Stroke

BRIEF REPORT

Improving Prehospital Stroke Diagnosis Using Natural Language Processing of Paramedic Reports

Anoop Mayampurath, PhD; Zahra Parnianpour, MS; Christopher T. Richards, MD, MS; William J. Meurer, MD, MS; Jungwha Lee, PhD, MPH; Bruce Ankenman, PhD; Ohad Perry, PhD; Scott J. Mendelson, MD, PhD; Jane L. Holl, MD, MPH; Shyam Prabhakaran, MD, MS

BACKGROUND AND PURPOSE: Accurate prehospital diagnosis of stroke by emergency medical services (EMS) can increase treatments rates, mitigate disability, and reduce stroke deaths. We aimed to develop a model that utilizes natural language processing of EMS reports and machine learning to improve prehospital stroke identification.

METHODS: We conducted a retrospective study of patients transported by the Chicago EMS to 17 regional primary and comprehensive stroke centers. Patients who were suspected of stroke by the EMS or had hospital-diagnosed stroke were included in our cohort. Text within EMS reports were converted to unigram features, which were given as input to a support-vector machine classifier that was trained on 70% of the cohort and tested on the remaining 30%. Outcomes included final diagnosis of stroke versus nonstroke, large vessel occlusion, severe stroke (National Institutes of Health Stroke Scale score >5), and comprehensive stroke center-eligible stroke (large vessel occlusion or hemorrhagic stroke).

RESULTS: Of 965 patients, 580 (60%) had confirmed acute stroke. In a test set of 289 patients, the text-based model predicted stroke nominally better than models based on the Cincinnati Prehospital Stroke Scale (c-statistic: 0.73 versus 0.67, P=0.165) and was superior to the 3-Item Stroke Scale (c-statistic: 0.73 versus 0.53, P<0.001) scores. Improvements in discrimination were also observed for the other outcomes.

CONCLUSIONS: We derived a model that utilizes clinical text from paramedic reports to identify stroke. Our results require validation but have the potential of improving prehospital routing protocols.

GRAPHIC ABSTRACT: An online graphic abstract is available for this article.

Key Words: diagnosis ■ machine learning ■ natural language processing ■ patient ■ retrospective studies

Stroke is a leading cause of disability and mortality, as well as a significant financial burden to the US health care system.[1] Early treatment is associated with better outcomes in patients with stroke; therefore, timely recognition by emergency medical system (EMS) personnel can better prepare hospitals for an incoming patient with stroke.[2,3] While screening tools, such as the Cincinnati Prehospital Stroke Scale (CPSS) and the 3-Item Stroke Scale (3I-SS), which are used in Chicago, IL, have been developed to identify stroke in the prehospital setting,[4,5] they have varying levels of accuracy and completeness in documentation.[6–9] We aimed to investigate whether information extracted from EMS notes can predict diagnosis of stroke, severe stroke, large vessel occlusion (LVO), and hemorrhagic stroke.

Correspondence to: Shyam Prabhakaran, MD, MS, 5841 S. Maryland Ave, MC 2030, A-223, Chicago, IL 60637-1470. Email shyam1@neurology.bsd.uchicago.edu
The Data Supplement is available with this article at https://www.ahajournals.org/doi/suppl/10.1161/STROKEAHA.120.033580.
For Sources of Funding and Disclosures, see page XXX.

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

Nonstandard Abbreviations and Acronyms

3I-SS	3-Item Stroke Scale
AIS	acute ischemic stroke
AUROC	area under the receiver operating characteristic curve
CPSS	Cincinnati Prehospital Stroke Scale
CSC	comprehensive stroke center
EMS	emergency medical system
LVO	large vessel occlusion

METHODS

Study Population

The data that support the findings of this study are available from the corresponding author upon reasonable request from qualified researchers trained in human subject confidentiality protocols. We accessed prehospital care reports for patients transported via EMS after a 9-1-1 call to one of 17 primary or comprehensive stroke centers (CSCs) in Chicago between November 28, 2018 and May 31, 2019. We also accessed the Get With The Guidelines-Stroke (IQVIA, Parsippany, NJ) registry at the 17 stroke centers. We considered an EMS-suspected stroke if "suspected stroke", transport decision to a stroke center, or abnormal prehospital stroke screening were documented in the prehospital record. Patients with hospital-confirmed stroke who arrived by Chicago EMS were included for analysis, even if paramedics did not initially suspect stroke. We excluded patients who were diagnosed with transient ischemic attack. Further details on cohort derivation are provided in the Data Supplement. The study was approved by University of Chicago Internal Review Board through waiver of consent (IRB19-0539).

