



# Post-stroke Anxiety Analysis via Machine Learning Methods

Jirui Wang<sup>1</sup>, Defeng Zhao<sup>2</sup>, Meiqing Lin<sup>1</sup>, Xinyu Huang<sup>3</sup> and Xiuli Shang<sup>1\*</sup>

<sup>1</sup> Department of Neurology, The First Affiliated Hospital, China Medical University, Shenyang, China, <sup>2</sup> The First Clinical Department, China Medical University, Shenyang, China, <sup>3</sup> Software College, Northeastern University, Shenyang, China

Post-stroke anxiety (PSA) has caused wide public concern in recent years, and the study on risk factors analysis and prediction is still an open issue. With the deepening of the research, machine learning has been widely applied to various scenarios and make great achievements increasingly, which brings new approaches to this field. In this paper, 395 patients with acute ischemic stroke are collected and evaluated by anxiety scales (i.e., HADS-A, HAMA, and SAS), hence the patients are divided into anxiety group and non-anxiety group. Afterward, the results of demographic data and general laboratory examination between the two groups are compared to identify the risk factors with statistical differences accordingly. Then the factors with statistical differences are incorporated into a multivariate logistic regression to obtain risk factors and protective factors of PSA. Statistical analysis shows great differences in gender, age, serious stroke, hypertension, diabetes mellitus, drinking, and HDL-C level between PSA group and non-anxiety group with HADS-A and HAMA evaluation. Meanwhile, as evaluated by SAS scale, gender, serious stroke, hypertension, diabetes mellitus, drinking, and HDL-C level differ in the PSA group and the non-anxiety group. Multivariate logistic regression analysis of HADS-A, HAMA, and SAS scales suggest that hypertension, diabetes mellitus, drinking, high NIHSS score, and low serum HDL-C level are related to PSA. In other words, gender, age, disability, hypertension, diabetes mellitus, HDL-C, and drinking are closely related to anxiety during the acute stage of ischemic stroke. Hypertension, diabetes mellitus, drinking, and disability increased the risk of PSA, and higher serum HDL-C level decreased the risk of PSA. Several machine learning methods are employed to predict PSA according to HADS-A, HAMA, and SAS scores, respectively. The experimental results indicate that random forest outperforms the competitive methods in PSA prediction, which contributes to early intervention for clinical treatment.

## OPEN ACCESS

### Edited by:

Guang H. Yue,  
Kessler Foundation, United States

### Reviewed by:

Julien Rossignol,  
Central Michigan University,  
United States  
Aurel Popa,  
Essen University Hospital, Germany

### \*Correspondence:

Xiuli Shang  
xlshangcmu@163.com

Received: 24 January 2021

Accepted: 14 May 2021

Published: 25 June 2021

### Citation:

Wang J, Zhao D, Lin M, Huang X and Shang X (2021) Post-stroke Anxiety Analysis via Machine Learning Methods. *Front. Aging Neurosci.* 13:657937. doi: 10.3389/fnagi.2021.657937

**Keywords:** post-stroke anxiety, acute ischemic stroke, machine learning, random forest, risk factors analysis

## 1. INTRODUCTION

Stroke is a medical condition in which poor blood flow to the brain results in cell death, associated with high morbidity, high disability, and high mortality across the world (Wolfe, 2000). Notably, approximately 2.5 million new stroke cases annually occur in China and the mortality rate has reached 11.48% (Sun et al., 2013; Chen et al., 2017). Mood problems such as depression, apathy, and distress are commonly reported with post-stroke (Hackett et al., 2014), but anxiety in stroke

patients has been relatively neglected both in clinical and research settings, in spite of its ubiquity in the general population (Remes et al., 2016). Post-stroke anxiety (PSA) refers that stroke patients extremely concern about the prognosis status, e.g., recurrence, re-working abilities, the occurrence of fall accidents, and so on (Gilworth et al., 2009). Once stroke onset, anxiety becomes common throughout the acute phase, after months, and even after years (Lincoln et al., 2013). A systematic review and meta-analysis shows that the prevalence of anxiety disorders is 29.3% post-stroke during the first year, with 36.7% in 2 weeks, 24.1% in 2 weeks to 3 months, and 23.8% in 3–12 months (Rafsten et al., 2018). Specifically, Knapp et al. (2020) collect and analyze 53 studies and report 25.5% of stroke patients developed PSA within 1 month of stroke, 23.6% in 1 and 5 months, and 21.5% in 6 months to 1 year. A plethora of studies indicate that PSA significantly influences the living quality (Lincoln et al., 2013), which is associated with the delaying recovery of neurological function (Chun et al., 2018), and the interventions on anxiety disorders have a positive impact on the incidence of both coronary artery disease and stroke (Pérez-Piñar et al., 2017).

Given the significant impact of PSA on patient outcomes, great emphasis has been placed on risk reduction and early detection. However, the pathophysiology of PSA is still unknown and the relevant risk factors are controversial. A systematic review on 18 observational studies with 8,130 patients suggests that pre-stroke depression, stroke severity, early anxiety, and dementia (or cognitive) impairment following stroke are the main predictors of PSA, while the lack of methodological and statistical rigorously affects the validity of predictive models, which indicates future research should focus on testing predictive models on both internal and external samples to ultimately inform future clinical practice (Menlove et al., 2015). Accurate individual patient risk prediction would allow for evaluation and intervention even earlier in the pathologic process. Notably, it is critical to identify risk factors associated with PSA and build models to predict PSA.

With the rapid development of advanced technology, artificial intelligence has been applied extensively in a variety of professions. As an important tool in artificial intelligence field, machine learning (Alpaydin, 2020) has received increasing attention in the last decades, which is widely utilized in medical image processing, autonomous driving, computer vision, and so on. Classic machine learning models such as linear models, decision trees (Kamiński et al., 2018), Bayesian classifiers (Kohavi, 1996), Support Vector Machines (SVM) (Cortes and Vapnik, 1995), neural networks (Müller et al., 2012), Stochastic Gradient Descent (denoted by SGD Classifier) (Zhang, 2004), Multilayer Perceptron (denoted by MLP) (Rumelhart et al., 1986), and random forests (Breiman, 2001) have exhibited certain specific usage, i.e., there are no methods suitable for solving problems at any real-life scenarios. Stochastic gradient descent is an iterative method for optimizing an objective function with suitable smoothness properties, which can be regarded as a stochastic approximation of gradient descent optimization (Saad, 1998). A multilayer perceptron is a class of feedforward artificial neural networks, which consists of at least three layers of nodes: an input layer, a hidden layer, and

an output layer. MLP utilizes a supervised learning technique called backpropagation for training, which can distinguish data that is not linearly separable (Hastie et al., 2009). An SVM maps training examples to points in space so as to maximize the width of the gap between the two categories. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall (Joachims, 1998). Random forest (RF), proposed by Breiman (2001), consists of a set of decision trees, each of which is a decision support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility (Kamiński et al., 2018). Random Forest can be used in the prediction of incident delirium (Corradi et al., 2018), malignancy of pulmonary nodules (Mei et al., 2018), survival from large echocardiography, electronic health record datasets (Samad et al., 2019), and so on. The basic thought is to determine the input sample by random sampling, and the sample data obtained will be handed over to each decision tree for judgment, thereby all the results will be voted, and the result with the most votes will be used as the output. Hence, the random forest also has ensemble learning, which can improve the accuracy of the predictive model.

Inspired by such new methods, this study plans to develop proper PSA prediction models using machine learning methods. To the best of our knowledge, this is the first study to apply machine learning to predicting anxiety for post-stroke patients. This work can identify anxiety patients after stroke at an early stage, thus benefits guiding appropriate prevention and treatments to avoid leading to severe outcomes.

The main contributions of this paper are listed as follows:

- (1) The main factors of PSA are analyzed in detail by traditional statistical methods between patients with/without PSA, and then all the factors with statistical difference are put into a multivariable logistic regression analysis to study in-depth.
- (2) Different anxiety test scales (i.e., HADS-A, HAMA, and SAS) are taken into consideration to evaluate the degree of PSA.
- (3) Classic machine learning methods such as decision tree and random forest are employed as predictive models to estimate PSA, and random forest outperforms the competitive approaches.

The rest of this paper is organized as follows. Section 2 introduces the material and methods for clinical data collection. Section 3 gives data analysis and experimental environment for machine learning methods. Section 4 exhibits experimental results of statistical analysis and PSA prediction comparison *via* different machine learning methods. Section 5 exhibits the discussion on the obtained results. Section 6 summarizes the whole paper and provides concluding remarks.

