



OPEN ACCESS

EDITED BY

Maurizio Acampa,
Siena University Hospital, Italy

REVIEWED BY

Ryuzaburo Kanazawa,
Nagareyama central hospital, Japan
Sheng-Feng Sung,
Ditmanson Medical Foundation Chia-Yi
Christian Hospital, Taiwan

*CORRESPONDENCE

Woo-Keun Seo
✉ mcastenosis@gmail.com
Jin-Man Jung
✉ dr.jinmanjung@gmail.com

[†]These authors have contributed equally to this work and share first authorship

RECEIVED 21 June 2023

ACCEPTED 17 October 2023

PUBLISHED 08 November 2023

CITATION

Jeon E-T, Jung SJ, Yeo TY, Seo W-K and Jung J-M (2023) Predicting short-term outcomes in atrial-fibrillation-related stroke using machine learning.
Front. Neurol. 14:1243700.
doi: 10.3389/fneur.2023.1243700

COPYRIGHT

© 2023 Jeon, Jung, Yeo, Seo and Jung. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Predicting short-term outcomes in atrial-fibrillation-related stroke using machine learning

Eun-Tae Jeon^{1†}, Seung Jin Jung^{2†}, Tae Young Yeo¹,
Woo-Keun Seo^{3*} and Jin-Man Jung^{1,4*}

¹Department of Neurology, Korea University Ansan Hospital, Korea University College of Medicine, Ansan, Republic of Korea, ²Department of Family Medicine, Gimpo Woori Hospital, Gimpo, Republic of Korea, ³Department of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea, ⁴Korea University Zebrafish Translational Medical Research Center, Ansan, Republic of Korea

Background: Prognostic prediction and the identification of prognostic factors are critical during the early period of atrial-fibrillation (AF)-related strokes as AF is associated with poor outcomes in stroke patients.

Methods: Two independent datasets, namely, the Korean Atrial Fibrillation Evaluation Registry in Ischemic Stroke Patients (K-ATTENTION) and the Korea University Stroke Registry (KUSR), were used for internal and external validation, respectively. These datasets include common variables such as demographic, laboratory, and imaging findings during early hospitalization. Outcomes were unfavorable functional status with modified Rankin scores of 3 or higher and mortality at 3 months. We developed two machine learning models, namely, a tree-based model and a multi-layer perceptron (MLP), along with a baseline logistic regression model. The area under the receiver operating characteristic curve (AUROC) was used as the outcome metric. The Shapley additive explanation (SHAP) method was used to evaluate the contributions of variables.

Results: Machine learning models outperformed logistic regression in predicting both outcomes. For 3-month unfavorable outcomes, MLP exhibited significantly higher AUROC values of 0.890 and 0.859 in internal and external validation sets, respectively, than those of logistic regression. For 3-month mortality, both machine learning models exhibited significantly higher AUROC values than the logistic regression for internal validation but not for external validation. The most significant predictor for both outcomes was the initial National Institute of Health and Stroke Scale.

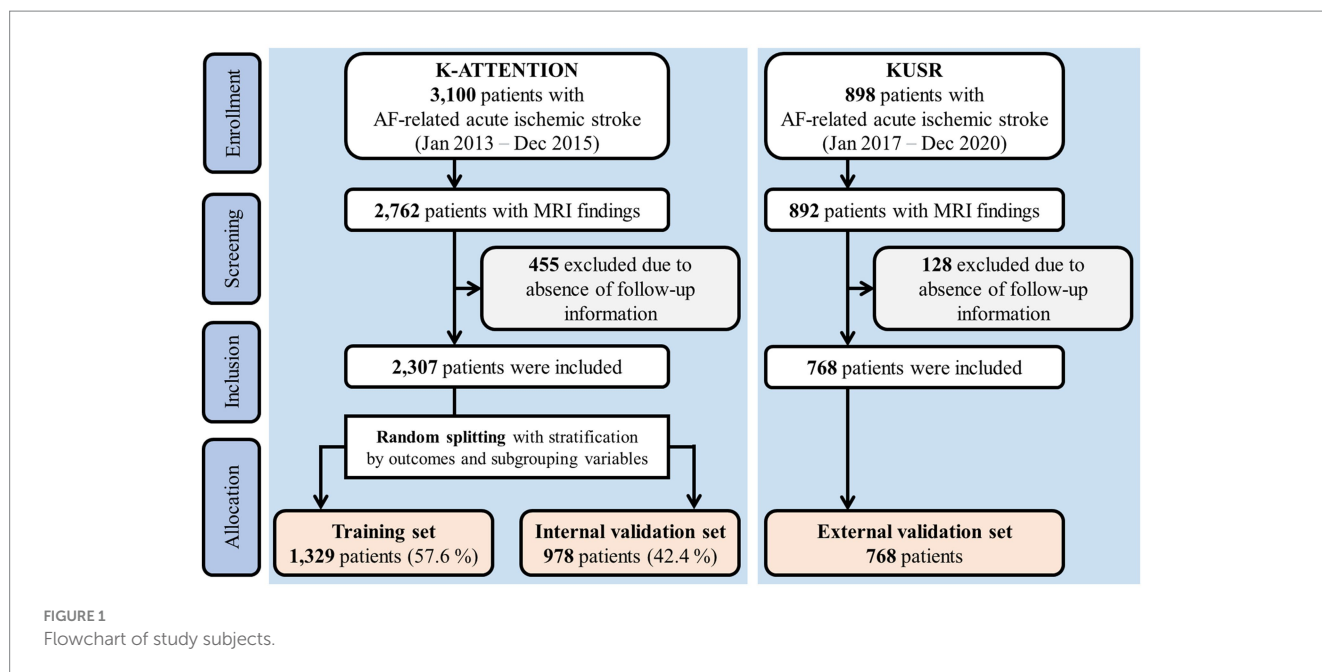
Conclusion: The explainable machine learning model can reliably predict short-term outcomes and identify high-risk patients with AF-related strokes.

KEYWORDS

atrial fibrillation, machine learning, outcome, prediction model, ischemic stroke

1. Introduction

Atrial fibrillation (AF) is a common cause of ischemic stroke, and AF-related strokes are associated with higher mortality and poorer functional outcomes than other ischemic stroke subtypes (1). The early identification of high-risk patients with poor functional outcomes is critical for maintaining a focus on available healthcare resources and improving outcomes in terms of prevention and early management. Accordingly, numerous studies have been conducted



to predict vascular events, mortality, and functional outcomes in patients experiencing AF-related stroke events. Although originally used for thromboembolic risk assessment in AF outpatients, the CHADS₂ score and its updated version, CHA₂DS₂-VASc, can be used to effectively predict the prognosis of an AF-related stroke event (2–6). However, these models are simplified and do not incorporate clinical and laboratory parameters such as the severity of stroke symptoms at admission, serum inflammatory markers, and image-based features derived from the early stages of stroke. Furthermore, the clinical implications of these scores remain controversial in stroke patients (7–9), and it may be challenging to interpret features and values extracted from other studies for practical clinical applications (10).

Clinical risk scoring systems such as CHADS₂, CHA₂DS₂-VASc, and ATRIA were originally developed for AF. However, their validation in stroke patients with AF, particularly in real-world settings with new oral anticoagulants (NOACs), has been rarely conducted. A nationwide multicenter study evaluated the scoring systems and presented unsatisfactory performance of the systems (9). This highlights the need for a new risk stratification approach tailored to secondary stroke prevention in AF patients.