Outcome and Variables

The primary outcome was acute stroke diagnosis. Secondary outcomes considered were severe stroke, defined as having stroke with a National Institutes of Health Stroke Scale score >5, acute ischemic stroke with large vessel occlusion (acute ischemic stroke [AIS]-LVO), and CSC-eligible stroke, defined as having either AIS-LVO or an intracerebral/subarachnoid hemorrhage. All measures about the primary and secondary outcomes were obtained from the Get With The Guidelines-Stroke-Stroke registry. Our primary features of interest were representations of single word (ie, unigram) occurrences within EMS paramedic reports. Unigrams are initially represented as counts of their presence in a particular patient paramedic report, following which count normalization is performed through term-frequency-inverse document frequency transformation (see Data Supplement).

Analysis Plan

We derived a support-vector-machine classifier with a linear kernel that utilized the text-based features from the paramedic notes to predict stroke diagnosis in patients. We split the cohort randomly into 70% and 30% for model development and test, respectively. Five-fold cross validation was used to determine the best cost hyper-parameter for the support-vector-machine within the development cohort. Baseline models were derived using logistic regression for both the CPSS and 3I-SS categorical scores (see Data Supplement). Model discrimination was assessed using area under the receiver operating characteristic curve (AUROC) and corresponding 95% CIs. Models were derived for other stroke outcomes using a similar strategy.

To examine which clinical terms were important for predicting stroke, we utilized the top 20 unigrams, based on support-vector-machine-weight, to derive following stroke-related phenotypes: "unilaterality", "weakness", "slurred speech", "facial droop", and "minutes ago". We assessed the association between outcomes and one or more of the above phenotypes depending on presence or absence of phenotype related search terms (see Data Supplement and Table I in the Data Supplement). We then compared the distribution of stroke and nonstroke patients across these phenotypes. All analyses were conducted using R version 3.6.2 and Python version 3.7.6, with $P<0.05$ indicating statistical significance.

RESULTS

Patient Characteristics

Our final cohort consisted of 965 patients, among whom 580 (60%) had confirmed stroke diagnosis. There were no observable differences in age, sex, race, initial heart rate, systolic blood pressure, and oxygen saturation, and in-hospital mortality between stroke and nonstroke patients (see Table 1).

Model Performance

Table 2 compares the predictive performance of the text-based models trained on EMS reports to models trained using the CPSS and the 3I-SS scores. On a test dataset of 289 patients, the text-based model performed nominally but not statistically better than the CPSS model (AUROC, 0.73 [95% CI, 0.67–0.79] versus 0.67 [95% CI, 0.61–0.73], P=0.165). However, the text-based model was superior to the 3I-SS model (AUROC, 0.73 [95% CI, 0.67–0.79] versus 0.53 [95% CI, 0.51–0.56], $P<0.001$) for stroke prediction. In a test cohort of 267 patients with and without severe stroke (National Institutes of Health Stroke Scale score >5), the text-based model predicted severe stroke more accurately than the CPSS (AUROC, 0.82 versus 0.70, P=0.006) and 3I-SS (AUROC, 0.82 versus 0.57, $P<0.001$) models. Slight improvements in identifying patients with AIS-LVO was observed for the text-based model in comparison to the other models but was not statistically significant (text-based model AUROC 0.76 versus CPSS AUROC 0.65, P=0.133; text-based model AUROC 0.76 versus 3I-SS AUROC 0.64, P=0.074) in a test set of 232 patients with and without AIS-LVO. The text-based model for predicting CSC-eligible stroke performed better in discriminating CSC-eligible patients compared with both the CPSS

Stroke. 2021;52:00–00. DOI: 10.1161/STROKEAHA.120.033580

Table 1. Comparison of Clinical Characteristics Between Patients With Final Diagnosis of Stroke and Patients With Other Final Diagnosis

