Stroke

CLINICAL AND POPULATION SCIENCES

Machine Learning Prediction of Stroke Mechanism in Embolic Strokes of Undetermined Source

Hooman Kamel, MD, MS; Babak B. Navi, MD, MS; Neal S. Parikh, MD, MS; Alexander E. Merkler, MD; Peter M. Okin, MD; Richard B. Devereux, MD; Jonathan W. Weinsaft, MD; Jiwon Kim, MD; Jim W. Cheung, MD; Luke K. Kim, MD; Barbara Casadei, MD, DPhil; Costantino Iadecola, MD; Mert R. Sabuncu, PhD; Ajay Gupta, MD; Iván Díaz, PhD

BACKGROUND AND PURPOSE: One-fifth of ischemic strokes are embolic strokes of undetermined source (ESUS). Their theoretical causes can be classified as cardioembolic versus noncardioembolic. This distinction has important implications, but the categories' proportions are unknown.

METHODS: Using data from the Cornell Acute Stroke Academic Registry, we trained a machine-learning algorithm to distinguish cardioembolic versus non-cardioembolic strokes, then applied the algorithm to ESUS cases to determine the predicted proportion with an occult cardioembolic source. A panel of neurologists adjudicated stroke etiologies using standard criteria. We trained a machine learning classifier using data on demographics, comorbidities, vitals, laboratory results, and echocardiograms. An ensemble predictive method including L1 regularization, gradient-boosted decision tree ensemble (XGBoost), random forests, and multivariate adaptive splines was used. Random search and cross-validation were used to tune hyperparameters. Model performance was assessed using cross-validation among cases of known etiology. We applied the final algorithm to an independent set of ESUS cases to determine the predicted mechanism (cardioembolic or not). To assess our classifier's validity, we correlated the predicted probability of a cardioembolic source with the eventual post-ESUS diagnosis of atrial fibrillation.

RESULTS: Among 1083 strokes with known etiologies, our classifier distinguished cardioembolic versus noncardioembolic cases with excellent accuracy (area under the curve, 0.85). Applied to 580 ESUS cases, the classifier predicted that 44% (95% credibility interval, 39%–49%) resulted from cardiac embolism. Individual ESUS patients' predicted likelihood of cardiac embolism was associated with eventual atrial fibrillation detection (OR per 10% increase, 1.27 [95% CI, 1.03–1.57]; *c-statistic, 0.68* [95% CI, 0.58–0.78]). ESUS patients with high predicted probability of cardiac embolism were older and had more coronary and peripheral vascular disease, lower ejection fractions, larger left atria, lower blood pressures, and higher creatinine levels.

CONCLUSIONS: A machine learning estimator that distinguished known cardioembolic versus noncardioembolic strokes indirectly estimated that 44% of ESUS cases were cardioembolic.

Key Words: atrial fibrillation ■ embolism ■ machine learning ■ probability ■ stroke

Stroke causes 10% of all deaths as well as a substantial burden of permanent disability.[1] The majority of strokes are ischemic, one-fifth of which lack a clear cause.[2] Such cryptogenic strokes often seem embolic and can be referred to as embolic strokes of undetermined source (ESUS).[3] The natural history of ESUS is not benign, with 1 in 20 patients each year experiencing a recurrent stroke,[4] which tends to be more severe than

Correspondence to: Hooman Kamel, MD, MS, 420 East 70th St, LH-413, New York, NY 10021. Email hok9010@med.cornell.edu

This manuscript was sent to Liping Liu, Guest Editor, for review by expert referees, editorial decision, and final disposition.

The Data Supplement is available with this article at https://www.ahajournals.org/doi/suppl/10.1161/STROKEAHA.120.029305.

For Sources of Funding and Disclosures, see page e209.

Stroke is available at www.ahajournals.org/journal/str

Nonstandard Abbreviations and Acronyms

AF	atrial fibrillation
CAESAR	Cornell Acute Stroke Academic Registry
ESUS	embolic stroke of undetermined source
TOAST	Trial of ORG 10172 in Acute Stroke Treatment

the initial stroke.[5] Despite a substantial amount of investigation,[4,6] better treatments to prevent recurrent stroke after ESUS remain elusive, highlighting the need for a better understanding of the pathophysiology of ESUS.[7]

There are numerous theoretical causes of ESUS.[3] Many of these potential mechanisms can be classified as cardioembolic—for example, atrial cardiopathy,[8] unrecognized myocardial infarction,[9] or patent foramen ovale[10]—in contrast to other common conditions such as nonstenosing atherosclerosis[11] or other forms of vasculopathy. This broad distinction between cardioembolic versus noncardioembolic sources may have important therapeutic implications.[12,13] However, the classification of ESUS into cardioembolic versus noncardioembolic etiologies remains difficult in the absence of a gold-standard diagnostic test. In such a setting, machine learning may provide novel insight by applying rules extracted from labeled cases (ie, strokes with a known etiology) to unlabeled cases (ie, ESUS).

METHODS

Design

We developed and tested a machine learning algorithm using data from ischemic stroke patients enrolled in the Cornell Acute Stroke Academic Registry (CAESAR). We trained the algorithm on known ischemic stroke subtypes in our registry and applied the algorithm to ESUS cases to assess the predicted proportion of cardioembolic versus noncardioembolic sources (Figure 1). The institutional review board at Weill Cornell Medicine approved this study and waived the requirement for informed consent. Deidentified data and analytic code are available from the corresponding author upon reasonable request.

