7-1-2021

Discrepancies in Stroke Distribution and Dataset Origin in Machine Learning for Stroke.

Lohit Velagapudi
Thomas Jefferson University

Nikolaos Mouchtouris
Thomas Jefferson University

Michael P Baldassari
Thomas Jefferson University

David Nauheim
Thomas Jefferson University

Omaditya Khanna
Thomas Jefferson University

See next page for additional authors

Recommended Citation

Velagapudi, Lohit; Mouchtouris, Nikolaos; Baldassari, Michael P; Nauheim, David; Khanna, Omaditya; Saiegh, Fadi Al; Herial, Nabeel; Gooch, M Reid; Tjoumakaris, Stavropoula; Rosenwasser, Robert H; and Jabbour, Pascal, "Discrepancies in Stroke Distribution and Dataset Origin in Machine Learning for Stroke." (2021). *Department of Neurosurgery Faculty Papers.* Paper 186.
https://jdc.jefferson.edu/neurosurgeryfp/186

Authors

Lohit Velagapudi, Nikolaos Mouchtouris, Michael P Baldassari, David Nauheim, Omaditya Khanna, Fadi Al Saiegh, Nabeel Herial, M Reid Gooch, Stavropoula Tjoumakaris, Robert H Rosenwasser, and Pascal Jabbour

Discrepancies in Stroke Distribution and Dataset Origin in Machine Learning for Stroke

Lohit Velagapudi, BS[1]; Nikolaos Mouchtouris, MD[1]; Michael P. Baldassari, BA[1]; David Nauheim, BA[1]; Omaditya Khanna, MD[1]; Fadi Al Saiegh, MD[1]; Nabeel Herial, MD, MPH[1]; M. Reid Gooch, MD[1]; Stavropoula Tjoumakaris MD[1]; Robert H. Rosenwasser MD, MBA[1]; Pascal Jabbour MD[1]

1. Department of Neurosurgery, Thomas Jefferson University, Philadelphia, PA

Corresponding Author:

Pascal Jabbour MD

909 Walnut Street, 2nd Floor

Philadelphia, Pennsylvania

19107

(215) 955 7000

Abstract Length: 304 Words

Text length: 1758 Words

Key Words: Machine learning, stroke, bias, epidemiology

1 Table, 2 Figures

Abstract

Background: Machine learning algorithms depend on accurate and representative datasets for training in order to become valuable clinical tools that are widely generalizable to a varied population. We aim to conduct a review of machine learning uses in stroke literature to assess the geographic distribution of datasets and patient cohorts used to train these models and compare them to stroke distribution to evaluate for disparities.

Aims: 582 studies were identified on initial searching of the PubMed database. Of these studies, 106 full texts were assessed after title and abstract screening which resulted in 489 papers excluded. Of these 106 studies, 79 were excluded due to using cohorts from outside the United States or being review articles or editorials. 27 studies were thus included in this analysis.

Summary of Review: Of the 27 studies included, 7 (25.9%) used patient data from California, 6 (22.2%) were multicenter, 3 (11.1%) were in Massachusetts, 2 (7.4%) each in Illinois, Missouri, and New York, and 1 (3.7%) each from South Carolina, Washington, West Virginia, and Wisconsin. 1 (3.7%) study used data from Utah and Texas. These were qualitatively compared to a CDC study showing the highest distribution of stroke in Mississippi (4.3%) followed by Oklahoma (3.4%), Washington D.C. (3.4%), Louisiana (3.3%), and Alabama (3.2%) while the prevalence in California was 2.6%.

Conclusions: It is clear that a strong disconnect exists between the datasets and patient cohorts used in training machine learning algorithms in clinical research and the stroke distribution in which clinical tools using these algorithms will be implemented. In order to ensure a lack of bias and increase generalizability and accuracy in future machine learning studies, datasets using a varied patient population that reflects the unequal distribution of stroke risk factors would greatly benefit the usability of these tools and ensure accuracy on a nationwide scale.

Introduction

Machine learning (ML) has markedly increased the capabilities of physician researchers to diagnose, treat, prognosticate, and even anticipate disease using clinical data. Different fields of medicine have embraced ML, and its implementation is set to escalate as clinicians become more comfortable with ML as an analytical technique[1–4]. In stroke research, multiple factors such as age, socioeconomic status, access to healthcare, and healthcare practices have been associated with differing prevalence of stroke. While ML is more resistant to biased analyses by reducing the human role in the analytical process, it is still subject to selection bias as large academic institutions with dedicated specialists in biostatistics and computer science are more likely to produce ML-based research. As has been described in previous literature, the quality and generalizability of machine learning algorithms depends heavily on dataset quality and relevance to the population that these clinical tools will be used for. Kaushal et al recently found that data used in ML research across multiple specialties was sourced disproportionately from geographic locations close to major academic centers, such as California, New York, and Massachusetts, and suggested that geographic distribution of datasets can be a major source of systematic bias in machine learning algorithms[5]. While this study assessed machine learning in multiple fields, including radiology, ophthalmology, dermatology, pathology, gastroenterology, and cardiology, such an analysis has not been performed on neurosurgical datasets used in machine learning, such as stroke, which are uniquely susceptible to bias due to socioeconomic and demographic factors affecting known stroke risk factors and comorbidities. We aim to assess geographic distribution of datasets used in studies on stroke using machine learning to identify disparities in dataset distribution and stroke prevalence.

