Received: 9 February 2022 | Accepted: 11 April 2022

DOI: 10.1111/ene.15356

european journal of neurology
the official journal of the european academy of neurology

ORIGINAL ARTICLE

Classification of patients with embolic stroke of undetermined source into cardioembolic and non-cardioembolic profile subgroups

Max Christian Martin[1] | Thorsten Sichtermann[2] | Kolja Schürmann[1] | Pardes Habib[1,3] | Martin Wiesmann[2] | Jörg B. Schulz[1,3] | Omid Nikoubashman[2] | João Pinho[1] | Arno Reich[1]

[1]Department of Neurology, University Hospital RWTH Aachen, Aachen, Germany

[2]Department of Diagnostic and Interventional Neuroradiology, University Hospital RWTH Aachen, Aachen, Germany

[3]JARA-Institute Molecular Neuroscience and Neuroimaging, Forschungszentrum Jülich, Jülich, Germany

Correspondence
João Pinho. Department of Neurology, University Hospital RWTH Aachen, Pauwelsstr. 30, 52062 Aachen, Germany.
Email: jferreiradep@ukaachen.de

Abstract

Background and purpose: It is currently thought that embolic stroke of undetermined source (ESUS) has diverse underlying hidden etiologies, of which cardioembolism is one of the most important. The subgroup of patients with this etiology could theoretically benefit from oral anticoagulation, but it remains unclear if these patients can be correctly identified from other ESUS subgroups and which markers should be used. We aimed to determine whether a machine-learning (ML) model could discriminate between ESUS patients with cardioembolic and those with non-cardioembolic profiles using baseline demographic and laboratory variables.

Methods: Based on a prospective registry of consecutive ischemic stroke patients submitted to acute revascularization therapies, an ML model was trained using the age, sex and 11 selected baseline laboratory parameters of patients with known stroke etiology, with the aim of correctly identifying patients with cardioembolic and non-cardioembolic etiologies. The resulting model was used to classify ESUS patients into those with cardioembolic and those with non-cardioembolic profiles.

Results: The ML model was able to distinguish patients with known stroke etiology into cardioembolic or non-cardioembolic profile groups with excellent accuracy (area under the curve = 0.82). When applied to ESUS patients, the model classified 40.3% as having cardioembolic profiles. ESUS patients with cardioembolic profiles were older, more frequently female, more frequently had hypertension, less frequently were active smokers, had higher CHA_2DS_2-VASc (Congestive heart failure or left ventricular systolic dysfunction, Hypertension, Age ≥ 75 [doubled], Diabetes, Stroke/transient ischemic attack [doubled], Vascular disease, Age 65–74, and Sex category) scores, and had more premature atrial complexes per hour.

Conclusions: An ML model based on baseline demographic and laboratory variables was able to classify ESUS patients into cardioembolic or non-cardioembolic profile groups and predicted that 40% of the ESUS patients had a cardioembolic profile.

João Pinho and Arno Reich contributed equally.

KEYWORDS
cardioembolism, embolic stroke of undetermined source, ischemic stroke, machine learning, stroke

INTRODUCTION

Embolic ischemic strokes of undetermined source (ESUS), also known as embolic strokes without an identified cause, impose a great burden on both individuals and society, and constitute 10%–25% of all ischemic strokes [1,2]. ESUS patients are reported to have experienced less severe strokes on admission, but are significantly younger than patients with other stroke subtypes, and therefore their quality of life and productivity are reduced the longest by stroke [1,3].

The stroke recurrence rate of ESUS patients (6.8 per 100 patient-years) is similar to that of cardioembolic stroke patients (7.0 per 100 patient-years) and higher than in all other stroke subtypes [4].

Hart et al. [5] introduced the acronym ESUS for embolic cryptogenic strokes in 2014, with the assumption that this group of patients could benefit from anticoagulation irrespective of the origin of the thrombus. However, in both the NAVIGATE-ESUS and RE-SPECT ESUS trials, no benefit of anticoagulation compared to antiplatelet monotherapy was found in terms of recurrent ischemic stroke risk in ESUS patients [6,7]. It became clear that ESUS represents a heterogenous group of patients in terms of stroke etiology, and that the identification of ESUS subgroups could allow tailored therapy for prevention of recurrent stroke [8]. The concept of atrial cardiopathy as the anatomical and functional substrate for atrial fibrillation has emerged as a possible cardioembolic cause of stroke or as a marker for increased risk of future atrial fibrillation detection [9]. Indeed, a secondary analysis of the NAVIGATE-ESUS trial showed that anticoagulation was superior to acetylsalicylic acid in a subset of ESUS patients who demonstrated moderate to severe left atrial enlargement [10], which is considered one of the markers for atrial cardiopathy.

