated the *p* values.<sup>30</sup> Since the DeLong test cannot be applied to the macro-averages of AUROC in multi-class classification, we only presented 95% confidence intervals calculated using the bootstrap method for the macro-averages of multi-class AUROC. All *p* values were two-tailed, and significance was set to *p* < 0.05. All statistical analyses were performed with EZR version 1.54 (Saitama Medical Center, Jichi Medical University, Saitama, Japan), which is based on R version 4.0.3. The “pROC” package was used to calculate the *p* values of the DeLong test.<sup>31</sup>

## Results

### Study participants

Figure 1 shows the study flowchart, including the number of patients corresponding to each outcome. Of the 1200 patients in aggregate, the median age was 71 years (IQR 51–81), 756 (63.0%) were male, 199 (16.6%) had severe TBI with GCS 3–8, 212 (17.7%) had severe extracranial injuries, and 43 (3.6%) experienced hypotension in the ED. The most common CT findings were tSAH (56.8%) and ASDH (47.4%). Emergency craniotomy was performed on 104 patients (8.7%). The median length of stay was 13.0 (IQR 4.0–28.0) days, and GOS ratings at discharge were as follows: GR in 514 (42.8%) patients, MD in 195 (16.3%), SD in 369 (30.8%), VS in 47 (3.9%), and death in 75 (6.2%). Missing data were common for D-dimer (*n* = 236) and glucose (*n* = 26; Supplementary Table S2). Even after imputing missing values, patient characteristics were similar (Supplementary Table S3). Baseline characteristics were similar between training, validation, and test datasets; however, a trend toward more tSAH was observed in the test dataset and one towards more pupillary abnormalities was observed in the validation dataset (Table 1). The distributions of GOSE and the 3-class outcomes showed similar trends on each dataset (Fig. 2A, 2B). Patient characteristics differed significantly among the six hospitals (Supplementary Table S4).

### Initial validation of the prediction models on the training dataset

On the training dataset, cross-validation results showed that the logistic regression model had the highest PPV and F1 score, while the XGBoost had the highest sensitivity, accuracy, AUROC, and AUC-PR. The complete results of the 10 XGBoost, 10 DNN, and logistic regression models are shown in Supplementary Table S5.

### Further validation of the prediction models on the validation and test datasets

On the validation dataset, the secondary validation results showed that the DNN with AUROC-maximizing hyperparameters had the highest sensitivity, PPV, accuracy, and F1 score of 83.7%, 85.0%, 79.9%, and 84.0%, respectively (Supplementary Table S6). The AUROC and AUC-PR of the logistic regression model were the highest, at 0.908 and 0.882, respectively. XGBoost with accuracy-maximizing hyperparameters and DNN with AUROC-maximizing hyperparameters had the highest AUC-PR among the XGBoost and DNN models—these were employed for further validation.

The test results showed that XGBoost with accuracy maximizing hyperparameters had the highest scores in terms of all metrics, with sensitivity of 69.5%, PPV of

**Table 1. Baseline Characteristics of Patients With TBI in the Training, Validation, and Test Datasets**

Characteristics	Training (n=1200)	Validation (n=80)	Test (n=120)	p value
Age, median (IQR), year	71 (50-81)	69.5 (53.8-81.3)	71 (54-82)	.90
Male sex, n (%)	628 (52.3)	52 (65)	76 (63.3)	.92
Pupillary abnormalities, n (%)	92 (7.7)	11 (14)	6 (5.0)	.10
SBP, median (IQR), mm Hg	142 (123-165)	142 (126-162)	143 (131-170)	.51
Hypotension (SBP <90 mm Hg), n (%)	36 (3.0)	3 (3.8)	4 (3.3)	.99
Major extracranial injury, n (%)	172 (14.3)	14 (17.5)	26 (21.7)	.48
Spinal injury, n (%)	21 (1.8)	0 (0)	0 (0)	.12
GCS eye score, median (IQR)	4 (3-4)	4 (3-4)	4 (3-4)	.33
GCS verbal score, median (IQR)	4 (3.8-5)	4 (3-5)	4.5 (4-5)	.29
GCS motor score, median (IQR)	6 (6-6)	6 (5-6)	6 (6-6)	.80
GCS total score, median (IQR)	14 (12-15)	14 (11-15)	14 (13-15)	.84
Severe TBI (GCS 3-8), n (%)	166 (13.8)	12 (15)	21 (17.5)	.90
CT findings				
AEDH, n (%)	91 (7.6)	6 (7.5)	17 (14.2)	.17
ASDH, n (%)	471 (39.3)	36 (45)	62 (51.7)	.58
Contusion, n (%)	308 (25.7)	22 (28)	39 (32.5)	.75
TSAH, n (%)	579 (48.3)	36 (45)	67 (55.8)	.08
Mass lesion, n (%)	122 (10.2)	9 (11)	18 (15)	.64
Midline shift, n (%)	84 (7.0)	7 (8.8)	9 (7.5)	.94
Absent cistern, n (%)	75 (6.3)	9 (11)	10 (8.3)	.48
Modified Rotterdam CT score, median (IQR)	3 (2-3)	3 (2-3)	3 (2-3)	.69
Laboratory findings				
CRP, median (IQR), mg/dL	0.07 (0.02-0.34)	0.06 (0.03-0.13)	0.10 (0.03-0.27)	.51
D-dimer, median (IQR), µg/mL	13.9 (4.2-41.7)	15.9 (6.7-44.2)	16.3 (4.8-46.5)	.45
Glucose, median (IQR), mg/dL	139 (117-177)	136 (123-172)	146 (117-179)	.72
Hemoglobin, median (IQR), mg/dL	13.0 (11.5-14.2)	13.0 (11.9-14.2)	13.3 (11.9-14.4)	.47
Treatment and outcomes				
Emergent craniotomy, n (%)	86 (7.2)	6 (7.5)	12 (10)	.81
GR or MD, n (%)	592 (49.3)	47 (59)	70 (58.3)	.98
SD or VS, n (%)	347 (28.9)	28 (35)	41 (34.2)	.99
Dead, n (%)	61 (5.1)	5 (6.3)	9 (7.5)	.84
Length of stay, median (IQR), day	13 (4-28)	12 (3-28)	15 (4.8-26)	.80

