Computers in Biology and Medicine 166 (2023) 107577

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

An interpretable ensemble learning model facilitates early risk stratification of ischemic stroke in intensive care unit: Development and external validation of ICU-ISPM

Wei Hu [a,1], Tingting Jin [a,1], Ziqi Pan [b,1], Huimin Xu [a], Lingyan Yu [a], Tingting Chen [a], Wei Zhang [b], Huifang Jiang [a], Wenjun Yang [a], Junjun Xu [a], Feng Zhu [a,b,**], Haibin Dai [a,c,*]

[a] Department of Pharmacy, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, 310009, China
[b] College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, 310058, China
[c] Clinical Pharmacy Research Center, Zhejiang University School of Medicine, Hangzhou, 310009, China

ARTICLE INFO

Keywords:
Critical care
Ischemic stroke
Ensemble learning
APACHE
Mortality

ABSTRACT

Ischemic stroke (IS) is a common and severe condition that requires intensive care unit (ICU) admission, with high mortality and variable prognosis. Accurate and reliable predictive tools that enable early risk stratification can facilitate interventions to improve patient outcomes; however, such tools are currently lacking. In this study, we developed and validated novel ensemble learning models based on soft voting and stacking methods to predict in-hospital mortality from IS in the ICU using two public databases: MIMIC-IV and eICU-CRD. Additionally, we identified the key predictors of mortality and developed a user-friendly online prediction tool for clinical use. The soft voting ensemble model, named ICU-ISPM, achieved an AUROC of 0.861 (95% CI: 0.829–0.892) and 0.844 (95% CI: 0.819–0.869) in the internal and external test cohorts, respectively. It significantly outperformed the APACHE scoring system and was more robust than individual models. ICU-ISPM obtained the highest performance compared to other models in similar studies. Using the SHAP method, the model was interpretable, revealing that GCS score, age, and intubation were the most important predictors of mortality. This model also provided a risk stratification system that can effectively distinguish between low-, medium-, and high-risk patients. Therefore, the ICU-ISPM is an accurate, reliable, interpretable, and clinically applicable tool, which is expected to assist clinicians in stratifying IS patients by the risk of mortality and rationally allocating medical resources. Based on ICU-ISPM, an online risk prediction tool was further developed, which was freely available at: http://ispm.idrblab.cn/.

1. Introduction

Globally, stroke is the second leading cause of death, accounting for nearly 6 million deaths annually, or 11.6% of all deaths [1]. Among patients admitted to ICUs, this rate is even higher [2]. As a limited medical resource, ICUs are indispensable for sustaining patients' lives [3,4]. However, ICUs are also a labor and financial burden, and the efficient utilization of ICU resources is crucial [5,6]. Ischemic stroke (IS) is a common condition entering the ICU, accounting for more than 87% of all strokes, with widely varying prognoses [7]. Therefore, clinical tools that provide accurate and reliable prognostic predictions for IS patients are important not only for early clinical decision-making, but also to facilitate the optimal allocation of available medical resources.

Currently, the National Institutes of Health Stroke Scale (NIHSS) is a widely used clinical tool to assess neurological deficits in stroke patients [8]. The NIHSS score can indirectly predict stroke mortality, but some items lack consistency or reliability due to subjective judgment and incompetence of assessors. Thus, several studies have developed prediction models attempting to identify early adverse progression of acute stroke [9–12]. In these studies, however, some limitations exist: 1) the simplicity of the model could lead to suboptimal performance and robustness; 2) a lack of reasonable risk stratification thresholds, which

* Corresponding author. Department of Pharmacy, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, 310009, China.
** Corresponding author. Department of Pharmacy, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, 310009, China.
E-mail addresses: zhufeng@zju.edu.cn (F. Zhu), haibindai@zju.edu.cn (H. Dai).
[1] These authors contributed equally to this work as co-first authors.

https://doi.org/10.1016/j.compbiomed.2023.107577
Received 7 August 2023; Received in revised form 13 September 2023; Accepted 11 October 2023
Available online 11 October 2023

affect clinical decision making; 3) no online prediction tool had been developed, making the model inapplicable in clinical practice.

Ensemble modeling is widely recognized as a pivotal area of research in machine learning (ML), owing to its ability to enhance prediction accuracy and robustness. Stacking [13] and voting [14] are two popular ensemble learning algorithms that can combine numerous homogeneous or heterogeneous regular base models. In this study, we developed ensemble learning models to predict in-hospital mortality from IS in the ICU and identified the associated predictors. We also contrasted the performance between our models and the widely used Acute Physiology and Chronic Health Assessment (APACHE) scoring system [15,16]. To improve the clinical acceptance and translational value of the model, interpretable machine learning methods were used to gain insight into the predictions or outcomes, and a user-friendly online prediction tool was developed.

2. Methods

2.1. Data source and outcome

The training and internal test cohort data were extracted from the Medical Information Mart for Intensive Care IV (MIMIC-IV) [17]. In addition, an independent external test cohort was derived from the eICU Collaborative Research database (eICU-CRD) [18]. Our study used the MIMIC-IV database version 2.1 (https://physionet.org/content/mimiciv/2.1/) and eICU-CRD database version 2.0 (https://physionet.org/content/eicu-crd/2.0/) publicly available through the PhysioNet website (certification ID: 11410188). The outcome of this study was in-hospital mortality. The MIMIC-IV and eICU-CRD databases have received ethical approval from the Institutional Review Boards (IRBs) at Beth Israel Deaconess Medical Center (BIDMC) and Massachusetts Institute of Technology (MIT). Since the database contains no protected health information, IRB approval included a waiver for informed consent. Reporting of this study followed the Transparent Reporting of Multivariate Predictive Models for Individual Prognosis or Diagnosis (TRIPOD) guidelines (Appendix Table S1) [19].

2.2. Study population

The inclusion criteria for this study were patients aged ≥18 years admitted to the ICU with IS as the major cause. The diagnostic criteria for IS are shown in Appendix Table S2. Patients with ICU stay less than 24 h and without an APACHE III/IVa score were excluded. Furthermore, we removed patients whose weights were outside 50–300 kg or their heights were outside 50–250 cm. As the outcome was in-hospital mortality, patients with missing a discharge status were excluded. The APACHE scoring system, which has been the gold standard for intensive care, is categorized into two modified versions known as APACHE-III and IVa [15]. MIMIC-IV and eICU-CRD databases use APACHE-III and APACHE-IVa, respectively. Here, for convenience in the following description, they are referred to collectively as APACHE.