There is also increasing evidence that non-stenosing atherosclerotic carotid plaques may play a role in the etiology of stroke in some ESUS patients [11–19], Hence, the identification of clinical, laboratory and imaging variables which may reveal the underlying etiologies is of paramount importance.

The aim of this study was to analyze whether a machine-learning (ML) model using a set of baseline demographic and laboratory variables could differentiate between cardioembolic and non-cardioembolic etiologies in patients with known stroke cause, and if it could be used to further classify subgroups of patients with ESUS.

METHODS

We conducted a retrospective analysis based on a local prospective registry of consecutive acute ischemic stroke patients who were admitted to our comprehensive stroke center during the period between March 2019 and October 2020. The beginning of the study period corresponds to the time at which our routine laboratory stroke diagnostic profile was adjusted and expanded to include more variables, such as N-terminal pro-brain natriuretic peptide (NTproBNP). Our analysis was approved by the Ethics Committee of the Medical Faculty of the RWTH Aachen University (approval reference 335/15).

We selected all adult patients who fulfilled the following inclusion criteria: (i) acute ischemic stroke; (ii) being submitted to intravenous thrombolysis (IVT) and/or endovascular stroke treatment (EVT); and (iii) having received a complete etiological investigation according to the definition of Hart et al. [5], including transthoracic echocardiography, brain computed tomography (CT) or magnetic resonance imaging (MRI), 12-lead electrocardiogram (ECG), cardiac monitoring for at least 24 h with automatic rhythm detection and imaging of the extracranial and intracranial arteries supplying the area of brain ischemia.

Exclusion criteria were as follows: (i) stroke of undetermined etiology because of the simultaneous presence of two or more major causes for stroke; (ii) acute lacunar infarct; (iii) small vessel disease as stroke etiology. The etiology of ischemic stroke was determined according to the Trial of Org 10172 in Acute Stroke Treatment (TOAST) [20] classification and according to the ESUS definition proposed by Hart et al. (non-lacunar ischemic stroke; absence of extracranial or intracranial atherosclerosis causing ≥50% luminal stenosis in arteries supplying the area of ischemia; no major-risk cardioembolic source of embolism; no other specific cause of stroke identified) [5] Acute lacunar infarcts were defined according to the criteria proposed by Hart et al. [5]

Data collection

Demographic information, comorbidities, vascular risk factors, baseline National Institutes of Health Stroke Scale (NIHSS) score and acute stroke treatments were collected from our local stroke registry.

Stroke etiology was reviewed by at least two independent experienced vascular neurologists in every case and classified as cardioembolic, non-cardioembolic or ESUS after review of individual patient records. Data on the following baseline laboratory blood variables were collected from the individual patient records: NTproBNP; high-sensitivity Troponin T; lactate dehydrogenase (LDH); D-dimers; glycated hemoglobin (HbA1c); low-density lipoprotein (LDL); high-density lipoprotein (HDL); C-reactive protein (CRP); glomerular filtration rate (GFR); red cell distribution width; and hematocrit. In our center, blood samples for routine laboratory stroke diagnostic are collected during the first 24–48 h after admission.

The presence of acute infarcts in multiple arterial territories and of old embolic infarcts in other arterial territories was collected after review of the CT or MRI scans in the acute phase. The presence of left atrial dilation was assessed from the reports of transthoracic or transesophageal echocardiograms. The number of atrial premature

complexes per hour was calculated based on the reports of 24- and 72-h Holter ECG.

Machine learning

Using data from cardioembolic and non-cardioembolic stroke patients, different ML models were trained and validated with the aim of distinguishing cardioembolic from non-cardioembolic stroke patients. For the development of the models, only a set of 11 selected baseline laboratory variables, as well as age and sex, were used (Table 1).

We used the PyCaret ML library for comparing a total number of 18 individual ML algorithms. For each model, data on 80% of the patients with known cardioembolic and non-cardioembolic stroke etiology were randomly selected as the training set, whereas data on the remaining 20% of patients subsequently formed the validation set. The validation group was used to test the performance of the created models for distinguishing between cardioembolic and non-cardioembolic stroke etiologies.

