RESEARCH

Development and validation of radiology-clinical statistical and machine learning model for stroke-associated pneumonia after first intracerebral haemorrhage

Wenru Zhang[1], Ying Zhou[1], Liuhui Xu[1], Chaomin Qiu[1], Zhixian Luo[1], Zhenghao Jiang[2], Xinyi Tao[2], Yingjie Wu[2], Shishi Yao[2], Hang Huang[1], Xinshi Wang[3,4], Yunjun Yang[1] and Ru Lin[1*]

Abstract

Background Society is burdened with stroke-associated pneumonia (SAP) after intracerebral haemorrhage (ICH). Cerebral small vessel disease (CSVD) complicates clinical manifestations of stroke. In this study, we redefined the CSVD burden score and incorporated it into a novel radiological-clinical prediction model for SAP.

Materials and methods A total of 1278 patients admitted to a tertiary hospital between 1 January 2010 and 31 December 2019 were included. The participants were divided into training and testing groups using fivefold cross-validation method. Four models, two traditional statistical models (logistic regression and ISAN) and two machine learning models (random forest and support vector machine), were established and evaluated. The outcomes and baseline characteristics were compared between the SAP and non-SAP groups.

Results Among the of 1278 patients, 281(22.0%) developed SAP after their first ICH. Multivariate analysis revealed that the logistic regression (LR) model was superior in predicting SAP in both the training and testing groups. Independent predictors of SAP after ICH included total CSVD burden score (OR, 1.29; 95% CI, 1.03–1.54), haematoma extension into ventricle (OR, 2.28; 95% CI, 1.87–3.31), haematoma with multilobar involvement (OR, 2.14; 95% CI, 1.44–3.18), transpharyngeal intubation operation (OR, 3.89; 95% CI, 2.7–5.62), admission NIHSS score ≥ 10 (OR, 2.06; 95% CI, 1.42–3.01), male sex (OR, 1.69; 95% CI, 1.16–2.52), and age ≥ 67 (OR, 2.24; 95% CI, 1.56–3.22). The patients in the SAP group had worse outcomes than those in the non-SAP group.

Conclusion This study established a clinically combined imaging model for predicting stroke-associated pneumonia and demonstrated superior performance compared with the existing ISAN model. Given the poor outcomes observed in patients with SAP, the use of individualised predictive nomograms is vital in clinical practice.

Keywords Intracerebral haemorrhage, Pneumonia, Cerebral vascular disease, Machine learning, Risk factor

*Correspondence:
Ru Lin
linru7459@gmail.com
Full list of author information is available at the end of the article

Zhang *et al. BMC Pulmonary Medicine*   (2024) 24:357

Page 2 of 11

Background

Intracerebral haemorrhage (ICH) ranks among the top causes of stroke-related morbidity and mortality globally, comprising 15% of all stroke cases [1–3]. Stroke-associated pneumonia (SAP), a common complication, worsens stroke outcomes, extending hospital stays, escalating healthcare costs, and increasing long-term mortality, particularly in affected patients [4–12].

Past research has identified several risk factors linked to SAP onset, including advanced age [8, 9, 13–16], male gender [8, 13, 15, 17], hypertension [18], heart failure [16, 17, 19], severe stroke [8, 13, 16, 20], dysphagia [14–16, 18, 21], elevated neutrophil-to-lymphocyte ratio [22], and high admission blood glucose levels [16, 23]. Several risk scores have been established based on the routinely collected parameters. For instance, the risk between ISAN scoring layered ischemia and hemorrhagic stroke, while A^2DS^2 and AIS-APS scores specially used for ischemic stroke depend on the clinical characteristics of easy use [1, 7, 15, 16]. However, these studies lack sufficient radiology and targeted prediction factors.

Cerebral and brain-brain small vascular disease (CSVD) is a chronic deductive cerebral vascular condition, which affects the global function and structure of the brain [24–28]. Dysfunction of endothelial cells, leading to blood–brain barrier dysfunction, disruption of blood flow homeostasis, and abnormal inflammatory responses, have been recognized as both the initial driver of CSVD and factor affecting systematic vascular inflammation, including pulmonary microvascular inflammation [29, 30]. Previous studies have shown that after ICH, there is positive correlation between CSVD burden, increased haematoma, and poor results [31–34]. A survey conducted by YY (2022) shows that combining CSVD's neural imaging features with A^2DS^2 scores is a promising method for predicting SAP and bad ending of patients with acute cerebral infarction [35]. However, the association between the CSVD burden and the occurrence of SAP after ICH remains to be explored.

Therefore, in this study, we aimed to (1) evaluate whether CSVD burden independently contributes to SAP development after first ICH, (2) implement an image scoring system to quantify the CSVD burden, and (3) develop and validate a radiology-clinical model for predicting SAP risk.

Methods

Patients and follow-up

All 1278 patients consecutively enrolled in this retrospective study were diagnosed with ICH. The diagnosis was confirmed using computed tomography (CT) at our institution. They were admitted to a tertiary hospital between 1 January 2010 and 31 December 2019. The exclusion criteria included: (1) patients diagnosed with primary intraventricular haemorrhage and/or those with multiple ICH foci resulting in difficulties in calculating haematoma volume; (2) patients with a history of stroke, possibly with complete or partial paralysis; (3) patients who did not undergo CT within 72 h post-stroke or those who did not have CT scans with required image quality for calculating haematoma volume; and (4) patients who were lost to follow-up or declined to participate in the study (see Fig. 1). Participant deaths were recorded using an electronic medical system and were supplemented by telephone interviews.

Clinicodemographic variables

Data on age, sex, current smoking status, and alcohol consumption were also collected. The coexisting diseases included hypertension, diabetes mellitus, ischaemic heart disease, atrial fibrillation, hyperlipidaemia, and hyperuricaemia. Additional clinical characteristics such as neutrophil-to-lymphocyte ratio, blood pressure, and National Institutes of Health Stroke Scale (NIHSS) score were also recorded upon admission. The body mass index (BMI) was also recorded. Other potential predictors, including post-stroke vomiting, dysphagia (evaluated as dysphagia or requiring dysphagia rehabilitation training), transpharyngeal intubation (orogastric, nasogastric, and endotracheal tubes), and post-stroke pump proton inhibitor (PPI) usage, were confirmed through the electronic medical system.

