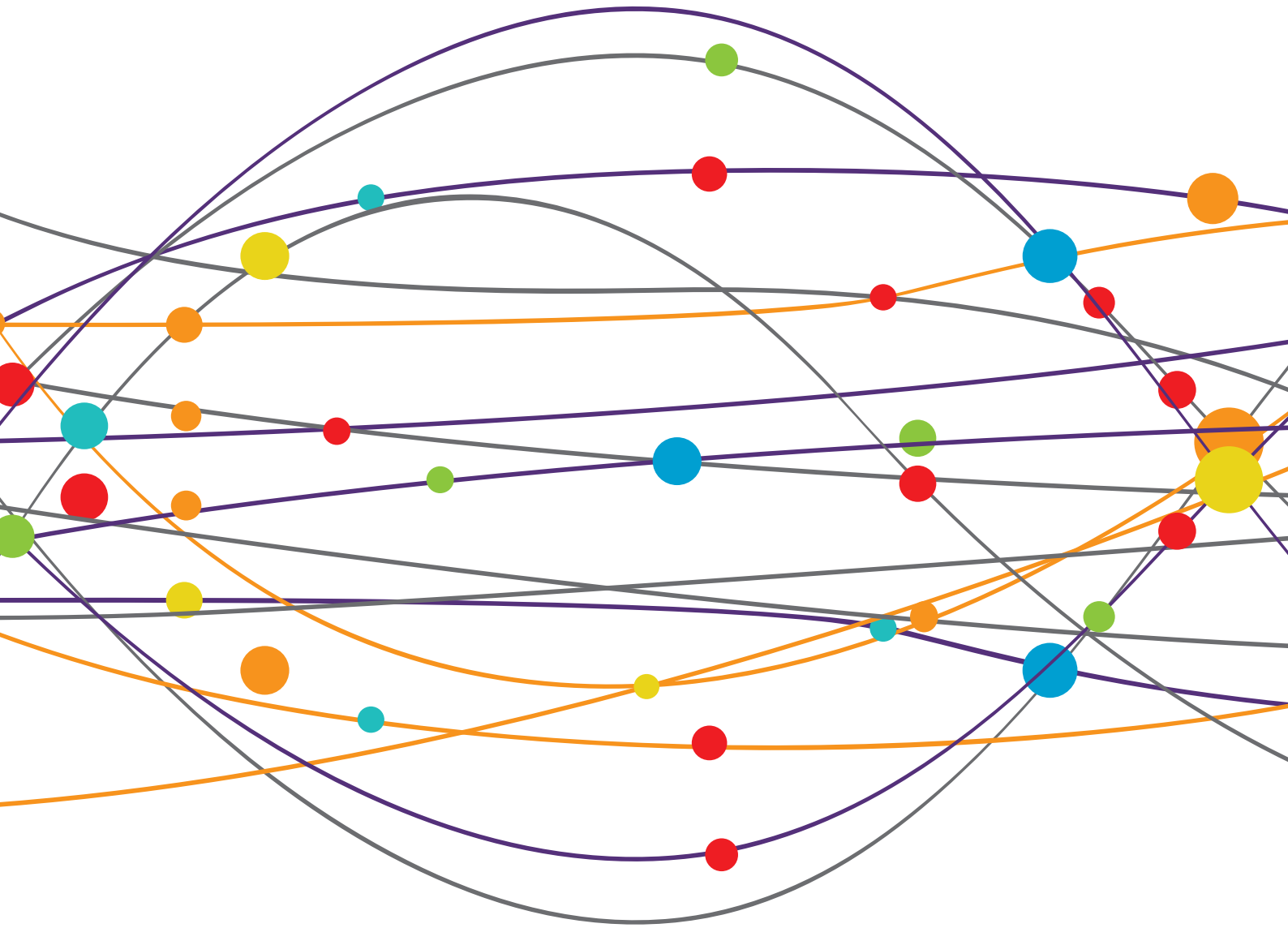
MACHINE LEARNING IN ACTION: STROKE DIAGNOSIS AND OUTCOME PREDICTION

EDITED BY: Ramin Zand, Vida Abedi, Jiang Li and Thanh G. Phan
PUBLISHED IN: Frontiers in Neurology



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-793-9
DOI 10.3389/978-2-88976-793-9

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

MACHINE LEARNING IN ACTION: STROKE DIAGNOSIS AND OUTCOME PREDICTION

Topic Editors:
Ramin Zand, Pennsylvania State University, United States
Vida Abedi, The Pennsylvania State University, United States
Jiang Li, Geisinger Medical Center, United States
Thanh G. Phans, Monash Health, Australia

Citation: Zand, R., Abedi, V., Li, J., Phan, T. G., eds. (2022). Machine Learning in Action: Stroke Diagnosis and Outcome Prediction. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-793-9

Table of Contents

frontiers | Frontiers in Neurology

TYPE Editorial
PUBLISHED 20 July 2022
DOI 10.3389/fneur.2022.984467

OPEN ACCESS

EDITED AND REVIEWED BY
Jean-Claude Baron,
University of Cambridge,
United Kingdom

*CORRESPONDENCE
Vida Abedi
vidaabedi@gmail.com

SPECIALTY SECTION
This article was submitted to Stroke, a section of the journal Frontiers in Neurology

RECEIVED 02 July 2022
ACCEPTED 05 July 2022
PUBLISHED 20 July 2022

CITATION
Abedi V, Kawamura Y, Li J, Phan TG and Zand R (2022) Editorial: Machine Learning in Action: Stroke Diagnosis and Outcome Prediction. *Front. Neurol.* 13:984467. doi: 10.3389/fneur.2022.984467

Editorial: Machine Learning in Action: Stroke Diagnosis and Outcome Prediction

Vida Abedi [1]*, Yuki Kawamura [2], Jiang Li [3], Thanh G. Phan[4,5] and Ramin Zand [6]

[1]Department of Public Health Sciences, College of Medicine, The Pennsylvania State University, Hershey, PA, United States, [2]Department of Medicine, University of Cambridge, Cambridge, United Kingdom, [3]Department of Molecular and Functional Genomics, Weis Center for Research, Geisinger Health System, Danville, PA, United States, [4]Stroke and Aging Research Group, Clinical Trials, Imaging and Informatics Division, School of Clinical Sciences at Monash Health, Melbourne, VIC, Australia, [5]Department of Neurology, Monash Health, Melbourne, VIC, Australia, [6]Department of Neurology, College of Medicine, The Pennsylvania State University, Hershey, PA, United States

KEYWORDS

machine learning, artificial intelligence, stroke, cerebrovascular events, electronic health records (EHR), electronic medical records (EMR), deep learning, natural language processing

Editorial on the Research Topic
Machine Learning in Action: Stroke Diagnosis and Outcome Prediction

Machine learning—the ability of computers to "learn" to perform a task rather than being explicitly programmed for the purpose—has seen significant developments in recent years. Biomedical research is no exception to its far-reaching impact and has seen more than a ten-fold increase in the number of publications related to machine learning in the last decade (1). In this Research Topic, we present recent advances in developing machine learning algorithms in the context of cerebrovascular diseases to highlight promising approaches that represent various areas of potential clinical utility in stroke care. The focus is on applications with high clinical value and a solid technical foundation.

Deployment of machine learning algorithms in the clinic principally involves four stages of the care workflow: primary prevention, acute-phase treatment, post-diagnosis prediction, and secondary prevention (2). Primary prevention includes personalized or stratified patient risk prediction and identification of gaps in care, whereas integration into acute phase treatment aims to aid physician diagnosis and referrals. Machine learning algorithms for post-diagnosis and secondary prediction can provide predicted outcomes that allow the identification of patients who would be responsive to treatment or require careful monitoring due to a higher risk of recurrent disease. Together, machine learning algorithms can aid clinical decision-making in each step by providing recommendations and pointing to possible missed cases for critical conditions. As suggested by Mainali et al., machine learning algorithms can have particular utility in alleviating two of the clinical challenges of stroke: the time-sensitive nature of the acute-phase treatment and the difficulty of predicting outcomes, especially in

the acute phase. Given these potential benefits, calibrating the algorithms to prevent excessive alerts and supporting physician autonomy through careful assessment of human-computer interaction is key to maximizing adoption (3).

Electronic health records (EHR) are one of the principal sources of standardized clinical information on a patient and can serve as a valuable starting point for algorithm development. The results of Rana et al. are encouraging, demonstrating that models trained on EHR data outperformed models trained on a limited number of features clinically associated with stroke, confirming the benefits of additional information obtained by data extraction from EHR. Using EHR data, Darabi et al. compared the performance of multiple machine learning models in predicting 30-day hospital readmission. Their models improved upon previous predictive models based on logistic regression and provided promising results that could direct targeted intervention for high-risk patients. Notably, features that their best predictive model indicated as being key predictors of 30-day readmission agree with results from independent studies (4, 5) and clinical intuition, underscoring the interpretability of their model.

Complementation of EHR data with additional modalities of clinical investigations holds promise in further improving prediction accuracy. Herein, Lineback et al. employ Natural Language Processing (NLP) to glean freeform textual data. In contrast, Rajashekar et al. combines MRI and CT imaging data to improve prediction models trained solely on EHR data. Multimodal approaches can require more sophisticated models to extract information from various data types but more closely approximate decision-making by physicians and better integrate multifaceted information collected *via* clinical investigations and examinations.

Imaging is a rich source of information. Imaging has critical clinical relevance in neurology and a high affinity for sophisticated deep learning models, such as convolutional neural networks. Indeed, many of the recent advances in machine learning in healthcare have centered on image analysis, including the use of retinal images for cardiometabolic disease prediction (6–9) and analysis of histopathological slides (10–15). Models focus on cerebrovascular disease, however, have been comparatively scant. McLouth et al. validate the performance of a commercially available deep learning software in assessing intracranial hemorrhage and large vessel occlusion using CT images. Implementing analysis software within the imaging workflow can provide venues where machine learning algorithms can seamlessly integrate into clinical decision-making. Furthermore, incorporating features from MRI scans, such as in the study by Xiao et al. predicting hypoperfusion in ischemic stroke patients, could define a clinically relevant threshold that directs decision-making in a facile manner. Integration of images in machine learning algorithms provides several benefits, including higher accuracy of diagnosis and improved objectivity compared to physical examinations. Given that imaging is routinely performed for stroke patients and is uniquely capable of providing functionally relevant anatomical information, image analysis models are promising candidates for clinical deployment in stroke care.

Machine learning can be an invaluable asset, especially in cases where diagnosis requires extensive examination or training or when the diagnosis is based on subtle features and are thus inherently prone to misdiagnosis. The algorithms described by Kim et al. to identify acute central dizziness and by Lin et al. to identify mild stroke patients at risk of disability exemplify the possibilities—supporting physicians in making challenging clinical decisions. Both models closely approximate or outperform existing risk scores without requiring extensive neurological examinations, allowing more patients to be screened and thus reducing the chances of a deteriorating patient escaping notice.

