Clinical Article

https://doi.org/10.3340/jkns.2023.0118

**Prediction Model for unfavorable Outcome in Spontaneous Intracerebral Hemorrhage Based on Machine Learning**

Shengli Li, Jianan Zhang, Xiaoqun Hou, Yongyi Wang, Tong Li, Zhiming Xu, Feng Chen, Yong Zhou, Weimin Wang, Mingxing Liu

Department of Neurosurgery, Qingdao Municipal Hospital (Headquarter), Qingdao, China

Department of Anesthesia Operating Room, Qingdao Hospital, University of Health and Rehabilitation Sciences (Qingdao Municipal Hospital), Qingdao, China

**Running Title:** Machine Learning for Intracerebral Hemorrhage

• Received : June 13, 2023 • Revised : August 5, 2023 • Accepted : August 21, 2023

Address for correspondence : **Mingxing Liu**

Department of Neurosurgery, Qingdao Municipal Hospital (Headquarter), No.1 Jiaozhou Road, Qingdao 266000, China

Tel : +86-, Fax : +86-, E-mail : lmx080501@163.com, ORCID : https://orcid.org/0000-0002-0030-3146

**Weimin Wang**

Department of Neurosurgery, Qingdao Municipal Hospital (Headquarter), No.1 Jiaozhou
Road, Qingdao 266000, China

Tel : +86-, Fax : +86-, E-mail : wangwm@sina.com, ORCID : https://orcid.org/0000-0003-0833-1739

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2023 The Korean Neurosurgical Society
Abstract

Objective: The spontaneous intracerebral hemorrhage (ICH) remains a significant cause of mortality and morbidity throughout the world. The purpose of this retrospective study is to develop multiple models for predicting ICH outcomes using machine learning (ML).

Methods: Between January 2014 and October 2021, we included ICH patients identified by CT or MRI and treated with surgery. At the 6-month check-up, outcomes were assessed using the modified Rankin Scale (mRS). In this study, four machine learning models, including Support Vector Machine (SVM), Decision Tree C5.0, Artificial Neural Network (ANN), Logistic Regression (LR) were used to build ICH prediction models. In order to evaluate the reliability and the ML models, we calculated the area under the receiver operating characteristic curve (AUC), specificity, sensitivity, accuracy, positive likelihood ratio (PLR), negative likelihood ratio (NLR), Diagnostic Odds Ratio (DOR).

Results: We identified 71 patients who had favorable outcomes and 156 who had unfavorable outcomes. The results showed that the SVM model achieved the best comprehensive prediction efficiency. For the SVM model, the AUC, accuracy, specificity, sensitivity, PLR, NLR and DOR were 0.91, 0.92, 0.92, 0.93, 11.63, 0.076 and 153.03 respectively. For the SVM model, we found the importance value of time to operating room (TOR) was higher significantly than other variables.

Conclusion: The analysis of clinical reliability showed that the SVM model achieved the best comprehensive prediction efficiency and the importance value of TOR was higher significantly than other variables.
**Key Words**: Intracerebral hemorrhage · Machine learning · Support vector machine · Area under the receiver operating characteristic curve · Time to operating room.

**INTRODUCTION**

Throughout the world, the spontaneous intracerebral hemorrhage (ICH) remains a significant cause of mortality and morbidity. Previous studies showed that excellent medical treatment, including timely removal of hematoma, management of blood pressure and glucose and control of intracranial pressure, can improve the outcome of ICH\(^6\)\(^{\text{-}10}\). In recent years, it was demonstrated that evacuation of hematoma with minimally invasive methods could potentially increase microcirculation in the afflicted brain areas\(^7\)\(^{\text{-}}\)\(^8\). Several studies have revealed that individuals who had their surgery performed sooner have better results. Other research has found that surgery during the acute phase is related with a greater likelihood of postoperative rebleeding\(^15\)\(^{\text{-}}\)\(^19\). Consequently, the optimal time of ICH surgical evacuation to produce favorable results has yet to be determined. All in all, the factors influenced the outcome of ICH were still controversial. The majority of ICH predictor studies have used univariable and multivariable logistic regression analysis and the accuracy of these models is generally poor\(^5\)\(^{\text{-}}\)\(^22\). Consequently, it is essential to develop new models for evaluating the results of ICH utilizing new technologies, such as machine learning (ML).