Radiological variables

The volume and location of the ICH were obtained and verified using CT within 72-h of stroke onset. Standardised window widths and levels were applied to the CT images to distinguish the haematomas from the brain tissue. The haematoma volume was measured using a manually outlined haematoma profile on each slice of non-enhanced CT (3D Slicer software version 4.10.2). The haematoma contours of each patient were delineated independently by two radiologists blinded to the clinical data (RL and RW). A senior radiologist (YY) was consulted to reach consensus when the contours differed. The location of the ICH was categorised as follows: cortical involvement (defined as any haemorrhage involving the cortex), deep involvement (including the basal ganglia and thalamus), infratentorial involvement (including the brainstem and cerebellum), and multilobar involvement (defined as a haematoma involving two or more lobes). CSVD burden included white matter lesions and cortical/central brain atrophy. The severity of the white matter lesions was assessed using the sum of the anterior/posterior white matter CT scores (0 none, 1–2, mild, 3–4 is severe) [36]. The degree of brain atrophy was measured
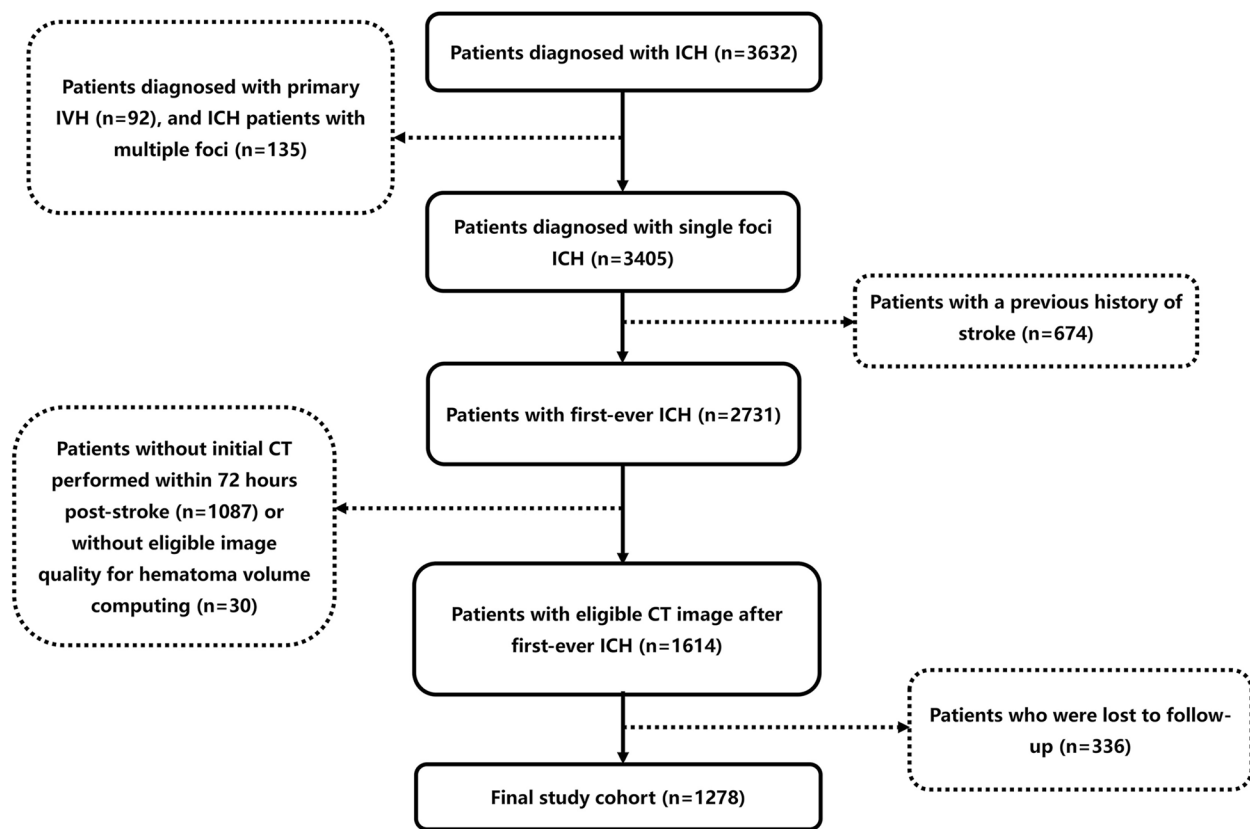
Fig. 1 Patient recruitment flowchart. ICH, intracerebral hemorrhage; IVH, intraventricular hemorrhage; CT, computed tomography

using the intercaudate distance to inner table width ratio (CC/IT) and temporal horn to choroid fissure distance [37, 38]. The total CSVD burden score was determined by collectively considering the score of CC/IT ($\geq$ cutoff value 0.15, 1 point), the temporal horn to choroid fissure distance ($\geq$ cutoff value 0.46, 1 point), and white matter lesion CT score (mild 1 point, severe 2 points).

Primary and secondary outcomes

ICH-associated pneumonia (ICH-SAP) was determined according to the SAP consensus (defined as the spectrum of lower respiratory tract infections within the first seven days after stroke onset) [39]. To avoid false positive outcome, SAP in the study was also verified by the following criteria: (1) absence of infection within two weeks before stroke onset; (2) diagnosis of pneumonia based on a combination of clinical presentation (fever, cough, etc.), positive laboratory findings (white blood cell count $\geq 11 * 10^9$/L, neutrophil count $\geq 7.5 * 10^9$/L etc.), and positive chest CT findings; (3) initiation of antibiotic therapy after pneumonia diagnosis. To compare the efficacies of ICH-SAP prediction between our model and the ISAN model [7], the ICH-SAP probability based on the ISAN model was calculated and recorded. Other

secondary outcomes included hospitalisation duration, modified Rankin scale (mRS) score at discharge, and all-cause mortality within 30 or 90 days after discharge.

Statistical analysis

In the univariate analysis, independent t-tests (for variables with a normal distribution) or Mann–Whitney U tests (for variables with a non-normal distribution) were used to compare continuous variables, whereas the chi-square test or Fisher's exact test was used for categorical variables between the SAP and non-SAP groups. The optimal cutoff value was selected by maximising the Youden index. Correlation analysis was performed among variables with a two-sided *p*-value < 0.1 and depicted using a correlation analysis heatmap (see Fig. 2).