The *p* values are calculated using chi-squared test or Kruskal–Wallis one-way analysis of variance. Outcomes are assessed at discharge according to the Glasgow outcome scale.

AEDH, acute epidural hematoma; ASDH, acute subdural hematoma; CRP, C-reactive protein; CT, computed tomography; GCS, Glasgow coma scale; GR, good recovery; IQR, interquartile range; MD, moderate disability; SBP, systolic blood pressure; SD, severe disability; TBI, traumatic brain injury; TSAH, traumatic subarachnoid hemorrhage; VS, vegetative state.

87.6%, accuracy of 82.5%, F1 score of 74.1%, AUROC of 0.901, and AUC-PR of 0.803 (Table 2). The hyperparameters corresponding to the best models are provided in Supplementary Table S7.

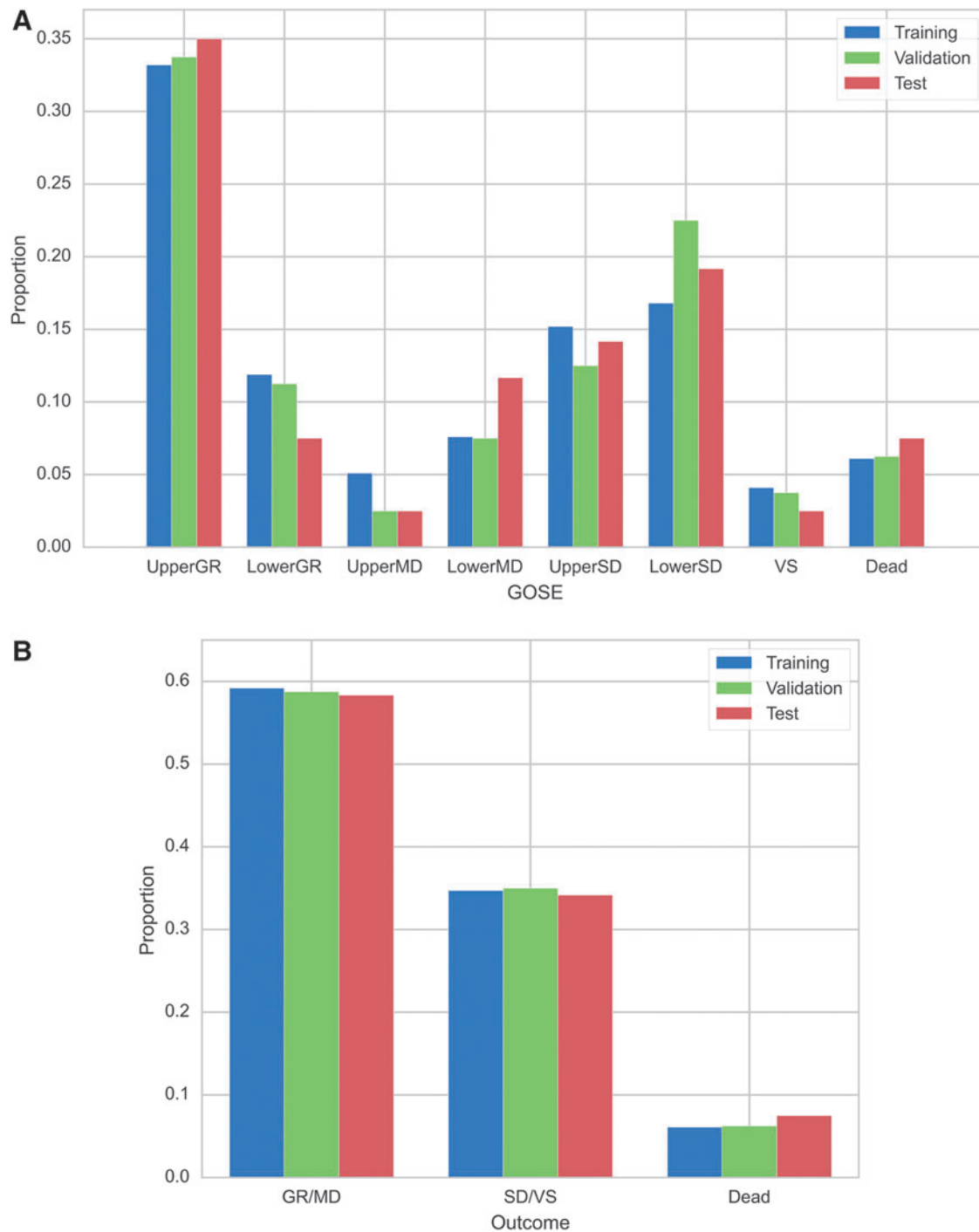
Based on 3-class ROC curve analysis, XGBoost was observed to slightly outperform DNN and logistic regression models, except for SD/VS prediction on the validation dataset (Fig. 3 and Supplementary Fig. S1). Comparison of the AUROC of the best XGBoost and logistic regression models on the test dataset revealed that the XGBoost model significantly outperformed the logistic regression model in predicting SD/VS (Supplementary Table S8). The 3-class confusion matrix analysis demonstrated that, while all models tended to predict GR/MD accurately, they tended to miss the prediction of death, particularly on the test dataset (Fig. 4). These analyses indicated that each prediction model behaved differently depending on the scenario, and the prediction targets on which it excelled did not appear to be consistent. For instance, on the validation dataset, XGBoost outperformed the other models in terms of PPV in predicting GR/MD and death, but it was less sensitive in predicting death.

