#### Check for updates

#### OPEN ACCESS

EDITED BY Fumiharu Ohka, Nagoya University, Japan

REVIEWED BY Adam James Wells, University of Adelaide, Australia Jacek Szczygielski, University of Rzeszow, Poland

\*CORRESPONDENCE Zhihui Ni [nizhihui@wmu.edu.cn](mailto:nizhihui@wmu.edu.cn) Jianjing Yang [yangjianjing@wmu.edu.cn](mailto:yangjianjing@wmu.edu.cn) Qichuan Zhuge [qc.zhuge@wmu.edu.cn](mailto:qc.zhuge@wmu.edu.cn)

RECEIVED 03 October 2023 ACCEPTED 28 March 2024 PUBLISHED 22 April 2024

#### CITATION

Ni Z, Zhu Y, Qian Y, Li X, Xing Z, Zhou Y, Chen Y, Huang L, Yang J and Zhuge Q (2024) Synthetic minority over-sampling techniqueenhanced machine learning models for predicting recurrence of postoperative chronic subdural hematoma. *Front. Neurol.* 15:1305543. [doi: 10.3389/fneur.2024.1305543](https://doi.org/10.3389/fneur.2024.1305543)

#### **COPYRIGHT**

© 2024 Ni, Zhu, Qian, Li, Xing, Zhou, Chen, Huang, Yang and Zhuge. This is an openaccess article distributed under the terms of the Creative Commons Attribution Licens [\(CC BY\)](http://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# [Synthetic minority over-sampling](https://www.frontiersin.org/articles/10.3389/fneur.2024.1305543/full)  [technique-enhanced machine](https://www.frontiersin.org/articles/10.3389/fneur.2024.1305543/full)  [learning models for predicting](https://www.frontiersin.org/articles/10.3389/fneur.2024.1305543/full)  [recurrence of postoperative](https://www.frontiersin.org/articles/10.3389/fneur.2024.1305543/full)  [chronic subdural hematoma](https://www.frontiersin.org/articles/10.3389/fneur.2024.1305543/full)

Zhihui Ni<sup>1,2\*</sup>, Yehao Zhu<sup>1,2</sup>, Yiwei Qian<sup>1,2</sup>, Xinbo Li<sup>1,2</sup>, Zhenqiu Xing<sup>1,2</sup>, Yinan Zhou<sup>1,2</sup>, Yu Chen<sup>1,2</sup>, Lijie Huang<sup>1,2</sup>, Jianjing Yang<sup>1,2\*</sup> and Qichuan Zhuge<sup>1,2\*</sup>

1 Department of Neurosurgery, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China, 2Zhejiang Provincial Key Laboratory of Aging and Neurological Disorder Research, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

Objective: Chronic subdural hematoma (CSDH) is a neurological condition with high recurrence rates, primarily observed in the elderly population. Although several risk factors have been identified, predicting CSDH recurrence remains a challenge. Given the potential of machine learning (ML) to extract meaningful insights from complex data sets, our study aims to develop and validate ML models capable of accurately predicting postoperative CSDH recurrence.

Methods: Data from 447 CSDH patients treated with consecutive burr-hole irrigations at Wenzhou Medical University's First Affiliated Hospital (December 2014-April 2019) were studied. 312 patients formed the development cohort, while 135 comprised the test cohort. The Least Absolute Shrinkage and Selection Operator (LASSO) method was employed to select crucial features associated with recurrence. Eight machine learning algorithms were used to construct prediction models for hematoma recurrence, using demographic, laboratory, and radiological features. The Border-line Synthetic Minority Over-sampling Technique (SMOTE) was applied to address data imbalance, and Shapley Additive Explanation (SHAP) analysis was utilized to improve model visualization and interpretability. Model performance was assessed using metrics such as AUROC, sensitivity, specificity, F1 score, calibration plots, and decision curve analysis (DCA).

Results: Our optimized ML models exhibited prediction accuracies ranging from 61.0% to 86.2% for hematoma recurrence in the validation set. Notably, the Random Forest (RF) model surpassed other algorithms, achieving an accuracy of 86.2%. SHAP analysis confirmed these results, highlighting key clinical predictors for CSDH recurrence risk, including age, alanine aminotransferase level, fibrinogen level, thrombin time, and maximum hematoma diameter. The RF model yielded an accuracy of 92.6% with an AUC value of 0.834 in the test dataset.

Conclusion: Our findings underscore the efficacy of machine learning algorithms, notably the integration of the RF model with SMOTE, in forecasting the recurrence of postoperative chronic subdural hematoma. Leveraging the RF model, we devised an online calculator that may serve as a pivotal instrument in tailoring therapeutic strategies and implementing timely preventive interventions for high-risk patients.

#### **KEYWORDS**

chronic subdural hematoma, machine learning, hematoma recurrence, shapley additive explanation, web calculator

# Introduction

Chronic subdural hematoma (CSDH) is a prevalent neurological disorder. Manifesting approximately 3weeks post-injury, its symptoms encompass focal neurological deficits, cognitive alterations, and signs of elevated intracranial pressure, primarily headaches and reduced consciousness [\(1](#page-11-0)). In severe cases, CSDH can be fatal [\(2\)](#page-11-1). The elderly population, particularly those above 65years of age, faces heightened risk due to widespread anticoagulation treatments, natural cerebral atrophy, and increased susceptibility to trauma [\(3,](#page-11-2) [4](#page-11-3)). While previously believed to be caused by slow venous bleeding from the brain's bridging veins following trauma, recent research indicates that CSDH's onset and recurrence are multifactorial, involving disruption to the cells lining the dura, inflammation, angiogenesis, coagulation disturbances, microbleeds, and exudation [\(5](#page-11-4)[–7\)](#page-11-5).