	Patients with stroke (n=580)	Patients without stroke (n=385)	*P* value
Age, y; mean (SD)	68 (15)	66 (18)	0.241
Sex, n (%)			
Female	270 (46.6)	173 (44.9)	0.885
Male	253 (43.6)	173 (44.9)	
Unknown	57 (9.8)	39 (10.1)	
Race, n (%)			
Black	226 (39.0)	158 (41.0)	0.528
White	190 (32.8)	106 (27.5)	
Hispanic/Latino	79 (13.6)	57 (14.8)	
Other	25 (4.3)	20 (5.2)	
Unknown	60 (10.3)	44 (11.4)	
Initial vital signs			
Heart rate, mean (SD)	87 (18)	86 (18)	0.241
Systolic blood pressure, mean (SD)	159 (34)	161 (34)	0.345
Oxygen saturation, mean (SD)	96 (6)	97 (5)	0.626
Died in-hospital, n (%)	22 (3.8)	12 (3.1)	0.853
Severe stroke, n (%)	264 (45.5)	…	
AIS-LVO, n (%)	84 (14.5)	…	
CSC-eligible stroke, n (%)	213 (36.7)	…	

Severe stroke is defined as having stroke with an NIHSS score >5. Large vessel occlusion was identified from computerized tomography angiography images (see Data Supplement). CSC-eligible stroke is defined as having either AIS-LVO or an intracerebral/subarachnoid hemorrhage. AIS-LVO indicates acute ischemic stroke with large vessel occlusion; CSC, Comprehensive Stroke Center; and NIHSS, National Institutes of Health Stroke Scale.

(AUROC, 0.88 versus 0.66, *P*<0.001) and the 3I-SS models (AUROC, 0.88 versus 0.54, *P*<0.001) in a test set of 270 patients. The model also demonstrated better sensitivity and specificity when compared with the other models (see Table II in the Data Supplement).

The Figure illustrates the distribution of stroke and nonstroke patients for “unilaterality”, “weakness”, “slurred speech”, “facial droop”, and “minutes ago” stroke-related phenotypes. Patients with stroke had an increased proportion of patients with mentions of stroke-related phenotypes within text as compared with nonstroke patients, thereby demonstrating that the model has good face validity.

DISCUSSION

In this study, we demonstrate that natural language processing with machine learning can extract information from EMS reports to accurately identify stroke and stroke subtypes in the prehospital setting. The text-based models nominally outperformed (but without statistical significance) the CPSS score-based model but outperformed the 3I-SS score-based models in identifying stroke and other stroke outcomes in a diverse, urban population. If validated in future studies in larger populations, our findings have implications for prehospital stroke care and could allow timely diagnosis of stroke facilitating appropriate triage to stroke centers.

Previous studies have developed and validated prehospital stroke scales such as the CPSS and the 3I-SS.[4,5] However, studies evaluating the real-world implementation of prehospital stroke screening have reported limitations in the accuracy, documentation, and reproducibility of prehospital these screens.[6–9] Recently, Uchida et al[10] developed models to predict different types of stroke (any stroke, LVO, intracerebral hemorrhage, and subarachnoid hemorrhage) using a list of 21 variables. However, these models depend on accurate assessment and discrete documentation of these variables, some of which such as medical history may be hard to reliably acquire in a prehospital setting.

To our knowledge, this is the first study to analyze clinical text within paramedic reports using natural language processing to identify stroke and stroke subtypes in a cohort of patients with suspected stroke. The score estimated by our model can be used as an indicator of the likelihood of a confirmed stroke diagnosis enabling informed decision-making about timely hospital triage and prenotification. Study limitations include the retrospective study design and a limited sample size from a single geographic region. Prospective validation and investigations into whether this approach can be generalizable to larger areas or other outcomes of stroke are needed. In addition, our model is dependent on accurate