### Importance of the clinical predictors

According to the SHAP values of clinical parameters based on the best XGBoost model on the test dataset, GCS total score, age, and D-dimer contributed the most to the prediction results (Fig. 5). Regarding imaging findings, ASDH exerted the greatest impact on outcome prediction.

### Discussion

This study revealed several new implications related to the application of machine learning in TBI research. First, we developed a more specialized machine learning model to predict in-hospital outcomes following TBI using three GOS-based classifications, and concluded that the XGBoost model exhibited a high AUROC of approximately 0.90. However, there is still scope for improvement, particularly concerning the sensitivity of death prediction, and developing a robust model with consistent performance in diverse scenarios remains challenging. Additionally, even on the training dataset, which was learned by the model, the best performing XGBoost model's prediction accuracy of 77.6% appeared to be insufficient, suggesting that the models may not yet be fully applicable to real-world settings. Next, although



**FIG. 2.** Distribution of the Glasgow outcome scale extended on each dataset **(A)**. Distribution of the three-class outcomes on each dataset **(B)**. GR, good recovery; MD, moderate disability; SD, severe disability; VS, vegetative state.

traditional logistic regression performed well on the validation dataset, the machine learning-based model outperformed it in the other two validation tests. Therefore, the machine learning-based models may have better generalization performance than logistic regression.

Our model achieved an AUROC of 0.90, which is a good level of discrimination. However, its prediction performance cannot be compared to that of any current model. The proposed model tended to miss death predictions, especially on the test dataset, resulting in a

**Table 2. Prediction Results for the Test Dataset With Bootstrapping of 1000 Repetitions**

<i>Model</i>	<i>Sensitivity</i>	<i>PPV</i>	<i>Accuracy</i>	<i>F1 score</i>	<i>AUROC</i>	<i>AUC-PR</i>
XGBoost [Accuracy]	<b>0.695(0.693-0.697)</b>	<b>0.876(0.875-0.878)</b>	<b>0.825(0.824-0.826)</b>	<b>0.741(0.739-0.743)</b>	<b>0.901(0.9-0.902)</b>	<b>0.803(0.801-0.805)</b>
Logistic regression	0.616(0.614-0.618)	0.709(0.706-0.712)	0.774(0.773-0.775)	0.637(0.635-0.639)	0.863(0.862-0.864)	0.742(0.74-0.744)
DNN [AUROC]	0.621(0.618-0.623)	0.74(0.738-0.743)	0.741(0.74-0.742)	0.65(0.648-0.652)	0.861(0.86-0.862)	0.735(0.733-0.737)

The figures in bold indicate the highest value for each performance metric.

Models are sorted by AUC-PR. Maximized metric in hyperparameter tuning is indicated in parentheses after the model's name (e.g., DNN [AUROC]). Metrics are calculated using macro averages of results for each of the three outcome classes and are reported as means and 95% confidence intervals from bootstrapping.

AUC-PR, area under the precision-recall curve; AUROC, area under the receiver operating characteristic curve; DNN, dense neural network; PPV, positive predictive value.

relatively low sensitivity of 69.5%. This seems attributable to the difficulty of predicting death caused by the worsening of underlying diseases, such as heart failure or cancer, based solely on ED information. Post-admission information can be a useful predictor; however, predicting outcomes only based on ED data would be more helpful to subsequent treatment decisions. To date, only a single study has predicted three categories of outcomes after TBI.<sup>32</sup> This study differs from the present study in that the outcome categories were defined somewhat loosely in terms of the types of discharge disposition, rather than in terms of GOS. Meanwhile, several studies have developed machine learning-based models to predict binary outcomes following TBI, such as in-hospital mortality,<sup>8,10,11,13,33-36</sup> poor outcomes at discharge,<sup>8,14,37</sup> and favorable outcomes after six months.<sup>38</sup>

These models suffer from certain limitations, such as including patients with chronic subdural hematoma,<sup>34</sup> excluding patients with mild TBI,<sup>13,33,35,38</sup> requiring information on post-admission interventions<sup>10</sup> or many underlying conditions as learning parameters,<sup>11,32,33</sup> and insufficient validation.<sup>8,10,14,33</sup> An alternative approach to address the multi-class classification problem is to combine these binary classifiers separately. However, if, for example, a binary prediction model for death is combined with one for an unfavorable outcome, some patients will be assigned contradictory outcomes—they may be predicted to die by one model but predicted to have a favorable outcome by the other. This issue may be resolved in the future if the prediction performance of the models improves drastically.