While machine learning holds promises, several challenges persist in implementing these technologies in healthcare. First, technical limitations can stem from the type and quality of the datasets available. EHR data can often be poorly standardized and sparse, posing problems in model generalizability. Investigators such as Rana et al. and Darabi et al. have only used administrative data from EHR with additional clinical variables such as NIHSS. By contrast, mining the free text in the patient chart (such as provider note, triage notes, discharge note, etc.) pose significant challenges. The free text is written by multiple clinicians, often with successive clinicians copying and pasting the written comments by the previous clinicians (16) in addition to auto-generated text that populate the patient chart. In addition, the tabular nature of clinical data extracted from EHR can often pose a difficulty even for advanced deep learning modalities, which often fail to surpass performances on simpler tree-based architectures (17). However, performance can be improved by extensive regularization (18). Sophisticated machine learning algorithms have had better success when applied to image datasets; however, even these complex deep learning algorithms can suffer from confounding factors, partially due to variation amongst institutions. Indeed, a recent study demonstrated that deep neural nets trained to predict SARS-CoV2 infection from X-ray images tend to select confounding "shortcuts" over signals in generating predictions (19). Attributes of datasets can limit the accuracy and generalizability of models, especially for external cohorts with different demographics and dataset characteristics. The development of standardized data protocols can aid the implementation of machine learning models that are more accurate and generalizable across multiple institutions. In addition to curating better datasets, models can also be adjusted for better generalizability; fine-tuning of pre-trained algorithms *via* transfer learning using site-specific data achieved superior results for external cohorts (20), and continuous domain adaptation has been explored to tackle temporal drifts in data (21, 22). It is essential to take all possible precautions

to ensure that the machine learning algorithms provide reliable, relevant, and interpretable results free from systemic biases. To achieve that, care must be taken to minimize confounding variations in the datasets that might affect generalizability and ensure fine-tuning approaches are integrated to allow the models to more closely approximate results for the underlying patient distribution.

Secondly, more complicated machine learning models can often be challenging to interpret, hindering the translation from prognosis to patient management. High-performing "black box" models lacking interpretability are of limited use in the clinic as they do little to inform physicians of actionable points. In particular, identifying modifiable risk factors is essential in the primary and secondary prevention of cerebrovascular events. To this end, Cui et al. used feature importance metrics to rank specific features mainly associated with predictive capability in each machine learning model. Analyses of feature importance could prove helpful in guiding intervention, especially if a factor is consistently listed as important across multiple models. For image analysis models, localization maps generated by methods such as Grad-CAM (23) could provide a limited level of interpretability. Separating interpretation from the prediction modeling to provide more flexibility is a strategy that has been getting more traction in recent years. Still, the usefulness of the algorithms can be diminished by confounding "shortcuts," as mentioned earlier. Since model depth is generally associated with better predictive capability, efforts must be made to create models that predict and inform. Desirable models should also consider workflow disruption or the possibility of causing "alert fatigue" before planning for implementation. Designing and training models so that interpretable features can be gleaned from model parameters and incorporating feedback from healthcare providers can improve the interpretability of models. In this respect, theoretical advances in model architecture and interpretation, combined with enhancing training data robustness, could prove fruitful.

Finally, ethical considerations must not be ignored. Model predictions can often be influenced by the socioeconomic, racial, and gender composition of the training datasets, the awareness of which is necessary to mitigate potential biases in models. For example, machine learning models were found to consistently underdiagnose patients in disadvantaged populations across three large chest X-ray datasets, especially where a patient was a member of more than one underserved group (24). The precedent of undertreatment in disadvantaged populations can further exacerbate biases by making it less likely for the algorithm to recommend treatment for members of the underprivileged sub-group of the population if similar patients were not provided treatment in the past. The performance of machine learning models must thus be thoroughly evaluated in different cohorts to assess the presence of systematic bias, which must be rectified before deployment. Further, while it is often possible to impute information that a patient declined to provide (e.g., smoking, HIV status, etc.), doing so can have ethical implications (25). Implementing machine learning algorithms in the clinic should proceed with special care to avoid unwittingly perpetuating health care inequalities in the training cohort. Finally, it is essential to reflect that algorithms are and will continue to be part of our medical system, including our medical education system. Thus, as a two-way street, we have to consider how such recommendations influence physicians' decisions and how this decision-making process potentially shifts with continued interaction.

In conclusion, recent developments in machine learning present ample opportunities for automated models that guide clinical decision-making and improve patient outcomes. The studies included herein represent selections of advances employing machine learning in various contexts in stroke care in our collective efforts to promote improved patient health through effective prevention, diagnosis, and intervention.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor J-CB declared a shared affiliation with the author YK at the time of review.

Publisher's note

References

1. Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. *NPJ Digital Med.* (2020) 3:126. doi: 10.1038/s41746-020-00333-z

2. Abedi V, Razavi S-M, Khan A, Avula V, Tompe A, Poursoroush A, et al. Artificial intelligence: a shifting paradigm in cardio-cerebrovascular medicine. *J Clin Med.* (2021) 10:5710. doi: 10.3390/jcm10235710

3. Abedi V, Khan A, Chaudhary D, Misra D, Avula V, Mathrawala D, et al. Using artificial intelligence for improving stroke diagnosis in emergency departments: a practical framework. *Ther Adv Neurol Disor.* (2020) 13:1756286420938962. doi: 10.1177/1756286420938962

4. Qiu X, Xue X, Xu R, Wang J, Zhang L, Zhang L, et al. Predictors, causes and outcome of 30-day readmission among acute ischemic stroke. *Neurol Res.* (2021) 43:9–14. doi: 10.1080/01616412.2020.1815954

5. Lichtman JH, Leifheit-Limson EC, Jones SB, Watanabe E, Bernheim SM, Phipps MS, et al. Predictors of hospital readmission after stroke. *Stroke.* (2010) 41:2525–33. doi: 10.1161/STROKEAHA.110.599159

6. Cheung CY, Xu D, Cheng C-Y, Sabanayagam C, Tham Y-C, Yu M, et al. A deep-learning system for the assessment of cardiovascular disease risk via the measurement of retinal-vessel calibre. *Nat Biomed Eng.* (2021) 5:498–508. doi: 10.1038/s41551-020-00626-4

7. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA.* (2016) 316:2402–10. doi: 10.1001/jama.2016.17216

8. Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng.* (2018) 2:158–64. doi: 10.1038/s41551-018-0195-0

9. Wolf RM, Channa R, Abramoff MD, Lehmann HP. Cost-effectiveness of autonomous point-of-care diabetic retinopathy screening for pediatric patients with diabetes. *JAMA Ophthalmol.* (2020) 138:1063–9. doi: 10.1001/jamaophthalmol.2020.3190

10. Diao JA, Wang JK, Chui WF, Mountain V, Gullapally SC, Srinivasan R, et al. Human-interpretable image features derived from densely mapped cancer pathology slides predict diverse molecular phenotypes. *Nat Commun.* (2021) 12:1613. doi: 10.1038/s41467-021-21896-9

11. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med.* (2019) 25:1301–9. doi: 10.1038/s41591-019-0508-1

12. Fu Y, Jung AW, Torne RV, Gonzalez S, Vöhringer H, Shmatko A, et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat Cancer.* (2020) 1:800–10. doi: 10.1038/s43018-020-0085-8

13. Jackson HW, Fischer JR, Zanotelli VRT, Ali HR, Mechera R, Soysal SD, et al. The single-cell pathology landscape of breast cancer. *Nature.* (2020) 578:615–20. doi: 10.1038/s41586-019-1876-x

14. Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med.* (2019) 25:1054–6. doi: 10.1038/s41591-019-0462-y

15. Schmauch B, Romagnoni A, Pronier E, Saillard C, Maillé P, Calderaro J, et al. A deep learning model to predict RNA-Seq expression of tumours from whole slide images. *Nat Commun.* (2020) 11:3877. doi: 10.1038/s41467-020-17678-4

16. Markel A. Copy and paste of electronic health records: a modern medical illness. *Am J Med.* (2010) 123:e9. doi: 10.1016/j.amjmed.2009.10.012

17. Shwartz-Ziv R, Armon A. *Tabular Data: Deep Learning is Not All You Need* (2021) [arXiv:2106.03253 p.]. Available online at: https://ui.adsabs.harvard.edu/abs/2021arXiv210603253S (accessed June 1, 2021).

18. Kadra A, Lindauer M, Hutter F, Grabocka J. *Well-tuned Simple Nets Excel on Tabular Datasets* (2021) [arXiv:2106.11189 p.]. Available online at: https://ui.adsabs.harvard.edu/abs/2021arXiv210611189K (accessed June 1, 2021).

19. DeGrave AJ, Janizek JD, Lee S-I. AI for radiographic COVID-19 detection selects shortcuts over signal. *Nat Mach Intel.* (2021) 3:610–9. doi: 10.1038/s42256-021-00338-7

20. Yang J, Soltan AAS, Clifton DA. Machine learning generalizability across healthcare settings: insights from multi-site COVID-19 screening. *NPJ Dig Med.* (2022) 5:69. doi: 10.1038/s41746-022-00614-9

21. Lao Q, Jiang X, Havaei M, Bengio Y. *Continuous Domain Adaptation with Variational Domain-Agnostic Feature Replay* (2020) [arXiv:2003.04382 p.]. Available online at: https://ui.adsabs.harvard.edu/abs/2020arXiv200304382L (accessed March 1, 2020).

22. Wang H, He H, Katabi D. *Continuously Indexed Domain Adaptation* (2020) [arXiv:2007.01807 p.]. Available online at: https://ui.adsabs.harvard.edu/abs/2020arXiv200701807W (accessed July 1, 2020).

23. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. *Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization* (2016) [arXiv:1610.02391 p.]. Available online at: https://ui.adsabs.harvard.edu/abs/2016arXiv161002391S (accessed October 1, 2016).

24. Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nat Med.* (2021) 27:2176–82. doi: 10.1038/s41591-021-01595-0

25. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med.* (2019) 25:1337–40. doi: 10.1038/s41591-019-0548-6

frontiers in Neurology

ORIGINAL RESEARCH
published: 31 March 2021
doi: 10.3389/fneur.2021.638267

Machine Learning-Enabled 30-Day Readmission Model for Stroke Patients

Negar Darabi[1], Niyousha Hosseinichimeh[1], Anthony Noto[2], Ramin Zand[2] and Vida Abedi[3,4]**

[1] Department of Industrial and Systems Engineering, Virginia Tech, Falls Church, VA, United States, [2] Geisinger Neuroscience Institute, Geisinger Health System, Danville, PA, United States, [3] Department of Molecular and Functional Genomics, Geisinger Health System, Danville, PA, United States, [4] Biocomplexity Institute, Virginia Tech, Blacksburg, VA, United States

OPEN ACCESS

Edited by:
Ping Zhou,
The University of Rehabilitation, China

Reviewed by:
Harshil Shah,
Guthrie Robert Packer Hospital, United States
Varun Kumar,
University of South Florida, United States
Jianjun Zou,
Nanjing Hospital Affiliated to Nanjing Medical University, China
Lucas Alexandre Ramos,
Academic Medical Center, Netherlands

***Correspondence:**
Vida Abedi
vidaabedi@gmail.com;
vabedi@geisinger.edu
Niyousha Hosseinichimeh
niyousha@vt.edu

Specialty section:
This article was submitted to Stroke, a section of the journal Frontiers in Neurology

Received: *05 December 2020*
Accepted: *08 March 2021*
Published: *31 March 2021*

Citation:
Darabi N, Hosseinichimeh N, Noto A, Zand R and Abedi V (2021) Machine Learning-Enabled 30-Day Readmission Model for Stroke Patients. Front. Neurol. 12:638267. doi: 10.3389/fneur.2021.638267

Background and Purpose: Hospital readmissions impose a substantial burden on the healthcare system. Reducing readmissions after stroke could lead to improved quality of care especially since stroke is associated with a high rate of readmission. The goal of this study is to enhance our understanding of the predictors of 30-day readmission after ischemic stroke and develop models to identify high-risk individuals for targeted interventions.

