Automating Ischemic Stroke Subtype Classification Using Machine Learning and Natural Language Processing

Ravi Garg, MS,* Elissa Oh, MS,* Andrew Naidech, MD, MSPH,* Konrad Kording, PhD,† and Shyam Prabhakaran, MD, MS‡

Objective: The manual adjudication of disease classification is time-consuming, error-prone, and limits scaling to large datasets. In ischemic stroke (IS), subtype classification is critical for management and outcome prediction. This study sought to use natural language processing of electronic health records (EHR) combined with machine learning methods to automate IS subtyping. *Methods*: Among IS patients from an observational registry with TOAST subtyping adjudicated by board-certified vascular neurologists, we analyzed unstructured text-based EHR data including neurology progress notes and neuroradiology reports using natural language processing. We performed several feature selection methods to reduce the high dimensionality of the features and 5-fold cross validation to test generalizability of our methods and minimize overfitting. We used several machine learning methods and calculated the kappa values for agreement between each machine learning approach to manual adjudication. We then performed a blinded testing of the best algorithm against a held-out subset of 50 cases. *Results*: Compared to manual classification, the best machine-based classification achieved a kappa of .25 using radiology reports alone, .57 using progress notes alone, and .57 using combined data. Kappa values varied by subtype being highest for cardioembolic (.64) and lowest for cryptogenic cases (.47). In the held-out test subset, machine-based classification agreed with rater classification in 40 of 50 cases (kappa .72). *Conclusions*: Automated machine learning approaches using textual data from the EHR shows agreement with manual TOAST classification. The automated pipeline, if externally validated, could enable large-scale stroke epidemiology research.

Key Words: Ischemic stroke—cryptogenic—cardioembolism—natural language processing—machine learning

Introduction

Ischemic stroke (IS) is a major cause of disability in the United States.[1] However, outcomes and management vary by ischemic stroke subtype. In addition, IS subtype has research implications since it informs selection into trials (eg, ongoing trials for cryptogenic stroke).[2] Classification of ischemic stroke subtype requires synthesis of historical, examination, laboratory, electrocardiographic, and imaging data to infer a mechanism and assign causal, etiologic, or phenotypic classification. Many different schemas have been proposed including the Trial of Org10172 in Acute Stroke (TOAST) classification,[3] Causative Classification System (CCS),[4] and Atherosclerosis, Small-vessel disease, Cardioembolism and Other causes (ASCO) system.[5] The most commonly used of these is the TOAST classification, which is simple, has modest inter-rater reliability, and is often used in stroke research projects. Compared to the TOAST system, the CSS and ASCO have higher inter-rater reliability but are more complex and therefore less commonly used in practice or in research.[6,7]

From the *Department of Neurology, Northwestern University, Feinberg School of Medicine, Chicago, Illinois; †University of Pennsylvania, Philadelphia, Pennsylvania; and ‡Department of Neurology, Pritzker School of Medicine, University of Chicago, Chicago, Illinois.

Received August 20, 2018; revision received January 30, 2019; accepted February 9, 2019.

Sources of Funding: None.

Address correspondence to Shyam Prabhakaran, MD, MS, Department of Neurology, Northwestern University, Feinberg School of Medicine, 633 St. Clair St. #2041, Chicago, IL 60611. E-mail: shyam1@uchicago.edu

1052-3057/$ - see front matter

https://doi.org/10.1016/j.jstrokecerebrovasdis.2019.02.004

Given the importance of IS subtyping, many curated registries manually abstract key historical details, clinical findings, laboratory tests, cardiac monitoring, and imaging results while also requiring a review of brain and neurovascular imaging by a trained rater, usually a board-certified stroke neurologist. This process is labor-intensive, introduces variability between raters, and cannot readily be scaled for large datasets of thousands or millions of records. Leveraging textual data in the electronic health record (EHR) and applying machine learning could overcome these limitations.

In this study, we tested the hypothesis that advanced machine learning methods performed on features identified using natural language processing (NLP) of unstructured textual data in the EHR could identify TOAST subtype with high concordance and inter-rater reliability compared to manually determined TOAST subtype performed by board-certified stroke neurologists.