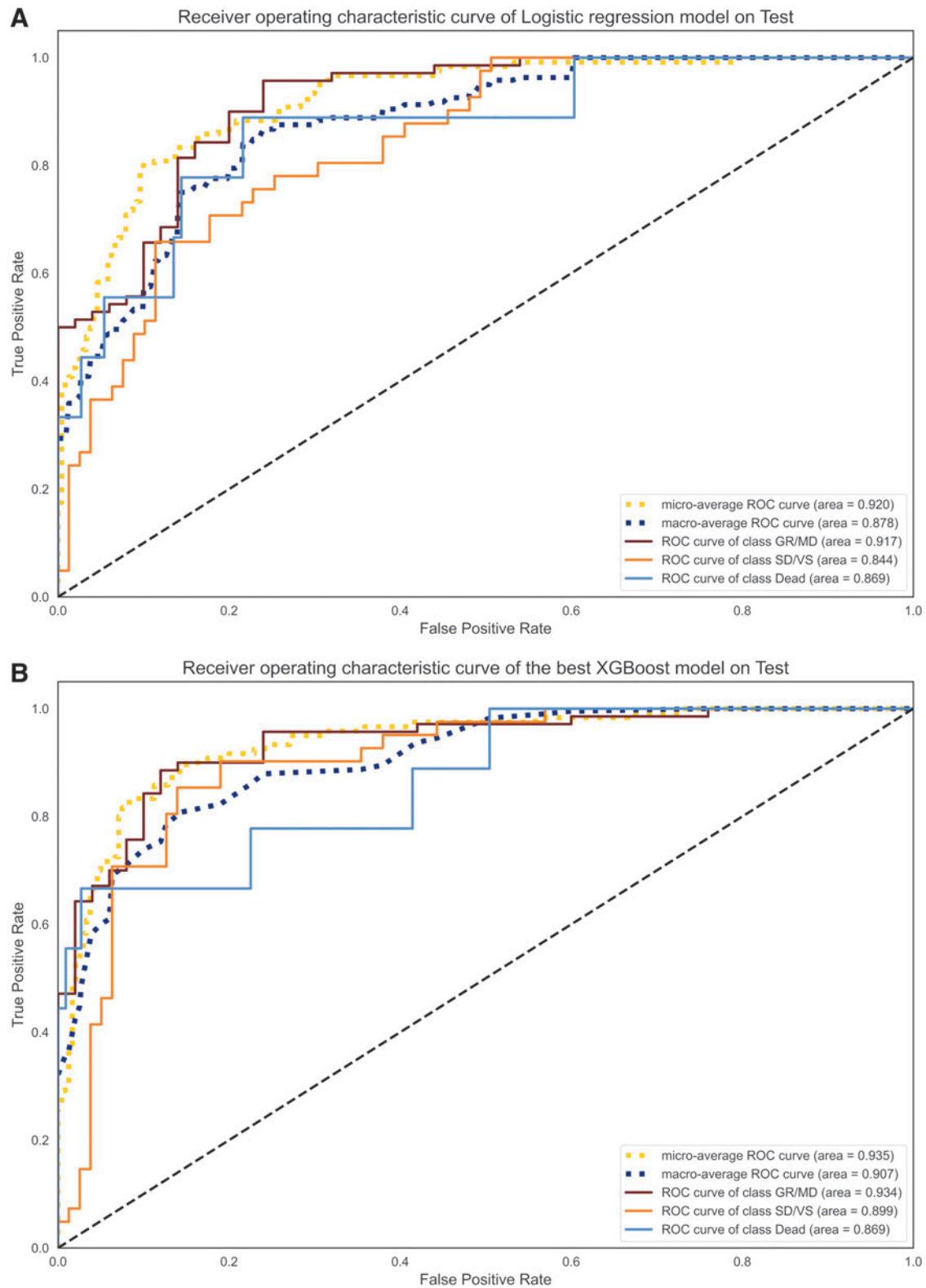
However, another problem is that most recent studies that make binary predictions define an unfavorable outcome as GOS SD or VS or death. Thus, in the case of a combined binary prediction model, if one constituent model predicts survival and the other predicts an unfavorable outcome, a discrepancy is introduced into the response regarding death because unfavorable outcomes include death. The ability to predict outcomes using more detailed categories may enable more effective personalized treatment. However, in our preliminary experiments, it was difficult to predict more subdivided outcomes accurately based solely on current clinical pa-

rameters (data not shown). As evidenced by the adoption of binary prediction, which is a simpler classification model than that proposed in the present study, in most previous studies, it is difficult to predict outcomes with greater subdivisions with high accuracy. To tackle this issue, the introduction of novel biomarkers or more advanced image analysis appear to be required.

Next, although logistic regression performed well on the validation dataset, the machine learning model outperformed it in the other two validation tests. This contradicts several studies that have indicated that machine learning models provide no benefit over regression models.<sup>12,39,40</sup> However, the machine learning models considered in these studies seem to have been inadequately tuned,<sup>40</sup> and the prediction targets of 6-month unfavorable outcomes (GOS <3) appear to be too facile to highlight the advantages of machine learning<sup>12</sup>—these would only be perceptible in more challenging classification problems or problems involving significantly more predictors. Nevertheless, we did not observe a large performance difference between the logistic regression and machine learning models; thus, this topic merits further investigation. Based on previous studies, the utilization of more outcome-relevant predictors should improve prediction performance more effectively than increasing the number of training samples.<sup>8,10,11,13,33,35,36</sup>

### Important predictors identified by machine learning model

The two highest-ranked features determined by the best XGBoost model were GCS total score and age, both of which are well-established predictors of outcome. Thus, it is reasonable that they contributed significantly. D-dimer was the third most influential clinical factor in predicting outcomes—in particular, it was more influential than GCS score or age in predicting death. Coagulopathy after TBI has been reported to occur in nearly two-thirds of patients with severe TBI<sup>41</sup> and is significantly associated with progressive hemorrhagic injury and increased mortality and disability rates at discharge.<sup>41,42</sup> Therefore, in this study, D-dimer most likely represented the degree of coagulation abnormalities associated with the severity of trauma. Since coagulopathy after TBI is complicated by multiple mediating



**FIG. 3.** Three-class receiver operating characteristic curve for the logistic regression model (**A**) and the best XGBoost model (**B**) on test datasets. GR, good recovery; ROC, receiver operating characteristic; MD, moderate disability; SD, severe disability; VS, vegetative state.

