Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/MBE

MBE, 20(1): 837–858.
DOI: 10.3934/mbe.2023038
Received: 19 August 2022
Revised: 29 September 2022
Accepted: 08 October 2022
Published: 17 October 2022

Research article

A data-driven medical knowledge discovery framework to predict the length of ICU stay for patients undergoing craniotomy based on electronic medical records

Shaobo Wang[1,3], Jun Li[2], Qiqi Wang[3], Zengtao Jiao[3], Jun Yan[3], Youjun Liu[1] and Rongguo Yu[2,*]

[1] College of Life Science and Bioengineering, Beijing University of Technology, Beijing, China
[2] Surgical Intensive Care Unit, Fujian Provincial Hospital, Fujian, China
[3] Yidu Cloud (Beijing) Technology Co. Ltd., Beijing, China

* **Correspondence:** Email: rongguoyu01@163.com; Tel: (+86)13805089943; Fax: (+86)059187557768.

Abstract: Craniotomy is an invasive operation with great trauma and many complications, and patients undergoing craniotomy should enter the ICU for monitoring and treatment. Based on electronic medical records (EMR), the discovery of high-risk multi-biomarkers rather than a single biomarker that may affect the length of ICU stay (LoICUS) can provide better decision-making or intervention suggestions for clinicians in ICU to reduce the high medical expenses of these patients and the medical burden as much as possible. The multi-biomarkers or medical decision rules can be discovered according to some interpretable predictive models, such as tree-based methods. Our study aimed to develop an interpretable framework based on real-world EMRs to predict the LoICUS and discover some high-risk medical rules of patients undergoing craniotomy. The EMR datasets of patients undergoing craniotomy in ICU were separated into preoperative and postoperative features. The paper proposes a framework called Rules-TabNet (RTN) based on the datasets. RTN is a rule-based classification model. High-risk medical rules can be discovered from RTN, and a risk analysis process is implemented to validate the rules discovered by RTN. The performance of the postoperative model was considerably better than that of the preoperative model. The postoperative RTN model had a better performance compared with the baseline model and achieved an accuracy of 0.76 and an AUC of 0.85 for the task. Twenty-four key decision rules that may have impact on the LoICUS of patients undergoing craniotomy are discovered and validated by our framework. The proposed postoperative RTN model in our framework can precisely predict whether the patients undergoing craniotomy are hospitalized for too long (more than 15 days) in the ICU. We also discovered and validated some key medical decision rules from our framework.

Keywords: craniotomy; electronic medical records; length of ICU stay; predictive model; TabNet; decision rules discovery

1. Introduction

Craniotomy is a common treatment of neurosurgical diseases. Patients undergoing craniotomy should enter the ICU for monitoring and treatment because of surgical trauma and the development of postoperative complications [1]. The increase in the length of ICU stay (LoICUS) of these patients results in high medical costs while suffering from physical pain. LoICUS is a common outcome used as an indicator of quality of care and resource use [2]. Therefore, from the perspective of personnel cost and resource management, predicting the LoICUS, especially for patients with long ICU hospitalization is an inevitable task, as the length of stay (LOS) outliers accounted for most ICU cost [3].

The vigorous development of machine learning (ML) or deep learning (DL) in the past decade has made ML- or DL-based predictive models (PM) widely used in various health care areas, which proves the feasibility of ML or DL methods in this field [4]. However, based on different problem definitions and medical scenarios, whether such methods are better than traditional models is a controversial issue. The prediction of the onset of diseases or medical outcomes, such as mortality, based on electronic medical records (EMRs) using DL or ML has been widely used in recent years [5]. ML- or DL-based methods are often implemented to ICU data for its rarity and value to predict outcomes [6]. Many scholars have focused on LoICUS prediction. Recent studies translated LoICUS prediction into a classification task but not a regression task. For example, Gentimis et al. [7] believed that day 5 has the greatest impact on LoICUS, and they made a classification model of whether the LoICUS is greater than 5 days based on neural network (NN) algorithm with 80% accuracy. In 2019, Harutyunyan et al. [8] made a PM to predict whether the LoICUS is greater than 7 days based on linear regression model and long short-term memory, and the accuracy of the classification model was 84%. They also noted that the task of LoICUS prediction is more difficult than other outcome prediction tasks, such as in-hospital mortality prediction, and they believed that predicting whether a patient would have an extended LoICUS (longer than 7 days) from only the first 24 hours of data might be more reasonable than other time points [8]. In 2021, Khalid et al. [9] applied six ML-based PMs for binary classification using the median patient population ICU stay of 2.64 days, and they considered that random forest (RF) is a good predictive model in this task with 65% accuracy. In conclusion, based on the research above and the sample size of our study, we decided to apply the classification method (classification boundary is the median of LoICUS) to predict the LoICUS in our research.

In recent years, the decision-making process of the PM is always the key point in interpretable learning, especially in the area of medical research. Miller defined interpretability as “the degree to which a human can understand the cause of a decision” [10]. In the field of medical research, such as the prediction tasks of medical outcome, establishing an interpretable model instead of a black-box one is more acceptable.

Gradient boosting-based tree models, such as gradient boosting decision tree (GBDT), XGBoost, or LightGBM [11,12], have presented many achievements and unusually brilliant results in data science competition, such as Kaggle, especially in the field of tabular data competition. Through this type of algorithm, we can obtain the decision-making process of the model conveniently in the form of decision rules, which are also known as cross features or feature interactions. In medical research, these rules can be understood as the combination of some clinical variables or biomarkers called multi-biomarkers. Previous studies have shown a higher prognostic accuracy using multi-biomarkers than an individual one [13]. Therefore, we believe that discovering some key decision rules or multi-biomarkers instead of a single biomarker according to self-interpretable methods, such as RF or GBDT, is a helpful method for medical workers.

In our study, we propose a supervised self-interpretable medical knowledge discovery framework, including a PM called Rules-TabNet (RTN), to predict the LoICUS, aiming to discover some medical rules that may affect the LoICUS of a patient undergoing craniotomy. RTN consists of a gradient boosting-based tree model to generate decision rules and a TabNet, the self-interpretable and NN-based model proposed by Sercan et al. [14] in 2019, to select some important decision rules. We also validated the medical rules discovered by RTN through risk analysis.

The rest of our paper is organized as follows. In the next section, the materials of our study and the methodology of our framework are described. In the third section, we present the statistics and experimental results of our study. The discussion and conclusion are stated in the last section.

2. Materials and methods

2.1. Task modeling

We aimed to discover some medical rules that may affect the LoICUS of a patient undergoing craniotomy according to our framework. We handled the prediction and discovery tasks based on the following steps, and the framework is shown in Figure 1.