## 2. MATERIALS AND METHODS

### 2.1. Patient Eligibility

The research protocol was accepted by the Regional Medical Scientific Research Ethics Committee of the First Affiliated Hospital of China Medical University (IRB no. 2020368). Written

informed consent was obtained from all patients after a complete description of all procedures of the study provided.

From August 2017 to September 2020, 516 patients with ischemic stroke who were consecutively admitted to the stroke unit of the Department of Neurology at The First Hospital of China Medical University in China were recruited. The inclusion criteria are as follows: (1) First-ever stroke with computed tomography or magnetic resonance imaging (MRI) scan upon admission and confirmed acute cerebral infarction within 7 days after stroke onset, which meets the diagnostic criteria of 2018 Chinese Guidelines for the Diagnosis and Treatment of acute ischemic stroke (Chinese Medical Association et al., 2018); (2) age 18 years or older; (3) stable temperature, pulse, respiration, and blood pressure; and (4) signed informed consent.

Exclusion criteria are as follows: (1) Stroke-like manifestation due to definite intracranial non-vascular factors (such as primary or metastatic tumors); (2) Concurrent diagnosis of terminal illness, dementia, depression, Parkinson's disease, or motor neuron disease, all of which have been shown to cause anxiety; (3) inability to complete the scale evaluation due to communication (e.g., aphasia) or cognitive disorders; (4) administered thrombolysis therapy; (5) anxiety diagnosis before stroke; (6) inability to give informed consent. The most common reasons for exclusion were cognitive impairment ( $n = 21$ ), depression ( $n = 19$ ), and inability to complete the scale ( $n = 18$ ). Thirty-two patients met other exclusion criteria and 31 patients refused to participate, leaving a total of 395 subjects (response 76.6%; 119 female, 276 male) for further analysis.

## 2.2. Collection of Clinical Data

All subjects' demographic data (gender, age, marital status, occupation, weight, and height), vascular risk factors (hypertension, coronary artery disease, diabetes mellitus, and tobacco smoking) and drinking history are collected and recorded by a trained research assistant at the time of admission. Based on the ESH/ESC hypertension guidelines recommendations (Kjeldsen et al., 2016; Cuspidi et al., 2018), hypertension is defined as systolic blood pressure (BP)  $\geq 140$  mm Hg or diastolic blood pressure higher than or equal to 90 mm Hg. Diabetes mellitus (DM) accords with the World Health Organization (WHO) diagnostic criteria for type 2 diabetes mellitus (Group et al., 1985). Smoking refers to more than 1 cigarette a day and continuous smoking for more than 3 months (Patkar et al., 2003). Drinking is distinguished by a history of more than 5 years, more than 3 times a week, and each time drinking more than 36 g of alcohol (Mazzaglia et al., 2001). Stroke severity and the level of disability were assessed using the National Institutes of Health Stroke Scale (NIHSS) (Lyden, 2017). The scores ranged from 0 (no impairment) to a maximum of 42 points. The higher the score, the more severe the neurological impairment. Scores with 4 or less are usually described as minor stroke while 21 or greater are usually described as severe stroke (Harrison et al., 2013). These measures are examined within 24 h of admission. Blood samples are obtained the morning after admission, and the serum levels of low-density lipoprotein (LDL), high-density lipoprotein (HDL), total cholesterol (TC), triglyceride (TG), glucose (GLU), uric acid

(UA), C-reactive protein (CRP), creatinine (Cr), red blood cell (RBC), hemoglobin (HB), platelet (PLT), D-dimer (D-D), fibrinogen (FIB), and homocysteine (HCY) are determined.

## 2.3. Assessment of Anxiety

The degree of PSA is estimated by the Hospital Anxiety and Depression Scale (HADS-A) scores (Zigmond and Snaith, 1983), Hamilton Anxiety Scale (HAMA) scores (Hamilton, 1959), and Self-Rating Anxiety Scale (SAS) scores (Zung, 1971). The Chinese versions are validated.

As for the above-mentioned scales, the HADS-A is the most commonly used rating scale for anxiety evaluation (Burton et al., 2013). Studies have shown HADS-A correlates significantly with the Stroke Specific Quality of Life (SSQOL), with scores for energy, mood, personality, social roles, family role, thinking, and work/productivity (Rafsten et al., 2018). The HADS is a classical self-assessment mood scale specifically designed for non-psychiatric hospital departments and is presented as a reliable and valid instrument for screening for anxiety and depression after stroke (Bjelland et al., 2002). The scale is frequently used for the assessment of depression and anxiety in stroke patients (Fure et al., 2006). It includes a total of 14 items each with a score of between 0 and 3. One-half of the items are related to anxiety (HADS-A) while the other half is specific for depression. Studies have found that HADS is performed well in the assessment of both symptom severity and diagnosis of anxiety at the recommended diagnostic cut-off of  $\geq 8$  (Zigmond and Snaith, 1983; Bjelland et al., 2002). The HADS has been previously validated in Nigeria (Abiodun, 1994) where the HADS-A was found to have a sensitivity in the range of 85.0–92.9% and a specificity of 86.5–90.0%.

In order to improve the predictive accuracy of machine learning models, HAMA and SAS are also employed to screen for PSA. HAMA is a 14-items rating scale that is developed to quantify the severity of anxiety symptoms. Each item is rated on a five-point scale, ranging from 0 (not present) to 4 (severe). Total scores on the HAMA range from 0 to 56 (Maier et al., 1988). Subjects with a HAMA score equal to or larger than 7 were considered to have anxiety symptoms. SAS is a norm-referenced scale that enjoys widespread usage as a screener for anxiety disorders since developed in 1971 (Dunstan and Scott, 2018). It contains 20 items, with the score of each item ranging from 1 to 4. The greater score indicates the higher degree of anxiety that involved the conversion of a total scale raw score (with a potential range of 20–80) to an index score with a potential range of 25–100. The index score is derived by dividing the sum of the values (raw scores) obtained on the 20 items by the maximum possible score of 80, converted to a decimal and multiplied by 100 (Zung, 1971). A raw score of 40 or an index score of 50 is the cut-off of the scale (Dunstan and Scott, 2020).

## 3. DATA ANALYSIS

### 3.1. Statistical Analysis

The SPSS 26.0 statistical package (SPSS Inc., Chicago, IL) is utilized for all statistical analysis. The comparison between patients with and without PSA of continuous variables is

analyzed by independent *t*-test or analyses of covariance. Univariate analyses of the association between categorical variables in both groups are performed *via* chi-square tests. Descriptive data are presented as mean and standard deviations (SD) or as 95% confidence intervals (95% CIs). When a correlation *P*-value is less than 0.15 for a variable, this variable is analyzed by multivariate logistic regression, and odds ratios (ORs) (with 95% CIs) are calculated for the relative risk of anxiety for each group. For all analyses, probability levels reported are two-tailed, and  $P < 0.05$  is considered as the statistically significant level.

### 3.2. Prediction With Machine Learning Methods

The experiments with machine learning methods are implemented in Python 3.8.3, with relevant library Scikit-learn 0.23.2. The operating system is 64 bit Windows 10, with configuration of Intel (R) Core (TM) i7-7700 CPU @ 3.60 GHz (8 CPUs), ~3.6 GHz and 16 GB ram installed.

Machine learning methods are efficient tools for prediction and classification problems in many real-life scenarios. The general process is performed with a previous treatment and then it can be utilized to predict new cases. As the patients have completed the anxiety tests, we record the scores of each patient [denoted by  $Y = (y_1, y_2, \dots, y_n)$ ] as a benchmark, the prediction result of machine methods  $r$  of the patients in the testing set can be represented as  $Pr = (pr_1, pr_2, \dots, pr_n)$ . The Euclidean Distance (Van Der Heijden et al., 2005) can be employed to measure the similarity between the prediction results and real test scores of the  $n$  patients in the testing set by

$$d(Pr, Y) = \sqrt{(pr_1 - y_1)^2 + (pr_2 - y_2)^2 + \dots + (pr_n - y_n)^2} \\ = \sqrt{\sum_{i=1}^n (pr_i - y_i)^2} \quad (1)$$

The Euclidean Distance is employed as a measure to compare the performance of each machine learning method in predicting PSA, where the smaller the Euclidean Distance, the better performance of a predicting method obtains.

## 4. RESULTS

### 4.1. Demographic and Clinical Characteristics Between Patients With and Without PSA

Note that 395 ischemic stroke patients (119 female and 276 male between 29 and 98 years of age) are taken into consideration in the analysis. Demographic and clinical characteristics between the two groups are summarized in **Tables 1–3**.