Methods

A review of the literature was performed on the PubMed Database of all literature published between January 1, 2010 and September 1, 2020. We sought to identify studies that utilized any sort of machine learning algorithm for the detection, prognostication, or etiological investigation of stroke. The search key words included ("machine learning"[Mesh] OR "deep learning" [tiab] OR "machine learning"[tiab]) AND ("ischemic lesion"[tiab] OR "stroke"[tiab] OR "stroke"[Mesh]). Articles that resulted from this search underwent title and abstract assessment for relevance to stroke and use of explicitly described machine learning algorithm. After this initial screening, the remaining full texts were screened by two independent reviews (L.V. and D.N.) for details describing the dataset, particularly geographic distribution. Results were supplemented by searching reference lists of publications and review papers. Only studies that included at least one dataset collected in the US were included. When not explicitly named, cohorts were attributed to the home state of the institution. Datasets created from >3 states, from NIH data, or from industry databases were marked as “multicenter” and analyzed independently.

Results

582 studies were identified on initial searching of the PubMed database. Of these studies, 106 full texts were assessed after title and abstract screening which resulted in 489 papers excluded. Of these 106 studies, 79 were excluded due to using cohorts from outside the United States or being review articles or editorials. 27 studies were thus included in this analysis (Figure 1).

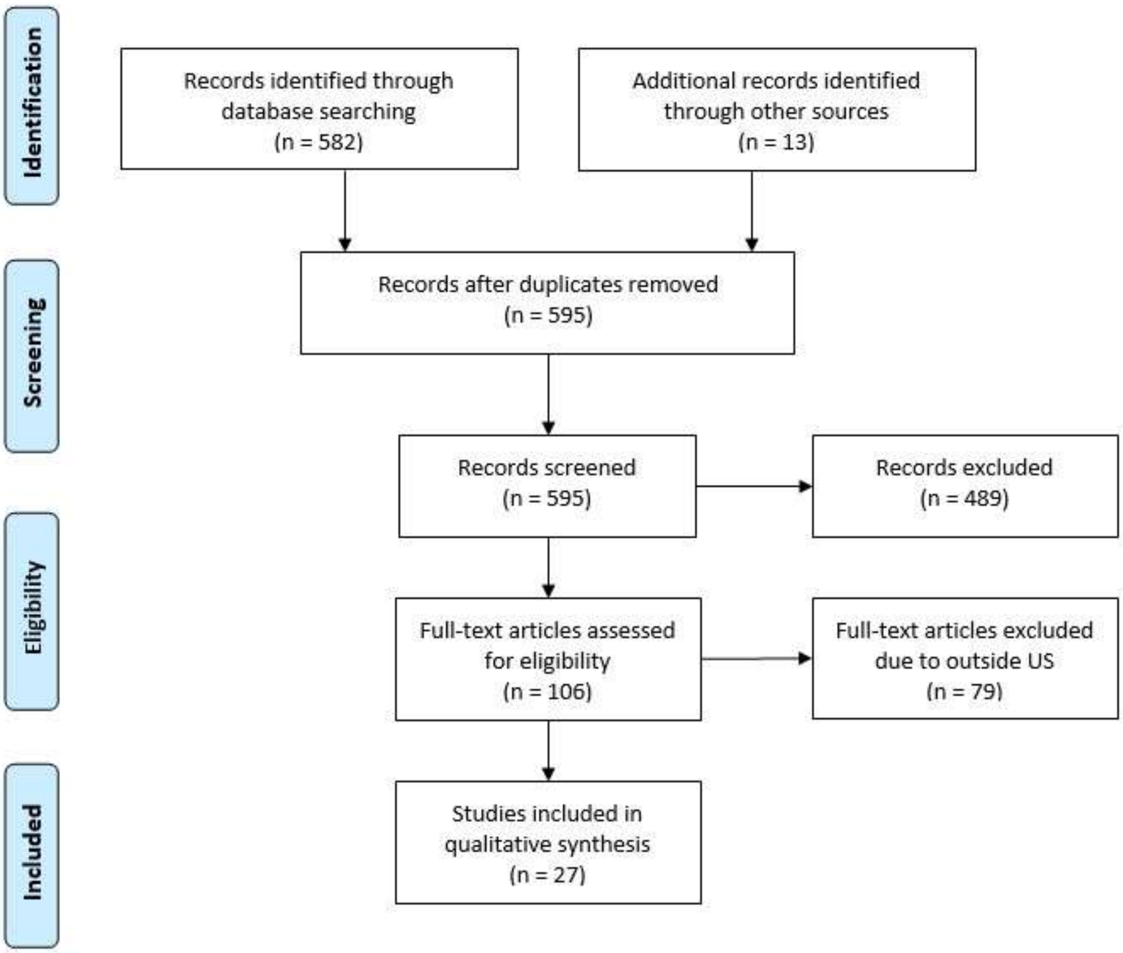

Figure 1: PRISMA flow diagram detailing literature search and exclusion

Of the 27 studies included, 7 (25.9%) used patient data from California, 6 (22.2%) were multicenter, 3 (11.1%) were in Massachusetts, 2 (7.4%) each in Illinois, Missouri, and New York, and 1 (3.7%) each from South Carolina, Washington, West Virginia, and Wisconsin. 1 (3.7%) study used data from Utah and Texas (Table 1).

Discussion

The use of machine learning within neurosurgery has recently expanded in an effort to optimize the generalizability and predictive value of research studies with small sample sizes. ML algorithms operate on a similar methodology to traditional statistical modeling studies: an initial dataset is required which is subjected to data cleaning, characterization, and preliminary descriptive analysis before being fed into, most commonly, a supervised learning algorithm that uses a portion of the data as a training set to optimize parameters that result in a predictive model. Optimizing these features results in a model that is then set against a testing set of data, which is used to evaluate model accuracy, sensitivity, and specificity. While many studies aim to evaluate optimal algorithms used in these use cases, the importance of the dataset is often understated. One of the reasons for this disconnect between research and clinical practice is due to a lack of characterization of reporting and quality outcomes of machine learning algorithms, which has been recently addressed by the SPIRIT-AI and CONSORT-AI consensus statements[33,34]. While multicenter studies have been used as seen in our study, the majority of datasets still originate from a single state or even a single institution, both of which do not match demographically to a wider population and thus are difficult to generalize outside of the specific population used. In addition, previous work has strongly characterized the adverse effect