Table 1
Baseline characteristics of the MIMIC IV and eICU-CRD cohorts.

Characteristic, (Units)	MIMIC IV cohort (n = 3149)			eICU-CRD cohort (n = 2464)		
	Survival	Death	*P*-value	Survival	Death	*P*-value
	n = 2583	n = 566		n = 2227	n = 237	
Age (years)	68.4 (16.3)	73.5 (14.2)	<0.001	68.1 (14.9)	74.1 (13.0)	<0.001
Intubation			<0.001			<0.001
No (n, %)	1915 (74.1)	239 (42.2)		1970 (88.5)	118 (49.8)	
Yes (n, %)	668 (25.9)	327 (57.8)		257 (11.5)	119 (50.2)	
Heart rate (beats/min)	80.6 (14.7)	86.4 (16.5)	<0.001	76.7 (14.1)	85.5 (17.6)	<0.001
Mean arterial pressure (mmHg)	81.9 (8.7)	77.9 (9.9)	<0.001	84.4 (8.5)	82.3 (9.7)	0.002
Respiratory rate (breaths/min)	19.2 (3.3)	20.7 (4.2)	<0.001	18.8 (3.7)	20.3 (5.1)	<0.001
Temperature (°C)	36.9 (0.4)	37.0 (0.8)	0.114	36.8 (0.4)	37.0 (0.5)	<0.001
Urine output (mL)	1761.4 (1248.7)	1535.0 (1295.0)	<0.001	1651.3 (1066.3)	1372.8 (1062.0)	0.001
Hemoglobin (g/dL)	34.0 (6.9)	31.8 (6.9)	<0.001	12.6 (2.1)	12.3 (2.2)	0.056
WBC (10^9/L)	12.2 (6.9)	15.7 (9.6)	<0.001	10.3 (5.2)	13.7 (6.6)	<0.001
Platelets (10^9/L)	216.7 (104.8)	197.0 (100.6)	<0.001	212.7 (69.5)	210.2 (94.5)	0.697
Albumin (g/dL)	3.6 (0.6)	3.2 (0.7)	<0.001	3.5 (0.5)	3.3 (0.7)	0.001
Anion gap (mEq/L)	15.8 (3.8)	17.8 (5.2)	<0.001	11.4 (4.0)	12.9 (5.1)	<0.001
Bicarbonate (mEq/L)	22.5 (3.8)	20.5 (4.9)	<0.001	24.0 (3.4)	22.3 (4.0)	<0.001
BUN (mg/dL)	22.9 (15.8)	32.3 (23.0)	<0.001	20.7 (12.3)	26.8 (16.3)	<0.001
Creatinine (mg/dL)	1.3 (2.2)	1.7 (1.7)	<0.001	1.2 (1.0)	1.4 (1.0)	0.013
Glucose (mEq/L)	155.2 (84.7)	190.8 (97.5)	<0.001	147.7 (76.0)	185.9 (87.4)	<0.001
ALP (IU/L)	90.6 (70.4)	112.6 (88.2)	<0.001	86.5 (42.7)	96.6 (41.2)	0.004
Bilirubin (mg/dL)	0.8 (1.9)	1.4 (3.6)	0.003	0.7 (0.4)	0.9 (0.7)	<0.001
CKD			<0.001			0.025
No (n, %)	2133 (82.6)	431 (76.1)		2134 (95.8)	219 (92.4)	
Yes (n, %)	450 (17.4)	135 (23.9)		93 (4.2)	18 (7.6)	
Antiarrhythmic			0.001			0.048
No (n, %)	1597 (62.0)	304 (54.2)		1753 (96.2)	184 (92.9)	
Yes (n, %)	978 (38.0)	257 (45.8)		70 (3.8)	14 (7.1)	
Antiplatelet			<0.001			0.116
No (n, %)	1553 (60.3)	389 (69.3)		1487 (81.6)	171 (86.4)	
Yes (n, %)	1022 (39.7)	172 (30.7)		336 (18.4)	27 (13.6)	
Sedative			<0.001			<0.001
No (n, %)	1028 (39.9)	118 (21.0)		1690 (92.7)	164 (82.8)	
Yes (n, %)	1547 (60.1)	443 (79.0)		133 (7.3)	34 (17.2)	
Antibacterial			<0.001			<0.001
No (n, %)	1686 (65.5)	283 (50.4)		1733 (95.1)	174 (87.9)	
Yes (n, %)	889 (34.5)	278 (49.6)		90 (4.9)	24 (12.1)	
GCS Score	11.6 (3.4)	8.6 (4.5)	<0.001	12.9 (2.9)	9.0 (4.2)	<0.001

GCS, Glasgow coma scale. WBC, white blood cell count. ALP, alkaline phosphatase. BUN, blood urea nitrogen. CKD, chronic kidney disease.

2.3. Feature extraction and selection

In the study, we collected clinical information in the first 24 h in the ICU, specifically, information on demographics, vital signs, laboratory test data, comorbidities, treatment medications, and clinical scores. Following Deshmukh et al. [20], we calculated the average or maximum and minimum values of vital signs and laboratory data within 24 h (Table 1 and Appendix Table S3). Information on treatment medications included antibacterial, vasopressors, antiarrhythmics, antiplatelets, anticoagulation, diuretics, and sedatives within 24 h. Comorbidities were included as binary variables including atrial fibrillation, diabetes mellitus, coronary artery disease, hypertension, congestive heart failure, dementia, chronic obstructive pulmonary disease, chronic kidney disease, liver failure, and metastatic cancer.

72 predictor variables were extracted and entered as candidate variables in the feature screening process (Fig. 1). Firstly, features with a missing rate >30% were excluded. Secondly, the features were examined for statistical significance with mortality, and non-significant features were also eliminated. Thirdly, features with high correlations were excluded to avoid data redundancy. Finally, a recursive feature elimination cross-validation (RFECV) algorithm based on an extreme gradient boosting model was used to select key variables and filter the optimal subset of patient features. A 10-fold cross-validation was used in this step, with AUROC as the scoring parameter. Appendix Table S3 described the reasons for features dropout.