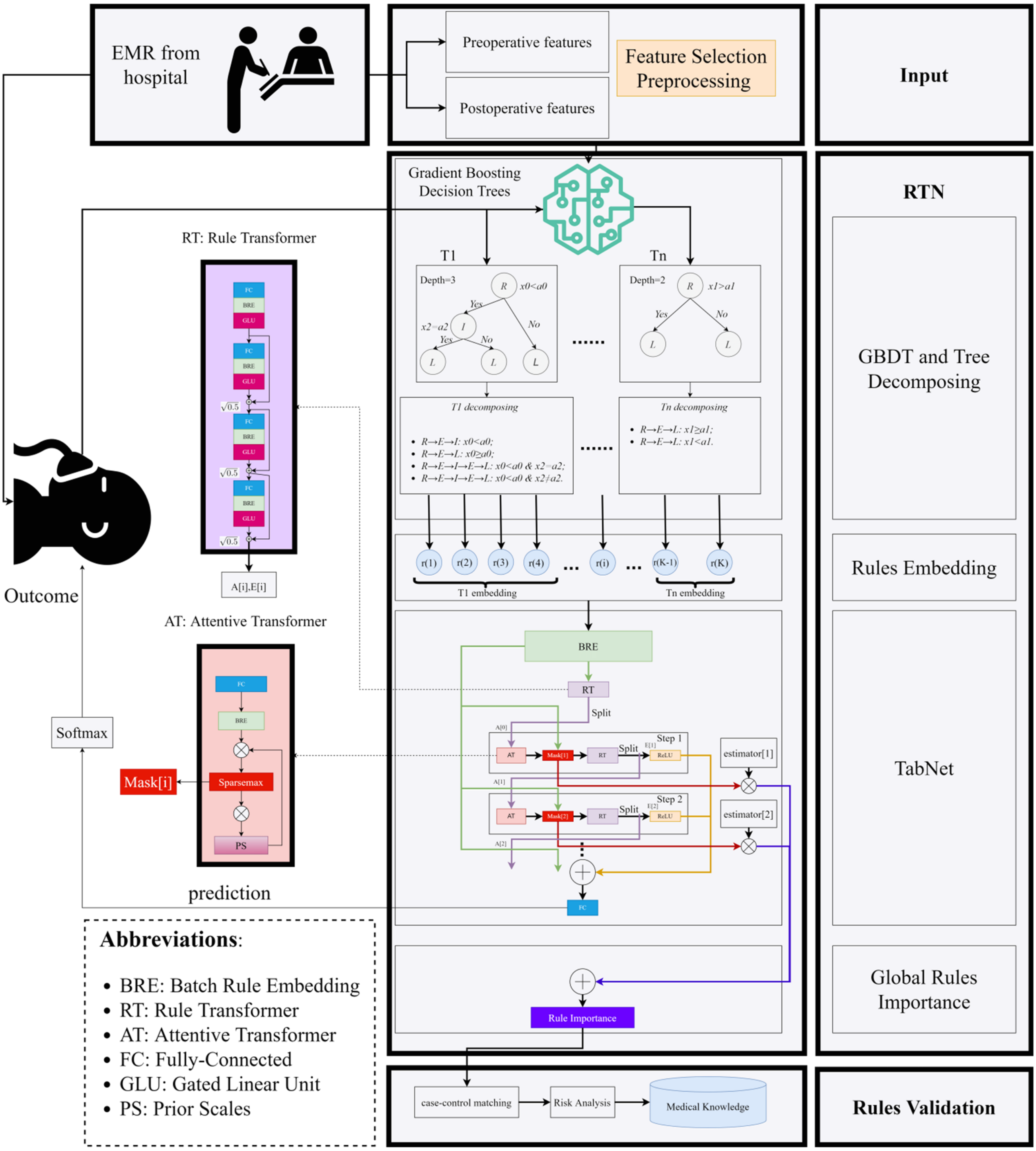


Figure 1. Study framework. The framework has three parts. The Input module contains data preprocessing and feature selection. The RTN module has four steps: the construction of GBDT and Tree Decomposing, Rules Embedding, TabNet construction, and Global Rules importance computation. The last part of the framework is Rules Validation.

- First, the dataset of patients undergoing craniotomy from real-world data is retrieved from the

EMR. The data is dichotomized according to the median of LoICUS, which is also the outcome of our study. Positive examples are the patients with longer LoICUS (above the median) after craniotomy, and negative examples are the patients with shorter LoICUS (less than the median). The data was entered into the Input module, and the preoperative and postoperative features were obtained after data preprocessing and feature selection.

- Second, we trained a supervised gradient boosting-based classification model, and the label is the outcome using the dataset of Step 1. After the hyperparameters of the model were tuned, we constructed a bunch of decision trees based on the model. Some coarse-grained rules can be generated by the tree-decomposing process. We then applied a rule-embedding method to generate rule features. A TabNet model was trained according to the rule features. TabNet has two major modules: Rule Transformer (RT) and Attentive Transformer (AT). The output of TabNet was entered into a Softmax for outcome prediction. During the training process of TabNet, the global importance of these rules can be calculated to evaluate them.
- Finally, a rule validation step, which includes case–control matching and risk analysis, is subsequently applied to validate the rules to analyze some risk factors that might affect the LoICUS of patients undergoing craniotomy.

2.2. Data retrieval

In this study, we retrospectively retrieved the EMR from the surgical ICU data platform of a hospital of China from 2005 to 2018. The variables were extracted from the surgical ICU special disease database, and the demographic information, vital signs, laboratory diagnosis, and other variables of patients from 6 hours before to 24 hours after entering the ICU were used as candidate variables for modeling. During the ICU stay of the patients, some variables, such as vital signs (blood pressure, heart rate, etc.) and laboratory results (platelets, white blood cells, etc.) were continuously changing in a short time. Therefore, these dynamic variables were represented by the first (the first collected value in ICU), maximum (the maximum value of the variable during ICU stay), and minimum (the minimum value of the variable during ICU stay) values of patients in the ICU as shown in Table S1.

Adult patients (age ≥ 18 years) undergoing craniotomy and requiring ICU treatments were recruited according to the inclusion and exclusion criteria of our study. We only retained data for at least one in-ICU test. For patients who entered the ICU multiple times, we only collected the data of the first ICU admission. Patients whose LoICUS is null or incorrectly recorded, too long (more than 365 days), or too short (less than 24 hours) were excluded. Based on the above criteria, we also excluded patients with an in-hospital outcome of death and patients whose discharge status is automatic discharge. The automatically discharged patient might die after discharge, and the therapeutic importance for such patient and his/her family members is little. Moreover, family members are unwilling to let the patient die in the hospital because of local culture and customs. In our previous study, we made a PM to predict the mortality of patients undergoing craniotomy in the ICU and discovered some high-risk factors (such as heart rate, temperature, etc.) that are closely related to the mortality of patients undergoing craniotomy [15]. Therefore, we excluded the dead samples in our dataset, which means that we removed the "death situation" in input features to reduce the impact of confounding factors. Besides, we performed an odds ratio (OR) test to analyze the relation between death situation and the LoICUS of such patients and a chi-square test to validate the OR and its 95% confidence interval (CI). We also tested the sensitivity of death situation to LoICUS. The main outcome of our study is the LoICUS. Through the investigation and other research [9], we divided the LoICUS into two categories based on the median LoICUS (15 days) to build a classification model, and we also made some regression models.

This study was approved by the Ethics Committee of Fujian Provincial Hospital, and all procedures performed in this study involving human participants were in accordance with its ethical standards. This study obtained the informed consent of all the participants.

2.3. Statistical analysis

After data retrieval, we selected the variables with filling rate greater than 70%. A total of 146 variables consisting of 20 preoperative and 126 postoperative variables were extracted from the data platform. The detailed information and statistical results are shown in Table S1.

In consideration of the feature engineering technology in ML, selecting all the variables as the input matrix for the predictive model is not a wise option, but variables related to clinical practices and medical facts need to be taken into account. We divided the patients into two groups according to whether the LoICUS is more than 15 days, which means the endpoint of our study is LoICUS groups. We made assumptions to see whether a variable is a statistical difference between the two groups. Chi-square test was applied to discrete variables, such as gender, and student t-test was used for continuous variables that conform to normal distribution in both groups (group 1: LoICUS $\geq$ 15 days; group 2: LoICUS $<$ 15 days). Otherwise, Wilcoxon test was used to judge the statistical difference between two groups. For discrete variables, we applied Yates's correction for chi-square tests according to the number of samples and frequency of variables. We performed 100 permutation tests to correct the multiple testing for each continuous variable, and adjusted $P < .05$ was used as the standard to express statistical significance. Feature selection was conducted according to the adjusted P-value, and variables related to clinical practices and medical facts were also considered.

The imputation of missing value is an inevitable process for many predictive models. Missing values were imputed by mean for continuous variables and mode for discrete variables.