Burr-hole irrigation has emerged as an effective treatment for symptomatic CSDH. Yet, despite post-evacuation closed-system drainage, recurrence remains a significant clinical challenge, with rates hovering around 9–33% [\(7\)](#page-11-5). This recurrence presents considerable clinical conundrums, especially for older patients at elevated risk of neurological and surgical complications [\(8](#page-11-6)). Several factors, such as age, previous bleeding episodes, cerebral atrophy, alcohol consumption, the presence of subdural air, radiological signs, and surgical techniques, have been identified as potential contributors to recurrence ([9](#page-11-7)–[11](#page-11-8)). Nevertheless, existing predictive measures have shown inconsistent results and have not seen widespread clinical adoption [\(12\)](#page-11-9). In recent years, several models predicting the recurrence of CSDH after surgery have been published. To evaluate the performance of these existing models in predicting postoperative recurrence in CSDH patients, Holl et al. [\(13\)](#page-12-0) utilized a retrospective database comprising data from 2,384 patients across three regions in the Netherlands. The study revealed that current predictive models perform poorly on the author's dataset, failing to effectively forecast the recurrence of hematomas following CSDH treatment. This research highlights the challenges of predicting hematoma recurrence after CSDH treatment and underscores the necessity of adopting appropriate modeling strategies to develop high-quality models.

Machine learning (ML) has gained increasing influence in medical research, offering the ability to uncover hidden patterns and correlations from vast datasets ([14](#page-12-1)). Its capability for identifying complex data associations, often overlooked by traditional statistical methods, provides invaluable insights for both clinicians and researchers [\(15\)](#page-12-2). Machine learning models hold significant potential in neurosurgical predictive analytics and have been extensively applied to forecast acute hematoma expansion in cerebral hemorrhage, predict meningioma grade, and prognosticate outcomes for glioma patients, among other uses ([16](#page-12-3)–[18\)](#page-12-4). These models empower physicians and patients to make more informed decisions and offer personalized medical services ([19](#page-12-5)). SMOTE is an approach designed to tackle the issue of data imbalance, particularly within the fields of machine learning and statistics. It operates by inserting new synthetic samples between minority class instances, thus oversampling the dataset to increase the number of minority class samples. While this method effectively mitigates data imbalance, it may lack precision when dealing with samples near the decision boundary. The core concept of border-line SMOTE is to concentrate on those minority class samples that are challenging to classify, specifically those located near the boundary between majority and minority classes. These samples are deemed crucial for constructing the decision boundary. Therefore, generating more synthetic samples around these boundary-line samples can aid classifiers in better learning the characteristics of these complex regions, thereby enhancing the classification performance for minority class samples ([20\)](#page-12-6). Considering this, our study aims to develop and assess ML models, supplemented with SMOTE, for the accurate prediction of postoperative CSDH recurrence.

## **Methods**

#### Study population

We conducted a retrospective analysis of medical records from patients diagnosed with CSDH who underwent surgical treatment at the Department of Neurosurgery, First Affiliated Hospital of Wenzhou Medical University, from December 2014 to January 2019. Diagnoses of CSDH were confirmed via head MRI and CT scans. This study received approval from the Committee for Ethics in Clinical Research, and due to its retrospective design, there was no need for informed patient consent. Out of 632 adults diagnosed with CSDH, 447 with unilateral CSDH who underwent burr-hole irrigation were included in the final analysis. We excluded patients with: (1) bilateral chronic subdural hematoma, (2) severe renal or hematological conditions, (3) significant surgical complications or in-hospital fatalities, (4) those who underwent craniotomy or bone flap replacements, and (5) cases with incomplete records or lost during follow-up ([Figure 1\)](#page-2-0).

#### Surgical procedures and management

Under general anesthesia, all patients underwent the standard burr-hole irrigation (BHI) procedure. A single burr hole was strategically placed at the hematoma's thickest region, followed by saline irrigation. Subsequently, a silicone catheter equipped with a closed subdural drainage system was positioned within the hematoma cavity. Typically, the catheter was extracted between 24 to 72h postoperatively, contingent upon drainage volume. Postoperative administration of atorvastatin, sustained for a minimum of 1month, facilitated the absorption of residual hematoma and minimized the risk of recurrence (this approach is grounded on the findings of a randomized clinical trial conducted by Jiang et al. ([21](#page-12-7)), which demonstrated the safety and efficacy of atorvastatin in promoting hematoma absorption in Chinese patients with CSDH).

#### Model input features selected

We collected data from patient records, encompassing 32 unique clinical, radiological, and laboratory test characteristics. These included demographic features (age, gender) and lifestyle behaviors such as smoking and alcohol consumption. Pertinent comorbidities captured were hypertension, diabetes, and heart disease, along with any history of cranial trauma. Blood pressure measurements, including systolic (SBP) and diastolic (DBP), were documented. Initial CT or MRI scans, reviewed independently by two authors (LH and XL), revealed the hematoma's location and maximum diameter. Admission laboratory metrics comprised white blood cell count

<span id="page-2-0"></span>

(WBC), neutrophil percentage and count (NEUT), lymphocyte count (LYM), red blood cell count (RBC), hemoglobin (HB), platelet count (PLT), prothrombin time (PT), international normalized ratio (INR), fibrinogen (FIB), activated partial thromboplastin time (APTT), thrombin time (TT), total bilirubin (TB), direct bilirubin (DB), indirect bilirubin (IB), albumin (ALB), alanine aminotransferase (ALT), aspartate aminotransferase (AST), blood glucose (GLU), urea nitrogen (UN), and creatinine (CR). Using the least absolute shrinkage and selection operator (LASSO) regression technique, we identified 10 features significantly correlated with CSDH recurrence ([Figure 2](#page-3-0)).

#### Assessment of CSDH recurrence

Patients underwent a head CT or MRI scan within the initial 48h following surgery, and again on days 6 or 7 post-operation. To ensure comprehensive evaluation, a subsequent CT scan or MRI was performed at the outpatient clinic 3months post-surgery. The recurrence of CSDH was determined using specific radiological criteria. Regardless of any subsequent surgical interventions, any observed increase in subdural fluid volume and concurrent brain compression in either hemisphere, within 3months post-surgery relative to the initial postoperative CT, was deemed a recurrence. This determination was collaboratively made by two seasoned neurosurgeons, uninformed of the study's particulars.

