Original Article

Web-based calculator using machine learning to predict intracranial hematoma in geriatric traumatic brain injury

Thara Tunthanathip[1]^, Nakornchai Phuenpathom[1]^, Apisorn Jongjit[2]^

[1]Division of Neurosurgery, Department of Surgery, Faculty of Medicine, Prince of Songkla University, Songkhla, Thailand; [2]Faculty of Medicine, Prince of Songkla University, Songkhla, Thailand

Contributions: (I) Conception and design: T Tunthanathip, N Phuenpathom; (II) Administrative support: T Tunthanathip, A Jongjit; (III) Provision of study materials or patients: T Tunthanathip, A Jongjit; (IV) Collection and assembly of data: T Tunthanathip, A Jongjit; (V) Data analysis and interpretation: T Tunthanathip, A Jongjit; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Thara Tunthanathip, MD, PhD. Division of Neurosurgery, Department of Surgery, Faculty of Medicine, Prince of Songkla University, Songkhla 90110, Thailand. Email: tsus4@hotmail.com.

Background: Traumatic brain injury (TBI) is a significant contributor to mortality and impairment among the general population. The elderly are at a higher risk of developing cerebral hematomas following TBI. Therefore, there has been an overuse of cranial computed tomography (CT) in this group. The purpose of this study was to assess the predictive ability of machine learning (ML) algorithms for traumatic intracranial hematoma prediction. The secondary objective was to explore the predictors associated with positive CT scans.

Methods: A retrospective cohort study was conducted to examine TBI patients aged 60 years and older. To train the ML models, 70% of the data was separated, with the remaining 30% being used for testing. The supervised techniques used for training the ML models were naïve Bayes (NB), support vector machines (SVM), k-nearest neighbor (KNN), decision trees (DT), random forests (RF), artificial neural networks (ANN), and extreme gradient boosting (XGB). Therefore, the testing dataset was used to evaluate the ML models' prediction capabilities.

Results: There were 2,052 patients in the total cohort and 403 (19.6%) of the cohort had positive CT scans. Ten clinical predictors were used for building ML models and testing their performance. The NB algorithm had acceptable discrimination; the area under the receiver operating characteristic curve (AUC) was 0.70. Moreover, the sensitivity and F1 score of NB were 0.97 and 0.91, respectively.

Conclusions: ML models have the potential to serve as a screening tool for predicting positive cranial CT scans in elderly TBI patients since they can assist clinicians in making clinical decisions. In practice, a web application would be a simple way to apply the predictive ML model. Furthermore, future studies should involve external validation to examine the generalizability of clinical prediction systems.

Keywords: Machine learning (ML); traumatic brain injury (TBI); elderly; clinical prediction tool; cranial computed tomography (cranial CT)

Received: 03 August 2023; Accepted: 22 September 2023; Published online: 08 November 2023.
doi: 10.21037/jhmhp-23-97
View this article at: https://dx.doi.org/10.21037/jhmhp-23-97

^ ORCID: Thara Tunthanathip, 0000-0002-6303-836X; Nakornchai Phuenpathom, 0009-0005-1464-1778; Apisorn Jongjit, 0000-0002-6998-3207.

Introduction

Traumatic brain injury (TBI) is a leading cause of death and disability in the general population, especially in low- and middle-income nations (1,2). From the literature review, age has been identified as one of the prognostic factors that have been documented (3-5). McIntyre *et al.* conducted a systematic review and meta-analysis and reported that the overall mortality rate among the elderly was 38.3%, and mortality was significantly associated with the patient's advanced age (6). Due to the high mortality associated with TBI in elderly patients, cranial computed tomography (CT) examinations are typically performed in this population. Cranial CT scans have been widely used to detect intracranial injury following TBI (7,8); however, the high cost and adverse effects, such as leukemia and brain tumors, must be weighed against each other in clinical practice (7,9).

A variety of clinical prediction tools have been created and are currently being used for outcome prediction in a variety of illnesses, including TBI (10), cancer (11), and surgical complications (12). In an era of disruptive technology, machine learning (ML) is one of the prediction techniques that has also been used to predict traumatic intracranial injury. Tunthanathip *et al.* used various ML algorithms to predict intracranial hematoma following TBI in children and reported that the random forest (RF) algorithm had the best predictive performance with an area under the curve (AUC) of 0.80 (13). Moreover, Abe *et al.* compared several ML algorithms to predict traumatic intracranial hematoma and found that extreme gradient boosting (XGB) had the highest AUC of 0.78–0.80 (14).

It has challenged us to strike a balance between excessive and optimal investigations in high-risk patients. Because under-investigation may result in missed intracranial injury, and the high expense of the over-investigation protocol imposes an economic burden in a low-resource setting. To the best of our knowledge, there is no documented method for using ML to predict intracranial hematoma in TBI elderly patients. In the face of this gap, the goal of the present study was to assess the predictive ability of ML algorithms for traumatic intracranial hematoma prediction. In addition, the secondary objective was to explore the predictors associated with intracranial hematoma in TBI elderly. We present this article in accordance with the TRIPOD reporting checklist (available at https://jhmhp.amegroups.com/article/view/10.21037/jhmhp-23-97/rc).

Highlight box

Key findings

- Machine learning (ML) algorithms demonstrated a high sensitivity and acceptable performance for predicting traumatic cerebral hematomas in the elderly, and they might be used as a screening tool to assist physicians.

What is known and what is new?

- The elderly have a higher risk of developing cerebral hematomas after traumatic brain injury (TBI); consequently, an overuse of cranial computed tomography has been observed in this group.
- ML has been used to predict outcomes in a range of diseases, including TBI. This manuscript assesses the predictive ability of various ML algorithms for traumatic intracranial hematoma prediction in the elderly.

What is the implication, and what should change now?

- The naïve Bayes model can guide physicians and healthcare organizations in deciding on optimal investigations and reducing unnecessary costs in general practice.