2.4. Data preprocessing and model development

Fig. 1 depicts the outline of our study's primary workflow. The MIMIC-IV dataset, 80% of patients were randomly assigned to the training set and 20% to the internal test set. Meanwhile, all patients in the eICU-CRD dataset were used as an external test set.

To avoid data leakage, the imputation of missing values and the normalization of the data were performed separately on the training and testing sets after the division of the data. We filled the continuous and categorical variables with missing values < 5% by using the mean and the plurality, respectively. For variables with missing values > 5%, we used the miceforest package to implement multiple imputation [21]. To improve the stability of the models, all continuous variables were normalized to achieve a distribution with a mean of 0 and a standard

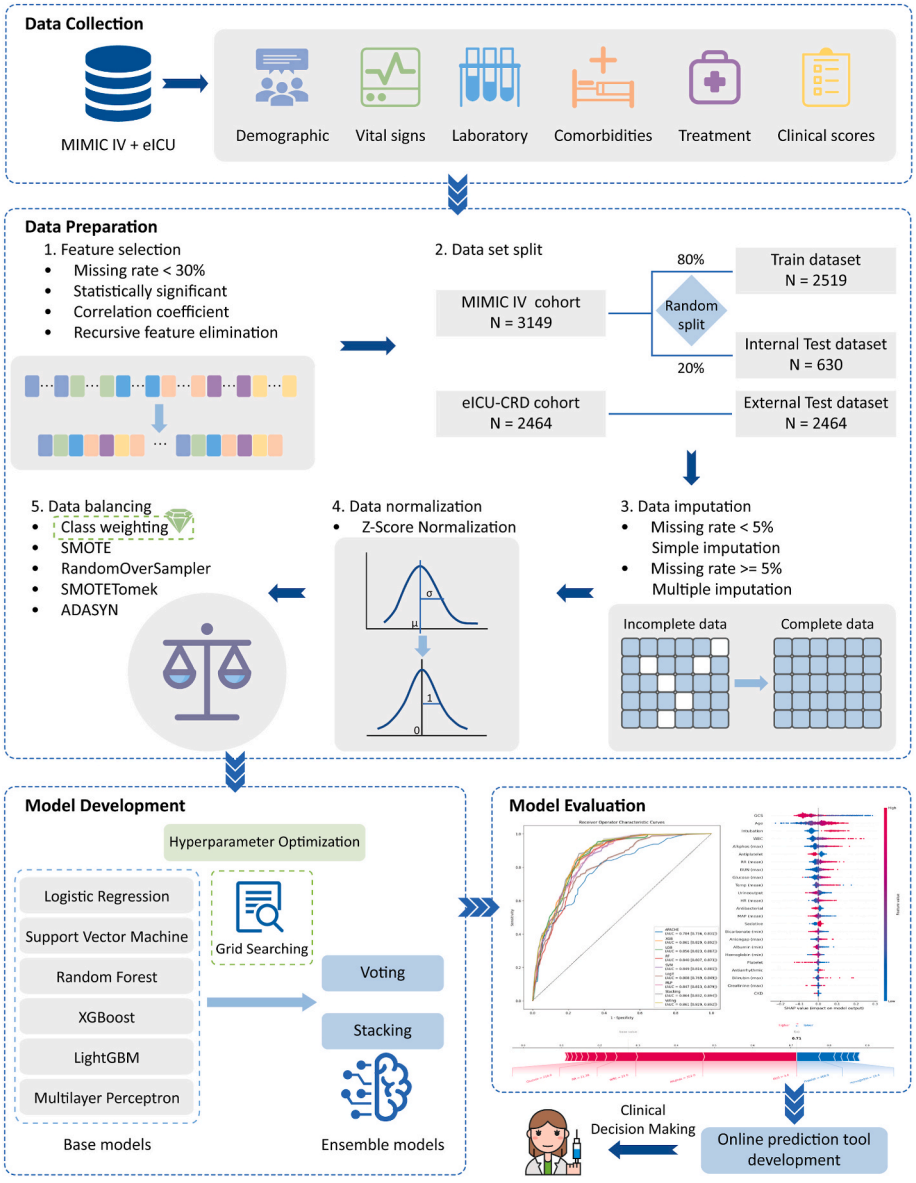

Fig. 1. Flowchart for the development of the early mortality prediction model for IS patients in the ICU. SMOTE, synthetic minority over-sampling technique. ADASYN, adaptive synthetic sampling.

deviation of 1 [22].

Data from MIMIC-IV and eICU-CRD are distributed imbalanced: the ratios of the death and survival groups were approximately 1:5 and 1:10, respectively. Therefore, data imbalances were handled using 5 methods: class-weighting method, synthetic minority oversampling technique (SMOTE) [23], random oversampling examples (ROSE) [24], adaptive synthetic sampling (ADASYN) [25], and SMOTETomek [26]. Finally, after comparing the model performance of the above methods, the class-weighting method was chosen for model development (Appendix Table S4).

In the present study, a two-layer ensemble model was applied. The first layer consists of six ML models: logistic regression (LR), support vector machine (SVM), random forest (RF), extreme gradient boosting model (XGBoost), light gradient boosting machine (LightGBM), and multilayer perceptron (MLP). We also fitted several other ML methods, including decision tree (DT), k-nearest neighbor (KNN) and naive bayes (NB). However, since their AUROC values were <0.8 in the initial experiments, they were all excluded (Appendix Table S5). In the second layer, a soft voting or stacking method was used to balance the prediction results from the first layer [27].

Soft voting is an ensemble learning method that applies a weighted average to the prediction results of multiple models and selects the class with the highest weighted average as the final prediction result [28]. The stacking ensemble method takes predictions from multiple models as input and then uses a meta model to train those predictions to generate the final result [29].

Hyperparameters were optimized during training using a grid search strategy and 10-fold cross-validation. The model with the highest average performance on all cross-validation sets was chosen. For the grid search process, AUROC was used as the scoring metric. Based on the completed risk prediction models, the optimal model was selected to develop and publish an online prediction tool.

2.5. Statistical analysis

Continuous variables were expressed as mean (standard deviation), and categorical variables were expressed as count and percentage. To test the statistical significance of the variables with patient mortality, two-tailed *t* tests and χ^2 tests were used. A two-tailed *P*-value <0.05 was considered statistically significant. For the ML model, Scikit-Learn v1.2.1, XGBoost v1.7.4, LightGBM v3.3.5 and MLXtend v0.21.0 were used to develop the models and tune the hyperparameters in Python v3.9. In addition, Pandas v1.5.2, and Scipy v1.9.1 were used for statistical analysis. Model performance was evaluated using AUROC and AUPRC. Bootstrap 2000 samples were used to estimate 95% confidence intervals (CIs) for the AUC.

