Background: Patients suffered aneurysmal subarachnoid hemorrhage (aSAH) usually develop poor survival and functional outcome. Evaluating aSAH patients at high risk of poor outcome is necessary for clinicians to make suitable therapeutic strategy. This study is conducted to develop prognostic model using XGBoost (extreme gradient boosting) algorithm in aSAH.

Methods: A total of 351 aSAH patients admitted to West China hospital were identified. Patients were divided into training set and test set with ratio of 7:3 to testify the predictive value of XGBoost based prognostic model. Additionally, logistic regression model was also constructed and compared with XGBoost based model. Area under the receiver operating characteristic curve (AUC), sensitivity and specificity were calculated to evaluate the value of XGBoost and logistic regression.

Results: There were 74 (21.1%) non-survivors and 148 (42.1%) patients with unfavorable functional outcome. Non-survivors had older age (p=0.025), lower Glasgow coma scale (GCS) (p<0.001), higher World Federation of Neurosurgical Societies WFNS score (p<0.001), mFisher score (p=0.001). The incidence of intraventricular hemorrhage (IVH) (p=0.025) and delayed cerebral ischemia (DCI) (p<0.001) was higher in non-survivors than survivors. The AUC of XGBoost model for predicting mortality and unfavorable functional outcome were 0.950 and 0.958, which were higher than 0.767 and 0.829 of logistic regression model.

Conclusion: XGBoost based model is more precise than logistic regression model in predicting outcome of aSAH patients. Using XGBoost prognostic model is helpful for clinicians to identify high-risk aSAH patients and therefore strengthen medical care.

Keywords: machine learning, artificial intelligence, extreme gradient boosting, aneurysmal subarachnoid hemorrhage, prognosis

Introduction
Occurred widely with the incidence of nearly 9.1 per 100 thousand persons annually, aneurysmal subarachnoid hemorrhage (aSAH) leads to high morbidity and mortality around the world.1,2 Nearly a half of aSAH patients suffer poor survival outcome or unrecovered functional independence.3–5 In addition to initial brain injury severity, secondary complications such as re-bleeding, delayed cerebral ischemia (DCI), acute hydrocephalus and seizure could lead to deteriorating outcome of aSAH patients.6–8 Therefore, predicting the outcome of aSAH patients in acute phase is essential for physicians to reasonably classify patients’ severity and sequentially adopt different treatment strategies.

Many clinical scores including Hunt and Hess score, World Federation of Neurosurgical Societies (WFNS) score and modified Fisher (mFisher) score has been utilized to assess brain severity of aSAH patients. These scores are mainly developed based on focal neurological signs, state of consciousness or radiological findings. Whereas intracranial complications and pathophysiological changes of extracranial organs function during hospitalizations which also play important role in disease progression of aSAH are not included into existing score. Some researches combined these complications and biochemical parameters into prognostic models of aSAH.9–14 However, these models were mainly...
constructed through linear regression such as Cox regression and logistic regression which strictly requires high quality of data and could not handle missing values efficiently. And some markers showing non-linear relationship with outcome of aSAH such as serum sodium and serum chloride may not be applicable to develop linear relationship based prediction.\textsuperscript{15–18}

As a kind of artificial intelligence, machine learning algorithms could improve event prediction accuracy due to their advantages in recognizing complex data pattern and dealing with nonlinear relationship efficiently. Recently, the importance of machine learning in clinical decision making has been gradually recognized by clinicians. Machine learning based prediction approaches could automatically provide warning signals to physicians about possible impending high-risk events of patients and assist physicians to make suitable therapeutic strategies. The prognostic value of some machine learning algorithms such as decision tree, random forest, support vector machine and Naive Bayes have been explored in aSAH patients.\textsuperscript{19–26} Developed by Tianqi Chen, the Extreme Gradient Boosting (XGBoost) algorithm performs well in both computational speed and model accuracy. One previous study explored performance of various machine learning algorithms including XGBoost on predicting mortality of non-traumatic SAH patients.\textsuperscript{23} While only records of laboratory tests and drugs usage were collected for training machine learning models without inclusion of complications such as DCI and hydrocephalus which may significantly affect prognosis of aSAH patients. The machine learning algorithm may show its superiority in outcome prediction when more clinically significant factors were included or framework of datasets were more complicated. Additionally, this study did not evaluate the value of XGBoost on predicting functional outcome of aSAH patients. Therefore, we designed this study to develop XGBoost models incorporating complications and other clinical variables to predict mortality and unfavorable functional outcome in aSAH patients and further compare XGBoost models’ value with conventional logistic regression based prediction approach.