Patient Population

All patients hospitalized at New York-Presbyterian Hospital/Weill Cornell Medical Center for acute stroke are prospectively enrolled in the American Heart Association's Get With The Guidelines—Stroke registry. Trained hospital analysts prospectively collect data on demographics, vascular risk factors and comorbidities, stroke severity, and in-hospital treatments and outcomes. CAESAR combines the Get With The Guidelines data plus additional clinical, laboratory, and radiographic data.[14] All clinically diagnosed cases of ischemic stroke are reviewed by a panel of at least three neurologists who adjudicate the etiology per the TOAST (Trial of ORG 10172 in Acute Stroke Treatment) classification[15] and the consensus definition of ESUS.[16] The adjudication panel comprises 4 board-certified neurologists with an average 6 years of post-training experience. Adjudications were performed by group discussion and consensus after review of all available data, including neuroimaging data, from the index hospitalization. In a validation analysis of a subset of 374 patients, double-blind adjudication by 2 independent reviewers indicated an interrater agreement rate of 86.4% and a kappa of 0.81. For this analysis, we included patients with ischemic stroke registered in CAESAR from 2011 through 2016. We excluded lacunar strokes because we were interested in better understanding the root causes of ESUS, which is by definition nonlacunar. We excluded patients with an incomplete stroke evaluation. We also excluded patients with multiple potential causes of stroke because the management of these patients often addresses all the known causes, rendering the delineation of the exact cause of stroke less important, and because we wanted to facilitate identification of mechanisms and risk factors unique to individual, well-defined stroke subtypes. Therefore, our analysis included patients with ESUS as well as patients with the following known ischemic stroke etiologies: large-artery atherosclerosis, cardiac embolism, or another determined etiology (eg, cervical artery dissection). Cases with a known ischemic stroke subtype were labeled as either cardioembolic or noncardioembolic (ie, large-artery atherosclerosis or another determined etiology).

Measurements

As features on which to train our machine learning classifier, we used data on patients' demographics, vascular risk factors and comorbidities, vital signs, laboratory results, and echocardiographic findings (Table I in the Data Supplement). We did not include a history of atrial fibrillation (AF) as a training variable since it was essentially completely collinear with a cardioembolic etiology and, by definition, was absent in all ESUS cases. Demographics and risk factors and comorbidities were collected by hospital analysts for the Get With The Guidelines—Stroke registry and were imported directly into the Microsoft SQL server which houses the CAESAR registry. Echocardiographic measurements were imported directly from our hospital's echocardiographic image management system (Xcelera, Philips Healthcare) into the CAESAR SQL server. Vital signs and laboratory results were imported directly into the CAESAR SQL server from our hospital's inpatient (Allscripts, Cerner) and outpatient (EPIC) electronic medical record systems. For patients with multiple measurements of a given vital sign or laboratory test, we used the first value after admission.

Statistical Analysis

We estimated the proportion of ESUS arising from occult cardiac sources using a Bayesian targeted machine learning algorithm.[17] The algorithm estimates the proportion in 3 steps: (1) build 2 predictive models: one for the probability that the source of stroke is undetermined, and one for the probability of a cardioembolic source among patients with a determined source; (2) tilt the probability of a cardioembolic source towards a debiased function to achieve the right bias-variance trade-off; and (3) incorporate prior information using a Bayesian updating

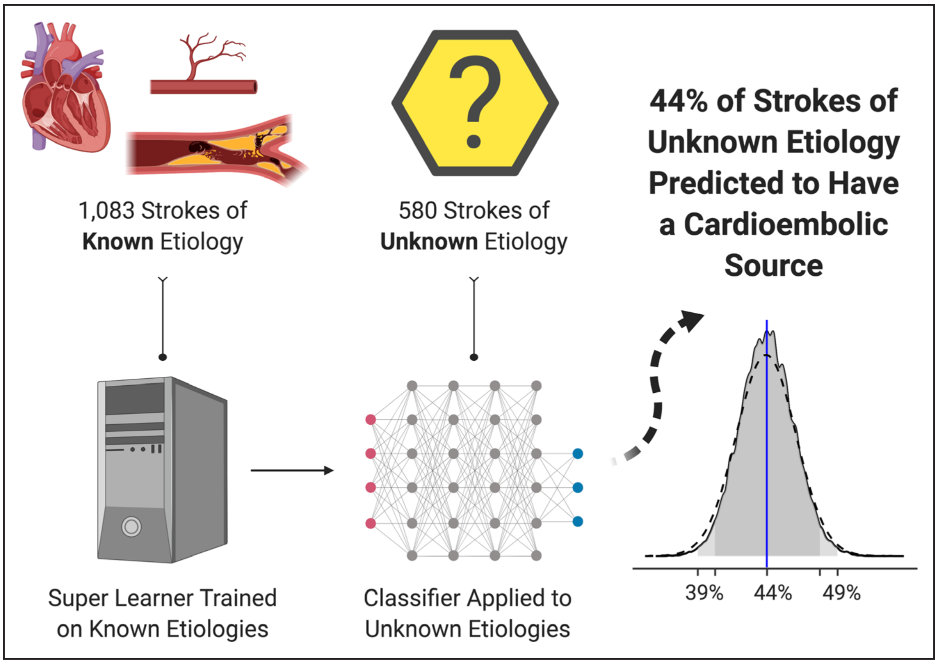
1,083 Strokes of
Known Etiology
580 Strokes of
Unknown Etiology
44% of Strokes of Unknown Etiology Predicted to Have a Cardioembolic Source
Super Learner Trained on Known Etiologies
Classifier Applied to Unknown Etiologies
39%
44%
49%