TABLE 1 Baseline characteristics of patients without embolic stroke of undetermined source, stratified by stroke etiology

	Cardioembolic etiology ($n = 183$)	Non-cardioembolic etiology ($n = 136$)	p
Median age, years (IQR)	81.0 (75.0–85.0)	67.5 (59.3–78.8)	<0.001
Female sex, *n* (%)	104 (56.8)	45 (33.1)	<0.001
Comorbidities, *n* (%)			
Hypertension	160 (87.4)	108 (79.4)	0.053
Diabetes mellitus	56 (30.6)	48 (35.3)	0.376
Dyslipidemia	86 (47.0)	62 (45.6)	0.763
Current smoking	17 (9.3)	40 (29.4)	<0.001
Heart failure	30 (16.4)	6 (4.4)	<0.001
Coronary heart disease or myocardial infarction	63 (34.6)	26 (19.1)	0.002
Peripheral artery disease	23 (12.6)	13 (9.6)	0.401
Median CHA_2DS_2-VASc score (IQR)	6 (5–7)	5 (4–6)	<0.001
Median baseline NIHSS score (IQR)	12 (7–17)	9 (4–15)	0.013
Hyperacute revascularization therapies, *n* (%)			
Intravenous thrombolysis	98 (53.6)	75 (55.1)	0.777
Endovascular treatment	137 (74.9)	101 (74.3)	0.903
Imaging characteristics, *n* (%)			
Acute infarcts in multiple arterial territories	25 (13.7)	16 (11.8)	0.617
Old embolic infarcts in other arterial territories	40 (21.9)	19 (14.0)	0,073
Admission laboratory variables			
Mean hematocrit, % (SD)	35.6 (±5.6)	36.6 (±6.1)	0.129
Median red cell distribution width, % (IQR)	13.8 (13.2–15.0)	13.5 (12.8–14.3)	0.001
Median D-Dimers, ng/ml (IQR)	3525 (1444–15212)	2133 (1114– 6193)	0.033
Median HbA1c, % (IQR)	5.8 (5.5–6.4)	5.8 (5.4–6.7)	0.748
Median HDL cholesterol, mg/dl (IQR)	43 (35–53)	41 (33–48)	0.048
Median LDL cholesterol, mg/dl (IQR)	95 (71–123)	105 (76–128)	0.080
Median LDH, U/L (IQR)	225 (194–274)	183 (165–219)	<0.001
Median high-sensitivity troponin T, pg/ml (IQR)	23 (14–46)	16 (8–30)	<0.001
Median NTproBNP, pg/ml (IQR)	1714 (842– 3546)	262 (125– 855)	<0.001
Median GFR, ml/min/1.73 m^2 (IQR)	69.4 (50.3– 81.7)	81.8 (60.0–96.3)	<0.001
Median CRP, mg/L (IQR)	7.6 (2.6–21.2)	3.8 (1.9–11.8)	0.007

Abbreviations: CRP, C-reactive protein; GFR, glomerular filtration rate; HbA1c, glycated hemoglobin; HDL, high-density lipoprotein; IQR, interquartile range; LDH, lactate dehydrogenase; LDL, low-density lipoprotein; NIHSS, National Institutes of Health Stroke Scale; NTproBNP, N-terminal pro-brain natriuretic peptide.

The process of selecting the best-performing model was accomplished using 10-fold cross-validation, whereby the training and testing steps mentioned above were carried out 10 times. For each execution, the 80% of patients used for training and the 20% designated for testing were randomly chosen.

Key performance indicators, such as area under the curve (AUC) and classification accuracy, were obtained for each model by calculating the arithmetic mean of the values of all 10 executions. AUC was chosen as the most important indicator due to its wide application in the evaluation of medical diagnostic tests [21]. On comparison of the AUCs, the CatBoost classifier returned the best results and was therefore used in the next step to classify ESUS patients into cardioembolic or non-cardioembolic profiles.

CatBoost classifier

We implemented categorical boosting (CatBoost), which is a competitive variant of gradient boosting, a supervised ML technique [22]. Gradient boosting iteratively constructs a set of functions out of training data to minimize the loss function, which is achieved by performing gradient descent in a functional space. The base models are decision trees trained on different splits of the dataset that are combined into one strong classifier with an ensemble technique [22].

CatBoost uses ordered boosting, a modification of standard gradient boosting, to prevent the so-called prediction shift during computation of the gradients, thereby improving the generalization ability on the training set. This is implemented by generating random permutations of the training dataset at the beginning and sampling one of the permutations for constructing a decision tree in each iteration.

CatBoost implements conversion of categorical features into numerical values based on the ordering principle. Category values are substituted by numerical values with the help of an introduced permutation variable based on the foregoing training iterations. In doing so, the whole training dataset can be used for the conversion, moreover, reducing prediction shift [22].The hyperparameters of the CatBoost Classifier were as follows: learning rate = 0.006, maximal depth = 6, subsample rate = 0.8, L2 regularization = 3, iterations = 1000.

Group comparisons

We compared the characteristics of patients in the cardioembolic and non-cardioembolic stroke etiology subgroups using chi-squared tests for categorical variables, Mann–Whitney *U*-tests for continuous variables, which had a non-normal distribution, and unpaired *T*-tests for the normally distributed continuous variable hematocrit. Using the same tests, we also compared the groups of ESUS patients classified as having cardioembolic and non-cardioembolic profiles. The level of significance was set at an alpha value of 0.05. Statistical analyses were performed with IBM SPSS Statistics 27 software (IBM Corp.).