Machine learning models offer various advantages over traditional parametric methods owing to their improved flexibility, capability to capture complex patterns, and good performance on large and high-dimensional datasets. All of these advantages are achieved without relying on strong assumptions, enabling machine learning to be utilized as a valuable tool in clinical practice. Therefore, it is desirable to develop novel prognostic methods that are easily interpretable and can improve risk stratification during the early stages of AF-related stroke.

Outcome prediction following ischemic stroke is generally performed using logistic regression as a statistical model using clinical and/or image-based features. However, the prediction results represent only the importance and linear directionality of the selected variables without any direct information regarding their priority. To overcome the limitations of conventional statistical and machine learning

models, more accurate high-level machine learning techniques must be developed and applied (11). The machine-learning-based prediction of outcomes using information obtained during the early period after hospital arrival—including clinical, laboratory, and imaging findings—is a feasible method of formulating therapeutic plans and prognoses (11, 12). In this study, machine learning models were constructed and validated for the prediction of short-term outcomes in AF-related stroke patients based on various features acquired during early hospitalization using two independent multicenter prospective hospital-based registries.

2. Methods

2.1. Study population

The internal and external validation sets used in this study are overviewed in Figure 1. One dataset was based on the Korean Atrial Fibrillation Evaluation Registry in Ischemic Stroke Patients (K-ATTENTION), which compiled medical information from AF-related stroke patients admitted within 7 days of symptom onset at 11 tertiary stroke centers in South Korea. K-ATTENTION has previously been used for model training, variable selection, and internal validation. The other dataset, used for external validation, comprised patients with AF-related stroke events extracted from the Korea University Stroke Registry (KUSR), collected from three Korean university hospitals (Anam, Ansan, and Guro branches). Common features among clinical, radiological, and laboratory findings acquired during early hospitalization were extracted from the two datasets to develop the models (11). Features with a loss of information pertaining to 3-month functional outcomes measured using the modified Rankin Scale (mRS) were excluded.

Approval was obtained from the institutional review boards of all participating centers. This study complied with the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis reporting guidelines (13).

2.2. Definition of outcomes

The outcomes of interest were short-term outcomes following 90 days of the index stroke. Primary outcomes were defined as unfavorable functional outcomes with an mRS ≥ 3 . The secondary outcome was mortality, defined as all-cause death within the 90-day period.

2.3. Data splitting and preprocessing

Approximately 40% of the patients from the K-ATTENTION registry were randomly selected and stratified by outcomes and variable subgroups (Figure 1). These data were used as the internal validation set, whereas the remaining data were allocated for training. Models were trained using a 10-fold cross-validation strategy.

Variables with a missing rate exceeding 20% were excluded, and missing values were imputed using multivariate imputation by chained equations (MICE) (14). Outliers were detected using an isolation forest (15) and replaced with the closest normal values from the training set. Supplementary Table S1 from the Supplementary Information lists all variables included in the analysis along with their missing rates.

2.4. Importance of variables and feature selection

The contribution of each variable to model prediction was evaluated using the Shapley additive explanations (SHAP) method (16). Positive and negative SHAP values indicate positive and negative effects, respectively, on the prediction score. A greedy backward selection method, namely, recursive feature elimination (17), was used to select the best feature set to maximize cross-validation performance by evaluating SHAP in every recursion. The mean absolute SHAP value for each variable was calculated as a ranking criterion representing the importance of each variable. The light gradient boosting machine (LightGBM) (18), a gradient-boosted tree-based model that can handle categorical variables, was used for SHAP evaluation.

The SHAP method was used to investigate the local interpretability of the developed LightGBM model. Model predictions were visualized for true-positive, true-negative, false-positive, and false-negative cases.

2.5. Model development

A logistic regression with L2 regularization was used as the baseline comparator. We tested two representative machine learning models, namely, LightGBM and the multi-layer perceptron (MLP). MLP is a feedforward neural network with fully connected layers. Hyperparameters were tuned using Bayesian optimization (19) to maximize the predictive performance of cross-validation. All details pertaining to hyperparameter settings are described in the Supplementary Methods and Supplementary Table S2. Models were calibrated using isotonic regression on the validation data during cross-validation. All methods used in this study were implemented in Python version 3.9.7 using Scikit-learn version 1.1.2.

2.6. Internal validation in different subgroup cohorts

The model performance of internal validation was evaluated in different subgroup cohorts. Age (≤ 64 , 65–74, and ≥ 75 years), sex, hypertension, diabetes mellitus, type of AF (sustained and paroxysmal AF), and stroke recurrence (first-ever and recurrent stroke) were defined as subgroups.

2.7. Statistical analysis

Descriptive statistics were expressed as numbers (percentages), means (standard deviations), or medians (interquartile ranges). The Shapiro–Wilk test was used for normality, and Levene's test was used for homoscedasticity. The chi-squared test, independent *t*-test, or Mann–Whitney *U*-test was used for comparison.

The area under the receiver operating characteristic curve (AUROC) was used as the primary outcome metric, as well as a guiding metric in all model training processes, including variable selection (recursive feature elimination) and hyperparameter tuning (Bayesian optimization). The AUROC was calculated and compared between models using the DeLong method (20). Furthermore, we evaluated the area under the precision–recall curve (AUPRC) of the models as an overall performance measure.

The detailed performance of each model was evaluated in terms of sensitivity, positive predictive value, and negative predictive value with net reclassification improvement (NRI) at low false positive rates (FPRs) of 5, 10, and 20%. A positive NRI value indicates superior reclassification performance by the new model compared to that of the reference model (21). We evaluated sensitivity at fixed FPR levels in the subgroups using the model that exhibited the best overall performance. Calibration errors were evaluated before and after calibration using the calibration curve and Brier score, which is the mean-squared error of predicted probability (22). We evaluated the model's capability of risk stratification for the prediction of 3-month mortality. Four quartile strata were obtained from the prediction scores of the best-performing model, and survival curves were plotted using Kaplan–Meier estimation and compared using the log-rank test. The cutoff thresholds for investigating local interpretability were determined at an FPR level of 20%.

p-values were adjusted using the Benjamini–Hochberg method for multiple comparisons. The significance level was set at a *p*-value of < 0.05 .

3. Results

3.1. Baseline characteristics of study subjects

A total of 2,307 and 898 patients were included in the K-ATTENTION and KUSR groups, respectively (Figure 1). Table 1 presents a comparison between the two registries, including the general characteristics of the study participants, revealing an unfavorable functional outcome rate of 49.8% and the mortality rate of 11.6% 3 months following the index stroke. These results are compatible with unfavorable functional outcomes at 3 months

TABLE 1 General characteristics of study participants.