Figure 1. Design of machine learning analysis of mechanism in embolic strokes of undetermined source.

step. A uniform distribution in the interval [0, 1] was used as an uninformative prior for the proportion of ESUS or cardiac source. The predictive models in step (1) were fitted using an ensemble machine learning method known as the super learner. The super learner[18,19] algorithm proceeds by splitting the data into 10-fold validation sets. For each validation set, the corresponding training set is given by the remaining 9 folds. Each candidate model is fitted in each training set; hyper-parameters are tuned using 8-fold nested cross-validation. In each training set, the outcome is regressed on the out-of-sample predictions of each candidate model, to find weights in an ensemble that minimize the cross-entropy loss function. This ensemble is used to obtain predictions in the corresponding validation sets. The ensemble consisted of gradient-boosted decision trees, random forests, L1 regularized logistic regression, multivariate adaptive regression splines, and vanilla logistic regression. A screening step selecting the 50 variables with the largest univariate correlations with the outcome was used for the latter 2 algorithms to avoid over-fitting. Cross-validation was used to calculate the area under the curve for out-of-sample predictions of known stroke etiologies. These predictive models were then used to compute a Bayesian targeted machine learning estimator for the proportion of ESUS arising from an occult cardioembolic source.

Validation

To assess the validity of our classifier among ESUS patients, we correlated our machine learning estimator's predicted probability of an occult cardioembolic source with eventual diagnosis of AF on postdischarge heart-rhythm monitoring tests. After hospital discharge, ESUS patients at our medical center routinely undergo 30 days of continuous heart-rhythm monitoring using an external loop recorder with a highly sensitive automated AF detection algorithm. All automated AF detections are reviewed and adjudicated as true or false AF detections by a cardiac electrophysiologist. AF is defined as sustained AF lasting >30 seconds. We considered eventual poststroke AF detection to be a reasonable construct for testing the validity of our estimator, as we would expect patients with a higher predicted probability of an occult cardioembolic source to more often eventually manifest AF than patients with a lower predicted probability of an occult cardioembolic source.

Sensitivity Analysis

We performed 2 sensitivity analyses. First, we included as training features additional data from a radiologist's interpretation of brain imaging, which were available for a subset of patients enrolled during 2011 to 2015. These brain imaging features were abstracted by a single radiologist blinded to the machine learning classifier's output and included the age-related White Matter Changes score, the Fazekas score, the number and location of acute infarcts, and the number and location of microbleeds (Table I in the Data Supplement). Second, we excluded patients who were transferred from other hospitals, thereby essentially limiting our population to local patients from neighboring zip codes in Manhattan. Third, we included AF as a training feature.

RESULTS

Among 2116 patients with ischemic stroke registered in CAESAR from 2011 through 2016, we excluded 202 patients with stroke because of small-vessel occlusion, 139 patients with an incomplete evaluation, and 112 patients with two or more potential causes of stroke. Among the remaining 1663 patients, 1083 had a known stroke etiology (291 had stroke due to large-artery atherosclerosis, 688 due to cardiac embolism, and 104 due to another determined etiology), while 580 had cryptogenic stroke, which was defined per the ESUS consensus definition based on information available at the time of the index hospitalization (ie, before postdischarge AF monitoring; Table 1).

Among the 1663 patients with stroke whose etiology was determined, our predictive model distinguished cardioembolic from noncardioembolic etiologies with excellent validity (area under the curve, 0.85; Figure 2; Figure I in the Data Supplement). The Bayesian targeted machine learning estimator which incorporated this predictive model estimated that 44% (95% credibility interval, 39%–49%) of ESUS cases were due to an occult cardioembolic source (Figure 3). Compared with patients with ESUS whose predicted likelihood of an occult cardioembolic source was in the lowest quartile (<25%), patients with ESUS with a predicted likelihood of cardiac embolism in the highest quartile (≥75%) were older and more often White; more often had coronary artery disease, heart failure, and peripheral vascular disease; were less likely to be active smokers; and had more severe strokes, lower ejection fractions, larger left atria, lower blood pressure, and higher serum creatinine levels (Table 2).

In our validation step, among the 580 patients who were all classified as ESUS after review of all available data including neuroimaging data, our machine learning classifier's estimated likelihood of an occult cardioembolic source was associated with the eventual detection of AF (OR per 10% increase in predicted probability of cardioembolic source, 1.27 [95% CI, 1.03–1.57]; *c-statistic, 0.68* [95% CI, 0.58–0.78]); AF was found in 9.1% of those in the top quartile versus 2.9% of those in the bottom quartile of predicted likelihood of a cardioembolic source. In a sensitivity analysis adding as a training feature brain imaging data from the 1369 patients (82.3%) enrolled during 2011 to 2015, the predicted proportion of patients with a cardioembolic source remained 44%. Our results were also similar after excluding the 352 patients (21.2%) who were transferred from another hospital and when including AF as a training feature.