of biased datasets on predictive algorithms, such as Gijsberts et. al showing the decreased generalizability of the Framingham study on populations not matching the initial study population and a study by Neighbors et. al showing overdiagnosis of schizophrenia in African American patients based on biased datasets[35,,36]. In addition, both Obermeyer et. al and Char et. al. have expounded on the adverse influence of racially biased datasets on healthcare algorithm predictive capability, which is possible with geographically biased datasets[37,38,39].

Ischemic stroke in particular is a pathology that is susceptible to this bias in machine learning datasets. As seen in our study, datasets were disproportionately trained across the 50 states, with weighting heavily towards California, Massachusetts, and less so toward Illinois, Missouri, and New York. As Kaushal et. al. mention, these states may have socioeconomic or ethnocultural differences from the rest of the nation that affect their outcomes. While that study looked across a broad range of diseases in multiple specialties and thus could not identify specific factors that are unequally distributed, these factors are well elucidated in the case of stroke, such as age, sex, co-morbities such as diabetes, hypertension, hyperlipidemia, and atrial fibrillation, and pre-operative factors such as access to preventative medications (aspirin, statins) and distance to stroke centers (last seen normal time, door-to-needle time in the case of thrombectomy, etc.)[40]. All of these variables are known not to be distributed equally across the United States, solidifying the need for a broader distribution of studies sampling from many different populations.

It is unlikely that these factors are equally distributed across the nation, as seen in a CDC study in 2005 that showed substantial differences in stroke prevalence by state, race/ethnicity, age group, and education level. This study noted the highest prevalence of stroke in Mississippi (4.3%) followed by Oklahoma (3.4%), Washington D.C. (3.4%), Louisiana (3.3%), and Alabama (3.2%) while the prevalence in California was 2.6% (Figure 2A)[41, 42]. These findings have been echoed by numerous studies on the so-called “stroke belt”, a group of states including Alabama, Arkansas, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, and Tennessee that have been noted to have an unusually high stroke prevalence[43, 44]. Of these stroke belt states, only South Carolina is included in the machine learning algorithm of one study (Figure 2B)[28]. There is thus a concerning discrepancy between the study populations used for predictive modeling and the states

most afflicted by ischemic stroke. Challen et. al describes the so-called "distributional shift", with training data not matching ongoing patient data in continuously trained algorithms, which can have an analogous influence with the disparity in stroke distributions seen here[45]. As noted by Chen et. al and Braveman et. al, geographic distribution of patients used to train datasets can be a major contributor to biases in machine learning, as historically underserved populations and groups without robust healthcare access can have adverse events or worse outcomes that are mistakenly attributed to causes such as comorbidities or compliance[46, 47]. In addition, with stroke outcomes being dependent on factors such as distance from a healthcare center and time to reperfusion, the impact of geographic bias on stroke outcomes is clear and can be perpetuated by geographically biased algorithms.