On the whole, in the PSA group, the mean age of patients is relatively younger and serum HDL-C level is lower. The proportion of male patients, serious stroke, hypertension, diabetes mellitus, and drinking were significantly higher in the PSA group than the non-anxiety group.

### 4.2. Multivariate Logistic Regression Analyses of the Risk Factors Associated With PSA

As exhibited in **Table 4**, gender, age, NIHSS score, hypertension, CHD, DM, drinking, and HDL-C are fed into the multivariate logistic regression model by HADS-A. As exhibited in **Tables 5, 6**, gender, age, NIHSS score, hypertension, DM, drinking, and HDL-C are fed into the multivariate logistic regression model by HAMA and SAS scale, respectively. Multivariate logistic regression (stepwise forward) analysis indicates that hypertension, diabetes mellitus, drinking, high NIHSS score, and low serum HDL-C level are associated with PSA, as shown in **Tables 4–6**.

### 4.3. Post-stroke Anxiety *via* Machine Learning Methods

To compare the performance of classic machine learning methods in PSA prediction, we carry out *k*-fold cross-validation by splitting the dataset into *k* parts. One part is assigned as the testing set and the remaining parts are regarded as training set each time until each part has already been calculated. Therefore, the validation process needs *k*-times comparison. Suppose  $k = 10$ , each machine learning method is employed to predict anxiety test in the *k*-cross validation test, and the results are shown in **Table 7**.

As shown in **Table 7**, the averaging Euclidean Distance obtained by RandomForest method is 18.6254, which outperforms than methods (22.6079, 103.3264, 104.2886, and 130.7300, respectively) on HADS-A test. Likewise, the competitive methods are evaluated by HAMA and SAS tests, which also suggests the superiority of random forest. We plotted the averaging Euclidean distance of each machine learning methods as shown in **Figure 1**.

As shown in **Figure 1**, the random forest method has a lower averaging Euclidean distance than the competitors, which indicates its superiority in predicting PSA. To compare the predicting accuracy of the above-mentioned methods, we compared the results and the boxplot is shown in **Figure 2**.

As shown in **Figure 2**, the decision tree and random forest methods are holding higher accuracy in predicting PSA. Specifically, the abnormal values of random forest method are higher than that of decision tree, which shows the superiority of ensemble learning. To analyze the relationship between accuracy and the varying *k* in the *k*-cross-validation process, we conduct the experiments with different *k*, i.e.,  $k = \{5, 10, 15, 20, 25\}$ , and plot the result as shown in **Figure 3**.

As shown in **Figure 3**, the random forest method (marked with green triangles) outperforms the competitive methods with *k* increasing in all the three anxiety scales. The DecisionTree method is second to RandomForest, and MLP (marked with red rhombus) and SVM (marked with purple triangles) are at the same level and show robustness with *k* increasing. The SGDClassifier (marked with blue square) is inferior to the other methods in general, which suggests more

**TABLE 1** | Significant characteristics between patients with and without post-stroke anxiety (PSA) by Hospital Anxiety and Depression Scale (HADS-A) ( $n = 395$ ).

	A group ( $n = 131$ )	NA group ( $n = 264$ )	Statistics	P-value
Male (yes/no)	101/30	175/89	$\chi^2 = 4.862$	0.027*
Age, year	60.59 $\pm$ 9.87	63.16 $\pm$ 11.40	$t = 2.203$	0.028*
Occupation	23/16/19/27/20/26	32/52/39/51/48/42	$\chi^2 = 6.005$	0.306
Marital status	27/34/31/39	64/79/62/59	$\chi^2 = 2.929$	0.403
BMI score (kg/m <sup>2</sup> )	23.95 $\pm$ 1.64	23.89 $\pm$ 1.26	$t = -0.442$	0.659
NIHSS score	4.26 $\pm$ 3.14	3.31 $\pm$ 2.70	$t = -3.129$	0.002*
NIHSS score group	81/50	204/60	$\chi^2 = 10.389$	0.001*
Hypertension (yes/no)	117/14	162/102	$\chi^2 = 32.973$	0.000*
CHD (yes/no)	14/117	17/247	$\chi^2 = 2.184$	0.139
DM (yes/no)	63/68	85/179	$\chi^2 = 9.441$	0.002*
Smoking (yes/no)	45/86	75/189	$\chi^2 = 1.462$	0.227
Drinking (yes/no)	34/97	36/228	$\chi^2 = 9.111$	0.003*
Lesion location	19/23/65/9/5/10	37/45/141/14/4/23	$\chi^2 = 2.796$	0.731
Lesion side	64/48/19	102/113/49	$\chi^2 = 3.828$	0.147
LDL-C, mmol/L	3.02 $\pm$ 1.00	3.01 $\pm$ 0.97	$t = -0.037$	0.970
HDL-C, mmol/L	1.05 $\pm$ 0.28	1.12 $\pm$ 0.28	$t = 2.235$	0.026*
TC, mmol/L	4.58 $\pm$ 1.33	4.61 $\pm$ 1.25	$t = 0.186$	0.852
TG, mmol/L	1.97 $\pm$ 1.48	1.80 $\pm$ 1.47	$t = -1.043$	0.298
GLU, mmol/L	7.59 $\pm$ 3.12	7.22 $\pm$ 3.06	$t = -1.096$	0.274
UA, umol/L	332.05 $\pm$ 70.88	339.30 $\pm$ 63.51	$t = 0.991$	0.323
CRP, mg/L	3.97 $\pm$ 2.20	3.75 $\pm$ 2.24	$t = -0.922$	0.357
Cr, umol/L	78.26 $\pm$ 26.16	78.19 $\pm$ 14.63	$t = -0.034$	0.973
RBC, *10 <sup>12</sup> /L	5.51 $\pm$ 0.64	5.46 $\pm$ 0.52	$t = -0.800$	0.424
HB, g/L	153.27 $\pm$ 15.81	154.07 $\pm$ 13.12	$t = 0.530$	0.596
PLT, *10 <sup>9</sup> /L	197.42 $\pm$ 39.96	193.78 $\pm$ 35.70	$t = -0.406$	0.685
D-D, ug/ml	0.76 $\pm$ 0.34	0.74 $\pm$ 0.32	$t = -0.885$	0.377
FIB, g/L	3.96 $\pm$ 0.82	4.05 $\pm$ 0.80	$t = -0.554$	0.580
HCY, umol/L	16.25 $\pm$ 4.71	16.24 $\pm$ 4.27	$t = -0.012$	0.991

Occupation includes staff/teacher/worker/retired/farmer/others; marital status includes divorced or separated/married/widowed/others; NIHSS score groups are divided by the number of patients with NIHSS score  $\leq 4$ /the number of patients with NIHSS score  $> 4$ . Lesion location includes lobe/basal ganglia or lateral ventricles/brainstem/cerebellum/thalamus/other; lesion side includes left/right/left and right. The data with \*denote  $P < 0.05$ , which is considered as the statistically significant level. A group, anxiety group; NA group, non-anxiety group; BMI, body mass index; NIHSS, National Institutes of Health Stroke Scale; CHD, coronary heart disease; DM, diabetes mellitus; LDL, low-density lipoprotein; HDL, high-density lipoprotein; TC, total cholesterol; TG, triglyceride; GLU, glucose; UA, uric acid; CRP, C-reactive protein; Cr, creatinine; RBC, red blood cell; HB, hemoglobin; PLT, platelet; D-D, D-dimer; FIB, fibrinogen; HCY, homocysteine.

treatment or optimization of SGDClassifier is needed for further analysis.

## 5. DISCUSSION

With the development of stroke relevant research, post-stroke emotional disorder has attracted more and more attention. The identification of risk factors benefits detecting PSA at an early stage and achieving timely intervention. The statistics in this study show that the frequency of PSA is 33.16% evaluated by HADS-A, 33.67% by HAMA, and 30.38% by SAS, respectively. Our results are consistent with the findings in previous studies (Burton et al., 2013; Broomfield et al., 2014; Knapp et al., 2017, 2020; Rafsten et al., 2018), i.e., 18–36.7% of patients with acute ischemic stroke experienced anxiety 0–2 weeks after stroke onset.