RESULTS

During the study period, 1885 patients with acute ischemic stroke were admitted to our hospital, among these, 543 patients received IVT and/or EVT. We excluded 46 patients with lacunar stroke and small vessel disease, 41 patients with simultaneous cardioembolic and non-cardioembolic cause of stroke and eight patients with incomplete etiological investigation (Figure 1). Hence, our study population consisted of 448 patients with a median (interquartile range [IQR]) age of 76 (64–83) years and a median (IQR) NIHSS score of 9 (4–15). A total of 46.7% of patients were female, 36.2% received IVT only, 37.9% received EVT only, and 25.9% underwent both IVT and EVT. With regard to etiology, 183 patients had a cardioembolic etiology, 136 patients had a non-cardioembolic etiology and 129 patients had ESUS.

Training set

For the ML training, only non-ESUS patients were used. Table 1 shows the characteristics of non-ESUS patients according to the presence of cardioembolic and non-cardioembolic etiologies. Patients with cardioembolic strokes were significantly older, were more often female, had higher CHA_2DS_2-VASc (Congestive heart failure or left ventricular systolic dysfunction, Hypertension, Age ≥ 75 [doubled], Diabetes, Stroke/transient ischemic attack [doubled], Vascular disease, Age 65–74, and Sex category) scores, and had more severe strokes (median [IQR] NIHSS score 12 [7–17] vs. 9 [4.0–15.25]; $p = 0.013$) when compared to patients with non-cardioembolic stroke. The prevalence of heart failure, coronary heart disease or myocardial infarction was higher in cardioembolic stroke patients, but current smoking was higher in non-cardioembolic stroke patients. In addition, the group of cardioembolic patients had significantly higher values for seven of the 11 laboratory variables examined (NTproBNP, high-sensitivity troponin T, D-dimers, CRP, HDL cholesterol, LDH and red cell distribution width), whereas they had significantly lower values in one variable (GFR).

Machine-learning classification

Using the 11 laboratory variables as well as age and sex, the resulting CatBoost Classifier ML model could categorize non-ESUS stroke patients into cardioembolic or non-cardioembolic etiology with excellent accuracy (AUC 0.82). When applied to the ESUS patients, the ML algorithm classified 40.3% of individuals as having a cardioembolic profile, and 59.7% as having a non-cardioembolic profile (Figure 1).

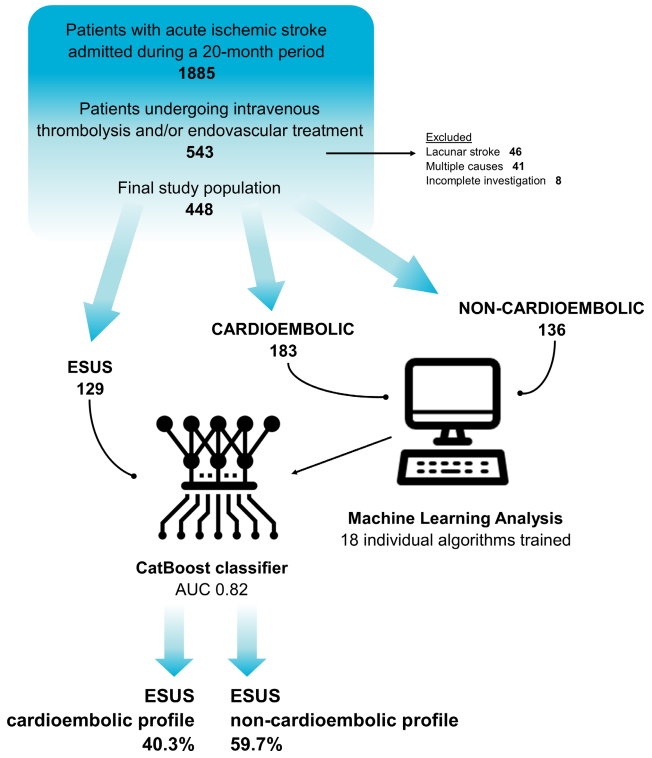

FIGURE 1 Concept of machine-learning classification in embolic stroke of undetermined source (ESUS) patients [Colour figure can be viewed at wileyonlinelibrary.com]

Patients with ESUS

Table 2 describes the baseline characteristics of ESUS patients stratified by cardioembolic or non-cardioembolic profile, originating from the CatBoost Classifier ML model based on age, sex, and baseline laboratory markers. In summary, ESUS patients classified as having a cardioembolic profile by our model were significantly older compared to those classified as non-cardioembolic. Furthermore, they were more often female, more likely to have arterial hypertension, were less frequently active smokers, had higher CHA_2DS_2-VASc scores, and had more premature atrial complexes per hour. Finally, significant differences in the majority of the laboratory variables which were used in the ML model were found between ESUS patients with cardioembolic and non-cardioembolic profiles (Table 2).