DISCUSSION

In a registry of patients with ischemic stroke, we developed a machine learning classifier that accurately sorted

Table 1. Baseline Characteristics of Eligible CAESAR Patients, Stratified by Ischemic Stroke Subtype

Characteristic*	Large-Artery Atherosclerosis (N=291)	Cardiac Embolism (N=688)	Other Determined Source (N=104)	Undetermined Source (N=580)	*P* Value
Age, mean (SD), y	72 (12)	75 (14)	56 (18)	67 (17)	<0.001
Female	115 (39.5)	368 (53.5)	40 (38.5)	310 (53.5)	<0.001
White†	98 (50.5)	290 (65.9)	38 (55.1)	251 (59.5)	0.002
Insurance					<0.001
Commercial	121 (41.6)	298 (43.3)	67 (64.4)	298 (51.4)	
Medicare	120 (41.2)	299 (43.5)	17 (16.4)	201 (34.7)	
Medicaid	44 (15.1)	82 (11.9)	16 (15.4)	66 (11.4)	
Other	6 (2.1)	9 (1.3)	4 (3.9)	15 (2.6)	
Hypertension	238 (81.8)	475 (69.0)	52 (50.0)	349 (60.2)	<0.001
Diabetes mellitus	98 (33.7)	138 (20.1)	16 (15.4)	127 (21.9)	<0.001
Active tobacco use	164 (56.4)	328 (47.7)	32 (30.8)	257 (44.3)	<0.001
Coronary artery disease	40 (13.8)	39 (5.7)	11 (10.6)	36 (6.2)	<0.001
Heart failure	54 (18.6)	179 (26.0)	19 (18.3)	86 (14.8)	<0.001
Chronic kidney disease	9 (3.1)	70 (10.2)	1 (1.0)	15 (2.6)	0.90
Peripheral vascular disease	16 (5.5)	44 (6.4)	4 (3.9)	23 (4.0)	0.24
Prior stroke	16 (5.5)	44 (6.4)	4 (3.9)	23 (4.0)	0.004
Onset to arrival, median (IQR), h	1.7 (0.3–14.6)	0.9 (0.1–4.9)	1.3 (0.2–9.8)	2.5 (0.4–12.9)	<0.001
NIHSS score on arrival, median (IQR)	4 (1–9)	8 (2–18)	4 (2–14)	3 (1–9)	<0.001
Ejection fraction, mean (SD), %	62 (9)	53 (17)	63 (7)	62 (8)	<0.001
LAVI, mean (SD), mL/m^2	31 (11)	50 (33)	32 (14)	34 (14)	<0.001
E/e' lateral	9.9 (3.7)	11.8 (7.2)	8.5 (3.3)	9.6 (5.1)	0.003
E/e' medial	13.9 (5.4)	14.5 (7.1)	10.2 (4.3)	13.2 (7.0)	0.05
SBP, mean (SD), mm Hg	157 (31)	145 (30)	152 (48)	150 (30)	<0.001
Serum creatinine, mean (SD), mg/dL	1.2 (0.9)	1.2 (1.0)	1.2 (1.1)	1.1 (0.9)	0.01

CAESAR indicates Cornell Acute Stroke Academic Registry; IQR, interquartile range; LAVI, left atrial volume index; NIHSS, National Institutes of Health Stroke Scale; and SBP, systolic blood pressure.

*Data are presented as number (%) unless otherwise specified.

†Numbers do not sum up to column total because of patients with missing data on race.

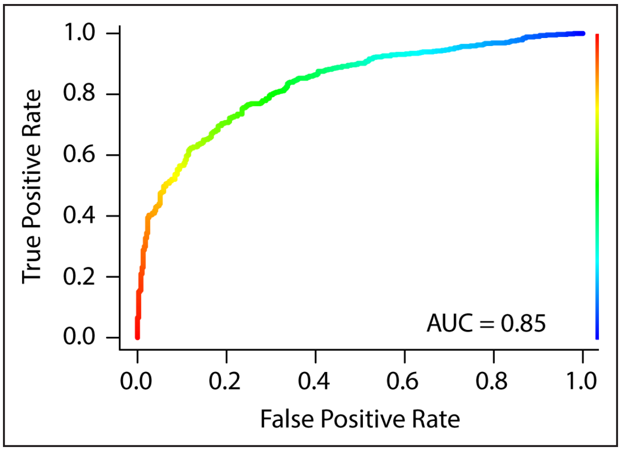

Figure 2. Receiver operating characteristics curve for Bayesian targeted machine learning estimator of probability of occult cardioembolic source.
AUC indicates area under the curve.