Model development

The training and validation groups were created using a fivefold cross-validation method with 400 repetitions, ensuring that the sampling was proportional to the original dataset. Logistic regression (LR) was employed alongside two machine learning (ML) methods, namely, support vector machine (SVM) and random forest (RF), to establish the SAP prediction model. A grid search
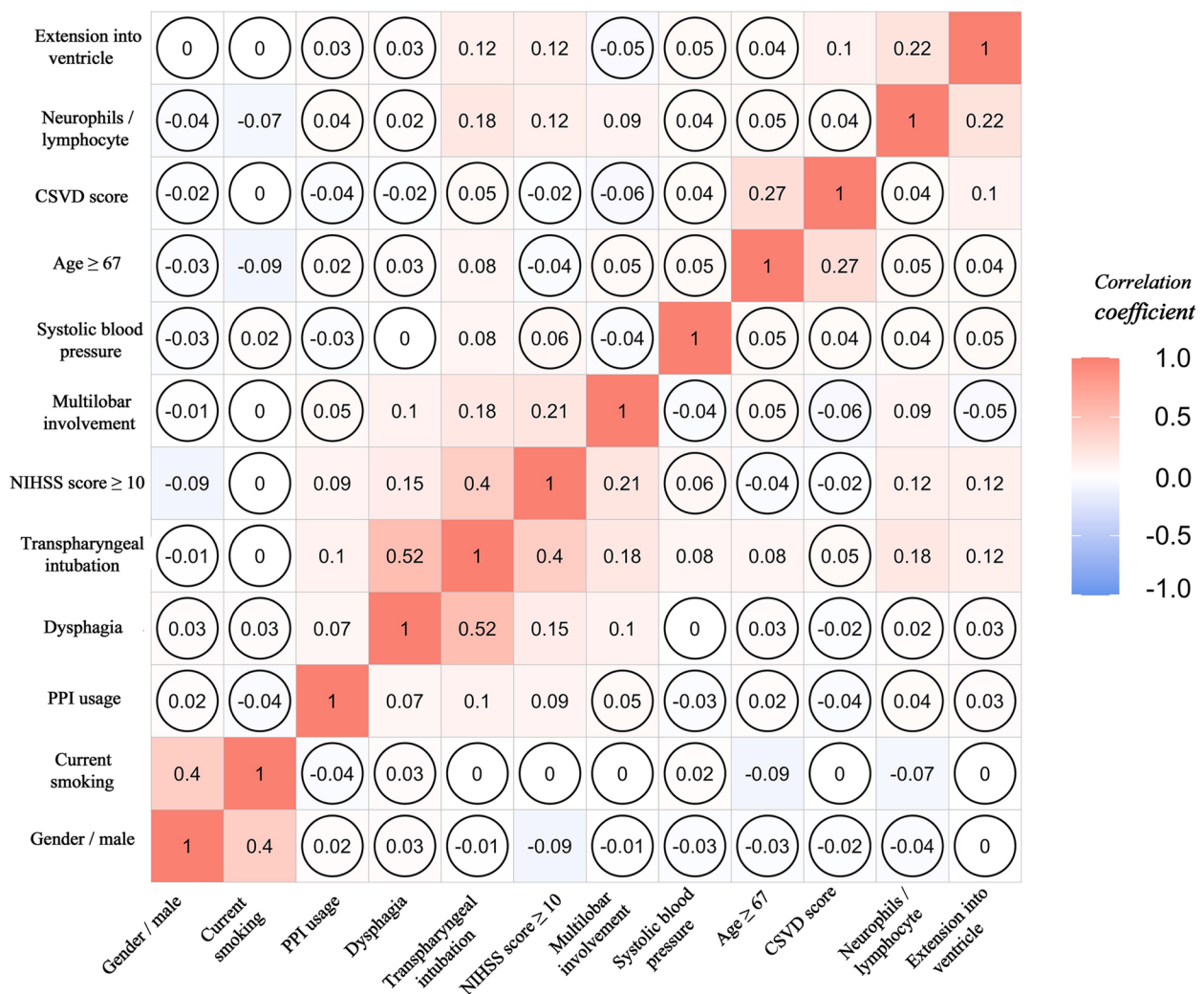
Fig. 2 Heatmap for correlation analysis within variables significant in univariable analysis. Different colors filling in the square represented different correlation coefficients. The square with a circle inside it represented a non-significant correlation between two corresponding variables. Numbers in the square represented the correlation coefficient. CSVD, cerebral small vessel disease; NIHSS, National Institutes of Health Stroke Scale; PPI, proton pump inhibitor

method was used to identify the best gamma and cost coefficients in the SVM model (Figure S1). The 'best_mtry', 'best_ntree', 'best_maxnodes' and 'best_nodesize' were determined to establish a reasonable RF model with hyperparameter optimisation (see Figure S2). Moreover, in developing the LR model, multivariate analysis was conducted using a forward stepwise regression approach with the maximum likelihood ratio method to assess the independent predictors of SAP.

Model evaluation

The area under the curve (AUC) and its corresponding 95% confidence interval (95% CI) were calculated and are shown in Table 2. The probability of SAP risk obtained

from the ISAN scoring system was used to calculate the corresponding AUC value. Furthermore, based on the results of the LR model, a forest plot was generated to illustrate the odds ratios (OR) and their 95% confidence intervals (CI) for independent predictors of SAP, and a nomogram with a gradient colour style was created to calculate individualised ICH-SAP risk. Additionally, a calibration curve with 80–95% CI was generated for both the training and testing groups.

Outcome analysis

Differences in hospitalisation duration, mRS ≥ 3 at discharge, and all-cause mortality of 30 days or 90 days after discharge between the SAP and non-SAP groups

Zhang *et al. BMC Pulmonary Medicine*   (2024) 24:357

Page 5 of 11

were compared. All statistical analyses were conducted using the R software (Version 4.1.2). The packages utilised in the analysis are described in the Supplementary Methods.

Result

Baseline patient characteristics

Among the 1278 patients finally enrolled in the study, the SAP occurrence rate after the first ICH was 22.0% ($n = 281/1278$). Thirty one point seven percent ($n = 405/1278$) of patients aged ≥ 67 years(The cutoff value calculated based on the Youden index.). Seventy percent of patients the total cohort were male. The other characteristics of the cohort are shown in Table S1.

Results of univariable analysis

Age ≥ 67 years, dysphagia, transpharyngeal intubation, neutrophils-to-lymphocyte ratio, admission NIHSS score ≥ 10, multilobar involvement, haemorrhage extension into the ventricle, and high CSVD burden score were statistically different between patients with and without SAP ($p < 0.001$). Patients with SAP were more likely to have a higher systolic blood pressure on admission (164.83 vs. 160.15, $p = 0.044$). Additionally, the proportions of male patients and current smokers were higher in the SAP group (74.4% vs. 68.7%, $p = 0.067$; and 33.8% vs. 28.1%, $p = 0.063$, respectively), although the difference was not statistically significant (Table 1). All variables mentioned above ($p < 0.1$) in the univariable analysis were further included in the correlation and model establishment analyses.