The algorithms used in machine learning (ML) could extract patterns from huge datasets with numerous variables. Because of its capacity to execute complicated pattern recognition and uncover nonlinear contributions in big data sets, ML algorithms have the potential to enhance
event and outcome prediction\textsuperscript{4}. Once the result label is established, ML algorithms may autonomously optimize its parameters with minimal supervision. Unlike regression models, ML can deal with vast volumes of data and patient features while accounting for all of their relationships. As a result, ML algorithms have the potential to outperform regression models in terms of predicted accuracy.

The purpose of this retrospective study is to develop multiple models for predicting ICH outcomes using ML. We also examined the prediction accuracy of the ML models and selected the best model for clinical use.

MATERIALS AND METHODS

Subjects

Patient information was obtained from the Hospital’s computerized medical records. The Hospital’s Institutional Review Board (IRB) authorized this study (ID:082). Because the study was retrospective, the IRB did not require informed consent.

Between January 2014 and October 2021, we included ICH patients identified by CT or MRI and treated with surgery, including minimally invasive aspiration and standard craniotomies. We ruled out the following scenarios: 1) Intracerebral bleeding caused by an aneurysm, arteriovenous malformation, or tumor 2) Patients suffering from severe end-organ failure, intracranial and systemic infections, and blood disorders 3) Patients lost to follow-up after 6 months (Figure 1).

Age, sex, BMI, systolic and diastolic blood pressure, midline shift, operation type and time,
residual hematoma volume, and time to operating room, as well as pre-morbid chronic conditions (diabetes, hypertension, hyperlipidemia), were gathered.

**Surgery and follow-up**

The recommendations from the American Heart Association and American Stroke Association published in *Stroke*, 2015 were used to treat every patient\(^{20}\). Our hospital’s on-call neurosurgeons choose between doing a normal craniotomy or a minimally invasive aspiration employing neuron navigation (Medtronic, XPS Nexus) depending on the patient’s circumstances. Every patient was monitored for at least six months. At the 6-month check-up, outcomes were assessed using the modified Rankin Scale (mRS). In this investigation, we used the mRS score to categorize patient outcomes, with mRS 0–2 representing positive results and mRS 3-5 representing poor outcomes.

**Machine learning**

In this study, four machine learning models, including Support Vector Machine (SVM), Decision Tree C5.0, Artificial Neural Network (ANN), Logistic Regression (LR) were used to build ICH prediction models. We evaluated each model’s performance using the area under the receiver-operator curve (AUC). The machine learning was performed with IBM SPSS Modeler 14.1 (IBM Corp., Armonk, New York). A classifier known as SVM uses kernels to turn input data into a multidimensional hyperplane in order to distinguish between two groups. By deriving decision rules from the training data, decision trees predict class membership. The term "artificial neural network" (ANN) refers to a large parallel and linked network made
up of straightforward adaptable components. LR is a binary dependent variable regression model (Figure 2).

**Statistical analysis**

The categorical data were expressed as percentage and the continuous data were expressed as mean with standard deviation. We used the t test or Mann-Whitney U test for categorical data to compare the groups. Fisher's exact tests or the chi-square test were applied to continuous variables. In order to evaluate the reliability and the ML models, we calculated the area under the receiver operating characteristic curve (AUC), specificity, sensitivity, accuracy, positive likelihood ratio (PLR), negative likelihood ratio (NLR), Diagnostic Odds Ratio (DOR). If the p value was less than 0.05, we regarded the result as statistically significant. The statistical analyses were performed with IBM SPSS Statistics 26 (IBM Corp., Armonk, New York).

**RESULTS**

**Characteristics of the population**

We identified 71 patients who had favorable outcomes and 156 who had unfavorable outcomes (Figure 1). Between the two groups, there were no significant differences in terms of age, hypertension, sex, BMI, hyperlipidemia, hematoma side and location, residual hematoma volume, operation time, and type (Table 1). However, two groups showed significant differences in the percentage of diabetes, SBP, DBP, GCS, midline shift and hematoma volume. The details were showed in the Table 1. As the
continuous data, we found there was no significant difference in the time to operating room (TOR) between two groups. Following that, we divided the cases into three categories: (1) <12 hours, (2) >=12 and <=36 hours, and (3) >36 hours. We discovered the patients who had surgery between 12 and 36 hours were more likely to have positive outcomes than the other groups. (p<0.001) (Table 1).