3. Results

3.1. Characteristics of the study participants

According to the inclusion and exclusion criteria of participants, 3149 and 2464 IS patients were included in the MIMIC-IV and eICU-CRD cohorts, respectively (Appendix Fig. S1). After a rigorous feature selection process (Appendix Fig. S2), 24 important features including age, temperature, heart rate (HR), mean aortic pressure (MAP), respiratory rate (RR), GCS score, intubation, urine output in the first 24 h, hemoglobin, white blood cell count (WBC), platelet count (PLT), albumin, anion gap, bicarbonate, blood urine nitrogen (BUN), creatinine, blood glucose, alkaline phosphatase (ALP), total bilirubin, chronic kidney disease (CKD), antiarrhythmic, antiplatelet, sedative and antibacterial were determined for the development of a compact model. The in-hospital mortality rates were 17.97% (n = 566) and 9.62% (n = 237) for the MIMIC-IV and eICU-CRD cohorts, respectively. In both cohorts, the mean age of the death group was 73.5 and 74.1, which was significantly higher than the 68.4 and 68.1 for the survival group ($p < 0.001$). The demographic and clinical features differences are outlined in Table 1.

3.2. Performance of the machine learning models

A grid search strategy and 10-fold cross-validation were used to select optimal hyperparameters for all machine learning models developed on the training set (2519 patients). The final optimal parameters for each model are shown in Appendix Tables S6–S7.

The AUROC values of the 6 individual models were superior to the APACHE scoring system in both internal (630 patients) and external tests (2464 patients), as shown in Table 2. The two ensemble models based on stacking and soft voting are more robust and consistent than the individual models, significantly outperforming APACHE in both test sets ($p < 0.05$). Similarly, all the models had higher AUPRC than APACHE, with the stacking and soft voting models being 1st and 2nd, respectively, on the internal test set, and 3rd and 1st on the external test set. Fig. 2 displays the AUROC and AUPRC of these models in two independent test sets.

3.3. Risk stratification and development of application

Both the stacking and soft voting classifiers demonstrated excellent performance in the task of predicting mortality in IS patients, and the soft voting classifier with better robustness was chosen as the final prediction model. Risk stratification was developed in our model to make it more clinically applicable. Based on three-fold cross-validation of the training cohort, risk cutoff values of 0.172 and 0.581 were selected at 95% sensitivity and 95% specificity, respectively (Appendix Fig. S3). Patients were divided into three risk categories: low-, medium-, and high-risk. In the two test cohorts, these risk cutoff values were evaluated independently. The confusion matrixes demonstrated that the risk stratification system can effectively distinguish between patients who are low-, medium-, and high-risk (Fig. 3A). The likelihood of in-hospital mortality differed by risk group (Fig. 3B). In the internal test cohort, the in-hospital mortality rates for low-, medium-, and high-risk patients were 1.44% (0.011–0.018), 25.0% (0.229–0.272), and 56.4% (0.536–0.593), respectively. Similarly, in the external test cohort, these values were 1.86% (0.015–0.022), 11.69% (0.110–0.124), and 36.38% (0.348–0.380), respectively.

The soft voting ensemble learning algorithm was used to build the risk prediction model, which we named the *ICU Ischemic Stroke Prediction Model* (ICU-ISPM). Based on ICU-ISPM, an online risk prediction tool was further developed (http://ispm.idrblab.cn/; Appendix Fig. S4). By entering all necessary values, users can obtain information about in-hospital mortality and risk stratification for IS patients in the ICU.

3.4. Feature importance and model interpretations

The ICU-ISPM was interpretable using the Shapley additive explain (SHAP) method, which allows the contribution and impact of each feature on the final prediction to be calculated precisely [30]. In Fig. 3C, we plot the importance of features for the ICU-ISPM as well as the relative weights of features in decision-making. According to the weight of the features, the top three most important features are GCS score, age, and intubation. The ICU-ISPM's SHAP summary graph further illustrates the relationship between the high and low values of each feature and SHAP values. As shown in Fig. 3D, a low GCS score, older age, and intubation increase death risk. The SHAP force graph in Fig. S5 analyzes the predicted outcomes for specific patients at the local level. For example, we randomly selected two cases from high- and low-risk groups that accurately predicted death and survival (See Appendix Fig. S5 for details).

Table 2
Performance of machine learning models and APACHE scoring system for early mortality prediction.

Model name	Internal test set				External test set			
	AUROC	Confidence Interval (95%)		*P* value	AUROC	Confidence Interval (95%)		*P* value
APACHE	0.784	0.736	0.831	–	0.799	0.770	0.830	–
XGBoost	0.861	0.829	0.892	**0.01**	0.832	0.807	0.858	0.10
LightGBM	0.856	0.823	0.887	**0.01**	0.832	0.808	0.859	0.10
RF	0.840	0.807	0.873	0.06	0.840	0.815	0.866	**0.04**
SVM	0.849	0.814	0.881	**0.03**	0.836	0.910	0.860	0.06
LR	0.808	0.769	0.849	0.41	0.845	0.820	0.870	**0.02**
MLP	0.847	0.813	0.879	**0.03**	0.818	0.792	0.846	0.36
Stacking	0.864	0.832	0.894	**0.01**	0.840	0.815	0.866	**0.04**
Voting	0.861	0.829	0.892	**0.01**	0.844	0.819	0.869	**0.02**

APACHE, acute physiology and chronic health assessment. LR, logistic regression. SVM, support vector machine. RF, random forest. XGBoost, extreme gradient boosting. LightGBM, light gradient boosting machine. MLP, Multilayer Perceptron.