DISCUSSION

We demonstrated that by training an ML model using only baseline characteristics such as age, sex, and a set of routine laboratory variables, it is possible to accurately classify stroke etiology into either cardioembolic or non-cardioembolic in patients with known stroke etiology. This model may help to distinguish cardioembolic and non-cardioembolic profiles in ESUS patients, with approximately 40% of the ESUS patients in our dataset having been attributed a cardioembolic profile by our ML algorithm. Thus, our model further supports the hypothesis that ESUS represents a heterogeneous patient population, fewer than half of whom may have stroke attributable to cardioembolic pathophysiology. Our findings are not in line with the initial hypothesis by Hart et al. [5] that oral anticoagulation could lead to fewer stroke recurrences in all ESUS patients, and help to explain the results of the RE-SPECT ESUS and NAVIGATE-ESUS trials, in which not all patients benefited from anticoagulation [6,7].

This is in accordance with the review by Kamel et al. [8], in which the authors conclude that ESUS has different underlying occult etiologies. This perspective was assumed in two ongoing clinical trials examining the benefit of anticoagulation in subgroups of patients with ESUS believed to have a higher risk of recurrence or to be associated with a higher likelihood of cardioembolic pathophysiology, the ATTICUS [23] and the ARCADIA [24] trials. For these two randomized controlled trials, several clinical, laboratorial, electrophysiological, and echocardiographic markers were used to identify higher risk for cardiac embolism. It is known that several markers, such as left atrial dilation, higher NTproBNP values and frequent premature atrial complexes, are associated with a greater likelihood of incident atrial fibrillation in patients with cryptogenic stroke [25]. Even though atrial cardiopathy is associated with increased detection of atrial fibrillation during long-term follow-up of ESUS patients [26], it is uncertain if these patients could benefit from oral anticoagulation. The challenge is to identify markers that predict incident atrial fibrillation and stroke recurrence not only with good sensitivity but also with a good positive predictive value to avoid unnecessary exposure of ESUS patients to the risks of anticoagulation.

Because of their mathematical complexity, automated solutions require critical scientific judgement for correct implementation and interpretation [27]. Therefore, it is essential to compare inter-study and inter-population results of ML applications. The ML model of a similar study by Kamel et al. [28], who also developed an algorithm to classify ESUS patients, predicts occurrence of cardioembolic ESUS in 44% of patients, which is similar to the results of our model.

In the testing phase, which measured the accuracy of the ML model in classifying patients with known stroke etiology, our model presented similar diagnostic accuracy to that of the model developed by Kamel et al. It is important to point out here that only age, sex, and a set of 11 baseline laboratory blood variables were used to train our algorithm, whereas Kamel et al. employed a total of 174 features, including a high number of echocardiographic and brain imaging variables, many of which are not part of routine diagnostics [28].

The main limitations of our study are related to its retrospective nature and the relatively small study population size. Additionally, the ML method used does not provide cut-off values for each variable included in the model or the weight of each variable in the model, which preclude the construction of a score for etiological stratification of individual patients based solely on the current results. The advantage of our model is that it uses parameters which can be obtained quickly in newly admitted patients. It demonstrates that it may be possible to use a set of baseline readily available parameters to develop rapid, easy-to-perform, and cost-effective diagnostic tools to distinguish between different ESUS subtypes as well

TABLE 2 Baseline characteristics of patients with embolic stroke of undetermined source, stratified by machine-learning classification