Methods

Study designs and study population

The retrospective cohort study started with a review of electronic medical records of TBI patients aged 60 years and older who were admitted to an urban trauma center hospital in southern Thailand between January 2015 and December 2019. Clinical characteristics and imaging findings were collected. Patients who did not have a preoperative cranial CT scan or whose official CT scan reports were unavailable were excluded from the study. In addition, the AUC formula was used for sample size calculation (15). Based on Abe *et al.*, various parameters were calculated as follows: AUC of 0.80, alpha of 0.05, and estimation error of 0.05 (14). Therefore, the sample size of the study population was at least 368 patients.

Operational definition

Baseline clinical characteristics and cranial CT findings were reviewed for analysis. Because hypotension produces a misinterpretation of the Glasgow coma scale (GCS) score due to inadequate cerebral perfusion, the GCS score collected in the current investigation was the patient's GCS score with stable vital signs following emergency department resuscitation (16). Based on the GCS score, the severity of TBI was classified as follows: mild TBI (GCS scores 13–15), moderate TBI (GCS scores 9–12), and severe

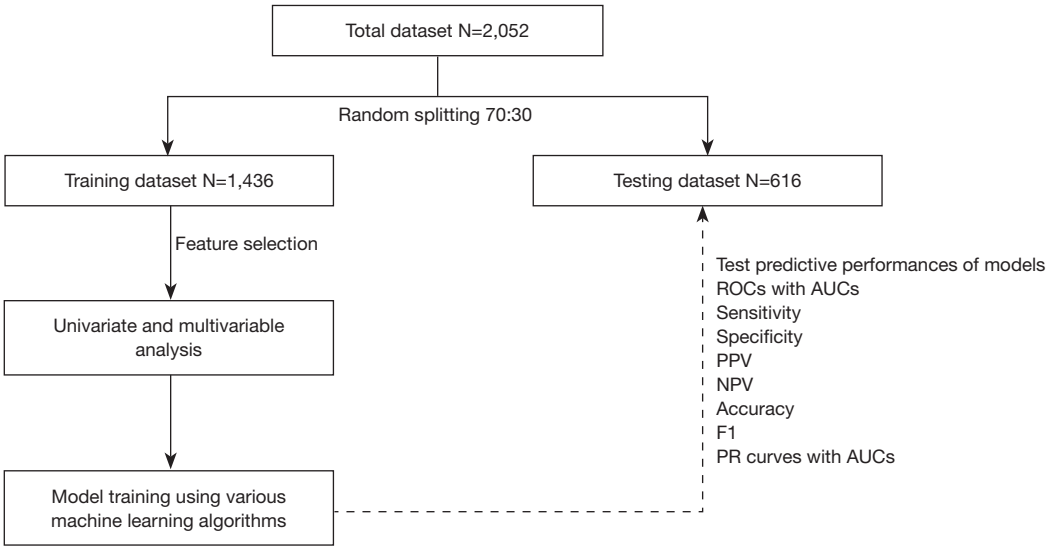

Figure 1 Workflow diagram. ROC, receiver operating characteristic; AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value; PR, precision-recall.

TBI (GCS scores 3–8).

Two neurosurgeons assessed the cranial CT findings, skull fracture, type of intracranial hematoma, midline displacement, and obliteration of the basal cistern. As a result, the present study's findings included the following intracranial hematomas: epidural hematoma (EDH), subdural hematoma (SDH), cerebral contusion, traumatic intracerebral hematoma, subarachnoid hemorrhage (SAH), intraventricular hemorrhage (IVH), and brainstem hematoma. The present study did not include skull fractures as an endpoint. Additionally, diffuse axonal injury (DAI) was characterized by Vieira *et al.* as those who showed symptoms of DAI on a CT scan or magnetic resonance imaging (17).

Ethical considerations

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the human research ethics committee board of the Faculty of Medicine, Prince of Songkla University (REC 65-138-10-1). The informed consent of the patients was not necessary for the present study because it was a retrospective analysis. However, patient identification numbers were encoded before analysis.

Statistical analysis

The workflow diagram of the present study is shown in *Figure 1*. Using descriptive analysis, patient characteristics, mechanism of injury, and intracranial injury of the total dataset were determined and presented as proportions with percentage, and mean with standard deviation (SD). In the present study, the complete case strategy was employed for missing value management prior to training the ML model.

Using a random data split, 70% of the total data was utilized to train the ML models, while the remaining 30% was used to test the models' performance. For the feature selection, various clinical characteristics were analyzed by Chi-square test and independent *t*-test. The Chi-square test was used to examine differences in proportions for categorical variables, while the independent *t*-test was used to compare the means of continuous variables between the positive CT scans and negative CT scan groups. Therefore, clinical variables with P values less than 0.05 were selected for training ML models. Additionally, the Hosmer-Lemeshow test was used to estimate the multivariable model for the calibration model utilizing binary classifiers, and the P value of the test greater than 0.05 indicated a good-fitting model (18,19).

Supervised algorithms with 10-fold cross-validation including naïve Bayes (NB), support vector machines (SVM), artificial neural networks (ANN), k-nearest neighbor (KNN), decision tree (DT), RF, and XGB were used for training the models from the training dataset. The "caret" package also optimized the parameters of each algorithm based on its accuracy score (20). The criterion

parameter and the maximum depth of the RF and XGB algorithms were also fine-tuned. The number of neighbors for KNN was tuned, whilst the SVM algorithm parameters were altered as follows: C and kernel. Furthermore, the activation, hidden layer sizes, learning rate, and solver of the ANN algorithm were modified.

Each algorithm's performance was measured using sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, and F1 score. The receiver operating characteristic (ROC) curves with AUCs were also calculated. An AUC of 0.7 indicated acceptable discrimination, 0.8 indicated outstanding discrimination, and 0.9 indicated great discrimination (21). In addition, the imbalanced number of endpoint outcomes is a common problem in clinical research; therefore, precision-recall (PR) curves with AUC and F1 scores have been reported to resolve imbalanced outcomes (22). The AUC of PR curves was used as a comparison metric among ML models. There is no standard cutoff value for the AUC of PR curves. Clearly, the greater the value of PR's AUC and the closer it is to 1.0, the better. The statistical analysis was performed using the R version 4.4.0 software (R Foundation, Vienna, Austria). Additionally, we constructed web-based applications of various algorithms and deployed them via the shiny platform to validate the ML models in the future (R studio, Boston, MA, USA).