Variable	All (n = 3,125)	K-ATTENTION (n = 2,307)	KUSR (n = 818)	p value ^a
<i>Demographics</i>				
Age, years	75.0 [67.0–80.0]	74.0 [67.0–80.0]	76.0 [67.0–82.0]	< 0.001***
Male sex, n (%)	1,628 (53.0%)	1,222 (53.0%)	406 (53.0%)	1.000
BMI, kg/m ²	23.4 (3.4)	23.3 (3.4)	23.8 (3.5)	< 0.001***
<i>Initial clinical status</i>				
Initial DBP, mmHg	85.5 (15.7)	84.7 (14.9)	87.8 (17.6)	< 0.001***
Initial SBP, mmHg	145.5 (27.7)	144.3 (27.7)	148.9 (27.4)	< 0.001***
Initial pulse rate, beats per min	83.7 (21.4)	83.4 (21.4)	84.6 (21.4)	0.168
Initial NIHSS	7.0 [2.0–15.0]	7.0 [2.0–15.0]	6.0 [2.0–15.0]	0.412
<i>Pre-existing clinical status</i>				
Pre-stroke mRS	0 [0–1]	0 [0–1]	0 [0–3]	< 0.001***
CHADS ₂	2 [1–3]	2 [1–3]	2 [1–4]	< 0.001***
CHA ₂ DS ₂ VASc	3 [2–5]	3 [2–4]	4 [3–5]	< 0.001***
Stroke onset time, h	14.4 (18.0)	13.2 (6.3)	17.7 (33.8)	< 0.001***
<i>Pre-existing comorbidities</i>				
Stroke	916 (29.8%)	749 (32.5%)	167 (21.8%)	< 0.001***
Sustained AF	1,599 (52.0%)	1,174 (50.9%)	425 (55.5%)	0.030*
CHF	173 (5.6%)	95 (4.1%)	78 (10.2%)	< 0.001***
HTN	2,121 (69.0%)	1,587 (68.8%)	534 (69.7%)	0.665
DM	843 (27.4%)	603 (26.1%)	240 (31.3%)	0.006**
CAD	454 (14.8%)	332 (14.4%)	122 (15.9%)	0.327
PAD	41 (1.3%)	33 (1.4%)	8 (1.0%)	0.532
<i>Initial image findings</i>				
Lesion lateralization				< 0.001***
Rt. anterior	904 (29.4%)	817 (35.4%)	87 (11.4%)	
Lt. anterior	1,038 (33.8%)	798 (34.6%)	240 (31.3%)	
Posterior	670 (21.8%)	385 (16.7%)	285 (37.2%)	
Bilateral or diffuse multifocal	461 (15.0%)	307 (13.3%)	154 (20.1%)	
DWI lesion pattern				< 0.001***
Single corticosubcortical	669 (21.8%)	552 (23.9%)	117 (15.3%)	
Cortical	289 (9.4%)	201 (8.7%)	88 (11.5%)	
Subcortical (≥15 mm)	185 (6.0%)	152 (6.6%)	33 (4.3%)	
Subcortical (<15 mm)	190 (6.2%)	115 (5.0%)	75 (9.8%)	
Small scattered lesion in one vascular territory	330 (10.7%)	221 (9.6%)	109 (14.2%)	
Confluent and an additional lesion in one vascular territory	521 (17.0%)	405 (17.6%)	116 (15.1%)	
Multiple lesions in multiple vascular territories	646 (21.0%)	418 (18.1%)	228 (29.8%)	
Concomitant ICAS	1,466 (47.9%)	1,235 (53.5%)	231 (30.2%)	< 0.001***
Concomitant ECAS	611 (20.1%)	548 (23.8%)	63 (8.2%)	< 0.001***
<i>Laboratory findings</i>				
WBC, 10 ³ /μL	8.3 (3.2)	8.3 (3.1)	8.3 (3.4)	0.903
Hb, g/dL	13.5 (2.0)	13.5 (2.0)	13.5 (2.0)	0.876
PLT, 10 ³ /mL	205.0 (72.9)	205.0 (76.5)	204.8 (60.9)	0.929

(Continued)

TABLE 1 (Continued)

Variable	All (n = 3,125)	K-ATTENTION (n = 2,307)	KUSR (n = 818)	p value ^a
hs-CRP, mg/L	4.3 (16.0)	2.8 (11.3)	8.6 (24.6)	< 0.001***
Initial glucose, mg/dL	142.8 (72.5)	141.2 (76.6)	147.3 (58.8)	0.047*
Fasting glucose, mg/dL	122.4 (43.9)	121.9 (43.6)	123.7 (44.7)	0.335
HbA1c, %	6.1 (2.5)	6.1 (2.1)	6.3 (3.2)	0.018*
Total cholesterol, mg/dL	162.0 (38.9)	162.8 (37.8)	159.7 (41.6)	0.071
TG, mg/dL	100.0 (65.2)	97.6 (61.8)	106.5 (73.4)	0.001**
HDL, mg/dL	46.6 (14.5)	46.9 (14.9)	46.0 (13.3)	0.148
LDL, mg/dL	99.4 (33.6)	100.7 (33.2)	95.8 (34.7)	< 0.001***
AST, U/L	26.0 [21.0–33.0]	26.0 [21.0–33.0]	26.0 [21.0–33.0]	0.908
ALT, U/L	19.0 [14.0–27.0]	18.0 [14.0–26.0]	19.0 [14.0–28.0]	0.546
ALP, IU/L	93.4 (65.1)	99.0 (73.0)	79.6 (35.2)	< 0.001***
Total bilirubin, mg/dL	0.9 (0.5)	0.9 (0.5)	0.8 (0.5)	< 0.001***
Uric acid, mg/dL	5.4 (2.6)	5.5 (2.8)	5.1 (1.7)	0.003**
Serum creatinine, mg/dL	0.9 [0.7–1.1]	0.9 [0.7–1.1]	0.9 [0.8–1.2]	< 0.001***
CrCl, mL/min	64.8 [46.2–85.4]	61.4 [42.7–81.2]	75.5 [57.4–91.2]	< 0.001***
Fibrinogen, mg/dL	321.3 (120.9)	316.9 (125.0)	337.9 (102.9)	< 0.001***
Homocysteine, μmol/L	12.7 (36.8)	13.2 (43.4)	11.6 (6.0)	0.348
CK-MB, ng/mL	3.6 (6.4)	3.5 (6.4)	3.9 (6.4)	0.163
FFA, μEq/L	866.1 (467.7)	894.2 (557.1)	848.6 (401.2)	0.129
<i>Thrombolytic treatment</i>				
Recanalization therapy				< 0.001***
None	2,162 (70.4%)	1,641 (71.1%)	521 (68.0%)	
IV	505 (16.4%)	393 (17.0%)	112 (14.6%)	
IA	195 (6.3%)	133 (5.8%)	62 (8.1%)	
IV + IA	201 (6.5%)	140 (6.1%)	61 (8.0%)	
<i>3-month outcomes</i>				
Mortality	355 (11.6%)	294 (12.7%)	61 (8.0%)	< 0.001***
Unfavorable outcomes	1,530 (49.8%)	1,136 (49.2%)	394 (51.4%)	0.312

^ap-value of the chi-squared test, independent *t*-test, or Mann–Whitney U test for comparing the difference between the internal validation set (K-ATTENTION) and the external validation set (KUSR). **p* < 0.05, ***p* < 0.01, ****p* < 0.001.