Methods: We used patient-level data from electronic health records (EHR), five machine learning algorithms (random forest, gradient boosting machine, extreme gradient boosting–XGBoost, support vector machine, and logistic regression-LR), data-driven feature selection strategy, and adaptive sampling to develop 15 models of 30-day readmission after ischemic stroke. We further identified important clinical variables.

Results: We included 3,184 patients with ischemic stroke (mean age: 71 ± 13.90 years, men: 51.06%). Among the 61 clinical variables included in the model, the National Institutes of Health Stroke Scale score above 24, insert indwelling urinary catheter, hypercoagulable state, and percutaneous gastrostomy had the highest importance score. The Model's AUC (area under the curve) for predicting 30-day readmission was 0.74 (95%CI: 0.64–0.78) with PPV of 0.43 when the XGBoost algorithm was used with ROSE-sampling. The balance between specificity and sensitivity improved through the sampling strategy. The best sensitivity was achieved with LR when optimized with feature selection and ROSE-sampling (AUC: 0.64, sensitivity: 0.53, specificity: 0.69).

Conclusions: Machine learning-based models can be designed to predict 30-day readmission after stroke using structured data from EHR. Among the algorithms analyzed, XGBoost with ROSE-sampling had the best performance in terms of AUC while LR with ROSE-sampling and feature selection had the best sensitivity. Clinical variables highly associated with 30-day readmission could be targeted for personalized interventions. Depending on healthcare systems' resources and criteria, models with optimized performance metrics can be implemented to improve outcomes.

Keywords: ischemic stroke, 30-day readmissions, machine learning, statistical analysis, patient readmission

INTRODUCTION

Hospital readmissions impose a substantial financial burden, costing Medicare about $26 billion annually (1). Centers for Medicare and Medicaid Services (CMS) has made reducing 30-day readmission rates a national healthcare reform goal (2) as a way to improve hospital care. Reducing readmissions after stroke could lead to improved quality of care especially since stroke is associated with a high rate of readmission (3).

Studies have found that stroke severity (3, 4), being discharged to skilled nursing, intermediate care facility, hospice, or left against doctor's advice (2, 3, 5–7), being enrolled in Medicaid/Medicare (4, 6, 8, 9), and being married (5) were associated with higher readmissions. A longer length of hospital stay was associated with lower readmissions among stroke patients (5). Heart failure (2, 6, 9), coronary artery disease (10, 11), and dysphagia (4) were also correlated with stroke readmissions. Additionally, patients with anemia, dementia, malnutrition, and diabetes were more likely to be readmitted within 30-day (2, 5, 6, 9).

However, previous studies [**Supplementary Table I** (12)] included a limited number of variables and used logistic regression which restricts the number of included interactions among the variables (13, 14), thus limiting the model performance. Machine learning (ML), more appropriate for high-dimensional datasets (15, 16), has been successfully applied for predicting readmissions after heart failure (17–19), heart attack (20), and other causes of readmissions (21, 22). The goal of this study was to develop prediction models of 30-day readmission among patients with ischemic stroke and identify associated predictors for the development of a more targeted intervention.

METHODS

Study Population

This study was based on the retrospective analysis of prospectively collected data from acute ischemic stroke (AIS) patients at two tertiary centers in Geisinger Health System between January 1, 2015, and October 7, 2018 (23). The data were extracted from electronic health records and de-identified. As a part of the de-identification process, the age of patients older than 89 years old was masked. Patients younger than 18 years of age were excluded from this study. Patients with transient ischemic attack were not included in this study due to the high rate of overdiagnosis (24). The study was reviewed and approved by the Geisinger Institutional Review Board to meet "Non-human subject research," for using de-identified information.

Data Elements

The outcome measure was hospital readmission within 30-day after discharge among patients with AIS. Independent variables included patient age, length of stay (LOS), gender, marital status (married, single, and previously married), and the National Institutes of Health Stroke Scale (NIHSS). The types of health insurance at the time of first admission (Medicare, Medicaid, private, direct employer contract, self-pay, worker compensation, and other government payers) were also included. Other variables in this study were six discharge destinations (discharged to the home health organizations; discharged to home, court, or against medical advice; discharged to hospice-home/hospice-medical facility; discharged or transferred to other facilities; discharged or transferred to Skilled Nursing Facility, SNF; discharged or transferred to another rehab facility), and five clinical interventions (intravenous thrombolysis; insert indwelling urinary catheter; endotracheal tube; percutaneous gastrostomy; and hemodialysis). In addition, a total of 47 comorbidities were included (see **Table 1**).

Data Processing, Feature Selection, and Sampling

Pearson's correlation coefficient was applied to continuous variables to identify those with high collinearity. The correlation matrix between all the predictors along with a list of correlations above 30 and 50% is provided in **Supplementary Figure I** and **Supplementary Table II** (12), respectively. The complete list of variables along with their descriptive statistics and level of missingness was provided in **Table 1**. Student's *t*-test was applied to identify the significant difference between two groups of patients (i.e., readmitted and not readmitted) for each predictor and the test statistics and *P*-values were reported in **Table 1**.

Some of the variables were suffering from missing observations (see **Table 1**). Imputation, using Multivariate Imputation by Chained Equations (MICE) package in R (25), was performed separately on the training and testing sets to ensure an unbiased evaluation of the final model. For the variables with high missingness, we performed an assessment of the distribution of the variable before and after imputation. We used two sets of variables, set one was the comprehensive set including all the variables, and set two included variables selected based on data-driven feature selection, where variables with high collinearity were removed. We used the random forest classification algorithm by Boruta package in RStudio (26) for our data-driven feature selection. Further, to avoid the poor performance of the minority class compared to the dominant class, we applied an adaptive sampling strategy, where we balanced the dataset by applying the Random Over-Sampling Examples (ROSE) algorithm on the minority class (27). The data cleaning and preparation were performed in STATA 14.0 (28) and the analyses were performed using R 3.6.0 (29) in R studio. **Figure 1** shows the processing and modeling pipeline.

Model Development

The de-identified dataset was randomly split into the train set (80%) and test set (20%). We developed models to predict 30-day readmission of ischemic stroke using the training dataset and used ten-fold cross-validation to select the best performing model. Overall, we built fifteen models – based on five different algorithms – following three study designs (Design 1, 2, and 3, see **Figure 1**). The five algorithms included logistic regression (LR), random forest (RF), gradient boosting machine (GBM), extreme gradient boosting (XGBoost), and support vector machines (SVM). Parameter tuning was performed by an automatic grid search with ten different values to randomly try for each

TABLE 1 | Descriptive statistics of variables.

Variables		Missing	Not Readmitted ($n = 2883$)	Readmitted ($n = 301$)	*t* statistics	*P*-value
	Age (y), Mean (SD)	12	71.10(13.90)	71.50(12.90)	−0.52	0.60
	LOS[a] (d), Median (IQR)	319	3 (1, 76)	4 (1, 24)	−4.03	0.00
	All gender	12	2873	299	–	–
Gender, *n* (%)	Female		1406 (48.80)	155 (51.50)	−0.90	0.39
	Male		1467 (50.90)	144 (47.80)	1.00	0.31
	Total	12	2873	299		
Marital Status, *n* (%)	Married		1349 (46.80)	124 (41.20)	1.85	0.06
	Single		425 (14.70)	45 (15.00)	−010	0.92
	Previously married		1099 (38.10)	130 (43.20)	−1.72	0.08
	0 to above 24	2545	580	59	–	–
NIHSS Score, *n* (%)	0 to 4		330 (11.40)	31 (10.30)	0.60	0.55
	5 to 11		150 (5.20)	17 (5.60)	−0.33	0.74
	12 to 23		75 (2.60)	10 (3.30)	−0.74	0.46
	Above 24		25 (0.90)	1 (0.30)	0.98	0.33
	Total	-	291	36	–	–
Procedures, *n* (%)	Intravenous thrombolysis		71 (3.00)	2 (0.80)	2.06	0.04
	Insert indwelling urinary catheters		3 (0.10)	1 (0.30)	−1.07	0.28
	Insert endotracheal tube		148 (5.10)	10 (3.30)	1.35	0.17
	Percutaneous gastrostomy		42 (1.50)	16 (5.40)	−4.81	0.00
	Hemodialysis		27 (0.90)	7 (2.30)	−2.25	0.02
	All centers	-	2883	301	–	–
Hospital, *n* (%)	GMC[b]		1784 (61.90)	176 (58.50)	1.08	0.27
	GWV[c]		1099 (38.10)	125 (41.50)	−1.16	0.25
	Total	492	2418	274	–	–
Discharge Status, *n* (%)	Discharged to home health organization		346 (12.00)	37 (12.30)	−0.15	0.88
	Discharged to home, court, or against medical advice		902 (31.30)	57 (18.90)	4.46	0.00
	Discharged to hospice-home/hospice-medical facility		82 (2.80)	4 (1.30)	1.54	0.12
	Discharged/transferred to other facilities		27 (0.90)	2 (0.70)	0.47	0.63
	Discharged/transferred to *SNF*[d]		447 (15.50)	91 (30.20)	−6.53	0.00
	Discharged/transferred to another rehab facility		614 (21.30)	83 (27.60)	−2.51	0.01
	Total	51	2836	297	–	–
Payer, *n* (%)	Direct employer contract		65 (2.30)	9 (3.00)	−0.80	0.42
	Medicaid		216 (7.50)	22 (7.30)	0.11	0.91
	Medicare		2037 (70.70)	230 (76.40)	−2.10	0.03
	Other government payers		58 (2.00)	4 (1.30)	0.81	0.41
	Private		425 (14.70)	30 (1.00)	2.25	0.02
	Self-pay		32 (1.10)	1 (0.30)	1.27	0.20
	Workers compensation		3 (0.10)	1 (0.30)	−1.06	0.29
Diagnoses, *n* (%)	Anemia	-	319 (13.70)	67 (26.40)	−5.42	0.00
	Atrial fibrillation		691 (29.40)	93 (36.30)	−2.29	0.02
	Anxiety disorders		328 (14.00)	53 (20.70)	−2.90	0.00
	Cerebral arterial dissection		23 (1.00)	6 (2.30)	−1.98	0.05
	Coronary artery disease		684 (29.10)	77 (30.10)	−0.32	0.75
	Delirium		44 (1.90)	12 (4.70)	−2.95	0.00
	Dementia		222 (9.50)	35 (13.70)	−2.15	0.03
	Diabetes		642 (27.50)	86 (33.90)	−2.13	0.03
	Dysphagia		150 (6.40)	29 (11.30)	−2.97	0.00
	Heart failure		426 (18.30)	64 (25.20)	−2.67	0.00

(Continued)