Results

Clinical characteristics and imaging findings

A total of 2,052 patients were enrolled, and *Table 1* presents the baseline characteristics of the training and testing datasets. After splitting the data, 1,436 patients were considered suitable for the training dataset and the remaining was the testing dataset. The mean age of the TBI patients was 75.58 (SD 9.02) years in the training dataset, while the average age of the testing dataset was 75.63 (SD 8.75) years. The most prevalent mechanism of TBI in both datasets was a fall to the ground, while traffic injuries were observed in 20.5% and 20.8% of the training and testing datasets, respectively. The study observed that the GCS score ranging from 3 to 8 was present in around 2.3% and 2.9% of the cohorts, whereas moderate TBI was identified in approximately 2.9% and 3.4% of the cohorts.

For underlying disease, diabetes mellitus, hypertension, and cerebrovascular disease were common in both cohorts, while the use of aspirin was common as a current drug before TBI in approximately 8.3–9.4% of cases. The incidence of scalp injury after TBI was seen to vary between 52.2% and 53.6% across instances, while post-traumatic seizure occurred in approximately 0.9% to 1.0% of TBI patients.

As a result, cranial CT scans revealed positive findings in 19.6% (403/2,052) of the total cohort, and positive CT scans in the training dataset and testing dataset were 20.1% and 18.5%, respectively. Acute SDH was the most common intracranial hematoma, occurring in 8.3% and 11.4% of TBI patients, whereas EDH, SAH, IVH, and DAI were found in 1.0% and 1.4%, 8.1% and 8.1%, 1.5% and 1.7%, and 1.4% and 1.5%, respectively.

Feature selection

Clinical variables were analyzed using the Chi-square test and *t*-test. Therefore, the following 20 variables were selected for training ML models as follows: gender, traffic injury, aspirin, warfarin, stroke, ischemic heart disease, thrombocytopenia, headache, loss of consciousness, amnesia, vomiting, scalp injury, seizure, motor weakness, hypotension, hypoxia, bradycardia, bleeding per nose/ear, GCS score, pupillary light reflex, as shown in *Table 2*. Furthermore, the Hosmer-Lemeshow test was run, and the P value of the test was 0.7, indicating that the multivariate model fit well.

ML

During the training processes, the parameters of the ML models were optimized. In detail, the SVM model was optimized with a radial kernel and the regularization parameter (C parameter) of 0.0889. For the KNN algorithm, the model was also optimized with five neighbors. The optimized DT model had three nodes, while the optimized RF model comprised two maximum depths of the tree with 500 trees in the forest. Optimization of the ANN model comprised of five hidden layers, a logistic activation function, and an alpha of 0.1.

The NB model achieved the highest AUC of ROC when testing ML models, at 0.846, while XGB, ANN, and RF also had AUCs greater than 0.8, as shown in *Figure 2*. Furthermore, almost all ML models exhibited a notable degree of sensitivity, making them well-suited for utilization as screening tools. Due to the observed imbalance in endpoints in both cohorts, the F1 score and AUC of PR curves were utilized to estimate performance models in a problem with imbalanced outcomes. As shown in *Table 3*,

Table 1 Baseline characteristics of training and testing datasets

Characteristics	Training dataset (n=1,436)	Testing dataset (n=616)
Average age, years	75.58 [9.02]	75.63 [8.75]
Age group, years		
60–69	446 (31.1)	175 (28.4)
70–79	482 (33.6)	235 (38.1)
80–89	410 (28.6)	166 (26.9)
≥90	98 (6.8)	40 (6.5)
Male	722 (50.3)	289 (46.9)
Mechanism of injury		
Ground-level fall	1,037 (72.2)	441 (71.6)
Fall from height	34 (2.4)	22 (3.6)
Motorcycle crash	253 (17.6)	105 (17.0)
Car crash	29 (2.0)	15 (2.4)
Pedestrian injury	13 (0.9)	8 (1.3)
Object stuck at head	38 (2.6)	11 (1.8)
Penetrating injury	2 (0.1)	1 (0.2)
Body assault	7 (0.5)	5 (0.8)
Bicycle accident	13 (0.9)	5 (0.8)
Other	10 (0.7)	3 (0.5)
Traffic injury	295 (20.5)	128 (20.8)
Medication		
Aspirin	119 (8.3)	58 (9.4)
Clopidogrel	27 (1.9)	11 (1.8)
Warfarin	24 (1.7)	3 (0.5)
Underlying disease		
Hypertension	107 (7.5)	45 (7.3)
Diabetes mellitus	140 (9.7)	67 (10.9)
Dyslipidemia	88 (6.1)	30 (4.9)
Stroke	139 (9.7)	63 (10.2)
Ischemic heart disease	42 (2.9)	22 (3.6)
Thrombocytopenia	6 (0.4)	1 (0.2)
Renal failure	6 (0.4)	3 (0.5)
Signs and symptoms		
Scalp injury	749 (52.2)	330 (53.6)
Headache	349 (24.3)	152 (24.7)
Loss of consciousness	301 (21.0)	137 (22.2)
Amnesia	272 (18.9)	125 (20.3)
Vomiting	27 (1.9)	10 (1.6)
Seizure	13 (0.9)	6 (1.0)
Motor weakness	51 (3.6)	26 (4.2)
Hypotension	9 (0.6)	4 (0.6)
Hypoxia	8 (0.6)	4 (0.6)
Bradycardia	6 (0.4)	3 (0.5)
Bleeding per nose/ear	39 (2.7)	14 (2.3)
Glasgow coma scale score		
13–15	1,361 (94.8)	577 (93.7)
9–12	42 (2.9)	21 (3.4)
3–8	33 (2.3)	18 (2.9)
Pupillary light reflex		
Fixed both eyes	11 (0.8)	6 (1.0)
Fixed one eye	14 (1.0)	6 (1.0)
React both eyes	1,411 (98.3)	604 (98.1)
Cranial computed tomography finding		
Skull fracture	54 (3.8)	16 (2.6)
Intracerebral hematoma	289 (20.1)	114 (18.5)
Epidural hematoma	20 (1.4)	6 (1.0)
Acute subdural hematoma	163 (11.4)	51 (8.3)
Chronic subdural hematoma	51 (3.6)	23 (3.7)
Contusion/intracerebral hematoma	93 (6.5)	34 (5.5)
Subarachnoid hemorrhage	117 (8.1)	50 (8.1)
Intraventricular hemorrhage	25 (1.7)	9 (1.5)
Brainstem contusion	1 (0.1)	2 (0.3)
Diffuse axonal injury	20 (1.4)	9 (1.5)
Basal cistern obliteration	52 (3.6)	15 (2.4)
Average midline shift, mm	0.39 [2.15]	0.31 [1.77]
Positive CT scans	289 (20.1)	114 (18.5)

Data were presented as mean [standard deviation] or n (%). CT, computed tomography.