Values are represented as *n* (%), mean (SD), or median [95% CIs]. BMI, body mass index; DBP, diastolic blood pressure; SBP, systolic blood pressure; NIHSS, National Institute of Health and Stroke Scale; mRS, modified Rankin Scale; AF, atrial fibrillation; CHF, congestive heart failure; DM, diabetes mellitus; CAD, coronary artery disease; PAD, peripheral artery disease; DWI, diffusion-weighted image; ICAS, intracranial artery stenosis; ECAS, extracranial artery stenosis; WBC, white blood cell; PLT, platelet; TG, triglyceride; HDL, high-density lipoprotein; LDL, low-density lipoprotein; CrCl, creatinine clearance; FFA, free fatty acid; IV, intravenous; IA, intra-arterial.

poststroke (*p* > 0.05), whereas the models significantly differed in terms of 3-month mortality (*p* < 0.05). The patients in the two registries were comparable in terms of sex, initial National Institute of Health and Stroke Scale (NIHSS) score, hypertension, and history of vascular diseases, including coronary and peripheral artery diseases. However, patients in the KUSR group were generally older and had higher initial blood pressure and body mass index (BMI) values than those in the K-ATTENTION group. Furthermore, patients in the KUSR group were more likely to exhibit a severe pre-stroke functional status, persistent AF, and a history of congestive heart failure and diabetes mellitus, whereas they had a lower previous history of stroke. The initial imaging findings revealed significant differences in lesion lateralization, diffusion-weighted imaging (DWI) lesion patterns, and concomitant intracranial and extracranial artery stenosis.

3.2. Model performance

An evaluation of model performance is presented in [Figure 2](#), indicating that the machine learning models outperform logistic regression in predicting both outcomes. In the prediction of unfavorable functional outcomes, MLP obtained an AUROC value of 0.890 on the internal validation set, representing a significant improvement over the AUROC value of 0.874 obtained by logistic regression. In the external validation set, both LightGBM (0.873) and MLP (0.859) achieved significantly higher AUROC values than those of logistic regression (0.834).

In the prediction of mortality, both LightGBM (0.839) and MLP (0.842) attained significantly higher AUROC values than logistic regression (0.803) on the internal validation set. On the external validation set, the AUROC values of LightGBM (0.805)

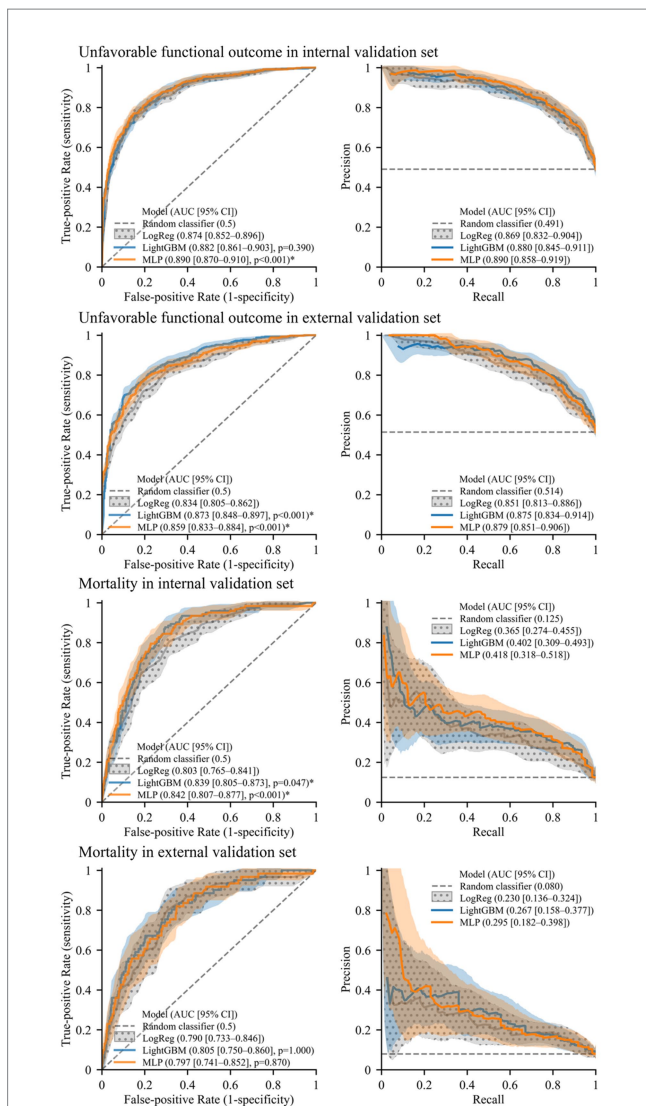


FIGURE 2 Receiver operating characteristics curves and precision–recall curves illustrating model performance. Solid lines and shades represent mean curves and 95% confidence interval areas, respectively. For the baseline model (logistic regression, “LogReg”), confidence intervals are represented with a polka dot pattern. An asterisk (*) indicates significantly higher AUROC than the baseline model ($p < 0.05$, Benjamini–Hochberg corrected).

and MLP (0.797) were also higher than that of logistic regression (0.790) although not significantly so. Furthermore, the machine learning models exhibited higher AUPRC values than those of logistic regression for predicting both outcomes in each validation set.

The cross-validation performance and calibration curves with the Brier scores of the models are presented in [Supplementary Table S3](#) and [Supplementary Figure S1](#), respectively. MLP was used to obtain four quartile strata, and mortality risk stratification was evaluated with pairwise comparisons between the survival curves of the strata using the log-rank test ([Supplementary Figure S2](#)). For the internal validation set, all p -values for pairwise comparisons were less than 0.0001, except for those between the first and second quartiles. For the external validation set, p -values for the pairwise comparisons were 0.0002 or less, except for those between adjacent quartiles.

3.3. Performance comparison according to subgroup

Model performance in predicting unfavorable functional outcomes was consistent across all subgroups, with the lowest AUROC and AUPRC values in patients with recurrent stroke and those aged <65 years, respectively. Both machine learning models exhibited comparable or superior performance to that of logistic regression, with MLP significantly outperforming logistic regression in most subgroups ($p < 0.05$) except for patients aged <65 and >74 years and those with recurrent strokes. No significant differences were observed between the two machine learning models or between logistic regression and LightGBM.

[Supplementary Figure S3](#) shows the results of a performance comparison in the prediction of 3-month mortality across subgroups. Performance was also consistent across subgroups, with machine learning models exhibiting superior performance for all subgroups except patients aged <65 years.

Low-FPR sensitivity results in the subgroup cohorts for predicting unfavorable functional outcomes and mortality are presented in [Supplementary Figures S4, S5](#), respectively. Detailed performance and NRI at low FPRs for the prediction of unfavorable functional outcomes and mortality are presented in [Supplementary Tables S4, S5](#), respectively (see [Figure 3](#)).

3.4. Importance of variables for prediction of unfavorable outcomes

Out of the 43 variables, 34 were selected in the variable selection process, with the 10 most important variables summarized in [Figure 4A](#). The most important variables were the initial NIHSS score, followed by DWI lesion pattern, pre-stroke mRS, and hs-CRP.

Partial SHAP dependence plots for the four representative variables among the top 10 are displayed in [Figure 4B](#), with those for the other six variables presented in [Supplementary Figure S6](#). Thus, ≥ 7.4 points in the initial NIHSS score, having a specific pattern of lesion, including single corticosubcortical lesion, confluent and an additional lesion in one vascular territory, or multiple lesions in multiple vascular territories, $\geq 8,700$ cells/ μL in white blood cell (WBC) count, ≥ 318.3 mg/dL in fibrinogen, ≤ 22.5 kg/m² in BMI, ≥ 3 in pre-stroke mRS, bilateral or diffuse multifocal lesion lateralization, ≤ 12.9 g/dL in hemoglobin, ≥ 74.3 years in age, and ≤ 51.8 mL/min in creatine clearance contributed to a higher risk of the unfavorable functional outcome. The SHAP values associated with the 10 most important variables exhibited a pattern comparable to linear association. WBC revealed a sigmoid pattern, fibrinogen displayed a J-shaped pattern, and other variables revealed complex linear, sigmoid, and J-shaped patterns.