<

**FIG. 4.** Three-class confusion matrix for the XGBoost, dense neural network, and logistic regression models in validation (A) and test (B) datasets. For each predicted target ("GR/MD," "SD/VS," and "death"), the highest sensitivity and PPV are highlighted in orange. GR, good recovery; MD, moderate disability; PPV, positive predictive value; SD, severe disability; VS, vegetative state.

mechanisms, it is not yet clear which parameter is the most informative.<sup>41,42</sup> However, considering only prothrombin time (PT) and activated partial thromboplastin time (aPTT) may not be adequate to assess the degree of coagulopathy,<sup>43</sup> and D-dimer and fibrin/fibrinogen degradation products (FDP) are associated with a higher mortality risk than aPTT or platelet count.<sup>44</sup> Additionally, FDP and D-dimer may be less susceptible to modification by oral anticoagulants than PT and aPTT. However, this study did not examine the influence of oral anticoagulants—this topic requires further research.

Second, pupillary abnormalities, which have been suggested to be associated with outcome in previous studies, exhibited a smaller association with outcome in this study. This may be attributed to the following factors. First, pupillary abnormalities were present in only 9.1% of all patients. In addition to this, the present study may have included pupillary abnormalities that differ in pathophysiology from pupillary abnormalities induced as symptoms of brain herniation (e.g., pupillary abnormalities due to traumatic optic neuropathy, original pupillary abnormalities due to previous ocular surgery, or Horner's syndrome due to neck trauma). Thus, this observation may not necessarily reflect brain herniation and may have become relatively less important because other features, such as the absent cistern or the Rotterdam CT score, which reflect brain herniation similarly, are used as features.

### Potential clinical applications of the proposed model

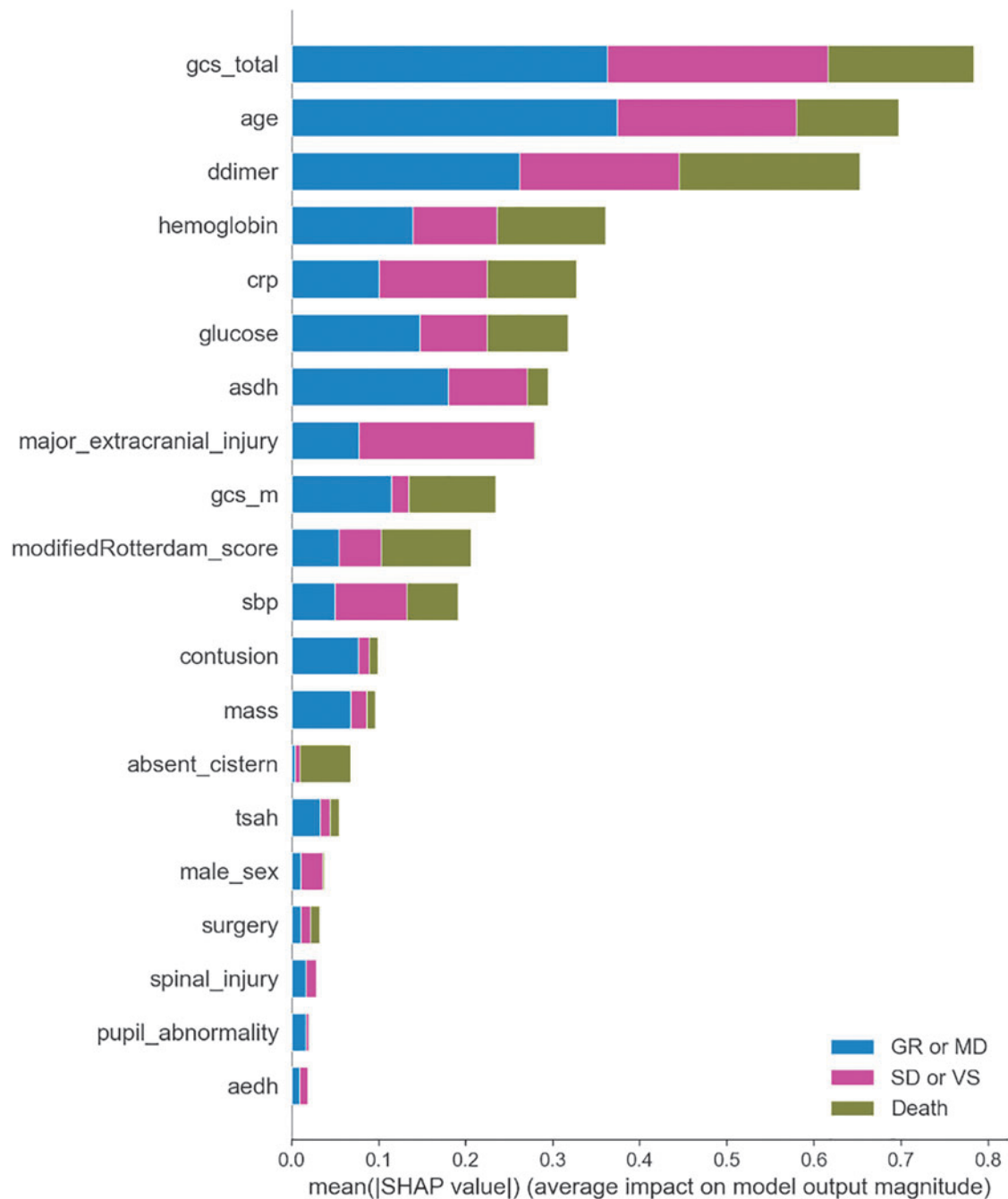
Based on data in the ED and the surgeon's judgment of the necessity of an emergency craniotomy, our model can calculate the probability of three classes of outcomes with and without craniotomy, which may help determine