TABLE 1 | Continued

Variables		Missing	Not Readmitted ($n = 2883$)	Readmitted ($n = 301$)	*t* statistics	*P*-value
	Hypercoagulable state		17 (0.70)	7 (2.70)	−3.20	0.00
	Hypertension		1308 (56.10)	131 (51.60)	1.38	0.17
	Hypotension		55 (2.30)	13 (5.10)	−2.61	0.00
	Kidney disease		670 (28.50)	101 (39.50)	−3.65	0.00
	Malignancy		286 (12.20)	47 (18.40)	−2.82	0.00
	Malnutrition		105 (4.50)	32 (12.50)	−5.49	0.00
	Migraine		69 (2.90)	13 (5.10)	−1.86	0.06
	Overweight		58 (2.50)	16 (6.20)	−3.46	0.00
	Tobacco use		1171 (49.90)	143 (55.90)	−1.83	0.07
	Venous thrombosis		81 (3.40)	18 (7.00)	−2.85	0.00
	Acute myocardial infarction		36 (1.50)	9 (3.50)	−2.31	0.02
	Alcohol use		95 (4.00)	10(3.90)	0.11	0.91
	Arrhythmias		99 (4.20)	15 (5.90)	−1.22	0.22
	Blindness		14 (0.60)	1 (0.40)	0.41	0.68
	Cardiac valvular disease		159 (6.80)	22 (8.60)	−1.09	0.27
	Cardiomyopathy		131 (5.60)	23 (9.00)	−2.20	0.03
	Cerebral atherosclerosis		93 (4.00)	13 (5.10)	−0.86	0.39
	Chronic kidney disease		583 (25.00)	89 (35.00)	−3.47	0.00
	Chronic liver disease		49 (2.10)	2 (0.80)	1.43	0.15
	Chronic lung disease		476 (20.30)	69 (27.00)	−2.50	0.01
	Dysautonomia		16 (0.70)	2 (0.80)	−0.18	0.85
	Hyperlipidemia		1537 (65.40)	177 (69.10)	−1.19	0.23
	Intracerebral hemorrhage		521 (22.20)	51 (19.90)	0.83	0.41
	Inflammatory disorders		66 (2.80)	5 (2.00)	0.80	0.42
	Mood disorders		372 (15.80)	58 (22.70)	−2.79	0.00
	Non-compliance		107 (4.60)	12 (4.70)	−0.09	0.92
	Normal weight		44 (1.90)	11 (4.30)	−2.56	0.01
	Obese		466 (19.80)	62 (24.20)	−1.65	0.09
	Palliative care on board		254 (10.80)	12 (4.70)	3.08	0.00
	Peripheral vascular disease		125 (5.30)	17 (6.60)	−0.88	0.38
	Respiratory failure		164 (7.00)	19 (7.40)	−0.26	0.79
	Seizure disorders		96 (4.10)	15 (5.90)	−1.33	0.18
	Sleep apnea		216 (9.20)	25 (9.80)	−0.30	0.76
	Systemic infection		63 (2.70)	16 (6.20)	−3.17	0.00
	Thyroid disease		445 (18.90)	54 (21.10)	−0.83	0.41
	Underweight		34 (1.40)	9 (3.50)	−2.47	0.01
	Use of steroids		72 (3.10)	11 (4.30)	−1.06	0.29
	Year, Median (IQR)	–	2016 (2015, 2018)	2016 (2015, 2018)	−0.36	0.72

[a], length of stay; [b], geisinger medical center, [c], Geisinger wyoming valley, [d], skilled nursing facility.

algorithm parameter. All the hyperparameter evaluation and model development were performed using the Caret package in R Studio (30). We ran the SVM with and without normalization of the dataset. In normalization, we scaled the data to calculate the standard deviation for an attribute and divided each value by that standard deviation. Then we centered the data to calculate the mean for an attribute and subtracted it from each value. The performance measures of the models were evaluated using the 20% test set. To compare the performance of the applied models, we calculated the area under the receiver operating characteristic curve (AUC). We also used other performance measures such as sensitivity or recall, specificity, and positive predictive value (PPV) as well as training time.

RESULTS

Study Design and Population Characteristics

A total number of 3,184 AIS patients [1,960 patients from Geisinger Medical Center (GMC) and 1,224 from Geisinger Wyoming Valley Medical Center (GWV)] were included in this study.

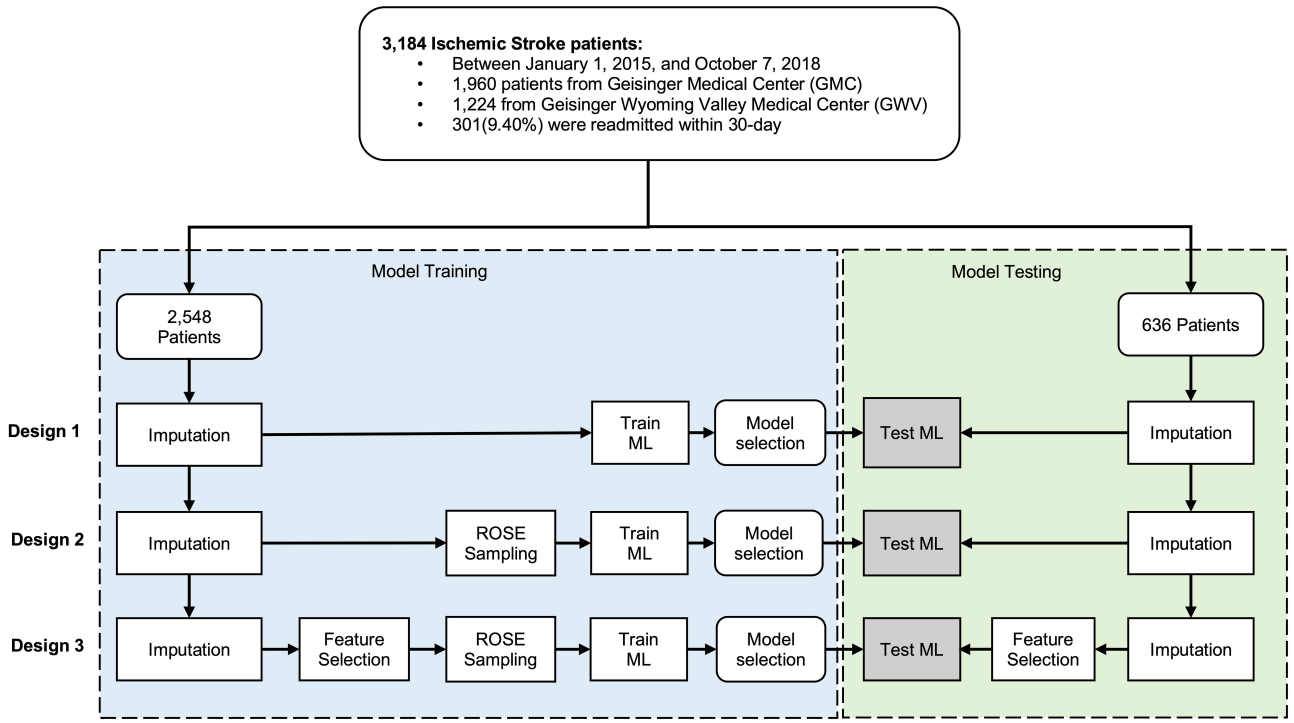


FIGURE 1 | Data processing flowchart.

Among 3,184 patients with ischemic stroke, 301(9.40%) were readmitted within 30-day. The train set and test set included 2,548 (80%) and 636 (20%) patient-level observations, respectively. In **Table 1**, the patients were compared based on diverse characteristics including demographic characteristics, medical history prior to the ischemic stroke event, and stroke severity using the NIHSS score. Continuous variables were presented as mean and standard deviation and as median with interquartile range (IQR). The average age of patients was 71 (interquartile range, IQR: 18–89) and 1,611(50.60%) patients were men. There was a significant difference between patients who were readmitted and those who were not in terms of median LOS, being married or previously married, discharged to SNF or against medical advice, and having Medicare or private insurance.

Models Can Be Trained to Predict 30-Day Readmission Using EHR

The performance metrics—AUC and its 95% confidence interval (CI), sensitivity, specificity, PPV, and the training time —for all the 15 models with and without ROSE-sampling (Design 2, and 1), and with feature selection and ROSE-sampling (Design 3) were reported in **Table 2**. The CIs for the test sets were calculated using bootstrapping. We also provided the confusion matrices of all 15 models in **Supplementary Table V**. The results showed that applying ROSE for addressing class imbalance during the model training improved the AUC, PPV, and specificity of models during the testing phase. However, feature selection did not improve the results [see **Table 2**, and **Supplementary Figures II**, **III** (12)]. Feature selection was performed using the Boruta package which reduced the number of features from 52 to 14 [see green variables in **Supplementary Figure IV** (12)]. These 14 attributes were used in the third design while all features were included in the other designs.

The ROC curves for LR, RF, GBM, XGBoost, and SVM without feature selection and sampling (Design 1) and with ROSE-sampling (Design 2) were shown in the top and bottom side of **Figure 2** accordingly. In the absence of sampling and feature selection, GBM provided the highest AUC (0.68), specificity (0.95), and PPV (0.33) when compared to the other models (**Figure 2** and **Table 2**). However, the best AUC (0.74), PPV (0.43), and specificity (0.98) were reached when ROSE-sampling was applied. The optimal model parameters for ROSE-sampled XGBoost were max-depth = 4, subsample = 0.50, colsample_bytree = 0.80, gamma = 0, and min_child_weight = 10. In terms of AUC, specificity, and PPV, the LR in Design 2 had poor performance compared to XGBoost and GBM models. However, LR with feature selection and ROSE-sampling (Design 3) provided the highest sensitivity (0.53) relative to other models. We also performed SVM with normalized data and the results are provided in **Supplementary Table IV**.

The training times for LR, RF, and GBM were faster when compared to models based on XGBoost and SVM (see **Table 2**). The model training was performed using MacBook Pro14,2, four thunderbolt 3 ports with 3.1 GHz Dual-Core Intel Core i5, and 8 GB memory. Overall, the addition of the sampling step increased

TABLE 2 | Performance metrics for machine learning models.