# Balancing data for enhanced predictive modeling

Data from 312 out of 447 CSDH patients were utilized for the construction of the model, while the remaining 135 patients' data were employed to assess the model's predictive performance. In the training subset (312 patients), CSDH recurrence occurred in 15.1% (47 patients), whereas the remaining 84.9% (265 patients) experienced no recurrence. This significant data imbalance could compromise the performance of the predictive model. Balanced datasets are known to improve predictive accuracy. To address this issue, we employed the borderline-SMOTE, an advanced oversampling technique commonly used in machine learning. This approach improves upon the conventional SMOTE by targeting minority class samples located at the boundary between the majority and minority classes—areas where misclassification is prone to occur. By generating new samples around these crucial instances, Borderline-SMOTE enhances the classifier's discriminatory ability, particularly when class overlap is present. The application of borderline-SMOTE increased the minority class representation in our training dataset from 15.1 to 50%. This resulted in equal representation of CSDH patient groups, each with 265 instances of recurrence and non-recurrence (shown in [Figure 3\)](#page-3-1). This expanded dataset of 530 data points was subsequently divided into development (70%) and validation (30%) subsets for machine learning model construction.

<span id="page-3-0"></span>

<span id="page-3-1"></span>

## Statistical analysis

Statistical analyses were performed using R statistical software (version 3.6.3, <https://www.r-project.org/>) and Python software (version 3.7, [https://www.python.org/\)](https://www.python.org/). The Shapiro–Wilk test was employed to assess the distribution of continuous variables. Continuous variables with a Gaussian distribution were expressed as mean±SD and subjected to an independent-sample *t*-test. Non-normally distributed variables were represented as median with interquartile range (IQR) and analyzed using the Mann–Whitney *U* test. Categorical variables were presented as frequencies and percentages and analyzed with either the chi-square test or Fisher's exact test, as appropriate.

For optimal predictive performance, we constructed eight models: extreme gradient boosting (XGBoost), logistic regression (LR), light gradient boosting machine (LGBM), random forest (RF), adaptive boosting (AdaBoost), multi-layer perceptron (MLP), support vector machine (SVM), and Gaussian Naive Bayes (GNB). These models were built upon features selected through LASSO regression. Validation sets were used to evaluate the performance of each classifier, measuring the area under the receiver operating characteristic curve (AUROC). Sensitivity, specificity, F1 score, and overall accuracy were also assessed for each algorithm. The model with the highest AUC was deemed superior. Shapley additive explanation (SHAP) analysis further clarified the importance of each feature, enhancing the visualization and interpretability of the model.

# **Results**

#### Demographic and clinical features

We incorporated a cohort of 447 individuals into our study. Within 3 months post-surgery, 70 patients, representing 15.6% of the sample, experienced a recurrence of chronic subdural hematoma (CSDH). Intriguingly, the mean age of patients in the recurrence group was  $73.51 \pm 10.74$  years, significantly higher than that of the non-recurrence group at  $69.60 \pm 12.20$  years. Moreover, at admission, the recurrence group had a median platelet count of 194 [161, 232], which was notably lower than the 213 [170, 252] observed in the non-recurrence group. Further analysis indicated an increased susceptibility to hematoma recurrence in patients with lower fibrinogen concentrations or elevated urea nitrogen levels. Comprehensive clinical data is available in [Table 1.](#page-5-0) For robust analysis, the CSDH patient cohort was divided into training and validation subsets, adhering to a 7:3 ratio. Importantly, there were no significant differences in demographic or clinical characteristics between the training and test groups [\(Table 2](#page-6-0)).

#### Key variables

[Figures 3A,B](#page-3-1) presents the 10 features with nonzero coefficients determined via LASSO regression analysis, utilizing 10-fold crossvalidation to ascertain the optimal lambda value. The following 10 factors, significantly associated with CSDH recurrence, comprise urea nitrogen (UN), aspartate aminotransferase (AST), direct bilirubin (DB), thrombin time (TT), fibrinogen (FIB), systolic blood pressure (SBP), hematoma's widest diameter (HWD), diabetes, age, and a history of heart diseases.

#### Model performance

After identifying the 10 key variables, we utilized machine learning algorithms to predict CSDH recurrence. The predictive accuracy of these models was evaluated using essential metrics, including AUC, precision, recall, specificity, and F1 score. The results are provided in [Table 3](#page-7-0). The random forest (RF) model outperformed the others in the validation set, as evidenced by an AUC value of 0.928 from the ROC curve. [Figure 4](#page-7-1) compares the ROC curves of the eight models and displays their calibration plots. It also shows the decision curve analysis of the random forest model, indicating its net clinical benefit compared to a universal treatment strategy.

#### Relative importance of variables in RF model

SHAP analysis enabled an unbiased interpretation of the features. In the random forest (RF) model, the clinically significant variables were ranked as follows: age, aspartate aminotransferase (AST), fibrinogen (FIB), thrombin time (TT), hematoma's widest diameter (HWD), urea nitrogen (UN), direct bilirubin (DB), systolic blood pressure (SBP), and the presence of heart diseases and diabetes in medical history. The SHAP value of a feature directly corresponds to the likelihood of CSDH recurrence. High feature values are represented in red, average values in purple, and low values in blue (see [Figures 5A,B](#page-8-0)). [Figure 5C](#page-8-0) displays the forecasted recurrence probability for a high-risk CSDH patient, while [Figure 5D](#page-8-0) shows the forecast for a low-risk CSDH patient.

#### Performance evaluation of RF model using the test set

The RF model, tested on a set of 135 samples, demonstrated an impressive accuracy of 92.6%, precision of 84.2%, recall of 69.6%, an F1 score of 76.2%, and an AUC of 83.4%. [Figure 6](#page-9-0) depicts a bar chart providing a comprehensive visual overview of the model's predictive capabilities, including the ROC curve, confusion matrix, precisionrecall curve, and a classification report.

## Web-based calculator

Utilizing the RF model, we designed an online calculator available at<sup>1</sup> ([Figure 7\)](#page-10-0).

# **Discussion**

Postoperative recurrence of chronic subdural hematoma (CSDH) is a common challenge in neurology, particularly after interventions like burr-hole irrigation. The high recurrence rate complicates clinical management due to associated neurological impairments and potential increased mortality risks ([22](#page-12-8), [23](#page-12-9)). However, there are few precise models for predicting post-surgical CSDH recurrence. To address this gap, we used machine learning to predict CSDH recurrence postoperatively, demonstrating that the random forest (RF) algorithm was particularly accurate.