the treatment strategy. In the future, we might be able to apply this prognostic model as a severity index to identify subgroups of patients who would be most likely to benefit from specific interventions. In recent major commissions, such neuroinformatics-based clinical decision support in the field of TBI has been highly recommended for implementation.<sup>45</sup>

### Limitations

The first major limitation of this study is that the outcome was assessed in terms of GOSE at discharge. Patients with severe disabilities at discharge may recover after six months.<sup>46,47</sup> However, retrospective assessment of outcomes at a pre-determined time, such as 14 days or 6 months post-injury, excludes mild TBI patients who are discharged early and not followed comprehensively in an outpatient setting, rendering data from these mild TBI patients unavailable for model training. In this study, we aimed to develop a highly generalizable prediction model by minimizing restrictions on target patients to reduce learning bias as much as possible. Consequently, it was imperative to include data from mild TBI patients in the dataset, and outcomes at discharge were used for outcome assessment. In addition, there may be inter-rater variation in the assessment of GOSE.<sup>48</sup> Further study is needed to determine the optimal timing for prognostication. Second, this study included patients who died of non-traumatic causes. It might be reasonable to exclude patients who died of underlying diseases from the training data.

Finally, all training data included in this study were obtained from Japanese hospitals that provide standard treatment for TBI. Japan has one of the highest aging populations in the world, with more than a quarter of the population being above 65 years in age.<sup>49</sup>



**FIG. 5.** SHapley Additive exPlanations (SHAP) values on the test dataset calculated using the best XGBoost model. asdh, acute subdural hematoma; crp, C-related protein; gcs, Glasgow coma scale; sbp, systolic blood pressure; tsah, traumatic subarachnoid hemorrhage; aedh, acute epidural hematoma.

Additionally, the median age of the patients in this study was 71 years, which could be attributed to the participation of several rural hospitals with higher-than-average elderly populations.<sup>50</sup> Thus, prediction accuracy might be affected corresponding to patient data obtained from low-resource hospitals or non-Japanese populations.<sup>51</sup> Prospective studies with larger datasets including more young patients are

required to confirm the usefulness of our prediction model and to develop a more robust alternative.

## Conclusions

In this study, we developed the first machine learning-based model to predict in-hospital outcomes after TBI

using three GOS-based classifications. The XGBoost model exhibited a high AUROC of approximately 0.90. This model can potentially be more impactful in the development of appropriate patient stratification methods in future TBI studies than conventional binary prognostic models. Further, outcomes were predicted based solely on clinical data from the ED. However, there is still scope for improvement, particularly concerning the sensitivity of death prediction. Developing a robust model with consistent performance in diverse scenarios remains challenging, suggesting that the proposed models may not yet be fully applicable to real-world settings. Additional innovations are required to improve generalization performance.

### Acknowledgments

We would like to thank Yusuke Okamura, Ayaka Shibano, Subaru Umeda, Shunsuke Yamanishi, Junichi Sakata, Daisuke Yamamoto, and Satoshi Nakamizo for their assistance in data collection and Editage (www.editage.com) for English language editing.

### Authors' Contributions

KM: Conceptualization (lead); methodology; software; formal analysis; writing—original draft. HA: Conceptualization (supporting); resources; writing—review and editing (equal); supervision. YH: Resources; writing—review and editing (equal). AM: Investigation; writing—review and editing (equal). YS: Investigation; writing—review and editing (equal). SM: Investigation; writing—review and editing (equal). ST: Investigation; writing—review and editing (equal). SI: Resources; writing—review and editing (equal). YT: Resources; writing—review and editing (equal). HY: Resources; writing—review and editing (equal). TS: Writing—review and editing (equal); project administration.

### Funding Information

This research was supported by research funds from JSPS Grant-in-Aid for Early-Career Scientists Grant Number 21K18079 (to K.M.); by the General Insurance Association of Japan (to K.M.); and by the ZENKYOREN (National Mutual Insurance Federation of Agricultural Cooperatives) (to K.M.). No funding body had any role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Author Disclosure Statement

No competing financial interest exist.

### Supplementary Material

Supplementary Method  
Supplementary Table S1  
Supplementary Table S2  
Supplementary Table S3  
Supplementary Table S4

Supplementary Table S5  
Supplementary Table S6  
Supplementary Table S7  
Supplementary Table S8  
Supplementary Figure S1