Method	Train set		Test set				
	AUC	Training Time (s)	AUC	95% CI for AUC	Sensitivity	Specificity	PPV
	No feature selection and sampling (Design 1)						
LR	0.76	3	0.60	(0.52, 0.67)	0.32	0.86	0.19
RF	0.82	42	0.57	(0.50, 0.64)	0.09	0.93	0.12
GBM	0.70	48	0.68	(0.52, 0.76)	0.23	0.95	0.33
XGBoost	0.76	1,752	0.62	(0.56, 0.69)	0.30	0.88	0.21
SVM	0.98	990	0.62	(0.56, 0.70)	0.30	0.86	0.18
	With ROSE-sampling (Design 2)						
LR	0.74	3	0.63	(0.55, 0.70)	0.38	0.72	0.12
RF	0.74	33	0.67	(0.51, 0.76)	0.09	0.97	0.26
GBM	0.74	48	0.70	(0.61, 0.75)	0.09	0.98	0.45
XGBoost	0.76	2,340	0.74	(0.64, 0.78)	0.20	0.98	0.43
SVM	0.83	1,689	0.67	(0.59, 0.74)	0.38	0.89	0.27
	With feature selection and ROSE-sampling (Design 3)						
LR	0.70	3	0.64	(0.56, 0.72)	0.53	0.69	0.15
RF	0.70	12	0.65	(0.56, 0.70)	0.30	0.89	0.24
GBM	0.69	30	0.66	(0.58, 0.74)	0.17	0.95	0.26
XGBoost	0.70	2,130	0.65	(0.56, 0.73)	0.17	0.95	0.27
SVM	0.72	960	0.64	(0.56, 0.72)	0.42	0.77	0.16

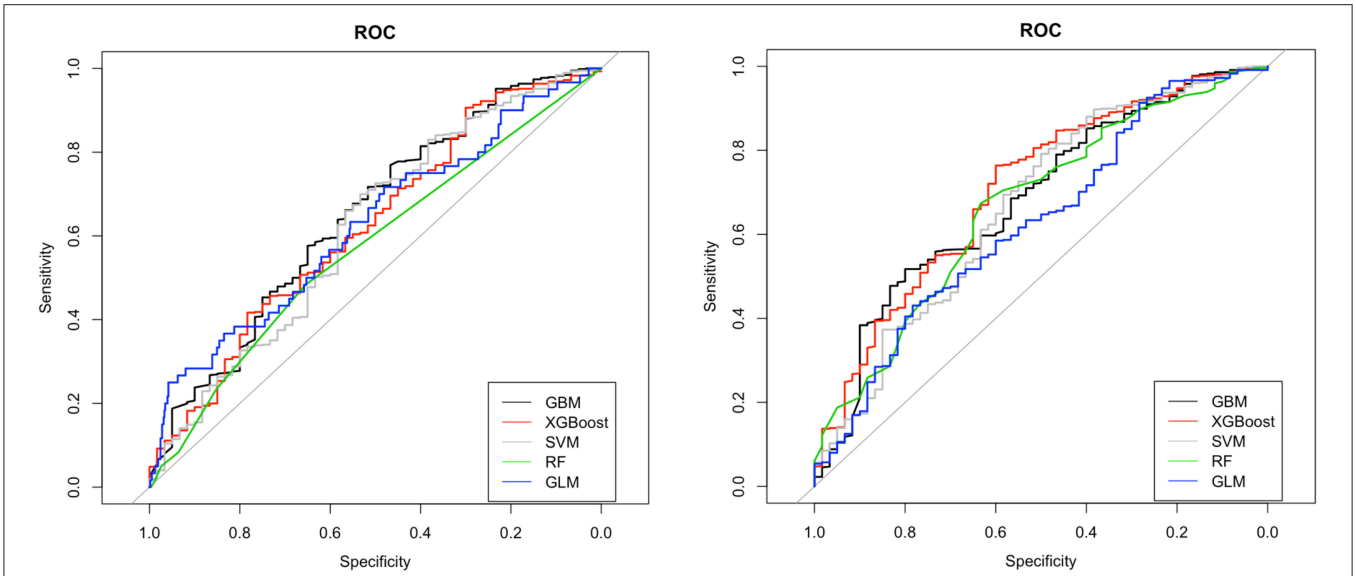


FIGURE 2 | ROC curves for machine learning models with (bottom) and without (top) ROSE-sampling. GBM, gradient boosting machine; XGBoost, extreme gradient boosting; SVM, support vector machines; RF, random forest; and GLM, generalized linear model with logit link which is logistic regression in our study.

the training time, while having fewer features resulted in faster training as expected.

NIHSS, Insert Indwelling Urinary Catheter, Hypercoagulable State, and Percutaneous Gastrostomy Are the Top Predictors of 30-Day Readmission

Using XGBoost in Design 2, the best predictive model, we identified the most important predictors of 30-day readmission. According to the variable importance scores for XGBoost in Design 2 (**Table 3**), the top 10 predictors of 30-day readmission were NIHSS above 24, insert indwelling urinary catheter, hypercoagulable state, percutaneous gastrostomy, using workers compensation as insurance, hemodialysis, overweight, cerebral arterial dissection, malnutrition, intravenous thrombolysis, and venous thrombosis.

We also reported the result of LR in the third design. In the latter, the multicollinearity was addressed by feature selection (**Table 4**). The odds ratios (OR), log odds, 95% CI,

TABLE 3 | Variable importance scores of the XGBoost model with ROSE-sampling (Design 2).

No.	Variable	Importance Score (out of 100)
1	NIHSS above 24	100.00
2	Insert Indwelling Urinary Catheter	89.94
3	Hypercoagulable State	61.95
4	Percutaneous Gastrostomy	40.04
5	Payer Workers Compensation	37.78
6	Hemodialysis	35.13
7	Overweight	33.35
8	Cerebral Arterial Dissection	30.61
9	Malnutrition	25.74
10	intravenous thrombolysis	24.15
11	Venous Thrombosis	22.94
12	Discharged to Hospice-Home or Hospice-Medical Facility	17.15
13	Palliative Care on Board	14.92
14	Delirium	14.54
15	Payer Self-Pay	14.42

TABLE 4 | Logistic regression results for predictors of 30-day readmission in ischemic stroke patients (Design 3).

Variables	OR	Log Odds	95% CI (2.5%, 97.5%)		*P*-value
(Intercept)	0.25	−1.39	−2.55	−0.30	0.01
Age	0.99	−0.01	−0.03	0.00	0.07
Discharged to home health organization	1.61	0.47	−0.01	0.94	0.05
Discharged to hospice-home or hospice-medical facility	0.44	−0.83	−2.04	0.28	0.16
Discharged/transferred to other facilities	0.49	−0.70	−2.62	0.66	0.38
Discharged/transferred to another rehab facility	1.79	0.58	0.17	0.99	0.01
Discharged/transferred to SNF	2.77	1.02	0.56	1.48	0.00
Medicaid	0.41	−0.89	−1.79	0.07	0.06
Medicare	0.57	−0.55	−1.31	0.30	0.17
Other government payers	0.39	−0.93	−2.33	0.32	0.16
Private insurance	0.40	−0.91	−1.75	−0.00	0.04
Self-pay	0.29	−1.23	−4.21	0.59	0.27
Workers compensation	1.82	0.60	−2.59	3.04	0.65
Chronic kidney disease	1.28	0.24	−0.36	0.89	0.44
Hypercoagulable state	3.09	1.13	0.18	1.99	0.01
Kidney disease	1.15	0.14	−0.49	0.72	0.65
Malignancy	2.10	0.74	0.37	1.09	0.00
Malnutrition	2.51	0.92	0.39	1.42	0.00
Palliative care on board	1.13	0.13	−0.73	0.90	0.76
Respiratory failure	0.74	−0.30	−1.06	0.39	0.41
Underweight	1.00	0.00	−1.01	0.91	0.99
Insert endotracheal tube	0.86	−0.15	−1.11	0.74	0.76
Percutaneous gastrostomy	1.39	0.33	−0.51	1.09	0.42
NIHSS 12 to 23	1.76	0.57	0.14	0.99	0.01
NIHSS 5 to 11	0.65	−0.43	−0.79	−0.08	0.02
NIHSS above 24	0.17	−1.77	−3.25	−0.65	0.01

and *P*-values were reported in this table. This analysis revealed that being discharged to SNF, malignancy, and malnutrition were significantly associated with stroke readmission within 30-day (p-value < 0.0001). Also, being discharged to a rehabilitation facility and stroke severity above twelve were significantly associated with 30-day readmission at 0.001 significance level.

DISCUSSION

We have taken a comprehensive approach to identify and prioritize factors associated with 30-day readmissions after ischemic stroke. We aimed to find the most effective predictive model by comparing the results of different ML techniques and LR. There have been multiple readmission studies that developed predictive models for the chances of 30-day readmission in stroke patients. However, most of these models used LR (31) which limits the inclusion of higher-order interactions among variables and does not perform well in the presence of collinearity. Also, many studies considered readmission after 90 days or 1 year as a dependent variable which is a long follow-up period, as CMS penalizes healthcare systems for readmission under 30 days. In this study, we addressed these gaps and improved the prediction performance of readmissions in stroke patients using a wide range of potential risk factors and the proper ML techniques. Our results show that depending on the resources and criteria of healthcare systems, a predictive model with optimized performance metrics can be used to improve decision making.

Machine Learning-Based Models Can Be Trained to Predict 30-Day Readmission

The results of this study indicate that ML-based models can be designed to predict 30-day readmission after stroke using structured data from EHR. ML algorithms can include higher-order interactions among variables, handle multicollinearity, and improve readmission predictions when applied to large and high-dimensional datasets (15). This study was the first in predicting the associated variables of 30-day ischemic stroke readmission using ML techniques. Our findings indicated that the best performance in terms of AUC, specificity, and PPV was obtained when XGBoost was used with ROSE-sampling.

Past studies that used ML techniques to improve the prediction power, either performed their analysis on readmission more than 30-day or studied other causes of readmission such as heart failure (17, 18, 21). However, our best performing model (XGBoost in Design 2) provides higher AUC and PPV compared to these studies [See **Supplementary Table III** (12)].

Clinical Features Highly Associated With 30-Day Readmission

The results of our best performing model (XGBoost in Design 2) showed that NIHSS score above 24, insert indwelling urinary

catheter, hypercoagulable state, percutaneous gastrostomy, and insurance type are among factors with the highest importance. The common significant predictors of the 30-day readmission in both XGBoost in Design 2 and LR in Design 3 included NIHSS score above 24, hypercoagulable state, and malnutrition. Since NIHSS is an important variable and this variable also suffered from high missingness, we assessed its distribution before and after imputation for both train and test sets. Our results corroborate that the distribution of this variable remains the same after applying imputation (see **Supplementary Table VI**, **Supplementary Figures V, VI**).

Additionally, malignancy, NIHSS scores between 5 and 23, private insurance type, and being discharged to a rehabilitation facility or SNF were only significant in the LR, and they had low importance scores in the XGBoost model. Among all variables, stroke severity and malnutrition were found significant predictors of 30-day readmission in ischemic stroke patients in past studies and our results corroborated the previous findings (2–6, 9).

It has been shown in previous studies that heart failure and being Medicare or Medicaid user were significantly correlated with 30-day readmission (2, 4, 6, 8, 9). However, we found no evidence in favor of these assertions. Past studies provided mixed results on the importance of age, hypertension, and gender; some studies found that patients of older age were more likely to be readmitted (2, 5) while others showed that age was not a significant predictor (3, 32). Also, hypertension was found as a significant risk factor of readmission in a study (8) while in other works authors claimed that hypertension was not significantly associated with 30-day readmission (13, 32). Several studies conducted on data from Taipei, China, and Western Australia found that gender of patients was not significantly associated with the chances of being readmitted (3, 5, 32); however, studies based on U.S. data have found women were significantly at higher risk of readmission (2, 8, 13). The results of the ROSE-sampled XGBoost model indicated that age, hypertension, and gender–in this specific cohort–were not significantly associated with 30-day readmission after ischemic stroke. We have also performed a detailed analysis of our Geisinger cohort and identified that sex was not an independent risk factor for all-cause mortality and ischemic stroke recurrence (33). Finally, the identification of malnutrition provides potential new venues to improve secondary prevention and outcome (34).