Machine learning models

We established four ML models. For the SVM model, the AUC, accuracy, specificity, sensitivity, PLR, NLR and DOR were 0.91, 0.92, 0.92, 0.93, 11.63, 0.076 and 153.03 respectively. For the Decision tree C5.0, the AUC, accuracy, specificity, sensitivity, PLR, NLR and DOR were 0.96, 0.87, 0.88, 0.87, 7.25, 0.148 and 48.99 respectively. For the ANN model, the AUC, accuracy, specificity, sensitivity, PLR, NLR and DOR were 0.91, 0.91, 0.82, 0.97, 5.39, 0.037 and 145.68 respectively. For the LR model, the AUC, accuracy, specificity, sensitivity, PLR, NLR and DOR were 0.77, 0.81, 0.73, 0.85, 3.15, 0.205 and 15.37 respectively. The results above showed that the SVM model achieved the best comprehensive prediction efficiency (Figure 3) (Table 2).

We also calculated the variable importance values in the four models. For the SVM and LR model, we found the importance value of TOR was higher significantly than other variables. For the Decision tree C5.0 model, the sequence of importance value was midline shift, GCS, SBP and TOR. For the ANN model, the sequence of importance value was midline shift, TOR, DBP, and SBP. Moreover, the importance values of the above variables in the Decision tree C5.0 and ANN model were similar (Figure 4).
DISCUSSION

In the study, we established four ML models including the SVM, Decision tree C5.0, ANN, and LR models. The analysis of clinical reliability showed that the SVM model achieved the best comprehensive prediction efficiency. For the SVM and LR model, we found the importance value of TOR was higher significantly than other variables. While, the importance values of the variables in the Decision tree C5.0 and ANN model were similar.

Spontaneous ICH is a form of stroke that is associated with a high death and disability rate. Regardless of therapy, physicians, families, and patients are most concerned with prognosis. Previous research has linked a variety of characteristics to a poorer prognosis following ICH, including advancing years, a lower GCS score, a greater hematoma volume and infratentorial hemorrhage\textsuperscript{18, 24}. Cheung and colleagues demonstrated that in a sample of 142 patients, the ICH score comprising the GCS, age, and ICH volume, helped to predict favorable prognosis with a sensitivity of 93.5\%\textsuperscript{3}. In 310 patients, Ruiz-Sandoval et al published an ICH score system with sensitivity of 70.0\% for 30-day excellent outcome\textsuperscript{21}. Previously described prediction approaches, on the whole, concentrated on short-term prognosis and had very poor accuracies. In our study, we found that the SVM model achieved the best comprehensive prediction efficiency and the importance value of TOR was higher significantly than other variables. While, the timing of surgery for ICH is still debatable. The previous study showed that the TOR<12 hours was early time windows\textsuperscript{12}. Wang et al. investigated ultra-early (or =7 h), early (7-24 h), and delayed (>24 h) surgery in spontaneous ICH and discovered that
surgery conducted within the 7-24h window resulted in the best outcome\textsuperscript{25}. Early surgery was linked to a greater risk of rebleeding, whereas delayed surgery was linked to severe consequences\textsuperscript{11}. Moreover, in my previous paper, I found that the 21 h of time to operation room was the turning point of outcome and the best outcome is obtainable at 12-36h surgery\textsuperscript{14}. Consequently, we TOR was divided into three categories: (1) <12 hours, (2) \(\geq\)12 and \(\leq\)36 hours, and (3) >36 hours. Our results were consistent with previous studies.

ML techniques have been employed in medical research and frequently outperform traditional statistical models. Machine learning has recently been used to the severity or outcome prediction model for neurological illnesses such as ischemic stroke and traumatic brain injury in research\textsuperscript{2,13,23}. In aneurysmal subarachnoid hemorrhage, Maldaner et al. discovered that machine learning models exhibit strong discrimination and calibration for a favorable functional result\textsuperscript{16}. Using decision tree analysis, Andrew et al. discovered that patients with a high Glasgow Coma Scale score, less than 44.5 mL hematoma volume, and a relatively low premorbid modified Rankin Score have a changeable prognosis\textsuperscript{9}. However, the use of machine learning to predict outcomes following ICH is still uncommon. The majority of research looking for ICH predictors have used univariable and multivariable logistic regression analysis. In general, the accuracy of these regression models is poor. In our study, we found that the SVM model achieved the best comprehensive prediction efficiency which was better than previous studies. SVM is a promising supervised ML that can learn from observing data sets, build complex models to capture the inherent relationships between input and output variables, and make data-driven predictions or decisions. It has been used successfully in a variety of biomedicine fields and has performed admirably\textsuperscript{1, 17, 26}. The
application of the SVM machine learning approach to predict the outcome of ICH was innovative in our studies. Our model achieved a sensitivity of 0.93, specificity of 0.92, and AUC of 0.91. This indicated that our model performed well in predicting the outcome of ICH and could be widely used in clinics.