The averaging age of PSA patients is younger than that of patients without PSA by HADS-A and HAMA scale, but there is no statistical difference in SAS evaluation, as shown in **Tables 1–3**. A systematic review of observational studies revealed that older age was the most consistent factor not predictive of PSA (Menlove et al., 2015). This may result from a combination of anxiety disorders being much less common in older adults while an increasing proportion risk of stroke in older adults (McEvoy et al., 2011). Researchers also propose that younger people especially those with a history of anxiety or depression are more probably to have PSA (Chun et al., 2018). Thus, the above-mentioned different viewpoints seem to explain the results of the multivariate analysis in section 4, which shows that age is not a risk factor of PSA. This study exhibits that men are more probably to be anxious after stroke than women, as shown in single factor analysis in **Tables 1, 2**. Burton et al. (2013) declare that 51–64% of PSA are male

**TABLE 2 |** Significant characteristics between patients with and without post-stroke anxiety (PSA) by Hamilton Anxiety Scale (HAMA) scale ( $n = 395$ ).

	A group ( $n = 133$ )	NA group ( $n = 262$ )	Statistics	P-value
Male (yes/no)	104/29	172/90	$\chi^2 = 6.597$	0.010*
Age, year	60.68 $\pm$ 10.06	63.13 $\pm$ 11.33	$t = 2.102$	0.036*
Occupation	23/17/18/26/21/28	32/51/40/52/44/40	$\chi^2 = 6.060$	0.300
Marital status	26/34/32/41	65/79/61/57	$\chi^2 = 4.658$	0.199
BMI score (kg/m <sup>2</sup> )	24.02 $\pm$ 1.58	23.85 $\pm$ 1.29	$t = -1.181$	0.238
NIHSS score	4.20 $\pm$ 3.12	3.33 $\pm$ 2.71	$t = -2.879$	0.004*
NIHSS score group	84/49	201/61	$\chi^2 = 8.073$	0.004*
Hypertension(yes/no)	117/16	162/100	$\chi^2 = 29.055$	0.000*
CHD (yes/no)	12/121	19/243	$\chi^2 = 0.382$	0.536
DM (yes/no)	63/70	85/177	$\chi^2 = 8.388$	0.004*
Smoking (yes/no)	45/88	75/187	$\chi^2 = 1.132$	0.287
Drinking (yes/no)	34/99	36/226	$\chi^2 = 8.458$	0.004*
Lesion location	20/24/63/9/5/12	36/44/143/14/4/21	$\chi^2 = 3.409$	0.637
Lesion side	64/50/19	102/111/49	$\chi^2 = 3.265$	0.195
LDL-C, mmol/L	3.00 $\pm$ 1.01	3.02 $\pm$ 0.96	$t = 0.210$	0.833
HDL-C, mmol/L	1.05 $\pm$ 0.28	1.12 $\pm$ 0.28	$t = 2.427$	0.016*
TC, mmol/L	4.56 $\pm$ 1.34	4.62 $\pm$ 1.24	$t = 0.452$	0.651
TG, mmol/L	1.95 $\pm$ 1.47	1.81 $\pm$ 1.48	$t = -0.924$	0.356
GLU, mmol/L	7.55 $\pm$ 3.10	7.24 $\pm$ 3.07	$t = -0.939$	0.348
UA, umol/L	334.49 $\pm$ 69.93	338.12 $\pm$ 64.08	$t = 0.501$	0.617
CRP, mg/L	4.00 $\pm$ 2.20	3.74 $\pm$ 2.24	$t = -1.122$	0.263
Cr, umol/L	78.94 $\pm$ 25.65	77.84 $\pm$ 14.94	$t = -0.536$	0.592
RBC, $\times 10^{12}/L$	5.54 $\pm$ 0.62	5.45 $\pm$ 0.53	$t = -1.386$	0.167
HB, g/L	153.87 $\pm$ 13.71	153.77 $\pm$ 14.25	$t = -0.065$	0.948
PLT, $\times 10^9/L$	187.40 $\pm$ 37.10	187.93 $\pm$ 35.53	$t = 0.138$	0.891
D-D, ug/ml	0.75 $\pm$ 0.34	0.75 $\pm$ 0.31	$t = -0.068$	0.946
FIB, g/L	4.00 $\pm$ 0.82	4.03 $\pm$ 0.80	$t = 0.363$	0.717
HCY, umol/L	16.44 $\pm$ 4.61	16.14 $\pm$ 4.32	$t = -0.638$	0.524

Occupation includes staff/teacher/worker/retired/farmer/others; marital status includes divorced or separated/married/widowed/others; NIHSS score groups are divided by the number of patients with NIHSS score  $\leq 4$ /the number of patients with NIHSS score  $> 4$ . Lesion location includes lobe/basal ganglia or lateral ventricles/brainstem/cerebellum/thalamus/other; lesion side includes left/right/left and right. The data with \*denote  $P < 0.05$ , which is considered as the statistically significant level. A group, anxiety group; NA group, non-anxiety group; BMI, body mass index; NIHSS, National Institutes of Health Stroke Scale; CHD, coronary heart disease; DM, diabetes mellitus; LDL, low-density lipoprotein; HDL, high-density lipoprotein; TC, total cholesterol; TG, triglyceride; GLU, glucose; UA, uric acid; CRP, C-reactive protein; Cr, creatinine; RBC, red blood cell; HB, hemoglobin; PLT, platelet; D-D, D-dimer; FIB, fibrinogen; HCY, homocysteine.

patients. As shown in multivariate logistic analysis, drinking is an independent risk factor of PSA. Notably, male patients are more likely to drink, which may support Burton's viewpoint. On the contrary, Beauchamp et al. (2020) insist that PSA is more common in female stroke patients while gender is analyzed in univariable analysis, but gender is not a statistically significant factor in their or our multivariable analysis. Thus, just as other precious researchers suggested (Astrom, 1996; Schultz et al., 1997; Leppävuori et al., 2003; Shuibin, 2006; Barker-Collo, 2007; Carod-Artal et al., 2009; Sagen et al., 2010), the relation between age and PSA is not sure and caution should be observed when making conclusions on the association of gender and PSA. Besides, the NIHSS scores of PSA patients are more prone to be higher than that of patients without anxiety in our study, suggesting that the severity of stroke is a risk factor for PSA, which is consistent with previous results (Menlove et al., 2015). It is clear that social isolation, loneliness, and

single status are linked to higher rates of cardiovascular disease and stroke mortality and morbidity (Tillmann et al., 2017; Hakulinen et al., 2018), and the association between PSA and non-married status has also been revealed (Beauchamp et al., 2020). As for marital status in our study, different from our subjective clinical experience or above studies, there are no significant differences between different marital statuses. Since a large number of patients are reluctant to reveal their concrete marital status and we grouped these patients into "others," this result may generate false-negative data. The relationship between PSA and affected brain regions is controversial. We find no association between lesion location or lesion side and PSA, which is the same with Chun et al. (2018), and a meta-analysis about PSA (Burton et al., 2013) summarized that no association was observed between PSA and lesion location in five of six studies (Astrom, 1996; Ghika-Schmid et al., 1999; Leppävuori et al., 2003; Fure et al., 2006; Barker-Collo, 2007).

**TABLE 3 |** Significant characteristics between patients with and without post-stroke anxiety (PSA) by SAS scale ( $n = 395$ ).