The PMs were constructed using the scikit-learn (version 0.23.0) and PyTorch (version 1.3.1) packages in Python (version 3.7.3 Python Software Foundation), and statistical analysis was conducted using the SciPy (version 1.5.0) package in Python (version 3.7.3). The statistical values and ranges of different variables were different. Continuous variables that conformed to normal distribution were analyzed by mean and standard deviation (mean ± std), whereas other continuous variables were analyzed using median and interquartile range (IQR). The discrete variables are presented as the percentages of positive examples (%).

2.4. RTN model

2.4.1. Rules generated by gradient boosting-based tree model

A GBDT algorithm was adopted to train the predictive model by generating multiple additive trees because of the good interpretability and advantage in training time of tree-based models compared with NN-based black-box models,.

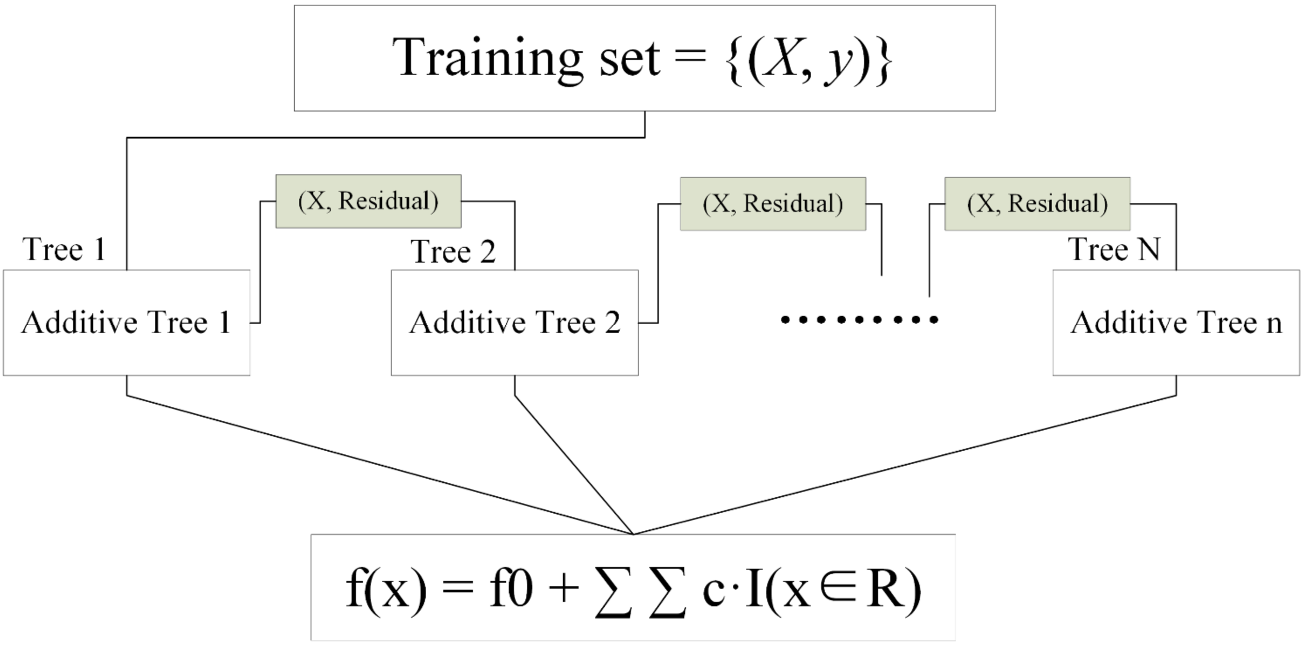


Figure 2. Framework of the gradient boosting-based tree model. *N* additive trees that are not independent are trained, and the later tree (Tree_i) learns the residual error based on the previous one (Tree_{i-1}), which means (X_m, Residual_{i-1}) is the input of Tree_i. f_0 is the initial value. R is the terminal region to each additive tree, and c is the minimum square loss of R. I represents the indicating function. $f(x)$ is the output.

Figure 2 illustrates the framework of the GBDT model. The first step is to initialize $f_0(X) = \arg\min \sum_{i=1}^{m} L(y_i, c)$. N additive trees are trained iteratively. The input of the later tree (Tree_i) is (X_m, Residual_{i-1}). The residual is simulated approximately by negative gradient. The terminal region corresponding to each tree n is $R_{nj}, j = 1, 2, \ldots, J$, where J is the number of leaf nodes. As for the J leaf nodes, $c_{nj} = \arg\min \sum_{X_i \in R_{nj}} L(y_i, f_{n-1}(X_i) + c)$ is the minimum loss of R_{nj}. $I(X \in R_{nj})$ is the indicating function, $I = 1$ when $X \in R_{nj}$; otherwise, $I = 0$. The output of the model is shown in Eq (1). The final classification result is $y = 1 / \{1 + \exp[-f(X)]\}$.

$$f(X) = f_0(x) + \sum_{n=1}^{N} \sum_{j=1}^{J} c_{nj} I(x \epsilon R_{nj}) \tag{1}$$

After the parameter tuning of the model, we generated the medical rules from the model. All the additive trees in the model can be denoted as decision trees, and a tree model can be represented as nodes (N_{R} is the root node, N_{I} is the internal node, and N_{L} is the leaf node) and edges (E), $T_i = \{N_{\mathrm{R}}, N_{\mathrm{I}}, N_{\mathrm{L}}, E\}$. We decomposed the trees into decision rules by connecting N_{I} or N_{L} by E from N_R. Any path through the N_R can be converted into a decision rule, and the coarse-grained medical rule base R_K is obtained and consisted of K rules. The strategy in [16] was adopted to limit the tree size to reduce the complexity of the model. If the maximum depth of the model is set to 2 and the three types of rules in R_K as shown in Eq (2). N is the number of trees, K is the number of rules derived from the model, and t_n is the number of N_{L} within Tree_n.

$$R_K = \begin{cases} N_R \rightarrow E \rightarrow N_I \\ N_R \rightarrow E \rightarrow N_I \rightarrow E \rightarrow N_L \, ; K = \sum_{n=1}^{N} 2(t_n - 1) \\ N_R \rightarrow E \rightarrow N_L \end{cases} \tag{2}$$

We take an example to illustrate the process of tree decomposing. As shown in Figure 1, $T_1 = \{1 \times N_{\mathrm{R}}, 1 \times N_{\mathrm{I}}, 3 \times N_{\mathrm{L}}, 4 \times E\}$. The tree has 3 layers, and four rules can be generated as follows:

- $N_{\mathrm{R}} \rightarrow E \rightarrow N_{\mathrm{I}}$: x0 < a0;
- $N_{\mathrm{R}} \rightarrow E \rightarrow N_{\mathrm{L}}$: x0 ≥ a0;
- $N_{\mathrm{R}} \rightarrow E \rightarrow N_{\mathrm{I}} \rightarrow E \rightarrow N_{\mathrm{L}}$: x0 < a0 & x2 = a2;
- $N_{\mathrm{R}} \rightarrow E \rightarrow N_{\mathrm{I}} \rightarrow E \rightarrow N_{\mathrm{L}} \rightarrow$: x0 < a0 & x2 ≠ a2.

x0 denotes the split node of the first layer, and x2 denotes the split node of the second layer. a0 and a2 represent the range of split nodes of x0 and x1, respectively.

Given the K coarse-grained medical rules derived by GBDT, we transfered each rule R in R_K into an embedding vector $r(i)_{1\times n} = [r_1, r_2, \dots, r_m, \dots, r_n]$, where n is the embedding size, also known as the samples of the dataset, and r_m is the embedding representation of the m-th patient. The rule features of the data can be denoted as $X_{n\times K}$.

$$r_m^k = \prod_{j=1}^{p} \hat{I}[r_m^k(j)] \tag{3}$$

Rule R can also be described as a cross feature, which meaning that R had p features, such as x0 and x2 of r (1) in Figure 1. $\hat{I}(.)$ is the indicator function and is 0 or 1 like $I(.)$ in [17]. r_m^k is the representation of the k_{th} rule in r_m as shown in Eq (3).