	Classified as cardioembolic ($n = 52$)	Classified as non-cardioembolic ($n = 77$)	p
Median age, years (IQR)	80.5 (74.0–83.0)	62.0 (52.5–71.5)	<0.001
Female sex, *n* (%)	35 (67.3)	25 (32.5)	<0.001
Comorbidities, *n* (%)			
Arterial hypertension	43 (82.7)	50 (64.9)	0.027
Diabetes mellitus	11 (21.2)	18 (23.4)	0.767
Dyslipidemia	22 (42.3)	37 (48.1)	0.521
Current smoking	3 (5.8)	14 (18.2)	0.041
Coronary heart disease or myocardial infarction	10 (19.2)	7 (9.1)	0.095
Peripheral artery disease	2 (3.8)	0 (0)	0.083
Median CHA_2DS_2-VASc score (IQR)	6 (5–6)	4 (3–5)	<0.001
Median baseline NIHSS (IQR)	5 (4–9)	4 (2–9)	0.281
Hyperacute revascularization therapies, *n* (%)			
Intravenous thrombolysis	41 (78.8)	64 (83.1)	0.541
Endovascular treatment	22 (42.3)	26 (33.8)	0.325
Imaging characteristics			
Acute infarcts in multiple arterial territories	10 (19.2)	7 (9.1)	0.095
Old embolic infarcts in other arterial territories	9 (17.3)	9 (11.7)	0.366
Severe left atrial dilation	0 (0)	0 (0)	-
Median premature atrial complexes per hour (IQR)	5.5 (0.5–32.5)	0.9 (0.3–3.3)	0.006
Admission laboratory variables			
Mean hematocrit, % (SD)	35.4 (±4.5)	39.4 (±4.5)	<0.001
Median red cell distribution width, % (IQR)	13.2 (12.8–14.3)	13.0 (12.5–13.5)	0.016
Median D-Dimers, ng/ml (IQR)	5346 (1887–11167)	1558 (857–2184)	<0.001
Median HbA1c, % (IQR)	5.6 (5.5–6.0)	5.5 (5.3–5.8)	0.171
Median HDL cholesterol, mg/dl (IQR)	47 (41–67)	44 (36–54)	0.011
Median LDL cholesterol, mg/dl (IQR)	103 (80–124)	103 (76–135)	0.708
Median LDH, U/l (IQR)	217 (189–268)	172 (159–194)	<0.001
Median high sensitivity troponin T, pg/ml (IQR)	18 (10–32)	9 (4–14)	<0.001
Median NTproBNP, pg/ml (IQR)	874 (559–1647)	119 (48–241)	<0.001
Median GFR, ml/min/1.73 m^2 (IQR)	69.0 (54.7–81.6)	93.0 (80.5–102.1)	<0.001
Median CRP, mg/l (IQR)	3.4 (1.3–16.3)	1.8 (0.9–5.4)	0.041

Abbreviations: CRP, C-reactive protein; GFR, glomerular filtration rate; HbA1c, glycated hemoglobin; HDL, high-density lipoprotein; LDH, lactate dehydrogenase; LDL, low-density lipoprotein; IQR, interquartile range; NTproBNP, N-terminal pro-brain natriuretic peptide; SD, standard deviation.

as different stroke types in general. In this context, ML algorithms can help to find more individualized and thus more precise diagnoses and therapies for ESUS patients, which is an important step in the light of personalized medicine. We predict that ML models may become broadly used to develop clinical decision support systems, which will be based on large multicentric datasets. One of the most important advantages of ML methods is that they provide the opportunity to analyze non-linear interactions among a high number of variables, which can provide models that are more robust and that follow more reliably the clinical observations. There are also key challenges for their clinical application, such as the need for large high-quality datasets originating from multiple centers which accurately represent the target population, the need to accommodate dataset shift across time, and the need for early recognition and elimination of accidental confounder fitting and of algorithmic bias, among others [29].

The main conclusion of our study is that it is possible to classify subgroups of patients with ESUS into cardioembolic and non-cardioembolic profile groups by using limited demographic information and a set of baseline blood parameters. Patients with ESUS and a non-cardioembolic profile comprise the majority of ESUS patients. Future studies addressing not only isolated markers but also sets of readily available parameters with high sensitivity and specificity for higher risk of stroke recurrence or incident atrial fibrillation

in ESUS patients are required. For this purpose, we suggest that a score including not only demographic variables (e.g., age and sex), comorbidities (e.g., arterial hypertension) and blood biomarkers (e.g., NTproBNP, D-dimers), but also echocardiography markers (e.g., left atrial size, left atrial spontaneous contrast), electrophysiological markers (e.g., burden of atrial premature complexes, atrial runs) and imaging markers (e.g., left atrial fibrosis) could be useful. In addition, further investigation is needed to assess whether the strategy to anticoagulate ESUS patients with such cardioembolic profiles may be beneficial.

ACKNOWLEDGEMENT

Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors report no conflict of interests.

AUTHOR CONTRIBUTIONS

Max Christian Martin: Data curation (lead); Formal analysis (lead); Investigation (supporting); Methodology (equal); Resources (supporting); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (lead); Writing – review and editing (equal). **Thorsten Sichtermann:** Formal analysis (supporting); Investigation (supporting); Methodology (equal); Software (equal); Writing – original draft (supporting). **Kolja Schürmann:** Writing – review and editing (supporting). **Pardes Habib:** Methodology (supporting); Writing – review and editing (supporting). **Martin Wiesmann:** Validation (equal); Writing – review & editing (supporting). **Jörg B. Schulz:** Validation (equal); Writing – review and editing (supporting). **Omid Nikoubashman:** Investigation (supporting); Methodology (supporting); Writing – review and editing (equal). **João Pinho:** Conceptualization (equal); Data curation (supporting); Formal analysis (supporting); Investigation (supporting); Methodology (equal); Project administration (equal); Resources (equal); Supervision (lead); Validation (equal); Visualization (equal); Writing – review and editing (lead). **Arno Reich:** Conceptualization (equal); Investigation (lead); Methodology (supporting); Project administration (equal); Resources (lead); Supervision (equal); Validation (supporting); Writing – review and editing (equal).