**Patients**

Patients admitted to West China hospital and received treatments in neuro-intensive care unit (NICU) for aSAH between January 2017 and June 2019 were identified. The diagnosis of aSAH was confirmed by computed tomography angiography (CTA) or digital subtraction angiography (DSA). The exclusion criteria were presented as follows: (1) SAH caused by other diseases including cerebral vascular malformation, Moyamoya disease, trauma; (2) history of other neurologic diseases such as intracranial tumor and stroke; (3) admitted to our hospital 48 hours after initial symptoms onset; (4) transferred from other medical centers; (5) Incomplete records of included variables. Finally, 351 aSAH patients were included in this study. This study obtained approval from the West China hospital ethics committee. All procedures involved in this study were in accordance with the ethical standards of the Helsinki declaration. We have obtained the informed consent from all patients or their legal guardians.

**Data Collection**

Collected variables included (1) patients demographics and vital signs on admission; (2) comorbidities (diabetes mellitus, hypertension); (3) GCS, WFNS, mFishers score; (4) aneurysm location and occurrence of Intraventricular hemorrhage; (5) laboratory tests (glucose, hemoglobin, platelet) analyzed from the first blood sample once admitted; (6) complications during hospitalization including acute hydrocephalus, delayed cerebral ischemia (DCI) and intracranial infection. The DCI was diagnosed based on one of the following criteria: (1) occurrence of focal neurological impairment; (2) a dramatic decrease of GCS at least 2 points (decrease of total score or on one of GCS individual components, and excluded other causes discovered by clinical assessments and brain image). The outcome of this study was mortality and unfavorable functional outcome (defined by GOS<4). Patients were followed up until 3 months after admission.

**Statistical Analysis**

Normality of included variables was verified by Kolmogorov–Smirnov test. Normally distributed and non-normally distributed variables were presented in the form of mean ± standard deviation and median (interquartile range), respectively. And categorical variables were shown as numbers (percentage). Independent Student’s \(t\)-test and Mann–Whitney \(U\)-test were utilized to compare the difference between two groups of normally distributed and non-normally distributed variables. \(\chi^2\) test or Fisher test was performed to analyze the difference of categorical variables.
Based on gradient boosted decision trees, the XGBoost algorithm applies a second-order Taylor expansion to calculate the loss function and has an excellent performance in both computational speed and prediction accuracy. The XGBoost continuously split features to grow trees to iteratively fit the residual of the last prediction model. K trees would be developed after training process. The characteristics of one specific sample correspond to specific leaf node of each tree with a score. The predicted value of one specific sample would be obtained by adding up scores of each tree. Gathering a strong classifier from a set of weak classifiers, XGBoost displays following advantages: (1) handling missing values effectively; (2) preventing overfitting; (3) reducing running time by the parallel and distributed calculation. The hyper-parameters of XGBoost were set with max depth of 10, learning rate of 0.3 and iterations of 50, considering both accuracy and over fitting. To assess the predictive value of XGBoost algorithm in aSAH patients, included patients were divided into training set and test set with ratio of 7:3. Logistic regression prediction model was also developed by univariate and multivariate logistic regression with forward stepwise method. To compare the predictive value of XGBoost and logistic regression, receiver operating characteristics (ROC) curves of these two models were drawn and area under the ROC curve (AUC) was calculated. In addition, sensitivity, specificity, accuracy, false positive rate (FPR), false negative rate (FNR), positive predictive value (PPV) and negative predictive value (NPV) of these two models were calculated and compared. Two-sided P value <0.05 was considered being statistically significant. R (version 3.6.1; R Foundation) was applied for all statistical analyses and figures drawing. And the XGBoost model was developed using “xgboost” package in R.

**Results**

**Patients Characteristics**

A total of 351 aSAH patients were included into this study with mortality of 21.1% (Table 1). Non-survivors had higher age than survivors (60 vs 55, p=0.025). The complicated rate of diabetes mellitus and hypertension was 4.8% and 43.9%, respectively. GCS was significantly lower in non-survivors (7 vs 15, p<0.001) while WFNS (4 vs 1, p<0.001) and mFisher (3 vs 2, p<0.001) were significantly higher in non-survivors. Regarding the aneurysm location, anterior circulation accounted for most of patients with incidence of 76.9%. Compared with survivors, non-survivors had higher incidence of posterior circulation aneurysm (10.8% vs 3.2%, p=0.024). In addition, intraventricular hemorrhage was also more likely to be observed in non-survivors (12.2% vs 4.3%, p=0.025). Results of laboratory tests showed non-survivors had higher level of blood glucose (7.57 vs 6.62, p<0.001). Finally, the incidence of delayed cerebral ischemia was higher in non-survivors group (54.1% vs 13.0%, p<0.001) while hydrocephalus and intracranial infection did not show statistical significance between these two groups. The incidence of unfavorable functional outcome in this study was 42.1%.