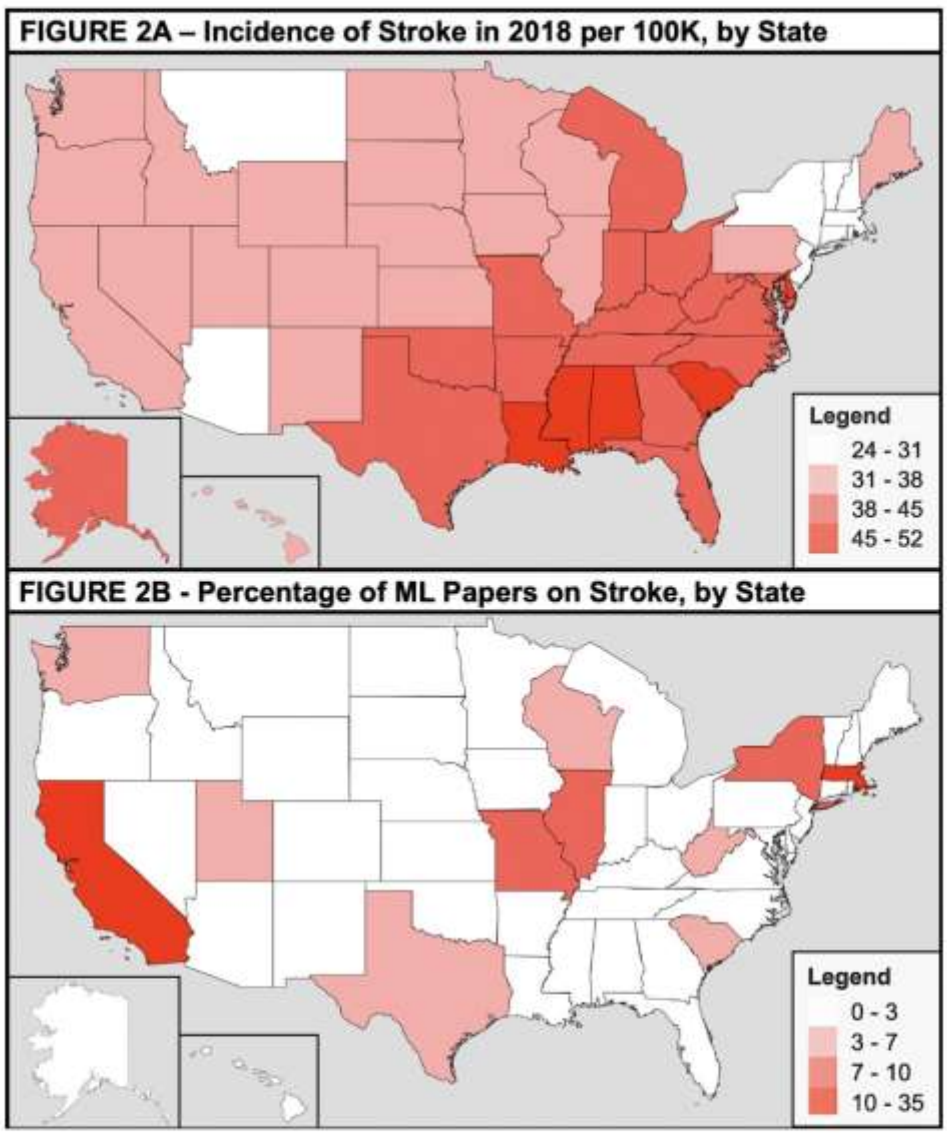

Figure 2: Comparison of Stroke Incidence with Dataset Distribution in Stroke Literature[6]

As the use of big data and machine learning algorithms increases in neurosurgery, careful consideration as to the quality and generalizability of these tools is imperative[48,49]. Increasing numbers of studies have clearly demonstrated the capabilities of these predictive algorithms, but the introduction of bias into these predictive algorithms has the potential to cause devastating effects in terms of inaccurate predictions or perpetuating systemic healthcare biases in a so-called "objective" methodology. Our study demonstrates one avenue through which bias can be introduced into machine learning studies, and the increasing interest in formal characterization and guidelines surrounding the use of machine learning, such as the CONSORT-AI and SPIRIT-AI consensus decisions, is a promising regulatory measure curtailing the implementation of biased algorithms. Further analyses into dataset quality, perhaps on an international scale, are warranted to develop stringent guideline which both aid the machine learning and research community by providing set benchmarks to measure against and also aid the clinician and patient communities by vouchsafing any clinical tools that do gain widespread acceptance.

Some limitations to this study include the varied nature of the stroke studies included as well as the analysis of machine learning outcomes using local and varied datasets. We included a broad variety of stroke machine learning studies, including for prognosis, diagnosis, and imaging analysis, due to the need for a corpus of work to analyze dataset origin. As more machine learning studies within a particular area are published, future studies to analyze dataset qualities that are limited to one area of research would be warranted. In addition, comparative studies between algorithms trained on data that is from the clinical population that the algorithm is applied to and algorithms trained on national or foreign data would be valuable to evaluate the quantitative difference that dataset origin can make.