2.4.2. TabNet and global rule importance

We applied a TabNet encoder, which is an additive model consisting of several steps to predict the outcome and compute the global importance of all rules. The embedding vector of all rules was input into the TabNet, and we merely selected a batch rule embedding (BRE), $X_{B\times K}^{\mathrm{BRE}}$, into the rule transformer (RT) and every step, where B is the batch size.

The RT module has four F-B-G layers, and each of which is composed of a fully-connected (FC) layer, BRE, and a gated linear unit (GLU) [18]. A skip-connection process was applied to each F-B-G, and a normalization with $\sqrt{0.5}$ after the last three F-B-G was used to stabilize the learning process of the network [19]. $(A[i], E[i])$ is the output of RT after splitting in step_i, $A[i]$ is the input of the fellow module, and $E[i]$ is used to generate the outcome of step_i.

$A[i]$ is the input of attentive transformer (AT), and the $\mathrm{Mask}[i]$ module of step_i, which is shown in Eq (4), was employed for rule selection. $\text{Mask}[i]$ has the same dimension as $X_{B\times K}^{\text{BRE}}(i)$, and the value range of $\text{Mask}[i]$ is [0,1].

$$\text{Mask}[i] = \text{Sparemax}\left(\prod_{j=1}^{i-1}(\gamma - \text{Mask}[j]) \cdot h_i(A[i-1])\right), \text{Mask[i]} \in [0,1] \tag{4}$$

Compared with Softmax, Sparemax is a normalization method that can obtain more sparse results [20]. γ is the relaxation parameter, which means that the feature can be used again with higher weight in subsequent steps if $\gamma > 1$. $h_i(A[i-1])$ is the output of the F-B of AT.

After the AT to $\text{Mask}[i]$ to RT in step_i, $A[i]$ was input into the AT of step_{i+1}, and $E[i]$ was input into the ReLU of step_i to get the output of the i-th estimator.

Finally, the output of RTN is the sum of the results of all estimators, and the output of the i-th estimator is $\text{ReLU}(E[i])$. The final output of RTN is shown in Eq (5). Softmax was used to classify the LoICUS of patients undergoing craniotomy.

$$\text{RTN}_{\text{output}} = FC\left[\sum_{i=1}^{N} \text{ReLU}(E[i])\right] \tag{5}$$

The GBDT architect and Mask module in our study make RTN have a stronger self-interpretability compared with the traditional NN-based black-box model. The calculation method of rule importance is described below.

As for a sample b of step_i, $E_b[i]$ is one of the outputs of RT, and the dimension of $E_b[i]$ is $B \times N_E$. The output of step_i is $(E[i]) = 0$ when $E_{b,j}[i] \leq 0$, where j is a one-dimensional rule-feature in N_E. The contribution of sample b in step_i $c_b[i]$ is shown in Eq (6)

$$c_b[i] = \sum_{j=1}^{N_E} ReLU(E_{b,j}[i]) \tag{6}$$

The larger the $c_b[i]$, the more obvious its impact on the outcome. $c_b[i]$ is also known as the weight of step_i in RTN. Therefore, the global importance of rule j in sample b is the sum of the weights of Masks in all steps. The normalized representation of rule importance if RTN has N steps is shown in Eq (7).

$$I_{b,j} = \frac{\sum_{i=1}^{N} c_b[i] \cdot \text{Mask}_{b,j}[i]}{\sum_{j=1}^{K}\sum_{i=1}^{N} c_b[i] \cdot \text{Mask}_{b,j}[i]} \tag{7}$$

2.5. Rule validation

According to the global importance of all rules computed by RTN, the rules R_{RTN} with nonzero global importance will be left for the subsequent validation. For each rule $r(f)$ in R_{RTN}, f represents the feature set contained in rule *r*. We adopted an analogous case-control strategy, which is similar to the case-control study that is commonly used in medical research to construct the validation data of the risk factor. We performed a risk analysis to validate the correlation between the rule and outcome.

Case-control study is widely used in risk factor detection [21]. It is very unfriendly, especially for small sample data, such as in our research, because of the precise inclusion and exclusion criteria of such study. We conducted an analogous case-control matching method of each rule automatically in terms of Propensity Score Matching (PSM), which is a statistical method that deals with biases and confounding factors [22]. We constructed two groups for each rule: group A (patients complying to $r(f)$) and group B (patients not complying to $r(f)$). Propensity score (PS) was calculated for each patient in terms of the predicted probability of the logistic regression (LR) model shown in Eq (8). F is the representation of all original feature set, and $(\mathrm{F}-f)$ means the different sets of the original feature set and the features contained in the rule $r(f)$.

$$\mathrm{PS} = \frac{1}{1+e^{-\beta(\mathrm{F}-f)}} \tag{8}$$

Two patients from groups A and B with the closest PS were selected and entered to a new patient cluster. Case and control groups were generated from the new cluster according to the outcome [23]. The number of people in the case group is at least 100, and the number of people in the control group is greater than or equal to that in the case group.

The OR and its 95% CI were computed to analyze the risk of each rule. Here, we made an example to illustrate the process of risk analysis. As for a rule $r(f) = \min \mathrm{PCT} \leq 0.16\,\&\,\mathrm{tracheotomy}$ in R_{RTN}, which means the patient has a minimum PCT of less than 0.16 and a tracheotomy is either performed. We built a fourfold table to calculate the OR value as shown in Table 1. The case–control pair of the rule will be divided into four parts: patients that comply with $r(f)$ and the outcome is LoICUS ≥ 15 (A); patients that do not comply with $r(f)$ and the outcome is LoICUS ≥ 15 (B); patients comply with $r(f)$ and the outcome is LoICUS < 15 (C); and patients do that not comply with $r(f)$ and the outcome is LoICUS < 15 (D). The calculation process of the 95% CI of the OR is shown in Eq (9).

Table 1. Fourfold table of $\boldsymbol{r(f)}$.

	Case	Control	Total
LoICUS ≥ 15	A = 119	B = 73	192
LoICUS < 15	C = 22	D = 68	90
Total	141	141	282

$$\mathrm{OR} = \frac{A/B}{C/D} = 5.04;\ 95\%\ \mathrm{CI} = e^{\ln(\mathrm{OR}) \pm (1.96\sqrt{\frac{1}{A}+\frac{1}{B}+\frac{1}{C}+\frac{1}{D}})} = (2.87, 8.84) \tag{9}$$

We eliminated the rules with vague OR (the left and right CIs of OR strides across 1). The higher the OR corresponding to a rule, the higher the risk impact of the rule on the LoICUS.

2.6. Experimental settings

We separated the experiments into two groups (preoperative group and postoperative group) according to the operation time of the patients. We split the dataset into training and testing sets in the proportion of 9:1, and a 10-fold cross validation is performed on the training set in our experiment to find the optimized hyperparameters, and the optimized hyperparameters of models are shown in Table S3. After rule-embedding, in the process of DL model training, we split the rule-features in to training, validation, and testing set in the proportion of 8:1:1. The classification result is measured by accuracy, precision, recall, f1-score, and AUC of the receiver operating characteristic (ROC) curve.

There are two types of parameters in RTN, including the parameters of tree-based model and the hyper-parameters of TabNet. Table 2 presents the parameters applied to preoperative model and postoperative model.

Specifically, as for tree-based model, the learning rate is tuned first to increase the convergence speed. $n\_estimators$ denotes the number of additive trees that are trained to generate the rules, max_depth is the limitation of the depth of the tree model, $min\_samples\_split$ represents the minimum number of samples required to split N_{I}, $min\_samples\_leaf$ means the minimum number of samples required to be at a N_{L}, and $max\_features$ is the number of features to consider when looking for the best split.