Development, validation, and evaluation of different model

The variables used to build the SVM and RF models were not highly correlated (Fig. 2). The classified SVM model was constructed using a radial kernel with the best gamma (0.1) and cost (1) parameters. The AUC for the training and testing groups in the SVM model were 0.812 and 0.674, respectively (Table 2). The RF model was established with parameters 'mtry', 'ntree', maxnodes', and 'nodesize' equal to 10, 580, 7, and 3, respectively (see Figure S2). The AUC of the values for the training and testing groups were 0.684 and 0.652, respectively (Table 2). The final LR model indicated that total CSVD burden score (OR, 1.29; 95% CI, 1.03–1.54; $p = 0.004$), haematoma extension into ventricle (OR, 2.28; 95% CI, 1.87–3.31; $p < 0.001$), haematoma with multilobar involvement (OR, 2.14; 95% CI, 1.44–3.18; $p < 0.001$), transpharyngeal intubation operation (OR, 3.89; 95% CI, 2.7–5.62; $p < 0.001$), admission NIHSS score ≥ 10 (OR, 2.06; 95% CI, 1.42–3.01; $p < 0.001$), male sex (OR, 1.69; 95% CI, 1.16–2.52; $p = 0.007$), and age ≥ 67 (OR, 2.24; 95%

CI, 1.56–3.22; $p < 0.001$) were independent predictors of SAP after ICH (see Fig. 4B). The AUC for the training and testing groups in the LR model were 0.796 and 0.746, respectively (Table 2). Figure 3 shows a comparison of the AUC values for the cohort groups of the three models. Figure 4A indicates that the LR model did not overestimate or underestimate the SAP risk at 80% CI and 95% CI in either the training or testing groups. A practical nomogram for predicting the probability of developing SAP was created (Fig. 4C). The AUC of the ISAN model based on the ISAN SAP risk probability was 0.688. The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1 score, and accuracy of each model are presented in Table 2.

Results of short and long-term outcomes analysis

Patients with SAP had a significantly longer duration of hospitalisation than those without SAP ($p < 0.001$). Moreover, patients with SAP exhibited a higher frequency of poor outcomes, defined as mRS ≥ 3 at discharge, compared to that of patients without SAP (81.5% vs. 51.2%, $p < 0.001$, Table 1). Additionally, the 30-day or 90-day discharge mortality rates were higher in the SAP group than in the non-SAP group (0.7% vs. 0.3%, $p = 0.665$, and 2.1% vs. 0.8%, $p = 0.116$, respectively). The detailed distribution of functional outcomes at the time of discharge in patients with and without SAP is presented in Fig. 5.

Discussion

Our study indicated that the multivariable logistic regression model achieved superior SAP prediction ability compared to both the ML models and the existing ISAN model. The following risk factors: age ≥ 67 years, male sex, transpharyngeal intubation, NHSS score ≥ 10 on admission, haematoma involving multiple lobes or extending into ventricle, and high total CSVD burden score were identified as independent predictors of SAP after first ICH.

Advanced age, male gender, significant neurofunctional deficits, and extension of the haematoma into the ventricle were predictive factors for SAP occurrence, aligning with previously reported results [8, 14–17]. The presence of multilobar involvement in the haematoma, indicating a larger haematoma volume, strongly correlated with SAP development in our study. This finding supports previous research, which showed that patients with a 1 ml increase in hematoma volume were 1.02 times more likely to develop SAP compared to those without it [1]. Our study demonstrated that transpharyngeal intubation, an easily judged and accessible indicator in clinical practice, independently increased the risk of SAP, whereas dysphagia did not reach statistical significance in the final model after controlling for confounding factors.

Zhang *et al. BMC Pulmonary Medicine*   (2024) 24:357

Page 6 of 11

Table 1 Predictors of SAP: univariable analysis

Variables	SAP(*n* = 281)	Non-SAP(*n* = 997)	*P*-value
Demographics			
Age ≥ 67	45.9% (*n* = 129/281)	27.7% (276/997)	0.000
Gender (male/total)	74.4% (*n* = 209/281)	68.7% (*n* = 685/997)	0.067
Lifestyle-related variables			
Current drinking	35.9% (*n = 101/281*)	34.9% (*n* = 348/997)	0.747
Current smoking	33.8% (*n* = 95/281)	28.1% (*n* = 280/997)	0.063
Clinical variables			
post-stroke vomiting	27.9% (*n* = 78/280)	24.6% (*n* = 245/996)	0.268
Dysphagia	17.5% (*n* = 48/275)	7.4% (*n* = 73/980)	0.000
Transpharyngeal intubation	59% (*n* = 164/278)	19.7% (*n* = 194/984)	0.000
Post-stroke PPI usage	46.7% (*n* = 128/281)	37.1% (*n* = 364/997)	0.004
Coexisting disease			
Hypertension	90.7% (*n* = 255/281)	90.3% (*n* = 900/997)	0.811
Diabetes mellitus	18.9% (*n* = 53/281)	19.8% (*n* = 197/997)	0.737
Ischemic heart disease	5.3% (*n* = 15/281)	5.6% (*n* = 56/997)	0.857
Atrial fibrillation	3.2% (*n* = 9/281)	1.8% (*n* = 18/997)	0.150
Hyperlipidemia	22.8% (*n* = 64/281)	26.7% (*n* = 266/997)	0.187
Hyperuricemia	5.0% (*n* = 14/281)	4.5% (*n* = 45/997)	0.741
Laboratory index			
BMI index, n (%)			0.297
< 18.5	9.9% (*n* = 8/81)	5.5% (*n* = 23/416)	
18.5–24	40.7% (*n* = 33/81)	45.7% (*n* = 190/416)	
≥ 24	49.4% (*n* = 40/81)	48.8% (*n* = 203/416)	
Neutrophils/lymphocyte, median (IQR)	6.09 (10.14, 3.9)	4.25 (2.67, 7.57)	0.000
Admission systolic blood pressure, mean ± SD	164.83 ± 25.85	160.15 ± 24.12	0.044
Admission diastolic blood pressure, median (IQR)	91 (80, 100)	90 (80, 100)	0.812
Discharge mRS ≥ 3	81.5% (*n* = 229/281)	51.2% (*n* = 510/997)	0.000
Hospitalization duration (day), median (IQR)	17 (13, 22)	14 (11, 18)	0.000
Death within 30-day discharge	0.7%(*n* = 2/281)	0.3%(*n* = 3/997)	0.665
Death within 90-day discharge	2.1% (*n* = 6/281)	0.8% (*n* = 8/997)	0.116
NIHSS score ≥ 10	59.1% (*n* = 166/281)	29.4% (*n* = 293/997)	0.000
Radiological variables			
Haematoma volume (ml), median (IQR)	7.83(2.94,15.27)	7.62(3.22,14.22)	0.789
Cortical involvement	35.2% (*n* = 99/281)	32.8% (*n* = 327/997)	0.445
Deep involvement	78.3% (*n* = 220/281)	75.4% (*n* = 752/997)	0.320
Infratentorial involvement	8.9% (*n* = 25/281)	12.0% (*n* = 120/997)	0.166
Multilobar involvement	34.2% (*n* = 96/281)	15.5% (*n* = 155/997)	0.000
Extension into ventricle	35.6% (*n* = 100/281)	19.4% (*n* = 193/997)	0.000
Total CSVD burden score	1 (1, 2)	1 (1, 2)	0.000
Follow-up duration (day), median (IQR)	1624 (1065, 2544)	1945 (1274, 2715)	0.000

SD Standard Deviation, *IQR* Interquartile Range, *PPI* proton pump inhibitor, *BMI* Body Mass Index, *mRS* modified Rankin scale, *NIHSS* National Institutes of Health Stroke Scale, *CSVD*, cerebral small vessel disease; Continuous variables were expressed as mean ± standard deviation or median (IQR). Categorical variables were expressed as counts and percentages

This may be attributed to the fact that manipulating the transoropharyngeal route increases the risk of accidental aspiration.