	A group ( $n = 120$ )	NA group ( $n = 275$ )	Statistics	P-value
Male (yes/no)	93/27	183/92	$\chi^2 = 4.763$	0.029*
Age, year	60.84 $\pm$ 10.02	62.95 $\pm$ 11.32	$t = 1.757$	0.080
Occupation	22/15/17/24/18/24	33/53/41/54/50/44	$\chi^2 = 5.937$	0.312
Marital status	25/28/31/36	66/85/62/62	$\chi^2 = 4.295$	0.231
BMI score (kg/m <sup>2</sup> )	23.93 $\pm$ 1.48	23.89 $\pm$ 1.36	$t = -0.272$	0.786
NIHSS score	4.14 $\pm$ 3.03	3.40 $\pm$ 2.79	$t = -2.379$	0.018*
NIHSS score group	74/46	211/64	$\chi^2 = 9.431$	0.002*
Hypertension(yes/no)	108/12	171/104	$\chi^2 = 31.168$	0.000*
CHD (yes/no)	11/109	20/255	$\chi^2 = 0.414$	0.520
DM (yes/no)	61/59	87/188	$\chi^2 = 13.141$	0.000*
Smoking (yes/no)	41/79	79/196	$\chi^2 = 1.169$	0.280
Drinking (yes/no)	31/89	39/236	$\chi^2 = 7.778$	0.005*
Lesion location	18/22/57/9/5/9	38/46/149/14/4/24	$\chi^2 = 4.603$	0.466
Lesion side	57/44/19	109/117/49	$\chi^2 = 2.129$	0.345
LDL-C, mmol/L	3.05 $\pm$ 1.01	3.00 $\pm$ 0.97	$t = -0.512$	0.609
HDL-C, mmol/L	1.04 $\pm$ 0.27	1.12 $\pm$ 0.28	$t = 2.552$	0.011*
TC, mmol/L	4.59 $\pm$ 1.33	4.60 $\pm$ 1.25	$t = 0.079$	0.937
TG, mmol/L	1.98 $\pm$ 1.53	1.80 $\pm$ 1.45	$t = -1.100$	0.272
GLU, mmol/L	7.55 $\pm$ 3.08	7.25 $\pm$ 3.08	$t = -0.896$	0.371
UA, umol/L	332.85 $\pm$ 69.22	338.66 $\pm$ 64.65	$t = 0.804$	0.422
CRP, mg/L	3.96 $\pm$ 2.20	3.77 $\pm$ 2.24	$t = -0.766$	0.444
Cr, umol/L	79.23 $\pm$ 26.62	77.77 $\pm$ 14.90	$t = -0.697$	0.486
RBC, $\times 10^{12}/L$	5.53 $\pm$ 0.60	5.46 $\pm$ 0.55	$t = -1.062$	0.289
HB, g/L	153.82 $\pm$ 13.83	153.80 $\pm$ 14.17	$t = -0.008$	0.993
PLT, $\times 10^9/L$	187.41 $\pm$ 36.33	187.90 $\pm$ 35.95	$t = 0.124$	0.901
D-D, ug/ml	0.77 $\pm$ 0.34	0.74 $\pm$ 0.31	$t = -0.723$	0.470
FIB, g/L	4.01 $\pm$ 0.80	4.03 $\pm$ 0.81	$t = 0.156$	0.876
HCY, umol/L	16.53 $\pm$ 4.69	16.12 $\pm$ 4.30	$t = -0.847$	0.397

Occupation includes staff/teacher/worker/retired/farmer/others; marital status includes divorced or separated/married/widowed/others; NIHSS score groups are divided by the number of patients with NIHSS score  $\leq 4$ /the number of patients with NIHSS score  $> 4$ . Lesion location includes lobe/basal ganglia or lateral ventricles/brainstem/cerebellum/thalamus/other; lesion side includes left/right/left and right. The data with \*denote  $P < 0.05$ , which is considered as the statistically significant level. A group, anxiety group; NA group, non-anxiety group; BMI, body mass index; NIHSS, National Institutes of Health Stroke Scale; CHD, coronary heart disease; DM, diabetes mellitus; LDL, low-density lipoprotein; HDL, high-density lipoprotein; TC, total cholesterol; TG, triglyceride; GLU, glucose; UA, uric acid; CRP, C-reactive protein; Cr, creatinine; RBC, red blood cell; HB, hemoglobin; PLT, platelet; D-D, D-dimer; FIB, fibrinogen; HCY, homocysteine.

**TABLE 4 |** Multivariate logistic regression analyses of the risk factors associated with post-stroke anxiety (PSA) evaluated by Hospital Anxiety and Depression Scale (HADS-A) scale.

	B	S.E.	Wald	OR	OR 95% CI	P-value
Constant	-1.882	0.583	10.473	0.152		0.001
Hypertension	1.843	0.333	30.636	6.318	3.289-12.136	0.000
DM	0.756	0.241	9.814	2.129	1.327-3.416	0.002
Drinking	1.171	0.306	14.603	3.324	1.769-5.877	0.000
HDL-C	-0.961	0.442	4.733	0.382	0.161-0.909	0.030
NIHSS score group	0.863	0.225	11.429	2.371	1.437-3.912	0.001

NIHSS score groups are divided by the number of patients with NIHSS score  $\leq 4$ /the number of patients with NIHSS score  $> 4$ . OR, odds ratio; DM, diabetes mellitus; HDL, high-density lipoprotein; NIHSS, National Institutes of Health Stroke Scale.

On the contrary, Tang et al. (2012) reported that PSA patients were more likely to have right frontal acute infarcts compared with non-PSA group. Differences in the above results may due to the small sample size, lack of detailed assessment of lesion

locations and the diversity between CT and MRI scans to estimate lesion locations.

This study also revealed that patients with hypertension or diabetes mellitus are more prone to have PSA. A cross-sectional

**TABLE 5 |** Multivariate logistic regression analyses of the risk factors associated with post-stroke anxiety (PSA) evaluated by Hamilton Anxiety Scale (HAMA) scale.

	<b>B</b>	<b>S.E.</b>	<b>Wald</b>	<b>OR</b>	<b>OR 95% CI</b>	<b>P-value</b>
Constant	-1.535	0.566	7.363	0.216		0.007
Hypertension	1.647	0.315	27.419	5.193	2.803–9.621	0.000
DM	0.689	0.237	8.423	1.991	1.251–3.170	0.004
Drinking	1.105	0.301	13.515	3.020	1.675–5.444	0.000
HDL-C	-1.034	0.437	5.592	0.356	0.151–0.838	0.018
NIHSS score group	0.744	0.251	8.752	2.104	1.285–3.444	0.003

NIHSS score groups are divided by the number of patients with NIHSS score  $\leq 4$ /the number of patients with NIHSS score  $> 4$ . OR, odds ratio; DM, diabetes mellitus; HDL, high-density lipoprotein; NIHSS, National Institutes of Health Stroke Scale.

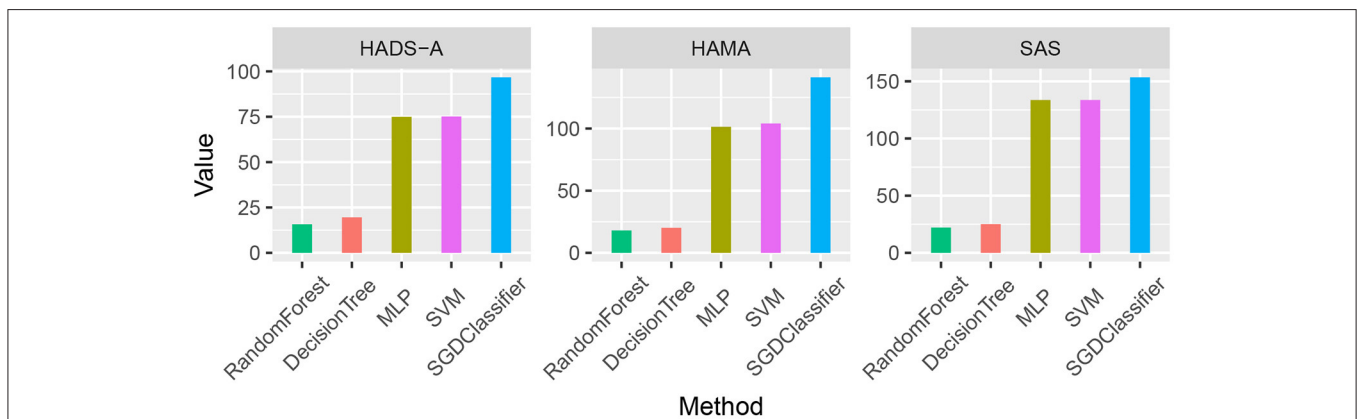
**TABLE 6 |** Multivariate logistic regression analyses of the risk factors associated with post-stroke anxiety (PSA) evaluated by SAS scale.

	<b>B</b>	<b>S.E.</b>	<b>Wald</b>	<b>OR</b>	<b>OR 95% CI</b>	<b>P-value</b>
Constant	-1.949	0.605	10.359	0.142		0.001
Hypertension	1.888	0.353	28.530	6.605	3.304–13.205	0.000
DM	0.906	0.247	13.474	2.474	1.525–4.013	0.000
Drinking	1.145	0.311	13.552	3.141	1.708–5.777	0.000
HDL-C	-1.133	0.460	6.077	0.322	0.131–0.793	0.014
NIHSS score group	0.851	0.260	10.680	2.342	1.406–3.902	0.001

NIHSS score groups are divided by the number of patients with NIHSS score  $\leq 4$ /the number of patients with NIHSS score  $> 4$ . OR, odds ratio; DM, diabetes mellitus; HDL, high-density lipoprotein; NIHSS, National Institutes of Health Stroke Scale.