[Supplementary Figure S7](#) presents the local interpretability of LightGBM for the prediction of unfavorable outcomes with individual cases on an external validation set.

3.5. Importance of variables for prediction of mortality

Out of the 16 selected variables, the 10 most important are summarized in [Figure 4C](#). The most important variable was the initial

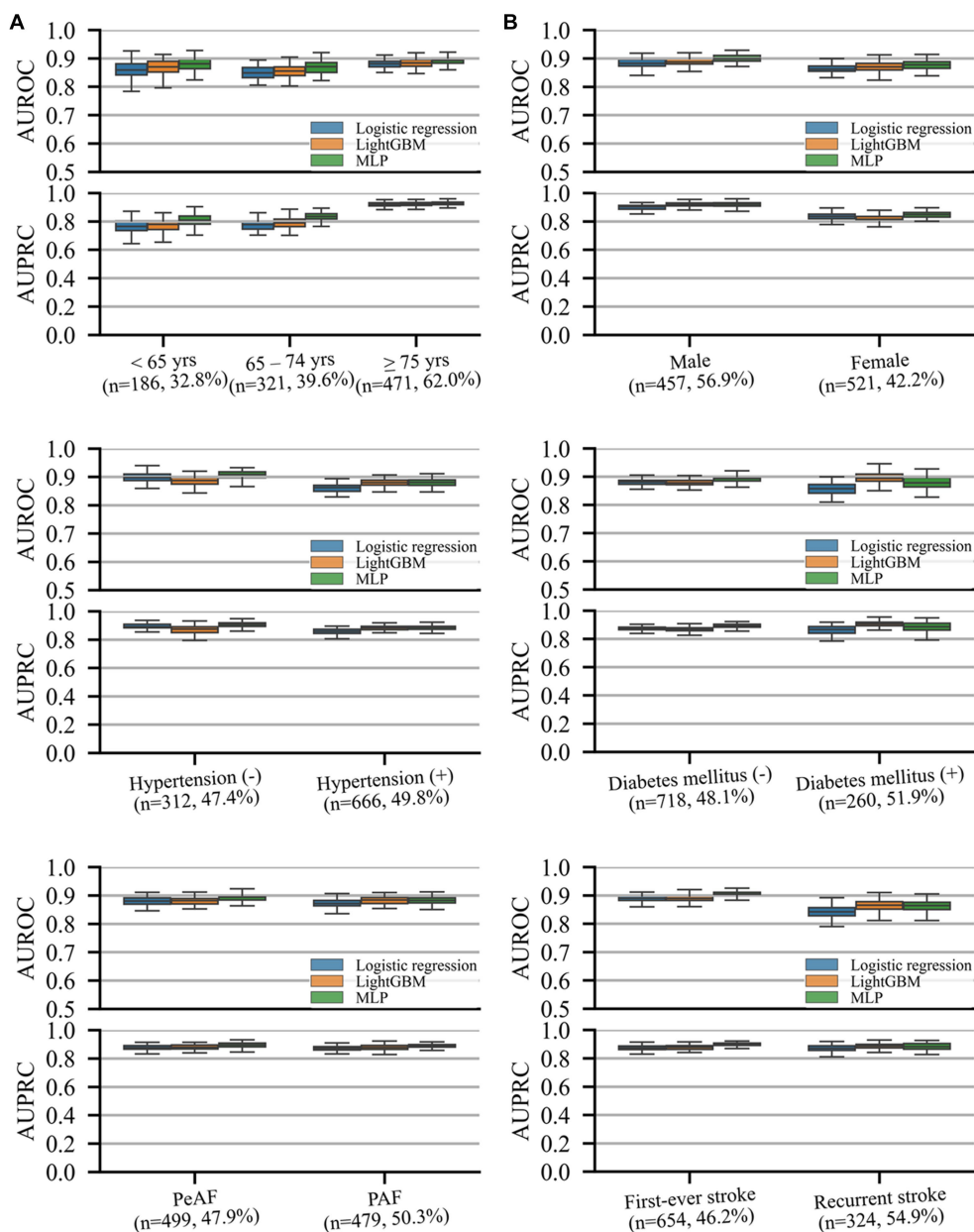


FIGURE 3
 AUROC and precision–recall curves representing performance on different subgroup cohorts for the prediction of unfavorable functional outcomes. The (*n* = A, B%) notation for each subgroup indicates the number of samples in the test set (A) and prevalence rate of the outcome (B) of the subgroup. Box plots are plotted with whiskers of 1.5 times the interquartile ranges. AUPRC, area under the precision–recall curve; AUROC, area under the receiver operating characteristics curve; DM, diabetic mellitus; HTN, hypertension; PAF, paroxysmal atrial fibrillation (AF); PeAF, persistent atrial fibrillation (AF).

NIHSS score, followed by age, concomitant intracranial/extracranial steno-occlusion, fasting glucose, and creatinine clearance.