CSVD has been extensively studied, whereas the association between CSVD and SAP has rarely been discussed [24, 25, 27, 28, 31, 32]. A previous study hypothesised that older patients with a history of severe cerebral atrophy were likely to develop lower respiratory tract infections [40]. Nam et al. found that patients with severe leukoaraiosis were likely to develop SAP

Zhang *et al. BMC Pulmonary Medicine*   (2024) 24:357

Page 7 of 11

Table 2 Multiple model evaluation indexes for different model

| Model evaluation index	Support vector machine		Random forest		Logistic regression		ISAN
	Training group	Testing group	Training group	Testing group	Training group	Testing group	Total group
AUC	0.812	0.674	0.684	0.652	0.796	0.746	0.688
95% CI	0.783–0.842	0.602–0.747	0.650–0.718	0.580–0.724	0.762–0.830	0.677–0.815	——
Sensitivity	0.793	0.395	0.471	0.463	0.387	0.395	0.149
Specificity	0.832	0.874	0.897	0.841	0.940	0.974	0.904
Positive predictive value	0.573	0.593	0.566	0.439	0.644	0.593	0.304
Negative predictive value	0.934	0.756	0.856	0.854	0.844	0.756	0.790
F1 score	0.665	0.474	0.514	0.450	0.483	0.474	0.200
Accuracy	0.823	0.722	0.803	0.761	0.818	0.722	0.527
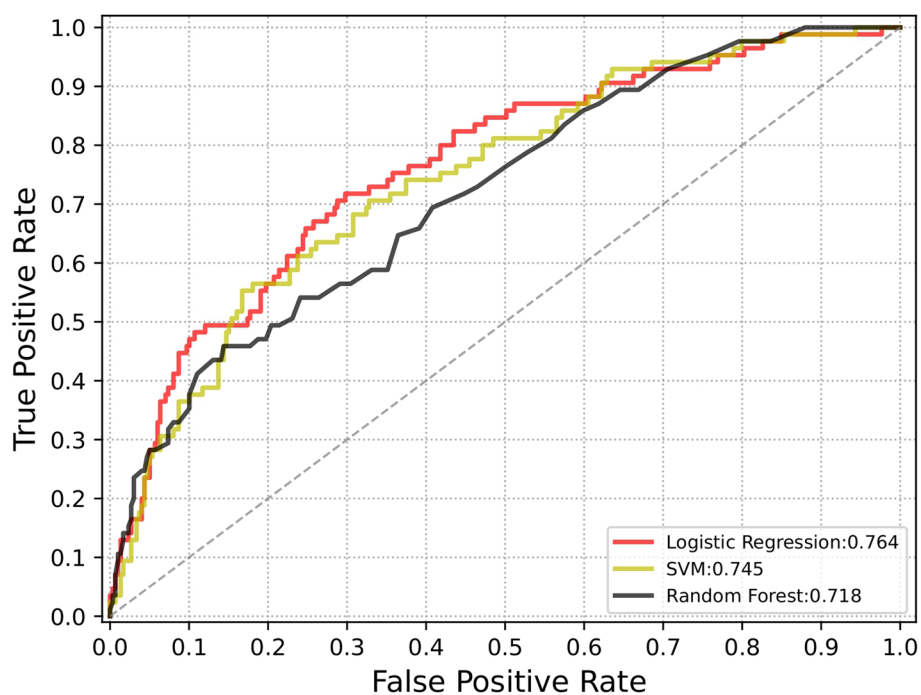
Fig. 3 ROC curve of different models in the cohort group. AUC, area under the curve

(See figure on next page.)

Fig. 4 Information of multivariable logistic regression model. **A** displayed the calibration curve belt of the logistic regression model in the training and testing groups, respectively. The gray belt region represented a well-calibrated situation of both training and testing groups within an 80% to 95% confidence interval range. **B** was a forest plot displaying the results of the multivariable analysis. **C** was a nomogram with gradient colors representing different total scores and SAP risk. CSVD, cerebral small vessel disease; NIHSS, National Institutes of Health Stroke Scale; OR, odds ratio; 95% CI, 95% confidence interval; SAP, stroke-associated pneumonia