**Predictive Value of XGBoost and Logistic Regression**

The feature importance of XGBoost for predicting mortality and unfavorable functional outcome was shown as Figure 1. GCS, DCI and heart rate ranked the top three of feature importance for predicting mortality while DCI, GCS and age ranked the top three of feature importance for predicting unfavorable functional outcome. The detail of logistic models was shown in Supplementary Table 1 and Supplementary Table 2. The AUC of XGBoost and logistic regression for predicting mortality was 0.950 and 0.767, respectively (Table 2) (Figure 2). XGBoost had higher predictive accuracy than logistic regression (0.981 vs 0.830). Sensitivity and specificity of XGBoost was 1.000 and 0.900, which was also higher than 0.837 and 0.750 of logistic regression. As for predicting unfavorable functional outcome, the AUC of XGBoost and logistic regression was 0.958 and 0.829, respectively. XGBoost had higher predictive accuracy than logistic regression (0.962 vs 0.764). Sensitivity and specificity of XGBoost was 0.984 and 0.933, which was higher than 0.836 and 0.756 of logistic regression.

**Discussion**

Our study indicated that XGBoost based prediction approach was more precise than conventional logistic regression prediction in evaluating prognosis of aSAH patients. The AUC of XGBoost approach for predicting mortality and unfavorable functional outcome was 0.950 and 0.958, which indicated excellent prediction accuracy. There were several studies exploring the prognostic value of other machine learning algorithm in aSAH patients. A study conducted by Nora

---

*References and Notes*

1. **Patients Characteristics**

A total of 351 aSAH patients were included into this study with mortality of 21.1% (Table 1). Non-survivors had higher age than survivors (60 vs 55, p=0.025). The complicated rate of diabetes mellitus and hypertension was 4.8% and 43.9%, respectively. GCS was significantly lower in non-survivors (7 vs 15, p<0.001) while WFNS (4 vs 1, p<0.001) and mFisher (3 vs 2, p<0.001) were significantly higher in non-survivors. Regarding the aneurysm location, anterior circulation accounted for most of patients with incidence of 76.9%. Compared with survivors, non-survivors had higher incidence of posterior circulation aneurysm (10.8% vs 3.2%, p=0.024). In addition, intraventricular hemorrhage was also more likely to be observed in non-survivors (12.2% vs 4.3%, p=0.025). Results of laboratory tests showed non-survivors had higher level of blood glucose (7.57 vs 6.62, p<0.001). Finally, the incidence of delayed cerebral ischemia was higher in non-survivors group (54.1% vs 13.0%, p<0.001) while hydrocephalus and intracranial infection did not show statistical significance between these two groups. The incidence of unfavorable functional outcome in this study was 42.1%.

**Predictive Value of XGBoost and Logistic Regression**

The feature importance of XGBoost for predicting mortality and unfavorable functional outcome was shown as Figure 1. GCS, DCI and heart rate ranked the top three of feature importance for predicting mortality while DCI, GCS and age ranked the top three of feature importance for predicting unfavorable functional outcome. The detail of logistic models was shown in Supplementary Table 1 and Supplementary Table 2. The AUC of XGBoost and logistic regression for predicting mortality was 0.950 and 0.767, respectively (Table 2) (Figure 2). XGBoost had higher predictive accuracy than logistic regression (0.981 vs 0.830). Sensitivity and specificity of XGBoost was 1.000 and 0.900, which was also higher than 0.837 and 0.750 of logistic regression. As for predicting unfavorable functional outcome, the AUC of XGBoost and logistic regression was 0.958 and 0.829, respectively. XGBoost had higher predictive accuracy than logistic regression (0.962 vs 0.764). Sensitivity and specificity of XGBoost was 0.984 and 0.933, which was higher than 0.836 and 0.756 of logistic regression.