Model Performance Metrics Optimized Based on the Target Goals

According to our results, the best performing predictive model, which was ROSE-sampled XGBoost, had a 17.5% improvement in AUC compared to LR. This XGBoost performed better in comparison with other models of 30-day readmission in the literature (17, 18, 21). We improved the AUC up to 0.74 (95% CI: 0.64, 0.78) for the test set with 0.43 PPV (see Design 2 in **Table 2**). In the absence of sampling and feature selection, GBM returned very close AUC for the training and testing sets, corroborating that the models did not suffer from overfitting (Design 1 in **Table 2**). XGBoost and GBM with ROSE-sampling achieved comparable AUC for the testing and training sets, confirming that these models did not suffer from overfitting (Design 2 in **Table 2**). However, the SVM-based models had the largest difference between testing and training AUC, leading to the possibility of overfitting given this dataset. Overall, ML-based models such as GBM and XGBoost improved the prediction of 30-day readmission in stroke patients compared to traditional LR [see **Table 2**, **Supplementary Figures II**, **III** (12)]. However, LR with feature selection and ROSE-sampling provided the best sensitivity which implies that healthcare systems can choose their decision models based on their resources and criteria.

LIMITATIONS

One of the important strengths of this study was that we analyzed a diverse list of potential predictors including an extensive number of clinical interventions and patient's comorbidities. To the best of our knowledge, this was the first attempt to apply ML techniques to predict the 30-day readmission for ischemic stroke patients. Considering a large number of included variables in our dataset, these ML techniques could include higher-order interactions among variables, and improve the prediction power when compared to LR.

Our analysis had several limitations. Although our dataset was rich in the number of variables, the number of patients was relatively small compared to the included independent variables. Therefore, the small number of observations might result in overfitting in the models. However, comparable AUC measures provided by XGBoost for the testing and training sets rule out the possibility of overfitting in this model. Another limitation of this work was missing data specifically for the NIHSS score. The most missing data points belonged to the NIHSS score before 2016 and we applied imputation to not lose any observation or cause sampling bias. Additionally, due to the unique demographic characteristics of this dataset (the majority of patients were white and from non-urban areas), the results may not be generalizable to other health systems.

FUTURE DIRECTIONS

In this study, we only considered ischemic stroke as the cause of readmission. Therefore, future avenues of research can be done by considering other stroke types and subtypes. However, considering the size of our dataset which came from two health centers from central Pennsylvania, further work needs to focus on a larger population with diverse demographics to introduce a generalizable model. Additionally, to improve the prediction power, future studies may include the application of deep learning techniques (35) as well as the integration of features from unstructured sources such as clinical notes and imaging reports. Finally, improvement in parameter optimization, by using sensitivity analysis (SA)-based approaches (36, 37) and improving the imputation for laboratory values for EHR-mining (38) can lead to an improvement in outcome prediction models using administrative datasets. These strategies will help in model

generalizability, improve patient representation, and reduce algorithmic bias.

CONCLUSION

Our results showed that machine learning-based predictive models perform better than traditional logistic regression, enabling the inclusion of a more comprehensive set of variables into the model. The insights from this work can assist with the identification of ischemic stroke patients who are at higher risk of readmission for more targeted preventive strategies. Our study also indicated the importance of including multiple performance metrics for empowering the healthcare system to choose a predictive model for implementation as an assistive decision support tool into their EHR based on their resources and criteria.

DATA AVAILABILITY STATEMENT

All relevant data are available in the article/**Supplementary Material**. Due to privacy and other restrictions, the primary data cannot be made openly available. Deidentified data may be available subject to data-sharing agreement with Geisinger Health System. Details about requesting access to the data are available from the Geisinger's corresponding author Vida Abedi.

AUTHOR CONTRIBUTIONS

NH, VA, and RZ: conception and design of the study. VA and NH: supervision of the project. AN and VA acquisition of the data. ND, VA, and NH: analysis of the data. ND: implementation of the code and Drafting a significant portion of the manuscript or figures. NH, VA, ND and RZ: interpretation of the findings. VA, RZ, and NH: editing the manuscript. ND, NH, VA, RZ, and AN: participation in discussions on the model and results. All authors contributed to the article and approved the submitted version.

FUNDING

VA had financial research support from the National Institute of Health (NIH) grant no. R56HL116832 sub-awarded to Geisinger during the study period. RZ had financial research support from Bucknell University Initiative Program, Roche – Genentech Biotechnology Company, the Geisinger Health Plan Quality fund, and receives institutional support from Geisinger Health System during the study period.

ACKNOWLEDGMENTS

The authors would like to thank Venkatesh Avula, Durgesh Chaudhary, and Jiang Li for thoughtful discussion during the modeling development and Matthew C. Gass for data de-identification.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fneur.2021.638267/full#supplementary-material

REFERENCES

1. LaPointe J. *3 Strategies to Reduce Hospital Readmission Rates, Costs*. (2018). Available online at: https://revcycleintelligence.com/news/3-strategies-to-reduce-hospital-readmission-rates-costs (accessed January 08, 2018).
2. Lichtman JH, Leifheit-Limson EC, Jones SB, Wang Y, Goldstein LB. Preventable readmissions within 30 days of ischemic stroke among medicare beneficiaries. *Stroke*. (2013) 44:3429–35. doi: 10.1161/STROKEAHA.113.003165
3. Chuang K-Y, Wu S-C, Ma A-HS, Chen Y-H, Wu C-L. Identifying factors associated with hospital readmissions among stroke patients in Taipei. *J Nurs Res*. (2005) 13:117–28. doi: 10.1097/01.JNR.0000387533.07395.42
4. Jia H, Zheng Y, Reker DM, Cowper DC, Wu SS, Vogel WB, et al. Multiple system utilization and mortality for veterans with stroke. *Stroke*. (2007) 38:355–60. doi: 10.1161/01.STR.0000254457.38901.fb
5. Wen T, Liu B, Wan X, Zhang X, Zhang J, Zhou X, et al. Risk factors associated with 31-day unplanned readmission in 50,912 discharged patients after stroke in China. *BMC Neurol*. (2018) 18:218. doi: 10.1186/s12883-018-1209-y
6. Smith MA, Liou J-I, Frytak JR, Finch MD. 30-day survival and rehospitalization for stroke patients according to physician specialty. *Cerebrovasc Dis*. (2006) 22:21–6. doi: 10.1159/000092333
7. Burke JF, Skolarus LE, Adelman EE, Reeves MJ, Brown DL. Influence of hospital-level practices on readmission after ischemic stroke. *Neurology*. (2014) 82:2196–204. doi: 10.1212/WNL.0000000000000514
8. Kennedy BS. Does race predict stroke readmission? An analysis using the truncated negative binomial model. *J Natl Med Assoc*. (2005) 97:699.
9. Smith MA, Frytak JR, Liou J-I, Finch MD. Rehospitalization and survival for stroke patients in managed care and traditional medicare plans. *Med Care*. (2005) 43:902. doi: 10.1097/01.mlr.0000173597.97232.a0
10. Heller RF, Fisher JD, O'Este CA, Lim LLY, Dobson AJ, Porter R. Death and readmission in the year after hospital admission with cardiovascular disease: the hunter area heart and stroke register. *Med J Aust*. (2000) 172:261–5. doi: 10.5694/j.1326-5377.2000.tb123940.x
11. Lin H-J, Chang W-L, Tseng M-C. Readmission after stroke in a hospital-based registry: risk, etiologies, and risk factors. *Neurology*. (2011) 76:438–43. doi: 10.1212/WNL.0b013e31820a0cd8
12. *Supplemental Material*. Available online at: https://www.ahajournals.org/journal/str (accessed March 19, 2021).
13. Lichtman JH, Leifheit-Limson EC, Jones SB, Watanabe E, Bernheim SM, Phipps MS, et al. Predictors of hospital readmission after stroke: a systematic review. *Stroke*. (2010) 41:2525–33. doi: 10.1161/STROKEAHA.110.599159
14. Ouwerkerk W, Voors AA, Zwinderman AH. Factors influencing the predictive power of models for predicting mortality and/or heart failure hospitalization in patients with heart failure. *JACC Heart Fail*. (2014) 2:429–36. doi: 10.1016/j.jchf.2014.04.006
15. Friedman J, Hastie T, Tibshirani R. *The Elements of Statistical Learning*. New York, NY; Springer series in statistics (2001).
16. Noorbakhsh-Sabet N, Zand R, Zhang Y, Abedi V. Artificial intelligence transforms the future of health care. *Am J Med*. (2019) 132:795–801. doi: 10.1016/j.amjmed.2019.01.017
17. Mortazavi BJ, Downing NS, Bucholz EM, Dharmarajan K, Manhapra A, Li S-X, et al. Analysis of machine learning techniques for heart

failure readmissions. *Circ Cardiovasc Qual Outcomes.* (2016) 9:629–40. doi: 10.1161/CIRCOUTCOMES.116.003039
18. Golas SB, Shibahara T, Agboola S, Otaki H, Sato J, Nakae T, et al. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. *BMC Med Inform Decis Mak.* (2018) 18:44. doi: 10.1186/s12911-018-0620-z
19. Frizzell JD, Liang L, Schulte PJ, Yancy CW, Heidenreich PA, Hernandez AF, et al. Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: comparison of machine learning and other statistical approaches. *JAMA Cardiol.* (2017) 2:204–9. doi: 10.1001/jamacardio.2016.3956
20. Francisco A, Stabler ME, Hisey W, Mackenzie TA, Dorn C, Denton J, et al. Using machine learning to predict 30-day readmission of patients hospitalized with an acute myocardial infarction. *Circulation.* (2018) 138(Suppl. 1):A15808.
21. Wolff P, Graña M, Ríos SA, Yarza MB. Machine learning readmission risk modeling: a pediatric case study. *Bio Med Res Int.* (2019) 2019:1–19. doi: 10.1155/2019/8532892
22. Kalagara S, Eltorai AE, Durand WM, DePasse JM, Daniels AH. Machine learning modeling for predicting hospital readmission following lumbar laminectomy. *J Neurosurg Spine.* (2018) 30:344–52. doi: 10.3171/2018.8.SPINE1869
23. Chaudhary D, Khan A, Shahjouei S, Gupta M, Lambert C, Avula V, et al. Trends in ischemic stroke outcomes in a rural population in the United States. *J Neurol Sci.* (2021) 422:117339. doi: 10.1016/j.jns.2021.117339
24. Sadighi A, Stanciu A, Banciu M, Abedi V, El Andary N, Holland N, et al. Rate and associated factors of transient ischemic attack misdiagnosis. *Eneurologicalsci.* (2019) 15:100193. doi: 10.1016/j.ensci.2019.100193
25. Zhang Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. *Ann Transl Med.* (2016) 4:30. doi: 10.3978/j.issn.2305-5839.2015.12.63
26. Kursa MB, Rudnicki WR. Feature selection with the boruta package. *J Stat Softw.* (2010) 36:1–13. doi: 10.18637/jss.v036.i11
27. Lunardon N, Menardi G, Torelli N. ROSE: a package for binary imbalanced learning. *R journal.* (2014) 6:79–89. doi: 10.32614/RJ-2014-008
28. STATA. *STATA 14*. (2015). Available online at: https://www.stata.com/stata14/ (accessed March 19, 2021).
29. *R. R 3.6.0*. (2019). Available online at: https://cran.r-project.org/bin/windows/base/old/3.6.0/ (accessed April 26, 2019).
30. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw.* (2008) 28:1–26. doi: 10.18637/jss.v028.i05
31. Bambhroliya AB, Donnelly JP, Thomas EJ, Tyson JE, Miller CC, McCullough LD, et al. Estimates and temporal trend for US nationwide 30-day hospital readmission among patients with ischemic and hemorrhagic stroke. *JAMA Netw open.* (2018) 1:e181190. doi: 10.1001/jamanetworkopen.2018.1190
32. Lee AH, Yau KK, Wang K. Recurrent ischaemic stroke hospitalisations: a retrospective cohort study using Western Australia linked patient records. *Eur J Epidemiol.* (2004) 19:999–1003. doi: 10.1007/s10654-004-0157-6
33. Lambert C, Chaudhary D, Olulana O, Shahjouei S, Avula V, Li J, Abedi V, Zand R. Sex Disparity in Long-term Stroke Recurrence and Mortality in a Rural Population in the United States. *Ther Adv Neurol Disord.* (2020) 13:1–12. doi: 10.1177/1756286420971895
34. Sharma V, Sharma V, Khan A, Wassmer DJ, Schoenholtz MD, Hontecillas R, et al. Malnutrition, health and the role of machine learning in clinical setting. *Front Nutr.* (2020) 7:44. doi: 10.3389/fnut.2020.00044
35. Ding L, Liu C, Li Z, Wang Y. Incorporating artificial intelligence into stroke care and research. *Stroke.* (2020) 51:e351–4. doi: 10.1161/STROKEAHA.120.031295
36. Alam M, Deng X, Philipson C, Bassaganya-Riera J, Bisset K, Carbo A, et al. Sensitivity analysis of an enteric immunity simulator (ENISI)-based model of immune responses to helicobacter pylori infection. *PLoS ONE.* (2015) 10:e0136139. doi: 10.1371/journal.pone.0136139
37. Chen X, Wang W, Xie G, Hontecillas R, Verma M, Leber A, et al. Multi-resolution sensitivity analysis of model of immune response to helicobacter pylori infection via spatio-temporal metamodeling. *Front Appl Math Stat.* (2019) 5:4. doi: 10.3389/fams.2019.00004
38. Abedi V, Li J, Shivakumar MK, Avula V, Chaudhary DP, Shellenberger MJ, et al. Increasing the density of laboratory measures for machine learning applications. *J Clin Med.* (2021) 10:103. doi: 10.3390/jcm10010103