Methods

Cohort

The local Institutional Review Board approved the study. Consecutive patients ≥18 years with a confirmed acute ischemic stroke over a period of 24 months (August 1, 2012 through July 31, 2014) were enrolled in the Northwestern University Brain Attack Registry with informed consent, as previously described.[8] Diagnosis of stroke was made by a board-certified stroke neurologist at our institution in each case. Board-certified neurologists prospectively and independently reviewed clinical and radiographic data to determine TOAST subtype by consensus adjudication.[9] In a subset of patients in year 1 of the study, a single rater applied the CCS method for stroke subtype classification using the website created for this purpose (https://ccs.mgh.harvard.edu/ccs_intro.php).

Data Extraction and Preparation

We obtained free-form text notes from the EHR (Cerner, Kansas City, MO; Epic, Madison, WI) pertaining to the index stroke hospitalization for all patients using the Northwestern Enterprise Data Warehouse (EDW), a database that collects and integrates data from the EHR at NMH and other Northwestern Healthcare System practice settings (Fig 1). The EDW currently contains clinical data on nearly 3 million patients dating back to the 1970s, which can be easily queried at the individual patient level or for aggregate data and can link laboratory tests, procedures, therapies, and clinical data with clinical outcomes at specific points in time.

A data analyst extracted the notes from the EDW. We preprocessed them to make it usable for machine learning. First, we combined the raw data with the adjudicated TOAST subtypes linking by a common identifier. We included only a small appropriate subset of report types to identify the stroke subtype: admission, progress, and discharge vascular neurology notes and radiology notes from brain magnetic resonance imaging (MRI) and magnetic resonance angiography (MRA) reports. Finally, we processed all the text for each type of report in order to remove language abnormalities. Specifically, we lowercased the text, removed punctuations, stop words, and nonalphanumeric words. After performing these preprocessing steps, we finally attained a clean corpus and labels for data analysis.

We transformed the textual corpus to a format to feed to machine learning algorithms. We first constructed a vocabulary for the corpus. All the reports for each patient are concatenated to form a large document. We considered all the noun phrases as potential candidates for the vocabulary. The noun phrases, which can be unigram, bigram or trigrams, were extracted using python-based natural language processing library spaCy. We also experimented with medical concepts extracted using package Apache cTAKES[10] and used them as a separate vocabulary. Next, we converted each document to a count-based matrix, where for each word/phrase in the vocabulary we record the number of times that word occurs in the document. In addition, we also experimented with frequency-based matrix where instead of counts we used term frequency in each document normalized by document frequency. Performing these feature-engineering steps, we obtained a feature matrix for machine learning algorithms to operate on. We then compared both the vocabularies (ie, medical concepts extracted using cTAKES and n-grams extracted using spaCy) with the baseline classifier of logistic regression. Finally, on the results obtained from the training set we discarded the features extracted using cTAKES.

Analysis

To deal with high dimensionality due to the large vocabulary, we used feature selection methods to reduce the dimensionality of the feature matrix. We performed principal component analysis (PCA) on the original matrix. PCA is a statistical procedure that converts observations of possibly correlated variables into a set of linearly uncorrelated variables called principal components.[11] The first principal component has the largest possible variance accounting for as much as variability in data set as possible. Each succeeding component has highest possible variance under condition that it is orthogonal to the components preceding it. We plotted the cumulative variance explained versus number of principal components and selected the first 150 principal components based on visual inspection.[12] We also performed XGBoost[13] based feature selection on the original feature-matrix to rank and select from original features. XGBoost is a decision tree-based method, which constructs many trees in forward greedy based fashion. This method then inherently