**Discussion**

Our study indicated that XGBoost based prediction approach was more precise than conventional logistic regression prediction in evaluating prognosis of aSAH patients. The AUC of XGBoost approach for predicting mortality and unfavorable functional outcome was 0.950 and 0.958, which indicated excellent prediction accuracy. There were several studies exploring the prognostic value of other machine learning algorithm in aSAH patients. A study conducted by Nora
Franziska Dengler in 2020 showed that machine learning algorithms including CatBoost tree boosting, support vector machine classifier, Naive Bayes classifier and multilayer perceptron artificial neural net was comparable but not superior to traditional logistic regression in predicting outcome of aSAH patients. And the predictive value of these algorithms was not superior than previously developed clinical-radiographic scores such as Hunt-Huss, WFNS, mFisher and Barrow

Table 1 Baseline Characteristics of Included aSAH Patients

<table>
<thead>
<tr>
<th>Variables</th>
<th>Overall Patients (n=351)</th>
<th>Survivors (n=277, 78.9%)</th>
<th>Non-Survivors (n=74, 21.1%)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>56 (49–65)</td>
<td>55 (49–64)</td>
<td>60 (51–68)</td>
<td>0.025</td>
</tr>
<tr>
<td>Female gender</td>
<td>231 (65.8%)</td>
<td>183 (66.1%)</td>
<td>48 (64.9%)</td>
<td>0.956</td>
</tr>
<tr>
<td>Prehospital time (hours)</td>
<td>9 (5–24)</td>
<td>9 (5–24)</td>
<td>9 (4–24)</td>
<td>0.875</td>
</tr>
<tr>
<td>Smoking</td>
<td>59 (16.8%)</td>
<td>47 (17.0%)</td>
<td>12 (16.2%)</td>
<td>1.000</td>
</tr>
<tr>
<td>Alcoholism</td>
<td>47 (13.4%)</td>
<td>41 (14.8%)</td>
<td>6 (8.1%)</td>
<td>0.190</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>17 (4.8%)</td>
<td>15 (5.4%)</td>
<td>2 (2.7%)</td>
<td>0.509</td>
</tr>
<tr>
<td>Hypertension</td>
<td>154 (43.9%)</td>
<td>125 (45.1%)</td>
<td>29 (39.2%)</td>
<td>0.434</td>
</tr>
<tr>
<td>Systolic blood pressure (mmHg)</td>
<td>150 (134–170)</td>
<td>152 (130–170)</td>
<td>146 (133–172)</td>
<td>0.996</td>
</tr>
<tr>
<td>Diastolic blood pressure (mmHg)</td>
<td>87 (78–98)</td>
<td>87 (77–98)</td>
<td>88 (79–97)</td>
<td>0.248</td>
</tr>
<tr>
<td>Heart rate (min⁻¹)</td>
<td>79 (70–89)</td>
<td>79 (70–88)</td>
<td>80 (65–97)</td>
<td>0.741</td>
</tr>
<tr>
<td>GCS</td>
<td>15 (9–15)</td>
<td>15 (12–15)</td>
<td>7 (5–12)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>WFNS</td>
<td>2 (1–4)</td>
<td>1 (1–4)</td>
<td>4 (4–5)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>mFisher</td>
<td>2 (2–3)</td>
<td>2 (2–3)</td>
<td>3 (2–4)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
<td></td>
<td>0.024</td>
</tr>
<tr>
<td>Anterior circulation</td>
<td>270 (76.9%)</td>
<td>218 (78.7%)</td>
<td>52 (70.3%)</td>
<td></td>
</tr>
<tr>
<td>Posterior circulation</td>
<td>17 (4.8%)</td>
<td>9 (3.2%)</td>
<td>8 (10.8%)</td>
<td></td>
</tr>
<tr>
<td>Multiple aneurysm</td>
<td>64 (18.2%)</td>
<td>50 (18.1%)</td>
<td>14 (18.9%)</td>
<td></td>
</tr>
<tr>
<td>Intraventricular hemorrhage</td>
<td>21 (6.0%)</td>
<td>12 (4.3%)</td>
<td>9 (12.2%)</td>
<td>0.025</td>
</tr>
<tr>
<td>Glucose (mmol/L)</td>
<td>6.93 (5.75–8.18)</td>
<td>6.62 (5.66–7.91)</td>
<td>7.57 (6.99–9.23)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Hemoglobin (g/L)</td>
<td>125 (110–135)</td>
<td>124 (109–135)</td>
<td>127 (114–136)</td>
<td>0.141</td>
</tr>
<tr>
<td>Platelet (×10⁹/L)</td>
<td>161.85 (52.24)</td>
<td>159.65 (48.72)</td>
<td>170.11 (63.40)</td>
<td>0.126</td>
</tr>
<tr>
<td>Hydrocephalus</td>
<td>34 (9.7%)</td>
<td>24 (8.7%)</td>
<td>10 (13.5%)</td>
<td>0.302</td>
</tr>
<tr>
<td>Delayed cerebral ischemia</td>
<td>76 (21.7%)</td>
<td>36 (13.0%)</td>
<td>40 (54.1%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Intracranial infection</td>
<td>37 (10.5%)</td>
<td>29 (10.5%)</td>
<td>8 (10.8%)</td>
<td>1.000</td>
</tr>
<tr>
<td>GOS</td>
<td>4 (3–4)</td>
<td>4 (3–4)</td>
<td>1 (1–1)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>GOS (1–3)</td>
<td>148 (42.1%)</td>
<td>74 (26.7%)</td>
<td>74 (100%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Length of ICU stay (days)</td>
<td>6 (4–14)</td>
<td>6 (4–15)</td>
<td>6 (2–10)</td>
<td>0.255</td>
</tr>
<tr>
<td>Length of hospital stay (days)</td>
<td>15 (10–23)</td>
<td>16 (11–24)</td>
<td>9 (5–16)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Abbreviations: GCS, Glasgow coma scale; WFNS, World Federation of Neurosurgical Societies; mFisher, modified Fisher; GOS, Glasgow outcome scale.
Neurological Institute (BNI) Score. However, this study only included basic clinical and radiographic features and did not collect intracranial complications and other biochemical parameters. The advantages of dealing with high dimensional dataset of machine learning algorithm may not be highlighted in datasets with simple framework. Instead, the research conducted by Nicolai Maldaner in 2020 applied machine learning algorithms into outcome prediction of aSAH and included complications during hospitalizations including rebleeding, DCI and epileptic seizure. And results of this study presented that Chi-square Automatic Interaction Detectors (CHAID) was the best algorithm for predicting favorable functional outcome at discharge of aSAH patients. Adding complications and medical treatments related information into machine learning based model could improve the AUC from 0.79 to 0.85. Guido de Jong performed a study comparing predictive value of feedforward artificial neural networks (ffANN) with Subarachnoid Haemorrhage International Trialists (SAHIT) and VASOGRADE scoring systems for predicting prognosis and DCI using basic demographics and aneurysm related variables. It was discovered that ffANN showed equal performance when compared with SAHIT and VASOGRADE. In summary, machine learning algorithms may show it’s superiority in outcome prediction when more clinically significant factors were included or framework of datasets were more complicated. Therefore, complications during hospitalizations and biochemical parameters on admission of aSAH patients were both included when we developed XGBoost based prediction approach.