after acute ischaemic stroke, which may be accounted for by studying the correlation between the degree of leukoaraiosis and the level of impaired brain structure and function (such as reduction of the cough reflex) [41]. Another study indicated that brain atrophy may contribute to the occurrence of SAP through indirect influence on swallowing reflex and that dilated perivascular spaces interpreted as blood–brain barrier dysfunction are associated with poor outcomes [35]. The degree of CSVD burden, which represents the level of endothelial cell dysfunction in microvascular inflammation, results in a highly inflammatory microenvironment in the
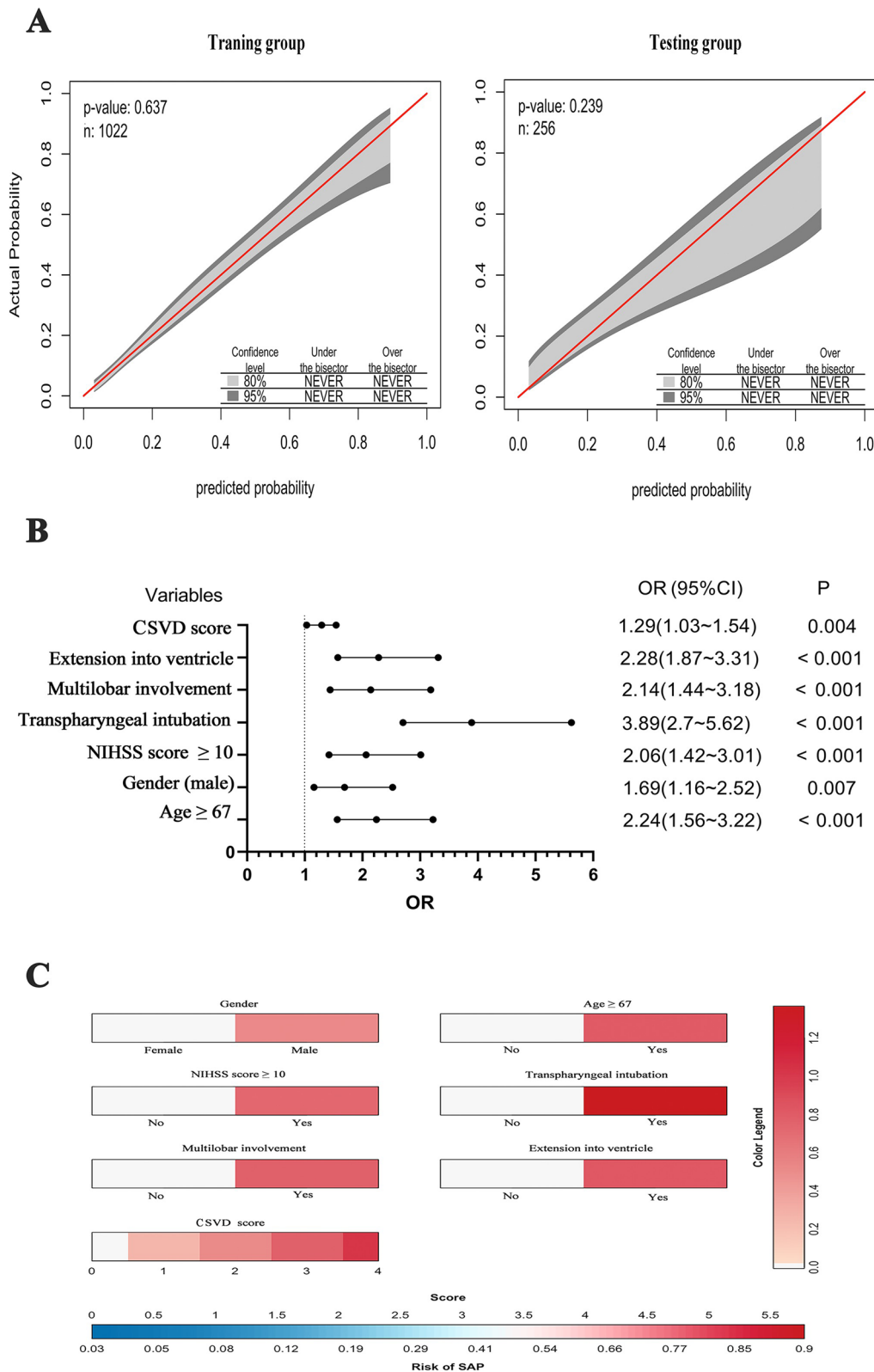
A

B

Variables	OR (95%CI)	P
CSVD score	1.29(1.03~1.54)	0.004
Extension into ventricle	2.28(1.87~3.31)	< 0.001
Multilobar involvement	2.14(1.44~3.18)	< 0.001
Transpharyngeal intubation	3.89(2.7~5.62)	< 0.001
NIHSS score ≥ 10	2.06(1.42~3.01)	< 0.001
Gender (male)	1.69(1.16~2.52)	0.007
Age ≥ 67	2.24(1.56~3.22)	< 0.001

C

Fig. 4 (See legend on previous page.)

Zhang *et al. BMC Pulmonary Medicine*   (2024) 24:357
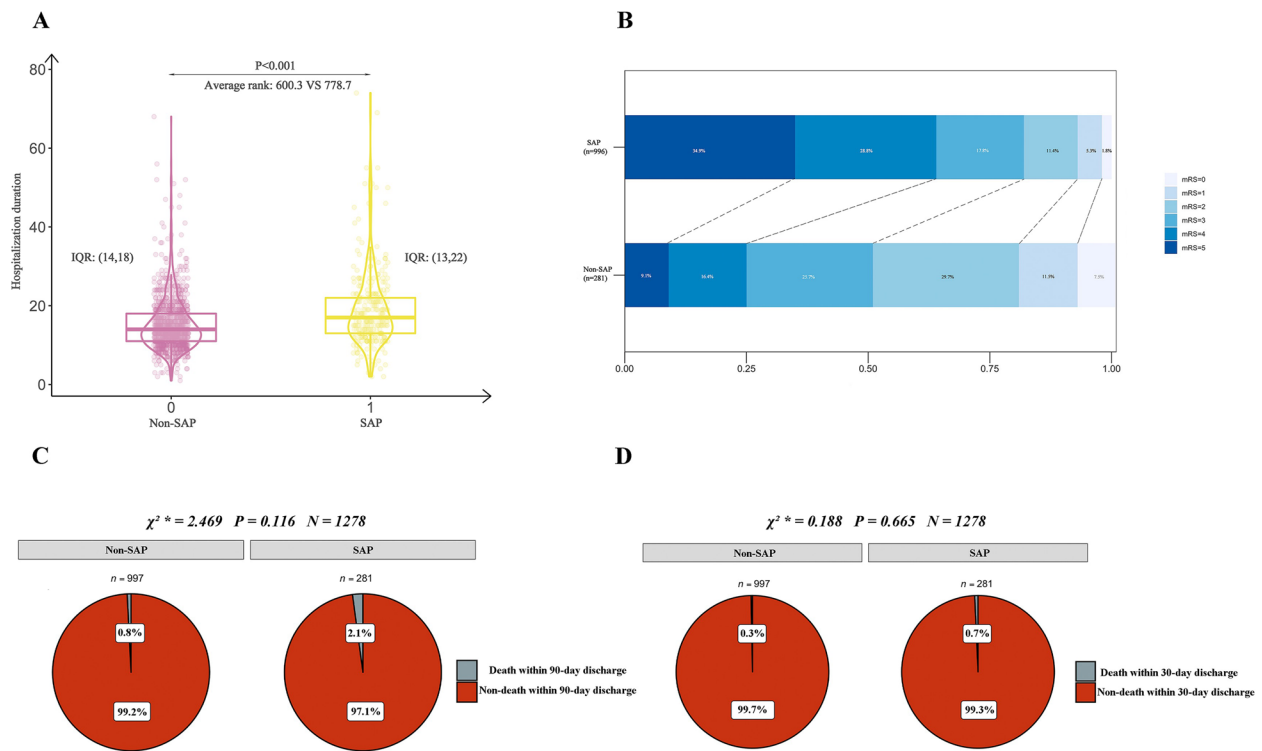
Page 9 of 11

Fig. 5 Short or long-term prognosis difference between SAP and non-SAP patients. **A** showed the difference in hospitalization duration between patients with and without SAP using Mann–Whitney u test. **B** displayed the proportion of the SAP or non-SAP patient's mRS at discharge. **C** and (**D**) were comparison results of death within 90-day or death within 30-day discharge difference between SAP and non-SAP group. SAP, stroke-associated pneumonia; IQR, interquartile range; mRS, modified Rankin scale

pulmonary glands [29, 42, 43]. In this study, CSVD was found to be independently associated with the development of SAP after ICH, after controlling for haematoma volume confounders. This relationship may be explained by several hypotheses. First, in line with existing assumptions, CSVD is always accompanied by impairment of neurofunction, indirectly affecting SAP occurrence to some extent. Second, CSVD, deemed as endothelial cell activation and dysfunction, accelerates the formation of the pulmonary inflammatory microenvironment, contributing to the development of SAP. Therefore, the possible cumulative effect of the CSVD burden on patients likely predicts ICH-SAP risk and helps in the clinical rationalisation of medical resources, designation of care plans, and implementation of targeted preventive strategies.