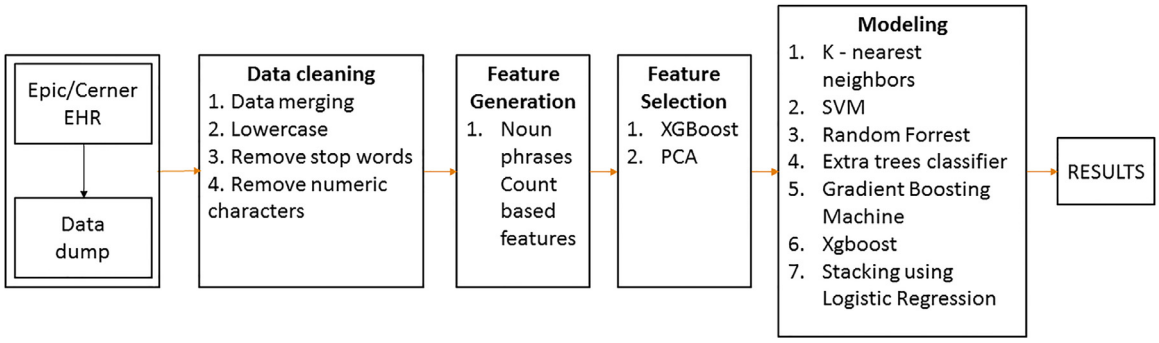

Figure 1. *Pipeline of steps applied for data extraction and analysis. Abbreviations: EHR, electronic health record; PCA, principal component analysis; SVM, support vector machines.*

ranks each feature or variable by how well it improves the purity of the node. Nodes with the greatest decrease in impurity happen at the start of the trees, while nodes with the least decrease in impurity occur at the end of trees. Thus, by pruning trees below a particular node, we create a subset of the most important features along with their relative importance. Once the features are ranked, we select the top 25% of features. We combined the features obtained from XGBoost with the 150 first principal components to obtain the final feature matrix. This process of feature selection is done within each cross-validation fold.

We used a variety of machine learning models for training the automated system. Specifically, we tried K-nearest neighbors (KNN),[14] Support vector machines (SVM).[15] random forests (RF),[16] extra randomized trees classifiers,[17] gradient boosting machines,[18] and extreme gradient boosting (XGBoost).[13] We also performed stacking[19] wherein prediction from these models is used by a meta-classifier to generate a new model. The stacked model or second-level model combines information from multiple predictive base models (first-level models) and often times outperforms each of individual base model due its smoothing nature and ability to highlight each base model where it performs best and discredit each base model where it performs poorly. In our experiments, we compared logistic regression and XGBoost as meta-classifiers with logistic regression performing better in our case.

To avoid over-fitting, we performed 5-fold cross validation within the training dataset. Cross validation,[20] also called rotation estimation, is a technique to evaluate predictive models by partitioning the original sample into a training set to train the model and a validation set to evaluate it. In k-fold cross validation, the original sample is randomly partitioned into k equal size subsamples. Of the k subsamples, a single subsample is retained as the validation data for testing the model, and the remaining k-1 subsamples are used as training data. The cross-validation process is then repeated k times (the folds), with each of the k subsamples used exactly once as the validation data. The k results from the folds can then be averaged (or otherwise combined) to produce a single estimation. In our case, we performed 5-fold cross validation within the training dataset and calculated inter-rater agreement (kappa, κ) and the discrimination (c-statistic) across each of 5 folds for each machine learning approach compared with manually adjudicated classification, adjusted for multiple comparisons using the Bonferroni method.[21] Finally, we applied the best performing classifier based on the highest kappa value in the cross-validation training dataset to the held-out batch of cases (n = 50) for blinded testing between machine classification and manual classification by one rater (S.P).

In sensitivity analysis, we assessed the performance of the best performing classifier on a subset of patients in whom CCS subtype had been determined.

Results

Between July 2012 and August 2014, 1091 confirmed stroke patients met our inclusion criteria (Fig 2). The