Table 2. Comparison of Model Performances for Various Outcomes

Model trained to predict outcome	AUROC (95% CI) for stroke (n=289)	AUROC (95% CI) for AIS-LVO (n=232)	AUROC (95% CI) for severe stroke (n=267)	AUROC (95% CI) for CSC eligible patients (n=270)
Text-based model (unigrams)	0.73 (0.67–0.79)	0.77 (0.66–0.87)	0.82 (0.77–0.87)	0.88 (0.82–0.94)
Text-based model (unigrams+bigrams)	0.72 (0.66–0.78)	0.75 (0.65–0.85)	0.81 (0.76–0.87)	0.86 (0.80–0.93)
CPSS	0.67 (0.61–0.73)	0.65 (0.53–0.77)	0.70 (0.63–0.76)*	0.65 (0.58–0.72)*
3I-SS	0.53 (0.51–0.56)*	0.64 (0.55–0.74)	0.57 (0.53–0.61)*	0.66 (0.59–0.73)*

Unigrams are single words (n-grams where n=1), while bigrams are a sequence of 2 consecutive words (n-grams where n=2). 3I-SS indicates 3-Item Stroke Scale; AIS-LVO, acute ischemic stroke with large vessel occlusion; AUROC, area under the receiver operating characteristic curve; CPSS, Cincinnati Prehospital Stroke Scale; and CSC, Comprehensive Stroke Center.

*P<0.001, when compared with the unigram text-based model.

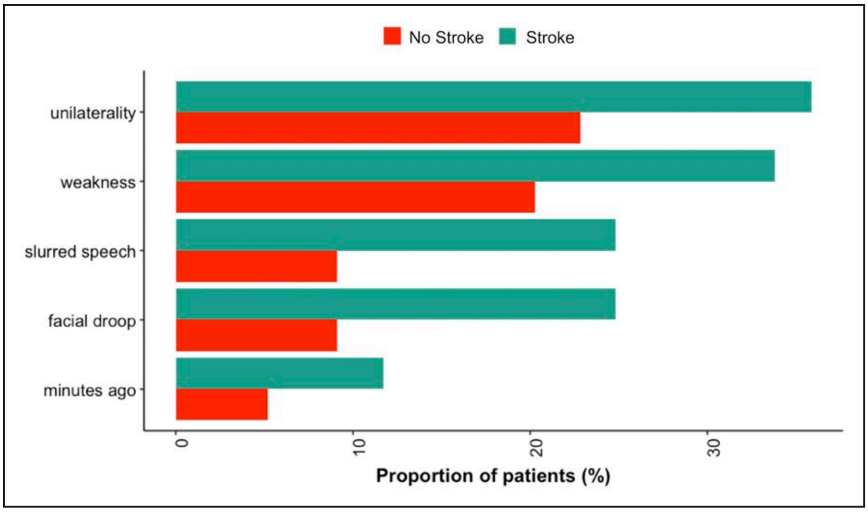

Figure. Distribution of stroke and nonstroke patients across stroke-related phenotypes.

documentation by the EMS providers including assessment of onset of stroke-like symptoms. However, our approach revealed novel phenotypes such as "minutes ago" that would not have been possible using stroke scales. Finally, there may be overlap between patients with CSC-eligible stroke and EMS-transport decisions. However, our outcome of CSC-eligible stroke is based on having either AIS-LVO or an intracerebral/subarachnoid hemorrhage and could not be influenced by paramedic behavior.

ARTICLE INFORMATON

Received October; 19, 2020; final revision received February 20, 2021; accepted March 31, 2021.

Affiliations

Department of Pediatrics (A.M.) and Department of Neurology (Z.P., S.J.M., J.L.H., S.P.), University of Chicago, IL. Department of Emergency Medicine, University of Cincinnati, OH (C.T.R.). Department of Emergency Medicine, University of Michigan, Ann Arbor, IL (W.J.M.). Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL (J.L.). Department of Industrial Engineering and Management Studies, Northwestern University (B.A., O.P.).