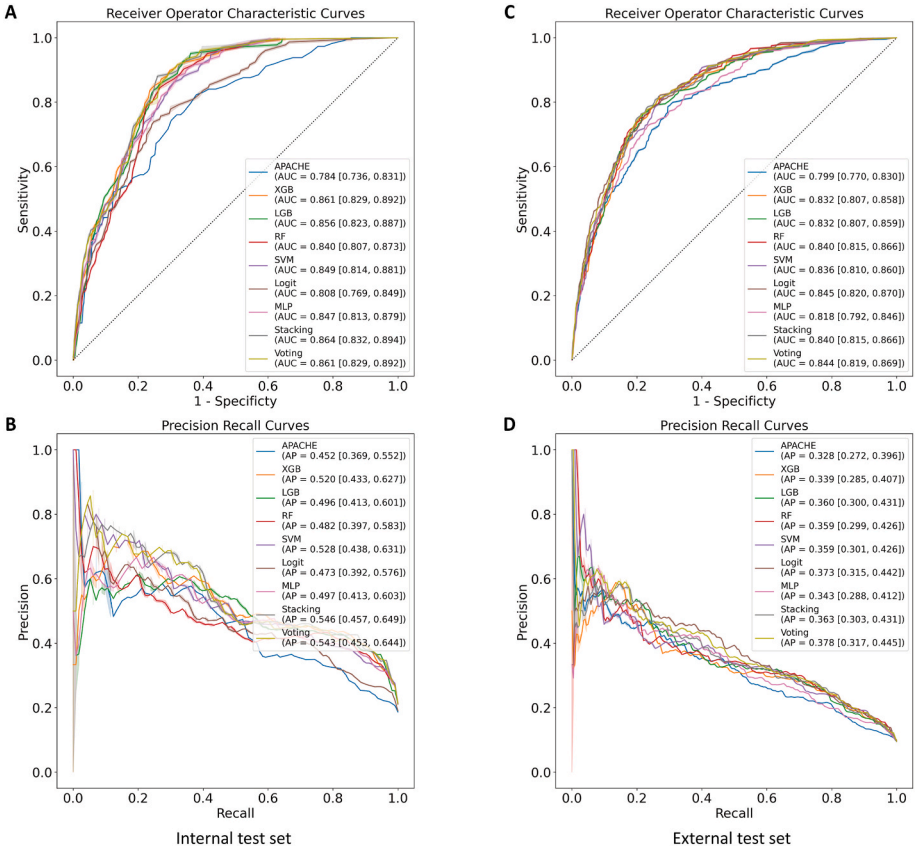

Fig. 2. Model performances are compared in internal test and external test sets. (A) AUROC results of internal test set. (B) AUPRC results of internal test set. (C) AUROC results of external test set. (D) AUPRC results of external test set. APACHE, acute physiology and chronic health assessment. XGB, extreme gradient boosting. LGB, light gradient boosting machine. RF, random forest. SVM, support vector machine. Logit, logistic regression. MLP, Multilayer Perceptron.

4. Discussion

Early in the course of IS, especially in patients admitted to the ICU, the mortality rate is highest [31]. An accurate, reliable, interpretable, and clinically applicable risk assessment model is essential for the management of IS patients. In this study, we developed a novel predictive model, ICU-ISPM, for predicting early mortality in IS patients. The innovation of the model was the utilization of an ensemble learning method. Additionally, the study's unique characteristics included a rigorous feature selection process, multiple imbalanced data processing methods, a sound risk stratification strategy and a user-friendly online prediction tool.

ICU-ISPM was comprehensively tested in independent internal and external test cohorts, and the results showed it's accurate, generalizable, and significantly superior to APACHE scoring system. In test sets, the AUROC value of ICU-ISPM improved by at least 4.5% compared to APACHE. The robustness performance of ICU-ISPM outperformed various classical ML models (LR, XGBoost, LightGBM, RF, SVM and MLP). Such results provide additional evidence that ensemble learning algorithms can enhance the performance and robustness of classification tasks involving structured clinical forms.

Currently, predicting mortality in stroke patients is an active area of research, and several studies have been published on predicting the risk of early mortality in IS patients. Soft voting and stacking methods produced satisfactory and reliable AUROC, AUPRC scores when compared to individual models. Table 3 compares the performance of our model to those of other similar studies [9–12]. Based on the comparison of AUROC metrics, our model had the best performance since not all

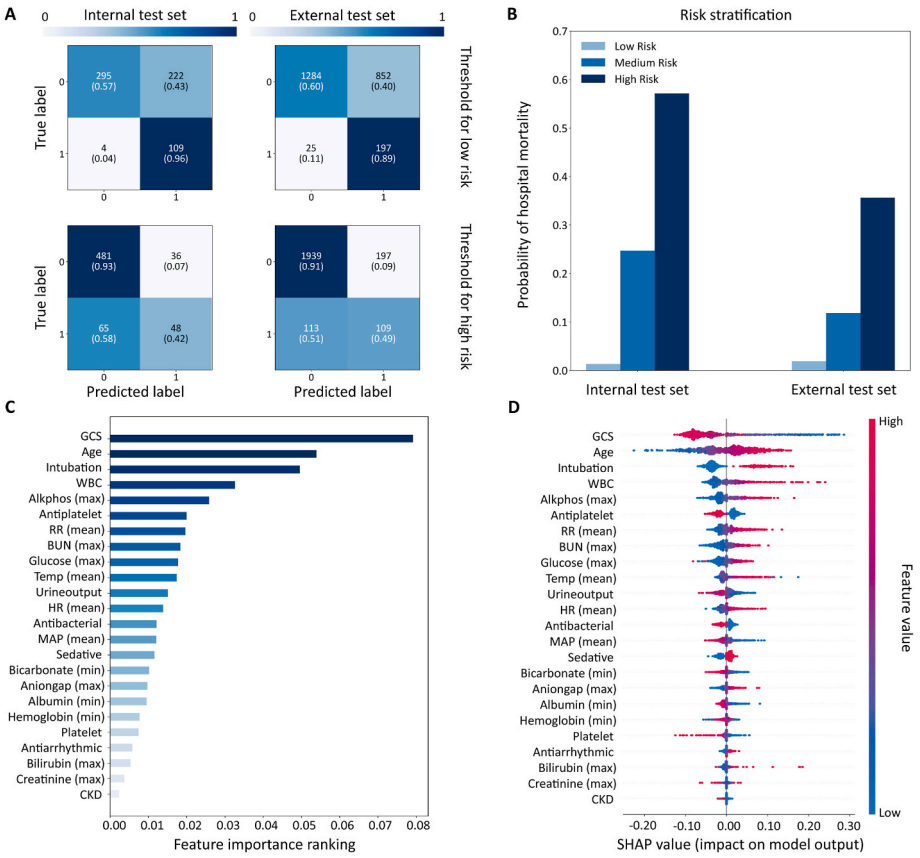

Fig. 3. Risk stratification and feature contribution analysis for ICU-ISPM. (A) Confusion matrices at cutoffs of 0.172 and 0.581 for the low- and high-risk groups, based on the internal test and external test sets. (B) The actual probability of early mortality in low-, medium- and high-risk groups in the internal and external test sets, respectively. (C) The weighting and ranking of the importance of the 24 features. (D) The SHAP summary graph. GCS, glasgow coma scale. WBC, white blood cell count. Alkphos, alkaline phosphatase. RR, respiratory rate. BUN, blood urea nitrogen. Temp, temperature. Urineoutput, 24 h urine output. HR, heart rate. MAP, mean arterial pressure. CKD, chronic kidney disease.