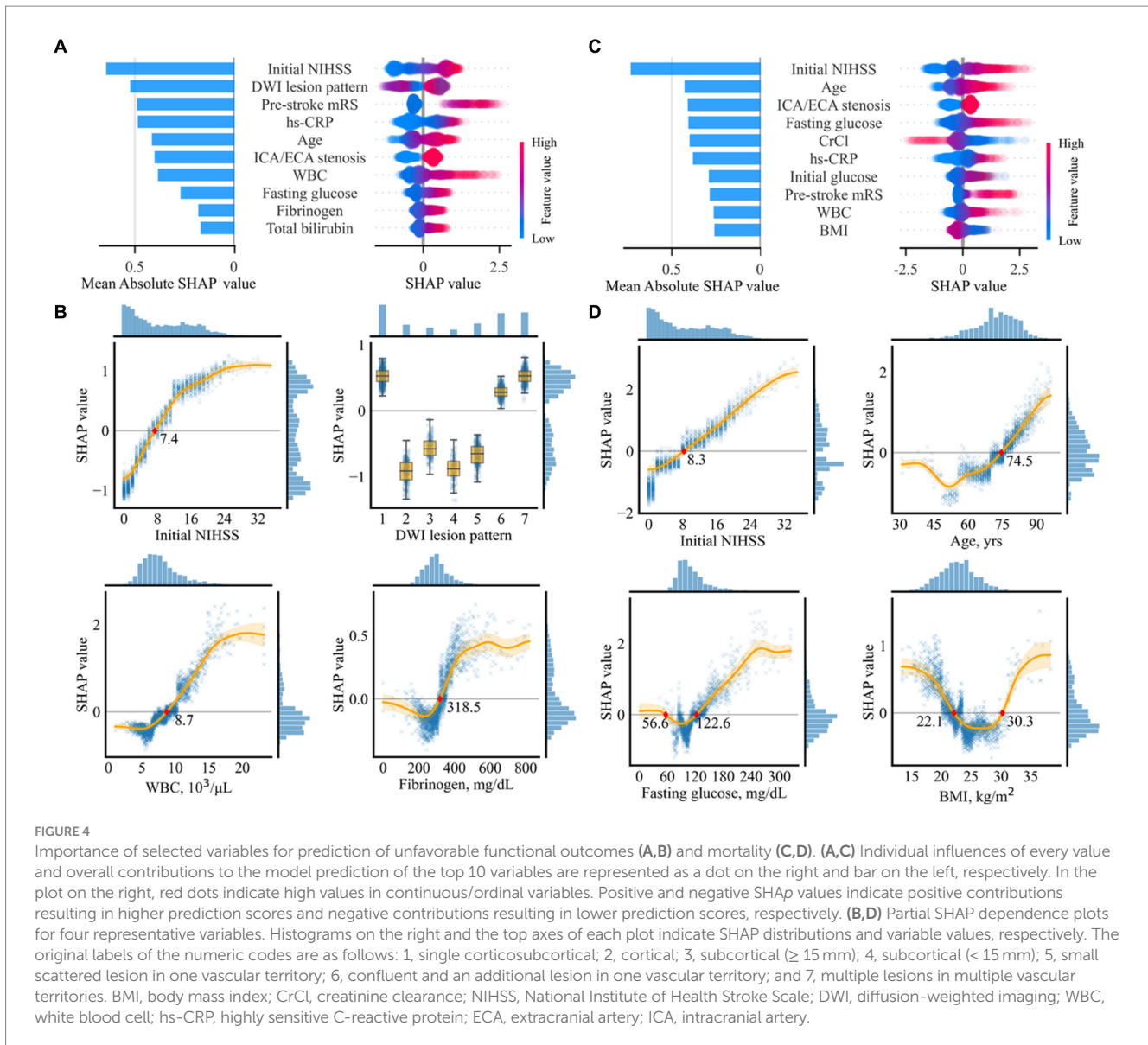
Partial SHAP dependence plots for four representative variables are displayed in Figure 4D, with those for the other six variables presented in Supplementary Figure S8. In summary, ≥ 8.2 in the initial NIHSS score, ≥ 74.5 years in age, ≤ 56.6 mg/dL or ≥ 122.6 mg/dL in fasting glucose, and ≤ 22.1 or ≥ 30.3 kg/m² in BMI, ≥ 4 in pre-stroke mRS, ≥ 87.2 IU/L in ALP, ≤ 4.0 mg/dL, or ≥ 7.0 mg/dL in uric acid, multiple lesions in multiple vascular territories, presence of concomitant intracranial or extracranial stenosis, and $\geq 8,300$ cells/ μ l in WBC count predicted mortality. Most association patterns between SHAP values and variables were near-linear, sigmoid, J-shaped, or

combinations of the three. BMI and uric acid levels revealed U-shaped patterns.

The local interpretability of LightGBM for the prediction of mortality is demonstrated in Supplementary Figure S9 by presenting individual cases through model prediction on the external validation set.

4. Discussion

We trained MLP and LightGBM machine learning models to predict unfavorable outcomes and mortality in AF-related stroke



patients over a 3-month period using two separate datasets. All models were validated internally and externally, with the machine learning models exhibiting higher predictive power than logistic regression for both outcomes. Similar trends were consistently observed across pre-specified subgroups, including age, sex, hypertension, diabetes mellitus, type of AF, and stroke recurrence. Overall, the machine learning models reliably predicted unfavorable outcomes and mortality in AF-related stroke patients. We identified influential variables through SHAP values to improve model explainability and identify high-risk patients with poor outcomes.

The initial NIHSS score, which reflects initial stroke severity, was the most influential variable with the highest SHAP value in determining short-term prognoses (unfavorable outcomes and mortality) following AF-related stroke. This finding is consistent with those of previous studies, which demonstrated an association between the initial NIHSS score and poor outcomes (10, 23) and mortality (8, 24) after the occurrence of stroke. Similarly, we observed linear associations between the initial NIHSS score and both poor functional outcomes and mortality, with cutoff values of 7.4 and 8.2, respectively.

Thus, the risk of mortality increased linearly with an initial NIHSS score exceeding 8.2.

Patient age was the second most significant variable affecting mortality, with a J-shaped association pattern. Patients aged 74.5 years and older exhibited a higher risk of mortality, which increased linearly with age. However, the magnitude of negative association decreased in patients aged 52 years and younger. The DWI lesion pattern was the second most influential variable for unfavorable functional outcomes. AF-related stroke patients exhibited a higher risk of poor functional outcomes when they had infarct patterns with single cortico-subcortical lesions, confluent and additional lesions in one vascular territory, or multiple lesions in different vascular territories. The size and number of ischemic lesions may indicate the burden of embolus, suggesting an association with functional outcomes.

Concomitant vascular diseases are frequently observed in AF patients because they share several risk factors and pathophysiological features with atherosclerosis (25, 26). Concomitant carotid atherosclerosis was identified as an important risk factor for short-term outcomes in this study. This result is consistent with previous

results, demonstrating that carotid atherosclerosis predicts recurrent vascular events and mortality in AF-related stroke patients (26). However, this study is the first to determine an association with poor functional outcomes. The pre-stroke mRS score, representing the degree of functional disability prior to the index stroke, is a well-known robust predictor of prognosis following stroke (27). We observed that a pre-stroke mRS of 3 or higher is associated with poor short-term functional outcomes, whereas the association with mortality increased almost linearly for each single-point increase in pre-stroke mRS of 4 or higher.

WBC and fibrinogen levels exhibited sigmoid patterns according to SHAP values. The leukocyte count, a marker of inflammatory response, is associated with short- and long-term clinical outcomes following acute strokes (28, 29). Fibrinogens play crucial roles in the coagulation cascade and inflammation. Increased fibrinogen levels are associated with functional outcomes (30, 31). Notably, the cutoff values for poor prognosis in this study were similar to those discovered previously. The association between fasting glucose levels revealed a J-shaped sigmoidal pattern. An association with lower mortality rates was also observed with glucose levels ranging from 56.6 to 122.6 mg/dL. The lower and upper cutoff values were comparable to the blood glucose levels for hypoglycemia and diagnostic criteria for diabetes mellitus, respectively. Hyperglycemia may contribute to poor outcomes in stroke patients through several mechanisms, including an increased risk of cerebral edema (32, 33), impaired blood flow regulation (34), and increased oxidative stress (35). Hypoglycemia may also result in poor outcomes owing to impaired brain function as the brain requires a constant supply of glucose. Additionally, hypoglycemia can induce ischemic stroke by increasing the levels of inflammatory markers, platelet activation, and fibrinogen formation (36, 37).

Machine learning may provide significant advantages in medical practice by uncovering complex non-linear patterns within medical data and enhancing predictive accuracy vital for optimizing patient care. Machine learning models are also adept at tailoring predictions to individual patient profiles, aligning with the principles of personalized medicine. Moreover, they serve as powerful clinical decision support tools, providing data-driven insights that enhance the decision-making capabilities of healthcare practitioners, ultimately improving patient outcomes.