As for the hyperparameters for TabNet, gamma is the coefficient for feature reusage in $\mathrm{Mask}[i]$ to increase the attention of features that the model has not been focused on in the next step. N_A and N_E are the dimensions of the outputs of RT, and $N_A = N_E$ is usually a good choice. n_steps means the number of steps in the architecture. TabNet was trained using gradient descent-based optimization, the optimizer is Adaptive Moment Estimation (Adam), and the learning rates are 0.02 for the postoperative model and 0.001 for the preoperative model. All rule-features were mapped to a single-dimensional trainable scalar with a learnable embedding, the loss function of the classification is softmax cross entropy, and the training process will stop until convergence. All the hyperparameters of TabNet were optimized on the validation set.

Table 2. Parameters of RTN in the experiments.

	Parameters	Preoperative model	Postoperative model
Tree-based model			
	learning rate	0.1	0.01
	n_estimators	20	40
	max_depth	2	2
	min_samples_split	124	158
	min_samples_leaf	9	57
	max_features	'sqrt'	'sqrt'
TabNet model			
	batch size	100	100
	epochs	71	48
	Gamma	1.8	1.4
	$\mathrm{N_A}$	56	40
	$\mathrm{N_E}$	56	40
	n_steps	9	6

We constructed some baseline models to prove the effectiveness of our model. In the rule generation process, we implemented two self-interpretable models, namely, LR and RF, which can generate some rules from the baseline model compared with GBDT. LR is widely used in many medical research and is considered a baseline model for classification task [24,25]. Ensembling methods, such as RF and GBDT, are a good choice for tabular data in recent years. We applied some evaluation metrics, including the area under the receiver operating characteristic curve (AUC), accuracy, precision, recall, and F1 score to evaluate the performance of the rule generation model.

RuleFit (GBDT based) was used as the baseline model compared with RTN in the fitting of the rule-based predictive model. RuleFit is a composite and self-interpretable model that can fit the rules generated by a tree-based model according to a linear model. It is a compound model that includes a rule generation model (GBDT) and a LR to screen the rules by global importance. The structure of RuleFit, including rule generation and rule fitting, has high similarity with our model. Therefore, we chose RuleFit as our main comparison model.

We also performed supplementary regression experiments to predict LoICUS using some regression models. We implemented some traditional models, such as linear regression, poisson regression, and hurdle regression, and some ML-based models, such as random forest regressor (RFR), gradient boosting regressor (GBR), and RuleFit (GBR based), to predict the number of days. Fitting predictive models for a dichotomous endpoint is much easier then predicting for time to event outcomes. Thus, we set supplementary experiments to illustrate the difficulty of regression in our tasks and why we defined the LoICU prediction task as a classification problem.

3. Results

3.1. Statistical results

The results of OR test to analyze the relation between death situation and the LoICUS (OR = 0.78 [0.46, 1.32], p = 0.421) revealed no significant difference between the death situation and LoICUS of such patients. Sixty-one death samples were excluded by our criteria from the original data. We focused on these data and extracted 61 non-death samples from the collected data through down-sampling. Based on the 122 samples, we trained a traditional LR model. The input features were consistent with the feature selection results of this study. The outcome is whether the LoICUS is greater than 15, and the evaluation metrics were defined as the AUC, sensitivity, and specificity of the model. We only focused on the model performance of the 61 death samples but did not split the data into training and testing sets. The experiment results are shown in Table S5. We compared whether the “death situation” was added to the model as a feature. We found that the “death situation” did not increase the sensitivity of the model but decreased the performance of the model. Hence, we excluded the dead samples in our study.

After data collection, 631 non-hospital dead patients (63.87% male) were enrolled into the study. The average age of the patients was 55.94 ± 15.15 years. The statistical information of the LoICUS is shown in Table 3. The positive examples are the patients with a LoICUS of more than 15 days after craniotomy (group one: 319 patients), and the negative examples are patients with a LoICUS of less than 15 days (group two: 312 patients). A total of 146 variables were extracted from the dataset. The variables were divided into the following categories: demographic information, previous medical history, brain injury inducements, infection sources, vital signs, Glasgow Coma Scale (GCS) score, hematoma properties, laboratory test, and therapeutic index.

The feature selection procedure revealed that 91 variables are statistically significant as shown in Table S1. The bold parts indicate the variables that are significantly different from the outcome. Six of them are preoperative variables (age, epilepsy, cerebral contusion and laceration, brain tumor, vascular diseases, and intracerebral hematoma), and 85 of them are postoperative variables. Patients in group one (58.14 ± 15.44) were older than those in group two (53.69 ± 14.53), and the two groups had a remarkable difference in age. The statistical results revealed that the LoICUS of men is longer than that of women. The statistically substantial variables are distributed in GCS scores, vital signs, and laboratory test indicators besides age. The detailed statistical results of the dataset are shown in Table S2.

Table 3. Statistical information of LoICUS.

Statistical information	Days
Mean of LoICUS	19.87
Std of LoICUS	19.87
Min of LoICUS	1.01
Q1 of LoICUS	5.53
Median of LoICUS	15.00
Q3 of LoICUS	28.02
Max of LoICUS	182.42

Note: Std: Standard deviation; Min: Minimum; Q1: Lower quantile; Q3: Higher quantile; Max: Maximum

3.2. Results of the predictive model

3.2.1. Results of rule generation

Table 4. Evaluation metrics of models.

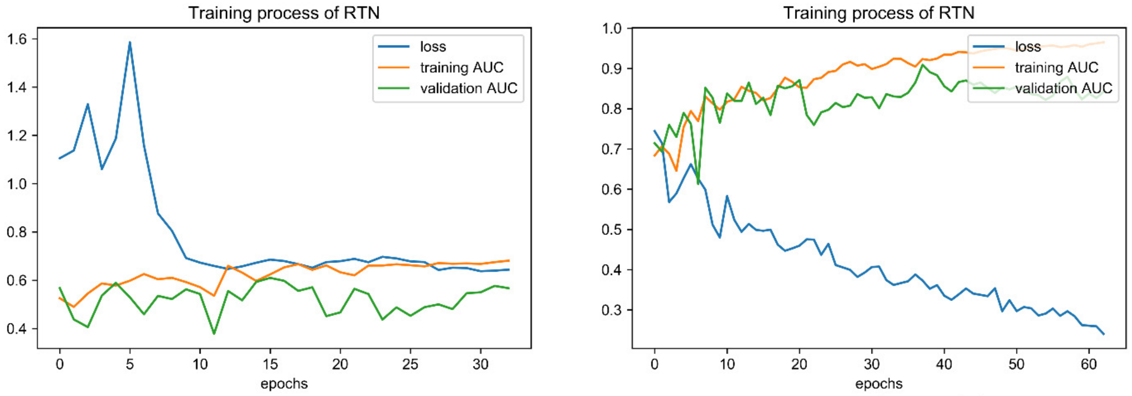
	PM	Accuracy	Precision	Recall	F1	AUC
Preoperative						
	LR	**0.59 (0.50, 0.73)**	0.66 (0.44, 0.92)	0.50 (0.16, 0.90)	**0.53 (0.24, 0.75)**	**0.61 (0.47, 0.77)**
	RF	0.58 (0.50, 0.69)	0.64 (0.43, 0.90)	**0.51 (0.10, 0.88)**	0.53 (0.17, 0.73)	0.60 (0.46, 0.74)
	GBDT	0.56 (0.44, 0.69)	**0.68 (0.39, 0.92)**	0.38 (0.04, 0.83)	0.43 (0.07, 0.70)	0.60 (0.44, 0.73)
Postoperative						
	LR	0.78 (0.67, 0.88)	0.82 (0.68, 0.93)	0.77 (0.53, 0.90)	0.79 (0.66, 0.88)	0.85 (0.72, 0.95)
	RF	0.80 (0.70, 0.89)	0.82 (0.70, 0.93)	0.79 (0.66, 0.93)	0.80 (0.70, 0.89)	0.88 (0.80, 0.95)
	GBDT	**0.81 (0.70, 0.88)**	**0.83 (0.70, 0.93)**	**0.79 (0.56, 0.93)**	**0.81 (0.68, 0.89)**	**0.91 (0.84, 0.97)**

The dataset was split into training and testing sets with a ratio of 9: 1 (training set: 567, testing set: 64). After the hyperparameters were tuned, the rules were generated based on GBDT. The results of GBDT compared with other baseline models are shown in Table 4. The optimized hyperparameters of the models are shown in Table S3.