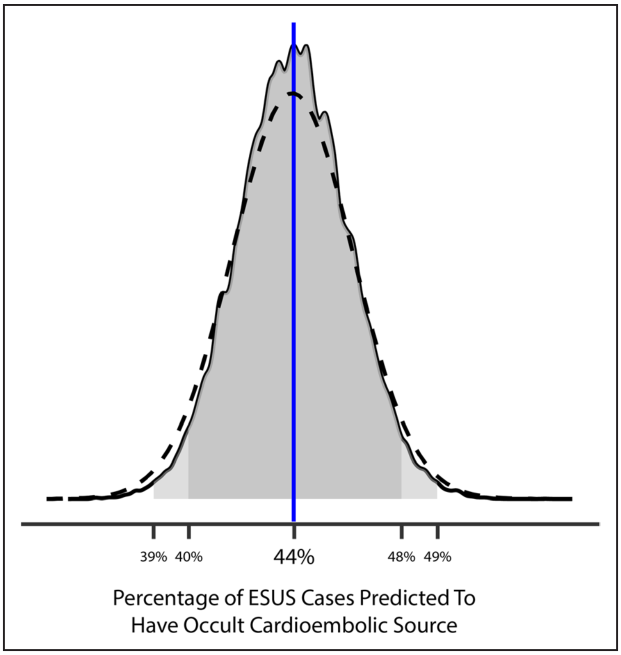

Figure 3. Distribution of percentage of embolic strokes of undetermined source (ESUS) cases predicted to have occult cardioembolic source.
The shaded gray areas indicate the 68% and 95% credibility intervals.

known stroke etiologies into cardioembolic versus noncardioembolic categories. Among ESUS cases, the predicted probability of an occult cardioembolic source, which was determined using only information available during the initial stroke hospitalization, was associated with the eventual diagnosis of paroxysmal AF after hospital discharge. When this classifier was applied to ESUS cases, it indirectly estimated that slightly fewer than half of currently unclassified strokes were due to an occult cardioembolic source.

Numerous studies have investigated individual conditions that may represent occult mechanisms of ESUS,[9,20–22] and several studies have investigated relationships among various mechanisms.[10,23,24] Such investigations make it increasingly clear that ESUS is a heterogeneous clinical condition with a variety of underlying mechanisms, some of which are cardioembolic and some of which are not.[7,11,25,26] There are few data on the relative proportions of cardioembolic versus noncardioembolic mechanisms of ESUS, or the characteristics of ESUS patients with presumed cardioembolic versus noncardioembolic mechanisms. In this context, our study provides several novel findings. First, our results suggest that slightly less than half of ESUS cases may represent occult cardiac embolism. Second, our results suggest that ESUS patients with a likely cardioembolic source have a distinct clinical profile compared with patients with a likely noncardioembolic source. Those with likely cardiac embolism seem to be an older group with more cardiac comorbidities, worse cardiac function, and more severe strokes, while those with likely noncardioembolic sources seem to be a younger group with higher blood pressure and a greater prevalence of active tobacco use.

Our study has several limitations. First, our data were from a single academic medical center and thus our algorithm and findings may not be generalizable to other populations. On the other hand, ≈80% of our patients with stroke are admitted through our emergency department and are residents of local neighborhoods in which the majority of strokes are admitted either to our medical center or to other nearby academic medical centers,[27] which may mitigate the risk of referral bias. Our findings were essentially identical when we limited our analysis to this local population. Second, in the absence of a gold standard, we cannot verify our machine learning estimator's prediction that 44% of ESUS cases represent a cardioembolic source. However, our estimator had excellent performance among known cases of cardioembolic and noncardioembolic stroke, and the validity of its predictions are supported by their association with the eventual diagnosis of AF. Nevertheless, poststroke AF diagnosis is not definitive proof that the mechanism of the index stroke was cardioembolic, so these findings provide only indirect corroboration. Some cases of poststroke AF may represent the results of stroke-induced damage to autonomic pathways, rather than a clue as to the underlying etiology of stroke. Third, we lacked pixel-level brain imaging data, and it is possible that the inclusion of such granular data would have improved the performance of our classifier and changed its findings when applied to ESUS cases. However, our findings were identical even with the inclusion of detailed data on brain imaging findings, which were available for the majority of our cohort. It may be that basic neuroimaging features such as infarct location and white matter disease grade may not contribute much to the algorithm once other features such as echocardiography findings are taken into account. Future studies are required to determine whether more granular neuroimaging features, including imaging of

Table 2. Baseline Characteristics of Patients With ESUS, Stratified by Machine Learning Prediction of a Cardioembolic Source

Characteristic*	Highest Predicted Probability of Cardiac Embolism† (N=71)	Lowest Predicted Probability of Cardiac Embolism‡ (N=149)	*P* Value
Age, mean (SD), y	78 (11)	60 (16)	<0.001
Female	38 (53.5)	68 (45.6)	0.27
White§	34 (75.6)	55 (53.4)	0.01
Payment source			0.09
Commercial	32 (45.1)	78 (52.4)	
Medicare	32 (45.1)	43 (28.9)	
Medicaid	5 (7.0)	20 (13.4)	
Other	2 (2.8)	8 (5.4)	
Hypertension	45 (63.4)	97 (65.1)	0.80
Diabetes mellitus	16 (22.5)	36 (24.2)	0.79
Active tobacco use	1 (1.4)	15 (10.1)	0.02
Coronary artery disease	21 (29.6)	13 (8.7)	<0.001
Heart failure	6 (8.5)	1 (0.7)	0.002
Chronic kidney disease	1 (1.4)	3 (2.0)	0.75
Peripheral vascular disease	9 (12.7)	3 (2.0)	0.001
Prior stroke	18 (25.4)	26 (17.5)	0.17
Onset to arrival, median (IQR), h	2.1 (1.0–5.5)	2.8 (0.4–20.0)	0.73
NIHSS score on arrival, median (IQR)	8 (3–16)	3 (1–6)	<0.001
Ejection fraction, mean (SD), %	54 (11)	65 (6)	<0.001
LAVI, mean (SD), mL/m^2	47 (19)	28 (9)	<0.001
E/e' lateral	10.3 (3.8)	9.2 (3.6)	0.38
E/e' medial	14.8 (5.1)	12.4 (4.6)	0.16
SBP, mean (SD), mm Hg	147 (29)	159 (30)	0.01
Serum creatinine, mean (SD), mg/dL	1.5 (1.7)	1.0 (0.8)	0.01