Table 3
Performance comparison with other similar studies.

Study	Algorithm/Feature number	Internal test		External test	
		AUROC (95% CI)	AUPRC (95% CI)	AUROC (95% CI)	AUPRC (95% CI)
Current study (ICU-ISPM)	Soft voting/24	**0.861 (0.829–0.892)**	**0.543 (0.453–0.644)**	**0.844 (0.819–0.869)**	**0.378 (0.317–0.445)**
Yang Ouyang (2023) [10]	Random Forest/18	0.799 (NA)	NA	0.733 (NA)	NA
Wei Liu (2022) [11]	Random Forest/58	0.806 (NA)	0.402 (NA)	0.838 (NA)	0.417 (NA)
Vida Abedi (2021) [12]	Random Forest/37	0.820 (NA)	NA	NA	NA
Xiaodan Li (2022) [9]	Cox regression/12	0.753 (NA)	NA	NA	NA

AUROC, area under the receiver operating characteristic curve. AUPRC, area under the precision-recall curve. CI, confidence interval.

studies provided AUPRC values.

Additionally, choosing the appropriate decision threshold based on the prediction model is a challenge. However, previous studies have lacked corresponding work. In our study, the population was classified as low-, medium-, or high-risk with 95% sensitivity and 95% specificity using risk cut-off values of 0.172 and 0.581. In other words, by using the cut-off values for low and high risk, we can accurately identify truly low- and high-risk populations. To make our study findings clinically useful, we developed an open-access online prediction tool based on ICU-ISPM. Users can easily and quickly estimate the probability of early mortality in IS patients by entering the relevant variables, as well as learn more about the risk stratification. Clinical decision-makers can use the output of the model to stratify patients early, identify low- or high-risk groups, make more optimal decisions, and allocate limited medical resources more efficiently [32,33]. In this way, ICU resources can be efficiently used and medical quality and efficiency can be improved.

Interpretive analysis of ICU-ISPM showed that a low GCS score, older age and intubation were strong predictors of early mortality in IS patients. The GCS score has been demonstrated to be a strong predictor of hospital mortality and poor neurological prognosis [34], with infarcted patients with a GCS $>$ 8 having a significantly better prognosis than those with a GCS $<$ 8 [35]. Similarly, our results indicate that the GCS score is the most important feature of ICU-ISPM. Typically, older patients have a higher risk of both morbidity and mortality from strokes. In-hospital mortality following a stroke increases after the age of 60 and reaches a peak of more than 18% after the age of 90 [36]. Mechanical ventilation or other types of respiratory support are typically needed for patients with acute brain injuries who are admitted to the ICU. Studies have shown that patients with severe IS requiring airway intubation are at high risk of death and have a poor functional prognosis [37,38]. White blood cell count, platelet count and hemoglobin level were important features that affected the ability of our model, and previous

studies have confirmed their significance [39–42].

This study identified additional serum biomarkers, especially ALP, which was not included in previous stroke mortality models despite the fact that a higher serum ALP level is associated with stroke mortality [43]. Furthermore, to the best of our knowledge, this is the first study to incorporate early medication as variables to predict IS mortality. Antiplatelet and antibacterial therapy both decreased mortality risk, but sedation and antiarrhythmic therapy increased such a risk. It is well known that guidelines recommend antiplatelet agents as the basic treatment for AIS, confirming their importance in treating IS and improving prognosis [44]. Patients with severe stroke often require more aggressive thrombolysis or thrombolytic therapy, which often delays antiplatelet therapy until after 24 h, so those who receive antiplatelet therapy early are generally less ill [44]. Stroke patients are more susceptible to infections, the most prevalent of which are pneumonia and urinary tract infections, as a result of impaired antimicrobial defenses [39]. Antibacterial agents used in the treatment or prevention of post-stroke infections may enhance recovery from stroke [39]. Currently, there are no reports indicating an association between antiarrhythmic or sedative agents and the prognosis of IS patients, but a hypothesis might be that such patients would have a greater burden of comorbidities leading to a poor prognosis [45].

This study has some limitations. First, our data sources were two U.S. public databases. Although the involved populations included a mix of different ethnicities, over half were White or Caucasian, which may have prevented the extrapolation of these findings to other mixed-ethnic patient groups. Second, due to the large number of missing values, some clinical characteristics [46,47], including d-dimer, lipoprotein and BNP, that have been observed to potentially affect the prognosis of IS patients had to be eliminated. Third, neither database provided information on NIHSS score and imaging, which are frequently used by clinicians to assess the severity of patients with IS [8]. As a result, these variables were excluded from our prediction model.

5. Conclusions

The present study developed and validated a novel risk prediction model to estimate early mortality of patients with IS in the ICU, surpassing the performance of existing scoring systems. By employing the soft voting ensemble method, both accuracy and robustness of ICU-ISPM were enhanced. Moreover, our interpretable model enables clinicians to comprehend the underlying factors contributing to predicted outcomes. ICU-ISPM effectively identifies low-, medium-, and high-risk IS patients at an early stage, thereby facilitating clinical decision-making and optimal allocation of medical resources.

Author contributions

Wei Hu conceived and designed the study. Wei Hu and Tingting Jin conducted the data collection and wrote the manuscript. Ziqi Pan developed a web server. Wei Hu, Huimin Xu, Lingyan Yu, Tingting Chen, and Wei Zhang participated in the development of the methodology. Wei Hu, Huifang Jiang, and Wenjun Yang performed the analysis and generated the figures and tables. Feng Zhu and Haibin Dai provided funding and supervision for this study; reviewed and revised the manuscript. The authors read and approved the final manuscript.