ETHICAL APPROVAL

The authors hereby state that all analyses of the study were performed with the understanding and written consent of each subject, and that the study conforms with the World Medical Association Declaration of Helsinki published on the website of the *Journal of the American Medical Association*. In addition, our analysis was approved by the Ethics Committee of the Medical Faculty of RWTH Aachen University (approval reference 335/15).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request and according to the local ethics committee guidelines.

ORCID

Max Christian Martin https://orcid.org/0000-0002-7033-0107
Kolja Schürmann https://orcid.org/0000-0002-3835-0327
Pardes Habib https://orcid.org/0000-0002-5771-216X
João Pinho https://orcid.org/0000-0002-3977-7146
Arno Reich https://orcid.org/0000-0002-9439-7963

REFERENCES

1. Hart RG, Catanese L, Perera KS, Ntaios G, Connolly SJ. Embolic stroke of undetermined source: a systematic review and clinical update. *Stroke*. 2017;48(4):867-872. doi:10.1161/STROKEAHA.116.016414
2. Bembenek JP, Karlinski MA, Kurkowska-Jastrzebska I, Czlonkowska A. Embolic strokes of undetermined source in a cohort of Polish stroke patients. *Neurol Sci*. 2018;39:1041-1047. doi:10.1007/s10072-018-3322-5
3. Tsivgoulis G, Kargiotis O, Katsanos AH, et al. Incidence, characteristics and outcomes in patients with embolic stroke of undetermined source: a population-based study. *J Neurol Sci*. 2019;401:5-11. doi:10.1016/j.jns.2019.04.008
4. Ntaios G, Papavasileiou V, Milionis H, et al. Embolic strokes of undetermined source in the Athens stroke registry: an outcome analysis. *Stroke*. 2015;46(8):2087-2093. doi:10.1161/STROKEAHA.115.009334
5. Hart RG, Diener H-C, Coutts SB, et al. Embolic strokes of undetermined source: the case for a new clinical construct. *Lancet Neurol*. 2014;13(4):429-438. doi:10.1016/s1474-4422(13)70310-7
6. Hart RG, Sharma M, Mundl H, et al. for Stroke prevention after embolic stroke of undetermined source. *N Engl J Med*. 2018;378(23):2191-2201. doi:10.1056/NEJMoa1802686
7. Diener HC, Sacco RL, Easton JD, et al. Dabigatran for prevention of stroke after embolic stroke of undetermined source. *N Engl J Med*. 2019;380(20):1906-1917. doi:10.1056/NEJMoa1813959
8. Kamel H, Merkler AE, Iadecola C, Gupta A, Navi BB. Tailoring the approach to embolic stroke of undetermined source: a review. *JAMA Neurol*. 2019;76(7):855-861. doi:10.1001/jamaneurol.2019.0591
9. Kamel H, Okin PM, Longstreth WTJr, Elkind MS, Soliman EZ. Atrial cardiopathy: a broadened concept of left atrial thromboembolism beyond atrial fibrillation. *Future Cardiol*. 2015;11(3):323-331. doi:10.2217/fca.15.22
10. Healey JS, Gladstone DJ, Swaminathan B, et al. Recurrent stroke with rivaroxaban compared with aspirin according to predictors of atrial fibrillation: secondary analysis of the NAVIGATE ESUS randomized clinical trial. *JAMA Neurol*. 2019;76(7):764-773. doi:10.1001/jamaneurol.2019.0617
11. Gupta A, Gialdini G, Giambrone AE, et al. Association between nonstenosing carotid artery plaque on MR angiography and acute ischemic stroke. *JACC Cardiovasc Imaging*. 2016;9(10):1228-1229. doi:10.1016/j.jcmg.2015.12.004
12. Gupta A, Gialdini G, Lerario MP, et al. Magnetic resonance angiography detection of abnormal carotid artery plaque in patients with cryptogenic stroke. *J Am Heart Assoc*. 2015;4(6):e002012. doi:10.1161/JAHA.115.002012
13. Buon R, Guidolin B, Jaffre A, et al. Carotid ultrasound for assessment of nonobstructive carotid atherosclerosis in young adults with cryptogenic stroke. *J Stroke Cerebrovasc Dis*. 2018;27(5):1212-1216. doi:10.1016/j.jstrokecerebrovasdis.2017.11.043
14. Coutinho JM, Derkatch S, Potvin ARJ, et al. Nonstenotic carotid plaque on CT angiography in patients with cryptogenic stroke. *Neurology*. 2016;87:665-672. doi:10.1212/WNL.0000000000002978
15. Komatsu T, Iguchi Y, Arai A, et al. Large but nonstenotic carotid artery plaque in patients with a history of embolic stroke