Conclusion

Machine learning has been increasingly applied across medical specialties and its potential in the prevention, diagnosis, treatment, prognostication of ischemic stroke is tremendous. Recent literature has shown an important geographic discrepancy between the states that cohorts are drawn from and the states most afflicted by ischemic stroke. There is a strong preponderance of studies based in California, Massachusetts, Illinois, Missouri, and New York, while the vast majority of states with the highest prevalence of stroke are not included in any study. There is thus a pressing need for improving our machine learning algorithms by minimizing selection bias and optimizing dataset quality.

Acknowledgements

The authors thank Paul Hunter, DMD, MLIS for his assistance with designing terms for database search.

References

1. Velagapudi L, D'Souza T, Matias CM, Sharan AD. Letter to the Editor: "Bridging Machine Learning and Clinical Practice in Neurosurgery: Hurdles and Solutions." *World Neurosurg*. 2020;134:678-679. doi:10.1016/j.wneu.2019.11.105

2. Harbaugh RE. Artificial neural networks for neurosurgical diagnosis, prognosis, and management. *Neurosurg Focus*. 2018;45(5):E3. doi:10.3171/2018.8.FOCUS18438

3. Brusko GD, Kolcun JPG, Wang MY. Machine-Learning Models: The Future of Predictive Analytics in Neurosurgery. *Clin Neurosurg*. 2018;83(1):E3-E4. doi:10.1093/neuros/nyy166

4. Senders JT, Zaki MM, Karhade A V., et al. An introduction and overview of machine learning in neurosurgical care. *Acta Neurochir (Wien)*. 2018;160(1):29-38. doi:10.1007/s00701-017-3385-8

5. Kaushal A, Altman R, Langlotz C. Geographic Distribution of US Cohorts Used to Train Deep Learning Algorithms. *JAMA J Am Med Assoc*. 2020;324(12):1212. doi:10.1001/jama.2020.12067

6. Ho KC hun., Speier W, El-Saden S, et al. Predicting discharge mortality after acute ischemic stroke using balanced data. *AMIA Annu Symp Proc*. 2014;2014:1787-1796. /pmc/articles/PMC4419881/?report=abstract. Accessed October 5, 2020.

7. Ho KC, Speier W, Zhang H, Scalzo F, El-Saden S, Arnold CW. A Machine Learning Approach for Classifying Ischemic Stroke Onset Time from Imaging. *IEEE Trans Med Imaging*. 2019;38(7):1666-1676. doi:10.1109/TMI.2019.2901445

8. Wang K, Shou Q, Ma SJ, et al. Deep Learning Detection of Penumbral Tissue on Arterial Spin Labeling in Stroke. *Stroke*. 2020;51(2):489-497. doi:10.1161/STROKEAHA.119.027457

9. Ho KC. Predicting ischemic stroke tissue fate using a deep convolutional neural network on source magnetic resonance perfusion images. *J Med Imaging*. 2019;6(02):1. doi:10.1117/1.jmi.6.2.026001

10. Ho KC, Speier W, El-Saden S, Arnold CW. Classifying Acute Ischemic Stroke Onset Time using Deep Imaging Features. *AMIA . Annu Symp proceedings AMIA Symp*. 2017;2017:892-901. /pmc/articles/PMC5977679/?report=abstract. Accessed October 5, 2020.

11. Xie Y, Jiang B, Gong E, et al. Use of gradient boosting machine learning to predict patient outcome in acute ischemic stroke on the basis of imaging, demographic, and clinical information. *Am J Roentgenol*. 2019;212(1):44-51. doi:10.2214/AJR.18.20260

12. Yu Y, Guo D, Lou M, Liebeskind D, Scalzo F. Prediction of hemorrhagic transformation severity in acute stroke from source perfusion MRI. *IEEE Trans Biomed Eng*. 2018;65(9):2058-2065. doi:10.1109/TBME.2017.2783241

13. Liu T, Fan W, Wu C. A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset. *Artif Intell Med*. 2019;101:101723. doi:10.1016/j.artmed.2019.101723

14. Kasasbeh AS, Christensen S, Parsons MW, Campbell B, Albers GW, Lansberg MG. Artificial Neural Network Computer Tomography Perfusion Prediction of Ischemic Core. *Stroke*. 2019;50(6):1578-1581. doi:10.1161/STROKEAHA.118.022649

15. Yu Y, Xie Y, Thamm T, et al. Use of Deep Learning to Predict Final Ischemic Stroke Lesions From Initial Magnetic Resonance Imaging. *JAMA Netw open*. 2020;3(3):e200772. doi:10.1001/jamanetworkopen.2020.0772

16. Ambale-Venkatesh B, Yang X, Wu CO, et al. Cardiovascular Event Prediction by Machine Learning: The Multi-Ethnic Study of Atherosclerosis. *Circ Res*. 2017;121(9):1092-1101. doi:10.1161/CIRCRESAHA.117.311312