ESUS indicates embolic strokes of undetermined source; IQR, interquartile range; LAVI, left atrial volume index; NIHSS, National Institutes of Health Stroke Scale; and SBP, systolic blood pressure.

*Data are presented as number (%) unless otherwise specified.

†Among the 580 patients with embolic stroke of undetermined source, these data are for the 71 patients to whom our algorithm assigned a predicted probability of a cardioembolic source ≥0.75.

‡Among the 580 patients with embolic stroke of undetermined source, these data are for the 149 patients to whom our algorithm assigned a predicted probability of a cardioembolic source <0.25.

§Numbers do not sum up to column total because of patients with missing data on race.

the intracranial and extracranial large arteries and aorta, can improve prediction. Fourth, we lacked data on recurrent stroke and thus could not examine the association between the predicted probability of a cardioembolic source and the risk of stroke recurrence. Fifth, we used cross-validation to obtain an efficient estimate of predictive error. The alternative approach of splitting the data into a training set and a testing set may have resulted in different estimates.

Our study may have several potential implications. First, our findings highlight the phenotypic heterogeneity of ESUS and support and inform personalized approaches to its treatment. For example, among many other factors, older patients and those with large left atria were over-represented among patients predicted to have a high risk of an occult cardioembolic source. These findings are concordant with secondary analyses of randomized clinical trials which found a benefit with anticoagulant therapy in these 2 subgroups,[6,13] although there was no benefit in the overall ESUS population.[4,6] Our study suggests that the findings of these subgroup analyses of the ESUS trials may indicate true therapeutic effects in these selected patient groups who seem to be enriched for underlying cardioembolic sources. Randomized trials of anticoagulant therapy in only ESUS patients with markers of cardiac disease, such as ARCADIA[28] and ATTICUS,[29] are ongoing. If positive, these trials will serve as proof-of-concept of personalized antithrombotic strategies for preventing stroke recurrence after ESUS, but ultimately more refined risk stratification of an underlying cardioembolic source will be necessary to guide trials of not just current oral anticoagulants but novel therapies such as emerging Factor XIa inhibitors. Our study may inform the findings of such future trials by suggesting other clinical selection criteria, such as older age, pre-existing coronary artery disease or peripheral vascular disease, and reduced ejection fraction. Second, our study suggests that machine learning models may have utility not just for predicting events but for interrogating disease states to shed light on the underlying pathophysiology. This latter application of machine learning has the advantage of easy dissemination via scientific publication, while prediction tools are often difficult to disseminate in the absence of a commercial software product or other platform for inputting the required data.

CONCLUSIONS

A machine learning algorithm trained on demographic, clinical, echocardiographic, and laboratory data was able to distinguish known cardioembolic strokes from known noncardioembolic strokes with excellent accuracy. Among ESUS cases, the predicted probability of occult cardiac embolism was significantly associated with the eventual diagnosis of AF. When applied to patients with ESUS, the algorithm indirectly estimated that slightly fewer than half of these strokes of unknown etiology were due to an occult cardioembolic source. Patients with ESUS with a high probability of an occult cardioembolic source were older, more often had cardiac disease, and had lower ejection fractions and larger left atria. These findings may shed light on the pathophysiology of stroke and inform the design of future trials on stroke prevention after ESUS.

ARTICLE INFORMATION

Received February 5, 2020; final revision received July 5, 2020; accepted July 14, 2020.

Affiliations

Clinical and Translational Neuroscience Unit, Department of Neurology, Feil Family Brain and Mind Research Institute (H.K., B.B.N., N.S.P., A.E.M., C.I.), Division of Cardiology (P.M.O., R.B.D., J.W.W., J.K., J.W.C., L.K.K.), Division of Biostatistics and Epidemiology (I.D.), and Department of Radiology (A.G.), Weill Cornell Medical College, New York. School of Electrical and Computer Engineering, and Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY (M.R.S.). Division of Cardiovascular Medicine, University of Oxford, United Kingdom (B.C.).

Sources of Funding

This study was supported by the NIH/NINDS (grant R01NS097443) and the NIH/NHLBI (grant R01HL144541).