**TABLE 7 |** Averaging Euclidean distance of 10-fold cross-validation on the five machine learning methods by comparing with Hospital Anxiety and Depression Scale (HADS-A), Hamilton Anxiety Scale (HAMA), and SAS test, respectively.

	<b>RandomForest</b>	<b>SVM</b>	<b>DecisionTree</b>	<b>SGDClassifier</b>	<b>MLP</b>
HADS-A	15.7433	17.7001	74.9658	75.1603	99.0425
HAMA	18.0221	20.6518	101.3920	104.0841	123.5836
SAS	22.1108	29.4717	133.6214	133.6214	169.5637

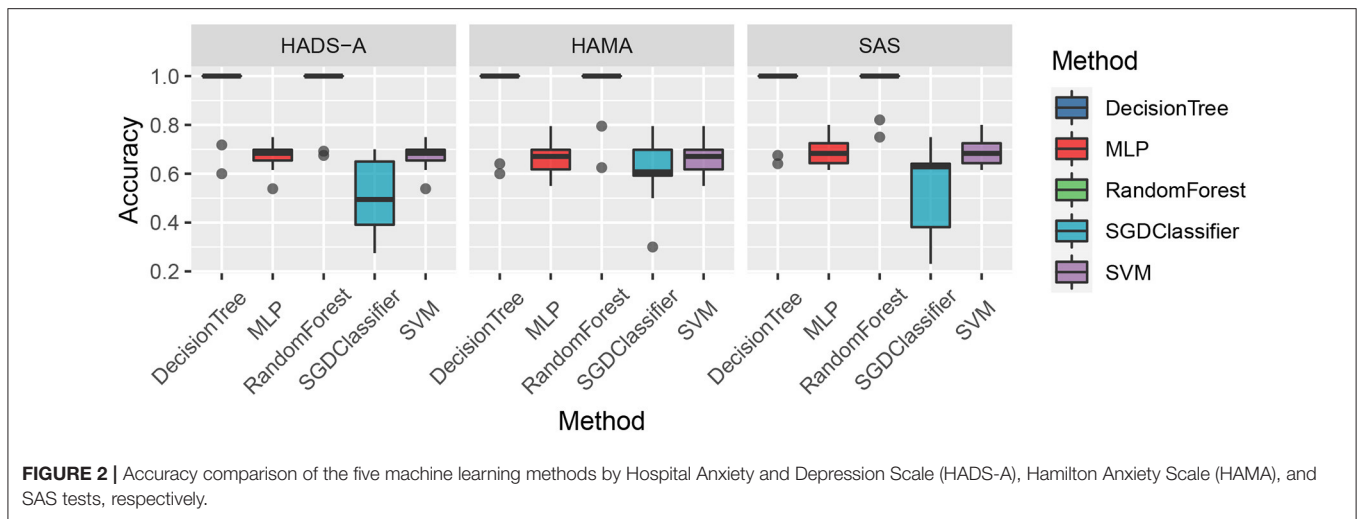


**FIGURE 1 |** Averaging Euclidean Distance comparison of the five machine learning methods by Hospital Anxiety and Depression Scale (HADS-A), Hamilton Anxiety Scale (HAMA), and SAS tests, respectively.

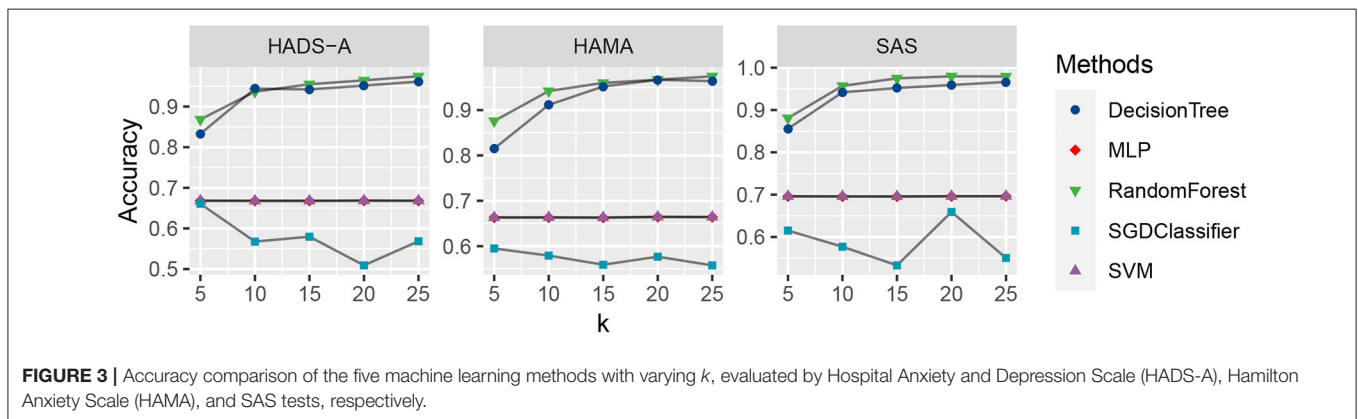
study also found that chronic physical diseases is an identified factor significantly associated with post-stroke mental health (Almhdawi et al., 2020). Interestingly, our study found that the level of HDL-C is independent protective factors

for PSA. Although relevant research is rare and the mechanism is unclear, it has been commonly acknowledged that a higher level of HDL-C is a protective factor for stroke, and the protective mechanism of HDL-C for PSA may be the same as that for stroke.





**FIGURE 2 |** Accuracy comparison of the five machine learning methods by Hospital Anxiety and Depression Scale (HADS-A), Hamilton Anxiety Scale (HAMA), and SAS tests, respectively.



**FIGURE 3 |** Accuracy comparison of the five machine learning methods with varying  $k$ , evaluated by Hospital Anxiety and Depression Scale (HADS-A), Hamilton Anxiety Scale (HAMA), and SAS tests, respectively.

Machine learning methods are commonly applied in modern medical research, such as image processing, computer-aided diagnosis, and so on. It is really a challenging task to predict PSA with limited clinical data and it is an open issue. To the best of our knowledge, there are bare research about predicting PSA *via* machine learning methods. In our study, machine learning methods are employed to predict PSA, and anxiety scales are utilized as evaluation benchmarks. As shown in **Figure 1**, random forest methods have the averaging Euclidean Distance of 18.6254 in HADS-A, HAMA, and SAS scales, being superior to that of DecisionTree, SVM, SGDClassifier, and MLP (i.e., 22.6079, 103.3264, 104.2886, and 130.7300, respectively), which is consistent with the findings in Tripathi et al. (2019). As plotted in **Figure 2**, decision tree and random forest methods show higher accuracy than the other three methods. In the  $k$ -fold cross-validation process, we set  $k = 10$  and collect all the accuracy results, the abnormal values of decision tree (i.e., 0.6410 and 0.7250 in HADS-A, 0.5500 and 0.6410 in HAMA, 0.5897 and 0.7250 in SAS) and random forest (i.e., 0.6750 and 0.6923 in HADS-A, 0.6250 and 0.7949 in HAMA, 0.7500 and 0.8205 in SAS) are obvious, which indicate the optimization of training process is expected in future works. In a word, with the aid of ensemble learning, random

forest can be applied in PSA prediction. **Figure 3** shows the relationship between  $k$  and accuracy from different anxiety scales. With the increasing of  $k$ , the predictive accuracy of machine learning methods improves gradually. The RandomForest and DecisionTree methods are at the same level, outperform the other three methods. RandomForest methods show superiority to DecisionTree, with a slender advantage in general, which further indicates the capability of RandomForest methods in predicting PSA.

The limitations of this study are listed as follows. In section 4, the risk factors of PSA are analyzed merely by tables, other variables potentially associated with PSA are not under consideration. For example, a review (Popa-Wagner et al., 2020) summarized plenty of articles pointed that lifestyle (such as high sugar diets, high fat diets or calorie restriction) can influence the onset, severity, and duration of the stroke, so it will be interesting and meaningful to study the relationship between lifestyle and PSA. Slevin et al. (2015) demonstrated that mCRP may be responsible for promoting dementia after ischemia stroke by sufficient *in vitro* experiments, murine models, and detailed histological studies, emphasizing the influence of inflammation on stroke and suggesting that the relationship between systemic inflammation and PSA should be further studied. Besides, since

previous studies (Burton et al., 2013; Menlove et al., 2015) have reported significant associations between PSA and Pre-stroke depression, aphasia, dementia, or cognitive impairment, we excluded these patients in our study and this selection bias may limit the generalizability of the findings. Thus, all the above would be further studied in the forthcoming research.