Based on gradient boosted decision trees, the XGBoost algorithm applied a second-order Taylor expansion to calculate the loss function and performs well in both computational speed and predictive precision. In the XGBoost based approach for predicting mortality of aSAH patients in our study, the top order of feature importance sequentially was GCS, DCI, heart rate, age, systolic blood pressure, glucose, platelet and prehospital time. And in the XGBoost based approach for predicting 3-month unfavorable functional outcome of aSAH patients, the top order of feature importance sequentially was DCI, GCS, age, WFNS, platelet, hemoglobin, systolic blood pressure, glucose, heart rate and prehospital time. The GCS score is a widely used assessment tool of brain injury severity for several decades. Whereas some limitations exist in the clinical practice including inability to accurately evaluate severity when patients suffered alcoholism, intubated or received sedatives. Consequently, many studies combine other risk factors with GCS to improve prognostic accuracy and stability. Characterized as local neurological deficit, decreased GCS and cerebral ischemia presented in radiological examinations, the DCI is a severe neurological complication which is closely correlated with poor prognosis of aSAH patients. Multiple factors may play significant role in the process of DCI including cerebral artery vasospasm, micro-thrombosis, cerebrovascular dysregulation, parenchymal inflammation and cortical spreading

**Figure 1** (A) Feature importance of factors in XGBoost for predicting mortality of aSAH patients; (B) Feature importance of factors in XGBoost for predicting 3-month unfavorable functional outcome of aSAH patients. The feature importance was automatically divided into three clusters by XGBoost according to the importance rank.
Table 2 The Prognostic Value Comparison Between Logistic Regression and Xgboost Algorithm