In this study, we utilized data from 312 patients to construct our model, while the remaining 135 patients' data were employed to assess the model's predictive performance. Despite the relatively limited sample size, we ensured the accuracy and reliability of our machine learning (ML) model through careful selection of the model, optimization of evaluation methods, and stringent overfitting control measures. Specifically, we chose the random forest (RF) model for its efficiency in handling the complexity of the dataset and its ability to minimize the risk of overfitting. The RF model is renowned for its robustness and its capacity to provide accurate predictions even in smaller datasets, thanks to its ensemble learning approach, which enhances prediction accuracy by combining multiple decision trees to effectively control overfitting. By implementing rigorous crossvalidation techniques and appropriate overfitting prevention measures, we further ensured the optimization of model performance. These meticulously designed steps enabled the RF model to achieve an accuracy rate of 92.6% and an area under the curve (AUC) value of 0.834 in the test dataset, thereby highlighting the model's high reliability and its capability to precisely predict recurrence risk.

Given the prediction of chronic subdural hematoma (CSDH) recurrence involves interpreting patients' demographic, laboratory, and radiological characteristics, the complexity of these datasets necessitates the use of a variety of algorithms to capture different information and relationships within the data. Each model has its unique method of processing data and learning patterns, thereby, a

<span id="page-4-0"></span><sup>1</sup> [https://www.xsmartanalysis.com/model/list/predict/model/html?mid=66](https://www.xsmartanalysis.com/model/list/predict/model/html?mid=6694&symbol=91xd69jB18oL581991bw) [94&symbol=91xd69jB18oL581991bw](https://www.xsmartanalysis.com/model/list/predict/model/html?mid=6694&symbol=91xd69jB18oL581991bw)

<span id="page-5-0"></span>TABLE 1 Comparison variables between non-recurrence group and recurrence group.



SBP, systolic blood pressure; DBP, diastolic blood pressure; HWD, hematoma widest diameter; WBC, white blood cell; NEUT, neutrophil; RBC, red blood cell; HB, hemoglobin; PLT, platelets; PT, prothrombin time; INR, international normalized ratio; APTT, activated partial thromboplastin time; TT, thrombin time; GLU, blood glucose; UN, urea nitrogen; CR, creatinine; ALT, alanine aminotransferase; AST, aspartate aminotransferase; TB, total bilirubin; DB, direct bilirubin; IB, indirect bilirubin.

multi-model strategy enhances the opportunity to capture all potential correlations within the data. This study adopted a comprehensive strategy to construct eight commonly used machine learning predictive models, aimed at assessing and utilizing the complex dataset from multiple perspectives to identify the most suitable model for predicting postoperative recurrence of CSDH. Through this approach, our research not only demonstrates the potential of machine learning in medical prediction but also provides a practical tool to help improve treatment outcomes for CSDH patients.

To enhance our machine learning model's accuracy in identifying patients at high risk of CSDH recurrence, we utilized the synthetic minority over-sampling technique (SMOTE) to counter imbalanced data. In many clinical datasets, an unequal distribution of instances between classes can impair machine learning model performance.

#### <span id="page-6-0"></span>TABLE 2 Clinical characteristics of patients.



SBP, systolic blood pressure; DBP, diastolic blood pressure; HWD, hematoma widest diameter; WBC, white blood cell; RBC, red blood cell; HB, hemoglobin; PLT, platelets; PT, prothrombin time; INR, international normalized ratio; APTT, activated partial thromboplastin time; TT, thrombin time; GLU, blood glucose; UN, urea nitrogen; CR, creatinine; ALT, alanine aminotransferase; AST, aspartate aminotransferase.

Often, the model may disproportionately favor the majority class, potentially resulting in suboptimal performance for the minority class ([24](#page-12-10)). The border-line SMOTE method tackles this by generating synthetic samples for the underrepresented class, balancing the dataset ([25\)](#page-12-11). Our study suggests that machine learning algorithms, specifically the RF algorithm, can accurately predict and assist healthcare professionals in identifying patients at a high risk of CSDH recurrence.

For feature selection, we employed LASSO regression, suitable for datasets with a high feature-to-observation ratio. LASSO penalizes the absolute values of coefficients, causing some of the less relevant feature coefficients to shrink to zero and effectively removing them from the model ([26](#page-12-12)). This allows the model to focus on genuinely significant predictive factors, improving both the comprehensibility and accuracy of predictions. Given its exceptional performance in our study, <span id="page-7-0"></span>TABLE 3 Performance of models in the validation set.



AUC, area under the receiver operating characteristic; XGBoost, extreme gradient boosting; LR, logistic regression; LightGBM, light gradient boosting machine; RF, random forest; AdaBoost, adaptive boosting; GNB, Gaussian NB; MLP, multilayer perceptron; SVM, support vector machine.

<span id="page-7-1"></span>

we selected the random forest model. During training, random forest builds multiple decision trees and generates classification patterns. The diverse decision trees, derived from various subsets of data and

features, enhance the RF model's ability to prevent overfitting and capture complex feature interactions, offering a more detailed insight into the factors contributing to CSDH recurrence.

<span id="page-8-0"></span>

We incorporated SHAP analysis into our predictive model, enhancing its interpretability. Transparency is vital in the medical field for understanding the reasoning behind predictions, gaining clinicians' trust, and encouraging practical application [\(27\)](#page-12-13). SHAP values provide this transparency by illuminating the individual contributions of each predictive factor. SHAP values elucidate the extent to which each feature influences the model's prediction outcomes. Positive SHAP values indicate that the presence or increase of a feature tends to elevate the value of the model's target variable (e.g., the likelihood of CSDH recurrence), signifying it as a risk factor. Conversely, negative SHAP values suggest that the presence or increase of a feature tends to decrease the model's target variable value, marking it as a protective factor. In [Figure 5A,](#page-8-0) we observed that older age, prolonged prothrombin time, thicker maximum hematoma width, elevated urea nitrogen (UN) levels, and higher direct bilirubin were identified as risk factors. On the other hand, heart disease, diabetes, elevated levels of aspartate aminotransferase, higher fibrinogen, and increased systolic blood pressure (SBP) were associated with protective factors. However, it is important to note that in the univariate analysis between the recurrence and non-recurrence groups, only differences in age, fibrinogen, and urea nitrogen were statistically significant ([Table 1](#page-5-0)). Therefore, for assessing the risk of hematoma recurrence in patients with chronic subdural hematoma, we recommend inputting all factors into our model for evaluation, rather than relying on any single factor to determine the risk of recurrence.