17. Kogan E, Twyman K, Heap J, Milentijevic D, Lin JH, Alberts M. Assessing stroke severity using electronic health record data: A machine learning approach. *BMC Med Inform Decis Mak*. 2020;20(1). doi:10.1186/s12911-019-1010-x

18. Wu O, Winzeck S, Giese AK, et al. Big Data Approaches to Phenotyping Acute Ischemic Stroke Using Automated Lesion Segmentation of Multi-Center Magnetic Resonance Imaging Data. *Stroke*. 2019;50(7):1734-1741. doi:10.1161/STROKEAHA.119.025373

19. Ong CJ, Orfanoudaki A, Zhang R, et al. Machine learning and natural language processing methods to identify ischemic stroke, acuity and location from radiology reports. *PLoS One*. 2020;15(6). doi:10.1371/journal.pone.0234908

20. Orfanoudaki A, Chesley E, Cadisch C, et al. Machine learning provides evidence that stroke risk is not linear: The non-linear Framingham stroke risk score. *PLoS One*. 2020;15(5). doi:10.1371/journal.pone.0232414

21. Forkert ND, Verleger T, Cheng B, Thomalla G, Hilgetag CC, Fiehler J. Multiclass support vector machine-based lesion mapping predicts functional outcome in ischemic stroke patients. *PLoS One*. 2015;10(6). doi:10.1371/journal.pone.0129569

22. Beecy AN, Chang Q, Anchouche K, et al. A Novel Deep Learning Approach for Automated Diagnosis of Acute Ischemic Infarction on Computed Tomography. *JACC Cardiovasc Imaging*. 2018;11(11):1723-1725. doi:10.1016/j.jcmg.2018.03.012

23. 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

24. Chen Y, Dhar R, Heitsch L, et al. Automated quantification of cerebral edema following hemispheric infarction: Application of a machine-learning algorithm to evaluate CSF shifts on serial head CTs. *NeuroImage Clin*. 2016;12:673-680. doi:10.1016/j.nicl.2016.09.018

25. Dhar R, Chen Y, An H, Lee JM. Application of machine learning to automated analysis of cerebral edema in large cohorts of ischemic stroke patients. *Front Neurol*. 2018;9(AUG). doi:10.3389/fneur.2018.00687

26. Harari Y, O'Brien MK, Lieber RL, Jayaraman A. Inpatient stroke rehabilitation: Prediction of clinical outcomes using a machine-learning approach. *J Neuroeng Rehabil*. 2020;17(1).

doi:10.1186/s12984-020-00704-3

27. Garg R, Oh E, Naidech A, Kording K, Prabhakaran S. Automating Ischemic Stroke Subtype Classification Using Machine Learning and Natural Language Processing. *J Stroke Cerebrovasc Dis*. 2019;28(7):2045-2051. doi:10.1016/j.jstrokecerebrovasdis.2019.02.004

28. Alawieh A, Zaraket F, Alawieh MB, Chatterjee AR, Spiotta A. Using machine learning to optimize selection of elderly patients for endovascular thrombectomy. *J Neurointerv Surg*. 2019;11(8):847-851. doi:10.1136/neurintsurg-2018-014381

29. Bochniewicz EM, Emmer G, McLeod A, Barth J, Dromerick AW, Lum P. Measuring Functional Arm Movement after Stroke Using a Single Wrist-Worn Sensor and Machine Learning. *J Stroke Cerebrovasc Dis*. 2017;26(12):2880-2887. doi:10.1016/j.jstrokecerebrovasdis.2017.07.004

30. O'Connell GC, Chantler PD, Barr TL. Stroke-associated pattern of gene expression previously identified by machine-learning is diagnostically robust in an independent patient population. *Genomics Data*. 2017;14:47-52. doi:10.1016/j.gdata.2017.08.006

31. Liu F, Jang H, Kijowski R, Bradshaw T, McMillan AB. Deep learning MR imaging-based attenuation correction for PET/MR imaging. *Radiology*. 2018;286(2):676-684. doi:10.1148/radiol.2017170700

32. Sheth SA, Lopez-Rivera V, Barman A, et al. Machine Learning-Enabled Automated Determination of Acute Ischemic Core From Computed Tomography Angiography. *Stroke*. 2019;50(11):3093-3100. doi:10.1161/STROKEAHA.119.026189

33. Cruz Rivera S, Liu X, Chan AW, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med*. 2020;26(9):1351-1363. doi:10.1038/s41591-020-1037-7

34. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med*. 2020;26(9):1364-1374. doi:10.1038/s41591-020-1034-x

35. Gijsberts CM, Groenewegen KA, Hoefer IE, et al. Race/ethnic differences in the associations of the Framingham risk factors with carotid IMT and cardiovascular events. *PLoS One*. 2015;10(7). doi:10.1371/journal.pone.0132321

36. Neighbors HW, Jackson JS, Campbell L, Williams D. The influence of racial factors on psychiatric diagnosis: A review and suggestions for research. *Community Ment Health J*. 1989;25(4):301-311. doi:10.1007/BF00755677

37. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science (80-)*. 2019;366(6464):447-453. doi:10.1126/science.aax2342

38. Char DS, Shah NH, Magnus D. Implementing Machine Learning in Health Care — Addressing Ethical Challenges. *N Engl J Med*. 2018;378(11):981-983. doi:10.1056/nejmp1714229

39. Char DS, Abràmoff MD, Feudtner C. Identifying Ethical Considerations for Machine Learning Healthcare Applications. *Am J Bioeth*. 2020;20(11):7-17. doi:10.1080/15265161.2020.1819469

40. Howard VJ, Cushman M, Pulley L, et al. The Reasons for Geographic and Racial Differences in Stroke Study: Objectives and Design. *Neuroepidemiology*. 2005;25(3):135-143. doi:10.1159/000086678

41. CDC. Prevalence of Stroke -- United States, 2005. *MMWR Wkly*. 2007;56(19):469-474. https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5619a2.htm#tab2. Accessed October 4, 2020.

42. Benjamin EJ, Virani SS, Callaway CW, et al. Heart disease and stroke statistics - 2018 update: A report from the American Heart Association. *Circulation*. 2018;137(12):E67-E492. doi:10.1161/CIR.0000000000000558

43. Howard G, Howard VJ. Twenty Years of Progress Toward Understanding the Stroke Belt. *Stroke*. 2020;51(3):742-750. doi:10.1161/STROKEAHA.119.024155

44. Karp DN, Wolff CS, Wiebe DJ, Branas CC, Carr BG, Mullen MT. Reassessing the stroke belt: Using small area spatial statistics to identify clusters of high stroke mortality in the United States. In: *Stroke*. Vol 47. Lippincott Williams and Wilkins; 2016:1939-1942. doi:10.1161/STROKEAHA.116.012997

45. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. *BMJ Qual Saf*. 2019;28(3):231-237. doi:10.1136/bmjqs-2018-008370

46. Chen IY, Pierson E, Rose S, Joshi S, Ferryman K, Ghassemi M. Ethical Machine Learning in Health Care. *Annu Rev Biomed Data Sci*. September 2020. http://arxiv.org/abs/2009.10576. Accessed December 30, 2020.

47. Braveman P. Health disparities and health equity: Concepts and measurement. *Annu Rev Public Health*. 2006;27:167-194. doi:10.1146/annurev.publhealth.27.021405.102103

48. Bydon M, Schirmer CM, Oermann EK, et al. Big Data Defined: A Practical Review for Neurosurgeons. *World Neurosurg*. 2020;133:e842-e849. doi:10.1016/j.wneu.2019.09.092

49. Oravec CS, Motiwala M, Reed K, et al. Big Data Research in Neurosurgery: A Critical Look at this Popular New Study Design. *Neurosurgery*. 2018;82(5):728-746. doi:10.1093/neuros/nyx328

Table 1: Studies Included in this Analysis

State	Study (n=sample size)	Algorithms Used	Study Outcome
California	Ho et. al[6] (n=190)	Naïve Bayes, Support Vector Machine, Decision Tree, Random Forests, and Logistic Regression	Compared the performance of SVM, PCA-SVM, DT, RF, NB, and LR models for predicting stroke patient mortality at discharge and determined SVM was the best based on relative c-statistic and F1-score
	Ho et. al[7] (n=131)	Logistic Regression, Random Forest, Gradient Boosted Regression Tree, Support Vector Machine, and Stepwise Multilinear Regression	Developed new imaging features from MR images, perfusion parameter maps, and deep AE feature maps, and showed that they can be utilized by machine learning models to classify TSS
	Wang et. al[8] (n=137)	Linear Regression Classifier, Ridge Regression Classifier, Kernel Ridge Regression classifier, Neural Network Classifier, Support Vector Machine with Radial Basis Function, and Random Forest Classifier	pCASL perfusion magnetic resonance imaging in conjunction with the DL algorithm provides a promising approach for assisting decision-making for endovascular treatment in patients with acute ischemic stroke
	Ho et. al[9] (n=44)	Deep Convolution Neural Networks (CNNs)	Deep convolution neural networks (CNNs) on predicting final stroke infarct volume using only the source perfusion images
	Ho et. al[10] (n=105)	Stepwise Multilinear Regression (SMR), Support Victor Machine (SVM), Random forest (RF), and Gradient Boosted Regression Tree (GBRT)	SMR, SVM, RF, and GBRT models were able to classify TSS, with SMR achieving the highest AUC
	Xie et. al[11] (n=512)	Extreme Gradient Boosting (XGB) and Gradient Boosting Machine (GBM)	Decision tree-based GBMs can predict the recovery outcome of stroke patients at admission with a high AUC
	Yu et. al[12] (n=165)	Support Vector Machines, Linear Regression, Decision Trees, Neural Networks, and Kernel Spectral Regression	A model can learn to extract imaging markers of HT directly from source PWI images rather than from pre-established metrics