Table 2 Factors associated with intracranial hematoma after cranial CT scans using the training dataset

Characteristics	Negative CT scans (n=1,147)	Intracranial hematoma (n=289)	P value
Age, years†	75.67 [9.00]	75.22 [8.98]	0.45
Gender‡			<0.001
Male	549 (47.9)	173 (59.9)	
Female	598 (52.1)	116 (40.1)	
Traffic injury‡			<0.001
No	939 (81.9)	202 (69.9)	
Traffic injury	208 (18.1)	87 (30.1)	
Aspirin‡			<0.001
No	1,076 (93.8)	241 (83.4)	
Yes	71 (6.2)	48 (16.6)	
Clopidogrel‡			0.44
No	1,127 (98.3)	282 (97.6)	
Yes	20 (1.7)	7 (2.4)	
Warfarin‡			0.002
No	1,134 (98.9)	278 (96.2)	
Yes	13 (1.1)	11 (3.8)	
Hypertension‡			0.90
No	1,062 (92.6)	267 (92.4)	
Yes	85 (7.4)	22 (7.6)	
Diabetes mellitus‡			0.25
No	1,030 (89.8)	266 (92.0)	
Yes	117 (10.2)	23 (8.0)	
Dyslipidemia‡			0.63
No	1,075 (93.7)	273 (94.5)	
Yes	72 (6.3)	16 (5.5)	
Stroke‡			<0.001
No	1,056 (92.1)	241 (83.4)	
Yes	91 (7.9)	48 (16.6)	
Ischemic heart disease‡			0.01
No	1,120 (97.6)	274 (94.8)	
Yes	27 (2.4)	15 (5.2)	
Thrombocytopenia‡			0.004
No	1,145 (99.8)	285 (98.6)	
Yes	2 (0.2)	4 (1.4)	
Headache‡			<0.001
No	839 (73.1)	248 (85.8)	
Yes	308 (26.9)	41 (14.2)	

Table 2 (*continued*)

Table 2 (*continued*)

Characteristics	Negative CT scans (n=1,147)	Intracranial hematoma (n=289)	P value
Loss of consciousness‡			<0.001
No	942 (82.1)	193 (66.8)	
Yes	205 (17.9)	96 (33.2)	
Amnesia‡			<0.001
No	978 (85.3)	186 (64.4)	
Yes	169 (14.7)	103 (35.6)	
Vomiting‡			<0.001
No	1,143 (99.7)	266 (92.0)	
Yes	4 (0.3)	23 (8.0)	
Scalp injury‡			<0.001
No	493 (43.0)	194 (67.1)	
Yes	654 (57.0)	95 (32.9)	
Seizure‡			<0.001
No	1,144 (99.7)	279 (96.5)	
Yes	3 (0.3)	10 (3.5)	
Motor weakness‡			<0.001
No	1,145 (99.8)	240 (83.0)	
Yes	2 (0.2)	49 (17.0)	
Hypotension‡			<0.001
No	1,145 (99.8)	282 (97.6)	
Yes	2 (0.2)	7 (2.4)	
Hypoxia‡			<0.001
No	1,145 (99.8)	283 (97.9)	
Yes	2 (0.2)	6 (2.1)	
Bradycardia‡			0.004
No	1,145 (99.8)	285 (98.6)	
Yes	2 (0.2)	4 (1.4)	
Bleeding per nose/ear‡			<0.001
No	1,143 (99.7)	254 (87.9)	
Yes	4 (0.3)	35 (12.1)	
Glasgow coma scale score‡			<0.001
13–15	1,127 (98.3)	234 (81.0)	
9–12	18 (1.6)	24 (8.3)	
3–8	2 (0.2)	31 (10.7)	
Pupillary light reflex‡			<0.001
Fixed both eyes	1 (0.1)	10 (3.5)	
Fixed one eye	0	14 (4.8)	
React both eyes	1,146 (99.9)	265 (91.7)	

Data were presented as mean [standard deviation] or n (%). †, independent *t*-test. ‡, Chi-square test. CT, computed tomography.