In our RF model's SHAP analysis, age emerged as the most significant predictive factor for CSDH recurrence, consistent with prior research indicating higher recurrence rates in older patients [\(28\)](#page-12-14).

Several factors may contribute to this pattern. First, brain atrophy could expand the subdural space, facilitating hematoma formation. Second, older patients are more likely to use anticoagulant medications, potentially increasing bleeding tendencies [\(23\)](#page-12-9). Finally, an increased susceptibility to injury may result in more frequent subdural hematomas. Apart from age, our model highlighted other factors associated with CSDH recurrence, emphasizing the condition's multifactorial nature [\(29\)](#page-12-15). AST levels may indicate liver or muscle damage, suggesting systemic inflammation or coagulation disorders that increase hematoma formation and recurrence risk ([30](#page-12-16)). Fibrinogen, a key clotting factor, may be linked to clotting disorders or hypercoagulability affecting CSDH recurrence [\(31\)](#page-12-17). Abnormal TT could signal coagulation disorders, predisposing individuals to recurrent bleeding [\(32\)](#page-12-18). HWD directly reflects hematoma size and influences recurrence. Larger hematomas might require more extensive surgical intervention, creating a larger post-drainage hematoma cavity and raising recurrence risk ([33](#page-12-19)). Elevated UN levels may indicate kidney dysfunction or dehydration, potentially affecting blood viscosity, coagulation, and CSDH recurrence [\(34](#page-12-20)). High DB levels might correlate with liver disease, affecting clotting factors and increasing recurrence risk ([34](#page-12-20)). Hypertension is known to increase the risk of rebleeding [\(35\)](#page-12-21). However, in our SHAP analysis, higher systolic blood pressure (SBP) exhibited protective tendencies, albeit without statistical significance. We hypothesize that in patients with chronic subdural hematoma (CSDH), elevated blood pressure might help maintain intracranial pressure, preventing the reaccumulation of blood. Alternatively, the CSDH patient population with higher SBP might exhibit differences in characteristics not thoroughly examined,

<span id="page-9-0"></span>

which could be related to the risk of recurrence. Nonetheless, it is important to underscore that these interpretations require validation through further research. The role of blood pressure management in CSDH is complex; high SBP may interact with various factors, influencing the risk of CSDH recurrence. Furthermore, hypertension itself is a significant risk factor for several cardiovascular diseases, and maintaining high SBP is not recommended in the long term. Therefore, caution should be exercised in interpreting this association, to avoid viewing high SBP as a protective factor while neglecting proper management of hypertension.

In our study, patients with heart disease and diabetes exhibited a tendency towards lower recurrence rates of chronic subdural hematoma (CSDH), although the difference was not statistically significant. This observation contradicts our prior understanding [\(23,](#page-12-9) [36](#page-12-22)). We speculate that one possible explanation for this phenomenon could be the long-term medication regimen these patients often require, which includes antihypertensives, glycemic control medications, and anticoagulant or antiplatelet drugs. It is conceivable that the initial CSDH in these patients may be associated with their medication use. Post-surgical treatment might then receive more personalized and cautious management, potentially reducing the likelihood of CSDH recurrence. Alternatively, the observed phenomenon in the study could be partially attributed to selection bias, suggesting that patients with heart disease and diabetes might receive more aggressive management and treatment for reasons not fully identified, which could indirectly lower their risk of CSDH recurrence.

Prior research has indicated that patients who take oral anticoagulants face a heightened risk of recurrent subdural hematoma (SDH) bleeding and exhibit larger hematoma volumes. Additionally, the use of these medications is associated with the chronic progression of SDH and an increased mortality rate [\(37\)](#page-12-23). On the other hand, a study found that antiplatelet drug use is related to a higher risk of postoperative rebleeding, whereas anticoagulant use does not show a similar correlation [\(38\)](#page-12-24). There is also evidence suggesting that early surgical intervention and the immediate resumption of antiplatelet medication after surgery can decrease the risk of thromboembolic complications without increasing the risk of chronic subdural hematoma recurrence. Moreover, extending the period of medication cessation before surgery does not offer significant benefits for patients already on antiplatelet therapy. However, patients who concurrently use antiplatelet and anticoagulant therapies may be at a higher risk for chronic subdural hematoma recurrence ([39](#page-12-25)). Conversely, studies have shown that ICU patients who underwent antiplatelet therapy before surgery might achieve better outcomes, possibly because platelet therapy is linked to better initial conditions [\(40\)](#page-12-26).

<span id="page-10-0"></span>

These systemic or localized factors underscore the complexity of CSDH recurrence. Interactions between diverse physiological and pathological conditions necessitate a comprehensive patient management approach. A deeper understanding of these contributing factors and their interactions is vital for effective prediction and prevention of CSDH recurrence.

However, certain limitations must be acknowledged. The retrospective design of our study and the reliance on data from a single medical institution might introduce bias and constrain generalizability. These factors could hinder the applicability of our results to wider populations or distinct healthcare settings, where differences in clinical practices, patient demographics, and institutional policies could influence CSDH recurrence patterns. Future research should prioritize multicenter prospective studies to validate our findings.

# Conclusion

Our investigation underscores the capacity of machine learning, especially the random forest (RF) model enhanced with the border-line synthetic minority over-sampling technique (SMOTE), for predicting postoperative chronic subdural hematoma (CSDH) recurrence with remarkable accuracy. Age is the most potent predictor of recurrence, consistent with previous studies. Other significant contributors include elevated levels of AST, abnormal TT, larger HWD, UN, DB, SBP, and history of heart diseases and diabetes. The combined application of LASSO

regression, SMOTE, and SHAP analysis enhances the RF model's precision and interpretability, offering valuable insights for optimizing therapeutic strategies and implementing preventive interventions for high-risk patients. However, our study has limitations, including its retrospective design and reliance on data from a single medical institution. Multicenter prospective studies with diverse populations and different medical contexts are needed to validate our findings.