The ISAN model has demonstrated high clinical applicability in patients with stroke, exhibiting the ability to accurately distinguish between SAP and non-SAP patients [7]. However, this study was primarily conducted in the ischaemic stroke group, with only 8% of the patients having ICH. In addition, this model did not include radiological elements, and the AUC value of the model in our study was only 0.688. The logistic regression

model, incorporating both clinical and radiological risk factors; showed good differentiation ability for SAP and was well-calibrated in both our study's training and testing groups. The high negative predictive value compared with the positive predictive value in our study, coupled with a well-calibrated belt indicating no over-or underestimation of SAP, suggesting that future prognostic models might benefit from attempts to achieve a more balanced distribution of differentiation utility between higher and lower values. Although our study did not include mRS scores at 3-month discharge, we observed that patients who experienced ICH-SAP tended to have longer hospitalisation durations and higher all-cause mortality during the longitudinal follow-up period.

Our study has several limitations. First, given the retrospective nature of the study, SAP may have been underestimated, despite the strict diagnostic criteria used. Additionally, patients with SAP exhibited a tendency towards poor outcomes in this study, although longitudinal mRS score follow-up was not conducted. Therefore, large-scale studies with standardised long-term observations are required. Second, only patients with first-ever ICH were included in this study, potentially leading to a selection bias, as pre-stroke dependence has been shown

Zhang *et al. BMC Pulmonary Medicine*　　(2024) 24:357

Page 10 of 11

to be a risk factor for SAP. Furthermore, the radiology-clinical model was not validated using an external dataset. Third, although the underlying mechanism by which CSVD influences ICH-SAP remains unclear, the current study demonstrates that CSVD is a predictive factor for ICH-SAP, providing guidance for future research.

In conclusion, we developed a novel radiological-clinical model to predict SAP after first ICH. Future studies are required to further explore and confirm the relationship between CSVD and SAP.

Conclusion

This study indicated that CSVD burden increased the risk of SAP after first ICH, independent of ICH volume. The novel radiology-clinical SAP model, incorporating the CSVD burden, was optimally established by logistic regression, surpassing two other machine learning models and the ISAN model in terms of performance. Patients developing SAP tended to have a poor prognosis in short- and long-term follow-ups. A nomogram with a gradient colour style was created based on a well-calibrated model to aid in the early identification of patients at a high risk of ICH-SAP in clinical practice. This tool assists in the selection of appropriate treatment and care strategies, thereby enhancing outcomes and potentially preventing SAP-related complications.

Abbreviations

ICH	Intracerebral haemorrhage
SAP	Stroke-associated pneumonia
CSVD	Cerebral small vessel disease
CT	Computed tomography
IVH	Intraventricular haemorrhage
NIHSS	National Institutes of Health Stroke Scale
BMI	Body mass index
PPI	Proton pump inhibitor
ICH-SAP	ICH associated-pneumonia
mRS	Modified Rankin scale
SVM	Support vector machine
RF	Random forest
LR	Logistic regression
AUC	Area under the curve
OR	Odds ratio
95% CI	95% Confidence interval
PPV	Positive predictive value
NPV	Negative predictive value

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12890-024-03160-0.

> Supplementary Material 1.

Authors' contributions

WZ and RL analyzed the data, framed the data, and drafted the manuscript. RL designed the study, analyzed the data, and drafted and revised this manuscript. XW and YY administrated and supervised the study and finalized the manuscript. All authors contributed to data collection. All authors reviewed and approved this manuscript.

Availability of data and material
The data supporting this study's findings are available at reasonable request from the corresponding author.

Declarations

Ethics approval and consent to participate
This study was approved by the local ethics committee of The Affiliated First Hospital of Wenzhou Medical University (No. 2021HZSY0057) (ChiCTR2300078239). Due to the retrospective nature of this study, informed consent was waived.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Radiology, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, China. [2]Wenzhou Medical University, Wenzhou, Zhejiang, China. [3]Department of Neurology, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, China. [4]Key Laboratory of Alzheimer's Disease of Zhejiang Province, Institute of Aging, Wenzhou Medical University, Wenzhou, Zhejiang, China.

References
1. Ji R, Shen H, Pan Y, et al. Risk score to predict hospital-acquired pneumonia after spontaneous intracerebral hemorrhage. Stroke. 2014;45(9):2620–8.
2. Schrag M, Kirshner H. Management of intracerebral hemorrhage: JACC focus seminar. J Am Coll Cardiol. 2020;75(15):1819–31.
3. O'Carroll CB, Brown BL, Freeman WD. Intracerebral hemorrhage: a common yet disproportionately deadly stroke subtype. Mayo Clin Proc. 2021;96(6):1639–54.
4. Ingeman A, Andersen G, Hundborg HH, et al. In-hospital medical complications, length of stay, and mortality among stroke unit patients. Stroke. 2011;42(11):3214–8.
5. Katzan IL, Dawson NV, Thomas CL, et al. The cost of pneumonia after acute stroke. Neurology. 2007;68(22):1938–43.
6. Zhao J, Li LQ, Zhen NX, et al. Microbiology and outcomes of institutionalized patients with stroke-associated pneumonia: an observational cohort study. Front Microbiol. 2021;12:720051.
7. Smith CJ, Bray BD, Hoffman A, et al. Can a novel clinical risk score improve pneumonia prediction in acute stroke care? A UK multicenter cohort study. J Am Heart Assoc. 2015;4(1):e001307.
8. Kammersgaard LP, Jørgensen HS, Reith J, et al. Early infection and prognosis after acute stroke: the Copenhagen Stroke Study. J Stroke Cerebrovasc Dis. 2001;10(5):217–21.
9. Kwan J, Hand P. Infection after acute stroke is associated with poor short-term outcome. Acta Neurol Scand. 2007;115(5):331–8.
10. Li J, Zhang P, Wu S, et al. Stroke-related complications in large hemisphere infarction: incidence and influence on unfavorable outcome. Ther Adv Neurol Disord. 2019;12:1756286419873264.
11. Westendorp WF, Dames C, Nederkoorn PJ, Meisel A. Immunodepression, Infections, and Functional Outcome in Ischemic Stroke. Stroke. 2022;53(5):1438–48.