The machine learning models achieved robustness through internal and external validation sets based on two separate datasets with distinct patient characteristics. However, this study had some limitations. First, our models may have been overfitted to the Korean population, which would present a challenge in generalizability. The enrolled population was Asian, and most patients were treated under the Korean medical system covered by national health insurance, which improved the accessibility of medical services and standardization of treatment processes. Asian populations have been associated with a higher bleeding tendency than thrombotic risk compared to Western populations (38). Ethnic differences in fibrinogen levels, one of the 10 highest SHAP variables, have also been reported (39, 40). However, these national and ethnic differences cannot be considered by machine learning models. Second, some variables associated with outcomes in AF-related stroke (41, 42), including D-dimer and N-terminal pro-B-type natriuretic peptides, were not considered, owing to an excessive number of missing values. To improve model interpretability, we used categorized ischemic

lesion patterns in place of raw brain magnetic scans. However, this approach may limit the utilization of potential information embedded in magnetic resonance images. Finally, although the machine learning models exhibited higher predictive power than logistic regression, a comparison in terms of AUROC and AUPRC did not reveal marked differences and these differences were statistically significant. Small differences were observed in terms of sensitivity at low FPR levels as each of the three models assumed the same number of variables. As displayed in Figure 4, most of the association patterns revealed roughly linear contributions. Generally, linear regression is a simple and powerful tool for modeling linear relations, whereas machine learning models can capture complex non-linear relationships between variables. Consequently, if the relationship between variables is primarily linear, the additional complexity of a machine learning model might not be conducive in terms of predictive accuracy. To improve model performance, artificial data generation techniques, such as synthetic minority oversampling, may be useful (43). The prevalence of outcomes for future model applications was not artificially increased, instead allowing the class imbalance to remain unmodified. To ensure that comparable performance can be expected when the models are applied to other independent datasets, we handled outliers using a popular automated method. Furthermore, we verified model performance in terms of AUPRC, with a more accurate reflection of practical performance than the AUROC (44).

Machine learning models can be used to predict short-term outcomes, including unfavorable outcomes and mortality over a 3-month period, and identify high-risk patients with poor outcomes in AF-related stroke. The initial NIHSS score was the most important factor influencing short-term prognosis. Because our results are restricted to Korean stroke patients, further validation is necessary to ensure that the models and selected features can be applied to all AF-related stroke patients.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Korea University Ansan Hospital IRB. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because this is a retrospective study.

Author contributions

W-KS and J-MJ designed the study and collected the data. E-TJ completed the machine learning and statistical analysis. E-TJ and SJ wrote the manuscript. J-MJ is responsible for the overall content as the guarantor. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the Korean Society of Neurosonology Grant and the K-Brain Project of the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. RS-2023-00265393). The funders had no role in the study design; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Conflict of interest

J-MJ has received lecture honoraria from Pfizer, Sanofi-Aventis, Ostuka, Dong-A, and Hanmi Pharmaceutical Co., Ltd. and consulting fees from Daewoong Pharmaceutical Co., Ltd. W-KS received honoraria for lectures from Pfizer, Sanofi-Aventis, Otsuka Korea, and Dong-A Pharmaceutical Co., Ltd.

References

- Alberts M, Chen YW, Lin JH, Kogan E, Twyman K, Milentijevic D. Risks of stroke and mortality in atrial fibrillation patients treated with rivaroxaban and warfarin. *Stroke*. (2020) 51:549–55. doi: 10.1161/STROKEAHA.119.025554
- Lahewala S, Arora S, Patel P, Kumar V, Patel N, Tripathi B, et al. Atrial fibrillation: utility of CHADS2 and CHA2DS2-VASc scores as predictors of readmission, mortality and resource utilization. *Int J Cardiol*. (2017) 245:162–7. doi: 10.1016/j.ijcard.2017.06.090
- Sato S, Yazawa Y, Itabashi R, Tsukita K, Fujiwara S, Furui E. Pre-admission CHADS2 score is related to severity and outcome of stroke. *J Neurol Sci*. (2011) 307:149–52. doi: 10.1016/j.jns.2011.04.018
- Tanaka K, Yamada T, Torii T, Furuta K, Matsumoto S, Yoshimura T, et al. Pre-admission CHADS2, CHA2DS2-VASc, and R2CHADS2 scores on severity and functional outcome in acute ischemic stroke with atrial fibrillation. *J Stroke Cerebrovasc Dis*. (2015) 24:1629–35. doi: 10.1016/j.jstrokecerebrovasdis.2015.03.036
- Tu HT, Campbell BC, Meretoja A, Churilov L, Lees KR, Donnan GA, et al. Pre-stroke CHADS2 and CHA2DS2-VASc scores are useful in stratifying three-month outcomes in patients with and without atrial fibrillation. *Cerebrovasc Dis*. (2013) 36:273–80. doi: 10.1159/000353670
- Acciarresi M, Paciaroni M, Agnelli G, Falocci N, Caso V, Becattini C, et al. Pre-stroke CHA2DS2-VASc score and severity of acute stroke in patients with atrial fibrillation: findings from RAF study. *J Stroke Cerebrovasc Dis*. (2017) 26:1363–8. doi: 10.1016/j.jstrokecerebrovasdis.2017.02.011
- Henriksson KM, Farahmand B, Johansson S, Asberg S, Terént A, Edvardsson N. Survival after stroke—the impact of CHADS2 score and atrial fibrillation. *Int J Cardiol*. (2010) 141:18–23. doi: 10.1016/j.ijcard.2008.11.122
- Li S, Zhao X, Wang C, Liu L, Liu G, Wang Y, et al. Risk factors for poor outcome and mortality at 3 months after the ischemic stroke in patients with atrial fibrillation. *J Stroke Cerebrovasc Dis*. (2013) 22:e419–25. doi: 10.1016/j.jstrokecerebrovasdis.2013.04.025
- Yu I, Song TJ, Kim BJ, Heo SH, Jung JM, Oh KM, et al. CHADS2, CHA2DS2-VASc, ATRIA, and Essen stroke risk scores in stroke with atrial fibrillation: a nationwide multicenter registry study. *Medicine*. (2021) 100:e24000. doi: 10.1097/MD.00000000000024000
- Maruyama K, Uchiyama S, Shiga T, Iijima M, Ishizuka K, Hoshino T, et al. Brain natriuretic peptide is a powerful predictor of outcome in stroke patients with atrial fibrillation. *Cerebrovasc Dis Extra*. (2017) 7:35–43. doi: 10.1159/000457808
- Kim SH, Jeon ET, Yu S, Oh K, Kim CK, Song TJ, et al. Interpretable machine learning for early neurological deterioration prediction in atrial fibrillation-related stroke. *Sci Rep*. (2021) 11:20610. doi: 10.1038/s41598-021-99920-7
- Lee S, Park HJ, Hwang J, Lee SW, Han KS, Kim WY, et al. Machine learning-based models for prediction of critical illness at community, paramedic, and hospital stages. *Emerg Med Int*. (2023) 2023:1221704. doi: 10.1155/2023/1221704
- Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Circulation*. (2015) 131:211–9. doi: 10.1161/CIRCULATIONAHA.114.014508
- Van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *J Stat Soft*. (2011) 45:1–67. doi: 10.18637/jss.v045.i03
- Liu FT, Ting KM, Zhou Z-H. Isolation-based anomaly detection. *ACM Trans Knowl Discov Data*. (2012) 6:1–39. doi: 10.1145/2133360.2133363
- Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. (2020) 2:56–67. doi: 10.1038/s42256-019-0138-9
- Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Mach Learn*. (2002) 46:389–422. doi: 10.1023/A:1012487302797
- Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. Light GBM: a highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst*. (2017) 30:3146–54. doi: 10.5555/3294996.3295074
- Snoek J, Larochelle H, Adams RP. Practical Bayesian optimization of machine learning algorithms. *Adv Neural Inf Process Syst*. (2012) 25:25. doi: 10.5555/2999325.2999464
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. (1988) 44: 837–845. doi: 10.2307/2531595
- Leening MJ, Vedder MM, Witteman JC, Pencina MJ, Steyerberg EW. Net reclassification improvement: computation, interpretation, and controversies: a literature review and clinician's guide. *Ann Intern Med*. (2014) 160:122–31. doi: 10.7326/M13-1522
- Rufibach K. Use of brier score to assess binary predictions. *J Clin Epidemiol*. (2010) 63:938–9. doi: 10.1016/j.jclinepi.2009.11.009
- Choi JH, Cha JK, Huh JT. Adenosine diphosphate-induced platelet aggregation might contribute to poor outcomes in atrial fibrillation-related ischemic stroke. *J Stroke Cerebrovasc Dis*. (2014) 23:e215–20. doi: 10.1016/j.jstrokecerebrovasdis.2013.10.011
- Smith EE, Shobha N, Dai D, Olson DM, Reeves MJ, Saver JL, et al. Risk score for in-hospital ischemic stroke mortality derived and validated within the get with the guidelines-stroke program. *Circulation*. (2010) 122:1496–504. doi: 10.1161/CIRCULATIONAHA.109.932822
- Jover E, Marin F, Roldán V, Montoro-García S, Valdés M, Lip GY. Atherosclerosis and thromboembolic risk in atrial fibrillation: focus on peripheral vascular disease. *Ann Med*. (2013) 45:274–90. doi: 10.3109/07853890.2012.732702
- Lehtola H, Airaksinen KEJ, Hartikainen P, Hartikainen JEK, Palomäki A, Nuotio I, et al. Stroke recurrence in patients with atrial fibrillation: concomitant carotid artery stenosis doubles the risk. *Eur J Neurol*. (2017) 24:719–25. doi: 10.1111/ene.13280
- Quinn TJ, Taylor-Rowan M, Coyte A, Clark AB, Musgrave SD, Metcalf AK, et al. Pre-stroke modified Rankin scale: evaluation of validity, prognostic accuracy, and association with treatment. *Front Neurol*. (2017) 8:275. doi: 10.3389/fneur.2017.00275
- Nardi K, Milia P, Eusebi P, Paciaroni M, Caso V, Agnelli G. Admission leukocytosis in acute cerebral ischemia: influence on early outcome. *J Stroke Cerebrovasc Dis*. (2012) 21:819–24. doi: 10.1016/j.jstrokecerebrovasdis.2011.04.015
- Quan K, Wang A, Zhang X, Wang Y. Leukocyte count and adverse clinical outcomes in acute ischemic stroke patients. *Front Neurol*. (2019) 10:1240. doi: 10.3389/fneur.2019.01240
- Swarowska M, Ferens A, Pera J, Slowik A, Dziedzic T. Can prediction of functional outcome after stroke be improved by adding fibrinogen to prognostic model? *J Stroke Cerebrovasc Dis*. (2016) 25:2752–5. doi: 10.1016/j.jstrokecerebrovasdis.2016.07.029