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neurology

ORIGINAL RESEARCH
published: 29 April 2021
doi: 10.3389/fneur.2021.656112

Validation of a Deep Learning Tool in the Detection of Intracranial Hemorrhage and Large Vessel Occlusion

Joel McLouth[1†], Sebastian Elstrott[1†], Yasmina Chaibi[2], Sarah Quenet[2], Peter D. Chang[1,3], Daniel S. Chow[1,3] and Jennifer E. Soun[1]*

[1] Department of Radiological Sciences, University of California, Irvine, Irvine, CA, United States, [2] Avicenna.AI, La Ciotat, France, [3] Center for Artificial Intelligence in Diagnostic Medicine, University of California, Irvine, Irvine, CA, United States

OPEN ACCESS

Edited by:
Vida Abedi,
Geisinger Health System,
United States

Reviewed by:
Aristeidis H. Katsanos,
McMaster University, Canada
Philipp Hendrix,
Geisinger Health System,
United States
Georgios Tsivgoulis,
National and Kapodistrian University of Athens, Greece

***Correspondence:**
Joel McLouth
mclouthj@hs.uci.edu

† These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Stroke, a section of the journal Frontiers in Neurology

Received: 20 January 2021
Accepted: 07 April 2021
Published: 29 April 2021

Citation:
McLouth J, Elstrott S, Chaibi Y, Quenet S, Chang PD, Chow DS and Soun JE (2021) Validation of a Deep Learning Tool in the Detection of Intracranial Hemorrhage and Large Vessel Occlusion. Front. Neurol. 12:656112. doi: 10.3389/fneur.2021.656112

Purpose: Recently developed machine-learning algorithms have demonstrated strong performance in the detection of intracranial hemorrhage (ICH) and large vessel occlusion (LVO). However, their generalizability is often limited by geographic bias of studies. The aim of this study was to validate a commercially available deep learning-based tool in the detection of both ICH and LVO across multiple hospital sites and vendors throughout the U.S.

Materials and Methods: This was a retrospective and multicenter study using anonymized data from two institutions. Eight hundred fourteen non-contrast CT cases and 378 CT angiography cases were analyzed to evaluate ICH and LVO, respectively. The tool's ability to detect and quantify ICH, LVO, and their various subtypes was assessed among multiple CT vendors and hospitals across the United States. Ground truth was based off imaging interpretations from two board-certified neuroradiologists.

Results: There were 255 positive and 559 negative ICH cases. Accuracy was 95.6%, sensitivity was 91.4%, and specificity was 97.5% for the ICH tool. ICH was further stratified into the following subtypes: intraparenchymal, intraventricular, epidural/subdural, and subarachnoid with true positive rates of 92.9, 100, 94.3, and 89.9%, respectively. ICH true positive rates by volume [small (<5 mL), medium (5–25 mL), and large (>25 mL)] were 71.8, 100, and 100%, respectively. There were 156 positive and 222 negative LVO cases. The LVO tool demonstrated an accuracy of 98.1%, sensitivity of 98.1%, and specificity of 98.2%. A subset of 55 randomly selected cases were also assessed for LVO detection at various sites, including the distal internal carotid artery, middle cerebral artery M1 segment, proximal middle cerebral artery M2 segment, and distal middle cerebral artery M2 segment with an accuracy of 97.0%, sensitivity of 94.3%, and specificity of 97.4%.

Conclusion: Deep learning tools can be effective in the detection of both ICH and LVO across a wide variety of hospital systems. While some limitations were identified, specifically in the detection of small ICH and distal M2 occlusion, this study highlights a deep learning tool that can assist radiologists in the detection of emergent findings in a variety of practice settings.

Keywords: deep learning, artificial intelligence, radiology, large vessel occlusion, neuroradiology, intracranial hemorrhage

INTRODUCTION

Timely diagnosis of acute cerebrovascular disease is critical to reduce patient mortality and morbidity. Two forms of stroke, intracranial hemorrhage (ICH) and ischemic stroke due to large vessel occlusion (LVO) are especially devastating. ICH 28-day mortality has been reported at 50.6% and 6-month mortality due to LVO at 26.2% (1, 2).

Prompt intervention of these entities is critical in achieving improved outcomes. For example, ICH hematoma expansion was significantly reduced with early blood pressure control (3). Regarding LVO, functional independence decreased with every hour delay to endovascular thrombectomy (4).

Deep learning, a subset of artificial intelligence, has recently emerged as a means to aid clinicians in the timely diagnosis of both ICH and LVO. Newly developed algorithms have demonstrated strong performance in the detection of each (5–12). However, limitations of most of these studies are that they are often performed at a single institution and have not been validated in different settings.

Given the potential for deep learning tools to aid physicians in the timely and accurate diagnosis of these emergencies, it is important to validate their uses across a variety of facilities. Prior studies examining the relationship between deep-learning based algorithms and imaging assessment have been limited by geographic bias introduced from their cohorts, with the majority of U.S. states lacking representation (13). The specific aim of this study is to validate a commercially available deep learning-based tool, CINA® v1.0 device (Avicenna.ai, La Ciotat, France) in the detection of both ICH and LVO from multiple hospital sites and vendors through a collaboration between the University of California, Irvine (UCI) and vRAD (Minneapolis, USA). In doing so, the goal was to evaluate the generalizability of this tool to eliminate possible geographic bias introduced in other similar studies.

MATERIALS AND METHODS

This was a retrospective study using anonymized data from UCI and vRAD. A waiver of consent was obtained from the local Institutional Review Board (IRB) at UCI for the UCI cases and the Western IRB for the vRAD cases. The CINA® v1.0 device (Avicenna.ai, La Ciotat, France) was used for standalone performance assessment in both the ICH and LVO validation studies. The statistics provided in this manuscript are derived from an external test set (the validation cohort) and are completely independent from a prior cohort used to train the CINA® v1.0 device. Specifically, the cohort used to train the tool was based off of 8,994 ICH cases acquired between November 2014–May 2018 and 566 LVO cases acquired between May 2018–November 2018. All of the training data was acquired from vRAD data only. No UCI data was used for the training cohort. Additionally, all vRAD cases used for the validation cohort were acquired in 2019 only.

Intracranial Hemorrhage

Patient Selection

A cohort of patients with suspected acute ICH on clinical grounds in whom non-contrast CT (NCCT) head studies had been performed from UCI and an American teleradiology service (vRAD) were assessed. In both UCI and vRAD cases, suspected acute ICH cases were identified with keywords such as "hemorrhage," "NCCT," and "head" in the clinical indication or Digital Imaging and Communications in Medicine (DICOM) header information of the NCCT studies. Only the initial scan obtained for ICH evaluation was assessed for patients in this validation cohort. vRAD cases were acquired in 2019 only, and UCI data from 2017 to 2019. Inclusion criteria for NCCT scans required a strict axial acquisition, 512 x 512 matrix, slice thickness of <5 mm, soft tissue reconstruction kernel, and kVp ranging between 100 and 160.

ICH cases were divided into intraparenchymal (IPH), intraventricular (IVH), subarachnoid (SAH), subdural (SDH), and epidural (EDH) subtypes. Multiple cases contained a combination of these subtypes and were categorized accordingly. Intracranial hemorrhages were further categorized into small ($<$5mL), medium (5–25 mL), and large ($>$25 mL) volumes. Positive cases for acute ICH ("ground truths") were assessed by two board-certified neuroradiologists, with consensus determined by a third board-certified neuroradiologist. The two board-certified neuroradiologists also determined ICH subtype and volume information.

Scanning Parameters

GE Healthcare, Philips, Siemens, and Canon (Formerly Toshiba) scanners were used among this cohort with 16, 7, 13, and 5 various scanner models, respectively. The number of detector rows (NDR) were divided into eight categories: $4 < \text{NDR} \leq 8$, $8 < \text{NDR} \leq 16$, $16 < \text{NDR} \leq 32$, $32 < \text{NDR} \leq 64$, 64