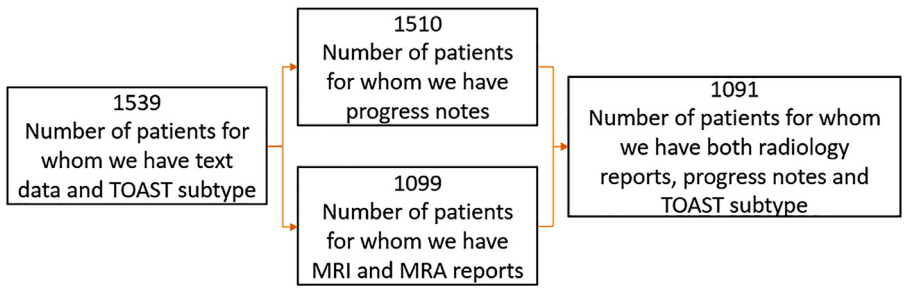

Figure 2. *Flowchart of cohort assembly used for final analysis.*

Table 1. *Baseline characteristics of the analyzed cohort (n = 1091)*

Age in years (SD)	66.0 (15.6)
Male, n (%)	555 (50.9)
Race, n (%)	
White	699 (64.1)
Black	339 (31.1)
Other	53 (4.9)
Hispanic, n (%)	88 (8.1)
Hypertension, n (%)	818 (75.0)
Hyperlipidemia, n (%)	707 (64.8)
Atrial fibrillation, n (%)	137 (12.6)
Coronary artery disease, n (%)	190 (17.4)
Diabetes mellitus, n (%)	303 (27.8)
Median NIHSS score (IQR)	2 (1-6)

baseline characteristics of the cohort are described in Table 1. The mean age was 66 years, 51% were male and 64% were white.

The best performing methods were stacking with logistic regression and XGBoost, followed by gradient boost trees, random forest, extra trees classifier, SVM, and KNN. Compared to manual classification (Table 2), the best machine-based classification (stacking) achieved kappa of .25 (95% confidence interval [CI] .19-.32) using radiology reports alone, .57 (95% CI .51-.62) using progress notes alone, and .57 (95% CI .49-.61) using combined data. The discrimination of TOAST subtypes (Supplemental Figure) was excellent (c-statistic >.8) using any of the machine learning approaches except KNN (c-statistic .67). The mean probability or certainty of the stacking-based classification was higher for concordant cases versus discordant cases (.74 versus .60). Calibration plots for each of 5 TOAST subtypes using the best performing classifier showed good fitness (Fig 3). In a blinded test set of 50 patients, there was 80% agreement between the stroke neurologist and machine-based TOAST classification with a kappa of .72 using stacking with logistic regression. The 10 discordant cases are described in Supplemental Table. The top features selected in the best performing stacked models were as follows: severe stenosis, atrial fibrillation, intracranial atherosclerotic, and lacunar.

Across TOAST subtypes, the machine-based approach had varying agreement with manual adjudication (Table 3) being highest kappa for cardioembolic stroke (stacking: κ = .64 [.53-.71]), large artery atherosclerosis (stacking: κ = .62 [.56-.68]), small artery disease (XGBoost: κ = .57 [.51-.64]), and other determined (stacking and gradient boost trees: κ = .56 [.44-.67]), and lowest for cryptogenic (stacking and XGBoost: κ = .47 [.42-.52]). The machine versus manual rater matrix of TOAST subtypes is shown in Table 4. Compared to manually assigned subtype, machine-based classification over-assigned patients into the cryptogenic stroke subtype (Table 4; 38.1% versus 31.8%). Table 2 provides simple kappa for classification across all 5 TOAST subtypes. Table 3 though provides kappa for each TOAST subtype classification compared against all the others (all $p < .001$ after adjustment for multiple comparisons [n = 5]).

In sensitivity analysis in 409 patients with CCS subtype, machine-based classification showed fair agreement (κ = .40 [.32-.48]), modest accuracy (.58 [.53-.63]), and excellent discrimination (.8 [.75-.84]) compared with website-based classification. Agreement was lowest for undetermined etiology (.14 [.04-.22]) and other causes (.25 [.20-.30]) and highest for large artery atherosclerosis (.56 [.47-.65]).

Discussion

In a prospectively adjudicated dataset of over 1000 records, machine learning techniques using natural language processing of textual data in the EHR demonstrated modest agreement and high discrimination against manual classification for subtyping ischemic stroke into 5 TOAST categories. Since the TOAST classification system is the most widely used stroke classification system across many hospitals and research studies, our results have wide-reaching implications. Our approach could pave the way for large-scale stroke epidemiologic projects wherein ischemic stroke subtyping could be derived from the EHR in thousands or even millions of patients, reducing manual effort and costs. Furthermore, a similar approach could be applied to other conditions wherein manual adjudication from EHR data is required. Prior studies

Table 2. *Agreement between machine learning approaches to TOAST classification using radiology reports, clinical notes, and combined data in the derivation and 5-fold validation dataset (n = 1091) and using combined data in the test dataset (n = 50) compared to manual adjudication*