of undetermined source. *Stroke*. 2018;49(12):3054-3056. doi:10.1161/STROKEAHA.118.022986
16. Hyafil F, Schindler A, Sepp D, et al. High-risk plaque features can be detected in non-stenotic carotid plaques of patients with ischaemic stroke classified as cryptogenic using combined (18)F-FDG PET/MR imaging. *Eur J Nucl Med Mol Imaging*. 2016;43(2):270-279. doi:10.1007/s00259-015-3201-8
17. Siegler JE, Thon JM, Woo JH, Do D, Messe SR, Cucchiara B. Prevalence of nonstenotic carotid plaque in stroke due to atrial fibrillation compared to embolic stroke of undetermined source. *J Stroke Cerebrovasc Dis*. 2019;28(10):104289. doi:10.1016/j.jstrokecerebrovasdis.2019.07.005
18. Kamel H, Gialdini G, Baradaran H, et al. Cryptogenic stroke and nonstenosing intracranial calcified atherosclerosis. *J Stroke Cerebrovasc Dis*. 2017;26(4):863-870. doi:10.1016/j.jstrokecerebrovasdis.2016.10.035
19. Cheung HM, Moody AR, Singh N, Bitar R, Zhan J, Leung G. Late stage complicated atheroma in low-grade stenotic carotid disease: MR imaging depiction—prevalence and risk factors. *Radiology*. 2011;260:841-847.
20. Adams HPJr, Bendixen BH, Kappelle LJ, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in acute stroke treatment. *Stroke*. 1993;24(1):35-41. doi:10.1161/01.str.24.1.35
21. Obuchowski N, Blackmore CC, Karlik S, Reinhold C. ROC analysis - fundamentals of clinical research for radiologists. *American Roentgen Ray Society*. 2004;14:364-372.
22. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. Presented at: arXiv:170609516v5 [csLG] Jan 2019.
23. Geisler T, Poli S, Meisner C, et al. Apixaban for treatment of embolic stroke of undetermined source (ATTICUS randomized trial): rationale and study design. *Int J Stroke*. 2017;12(9):985-990. doi:10.1177/1747493016681019
24. Kamel H, Longstreth WTJr, Tirschwell DL, et al. The AtRial cardiopathy and antithrombotic drugs in prevention after cryptogenic stroke randomized trial: rationale and methods. *Int J Stroke*. 2019;14(2):207-214. doi:10.1177/1747493018799981
25. Elkind MSV. Atrial cardiopathy and stroke prevention. *Curr Cardiol Rep*. 2018;20(11):103. doi:10.1007/s11886-018-1053-0
26. Silva AR, Pires C, Meira-Carvalho F, et al. Atrial cardiopathy predicts detection of atrial fibrillation in embolic stroke of undetermined source: real-life data. *Neurol Sci*. 2022;43(4):2383-2386. doi:10.1007/s10072-021-05692-3
27. Baştanlar Y, Özuysal M. Introduction to machine learning. In: Yousef M, Allmer J, eds. *miRNomics: MicroRNA Biology and Computational Analysis*. Humana Press; 2014: 105-128. Methods in Molecular Biology.
28. Kamel H, Navi BB, Parikh NS, et al. Machine learning prediction of stroke mechanism in embolic strokes of undetermined source. *Stroke*. 2020;51(9):e203-e210. doi:10.1161/STROKEAHA.120.029305
29. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2019;17:195. doi:10.1186/s12916-019-1426-2

How to cite this article: Martin MC, Sichtermann T, Schürmann K, et al. Classification of patients with embolic stroke of undetermined source into cardioembolic and non-cardioembolic profile subgroups. *Eur J Neurol*. 2022;29:2275-2282. doi: 10.1111/ene.15356

MANAGE-PD

Tool for Making Informed Decisions to **Aid Timely Management of Parkinson's Disease**

MANAGE-PD allows you to:

- Identify PD patients inadequately controlled on oral medications
- Determine which patients with PD may be adequately controlled on their current treatment regimen or may require changes to their treatment regimen

Scan the QR code to access to the web

Click here to access to the web

MANAGE-PD is an AbbVie Inc. registered Medical Device. It is a collaborative research and development effort between AbbVie Medical Affairs and Health Economics and Outcomes, the Parkinson's Foundation and an international panel of Movement Disorder Specialists.

PD: Parkinson´s Disease

Parkinson's Foundation

Veeva Code: ABBV-AA-00335-FM Approved: January 2022

abbvie