#### Transparency, rigor, and reproducibility summary

Our study utilized data from 447 CSDH patients treated with consecutive burr-hole irrigations at the First Affiliated Hospital of Wenzhou Medical University between December 2014 and April 2019. The Ethics Committee of this institution granted the research ethical approval. Due to the study's retrospective design, informed consent from participants was considered unnecessary. Upon a valid inquiry, the corresponding author will make available the data used or examined in this research. Of the patients, 312 were designated as the development cohort and 135 as the test cohort. The models incorporated demographic, laboratory, and radiological parameters. The border-line synthetic minority over-sampling technique (SMOTE) addressed data imbalance, while the LASSO regression method identified salient features associated with CSDH recurrence. We employed eight machine learning algorithms to predict hematoma recurrence. Model construction was executed via the XSmartAnalysis website [\(https://](https://www.xsmartanalysis.com/)

[www.xsmartanalysis.com/](https://www.xsmartanalysis.com/)) using R statistical 3.6.3 and Python 3.7. Evaluation metrics included AUROC, sensitivity, specificity, F1 score, calibration plots, and decision curve analysis (DCA). The RF model displayed exceptional accuracy. Shapley additive explanation (SHAP) analysis enhanced model visualization and interpretability, verifying results and highlighting critical clinical predictors. The rigorous methodology, complemented by diverse machine learning techniques, supports the replicability of our models in comparable clinical contexts. The RF model, with its exemplary performance, stands as a robust tool for predicting postoperative CSDH recurrence, offering valuable insights for therapeutic decision-making and preventive strategies for high-risk patients.

## Data availability statement

The data analyzed in this study is subject to the following licenses/ restrictions: the corresponding author can provide the data utilized and/or examined in the present investigation upon a reasonable inquiry. Requests to access these datasets should be directed to [nizhihui@wmu.edu.cn](mailto:nizhihui@wmu.edu.cn).

## Ethics statement

The studies involving humans were approved by Ethics Committee in Clinical Research (ECCR) of the First Affiliated Hospital of Wenzhou Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

# Author contributions

ZN: Conceptualization, Visualization, Writing – original draft. YeZ: Data curation, Formal Analysis, Writing – review & editing. YQ: Software, Validation, Writing – review & editing. XL: Data curation,

## References

<span id="page-11-0"></span>1. Bartley A, Bartek J Jr, Jakola AS, Sundblom J, Fält M, Förander P, et al. Effect of irrigation fluid temperature on recurrence in the evacuation of chronic subdural hematoma: a randomized clinical trial. *JAMA Neurol*. (2022) 80:58–63. doi: [10.1001/](https://doi.org/10.1001/jamaneurol.2022.4133) [jamaneurol.2022.4133](https://doi.org/10.1001/jamaneurol.2022.4133)

<span id="page-11-1"></span>2. Miranda LB, Braxton E, Hobbs J, Quigley MR. Chronic subdural hematoma in the elderly: not a benign disease. *J Neurosurg*. (2011) 114:72–6. doi: [10.3171/](https://doi.org/10.3171/2010.8.Jns10298) [2010.8.Jns10298](https://doi.org/10.3171/2010.8.Jns10298)

<span id="page-11-2"></span>3. Duerinck J, Van Der Veken J, Schuind S, Van Calenbergh FV, van Loon J, Du Four S, et al. Randomized trial comparing burr hole craniostomy, minicraniotomy, and twist drill craniostomy for treatment of chronic subdural hematoma. *Neurosurgery*. (2022) 91:304–11. doi: [10.1227/neu.0000000000001997](https://doi.org/10.1227/neu.0000000000001997)

<span id="page-11-3"></span>4. Gaist D, García Rodríguez LA, Hellfritzsch M, Poulsen FR, Halle B, Hallas J, et al. Association of antithrombotic drug use with subdural hematoma risk. *JAMA*. (2017) 317:836–46. doi: [10.1001/jama.2017.0639](https://doi.org/10.1001/jama.2017.0639)

<span id="page-11-4"></span>5. Edlmann E, Giorgi-Coll S, Whitfield PC, Carpenter KLH, Hutchinson PJ. Pathophysiology of chronic subdural haematoma: inflammation, angiogenesis and implications for pharmacotherapy. *J Neuroinflammation*. (2017) 14:108. doi: [10.1186/](https://doi.org/10.1186/s12974-017-0881-y) [s12974-017-0881-y](https://doi.org/10.1186/s12974-017-0881-y)

6. Edlmann E, Whitfield P, Kolias A, Hutchinson PJ. Pathogenesis of chronic subdural hematoma: a cohort evidencing *de novo* and transformational origins. *J Neurotrauma*. (2021) 38:2580–9. doi: [10.1089/neu.2020.7574](https://doi.org/10.1089/neu.2020.7574)

Writing – review & editing. ZX: Data curation, Writing – review & editing. YiZ: Data curation, Writing – review & editing. YC: Data curation, Writing – review & editing. LH: Writing – review & editing. JY: Supervision, Writing – review & editing. QZ: Supervision, Writing – review & editing.

# Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the National Natural Science Foundation of China (Nos. 82271345, 81820108011, and 82103216).

# Acknowledgments

The authors would like to express our heartfelt appreciation to the team at Home for Researchers [\(www.Home-for-researchers.com\)](http://www.Home-for-researchers.com) for their invaluable assistance with professional English language editing services.

# Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

# Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

<span id="page-11-7"></span>9. Mishra R, Deora H, Florez-Perdomo W, Moscote-Salazar LR, Garcia-Ballestas E, Rahman MM, et al. Clinical and radiological characteristics for recurrence of chronic subdural hematoma: a systematic review and meta-analysis. *Neurol Int*. (2022) 14:683–95. doi: [10.3390/neurolint14030057](https://doi.org/10.3390/neurolint14030057)

10. Ohba S, Kinoshita Y, Nakagawa T, Murakami H. The risk factors for recurrence of chronic subdural hematoma. *Neurosurg Rev*. (2013) 36:145–50. doi: [10.1007/](https://doi.org/10.1007/s10143-012-0396-z) [s10143-012-0396-z](https://doi.org/10.1007/s10143-012-0396-z)

<span id="page-11-8"></span>11. Andersen-Ranberg NC, Debrabant B, Poulsen FR, Bergholt B, Hundsholt T, Fugleholm K. The Danish chronic subdural hematoma study-predicting recurrence of chronic subdural hematoma. *Acta Neurochir*. (2019) 161:885–94. doi: [10.1007/](https://doi.org/10.1007/s00701-019-03858-9) [s00701-019-03858-9](https://doi.org/10.1007/s00701-019-03858-9)

<span id="page-11-9"></span>12. Dubinski D, Won SY, Behmanesh B, Cantré D, Mattes I, Trnovec S, et al. Significance of temporal muscle thickness in chronic subdural hematoma. *J Clin Med*. (2022) 11:6456. doi: [10.3390/jcm11216456](https://doi.org/10.3390/jcm11216456)

<span id="page-11-5"></span><sup>7.</sup> Nouri A, Gondar R, Schaller K, Meling T. Chronic subdural hematoma (cSDH): a review of the current state of the art. *Brain Spine*. (2021) 1:100300. doi: [10.1016/j.](https://doi.org/10.1016/j.bas.2021.100300) [bas.2021.100300](https://doi.org/10.1016/j.bas.2021.100300)

<span id="page-11-6"></span><sup>8.</sup> Shotar E, Meyblum L, Premat K, Lenck S, Degos V, Grand T, et al. Middle meningeal artery embolization reduces the post-operative recurrence rate of at-risk chronic subdural hematoma. *J Neurointerv Surg*. (2020) 12:1209–13. doi: [10.1136/](https://doi.org/10.1136/neurintsurg-2020-016048) [neurintsurg-2020-016048](https://doi.org/10.1136/neurintsurg-2020-016048)

<span id="page-12-0"></span>13. Holl DC, Mikolic A, Blaauw J, Lodewijkx R, Foppen M, Jellema K, et al. External validation of prognostic models predicting outcome after chronic subdural hematoma. *Acta Neurochir*. (2022) 164:2719–30. doi: [10.1007/s00701-022-05216-8](https://doi.org/10.1007/s00701-022-05216-8)

<span id="page-12-1"></span>14. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologistlevel classification of skin cancer with deep neural networks. *Nature*. (2017) 542:115–8. doi: [10.1038/nature21056](https://doi.org/10.1038/nature21056)

<span id="page-12-2"></span>15. Deo RC. Machine learning in medicine. *Circulation*. (2015) 132:1920–30. doi: [10.1161/circulationaha.115.001593](https://doi.org/10.1161/circulationaha.115.001593)

<span id="page-12-3"></span>16. Hale AT, Stonko DP, Wang L, Strother MK, Chambless LB. Machine learning analyses can differentiate meningioma grade by features on magnetic resonance imaging. *Neurosurg Focus*. (2018) 45:E4. doi: [10.3171/2018.8.Focus18191](https://doi.org/10.3171/2018.8.Focus18191)

17. Karabacak M, Jagtiani P, Carrasquilla A, Germano IM, Margetis K. Prognosis individualized: survival predictions for WHO grade II and III gliomas with a machine learning-based web application. *npj Digit Med*. (2023) 6:200. doi: [10.1038/](https://doi.org/10.1038/s41746-023-00948-y) [s41746-023-00948-y](https://doi.org/10.1038/s41746-023-00948-y)

<span id="page-12-4"></span>18. Brouwers HB, Chang Y, Falcone GJ, Cai X, Ayres AM, Battey TWK, et al. Predicting hematoma expansion after primary intracerebral hemorrhage. *JAMA Neurol*. (2014) 71:158–64. doi: [10.1001/jamaneurol.2013.5433](https://doi.org/10.1001/jamaneurol.2013.5433)

<span id="page-12-5"></span>19. Brusko G, Kolcun J, Wang M. Machine-learning models: the future of predictive analytics in neurosurgery. *Neurosurgery*. (2018) 83:E3–4. doi: [10.1093/neuros/nyy166](https://doi.org/10.1093/neuros/nyy166)

<span id="page-12-6"></span>20. Han H, Wang W-Y, Mao B-H. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning In: DS Huang, XP Zhang and GB Huang, editors. *Advances in intelligent computing. ICIC 2005. Lecture notes in computer science*. Berlin: Springer (2005)

<span id="page-12-7"></span>21. Jiang R, Zhao S, Wang R, Feng H, Zhang J, Li X, et al. Safety and efficacy of atorvastatin for chronic subdural hematoma in Chinese patients: a randomized clinicaltrial. *JAMA Neurol*. (2018) 75:1338–46. doi: [10.1001/jamaneurol.2018.2030](https://doi.org/10.1001/jamaneurol.2018.2030)

<span id="page-12-8"></span>22. Guilfoyle MR, Hutchinson PJ, Santarius T. Improved long-term survival with subdural drains following evacuation of chronic subdural haematoma. *Acta Neurochir*. (2017) 159:903–5. doi: [10.1007/s00701-017-3095-2](https://doi.org/10.1007/s00701-017-3095-2)

<span id="page-12-9"></span>23. Torihashi K, Sadamasa N, Yoshida K, Narumi O, Chin M, Yamagata S. Independent predictors for recurrence of chronic subdural hematoma: a review of 343 consecutive surgical cases. *Neurosurgery*. (2008) 63:1125–9. doi: [10.1227/01.Neu.0000335782.60059.17](https://doi.org/10.1227/01.Neu.0000335782.60059.17)

<span id="page-12-10"></span>24. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res*. (2002) 16:321. doi: [10.1613/jair.953](https://doi.org/10.1613/jair.953)

<span id="page-12-11"></span>25. Riafio D. Using Gabriel graphs in Borderline-SMOTE to deal with severe two-class imbalance problems on neural networks. In Artificial Intelligence Research and Development: Proceedings of the 15th International Conference of the Catalan Association for Artificial Intelligence 248:29. IOS Press (2012).