## 6. CONCLUSION

Anxiety after stroke is common and disabling (Chun et al., 2018), which may lead to severe effects and bring great troubles to patients. In this paper, we carry out a series of experiments to analyze the risk factors and employ machine learning methods to predict PSA. The experimental results suggest that hypertension, diabetes mellitus, drinking, disability, and low serum HDL-C levels are closely related to anxiety in acute ischemic stroke, and random forest can be applied in PSA prediction. These results not only provide insight into the possible factors related to PSA but also benefit predicting anxiety of acute ischemic stroke patients, providing a theoretical basis for the treatment of PSA. It is of great significance in lowering costs of care by shortening the course of treatment or reducing the possibility of anxiety with the aid of the findings in this work, and we hope it will shed light on more forthcoming researchers to further explore the uncharted part of this promising field.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

## REFERENCES

- Abiodun, O. (1994). A validity study of the hospital anxiety and depression scale in general hospital units and a community sample in Nigeria. *Br. J. Psychiatry* 165, 669–672. doi: 10.1192/bjp.165.5.669
- Almhdawi, K. A., Alazrai, A., Kanaan, S., Shyyab, A. A., Oteir, A. O., Mansour, Z. M., et al. (2020). Post-stroke depression, anxiety, and stress symptoms and their associated factors: a cross-sectional study. *Neuropsychological Rehabilitation*, pages 1–14. doi: 10.1080/09602011.2020.1760893
- Alpaydin, E. (2020). *Introduction to Machine Learning*. Cambridge, MA: MIT Press.
- Astrom, M. (1996). Generalized anxiety disorder in stroke patients: a 3-year longitudinal study. *Stroke* 27, 270–275. doi: 10.1161/01.STR.27.2.270
- Barker-Collo, S. L. (2007). Depression and anxiety 3 months post stroke: prevalence and correlates. *Arch. Clin. Neuropsychol.* 22, 519–531. doi: 10.1016/j.acn.2007.03.002
- Beauchamp, J. E. S., Montiel, T. C., Cai, C., Tallavajhula, S., Hinojosa, E., Okpala, M. N., et al. (2020). A retrospective study to identify novel factors associated with post-stroke anxiety. *J. Stroke Cerebrovasc. Dis.* 29:104582. doi: 10.1016/j.jstrokecerebrovasdis.2019.104582
- Bjelland, I., Dahl, A. A., Haug, T. T., and Neckelmann, D. (2002). The validity of the hospital anxiety and depression scale: an updated literature review. *J. Psychosom. Res.* 52, 69–77. doi: 10.1016/S0022-3999(01)00296-3
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324
- Broomfield, N. M., Quinn, T. J., Abdul-Rahim, A. H., Walters, M. R., and Evans, J. J. (2014). Depression and anxiety symptoms post-stroke/tia: prevalence and

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Regional Medical Scientific Research Ethics Committee of the First Affiliated Hospital of China Medical University (IRB no.2020368). The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

JW wrote the original draft. DZ designed the post-stroke anxiety evaluation for all the collected cases. ML revised the manuscript. XH conducted the machine learning experiments and plotted the figures. XS checked the manuscript and made final modifications. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the National Natural Science Foundation of China (81871104). The funding body supplied this manuscript with funding for data collection, analysis, and interpretation, as well as writing the manuscript.

## ACKNOWLEDGMENTS

We would like to thank the reviewers for their careful reading and useful comments that helped us to improve the final version of this paper.

associations in cross-sectional data from a regional stroke registry. *BMC Neurol.* 14:198. doi: 10.1186/s12883-014-0198-8

- Burton, C. A. C., Murray, J., Holmes, J., Astin, F., Greenwood, D., and Knapp, P. (2013). Frequency of anxiety after stroke: a systematic review and meta-analysis of observational studies. *Int. J. Stroke* 8, 545–559. doi: 10.1111/j.1747-4949.2012.00906.x
- Carod-Artal, F. J., Coral, L. F., Trizotto, D. S., and Moreira, C. M. (2009). Poststroke depression: prevalence and determinants in Brazilian stroke patients. *Cerebrovasc. Dis.* 28, 157–165. doi: 10.1159/000226114
- Chen, Z., Jiang, B., Ru, X., Sun, H., Sun, D., Liu, X., et al. (2017). Mortality of stroke and its subtypes in china: results from a nationwide population-based survey. *Neuroepidemiology* 48, 95–102. doi: 10.1159/000477494
- Chinese Medical Association (2018). Chinese guidelines for diagnosis and treatment of acute ischemic stroke 2018. *Chin. J. Neurol.* 51, 666–682. doi: 10.3760/cma.j.issn.1006-7876.2018.09.004
- Chun, H.-Y. Y., Whiteley, W. N., Dennis, M. S., Mead, G. E., and Carson, A. J. (2018). Anxiety after stroke: the importance of subtyping. *Stroke* 49, 556–564. doi: 10.1161/STROKEAHA.117.020078
- Corradi, J. P., Thompson, S., Mather, J. F., Waszynski, C. M., and Dicks, R. S. (2018). Prediction of incident delirium using a random forest classifier. *J. Med. Syst.* 42:261. doi: 10.1007/s10916-018-1109-0
- Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. doi: 10.1007/BF00994018
- Cuspidi, C., Tadic, M., Grassi, G., and Mancia, G. (2018). Treatment of hypertension: the ESH/ESC guidelines recommendations. *Pharmacol. Res.* 128, 315–321. doi: 10.1016/j.phrs.2017.10.003
- Dunstan, D. A., and Scott, N. (2018). Assigning clinical significance and symptom severity using the zung scales: levels of misclassification arising from