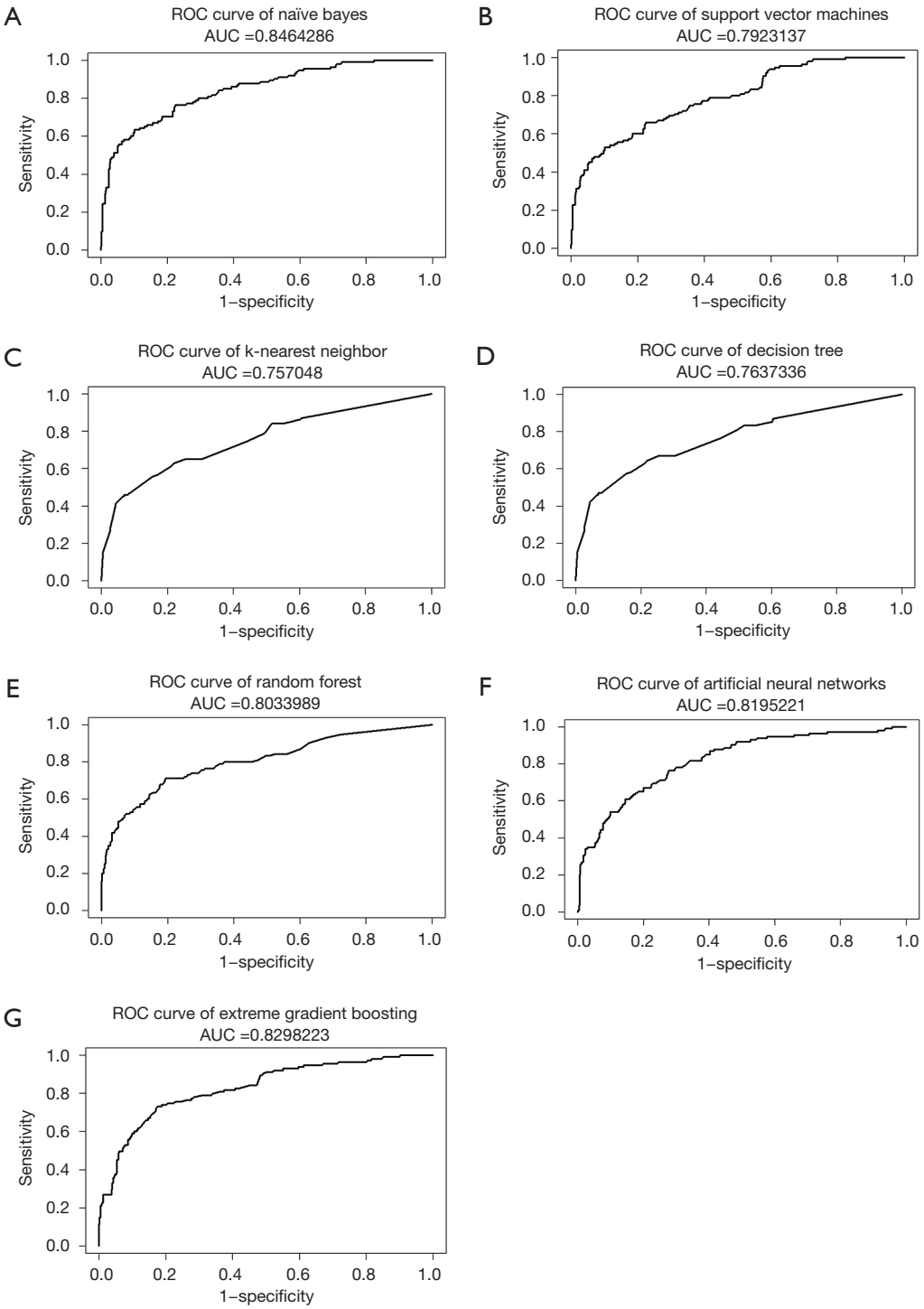

Figure 2 Receiver operating characteristic curves with area under the curves among various machine learning algorithms. (A) Naïve Bayes; (B) support vector machine; (C) k-nearest neighbors; (D) decision tree; (E) random forest; (F) artificial neural network; (G) extreme gradient boosting. ROC, receiver operating characteristic; AUC, area under the curve.

Table 3 Performances of ML algorithms for predicting intracranial hematoma after cranial CT scans

Algorithms	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Accuracy (95% CI)	F1 score (95% CI)
NB	0.95 (0.93, 0.97)	0.51 (0.42, 0.60)	0.89 (0.86, 0.92)	0.71 (0.61, 0.80)	0.87 (0.84, 0.89)	0.92 (0.90, 0.94)
SVM	0.97 (0.96, 0.98)	0.32 (0.23, 0.40)	0.86 (0.83, 0.89)	0.74 (0.61, 0.86)	0.85 (0.82, 0.88)	0.91 (0.90, 0.92)
KNN	0.95 (0.94, 0.97)	0.40 (0.31, 0.48)	0.87 (0.84, 0.90)	0.68 (0.57, 0.79)	0.85 (0.82, 0.88)	0.91 (0.89, 0.93)
DT	0.97 (0.96, 0.99)	0.31 (0.22, 0.39)	0.86 (0.83, 0.89)	0.76 (0.64, 0.88)	0.85 (0.82, 0.88)	0.91 (0.90, 0.92)
RF	0.96 (0.95, 0.98)	0.39 (0.30, 0.48)	0.87 (0.84, 0.90)	0.73 (0.62, 0.84)	0.86 (0.83, 0.88)	0.91 (0.90, 0.93)
ANN	0.97 (0.95, 0.98)	0.33 (0.25, 0.42)	0.86 (0.83, 0.89)	0.72 (0.60, 0.84)	0.85 (0.82, 0.88)	0.91 (0.90, 0.92)
XGB	0.96 (0.94, 0.97)	0.31 (0.22, 0.39)	0.85 (0.83, 0.88)	0.64 (0.51, 0.76)	0.84 (0.81, 0.86)	0.90 (0.89, 0.92)

ML, machine learning; CT, computed tomography; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; NB, naïve Bayes; SVM, support vector machine; KNN, k-nearest neighbors; DT, decision tree; RF, random forest; ANN, artificial neural network; XGB, extreme gradient boosting.

the NB model had the highest value of F1 score. Moreover, NB, XGB, and RF models had AUC of PR curves more than 0.6, as shown in *Figure 3*. Therefore, several ML models were launched as web applications to simplify implementation in general practice or external validation in the future. The web application can be accessed on both mobile phones and laptops via https://neurosx.shinyapps.io/TBI_elderly_ML/ or a quick response code scan. When ten clinical predictors of a new patient are entered, the prediction and probability of a positive cranial CT scan are displayed using several ML models, as shown in *Figure 4*.

Discussion

The incidence of intracranial hematoma following TBI in the elderly was found in 18.5–20.1%, which is consistent with prior studies (23,24). According to Paosaree *et al.*, the positive rate of cranial CT following TBI in patients older than 60 years was 21.6% (23). Moreover, acute SDH was the most common intracranial hematoma in the present study. These results are in concordance with other research reports. Heydari *et al.* (24). reported SDH in 27.6%, while another study reported that acute SDH was the most common intracranial injury in TBI patients ranging from 5.3% to 5.9% (25). Because shearing stain occurs during a head injury, a tear of the bridging vein leads to develop acute SDH (26).