<span id="page-12-12"></span>26. Ranstam J, Cook J. LASSO regression. *J Br Surg*. (2018) 105:1348–8. doi: [10.1002/](https://doi.org/10.1002/bjs.10895) [bjs.10895](https://doi.org/10.1002/bjs.10895)

<span id="page-12-13"></span>27. Yang Y, Yuan Y, Han Z, Liu G. Interpretability analysis for thermal sensation machine learning models: an exploration based on the SHAP approach. *Indoor Air*. (2022) 32:e12984. doi: [10.1111/ina.12984](https://doi.org/10.1111/ina.12984)

<span id="page-12-14"></span>28. Miah IP, Tank Y, Rosendaal FR, Peul WC, Dammers R, Lingsma HF, et al. Radiological prognostic factors of chronic subdural hematoma recurrence: a systematic review and meta-analysis. *Neuroradiology*. (2021) 63:27. doi: [10.1007/](https://doi.org/10.1007/s00234-020-02558-x) [s00234-020-02558-x](https://doi.org/10.1007/s00234-020-02558-x)

<span id="page-12-15"></span>29. Taussky P, Fandino J, Landolt H. Number of burr holes as independent predictor of postoperative recurrence in chronic subdural haematoma. *Br J Neurosurg*. (2008) 22:279–82. doi: [10.1080/02688690701818885](https://doi.org/10.1080/02688690701818885)

<span id="page-12-16"></span>30. Ko B-S, Lee J-K, Seo B-R, Moon SJ, Kim JH, Kim SH. Clinical analysis of risk factors related to recurrent chronic subdural hematoma. *J Korean Neurosurg Soc*. (2008) 43:11–5. doi: [10.3340/jkns.2008.43.1.11](https://doi.org/10.3340/jkns.2008.43.1.11)

<span id="page-12-17"></span>31. Park S-H, Kang D-H, Park J, Hwang JH, Hwang SK, Sung JK, et al. Fibrinogen and D-dimer analysis of chronic subdural hematomas and computed tomography findings: a prospective study. *Clin Neurol Neurosurg*. (2011) 113:272–6. doi: [10.1016/j.](https://doi.org/10.1016/j.clineuro.2010.11.014) [clineuro.2010.11.014](https://doi.org/10.1016/j.clineuro.2010.11.014)

<span id="page-12-18"></span>32. Yu X, Wu L, Ou Y, Xu L, Guo X, Yang C, et al. Clinical characteristics and outcomes of chronic subdural hematoma in patients with a history of antiplatelet therapy. *Clin Neurol Neurosurg*. (2021) 208:106817. doi: [10.1016/j.clineuro.2021.106817](https://doi.org/10.1016/j.clineuro.2021.106817)

<span id="page-12-19"></span>33. Amirjamshidi A, Abouzari M, Eftekhar B, Rashidi A, Rezaii J, Esfandiari K, et al. Outcomes and recurrence rates in chronic subdural haematoma. *Br J Neurosurg*. (2007) 21:272–5. doi: [10.1080/02688690701272232](https://doi.org/10.1080/02688690701272232)

<span id="page-12-20"></span>34. Wang N, Hu J, Oppong-Gyebi A, Zhu X, Li Y, Yang J, et al. Elevated blood urea nitrogen is associated with recurrence of post-operative chronic subdural hematoma. *BMC Neurol*. (2020) 20:1–10. doi: [10.1186/s12883-020-01985-w](https://doi.org/10.1186/s12883-020-01985-w)

<span id="page-12-21"></span>35. Weigel R, Hohenstein A, Schlickum L, Weiss C, Schilling L. Angiotensin converting enzyme inhibition for arterial hypertension reduces the risk of recurrence in patients with chronic subdural hematoma possibly by an antiangiogenic mechanism. *Neurosurgery*. (2007) 61:788–93. doi: [10.1227/01.NEU.0000298907.56012.E8](https://doi.org/10.1227/01.NEU.0000298907.56012.E8)

<span id="page-12-22"></span>36. Abdullah H, Dakurah T, Akoto H, Abaidoo B, JCB D, Yawson AE, et al. Predictors of recurrence of chronic subdural haematoma in a cohort study of patients presenting in a sub-Saharan African teaching hospital. *BMC Neurol*. (2022) 22:346. doi: [10.1186/](https://doi.org/10.1186/s12883-022-02857-1) [s12883-022-02857-1](https://doi.org/10.1186/s12883-022-02857-1)

<span id="page-12-23"></span>37. Pastor I, Para I, Vesa Ş, Florian IȘ. The impact of oral anticoagulants on the characteristics of subdural hematomas and other brain lesions in patients with traumatic brain injury. *Med Pharm Rep*. (2023) 96:269–73. doi: [10.15386/mpr-2535](https://doi.org/10.15386/mpr-2535)

<span id="page-12-24"></span>38. Poon M, Rea C, Kolias A, Brennan PM, Afshari FT, Ahmed AI, et al. Influence of antiplatelet and anticoagulant drug use on outcomes after chronic subdural hematoma drainage. *J Neurotrauma*. (2021) 38:1177–84. doi: [10.1089/neu.2018.6080](https://doi.org/10.1089/neu.2018.6080)

<span id="page-12-25"></span>39. Szczygielski J, Utter K, Oertel J. Response to Poon et al. (doi: 10.1089/ neu.2018.6080): acetylsalicylic acid and chronic subdural hematoma:is it really a bad couple? Influence of antiplatelet and anticoagulant drug use on outcomes after chronic subdural hematoma drainage. *J Neurotrauma*. (2020) 37:428–9. doi: [10.1089/](https://doi.org/10.1089/neu.2019.6528) [neu.2019.6528](https://doi.org/10.1089/neu.2019.6528)

<span id="page-12-26"></span>40. Szczygielski J, Gund S, Schwerdtfeger K, Steudel WI, Oertel J. Factors affecting outcome in treatment of chronic subdural hematoma in ICU patients: impact of anticoagulation. *World Neurosurg*. (2016) 92:426–33. doi: [10.1016/j.wneu.2016.05.049](https://doi.org/10.1016/j.wneu.2016.05.049)