For the preoperative model, we found that the overall effect of the prediction models is not good (the AUC of all models is lower than 0.70). Twenty estimators and two depths were set in GBDT, and we generated 38 rules according to the model. For the postoperative model, all the evaluation metrics of GBDT were higher than those of the other baseline models. We set 40 estimators and two depths in GBDT and generated 116 medical rules according to the model.

3.2.2. Performance of the predictive model

After rule embedding, we split the rule-feature into training, validation, and testing sets with a ratio of 8:1:1 (training set: 510, validation set: 57, testing set: 64). Figure 3 depicts the training process of RTN based on the constructed datasets. The horizontal axis is the number of epochs, and the vertical axis represents the AUCs of the training and validation sets.



Pre-operative model

Post-operative model

Figure 3. Training process of RTN. The blue curve stands for the training loss of RTN, the orange and green curves are the variation of the training and validation AUCs, respectively.

Table 5. Ten times 10-fold cross validation results of different models.

	PM	Accuracy	Precision	Recall	F1	AUC
Preoperative						
	RuleFit	0.57 (0.56, 0.58)	0.57 (0.56, 0.59)	0.60 (0.59, 0.62)	0.58 (0.57, 0.60)	0.59 (0.58, 0.60)
	RTN	**0.65 (0.64, 0.66)**	**0.71 (0.70, 0.72)**	**0.56 (0.55, 0.58)**	**0.60 (0.59, 0.61)**	**0.71 (0.70, 0.73)**
Postoperative						
	RuleFit	0.73 (0.72, 0.74)	0.73 (0.72, 0.74)	0.75 (0.74, 0.78)	0.73 (0.73, 0.74)	0.80 (0.79, 0.81)
	RTN	**0.79 (0.78, 0.80)**	**0.79 (0.78, 0.80)**	**0.80 (0.79, 0.81)**	**0.79 (0.78, 0.80)**	**0.90 (0.89, 0.91)**

Note: Bold font indicates the best value for each metric.

We found that the preoperative model is difficult to fit and the loss is about 0.60 after 30 epochs.

The postoperative model can converge quickly after 60 epochs, and the final loss is close to zero. We can draw the conclusion that the performances of the preoperative and postoperative models on the validation set is poor according to the figure. The 10 times 10-fold cross validation results of the different models are shown in Table 5. We used the mean and 95% confidence interval of all indicators. The RTN performed better on the training set compared with RuleFit, and all the evaluation metrics exceeded the baseline model.

Table 6 shows the performance on the test set of our model comparing to RuleFit (GBDT based). It is noticeable that the performance of postoperative models is much better than baseline model, and the five evaluating indicators of postoperative models have more superiority.

For the preoperative models, the advantage of RTN is not obvious compared with the baseline model. Compared with baseline model, RTN had an increased in AUC (6%, 0.57) and precision (4%, 0.61). RuleFit (GBDT-based model) had 0.56 accuracy, 0.68 recall, and 0.60 F1 score, which is higher than that of RTN.

We can see the advantage of our postoperative model (RTN) compared with the baseline model. Compared with previous studies on self-interpretable models, we can find remarkable performance improvements in our model.

The accuracy (0.76) and AUC (0.85) of the RTN exceeded the performance of the baseline model, and the precision (0.86) of the RTN was equal to that of the baseline model. Its recall and F1 score are slightly lower than those in the baseline model. Compared with RuleFit, which is also a rule-based prediction model, our model had few limitations, such as the premise assumption of data distribution. Therefore, the RTN proposed in our study performed better and had better robustness and applicability.

Table 6. Performance of RTN on the test set compared with the baseline model.

	PM	Accuracy	Precision	Recall	F1	AUC
preoperative						
	RuleFit	**0.56 (0.47, 0.66)**	0.57 (0.46, 0.71)	**0.68 (0.14, 0.92)**	**0.60 (0.22, 0.73)**	0.51 (0.38, 0.67)
	RTN	0.54 (0.41, 0.66)	**0.61 (0.35, 0.88)**	0.43 (0.01, 0.90)	0.46 (0.16, 0.71)	**0.57 (0.40, 0.71)**
postoperative						
	RuleFit	0.73 (0.63, 0.81)	0.86 (0.73, 0.96)	**0.70 (0.46, 0.85)**	**0.76 (0.61, 0.85)**	0.83 (0.72, 0.92)
	RTN	**0.76 (0.64, 0.86)**	**0.86 (0.72, 0.95)**	0.67 (0.40, 0.88)	0.74 (0.55, 0.87)	**0.85 (0.75, 0.95)**

Note: Bold font indicates the best value for each metric.

3.2.3. Results of supplementary regression experiments

The results of regression models are shown in Table S4. The evaluation metrics are explained variance score, mean squared error, root mean squared error, mean absolute error, R2 score, and adjusted R2 score. Overall, according to the results and performance in supplementary experiments, we drew the conclusion that regression predictive models are harder to fit compared with classification models on preoperative and postoperative tasks. Besides, considering that the follow-up tasks need the strong reliability of the prediction model, we believe that using a dichotomized endpoint rather than time to event data as the outcome of the study is a better choice. Therefore, we chose classification instead of regression methods to predict the LoICUS.

3.3. Result of postoperative risk analysis

Based on the outstanding performance of our postoperative model, we extracted and validated the medical rules generated by RTN. A total of 116 rules imported into the TabNet were extracted for subsequent verification. According to the global rule importance computed by RTN, we filtered out 57 rules with global importance of zero, and 59 rules were left. We match the case–control pairs of each rule according to the case–control matching strategy. After the rules with vague OR were eliminated, 24 rules were left, and 13 of which are risk factors. We listed seven representative decision rules in Table 7. All rules are sorted in descending order by OR value.

Table 7. Medical rules discovered by our framework.

No.	Medical decision rules	OR	95% CI low	95% CI high	I
1	max ALP > 160.0 & min RBC ≤ 2.52	5.12	1.59	16.42	0.0330
2	min WBC ≤ 5.33 & first Ca ≤ 2.25	4.29	2.05	8.96	0.0034
3	min PCT ≤ 0.30 & max PCT > 0.31	4.17	2.39	7.28	0.0069
4	min PT ≤ 12.65 & max PT ≤ 14.15	3.95	1.82	8.57	0.0594
5	min urine volume ≤ 32.5 & min DBIL ≤ 3.58	3.82	2.25	6.48	0.0741
6	min PCT ≤ 0.30 & max PCT > 0.44	3.54	2.12	5.91	0.0002
7	max GLO > 36.35	3.29	1.31	8.27	0.0001

Note: OR: odds ratio; 95% CI low: the lower limit of confidence interval of OR; 95% CI high: the higher limit of confidence interval of OR; I: global importance of the rule computed by RTN; ALP: alkaline phosphatase; RBC: red blood cells; WBC: white blood cells; Ca: calcium; PCT: procalcitonin; PT: prothrombin time; GLO: globulin.

4. Discussion

4.1. Discovery of medical rules

Based on the experiments results, we can draw the conclusion that in the LoICUS prediction task of patients undergoing craniotomy, postoperative variables, such as the vital signs and laboratory information of patients, may have a greater impact on the outcome compared with preoperative variables.

From the experiment result of medical rules discovery and validation in Table 6, we found that the global importance of the rules is not completely positively correlated with the results of risk analysis. Some interesting findings and inspirations can be analyzed from the results.