Following TBI, older age is a risk factor for intracranial injury, and cranial CT is generally recommended (27,28). Clinical predictors associated with traumatic intracranial hematoma in the present study are as follows: antiplatelet therapy, anticoagulant therapy, traffic injury, GCS score, amnesia, vomiting, seizure, weakness, and bleeding per nose/ear. These findings are consistent with those of previous studies. Mori *et al.* (25) reported that GCS score less than 13, anticoagulant therapy, focal neurologic symptoms, posttraumatic convulsions, penetrating injury, and depressed fracture were associated with traumatic cerebral hematoma. On the other hand, Paosaree *et al.* (23) found that diabetes mellitus, ischemic heart disease, GCS score, amnesia, loss of consciousness, vomiting, seizure, decline in GCS score greater than 2 points, and chronic alcoholism were significantly related to positive brain CT scan results in TBI elderly.

The NB algorithm performed exceptionally well in predicting traumatic intracerebral hematoma, as indicated by the AUC of ROC. However, imbalance classes of endpoints were observed in the present study. AUC of PR curves and F1 scores were used to estimate the predictive performance. As a result, the NB algorithm still had the highest AUC of the PR curve and F1 score among various ML algorithms. Since there are currently no established criteria for these indicators to indicate good predictive performance, greater values are preferable and those that are closer to 1.0 are better (29). The present study's predicted performance results are consistent with previous research. Abe *et al.* tested various ML models to predict intracerebral hematoma in the prehospital setting and discovered that the AUCs of the ROC and PC curves were 0.78–0.80 and 0.46–0.51, respectively.

Additionally, various ML algorithms have been studied in various neurosurgical conditions from the literature review (30-33). Tunthanathip *et al.* employed the NB algorithm to forecast surgical site infection, and

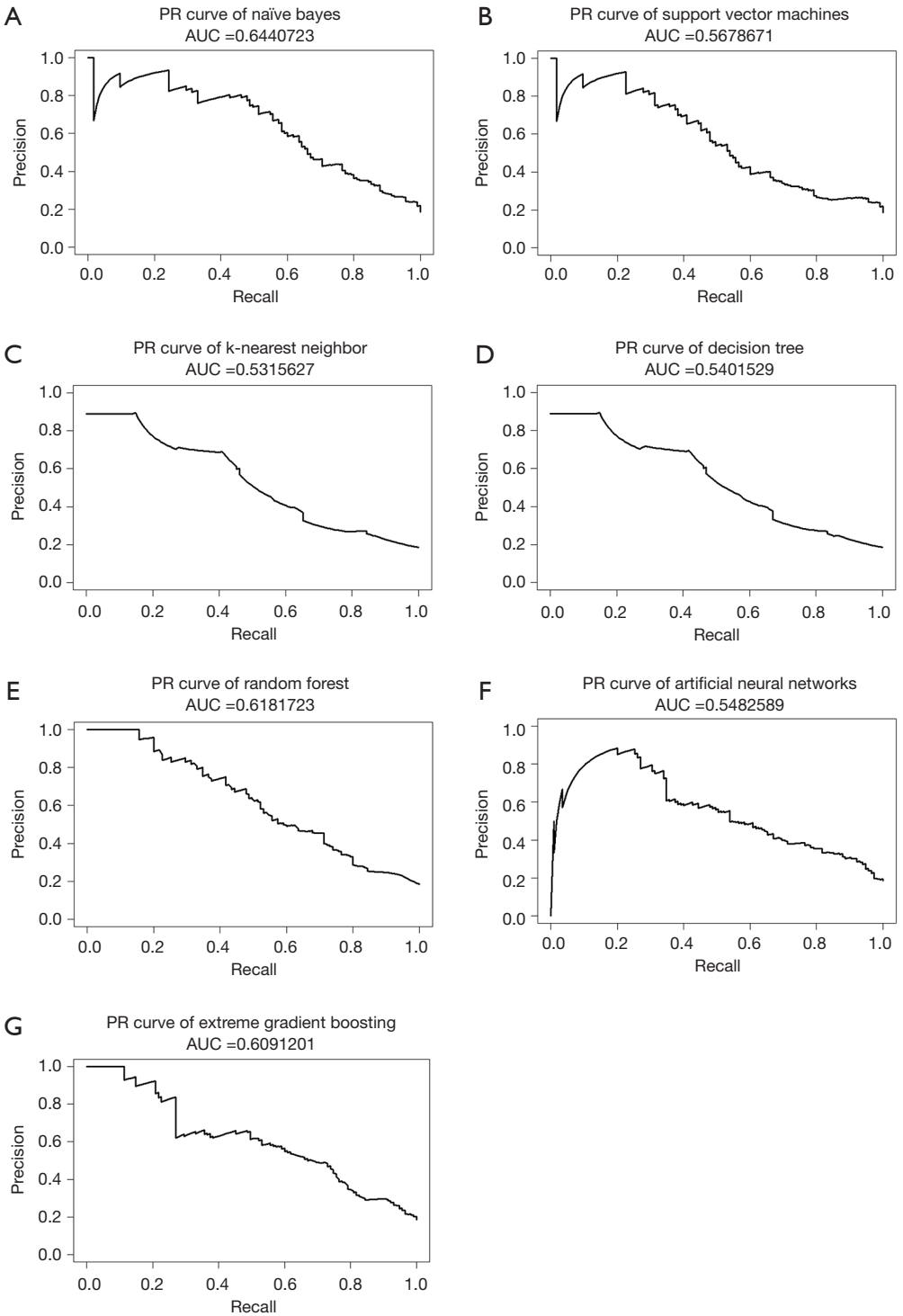

Figure 3 PR curves with area under the curves among various machine learning algorithms. (A) Naïve Bayes; (B) support vector machine; (C) k-nearest neighbors; (D) decision tree; (E) random forest; (F) artificial neural network; (G) extreme gradient boosting. PR, precision-recall; AUC, area under the curve.