There were several limitations to the study. To begin, this was retrospective research conducted at a single center. Large clinical data sets and multicenter validation are necessary to improve the model's performance and to make precise comparisons among machine learning-based models. Furthermore, our study did not seek to pinpoint the precise reason of a particular good or bad result. Although our approach cannot replace clinical judgment, result prediction based on long-term human experience may exceed the best algorithm. In the future, more complex machine learning approaches will be applied to increase AUC.

**CONCLUSION**

The analysis of clinical reliability showed that the SVM model achieved the best comprehensive prediction efficiency than other models. For the SVM model, we found the importance value of TOR was higher significantly than other variables. Using machine learning techniques, we may be able to identify the factors that have been previously neglected and to develop new treatment methods to improve outcomes according to these factors.

**AUTHORS' DECLARATION**
Conflicts of interest

No potential conflict of interest relevant to this article was reported.

Informed consent

This type of study does not require informed consent.

Data sharing

None

Preprint

None

ORCID

Shengli Li  https://orcid.org/0000-0002-4040-4363
Jianan Zhang https://orcid.org/0000-0003-1879-1907
Xiaoqun Hou https://orcid.org/0000-0002-2094-5410
Yongyi Wang https://orcid.org/0000-0003-1111-1294
Tong Li  https://orcid.org/0000-0002-2608-2855
Zhiming Xu  https://orcid.org/0000-0002-2426-7549
Feng Chen  https://orcid.org/0000-0002-4184-9719
Yong Zhou https://orcid.org/0000-0003-4261-3385

Weimin Wang https://orcid.org/0000-0003-0833-1739

Mingxing Liu https://orcid.org/0000-0002-0030-3146

• Acknowledgements

This work was supported by the China’s government under grant of National Natural Science Foundation (#82001184 & 82001253).

References


12. Kellner CP, Schupper AJ, Mocco J: Surgical Evacuation of Intracerebral Hemorrhage:


Fig. 1. Flow chart outlining inclusion and exclusion criteria.
Fig. 2. Modeling process using ML. Four machine learning models, including Support Vector Machine (SVM), Decision Tree C5.0, Artificial Neural Network (ANN), Logistic Regression (LR) were used to build ICH prediction models.
Fig. 3. The ROC analysis of ML models. We established four ML models. For the SVM, the Decision tree C5.0, ANN, and LR models, the AUC were 0.91, 0.96, 0.91, 0.77 respectively.
Fig. 4. The variable importance values in the four models. For the SVM and LR model, we found the importance value of TOR was higher significantly than other variables. For the Decision tree C5.0 model, the sequence of importance value was midline shift, GCS, SBP and TOR. For the ANN model, the sequence of importance value was midline shift, TOR, DBP, and SBP. Moreover, the importance values of the above variables in the Decision tree C5.0 and ANN model were similar.
### Table 1. Clinical Summary of Patients’ Data