### References

1. Maas AIR, Menon DK, Adelson PD, et al. Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. *Lancet Neurol* 2017;16(12):987–1048; doi: 10.1016/S1474-4422(17)30371-X
2. Kato Y, Narisawa A, Karibe H, et al. The age distribution of severe head injury in Japan Neurotrauma Data Bank: the changes among Project 1998, 2004, 2009, and 2015. *Neurotraumatology* 2019;42(2):160–167.
3. Rosenfeld JV, Maas AI, Bragge P, et al. Early management of severe traumatic brain injury. *Lancet* 2012;380(9847):1088–1098; doi: 10.1016/S0140-6736(12)60864-2
4. Basso A, Previgliano I, Servadei F. Traumatic brain injuries. In: *Neurological Disorders: Public Health Challenges*. (World Health Organization eds.) WHO Press: Geneva, Switzerland; 2006; pp. 164–173.
5. GBD 2016 Traumatic Brain Injury and Spinal Cord Injury Collaborators. Global, regional, and national burden of traumatic brain injury and spinal cord injury, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol* 2019;18(1):56–87; doi: 10.1016/S1474-4422(18)30415-0
6. Maas AI, Stocchetti N, Bullock R. Moderate and severe traumatic brain injury in adults. *Lancet Neurol* 2008;7(8):728–741; doi: 10.1016/S1474-4422(08)70164-9
7. CRASH-3 trial collaborators. Effects of tranexamic acid on death, disability, vascular occlusive events and other morbidities in patients with acute traumatic brain injury (CRASH-3): a randomised, placebo-controlled trial. *Lancet* 2019;394(10210):1713–1723; doi: 10.1016/S0140-6736(19)32233-0
8. Matsuo K, Aihara H, Nakai T, et al. Machine learning to predict in-hospital morbidity and mortality after traumatic brain injury. *J Neurotrauma* 2020;37(1):202–210; doi: 10.1089/neu.2018.6276
9. Amorim RL, Oliveira LM, Malbouissou LM, et al. Prediction of early TBI mortality using a machine learning approach in a LMIC population. *Front Neurol* 2020; 10:1366; doi: 10.3389/fneur.2019.01366
10. Abujaber A, Fadlalla A, Gammoh D, et al. Prediction of in-hospital mortality in patients with post traumatic brain injury using national trauma registry and machine learning approach. *Scand J Trauma Resusc Emerg Med* 2020;28(1):44; doi: 10.1186/s13049-020-00738-5
11. Warman PI, Seas A, Satyadev N, et al. Machine learning for predicting in-hospital mortality after traumatic brain injury in both high-income and low- and middle-income countries. *Neurosurgery* 2022;90(5):605–612; doi: 10.1227/neu.0000000000001898
12. Gravesteijn BY, Nieboer D, Ercole A, et al. Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury. *J Clin Epidemiol* 2020;122:95–107; doi: 10.1016/j.jclinepi.2020.03.005
13. Wang R, Wang L, Zhang J, et al. XGBoost machine learning algorithm performed better than regression models in predicting mortality of moderate-to-severe traumatic brain injury. *World Neurosurg* 2022;163:e617–e622; doi: 10.1016/j.wneu.2022.04.044
14. Adil SM, Elahi C, Patel DN, et al. Deep learning to predict traumatic brain injury outcomes in the low-resource setting. *World Neurosurg* 2022;164:e8–e16; doi: 10.1016/j.wneu.2022.02.097
15. Servia L, Montserrat N, Badia M, et al. Machine learning techniques for mortality prediction in critical traumatic patients: anatomic and physiologic variables from the RETRAUCI study. *BMC Med Res Methodol* 2020;20(1):262; doi: 10.1186/s12874-020-01151-3
16. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162(1):W1–W73; doi: 10.7326/M14-0698
17. Hawryluk GWJ, Rubiano AM, Totten AM, et al. Guidelines for the Management of Severe Traumatic Brain Injury: 2020 Update of the Decompressive Craniectomy Recommendations. *Neurosurgery* 2020;87(3):427–434; doi: 10.1093/neuros/nyaa278
18. Hawryluk GWJ, Aguilera S, Buki A, et al. A management algorithm for patients with intracranial pressure monitoring: the Seattle International Severe Traumatic Brain Injury Consensus Conference (SIBICC). *Intensive Care Med* 2019;45(12):1783–1794; doi: 10.1007/s00134-019-05805-9
19. Maas AI, Hukkelhoven CW, Marshall LF, et al. Prediction of outcome in traumatic brain injury with computed tomographic characteristics: a comparison between the computed tomographic classification and