No. 1 is the rule with the highest OR value (5.12) and high global feature importance (0.0330). No. 1 means that within 24 hours after the patient enters the ICU, we need pay more attention to patients with an ALP higher than 160 U/L and an RBC below 2.52 1012/L, because they might have a longer LoICUS. Research found that the gradual elevation in ALP might prolong hospitalization [26]. A study found that the postoperative length of hospitalization increases by 0.837% (95% CI, 0.249–1.425%) per RBC unit transfused [27]. Therefore, the LoICUS might increase when RBC is too low to require red blood cell transfusion.

No. 2 implies that when the patient's serum ionized calcium is lower than 2.25 mmol/L when he first entered the ICU, we also need to pay attention to whether the WBC of patient will drop below 5.33 109/L, which might increase the LoICUS for patients undergoing craniotomy. Satoshi et al. indicated that for patients after cardiopulmonary bypass in the ICU, ionized calcium can make a remarkable difference in LoICUS [28]. To the best of our knowledge, no study has reported on the effect of ionized calcium on the LoICUS of patients undergoing craniotomy when they first entered the ICU. A study found that for patients undergoing craniotomy, infection can be predicted within 4 days according to standard blood count data, such as WBC [29]. However, within 24 hours after the patient undergoing craniotomy enters the ICU, the effect of decreased WBC on LoICUS has not been studied before, and its rule needs to be paid attention.

No. 3 and No. 6 focuses on the minimum and maximum values of PCT for patients undergoing craniotomy, meaning that too low or too high PCT within 24 hours after the patient undergoing craniotomy enters the ICU will have a certain impact on the LoICUS. A study found that PCT is a valuable marker for patients undergoing craniotomy [30]. Therefore, high and low PCT values should be concerned for patients undergoing craniotomy, especially the LoICUS.

No. 4 indicates that short prothrombin time may affect the LoICUS after patients enter the ICU. It has the second highest global importance with an OR of 3.95. Traumatic injury is associated with coagulopathy [31]; thus, we consider that it might be a risk factor of patients undergoing craniotomy that has not been studied before.

No. 5 possesses the highest global importance (0.0741) among the rules. It means that the minimum urine volume of the patient is lower than 32.5 and the minimum direct bilirubin (DBiL) is lower than 3.58 μmol/L. The urine outcome is routinely measured in the ICU and may be associated with hospital mortality [32]. Hyperbilirubinemia is a common postoperative complication, and DBiL is related to some pathophysiologies [33]. However, research on declined DBiL, especially for patients undergoing craniotomy is few. We believe that for patients undergoing craniotomy who entered the ICU within 24 hours, the cross influence of declined DBiL and urine volume on LoICUS might need to be paid more attention.

No. 7 means that when the serum globulin (GLO) is higher than 36.35 g/L, the rule is a risk factor of LoICUS for patients undergoing craniotomy. The increase in serum GLO may be the enhancement of immune response caused by infection. We found that each infection after craniotomy is statistically substantial and might affect the LoICUS. Therefore, we should pay attention to whether patients have elevated serum GLO within 24 hours after entering the ICU.

4.2. Limitations and future works

Our study still has some limitations and improvements despite the compelling results. First, in the study of e-healthcare systems using health electronics records, privacy is an important factor that needs to be considered, and we need to integrate privacy-preserving methods as a future work for the proposed method [34,35]. Second, the lack of preoperative variables in our study leads to the poor effect of our preoperative model. However, considering the allocation of medical resources, discovering the preoperative rules of patients undergoing craniotomy may be more helpful to reduce the medical burden in the real world. Third, our model achieved a better performance than the baseline model, and it has a strong interpretability, especially the interpretation of feature interaction. However, its performance is slightly lower compared with the gradient boosting-based method, such as GBDT. Our research shed light on the results of [36], which considered that tree-based models still outperform DL on tabular data, especially on medium-sized data (< 10 K samples). They also found that NNs are not robust to uninformative features, which we could not remove, because the feature selection method is combined with the statistical results and recommendations of domain experts. In our study, we considered GBDT is a basic, typical, and common method in gradient boosting, and we might attempt to apply other ML algorithms. such as faster implementation methods, including LightGBM, instead of GBDT in rule generation procedure in the future. Besides, other ML algorithms, except for tree-based method (such as k-NN, SVM, etc.) are also used in medical research [37,38], and the rule generation framework based on these methods are expected to be studied. We considered that our approaches disturb the spatial continuity of the original structure of the trees to some extent in the process of tree decomposing and rule embedding. In the future work, we can handle this problem by regarding trees as directed acyclic graphs, and the representation of rules is the combination of vertices and edges of graphs, such as the Bayesian network. Moreover, missing data is an inevitable phenomenon in real-world study, and many of the ML packages we applied in our research do not accept missing data. According to the previous research, missing values have to be imputed using the population mean. However, mean imputation is more likely to introduce bias in the model; thus, we might build a model using an algorithm that is less affected by missing values in the future. Finally, verifying a piece of knowledge in medical research is difficult, especially on small-sample datasets. In our study, some possible high-risk rules may be filtered out because the experimental samples are too few.

5. Conclusions

Aiming at discovering some medical decision rules affecting the LoICUS of patients undergoing craniotomy, this paper proposes an interpretable framework including a PM (RTN) and a rule validation process based on real-world EMR data. The medical decision rules were generated and screened preliminary by RTN, and the validation procedure was implemented to verify and make further selection of the rules. The experiment results indicate that postoperative features have a greater impact on the LoICUS of patients undergoing craniotomy compared with preoperative features. The results also proved that RTN can achieve good performance and outperform other postoperative baseline models. The medical decision rules discovered by our framework are valuable in providing some clinical decision supports in the ICU to shorten the LoICUS and reduce the medical expenses of patients undergoing craniotomy.

Acknowledgments

This work was supported by the Fujian Province Intensive Medical Center Construction Project (2017-510) and the National Natural Science Foundation of China (Grant Nos. 11832003) and Major special project of "scientific and technological innovation 2025" in Ningbo (2021Z021). We sincerely appreciate Yidu Cloud (Beijing) Technology Co., Ltd. for providing technical support in data extracting and model developing.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. C. L. Beauregard, W. A. Friedman, Routine use of postoperative ICU care for elective craniotomy: a cost-benefit analysis, *Surg. Neurol.*, **60** (2003), 483–489. http://doi.org/10.1016/s0090-3019(03)00517-2
2. C. Li, L. Chen, J. Feng, D. Wu, W. Xu, Prediction of length of stay on the intensive care unit based on least absolute shrinkage and selection operator, *IEEE Access*, **7** (2019), 110710–110721. http://doi.org/10.1109/ACCESS.2019.2934166
3. D. Dahl, G. G. Wojtal, M. J. Breslow, R. Holl, D. Huguez, D. Stone, et al., The high cost of low-acuity ICU outliers, *J. Healthcare Manage.*, **57** (2012), 421–433.
4. S. Rose, Mortality risk score prediction in an elderly population using machine learning, *Am. J. Epidemiol.*, **177** (2013), 443–452. http://doi.org/10.1093/aje/kws241
5. B. P. Nguyen, H. N. Pham, H. Tran, N. Nghiem, Q. H. Nguyen, T. T. T. Do, et al., Predicting the onset of type 2 diabetes using wide and deep learning with electronic health records, *Comput. Methods Programs Biomed.*, **182** (2019), 105055. https://doi.org/10.1016/j.cmpb.2019.105055
6. A. H. T. Chia, M. S. Khoo, A. Z. Lim, K. E. Ong, Y. Sun, B. P. Nguyen, et al., Explainable machine learning prediction of ICU mortality, *Inf. Med. Unlocked*, **25** (2021), 100674. https://doi.org/10.1016/j.imu.2021.100674
7. T. Gentimis, A. J. Alnaser, A. Durante, K. Cook, R. Steele, Predicting hospital length of stay using neural networks on MIMIC III data, in *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress*, 2017. http://doi.org/10.1109/DASC-PICom-DataCom-CyberSciTec.2017.191
8. H. Harutyunyan, H. Khachatrian, D. C. Kale, G. V. Steeg, A. Galstyan, Multitask learning and benchmarking with clinical time series data, *Sci. Data*, **6** (2019), 96. http://doi.org/10.1038/s41597-019-0103-9