- confusion between index and raw scores. *Depress. Res. Treat.* 2018:9250972. doi: 10.1155/2018/9250972
- Dunstan, D. A., and Scott, N. (2020). Norms for Zung's self-rating anxiety scale. *BMC Psychiatry* 20:6. doi: 10.1186/s12888-019-2427-6
- Fure, B., Wyller, T. B., Engedal, K., and Thommessen, B. (2006). Emotional symptoms in acute ischemic stroke. *Int. J. Geriatr. Psychiatry* 21, 382–387. doi: 10.1002/gps.1482
- Ghika-Schmid, F., Van Melle, G., Guex, P., and Bogousslavsky, J. (1999). Subjective experience and behavior in acute stroke: the lausanne emotion in acute stroke study. *Neurology* 52, 22–22. doi: 10.1212/WNL.52.1.22
- Gilworth, G., Phil, M., Cert, A., Sansam, K., and Kent, R. (2009). Personal experiences of returning to work following stroke: an exploratory study. *Work* 34, 95–103. doi: 10.3233/WOR-2009-0906
- Hackett, M. L., Köhler, S., T O'Brien, J., and Mead, G. E. (2014). Neuropsychiatric outcomes of stroke. *Lancet Neurol.* 13, 525–534. doi: 10.1016/S1474-4422(14)70016-X
- Hakulinen, C., Pulkki-Råback, L., Virtanen, M., Jokela, M., Kivimäki, M., and Elovainio, M. (2018). Social isolation and loneliness as risk factors for myocardial infarction, stroke and mortality: UK biobank cohort study of 479 054 men and women. *Heart* 104, 1536–1542. doi: 10.1136/heartjnl-2017-312663
- Hamilton, M. (1959). The assessment of anxiety states by rating. *Br. J. Med. Psychol.* 32, 50–55. doi: 10.1111/j.2044-8341.1959.tb00467.x
- Harrison, J. K., McArthur, K. S., and Quinn, T. J. (2013). Assessment scales in stroke: clinimetric and clinical considerations. *Clin. Intervent. Aging* 8:201. doi: 10.2147/CIA.S32405
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY: Springer Science & Business Media.
- Joachims, T. (1998). "Text categorization with support vector machines: learning with many relevant features," in *European Conference on Machine Learning* (Chemnitz: Springer), 137–142. doi: 10.1007/BFb0026683
- Kamiński, B., Jakubczyk, M., and Szufel, P. (2018). A framework for sensitivity analysis of decision trees. *Central Eur. J. Oper. Res.* 26, 135–159. doi: 10.1007/s10100-017-0479-6
- Kjeldsen, S. E., Steneshjem, A., Os, I., Van de Borne, P., Burnier, M., Narkiewicz, K., et al. (2016). Treatment of high blood pressure in elderly and octogenarians: European society of hypertension statement on blood pressure targets. *Blood Press.* 25, 333–336. doi: 10.1080/08037051.2016.1236329
- Knapp, P., Burton, C. A. C., Holmes, J., Murray, J., Gillespie, D., Lightbody, C. E., et al. (2017). Interventions for treating anxiety after stroke. *Cochrane Database Syst. Rev.* 5:CD008860. doi: 10.1002/14651858.CD008860.pub3
- Knapp, P., Dunn-Roberts, A., Sahib, N., Cook, L., Astin, F., Kontou, E., et al. (2020). Frequency of anxiety after stroke: an updated systematic review and meta-analysis of observational studies. *Int. J. Stroke* 15, 244–255. doi: 10.1177/1747493019896958
- Kohavi, R. (1996). "Scaling up the accuracy of Naive-Bayes classifiers: a decision-tree hybrid," in *Kdd, Vol. 96*, Portland, Oregon, USA 202–207.
- Leppävuori, A., Pohjasvaara, T., Vataja, R., Kaste, M., and Erkinjuntti, T. (2003). Generalized anxiety disorders three to four months after ischemic stroke. *Cerebrovasc. Dis.* 16, 257–264. doi: 10.1159/000071125
- Lincoln, N., Brinkmann, N., Cunningham, S., Dejaeger, E., De Weerd, W., Jenni, W., et al. (2013). Anxiety and depression after stroke: a 5 year follow-up. *Disabil. Rehabil.* 35, 140–145. doi: 10.3109/09638288.2012.691939
- Lyden, P. (2017). Using the national institutes of health stroke scale: a cautionary tale. *Stroke* 48, 513–519. doi: 10.1161/STROKEAHA.116.015434
- Maier, W., Buller, R., Philipp, M., and Heuser, I. (1988). The hamilton anxiety scale: reliability, validity and sensitivity to change in anxiety and depressive disorders. *J. Affect. Disord.* 14, 61–68. doi: 10.1016/0165-0327(88)90072-9
- Mazzaglia, G., Britton, A. R., Altmann, D. R., and Chenet, L. (2001). Exploring the relationship between alcohol consumption and non-fatal or fatal stroke: a systematic review. *Addiction* 96, 1743–1756. doi: 10.1046/j.1360-0443.2001.961217434.x
- McEvoy, P. M., Grove, R., and Slade, T. (2011). Epidemiology of anxiety disorders in the Australian general population: findings of the 2007 Australian national survey of mental health and wellbeing. *Austral. N. Z. J. Psychiatry* 45, 957–967. doi: 10.3109/00048674.2011.624083
- Mei, X., Wang, R., Yang, W., Qian, F., Ye, X., Zhu, L., et al. (2018). Predicting malignancy of pulmonary ground-glass nodules and their invasiveness by random forest. *J. Thorac. Dis.* 10:458. doi: 10.21037/jtd.2018.01.88
- Menlove, L., Crayton, E., Kneebone, I., Allen-Crooks, R., Otto, E., and Harder, H. (2015). Predictors of anxiety after stroke: a systematic review of observational studies. *J. Stroke Cerebrovasc. Dis.* 24, 1107–1117. doi: 10.1016/j.jstrokecerebrovasdis.2014.12.036
- Müller, B., Reinhardt, J., and Strickland, M. T. (2012). *Neural Networks: An Introduction*. New York, NY: Springer Science & Business Media.
- Patkar, A. A., Hill, K., Batra, V., Vergare, M. J., and Leone, F. T. (2003). A comparison of smoking habits among medical and nursing students. *Chest* 124, 1415–1420. doi: 10.1378/chest.124.4.1415
- Pérez-Pi nar, M., Ayerbe, L., González, E., Mathur, R., Foguet-Boreu, Q., and Ayis, S. (2017). Anxiety disorders and risk of stroke: a systematic review and meta-analysis. *Eur. Psychiatry* 41, 102–108. doi: 10.1016/j.eurpsy.2016.11.004
- Popa-Wagner, A., Dumitrascu, D. I., Capitanescu, B., Petcu, E. B., Surugiu, R., Fang, W.-H., et al. (2020). Dietary habits, lifestyle factors and neurodegenerative diseases. *Neural Regen. Res.* 15:394. doi: 10.4103/1673-5374.266045
- Rafsten, L., Danielsson, A., and Sunnerhagen, K. S. (2018). Anxiety after stroke: a systematic review and meta-analysis. *J. Rehabil. Med.* 50, 769–778. doi: 10.2340/16501977-2384
- Remes, O., Brayne, C., Van Der Linde, R., and Lafortune, L. (2016). A systematic review of reviews on the prevalence of anxiety disorders in adult populations. *Brain Behav.* 6:e00497. doi: 10.1002/brb3.497
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature* 323, 533–536. doi: 10.1038/323533a0
- Saad, D. (1998). Online algorithms and stochastic approximations. *Online Learn.* 5, 6–3.
- Sagen, U., Finset, A., Moum, T., Mørland, T., Vik, T. G., Nagy, T., and Dammen, T. (2010). Early detection of patients at risk for anxiety, depression and apathy after stroke. *Gen. Hosp. Psychiatry* 32, 80–85. doi: 10.1016/j.genhosppsych.2009.10.001
- Samad, M. D., Ulloa, A., Wehner, G. J., Jing, L., Hartzel, D., Good, C. W., et al. (2019). Predicting survival from large echocardiography and electronic health record datasets: optimization with machine learning. *Cardiovasc. Imaging* 12, 681–689. doi: 10.1016/j.jcmg.2018.04.026
- Schultz, S. K., Castillo, C. S., Rosier, J. T., and Robinson, R. G. (1997). Generalized anxiety and depression: assessment over 2 years after stroke. *Am. J. Geriatr. Psychiatry* 5, 229–237. doi: 10.1097/00019442-199700530-00007
- Shuibin, L. (2006). Psychological mood and its related factors in patients with cerebral infarction. *Chinese J. Tissue Eng. Res.* 10, 186–188. doi: 10.3321/j.issn:1673-8225.2006.46.018
- Slevin, M., Matou, S., Zeinolabediny, Y., Corpas, R., Weston, R., Liu, D., et al. (2015). Monomeric c-reactive protein—a key molecule driving development of Alzheimer's disease associated with brain ischaemia? *Sci. Rep.* 5, 1–21. doi: 10.1038/srep13281
- Sun, H., Zou, X., and Liu, L. (2013). Epidemiological factors of stroke: a survey of the current status in china. *J. Stroke* 15:109. doi: 10.5853/jos.2013.15.2.109
- Tang, W. K., Chen, Y., Lu, J., Liang, H., Chu, W. C. W., Tong Mok, V. C., et al. (2012). Frontal infarcts and anxiety in stroke. *Stroke* 43, 1426–1428. doi: 10.1161/STROKEAHA.111.640482
- Tillmann, T., Pikhart, H., Peasey, A., Kubinova, R., Pajak, A., Tamosiunas, A., et al. (2017). Psychosocial and socioeconomic determinants of cardiovascular mortality in Eastern Europe: a multicentre prospective cohort study. *PLoS Med.* 14:e1002459. doi: 10.1371/journal.pmed.1002459
- Tripathi, A., Xu, Z. Z., Xue, J., Poulsen, O., Gonzalez, A., Humphrey, G., et al. (2019). Intermittent hypoxia and hypercapnia reproducibly change the gut microbiome and metabolome across rodent model systems. *MSystems* 4:e00058-19. doi: 10.1128/mSystems.00058-19
- Van Der Heijden, F., Duin, R. P., De Ridder, D., and Tax, D. M. (2005). *Classification, Parameter Estimation and State Estimation: An Engineering Approach Using MATLAB*. Chichester: John Wiley & Sons. doi: 10.1002/0470090154
- Wolfe, C. D. (2000). The impact of stroke. *Br. Med. Bull.* 56, 275–286. doi: 10.1258/0007142001903120

- World Health Organization Group (1985). *Diabetes Mellitus: Technical Report Series 727*. Geneva: World Health Organization.
- Zhang, T. (2004). "Solving large scale linear prediction problems using stochastic gradient descent algorithms," in *Proceedings of the Twenty-First International Conference on Machine Learning* (Banff, AB), 116. doi: 10.1145/1015330.1015332
- Zigmond, A. S., and Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatr. Scand.* 67, 361–370. doi: 10.1111/j.1600-0447.1983.tb09716.x
- Zung, W. W. (1971). A rating instrument for anxiety disorders. *Psychosomatics* 12, 371–379. doi: 10.1016/S0033-3182(71)71479-0

**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.

Copyright © 2021 Wang, Zhao, Lin, Huang and Shang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). 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.