	Radiology reports only	Clinical notes only	Combined data	Test data (n = 28)
K-nearest neighbor	κ = .09 (.05-.13)	κ = .18 (.15-.22)	κ = .21 (.15-.27)	κ = .17
Support vector machines	κ = .17 (.11-.23)	κ = .44 (.34-.54)	κ = .38 (.33-.43)	κ = .35
Random forest	κ = .20 (.14-.26)	κ = .46 (.39-.53)	κ = .48 (.44-.52)	κ = .57
Extra trees classifier	κ = .18 (.11-.26)	κ = .45 (.37-.53)	κ = .44 (.38-.49)	κ = .49
Gradient boosting trees	κ = .24 (.19-.30)	κ = .56 (.48-.63)	κ = .54 (.49-.60)	κ = .70
XGBoost	κ = .25 (.21-.30)	κ = .55 (.49-.60)	κ = .56 (.50-.61)	κ = .72
Stacking using logistic regression	κ = .25 (.19-.32)	κ = .57 (.51-.62)	κ = .57 (.49-.61)	κ = .72

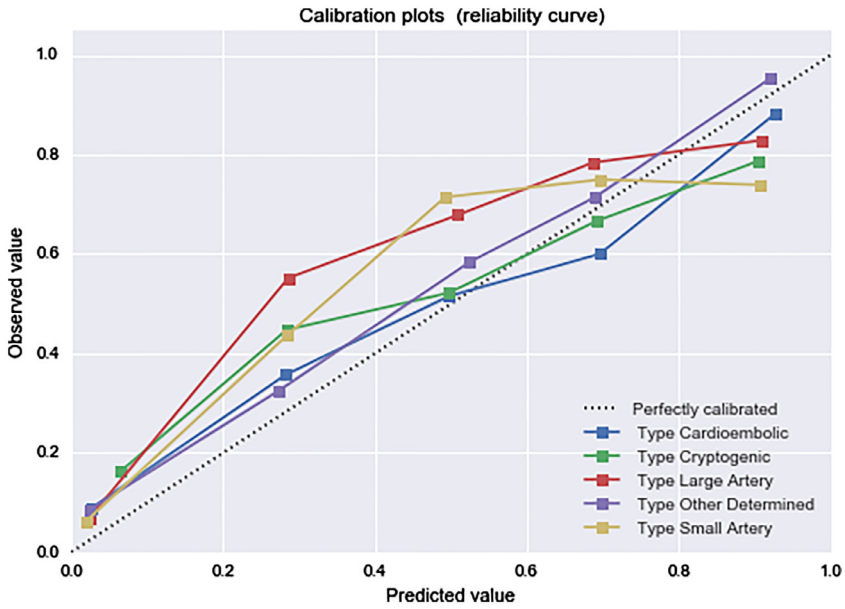

Figure 3. *Calibration plots of observed versus predicted probability for each TOAST subtype.*

using the TOAST classification system have noted modest overall inter-rater reliability between human raters, with kappa ranging from .4 for cryptogenic stroke to .8 for large artery atherosclerosis and cardioembolic stroke subtypes.[22] Others have tried to improve and modify the TOAST criteria with slightly improved kappa values.[9] Our results suggest that automated machine-based approaches have similar degree of agreement with manual classification by trained raters and with similar performance amongst TOAST subtypes. Besides the efficiency of our approach, another advantage of a machine-based approach is that it would not suffer from inter- or intra-rater variability since a single standardized algorithm would be uniformly applied each time.

In sensitivity analysis using the CCS schema, we found that a machine-based approach performed less well, being particularly poor for undetermined subtype. The CCS and TOAST showed only moderate agreement (κ = .59) in the largest multicenter study by McArdle et al.[23] The 2 schema have differences in criteria, required tests, and definitions. For example, many patients with possible cardioaortic embolism by CCS would be classified as cryptogenic stroke using the TOAST classification. Thus, it is not surprising that we found worse performance using machine learning for CCS than with TOAST.