Declaration of competing interest

None declared.

Acknowledgments

This work was funded by National Natural Science Foundation of China (82173789, 82271391 & 81872798); Natural Science Foundation of Zhejiang Province (LR21H300001); Leading Talent of “Ten Thousand Plan” of the National High-Level Talents Special Support Plan of China; Fundamental Research Fund of Central University (2018QNA7023); Key R&D Program of Zhejiang Province (2020C03010); Chinese “Double Top-Class” Universities (181201*194232101); Westlake Laboratory (Westlake Laboratory of Life Science & Biomedicine); Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare; Alibaba Cloud; Information Tech Center of Zhejiang University.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2023.107577.

References

[1] G.B.D.S. Collaborators, Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019, Lancet Neurol. 20 (2021) 795–820, https://doi.org/10.1016/S1474-4422(21)00252-0.
[2] L. Premraj, C. Camarda, N. White, et al., Tracheostomy timing and outcome in critically ill patients with stroke: a meta-analysis and meta-regression, Crit. Care 27 (2023) 132, https://doi.org/10.1186/s13054-023-04417-6.
[3] M.G. Kashiouris, C.N. Sessler, R. Qayyum, et al., Near-simultaneous intensive care unit (ICU) admissions and all-cause mortality: a cohort study, Intensive Care Med. 45 (2019) 1559–1569, https://doi.org/10.1007/s00134-019-05753-4.
[4] R. Pirracchio, M.L. Petersen, M. Carone, et al., Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study, Lancet Respir. Med. 3 (2015) 42–52, https://doi.org/10.1016/S2213-2600(14)70239-5.
[5] A. St Andre, The formation, elements of success, and challenges in managing a critical care program: part II, Crit. Care Med. 43 (2015) 1096–1101, https://doi.org/10.1097/CCM.0000000000000856.
[6] S.M. Fernando, B. Rochwerg, P.M. Reardon, et al., Emergency Department disposition decisions and associated mortality and costs in ICU patients with suspected infection, Crit. Care 22 (2018) 172, https://doi.org/10.1186/s13054-018-2096-8.
[7] C.W. Tsao, A.W. Aday, Z.I. Almarzooq, et al., Heart disease and stroke statistics-2022 update: a report from the American heart association, Circulation 145 (2022) e153–e639, https://doi.org/10.1161/CIR.0000000000001052.
[8] S.E. Kasner, Clinical interpretation and use of stroke scales, Lancet Neurol. 5 (2006) 603–612, https://doi.org/10.1016/S1474-4422(06)70495-1.
[9] X.D. Li, M.M. Li, A novel nomogram to predict mortality in patients with stroke: a survival analysis based on the MIMIC-III clinical database, BMC Med. Inf. Decis. Making 22 (2022) 92, https://doi.org/10.1186/s12911-022-01836-3.
[10] Y. Ouyang, M. Cheng, B. He, et al., Interpretable machine learning models for predicting in-hospital death in patients in the intensive care unit with cerebral infarction, Comput. Methods Progr. Biomed. 231 (2023), 107431, https://doi.org/10.1016/j.cmpb.2023.107431.
[11] W. Liu, W. Ma, N. Bai, et al., Identification of key predictors of hospital mortality in critically ill patients with embolic stroke using machine learning, Biosci. Rep. 42 (2022), https://doi.org/10.1042/BSR20220995.
[12] V. Abedi, V. Avula, S.M. Razavi, et al., Predicting short and long-term mortality after acute ischemic stroke using EHR, J. Neurol. Sci. 427 (2021), 117560, https://doi.org/10.1016/j.jns.2021.117560.
[13] J.J. Yang, J. Li, R. Shen, et al., Exploiting ensemble learning for automatic cataract detection and grading, Comput. Methods Progr. Biomed. 124 (2016) 45–57, https://doi.org/10.1016/j.cmpb.2015.10.007.
[14] N. An, H. Ding, J. Yang, R. Au, T.F.A. Ang, Deep ensemble learning for Alzheimer's disease classification, J. Biomed. Inf. 105 (2020), 103411, https://doi.org/10.1016/j.jbi.2020.103411.
[15] J.E. Zimmerman, A.A. Kramer, D.S. McNair, F.M. Malila, Acute Physiology and Chronic Health Evaluation (Apache) IV: hospital mortality assessment for today's critically ill patients, Crit. Care Med. 34 (2006) 1297–1310, https://doi.org/10.1097/01.CCM.0000215112.84523.F0.
[16] Z. Haddad, B.F. Falissard, K.C. Chokri, et al., Disparity in outcome prediction between Apache II, Apache III and Apache IV, Crit. Care 12 (2008) P501, https://doi.org/10.1186/cc6722.
[17] A.E.W. Johnson, L. Bulgarelli, L. Shen, et al., MIMIC-IV, a freely accessible electronic health record dataset, Sci. Data 10 (2023) 1, https://doi.org/10.1038/s41597-022-01899-x.
[18] T.J. Pollard, A.E.W. Johnson, J.D. Raffa, et al., The eICU Collaborative Research Database, a freely available multi-center database for critical care research, Sci. Data 5 (2018), 180178, https://doi.org/10.1038/sdata.2018.178.
[19] G.S. Collins, J.B. Reitsma, D.G. Altman, K.G. Moons, Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement, BMJ 350 (2015), g7594, https://doi.org/10.1136/bmj.g7594.
[20] F. Deshmukh, S.S. Merchant, Explainable machine learning model for predicting GI bleed mortality in the intensive care unit, Am. J. Gastroenterol. 115 (2020) 1657–1668, https://doi.org/10.14309/ajg.0000000000000632.
[21] H. Tang, Z. Jin, J. Deng, et al., Development and validation of a deep learning model to predict the survival of patients in ICU, J. Am. Med. Inf. Assoc. 29 (2022) 1567–1576, https://doi.org/10.1093/jamia/ocac098.