9. K. Alghatani, N. Ammar, A. Rezgui, A. Shaban-Nejad, Predicting intensive care unit length of stay and mortality using patient vital signs: machine learning model development and validation, *JMIR. Med. Inf.*, **9** (2021), e21347. http://doi.org/10.2196/21347
10. T. Miller, Explanation in artificial intelligence: Insights from the social sciences, *Artif. Intell.*, **267** (2019), 1–38. http://doi.org/10.1016/j.artint.2018.07.007
11. T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (2016), 13–17. http://doi.org/10.1145/2939672.2939785
12. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, et al., Lightgbm: A highly efficient gradient boosting decision tree, in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, (2017), 4–9.
13. B. G. Demissei, G. Cotter, M. F. Prescott, G. M. Felker, G. Filippatos, B. H. Greenberg, et al., A multimarker multi-time point based risk stratification strategy in acute heart failure: results from the RELAXAHF trial, *Eur. J. Heart Fail.*, **19** (2017), 1001–1010. http://doi.org/10.1002/ejhf.749
14. S. O. Arik, T. Pfister, TabNet: Attentive interpretable tabular learning, preprint, arXiv:1908.07442. https://doi.org/10.48550/arXiv.1908.07442
15. R. Yu, S. Wang, J. Xu, Q. Wang, X. He, J. Li, et al., Machine learning approaches-driven for mortality prediction for patients undergoing craniotomy in ICU, *Brain Inj.*, **35** (2021), 1658–1664. https://doi.org/10.1080/02699052.2021.2008491
16. J. H. Friedman, B. E. Popescu, Predictive learning via rule ensembles, *Ann. Appl. Stat.*, **2** (2008), 916–954. http://doi.org/10.1214/07-AOAS148
17. S. Wang, G. Liu, W. Zhu, Z. Jiao, H. Lv, J. Yan, et al., Interpretable knowledge mining for heart failure prognosis risk evaluation, *SEDAMI*, (2021), 3032.
18. Y. N. Dauphin, A. Fan, M. Auli, D. Grangier, Language modeling with gated convolutional networks, preprint, arXiv:1612.08083. https://doi.org/10.48550/arXiv.1612.08083
19. J. Gehring, M. Auli, D. Grangier, D. Yarats, Y. N. Dauphin, Convolutional Sequence to Sequence Learning, preprint, arXiv:1705.03122. https://doi.org/10.48550/arXiv.1705.03122
20. A. Martins, R. Astudillo, From softmax to sparsemax: A sparse model of attention and multi-label classification, preprint, arXiv:1602.02068. https://doi.org/10.48550/arXiv.1602.02068
21. P. Yadav, M. Steinbach, V. Kumar, G. Simon, Mining electronic health records (EHRs) A survey, *ACM Comput. Surv. (CSUR)*, **50** (2018), 1–40. http://doi.org/10.1145/3127881
22. D. N. Peikes, L. Moreno, S. M. Orzol, Propensity score matching: A note of caution for evaluators of social programs, *Am. Stat.*, **62** (2008), 222–231. http://doi.org/ 10.1198/000313008X332016
23. A. Thavaneswaran, L. Lix, Propensity score matching in observational studies, *Manit. Cent. Health Policy*, 2008.
24. S. Missios, P. Kalakoti, A. Nanda, K. Bekelis, Craniotomy for glioma resection: a predictive model, *World Neurosurg.*, **83** (2015), 957–964. http://doi.org/10.1016/j.wneu.2015.04.052

25. N. J. Goel, A. N. Mallela, P. Agarwal, K. G. Abdullah, O. A. Choudhri, D. K. Kung, et al., Complications predicting perioperative mortality in patients undergoing elective craniotomy: a population-based study, *World Neurosurg.*, **118** (2018), e195–e205. http://doi.org/10.1016/j.wneu.2018.06.153
26. M. Saeed, M. Saliaj, M. Khan, S. Kumar, Z. Khan, M. Bachan, Isolated elevation of serum alkaline phosphatase in ICU patients, *Chest*, **154** (2018), 370A. https://doi.org/10.1016/j.chest.2018.08.338
27. E. C. Vamvakas, J. H. Carven, RBC transfusion and postoperative length of stay in the hospital or the intensive care unit among patients undergoing coronary artery bypass graft surgery: the effects of confounding factors, *Transfusion*, **40** (2000), 832–839. https://doi.org/10.1046/j.1537-2995.2000.40070832.x
28. S. Kimura, T. Iwasaki, K. Oe, K. Shimizu, T. Suemori, T. Kanazawa, et al., High ionized calcium concentration is associated with prolonged length of stay in the intensive care unit for postoperative pediatric cardiac patients, *J. Cardiothorac. Vasc. Anesth.*, **32** (2018), 1667–1675. https://doi.org/10.1053/j.jvca.2017.11.006
29. N. Shinoura, R. Yamada, K. Okamoto, O. Nakamura, Early prediction of infection after craniotomy for brain tumours, *Br. J. Neurosurg.*, **18** (2004), 598–603. https://doi.org/10.1080/02688690400022771
30. H. Wang, Higher Procalcitonin level in cerebrospinal fluid than in serum is a feasible Indicator for diagnosis of intracranial infection, *Surg. Infect.*, **21** (2020), 704–708. http://doi.org/10.1089/sur.2019.194
31. A. B. Böhmer, K. S. Just, R. Lefering, T. Paffrath, B, Bouillon. R. Joppich, et al., Factors influencing lengths of stay in the intensive care unit for surviving trauma patients: A retrospective analysis of 30,157 cases, *Crit. Care.*, **18** (2014), 1–10. https://doi.org/10.1186/cc13976
32. Z. Zhang, X. Xu, H. Ni, H. Deng, Urine output on ICU entry is associated with hospital mortality in unselected critically ill patients, *J. Nephrol.*, **27** (2014), 65–71. http://doi.org/10.1007/s40620-013-0024-1
33. M. Nagae, M. Egi, K. Kubota, S. Makino, S. Mizobuchi, Association of direct bilirubin level with postoperative outcome in critically ill postoperative patients, *Korean J. Anesthesiol.*, **71** (2018), 30–36. https://doi.org/10.4097/kjae.2018.71.1.30
34. M. Zhang, Y. Chen, J. Lin, A privacy-preserving optimization of neighborhood-based recommendation for medical-aided diagnosis and treatment, *IEEE Internet Things J.*, **8** (2021), 10830–10842. http://doi.org/10.1109/JIOT.2021.3051060
35. C. Li, M. Dong, J. Li, G. Xu, X. Chen, W. Liu, et al., Efficient medical big data management with keyword-searchable encryption in healthchain, *IEEE Syst. J.*, 2022. http://doi.org/10.1109/JSYST.2022.3173538
36. L. Grinsztajn, E. Oyallon, G. Varoquaux, Why do tree-based models still outperform deep learning on tabular data, preprint, arXiv:2207.08815. https://doi.org/10.48550/arXiv.2207.08815
37. B. P. Nguyen, W. L. Tay, C. K. Chui, Robust biometric recognition from palm depth images for gloved hands, *IEEE Trans. Human Mach. Syst.*, **45** (2015), 799–804. http://doi.org/10.1109/THMS.2015.2453203

38. J. C. Ferrão, F. Janela, M. D. Oliveira, H. M. G. Martins, Using structured EHR data and SVM to support ICD-9-CM coding, in *2013 IEEE International Conference on Healthcare Informatics*, (2013), 511–516. http://doi.org/10.1109/ICHI.2013.79