Disclosures

Dr Kamel serves as co-PI for the NIH-funded ARCADIA trial which receives in-kind study drug from the BMS-Pfizer Alliance and in-kind study assays from Roche Diagnostics, serves as Deputy Editor for *JAMA* Neurology, serves as a steering committee member of Medtronic's Stroke AF trial (uncompensated), serves on an end point adjudication committee for a trial of empagliflozin for Boehringer-Ingelheim, and has served on an advisory board for Roivant Sciences related to Factor XI inhibition. Dr Navi serves as a member of the data and safety monitoring board for the PCORI-funded TRAVERSE trial and has received personal fees for medicolegal consulting on stroke. Dr Parikh reports grants from New York State Department of Health Empire Clinical Research Investigator Program, grants from the Leon Levy Foundation, and grants from the Florence Gould Foundation outside the submitted work. Dr Merkler reports grants from AHA and personal fees from medicolegal consulting outside the submitted work. Dr Devereux has received an honorarium from Acer pharmaceuticals for a one-day consultation about treatment of vascular Ehlers-Danlos syndrome. Dr Cheung has received personal fees from Biosense Webster, Boston Scientific, Abbott, and Biotronik. Dr Casadei has received nonfinancial support from Roche Diagnostics (free assays) and iRhythm (free ECG monitors), and grants from the British Heart Foundation and the UK NIHR. Dr Iadecola has received personal fees from Broadview Ventures outside the submitted work. Dr Gupta reports nonfinancial support from GE Healthcare and nonfinancial support from Siemens Medical Solutions, United States outside the submitted work.

Supplemental Materials

Expanded Methods and Results
Table I
Figure I

REFERENCES

1. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, Abraham J, Adair T, Aggarwal R, Ahn SY, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380:2095–2128. doi: 10.1016/S0140-6736(12)61728-0
2. Hart RG, Catanese L, Perera KS, Ntaios G, Connolly SJ. Embolic stroke of undetermined source: a systematic review and clinical update. *Stroke*. 2017;48:867–872. doi: 10.1161/STROKEAHA.116.016414
3. Hart RG, Diener HC, Coutts SB, Easton JD, Granger CB, O'Donnell MJ, Sacco RL, Connolly SJ; Cryptogenic Stroke/ESUS International Working Group. Embolic strokes of undetermined source: the case for a new clinical construct. *Lancet Neurol*. 2014;13:429–438. doi: 10.1016/S1474-4422(13)70310-7
4. Hart RG, Sharma M, Mundl H, Kasner SE, Bangdiwala SI, Berkowitz SD, Swaminathan B, Lavados P, Wang Y, Wang Y, et al; NAVIGATE ESUS Investigators. Rivaroxaban for stroke prevention after embolic stroke of undetermined source. *N Engl J Med*. 2018;378:2191–2201. doi: 10.1056/NEJMoa1802686
5. Jørgensen HS, Nakayama H, Reith J, Raaschou HO, Olsen TS. Stroke recurrence: predictors, severity, and prognosis. The Copenhagen Stroke Study. *Neurology*. 1997;48:891–895. doi: 10.1212/wnl.48.4.891
6. Diener HC, Sacco RL, Easton JD, Granger CB, Bernstein RA, Uchiyama S, Kreuzer J, Cronin L, Cotton D, Grauer C, et al; RE-SPECT ESUS Steering Committee and Investigators. Dabigatran for prevention of stroke after embolic stroke of undetermined source. *N Engl J Med*. 2019;380:1906–1917. doi: 10.1056/NEJMoa1813959
7. Paciaroni M, Kamel H. Do the results of RE-SPECT ESUS call for a revision of the embolic stroke of undetermined source definition? *Stroke*. 2019;50:1032–1033. doi: 10.1161/STROKEAHA.118.024160
8. Kamel H, Okin PM, Elkind MS, Iadecola C. Atrial fibrillation and mechanisms of stroke: time for a new model. *Stroke*. 2016;47:895–900. doi: 10.1161/STROKEAHA.115.012004
9. Merkler AE, Sigurdsson S, Eiriksdottir G, Safford MM, Phillips CL, Iadecola C, Gudnason V, Weinsaft JW, Kamel H, Arai AE, et al. Association between unrecognized myocardial infarction and cerebral infarction on magnetic resonance imaging. *JAMA Neurol*. 2019;76:956–961.
10. Ntaios G, Perlepe K, Lambrou D, Sirimarco G, Strambo D, Eskandari A, Karagkiozi E, Vemmou A, Koroboki E, Manios E, et al. Prevalence and overlap of potential embolic sources in patients with embolic stroke of undetermined source. *J Am Heart Assoc*. 2019;8:e012858. doi: 10.1161/JAHA.119.012858
11. Saba L, Saam T, Jäger HR, Yuan C, Hatsukami TS, Saloner D, Wasserman BA, Bonati LH, Wintermark M. Imaging biomarkers of vulnerable carotid plaques for stroke risk prediction and their potential clinical implications. *Lancet Neurol*. 2019;18:559–572. doi: 10.1016/S1474-4422(19)30035-3
12. Longstreth WT Jr, Kronmal RA, Thompson JL, Christenson RH, Levine SR, Gross R, Brey RL, Buchsbaum R, Elkind MS, Tirschwell DL, et al. Amino terminal pro-B-type natriuretic peptide, secondary stroke prevention, and choice of antithrombotic therapy. *Stroke*. 2013;44:714–719. doi: 10.1161/STROKEAHA.112.675942
13. Healey JS, Gladstone DJ, Swaminathan B, Eckstein J, Mundl H, Epstein AE, Haeusler KG, Mikulik R, Kasner SE, Toni D, 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;76:764–773. doi: 10.1001/jamaneurol.2019.0617
14. Sholle ET, Kabariti J, Johnson SB, Leonard JP, Pathak J, Varughese VI, Cole CL, Campion TR Jr. Secondary use of patients' electronic records (SUPER): an approach for meeting specific data needs of clinical and translational researchers. *AMIA Annu Symp Proc*. 2017;2017:1581–1588.