Zhang *et al. BMC Pulmonary Medicine*   (2024) 24:357

Page 11 of 11

12. Hotter B, Hoffmann S, Ulm L, et al. Inflammatory and stress markers predicting pneumonia, outcome, and etiology in patients with stroke: Biomarkers for predicting pneumonia, functional outcome, and death after stroke. Neurol Neuroimmunol Neuroinflamm. 2020;7(3):e692.

13. Aslanyan S, Weir CJ, Diener HC, et al. Pneumonia and urinary tract infection after acute ischaemic stroke: a tertiary analysis of the GAIN International trial. Eur J Neurol. 2004;11(1):49–53.

14. Chumbler NR, Williams LS, Wells CK, et al. Derivation and validation of a clinical system for predicting pneumonia in acute stroke. Neuroepidemiology. 2010;34(4):193–9.

15. Hoffmann S, Malzahn U, Harms H, et al. Development of a clinical score (A2DS2) to predict pneumonia in acute ischemic stroke. Stroke. 2012;43(10):2617–23.

16. Ji R, Shen H, Pan Y, et al. Novel risk score to predict pneumonia after acute ischemic stroke. Stroke. 2013;44(5):1303–9.

17. Katzan IL, Cebul RD, Husak SH, et al. The effect of pneumonia on mortality among patients hospitalized for acute stroke. Neurology. 2003;60(4):620–5.

18. Ding R, Logemann JA. Pneumonia in stroke patients: a retrospective study. Dysphagia. 2000;15(2):51–7.

19. Emsley HC, Hopkins SJ. Acute ischaemic stroke and infection: recent and emerging concepts. Lancet Neurol. 2008;7(4):341–53.

20. Vargas M, Horcajada JP, Obach V, et al. Clinical consequences of infection in patients with acute stroke: is it prime time for further antibiotic trials? Stroke. 2006;37(2):461–5.

21. Bray BD, Smith CJ, Cloud GC, et al. The association between delays in screening for and assessing dysphagia after acute stroke, and the risk of stroke-associated pneumonia. J Neurol Neurosurg Psychiatry. 2017;88(1):25–30.

22. Nam KW, Kim TJ, Lee JS, et al. High Neutrophil-to-Lymphocyte ratio predicts stroke-associated pneumonia. Stroke. 2018;49(8):1886–92.

23. Tao J, Hu Z, Lou F, et al. Higher stress hyperglycemia ratio is associated with a higher risk of stroke-associated pneumonia. Front Nutr. 2022;9:784114.

24. Wardlaw JM, Smith C, Dichgans M. Small vessel disease: mechanisms and clinical implications. Lancet Neurol. 2019;18(7):684–96.

25. Debette S, Schilling S, Duperron MG, et al. Clinical Significance of Magnetic Resonance Imaging Markers of Vascular Brain Injury: A Systematic Review and Meta-analysis. JAMA Neurol. 2019;76(1):81–94.

26. Pantoni L. Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges. Lancet Neurol. 2010;9(7):689–701.

27. Ter Telgte A, van Leijsen EMC, Wiegertjes K, et al. Cerebral small vessel disease: from a focal to a global perspective. Nat Rev Neurol. 2018;14(7):387–98.

28. Hamilton OKL, Backhouse EV, Janssen E, et al. Cognitive impairment in sporadic cerebral small vessel disease: a systematic review and meta-analysis. Alzheimers Dement. 2021;17(4):665–85.

29. Quick S, Moss J, Rajani RM, et al. A vessel for change: endothelial dysfunction in cerebral small vessel disease. Trends Neurosci. 2021;44(4):289–305.

30. Huertas A, Palange P. Circulating endothelial progenitor cells and chronic pulmonary diseases. Eur Respir J. 2011;37(2):426–31.

31. Uniken Venema SM, Marini S, Lena UK, et al. Impact of cerebral small vessel disease on functional recovery after intracerebral hemorrhage. Stroke. 2019;50(10):2722–8.

32. Xu M, Li B, Zhong D, et al. Cerebral small vessel disease load predicts functional outcome and stroke recurrence after intracerebral hemorrhage: a median follow-up of 5 years. Front Aging Neurosci. 2021;13:628271.

33. Boulouis G, van Etten ES, Charidimou A, et al. Association of Key magnetic resonance imaging markers of cerebral small vessel disease with hematoma volume and expansion in patients with lobar and deep intracerebral hemorrhage. JAMA Neurol. 2016;73(12):1440–7.

34. Chen X, Jin Y, Chen J, et al. Relationship between white matter hyperintensities and hematoma volume in patients with intracerebral hematoma. Aging Dis. 2018;9(6):999–1009.

35. Yu Y, Xia T, Tan Z, et al. A(2)DS(2) score combined with clinical and neuroimaging factors better predicts stroke-associated pneumonia in hyperacute cerebral infarction. Front Neurol. 2022;13:800614.

36. van Swieten JC, Hijdra A, Koudstaal PJ, et al. Grading white matter lesions on CT and MRI: a simple scale. J Neurol Neurosurg Psychiatry. 1990;53(12):1080–3.

37. Ho VB, Chuang HS, Rovira MJ, et al. Juvenile Huntington disease: CT and MR features. AJNR Am J Neuroradiol. 1995;16(7):1405–12.

38. Scheltens P, Launer LJ, Barkhof F, et al. Visual assessment of medial temporal lobe atrophy on magnetic resonance imaging: interobserver reliability. J Neurol. 1995;242(9):557–60.

39. Smith CJ, Kishore AK, Vail A, et al. Diagnosis of stroke-associated pneumonia: recommendations from the pneumonia in stroke consensus group. Stroke. 2015;46(8):2335–40.

40. Okada R, Okada T, Okada A, et al. Severe brain atrophy in the elderly as a risk factor for lower respiratory tract infection. Clin Interv Aging. 2012;7:481–7.

41. Nam KW, Kwon HM, Lim JS, et al. Leukoaraiosis is associated with pneumonia after acute ischemic stroke. BMC Neurol. 2017;17(1):51.

42. Ruhl L, Pink I, Kühne JF, et al. Endothelial dysfunction contributes to severe COVID-19 in combination with dysregulated lymphocyte responses and cytokine networks. Signal Transduct Target Ther. 2021;6(1):418.

43. Declercq M, de Zeeuw P, Conchinha NV, et al. Transcriptomic analysis of CFTR-impaired endothelial cells reveals a pro-inflammatory phenotype. Eur Respir J. 2021;57(4):2000261.

Publisher's Note