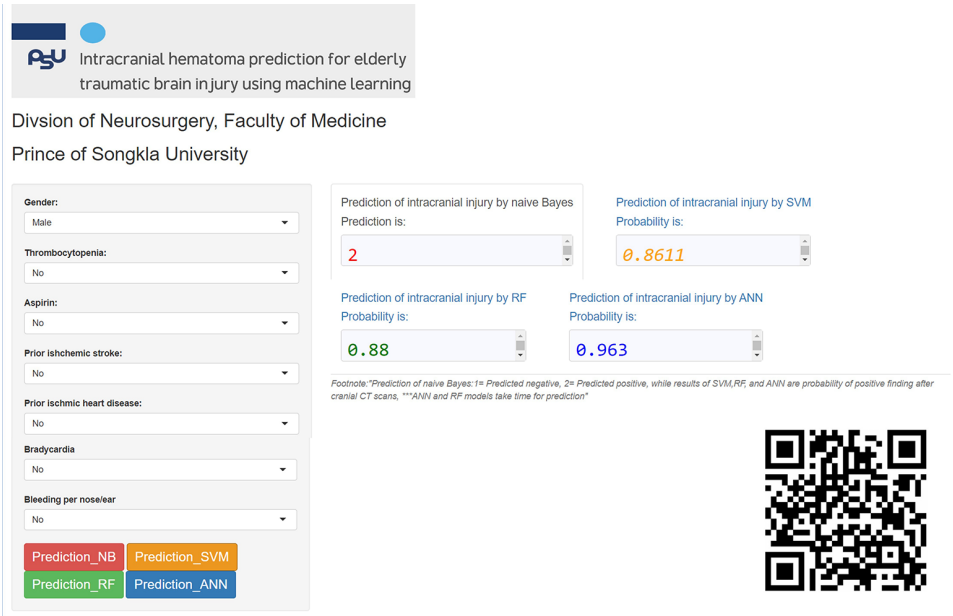

Figure 4 Screenshot of web application using machine learning algorithms for predicting intracranial hematoma in elderly traumatic brain injury.

this approach outperformed other ML algorithms (30). However, prior studies reported that the RF algorithm had the highest performance of positive cranial CT scans in pediatric TBI (13). Additionally, XGB and RF algorithms have been used for predicting intracranial pressure in hydrocephalus patients (31). From controversial results, the predictability of ML models needs to be compared using external validation with unseen data in the future.

Because the outstanding performance of ML models was high sensitivity, ML may be employed as the screening tool in real-world settings (32). A highly sensitive tool implies a low rate of false negative outcomes; few positive cranial CT scans are missed. This performance may assist physicians in determining the need for further investigation. Additionally, the present study also deployed various ML models in a cloud server as a web application. The online application would simplify the predictive ML model for use in general practice or external validation by other hospitals as the computerized clinical decision support systems (CDSS). A previous systematic review found that CCDSS improved the care process, including screening and treatment, and had an influence on patient outcomes, healthcare expenses, and patient safety (33). Moreover, CDSS can propose various ML models on the server that will support physicians in deciding for investigation by voting approach. The finalized predictions from several ML models involve the result of the majority.

To the best of the authors' knowledge, this is the first study that demonstrated and compared the predictive performance of several ML algorithms for a traumatic cerebral hematoma in the elderly. However, there were some limitations that should be acknowledged that the design of the study was a retrospective cohort study, which may have led to bias due to confounding variables. We attempted to control bias by adjusting for confounding variables using multivariable analysis (34-36). In the present study, an imbalance of endpoints was also observed; therefore, the F1 score was also calculated to assess predictability. As a result, ML models had an F1 score greater than 0.9 which means good performance (37). Due to the small number of positive brain CT scans observed in the current cohort, a multicenter study may be able to rectify this issue in order to enhance predictive performance. Also, future external validation with unobserved data is required to confirm the predictability of the ML models of the present investigation. Finally, several clinical prediction tools have been investigated for their ability to predict clinical outcomes. As an alternative, a nomogram is one of the clinical tools that has been used to forecast the course of many diseases (38,39). Future research should focus on comparing the prediction of ML-based models with nomograms.

Conclusions

ML models have the potential to serve as a screening tool for predicting positive cranial CT scans in elderly TBI patients since they can assist clinicians in making decisions in clinical practice. In practice, a web application would be a simple way to apply the predictive ML model. Furthermore, future studies should involve external validation to examine the generalizability of clinical prediction systems.

Acknowledgments

Funding: None.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://jhmhp.amegroups.com/article/view/10.21037/jhmhp-23-97/rc

Data Sharing Statement: Available at https://jhmhp.amegroups.com/article/view/10.21037/jhmhp-23-97/dss

Peer Review File: Available at https://jhmhp.amegroups.com/article/view/10.21037/jhmhp-23-97/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://jhmhp.amegroups.com/article/view/10.21037/jhmhp-23-97/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the human research ethics committee board of the Faculty of Medicine, Prince of Songkla University (REC 65-138-10-1). The informed consent of the patients was not necessary for the present study because it was a retrospective analysis. However, patient identification numbers were encoded before analysis.

References

1. Peden M, Oyegbite K, Ozanne-Smith J, et al. editors. World report on child injury prevention. Geneva: World Health Organization; 2008.
2. Tunthanathip T, Phuenpathom N. Impact of Road Traffic Injury to Pediatric Traumatic Brain Injury in Southern Thailand. J Neurosci Rural Pract 2017;8:601-8.
3. Fu WW, Fu TS, Jing R, et al. Predictors of falls and mortality among elderly adults with traumatic brain injury: A nationwide, population-based study. PLoS One 2017;12:e0175868.
4. Tran A, Saigle V, Manhas N, et al. Association of age with death and withdrawal of life-sustaining therapy after severe traumatic brain injury. Can J Surg 2023;66:E348-55.
5. Mosenthal AC, Livingston DH, Lavery RF, et al. The effect of age on functional outcome in mild traumatic brain injury: 6-month report of a prospective multicenter trial. J Trauma 2004;56:1042-8.
6. McIntyre A, Mehta S, Aubut J, et al. Mortality among older adults after a traumatic brain injury: a meta-analysis. Brain Inj 2013;27:31-40.
7. Melnick ER, Szlezak CM, Bentley SK, et al. CT overuse for mild traumatic brain injury. Jt Comm J Qual Patient Saf 2012;38:483-9.
8. Maxwell S, Ha NT, Bulsara MK, et al. Increasing use of CT requested by emergency department physicians in tertiary hospitals in Western Australia 2003-2015: an analysis of linked administrative data. BMJ Open 2021;11:e043315.
9. Pearce MS, Salotti JA, Little MP, et al. Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study. Lancet 2012;380:499-505.
10. Cui W, Ge S, Shi Y, et al. Death after discharge: prognostic model of 1-year mortality in traumatic brain injury patients undergoing decompressive craniectomy. Chin Neurosurg J 2021;7:24.
11. Tunthanathip T, Ratanalert S, Sae-Heng S, et al. Prognostic Factors and Nomogram Predicting Survival