Multicenter	Liu et. al[13] (n=43,400)	Random Forest Regression, Hyperparameter Optimization(AutoHPO) based on Deep Neural Network(DNN)	Reduced the false negative rate with a relatively high overall accuracy, which means a successful decrease in the misdiagnosis rate for stroke prediction
	Kasasbeh et. al[14] (n=128)	Artificial Neural Network	An ANN that integrates clinical and CTP data predicts the ischemic core with accuracy
	Yu et. al[15] (n=182)	Attention-gated U-Net	Deep learning model appears to have successfully predicted infarct lesions from baseline imaging without reperfusion information and achieved comparable performance to existing clinical methods
	Ambale-Venkatesh et al[16] (n=6,814)	Random Survival Forest	Machine learning in conjunction with deep phenotyping improves prediction accuracy in cardiovascular event prediction in an initially asymptomatic population
	Kogan et. al[17] (n=7,149)	Random Forest Model, Gradient Boosting Model, Neural Network, and Linear Regression	Machine learning models built on EHR data can be used to determine proxies for stroke severity
	Wu et. al[18] (n=3,301)	Automated Deep Learning Segmentation	Automated accurate clinical diffusion-weighted MRI lesion segmentation using deep learning algorithms trained with multi-center and diverse data is feasible
Massachusetts	Ong et. al[19] (n=17,864)	Logistic Regression, k-Nearest Neighbors (k-NN), Classification and Regression Trees, (CART) Optimal Classification Trees (OCT) with and without hyperplanes (OCT-H), Random Forest, and Recurrent Neural Networks (RNN)	Identifying salient stroke features from radiology text that can triage high-risk imaging findings and identify patient populations of interest for research
	Orfanoudaki et. al[20] (n=4,385)	N-SRS, R-FSRS, Logistic Regression, CART, Random Forest, XGBoost	Developed N-SRS, an accurate stroke risk calculator that outperforms, in accuracy and user-friendliness, the existing stroke risk prediction tool

	Forkert et. al[21] (n=68)	Support Vector Machine (SVM)	Graded SVM-based functional stroke outcome prediction using the problem-specific brain regions for lesion overlap quantification leads to promising results but needs to be further validated using an independent database to rule out a potential methodical bias and overfitting effects.
New York	Beecy et. al[22] (n=114)	Deep Learning	Machine-learned models using novel DL techniques enable highly accurate automated diagnosis of acute brain infarction.
	Kamel et. al[23] (n=1,083)	L1 regularization, Gradient-Boosted Decision Tree Ensemble (XGBoost), Random Forests, and Multivariate Adaptive Splines was used	Machine learning estimator that distinguished known cardioembolic versus noncardioembolic strokes indirectly estimated that 44% of ESUS cases were cardioembolic.
Missouri	Chen et. Al[24] (n=38)	Random Forest (RF), (HU) thresholding and RF Segmentation	Validated an automated CSF quantification approach which is accurate and reliable, and can be applied to scans from multiple centers
	Dhar et. al[25] (n=155)	Generalized estimating equation (GEE)	Proof-of-principle that we can automate brain imaging data analysis and obtain meaningful volumetric data on large cohorts of stroke patients.
Illinois	Harari et. al[26] (n=50)	Lasso regression	Models presented in this study could help clinicians and researchers to predict the discharge scores of clinical outcomes for individuals enrolled in an inpatient stroke rehabilitation program
	Garg et. al[27] (n=50)	K-nearest neighbors (KNN), Support Vector Machines (SVM), Random Forests (RF), Extra Randomized Trees Classifiers, Gradient Boosting Machines, and Extreme Gradient Boosting (XGBoost)	Automated machine learning approaches using textual data from the EHR shows agreement with manual TOAST classification

South Carolina	Alawieh et. al[28] (n=110)	Optimal Prognostic Mode	SPOT is a useful tool to determine which patients to exclude from ET, and has been implemented in an online calculator for public use
Washington	Bochniewicz et. al[29] (n=20)	Random Forest Model	Inexpensive and objective quantification of functional UE (upper extremity) use in hemiparesis, and for assessing the impact of UE treatments
West Virginia	O'Connell et. al[30] (n=46)	k-nearest neighbors (kNN)	Confirm the diagnostic robustness of the previously identified pattern of differential expression in an independent patient population, and further suggest that it is temporally stable over the first 24 h of stroke pathology
Wisconsin	Liu et. al[31] (n=10)	Imaging–Based attenuation correction using deep convolutional auto-encoder	Developed an automated approach that allows generation of discrete-valued pseudo CT scans (soft tissue, bone, and air) from a single high-spatial-resolution diagnostic-quality three-dimensional MR image and evaluated it in brain PET/MR imaging
Utah/Texas	Sheth et. al[32] (n=297)	Convolutional Neural Network	Information needed to perform the neuroimaging evaluation for endovascular therapy with comparable accuracy to advanced imaging modalities may be present in CTA, and the ability of machine learning to automate the analysis