<table>
<thead>
<tr>
<th></th>
<th>AUC</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Youden Index</th>
<th>Accuracy</th>
<th>FPR</th>
<th>FNR</th>
<th>PPV</th>
<th>NPV</th>
<th>F Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logistic regression</td>
<td>0.767</td>
<td>0.837</td>
<td>0.750</td>
<td>0.587</td>
<td>0.830</td>
<td>0.116</td>
<td>0.400</td>
<td>0.545</td>
<td>0.905</td>
<td>0.860</td>
</tr>
<tr>
<td>Xgboost algorithm</td>
<td>0.950</td>
<td>1.000</td>
<td>0.900</td>
<td>0.900</td>
<td>0.981</td>
<td>0.000</td>
<td>0.100</td>
<td>1.000</td>
<td>0.977</td>
<td>0.990</td>
</tr>
<tr>
<td>GOS (1–3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logistic regression</td>
<td>0.829</td>
<td>0.836</td>
<td>0.756</td>
<td>0.592</td>
<td>0.764</td>
<td>0.098</td>
<td>0.422</td>
<td>0.813</td>
<td>0.743</td>
<td>0.803</td>
</tr>
<tr>
<td>Xgboost algorithm</td>
<td>0.958</td>
<td>0.984</td>
<td>0.933</td>
<td>0.917</td>
<td>0.962</td>
<td>0.016</td>
<td>0.067</td>
<td>0.977</td>
<td>0.952</td>
<td>0.976</td>
</tr>
</tbody>
</table>

Abbreviations: AUC, area under the receiver operating characteristics curve; FPR, false positive rate; FNR, false negative rate; PPV, positive predictive value; NPV, negative predictive value; GOS, Glasgow outcome scale.
The abnormal change of heart rate and systolic blood pressure may reflect influenced cardiac function in aSAH patients. Actually, cardiac complications frequently occur in aSAH patients and are commonly manifested as electrocardiogram abnormalities and stress cardiomyopathy. Caused by multiple factors such as neurogenic activation of sympathetic system, immune response, the cardiac dysfunction may lead to cerebral hypoperfusion and DCI which in turn aggravate the adverse progress of aSAH patients. Increased age also indicates higher likelihood of poor prognosis due to more underlying diseases, worse nutritional status and higher incidence of frailty. Finally, glucose and platelet also showed a degree of feature importance in the XGBoost prediction approach. It has been confirmed hyperglycemia and platelet play critical role of in the pathophysiological process of aSAH and correlated with prognosis of these patients. In summary, our constructed XGBoost prediction approach incorporating above mentioned risk factors may reflect more comprehensive condition of aSAH patients and show better predictive accuracy and stability than conventional logistic regression prediction.

This study had several limitations. Firstly, this was an observational research conducted in single medical center. aSAH patients included in this study were those who received critical care in NICU. Further studies conducted in multiple medical centers with larger sample size are worthwhile to externally verify the predictive value of XGBoost algorithm in general aSAH patients and try to avoid selection bias. Secondly, though XGBoost algorithm was confirmed as a precise prediction approach in aSAH patients, it was not developed into a conveniently used program in this study. Future works should be performed to translate this algorithm into clinical practice. Thirdly, some complications of clinical significance such as seizure, re-bleeding, pneumonia were not included due to lack of evaluation and records. Studies collecting these variables may further improve the prediction accuracy of our proposed models.

Conclusion
The XGBoost algorithm based prediction approach is more effective than logistic regression model in prognostication of aSAH patients. Developing XGBoost prediction approach could instruct clinicians adjust therapeutic strategies and strengthen clinical care for aSAH patients.

Data Sharing Statement
The datasets used for the current study are available from the corresponding author on reasonable request.
Ethics Approval and Informed Consent
This study obtained approval from the West China hospital ethics committee (Number: 20180359). All procedures involved in this study were in accordance with the ethical standards of the Helsinki declaration.

Consent for Publication
We have obtained the informed consent from all patients or their legal guardians.

Author Contributions
All authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; took part in drafting the article or revising it critically for important intellectual content; agreed to submit to the current journal; gave final approval of the version to be published; and agree to be accountable for all aspects of the work.

Funding
This work was supported by Sichuan Science and Technology Innovation Seedling Project (No.21-YCG021, sub-No. 2021026), Key research and development project of science and technology department of Sichuan Province (2019YFS0392), and the 1.3.5 project for disciplines of excellence, West China Hospital, Sichuan University (ZYJC18007).

Disclosure
Ruoran Wang and Jiang Zhang are the co-first authors of this manuscript. The authors declare that they have no competing interests.

References