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Favorable outcome group (N=71)</th>
<th>Unfavorable outcome group (N=156)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (year)</td>
<td>62.24±7.79</td>
<td>61.76±12.47</td>
<td>0.070</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td>0.298</td>
</tr>
<tr>
<td>Male</td>
<td>49 (69.0%)</td>
<td>95 (60.9%)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>22 (31.0%)</td>
<td>61 (39.1%)</td>
<td></td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td>25.51±2.59</td>
<td>26.09±4.34</td>
<td>0.209</td>
</tr>
<tr>
<td>Diabetes (n, %)</td>
<td>51 (71.8%)</td>
<td>132 (84.6%)</td>
<td>0.030</td>
</tr>
<tr>
<td>Hypertension (n, %)</td>
<td>65 (91.5%)</td>
<td>148 (94.9%)</td>
<td>0.376</td>
</tr>
<tr>
<td>Hyperlipidemia (n, %)</td>
<td>8 (11.3%)</td>
<td>15 (9.6%)</td>
<td>0.813</td>
</tr>
<tr>
<td>SBP (mmHg)</td>
<td>154.37±29.40</td>
<td>169.23±32.87</td>
<td>0.001</td>
</tr>
<tr>
<td>DBP (mmHg)</td>
<td>88.06±15.39</td>
<td>93.10±18.76</td>
<td>0.034</td>
</tr>
<tr>
<td>GCS</td>
<td>10.58±5.02</td>
<td>6.74±3.54</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
<td>0.075</td>
</tr>
<tr>
<td>Thalamus, Basal ganglia</td>
<td>32 (45.1%)</td>
<td>95 (60.9%)</td>
<td></td>
</tr>
<tr>
<td>Frontal, temporal, parietal, occipital lobe</td>
<td>34 (47.9%)</td>
<td>53 (34.0%)</td>
<td></td>
</tr>
<tr>
<td>Cerebellum</td>
<td>5 (7.0%)</td>
<td>6 (3.8%)</td>
<td></td>
</tr>
<tr>
<td>Brainstem</td>
<td>0</td>
<td>2 (1.3%)</td>
<td></td>
</tr>
<tr>
<td>Affected side</td>
<td></td>
<td></td>
<td>0.474</td>
</tr>
<tr>
<td></td>
<td>Left</td>
<td>Right</td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------------</td>
<td>-----------------------</td>
<td>-------</td>
</tr>
<tr>
<td></td>
<td>41(57.7%)</td>
<td>81(51.9%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30(42.3%)</td>
<td>75(48.1%)</td>
<td></td>
</tr>
<tr>
<td>Midline shift (mm)</td>
<td>4.01±3.57</td>
<td>7.49±5.02</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Hematoma volume (ml)</td>
<td>42.56±26.62</td>
<td>62.17±36.47</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Residual Hematoma volume (ml)</td>
<td>10.10±11.81</td>
<td>10.91±13.92</td>
<td>0.670</td>
</tr>
<tr>
<td>Operation type</td>
<td>0622</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Craniotomy surgery</td>
<td>55(77.5%)</td>
<td>115(73.7%)</td>
<td></td>
</tr>
<tr>
<td>Minimally invasive surgery</td>
<td>16(22.5%)</td>
<td>41(26.3%)</td>
<td></td>
</tr>
<tr>
<td>Time to operation room (h)</td>
<td>26.83±22.26</td>
<td>29.66±47.09</td>
<td>0.539</td>
</tr>
<tr>
<td>Time to operation room group</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;=12h</td>
<td>14(19.7%)</td>
<td>110(70.5%)</td>
<td></td>
</tr>
<tr>
<td>&gt;12h, &lt;36h</td>
<td>46(64.8%)</td>
<td>15(9.6%)</td>
<td></td>
</tr>
<tr>
<td>&gt;=36h</td>
<td>11(15.5%)</td>
<td>31(19.9%)</td>
<td></td>
</tr>
<tr>
<td>Operation time (h)</td>
<td>153.03±57.13</td>
<td>165.71±70.86</td>
<td>0.187</td>
</tr>
</tbody>
</table>
Table 2. The comparison of prediction efficiency for four models

<table>
<thead>
<tr>
<th>Items</th>
<th>AUC</th>
<th>Accuracy</th>
<th>Specificity</th>
<th>Sensitivity</th>
<th>PLR</th>
<th>NLR</th>
<th>DOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>0.91</td>
<td>0.92</td>
<td>0.92</td>
<td>0.93</td>
<td>11.63</td>
<td>0.076</td>
<td>153.03</td>
</tr>
<tr>
<td>Decision tree</td>
<td>0.96</td>
<td>0.87</td>
<td>0.88</td>
<td>0.87</td>
<td>7.25</td>
<td>0.148</td>
<td>48.99</td>
</tr>
<tr>
<td>C5.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANN</td>
<td>0.91</td>
<td>0.91</td>
<td>0.82</td>
<td>0.97</td>
<td>5.39</td>
<td>0.037</td>
<td>145.68</td>
</tr>
<tr>
<td>LR</td>
<td>0.77</td>
<td>0.81</td>
<td>0.73</td>
<td>0.85</td>
<td>3.15</td>
<td>0.205</td>
<td>15.37</td>
</tr>
</tbody>
</table>

AUC: area under the curve, PLR: positive likelihood ratio, NLR: negative likelihood ratio,
DOR: diagnostic odds ratio, SVM: support vector machine, ANN: Artificial Neural Network, LR: logistic regression