[22] L. Al Shalabi, Z. Shaaban, Normalization as a preprocessing engine for data mining and the approach of preference matrix, DEPCOS-RELCOMEX (2006) 207–214, 2006, 10.1109/DEPCOS-RELCOMEX.2006.38pp.
[23] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357, https://doi.org/10.1613/jair.953.
[24] N. Lunardon, G. Menardi, N. Torelli, Rose, A package for binary imbalanced learning, R J 6 (2014) 79–89, https://doi.org/10.32614/RJ-2014-008.
[25] H.B. He, Y. Bai, E.A. Garcia, S. T. Li Adasyn, Adaptive synthetic sampling approach for imbalanced learning, Ieee Ijcnn (2008) 1322–1328, https://doi.org/10.1109/Ijcnn.2008.4633969.
[26] Z. Wang, C.H. Wu, K.F. Zheng, X.X. Niu, X.J. Wang, SMOTETomek-based resampling for personality recognition, IEEE Access 7 (2019) 129678–129689, https://doi.org/10.1109/Access.2019.2940061.
[27] J.A.M. Sidey-Gibbons, C.J. Sidey-Gibbons, Machine learning in medicine: a practical introduction, BMC Med. Res. Methodol. 19 (2019) 64, https://doi.org/10.1186/s12874-019-0681-4.
[28] Z. Wang, Z. Wu, M. Zou, et al., A voting-based ensemble deep learning method focused on multi-step prediction of food safety risk levels: applications in hazard analysis of heavy metals in grain processing products, Foods 11 (2022), https://doi.org/10.3390/foods11060823.
[29] A.I. Naimi, L.B. Balzer, Stacked generalization: an introduction to super learning, Eur. J. Epidemiol. 33 (2018) 459–464, https://doi.org/10.1007/s10654-018-0390-z.
[30] S.M. Lundberg, G. Erion, H. Chen, et al., From local explanations to global understanding with explainable AI for trees, Nat. Mach. Intell. 2 (2020) 56–67, https://doi.org/10.1038/s42256-019-0138-9.
[31] R.J. Singh, S. Chen, A. Ganesh, M.D. Hill, Long-term neurological, vascular, and mortality outcomes after stroke, Int. J. Stroke 13 (2018) 787–796, https://doi.org/10.1177/1747493018798526.
[32] A. Leigh, J.W. McEvoy, P. Garg, et al., Coronary artery calcium scores and atherosclerotic cardiovascular disease risk stratification in smokers, JACC Cardiovasc Imaging 12 (2019) 852–861, https://doi.org/10.1016/j.jcmg.2017.12.017.
[33] M.B. Downer, L. Li, S. Carter, S. Beebe, P.M. Rothwell, Associations of multimorbidity with stroke severity, subtype, premorbid disability, and early mortality: oxford vascular study, Neurology 101 (2023) e645–e652, https://doi.org/10.1212/WNL.0000000000207479.
[34] C.A.I. Akerlund, A. Holst, N. Stocchetti, et al., Clustering identifies endotypes of traumatic brain injury in an intensive care cohort: a CENTER-TBI study, Crit. Care 26 (2022) 228, https://doi.org/10.1186/s13054-022-04079-w.
[35] C.L. Tsai, H. Chu, G.S. Peng, et al., Preoperative Apache II and GCS scores as predictors of outcomes in patients with malignant MCA infarction after decompressive hemicraniectomy, Neurol. India 60 (2012) 608–612, https://doi.org/10.4103/0028-3886.105195.
[36] T. Gattringer, A. Posekany, K. Niederkorn, et al., Predicting early mortality of acute ischemic stroke, Stroke 50 (2019) 349–356, https://doi.org/10.1161/STROKEAHA.118.022863.
[37] J. Bosel, W.D. Niesen, F. Salih, et al., Effect of early vs standard approach to tracheostomy on functional outcome at 6 Months Among patients with severe stroke receiving mechanical ventilation: the SETPOINT2 randomized clinical trial, JAMA 327 (2022) 1899–1909, https://doi.org/10.1001/jama.2022.4798.
[38] K. Asehnoune, P. Rooze, C. Robba, et al., Mechanical ventilation in patients with acute brain injury: a systematic review with meta-analysis, Crit. Care 27 (2023) 221, https://doi.org/10.1186/s13054-023-04509-3.
[39] N.S. Parikh, A.E. Merkler, C. Iadecola, Inflammation, autoimmunity, infection, and stroke: epidemiology and lessons from therapeutic intervention, Stroke 51 (2020) 711–718, https://doi.org/10.1161/STROKEAHA.119.024157.
[40] S. You, Z. Ou, W. Zhang, et al., Combined utility of white blood cell count and blood glucose for predicting in-hospital outcomes in acute ischemic stroke, J. Neuroinflammation 16 (2019) 37, https://doi.org/10.1186/s12974-019-1422-7.
[41] Y. Liu, F. Li, H. Sun, et al., Combined prognostic significance of D-dimer level and platelet count in acute ischemic stroke, Thromb. Res. 194 (2020) 142–149, https://doi.org/10.1016/j.thromres.2020.05.021.
[42] A. Dayimu, W. Qian, B. Fan, et al., Trajectories of Haemoglobin and incident stroke risk: a longitudinal cohort study, BMC Publ. Health 19 (2019), 1395, https://doi.org/10.1186/s12889-019-7752-7.
[43] A.L. Brichacek, C.M. Brown, Alkaline phosphatase: a potential biomarker for stroke and implications for treatment, Metab. Brain Dis. 34 (2019) 3–19, https://doi.org/10.1007/s11011-018-0322-3.
[44] W.J. Powers, A.A. Rabinstein, T. Ackerson, et al., Guidelines for the early management of patients with acute ischemic stroke: a guideline for Healthcare professionals from the American heart association/American stroke association, Stroke 49 (2018) e46–e110, https://doi.org/10.1161/STR.0000000000000158, 2018.
[45] M. Galovic, C. Ferreira-Atuesta, L. Abraira, et al., Seizures and epilepsy after stroke: epidemiology, biomarkers and management, Drugs Aging 38 (2021) 285–299, https://doi.org/10.1007/s40266-021-00837-7.
[46] D. Battaglini, C. Robba, A. Lopes da Silva, et al., Brain-heart interaction after acute ischemic stroke, Crit. Care 24 (2020) 163, https://doi.org/10.1186/s13054-020-02885-8.
[47] K. Tziomalos, K. Katrini, M. Papagianni, et al., Impaired antioxidative activity of high-density lipoprotein is associated with more severe acute ischemic stroke, Metabolism 98 (2019) 49–52, https://doi.org/10.1016/j.metabol.2019.06.004.