- combinations of computed tomographic predictors. *Neurosurgery* 2005;57(6):1173–1182; doi: 10.1227/01.neu.0000186013.63046.6b
20. Wilson JT, Pettigrew LE, Teasdale GM. Structured interviews for the Glasgow Outcome Scale and the extended Glasgow Outcome Scale: guidelines for their use. *J Neurotrauma* 1998;15(8):573–585; doi: 10.1089/neu.1998.15.5730
  21. Wilson L, Boase K, Nelson LD, et al. A manual for the Glasgow Outcome Scale-Extended interview. *J Neurotrauma* 2021; 38(17):2435–2446; doi: 10.1089/neu.2020.7527
  22. Troyanskaya O, Cantor M, Sherlock G, et al. Missing value estimation methods for DNA microarrays. *Bioinformatics* 2001;17(6):520–525; doi: 10.1093/bioinformatics/17.6.520
  23. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; 2016; doi: 10.1145/2939672.2939785
  24. Farre P, Heurteau A, Cuvier O, et al. Dense neural networks for predicting chromatin conformation. *BMC Bioinformatics* 2018;19(1):372; doi: 10.1186/s12859-018-2286-z
  25. Bergstra J, Komer B, Eliasmith C, et al. Hyperopt: a Python library for model selection and hyperparameter optimization. *Comput Sci Discov* 2015;8(1):014008; doi: 10.1088/1749-4699/8/1/014008
  26. Hand DJ, Till RJ. A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning* 2001;45:171–186; doi: 10.1023/A:1010920819831
  27. Petrucci CJ. A primer for social worker researchers on how to conduct a multinomial logistic Rregression. *J Soc Serv Res* 2009;35(2):193–205.
  28. Hosmer Jr DW, Lemeshow S, Sturdivant RX. *Applied logistic regression* (Vol. 398). John Wiley & Sons: Hoboken, NJ.; 2013.
  29. Lemeshow S, Hosmer Jr DW. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 1982;115(1):92–106; doi: 10.1093/oxfordjournals.aje.a113284
  30. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44(3):837–845.
  31. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12:77; doi: 10.1186/1471-2105-12-77.
  32. Satyadev N, Warman PI, Seas A, et al. Machine learning for predicting discharge disposition after traumatic brain injury. *Neurosurgery* 2022;90(6):768–774; doi: 10.1227/neu.0000000000001911
  33. Rau CS, Kuo PJ, Chien PC, et al. Mortality prediction in patients with isolated moderate and severe traumatic brain injury using machine learning models. *PLoS One* 2018;13(11):e0207192; doi: 10.1371/journal.pone.0207192
  34. Shi HY, Hwang SL, Lee KT, et al. In-hospital mortality after traumatic brain injury surgery: a nationwide population-based comparison of mortality predictors used in artificial neural network and logistic regression models. *J Neurosurg* 2013;118(4):746–752; doi: 10.3171/2013.1.JNS121130
  35. Rughani AI, Dumont TM, Lu Z, et al. Use of an artificial neural network to predict head injury outcome. *J Neurosurg* 2010;113(3):585–590; doi: 10.3171/2009.11.JNS09857
  36. Tu KC, Eric Nyam TT, Wang CC, et al. A computer-assisted system for early mortality risk prediction in patients with traumatic brain injury using artificial intelligence algorithms in emergency room triage. *Brain Sci* 2022;12(5):612; doi: 10.3390/brainsci12050612
  37. Hernandez Rocha TA, Elahi C, Cristina da Silva N, et al. A traumatic brain injury prognostic model to support in-hospital triage in a low-income country: a machine learning-based approach. *J Neurosurg* 2019;132(6):1961–1969; doi: 10.3171/2019.2.JNS182098
  38. Pease M, Arefan D, Barber J, et al. Outcome prediction in patients with severe traumatic brain injury using deep learning from head CT scans. *Radiology* 2022;304(2):385–394; doi: 10.1148/radiol.212181
  39. Christodoulou E, Ma J, Collins GS, et al. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol* 2019;110:12–22; doi: 10.1016/j.jclinepi.2019.02.004
  40. van der Ploeg T, Nieboer D, Steyerberg EW. Modern modeling techniques had limited external validity in predicting mortality from traumatic brain injury. *J Clin Epidemiol* 2016;78:83–89; doi: 10.1016/j.jclinepi.2016.03.002
  41. Maegele M, Schöchl H, Menovsky T, et al. Coagulopathy and haemorrhagic progression in traumatic brain injury: advances in mechanisms, diagnosis, and management. *Lancet Neurol* 2017;16(8):630–647; doi: 10.1016/S1474-4422(17)30197-7
  42. Epstein DS, Mitra B, O'Reilly G, et al. Acute traumatic coagulopathy in the setting of isolated traumatic brain injury: a systematic review and meta-analysis. *Injury* 2014;45(5): 819–824; doi: 10.1016/j.injury.2014.01.011
  43. Haas T, Fries D, Tanaka KA, et al. Usefulness of standard plasma coagulation tests in the management of perioperative coagulopathic bleeding: is there any evidence? *Br J Anaesth* 2015;114(2):217–224; doi: 10.1093/bja/aeu303
  44. Saggat V, Mittal RS, Vyas MC. Hemostatic abnormalities in patients with closed head injuries and their role in predicting early mortality. *J Neurotrauma* 2009;26(10): 1665–1668; doi: 10.1089/neu.2008.0799
  45. Maas AIR, Menon DK, Manley GT, et al. Traumatic brain injury: progress and challenges in prevention, clinical care, and research. *Lancet Neurol* 2022;21(11):1004–1060; doi: 10.1016/s1474-4422(22)00309-x
  46. McCrea MA, Giacino JT, Barber J, et al. Functional outcomes over the first year after moderate to severe traumatic brain injury in the prospective, longitudinal TRACK-TBI study. *JAMA Neurol* 2021;78(8):982–992; doi: 10.1001/jamaneurol.2021.2043
  47. Suehiro E, Kiyohira M, Haji K, et al. Changes in outcomes after discharge from an acute hospital in severe traumatic brain injury. *Neurol Med Chir (Tokyo)* 2022;62(3):111–117; doi: 10.2176/nmc.oa.2021-0217
  48. Lu J, Marmarou A, Lapane K, et al. A method for reducing misclassification in the extended Glasgow Outcome Score. *J Neurotrauma* 2010;27(5):843–852; doi: 10.1089/neu.2010.1293
  49. Ritchie H, Roser M. OurWorldInData.org. Age Structure; 2019. Available from: <https://ourworldindata.org/age-structure>. [Last accessed: March 3, 2023].
  50. Okada Y, Kiguchi T, Iiduka R, et al. Association between the Japan Coma Scale scores at the scene of injury and in-hospital outcomes in trauma patients: an analysis from the nationwide trauma database in Japan. *BMJ open* 2019;9(7):e029706; doi: 10.1136/bmjopen-2019-029706
  51. Obermeyer Z, Powers B, Vogeli C, et al. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019;366(6464):447–453; doi: 10.1126/science.aax234