The remaining 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.

Publisher's note

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

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2023.1243700/full#supplementary-material>

31. Swarowska M, Janowska A, Polczak A, Klimkowicz-Mrowiec A, Pera J, Slowik A, et al. The sustained increase of plasma fibrinogen during ischemic stroke predicts worse outcome independently of baseline fibrinogen level. *Inflammation*. (2014) 37:1142–7. doi: 10.1007/s10753-014-9838-9
32. Alvarez-Sabin J, Molina CA, Ribó M, Arenillas JF, Montaner J, Huertas R, et al. Impact of admission hyperglycemia on stroke outcome after thrombolysis: risk stratification in relation to time to reperfusion. *Stroke*. (2004) 35:2493–8. doi: 10.1161/01.STR.0000143728.45516.c6
33. Parsons MW, Barber PA, Desmond PM, Baird TA, Darby DG, Byrnes G, et al. Acute hyperglycemia adversely affects stroke outcome: a magnetic resonance imaging and spectroscopy study. *Ann Neurol*. (2002) 52:20–8. doi: 10.1002/ana.10241
34. Gray CS, Hildreth AJ, Sandercock PA, O'Connell JE, Johnston DE, Cartlidge NE, et al. Glucose-potassium-insulin infusions in the management of post-stroke hyperglycaemia: the UK glucose insulin in stroke trial (GIST-UK). *Lancet Neurol*. (2007) 6:397–406. doi: 10.1016/S1474-4422(07)70080-7
35. Chen H, Yoshioka H, Kim GS, Jung JE, Okami N, Sakata H, et al. Oxidative stress in ischemic brain damage: mechanisms of cell death and potential molecular targets for neuroprotection. *Antioxid Redox Signal*. (2011) 14:1505–17. doi: 10.1089/ars.2010.3576
36. Smith L, Chakraborty D, Bhattacharya P, Sarmah D, Koch S, Dave KR. Exposure to hypoglycemia and risk of stroke. *Ann N Y Acad Sci*. (2018) 1431:25–34. doi: 10.1111/nyas.13872
37. Collins R, Tagliaferri AR, Lobue G, Meng W, Ismail M. Hypoglycemia-induced basal ganglia infarct: a rare case of metformin toxicity in a Hemodialysis patient. *Cureus*. (2022) 14:e32449. doi: 10.7759/cureus.32449
38. Jung SJ, Shim SR, Kim BJ, Jung JM. Antiplatelet regimens for Asian patients with ischemic stroke or transient ischemic attack: a systematic review and network meta-analysis. *Ann Transl Med*. (2021) 9:753. doi: 10.21037/atm-20-7951
39. Cook DG, Cappuccio FP, Atkinson RW, Wicks PD, Chitolie A, Nakandakare ER, et al. Ethnic differences in fibrinogen levels: the role of environmental factors and the beta-fibrinogen gene. *Am J Epidemiol*. (2001) 153:799–806. doi: 10.1093/aje/153.8.799
40. Kain K, Blaxill JM, Catto AJ, Grant PJ, Carter AM. Increased fibrinogen levels among south Asians versus whites in the United Kingdom are not explained by common polymorphisms. *Am J Epidemiol*. (2002) 156:174–9. doi: 10.1093/aje/kwf017
41. Ohara T, Farhoudi M, Bang OY, Koga M, Demchuk AM. The emerging value of serum D-dimer measurement in the work-up and management of ischemic stroke. *Int J Stroke*. (2020) 15:122–31. doi: 10.1177/1747493019876538
42. Shibasaki K, Kimura K, Iguchi Y, Aoki J, Sakai K, Kobayashi K. Plasma brain natriuretic peptide predicts death during hospitalization in acute ischaemic stroke and transient ischaemic attack patients with atrial fibrillation. *Eur J Neurol*. (2011) 18:165–9. doi: 10.1111/j.1468-1331.2010.03101.x
43. Xu Z, Shen D, Kou Y, Nie T. A synthetic minority oversampling technique based on Gaussian mixture model filtering for imbalanced data classification. *IEEE Trans Neural Netw Learn Syst*. (2022) 1–14. doi: 10.1109/TNNLS.2022.3197156
44. Davis J, Goadrich M. The relationship between precision-recall and ROC curves In: *Proceedings of the 23rd international conference on machine learning*. Association for Computing Machinery (2006). 233–40.