Acknowledgments

We thank the Chicago Fire Department and participating hospitals: Advocate Trinity Hospital, Community First Hospital (formerly Presence Our Lady of the Resurrection Medical Center), Holy Cross Hospital, Humboldt Park Health (formerly Norwegian American Hospital), John H. Stroger, Jr Hospital of Cook County, Mercy Hospital and Medical Center, Mount Sinai Hospital Medical Center, Northwestern Memorial Hospital, AMITA Health (Resurrection Medical Center, Saint Joseph Hospital, Saints Mary and Elizabeth Medical Center), Rush University Medical Center, Saint Anthony Hospital, Swedish Covenant Hospital, University of Chicago Hospitals and Health Systems, University of Illinois Hospital Health Sciences System at Chicago, and Louis A. Weiss Memorial Hospital.

Sources of Funding

Dr Mayampurath is supported by a career development award from the National Heart, Lung, and Blood Institute (K01HL148390). Dr Prabhakaran is supported by grants from NIH and Agency for Healthcare Research and Quality (AHRQ) (U24NS10723; R18HS025359; R18HS027264). All other authors are supported by grant from AHRQ R18HS025359.

Disclosures

Dr Mayampurath performed consulting services for Litmus Health, Austin, TX. Dr Richards has received personal fees from American College of Emergency Medicine, personal fees from American Heart Association, and personal fees from Illinois EMT Association outside the submitted work. Dr Meurer has received personal fees from Meurer Consulting, LLC outside the submitted work. Dr Prabhakaran has received personal fees from Abbvie, and personal fees from UpToDate outside the submitted work. The other authors report no conflicts.

Supplemental Materials

Expanded Methods and Results
Online Tables I and II
Online Figure I

REFERENCES

1. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, et al; American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics–2015 update: a report from the American Heart Association. *Circulation*. 2015;131:e29–322. doi: 10.1161/CIR.0000000000000152
2. Lees KR, Bluhmki E, von Kummer R, Brott TG, Toni D, Grotta JC, Albers GW, Kaste M, Marler JR, Hamilton SA, et al; ECASS, ATLANTIS, NINDS and EPITHET rt-PA Study Group. Time to treatment with intravenous alteplase and outcome in stroke: an updated pooled analysis of ECASS, ATLANTIS, NINDS, and EPITHET trials. *Lancet*. 2010;375:1695–1703. doi: 10.1016/S0140-6736(10)60491-6
3. Audebert HJ, Saver JL, Starkman S, Lees KR, Endres M. Prehospital stroke care: new prospects for treatment and clinical research. *Neurology*. 2013;81:501–508. doi: 10.1212/WNL.0b013e31829e0fdd
4. Kothari RU, Pancioli A, Liu T, Brott T, Broderick J. Cincinnati prehospital stroke scale: reproducibility and validity. *Ann Emerg Med*. 1999;33:373–378. doi: 10.1016/s0196-0644(99)70299-4
5. Singer OC, Dvorak F, du Mesnil de Rochemont R, Lanfermann H, Sitzer M, Neumann-Haefelin T. A simple 3-item stroke scale. *Stroke*. 2005;36:773–776.
6. Oostema JA, Konen J, Chassee T, Nasiri M, Reeves MJ. Clinical predictors of accurate prehospital stroke recognition. *Stroke*. 2015;46:1513–1517. doi: 10.1161/STROKEAHA.115.008650
7. English SW, Rabinstein AA, Mandrekar J, Klaas JP. Rethinking prehospital stroke notification: assessing utility of emergency medical services impression and cincinnati prehospital stroke scale. *J Stroke Cerebrovasc Dis*. 2018;27:919–925. doi: 10.1016/j.jstrokecerebrovasdis.2017.10.036
8. Brandler ES, Sharma M, Sinert RH, Levine SR. Prehospital stroke scales in urban environments: a systematic review. *Neurology*. 2014;82:2241–2249. doi: 10.1212/WNL.0000000000000523
9. Turc G, Maïer B, Naggara O, Seners P, Isabel C, Tisserand M, Raynouard I, Edjlali M, Calvet D, Baron JC, et al. Clinical scales do not reliably identify acute ischemic stroke patients with large-artery occlusion. *Stroke*. 2016;47:1466–1472.
10. Uchida K, Yoshimura S, Hiyama N, Oki Y, Matsumoto T, Tokuda R, Yamaura I, Saito S, Takeuchi M, Shigeta K, et al. Clinical prediction rules to classify types of stroke at prehospital stage. *Stroke*. 2018;49:1820–1827.