15. Adams HP Jr, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, Marsh EE 3rd. 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;24:35–41. doi: 10.1161/01.str.24.1.35
16. Hart RG, Diener HC, Coutts SB, Easton JD, Granger CB, O'Donnell MJ, Sacco RL, Connolly SJ; Cryptogenic Stroke/ESUS International Working Group. Embolic strokes of undetermined source: the case for a new clinical construct. *Lancet Neurol*. 2014;13:429–438. doi: 10.1016/S1474-4422(13)70310-7
17. Diaz I, Savenkov O, Kamel H. Nonparametric targeted Bayesian estimation of class proportions in unlabeled data. *Biostatistics*. 2020;kxaa022. https://academic.oup.com/biostatistics/article-abstract/doi/10.1093/biostatistics/kxaa022/5856304. Accessed July 14, 2020.
18. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol*. 2007;6:Article25. doi: 10.2202/1544-6115.1309
19. Pirracchio R, Petersen ML, Carone M, Rigon MR, Chevret S, van der Laan MJ. Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study. *Lancet Respir Med*. 2015;3:42–52. doi: 10.1016/S2213-2600(14)70239-5
20. Freilinger TM, Schindler A, Schmidt C, Grimm J, Cyran C, Schwarz F, Bamberg F, Linn J, Reiser M, Yuan C, et al. Prevalence of nonstenosing, complicated atherosclerotic plaques in cryptogenic stroke. *JACC Cardiovasc Imaging*. 2012;5:397–405. doi: 10.1016/j.jcmg.2012.01.012
21. Kamel H, Soliman EZ, Heckbert SR, Kronmal RA, Longstreth WT Jr, Nazarian S, Okin PM. P-wave morphology and the risk of incident ischemic stroke in the Multi-Ethnic Study of Atherosclerosis. *Stroke*. 2014;45:2786–2788. doi: 10.1161/STROKEAHA.114.006364
22. Li L, Yiin GS, Geraghty OC, Schulz UG, Kuker W, Mehta Z, Rothwell PM; Oxford Vascular Study. Incidence, outcome, risk factors, and long-term prognosis of cryptogenic transient ischaemic attack and ischaemic stroke: a population-based study. *Lancet Neurol*. 2015;14:903–913. doi: 10.1016/S1474-4422(15)00132-5
23. Ntaios G, Perlepe K, Sirimarco G, Strambo D, Eskandari A, Karagkiozi E, Vemmou A, Koroboki E, Manios E, Makaritsis K, et al. Carotid plaques and detection of atrial fibrillation in embolic stroke of undetermined source. *Neurology*. 2019;92:e2644–e2652. doi: 10.1212/WNL.0000000000007611
24. Kamel H, Pearce LA, Ntaios G, Gladstone DJ, Perera K, Roine RO, Meseguer E, Shoamanesh A, Berkowitz SD, Mundl H, et al. Atrial cardiopathy and nonstenosing large artery plaque in patients with embolic stroke of undetermined source. *Stroke*. 2020;51:938–943. doi: 10.1161/STROKEAHA.119.028154
25. Kamel H, Merkler AE, Iadecola C, Gupta A, Navi BB. Tailoring the approach to embolic stroke of undetermined source: a review. *JAMA Neurol*. 2019;76:855–861. doi: 10.1001/jamaneurol.2019.0591
26. Hankey GJ. Antithrombotic therapy for stroke prevention. *Circulation*. 2019;139:1131–1133. doi: 10.1161/CIRCULATIONAHA.118.036656
27. Shimizu H, Murakami Y, Inoue S, Ohta Y, Nakamura K, Katoh H, Sakne T, Takahashi N, Ohata S, Sugamori T, et al. High plasma brain natriuretic polypeptide level as a marker of risk for thromboembolism in patients with nonvalvular atrial fibrillation. *Stroke*. 2002;33:1005–1010. doi: 10.1161/hs0402.105657
28. Kamel H, Longstreth WT Jr, Tirschwell DL, Kronmal RA, Broderick JP, Palesch YY, Meinzer C, Dillon C, Ewing I, Spilker JA, et al. The AtRial cardiopathy and antithrombotic drugs in prevention after cryptogenic stroke randomized trial: rationale and methods. *Int J Stroke*. 2019;14:207–214. doi: 10.1177/1747493018799981
29. Geisler T, Poli S, Meisner C, Schreieck J, Zuern CS, Nägele T, Brachmann J, Jung W, Gahn G, Schmid E, et al. Apixaban for treatment of embolic stroke of undetermined source (ATTICUS randomized trial): Rationale and study design. *Int J Stroke*. 2017;12:985–990. doi: 10.1177/1747493016681019