The EHR provides a rich source of information for studying stroke, allowing researchers to leverage heterogeneous data that are collected in the normal processes of clinical care. However, much of the potentially useful information is in the form of unstructured, free text reports. We demonstrate that NLP and machine learning can be applied to the rich, vast stores of data to mine the EHR for clinical syndrome descriptors, test results and findings, and demographics and risk factors to predict the most plausible stroke subtype classification. While NLP has been applied for other conditions,[24,25] we are not aware of prior work in stroke. Furthermore, future work could parse cryptogenic stroke patients into sensible phenotypes based on similarity to other known subtypes (eg, atrial cardiopathy or large artery plaque).[26]

This line of investigation has implications for future stroke health services research. In its 2014

Table 3. *Agreement between machine learning approaches with manual classification for each TOAST subtype*

Classifier	Cardioembolic	Cryptogenic	Large artery atherosclerosis	Other determined	Small artery disease
K-nearest neighbor	κ = .21 (.18-.24)	κ = .09 (.02-.15)	κ = .31 (.24-.38)	κ = .33 (.20-.47)	κ = .21 (.03-.39)
Support vector machines	κ = .46 (.41-.51)	κ = .26 (.20-.32)	κ = .37 (.28-.46)	κ = .46 (.34-.58)	κ = .42 (.24-.60)
Random forest	κ = .58 (.57-.60)	κ = .38 (.33-.43)	κ = .58 (.48-.68)	κ = .50 (.37-.63)	κ = .40 (.34-.47)
Extra trees classifier	κ = .53 (.45-.60)	κ = .32 (.26-.39)	κ = .52 (.42-.61)	κ = .46 (.32-.60)	κ = .44 (.32-.56)
Gradient boosting trees	κ = .62 (.53-.70)	κ = .45 (.41-.49)	κ = .59 (.56-.61)	κ = .56 (.44-.68)	κ = .55 (.46-.64)
XGBoost	κ = .63 (.57-.70)	κ = .47 (.41-.52)	κ = .61 (.55-.68)	κ = .53 (.42-.65)	κ = .57 (.51-.64)
Stacking using logistic regression	κ = .64 (.53-.71)	κ = .47 (.42-.52)	κ = .62 (.56-.68)	κ = .56 (.44-.67)	κ = .55 (.51-.60)

Table 4. *Matrix of manual versus XGBoost-based classification of TOAST subtypes XGBoost-based classification of TOAST subtypes*

	Predicted subtype by XGBoost machine learning method					
Manually assigned subtype	Cardio embolic	Cryptogenic	Large artery atherosclerosis	Other causes	Small artery disease	Total
Cardioembolic	164	49	6	10	4	233
Cryptogenic	31	256	20	20	20	347
Large artery atherosclerosis	11	38	130	9	11	199
Other determined	16	47	8	88	7	166
Small artery disease	6	34	13	2	91	146
Total	228	424	177	129	133	1091

recommendations, the National Heart Lung, and Blood Institute Board of External Experts stated there is a need for "studies that validity, reliability, and scalability of electronic tools for primary data collection." Indeed, traditional research methods (eg, surveys, in-person visits, paper-based data collection forms, and manual data entry and maintenance) are labor intensive and costly and cannot easily be repurposed or scaled for nationwide and multihospital regional registries. Since stroke subtype is a major predictor of outcomes after stroke, our approach would facilitate large-scale electronic epidemiologic projects to study temporal and regional patterns, disparities, and impact of new treatments in practice.

Our study has several limitations. First, our method relies on the level of documentation and detail in the EHR. Our study was performed at an academic medical center, wherein board-certified stroke neurologists evaluate and document findings on each stroke patient with high fidelity. Our results may not generalize well to hospitals without highly trained stroke experts or where consultation on each patient is not required or feasible. Second, while we included text from commonly available notes and reports, we did not include the entire EHR (eg, cardiac imaging, laboratory, procedures) though these are often mentioned or summarized in the clinical notes. We suspect that a more comprehensive approach of all data elements would improve performance but with substantially more computational resource utilization. Third, we did not include CT-based radiology reports to reduce variability in the dataset. Including CT angiography may be particularly relevant for future development since this is increasingly used to identify vascular occlusion and stenosis. Fourth, the consolidated EDW can introduce inaccuracy and missingness, as a result of how its tables are created, patient identity maintenance and matching, and other multifaceted dimensions.