in Diffuse Astrocytoma. J Neurosci Rural Pract 2020;11:135-43.
12. Tang TY, Zong Y, Shen YN, et al. Predicting surgical site infections using a novel nomogram in patients with hepatocelluar carcinoma undergoing hepatectomy. World J Clin Cases 2019;7:2176-88.
13. Tunthanathip T, Duangsuwan J, Wattanakitrungroj N, et al. Comparison of intracranial injury predictability between machine learning algorithms and the nomogram in pediatric traumatic brain injury. Neurosurg Focus 2021;51:E7.
14. Abe D, Inaji M, Hase T, et al. A Prehospital Triage System to Detect Traumatic Intracranial Hemorrhage Using Machine Learning Algorithms. JAMA Netw Open 2022;5:e2216393.
15. Thai Thanh Truc. Statistics and Sample Size Pro. 2020. [cited 2023 Sep 2]. Available online: https://play.google.com/store/apps/details?id=thaithanhtruc.info.sass&hl=en_US
16. Taweesomboonyat T, Kaewborisutsakul A, Tunthanathip T, et al. Necessity of in-hospital neurological observation for mild traumatic brain injury patients with negative computed tomography brain scans. J Health Sci Med Res 2000;38:267-74.
17. Vieira RC, Paiva WS, de Oliveira DV, et al. Diffuse Axonal Injury: Epidemiology, Outcome and Associated Risk Factors. Front Neurol 2016;7:178.
18. Crowson CS, Atkinson EJ, Therneau TM. Assessing calibration of prognostic risk scores. Stat Methods Med Res 2016;25:1692-706.
19. Demler OV, Paynter NP, Cook NR. Tests of calibration and goodness-of-fit in the survival setting. Stat Med 2015;34:1659-80.
20. Kuhn M, Wing J, Weston S, et al. caret: Classification and Regression Training. 2023 [cited 2023 Jul 5]. Available online: https://cran.r-project.org/web/packages/caret/caret.pdf
21. Roelen CA, Bültmann U, van Rhenen W, et al. External validation of two prediction models identifying employees at risk of high sickness absence: cohort study with 1-year follow-up. BMC Public Health 2013;13:105.
22. Brabec J, Komárek T, Franc V, et al. On Model Evaluation Under Non-constant Class Imbalance. Computational Science – ICCS 2020 2020;12140:74-87.
23. Paosaree P. Clinical Factors Predictive for Intracranial Hemorrhage in Geriatric with Traumatic Brain Injury in Chumphae Hospital. JPMAT 2020;10:341-50.
24. Heydari F, Golban M, Majidinejad S. Traumatic Brain Injury in Older Adults Presenting to the Emergency Department: Epidemiology, Outcomes and Risk Factors Predicting the Prognosis. Adv J Emerg Med 2020;4:e19.
25. Mori K, Abe T, Matsumoto J, et al. Indications for Computed Tomography in Older Adult Patients With Minor Head Injury in the Emergency Department. Acad Emerg Med 2021;28:435-43.
26. Miller JD, Nader R. Acute subdural hematoma from bridging vein rupture: a potential mechanism for growth. J Neurosurg 2014;120:1378-84.
27. Stiell IG, Wells GA, Vandemheen K, et al. The Canadian CT Head Rule for patients with minor head injury. Lancet 2001;357:1391-6.
28. Wasson JH, Sox HC, Neff RK, et al. Clinical prediction rules. Applications and methodological standards. N Engl J Med 1985;313:793-9.
29. Boyd K, Santos Costa V, Davis J, et al. Unachievable Region in Precision-Recall Space and Its Effect on Empirical Evaluation. Proc Int Conf Mach Learn 2012;2012:349.
30. Tunthanathip T, Sae-Heng S, Oearsakul T, et al. Machine learning applications for the prediction of surgical site infection in neurological operations. Neurosurg Focus 2019;47:E7.
31. Trakulpanitkit A, Tunthanathip T. Comparison of intracranial pressure prediction in hydrocephalus patients among linear, non-linear, and machine learning regression models in Thailand. Acute Crit Care 2023;38:362-70.
32. Maxim LD, Niebo R, Utell MJ. Screening tests: a review with examples. Inhal Toxicol 2014;26:811-28.
33. Souza NM, Sebaldt RJ, Mackay JA, et al. Computerized clinical decision support systems for primary preventive care: a decision-maker-researcher partnership systematic review of effects on process of care and patient outcomes. Implement Sci 2011;6:87.
34. Kaewborisutsakul A, Tunthanathip T. Development and internal validation of a nomogram for predicting outcomes in children with traumatic subdural hematoma. Acute Crit Care 2022;37:429-37.
35. Tunthanathip T, Oearsakul T. Application of machine learning to predict the outcome of pediatric traumatic brain injury. Chin J Traumatol 2021;24:350-5.
36. Tunthanathip T, Sae-Heng S, Oearsakul T, et al. Economic impact of a machine learning-based strategy for preparation of blood products in brain tumor surgery. PLoS One 2022;17:e0270916.
37. Allwright S. What is a good F1 score and how do I

interpret it? 2023. [cited 2023 Jul 5]. Available online: https://stephenallwright.com/good-f1-score/
38. Oearsakul T, Tunthanathip T. Development of a nomogram to predict the outcome of moderate or severe pediatric traumatic brain injury. Turk J Emerg Med 2022;22:15-22.
39. Tunthanathip T, Duangsuwan J, Wattanakitrungroj N, et al. Clinical Nomogram Predicting Intracranial Injury in Pediatric Traumatic Brain Injury. J Pediatr Neurosci 2020;15:409-15.

doi: 10.21037/jhmhp-23-97
Cite this article as: Tunthanathip T, Phuenpathom N, Jongjit A. Web-based calculator using machine learning to predict intracranial hematoma in geriatric traumatic brain injury. J Hosp Manag Health Policy 2023;7:16.