Conclusion

In summary, we found that machine learning approaches using natural language processing of text data in the EHR perform moderately well in classifying TOAST subtypes compared to human rating. Our results have implications for future large-scale electronic epidemiologic research projects wherein manual adjudication of disease classification and outcomes is required. In addition, machine learning could help further parse ischemic stroke into etiologic phenotypes, especially for cryptogenic stroke wherein subgroups similar to cardioembolic and large-artery atherosclerosis subtypes are likely to be found.

Disclosures
None.

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.jstrokecerebrovasdis.2019.02.004.

References

1. Benjamin EJ, Blaha MJ, Chiuve SE, et al. Heart disease and stroke statistics—2017 Update: a report from the American Heart Association. Circulation 2017;135:e146-e603.
2. Diener HC, Bernstein R, Hart R. Secondary stroke prevention in cryptogenic stroke and embolic stroke of undetermined source (ESUS). Curr Neurol Neurosci Rep 2017;17:64.
3. Adams HP, 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;24:35-41.
4. Ay H, Benner T, Arsava EM, et al. A computerized algorithm for etiologic classification of ischemic stroke: the causative classification of stroke system. Stroke 2007;38:2979-2984.
5. Amarenco P, Bogousslavsky J, Caplan LR, et al. The ASCOD phenotyping of ischemic stroke (updated ASCO phenotyping). Cerebrovasc Dis 2013;36:1-5.
6. Gökçal E, Niftaliyev E, Asil T. Etiological classification of ischemic stroke in young patients: a comparative study of TOAST, CCS, and ASCO. Acta Neurol Belg 2017;117:643-648.
7. Marnane M, Duggan CA, Sheehan OC, et al. Stroke subtype classification to mechanism-specific and undetermined categories by TOAST, A-S-C-O, and causative classification

system: direct comparison in the North Dublin population stroke study. Stroke 2010;41:1579-1586.
8. Lin C, Lee J, Chatterjee N, et al. Predicting domain-specific health-related quality of life using acute infarct volume. Stroke 2017;48:1925-1931.
9. Goldstein LB, Jones MR, Matchar DB, et al. Improving the reliability of stroke subgroup classification using the Trial of ORG 10172 in Acute Stroke Treatment (TOAST) criteria. Stroke 2001;32:1091-1098.
10. Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc 2010;17:507-513.
11. Abdi H, Williams LJ. Principal component analysis. WIREs Comp Stat 2010;2:433-459.
12. Deerwester S, Dumais ST, Furnas GW, et al. Indexing by latent semantic analysis. J Am Soc Inform Sci 1990;41:391.
13. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016. p. 785-794.
14. Peterson LE. K-nearest neighbor. Scholarpedia 2009;4:1883.
15. Cortes C, Vapnik V. Support-vector networks. Mach Learning 1995;20:273-297.
16. Liaw A, Wiener M. Classification and regression by randomForest. R News 2002;2:18-22.
17. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learning 2006;63:3-42.
18. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat 2001;54:1189-1232.
19. Džeroski S, Ženko B. Is combining classifiers with stacking better than selecting the best one. Mach Learning 2004;54:255-273.
20. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence; 1995. p. 1137-1143.
21. Dunn OJ. Multiple comparisons among means. J Am Stat Assoc 1961;56:52-64.
22. Meschia JF, Barrett KM, Chukwudelunzu F, et al. Worrall BB and siblings with ischemic stroke study I. Interobserver agreement in the Trial of Org 10172 in Acute Stroke Treatment classification of stroke based on retrospective medical record review. J Stroke Cerebrovasc Dis 2006;15:266-272.
23. McArdle PF, Kittner SJ, Ay H, et al. Agreement between TOAST and CCS ischemic stroke classification: the NINDS SiGN study. Neurology 2014;83:1653-1660.
24. Agarwal A, Baechle C, Behara R, et al. A natural language processing framework for assessing hospital readmissions for patients with COPD. IEEE J Biomed Health Inform 2018;22:588-596.
25. Cai T, Giannopoulos AA, Yu S, et al. Natural language processing technologies in radiology research and clinical applications. Radiographics 2016;36:176-191.
26. Yaghi S, Kamel H, Elkind MSV. Atrial cardiopathy: a mechanism of cryptogenic stroke. Exp Rev Cardiovasc Ther 2017;15:591-599.