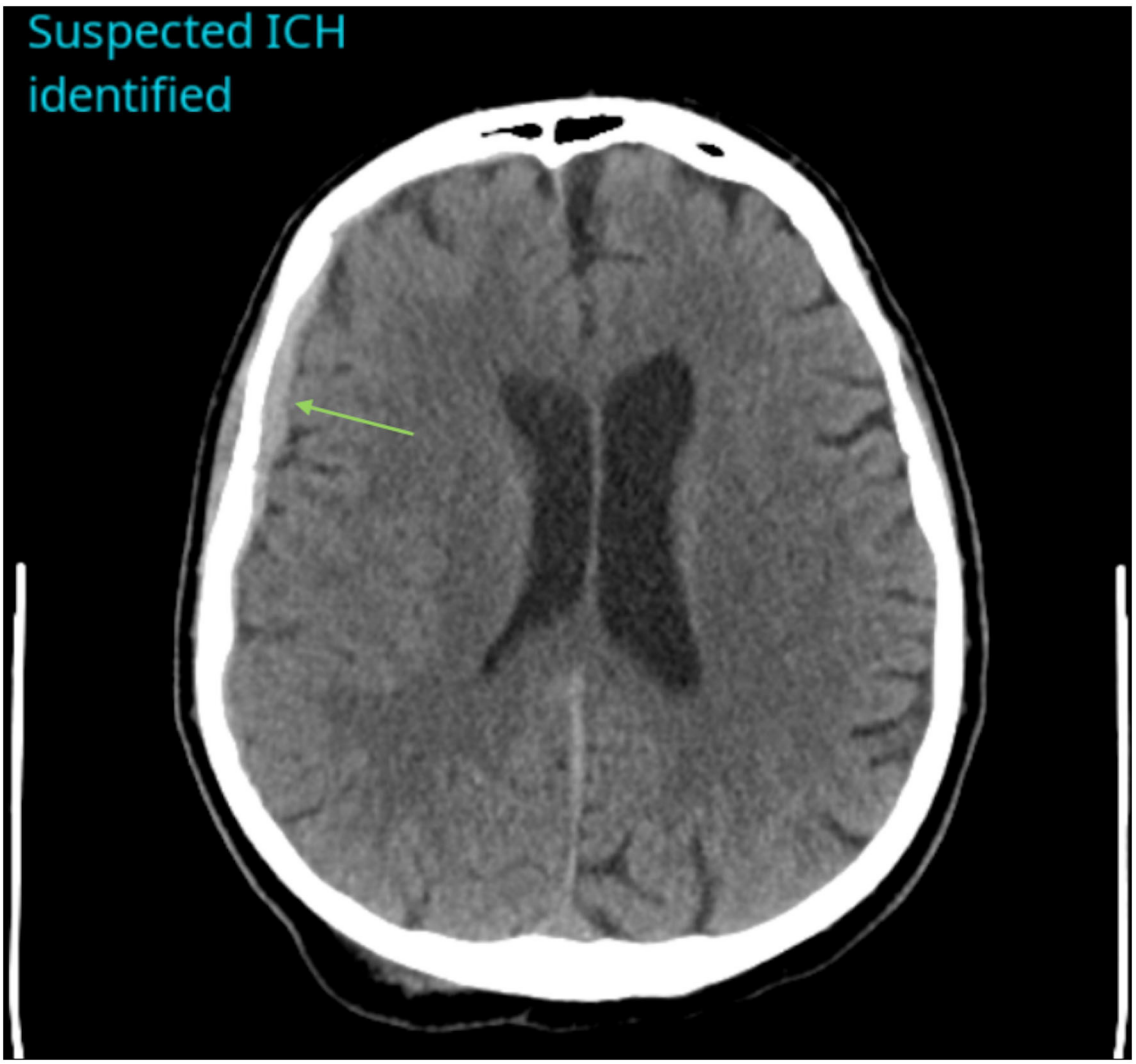


FIGURE 1 | Example of a positive ICH identified by CINA®. There is an acute subdural hemorrhage along the right cerebral convexity on non-contrast CT (green arrow).

$<$ NDR $\leq$ 128, 128 $<$ NDR $\leq$ 256, 256 $<$ NDR $\leq$ 320, and not available (NA) if this information was not attached to the case. Slice thickness (ST) was categorized as being $<$ 2.5 and 2.5 mm $\leq$ ST $\leq$ 5 mm. Radiation dose parameters were measured in kilovoltage peak (kVp) and milliampere-seconds (mAs). kVp was categorized as kVp $<$ 120, 120 $\leq$ kVp $\leq$ 140, and $>$140. mAs was categorized as $<$150, 150 $\leq$ mAs $\leq$ 400, and $>$400.

TABLE 1 | Performance metrics for overall cases of CINA-ICH application.

Statistical findings for ICH detection	Values [95% CI]
Sensitivity (%) [95% CI]	91.4% [87.2–94.5%]
Specificity (%) [95% CI]	97.5% [95.8–98.6%]

Large Vessel Occlusion

Patient Selection

A cohort of patients with suspected LVO on clinical grounds in whom CT angiography (CTA) head studies had been performed from UCI and vRAD were assessed. For both UCI and vRAD cases, suspected LVO cases were identified with keywords such as, "CTA," "head," and "large vessel occlusion" in the clinical indication or DICOM header information of the CTA studies. Only the initial scan obtained for LVO evaluation was assessed for patients in this validation cohort. vRAD cases were acquired in 2019 only and UCI cases from 2015 to 2019. Inclusion criteria for CTA scans included a strict axial acquisition, 512 x 512 matrix, slice thickness $\leq$1.25 mm, kVp to range between

80 and 140, arterial phase timing of contrast bolus confirmed by mini test bolus or automatic bolus tracking software, and arterial (or other sharp) reconstruction kernel. Positive cases for LVO ("ground truths") were assessed by two U.S. board-certified neuroradiologists, with consensus determined by a third board-certified neuroradiologist. Positive LVO cases were divided based on location into Distal Internal Carotid Artery (ICA), Middle Cerebral Artery (MCA)-M1, MCA-M2 Proximal, and MCA-M2 Distal.

Scanning Parameters

GE Medical Systems, Philips, Siemens, Canon (Formerly Toshiba), and NMS with 13, 4, 12, 4, and 1 various scanner models were included, respectively. The NDR were divided into the same eight categories as for the ICH cases. Slice thickness (ST) was $\leq$ 1.25 mm. Radiation dose parameters were measured in kilovoltage peak (kVp) and milliampere-seconds (mAs). kVp was categorized as $<$100, 100 $\leq$ kVp $\leq$ 120, and $>$120. mAs were categorized as $<$100, 100 $\leq$ mAs $\leq$ 400, and $>$400.

Statistical Analysis

Data was compared from the CINA® v1.0 device (Avicenna.ai, La Ciotat, France) to the ground truths determined by the board-certified neuroradiologists via a confusion matrix in order to obtain sensitivity, specificity, and accuracy. Positive predictive values (PPV) and negative predictive values (NPV) were computed with varying prevalence values (from 10 to 50%, increments of 5%). All of these statistics were performed using Excel and MedCalc version 19.7.2.

These statistics were performed for the total cases in both the ICH and LVO groups in addition to stratifications based on scanner models, NDR, slice thickness, radiation dose parameters, age, and sex, as well as ICH subtypes and volumes and LVO locations. The CINA® v1.0 device is not intended to discern ICH subtype or volume and only detects whether hemorrhage is present or not. Therefore, ICH subtypes and volume information were only assessed in positive cases by the two board-certified neuroradiologists. Only true positives and false negative values could be calculated and only true positive rate was provided for these classifications.

PASS sample size software was used to calculate the minimum number of cases needed to achieve a 95% CI lower bound of at least 80% assuming a point estimate of 90% (for sensitivity and specificity, separately). Using the binomial dichotomous endpoint for a one sample study, at least 137 positive and 137 negative anonymized cases were required (for ICH and LVO).

RESULTS

Intracranial Hemorrhage

Patient Selection

824 cases were selected for analysis from a pool of 400 retrospective anonymized cases from vRAD and 424 from UCI. 10 cases were excluded for the following reasons: 1 because slice thickness was not identical among the volume, 3 because the matrix was not 512 x 512, 1 because it contained a post-contrast series, 2 lacked a full field of view, 2 were uninterpretable due to significant motion artifact, and 1 was uninterpretable due to significant metal artifact. After exclusion, case distribution was 395 from vRAD and 419 from UCI for a total of 814 cases.

Overall Cases

ICH ground-truths were as follows: 204 positive ICH cases from vRAD and 51 from UCI. There were 191 negative ICH cases from vRAD and 368 from UCI. There was initial disagreement on 21 cases between the two neuroradiologists. However, a consensus was eventually reached for each of these cases.

The CINA® v1.0 algorithm identified 233 true positive ICH (**Figure 1**), 14 false positive ICH, 545 true negative ICH, and 22 false negative ICH. Accuracy was calculated as 95.6%. Sensitivity was 91.4 % [95% CI, 87.2–94.5%] and specificity was 97.5% [95% CI, 95.8–98.6%]. Performance metrics can be found in **Table 1**.

TABLE 3 | Performance metrics for ICH cases by geographic distribution in the United States.

Region category	Positive ICH	Negative ICH	Sensitivity (%)	Specificity (%)
Continental (n = 55)	29	26	93.1	100
Northeast (n = 187)	83	104	91.6	95.2
Pacific (n = 450)	67	383	89.5	97.6
Southeast (n = 96)	50	46	88	100
NA (n = 26)	26	0	–	–

NA, not available.

TABLE 2 | Performance metrics for ICH cases based on demographics.

Demographic		Positive ICH	Negative ICH	Sensitivity (%)	Specificity (%)
Age	18 $\leq$ Age $<$ 40 (n = 168)	37	131	89.2	100
	40 $\leq$ Age $\leq$ 70 (n = 316)	112	204	91.1	96.1
	Age $>$ 70 (n = 249)	98	151	92.9	97.4
	Age NA (n = 81)	8	73	–	–
Sex	Male (n = 206)	101	95	93.6	96.8
	Female (n = 188)	93	95	91.4	97.9
	NA (n = 421)	52	359	–	–

NA, Not Available.

Positive predictive values (PPV) and negative predictive values (NPV) were also assessed based on varying prevalence. PPV ranged from 80.2% (10% prevalence) to 97.3% (50% prevalence). NPV ranged from 99.0% (10% prevalence) to 91.9 % (50% prevalence).

Demographics

Performance metrics for age and sex can be found in **Table 2**. For age, sensitivity ranged from 89.2% ($18 \leq$ Age < 40, $n = 168$) to 92.9% (Age > 70; $n = 249$). Specificity ranged from 96.1% ($40 \leq$ Age ≤ 70, n=316) to 100% ($18 \leq$ Age < 40, n=168). For sex, sensitivity was 91.4% for females ($n = 188$) and 93.6% for males ($n = 206$). Specificity was 97.9% for females and 96.8% for males.

TABLE 4 | True positive rates for ICH cases based on subtype and volume.

ICH classification		True positive rate (%)
Subtype	IPH ($n = 99$)	92.9
	IVH ($n = 23$)	100
	EDH/SDH ($n = 122$)	94.3
	SAH ($n = 79$)	89.9
Volume	Small: <5 mL ($n = 78$)	71.8
	Medium: 5–25 mL ($n = 100$)	100
	Large: >25 mL ($n = 77$)	100

Site

The distribution, sensitivity, and specificity of the ICH tool based on geographic U.S. regions are shown in **Table 3**. Sensitivity ranged from 88% (Southeast U.S., $n = 96$) to 93.1% (Continental U.S., $n = 55$). Specificity ranged from 95.2% (Northeast U.S., n =187) to 100% (Continental and Southeast U.S.).

ICH Subtypes

ICH cases were additionally categorized based on subtypes: Intraparenchymal (IPH), Intraventricular (IVH), Subarachnoid (SAH), Subdural (SDH), and Epidural (EDH). The SDH and EDH subtypes were combined into one group for stratification purposes. In addition, some patients are represented across multiple categories (IPH, IVH, EDH/SDH, and SAH). Distribution is seen in **Table 4**. CINA® demonstrated a true positive rate of 92.9% for IPH, 100% for IVH, 94.3% for EDH/SDH, and 89.9% for SAH. ICH size distribution is also seen in **Table 4**. CINA® demonstrated a true positive rate of 71.8% for small (<5 mL) ICH, 100% for medium (5–25 mL), and 100% for large (>25 mL) ICH.

Scanner Models

Case distribution among the various scanner models are found in **Table 5**. Note that there were 41 different scanner models for the ICH data set with 16, 7, 13, and 5 different models for GE, Philips, Siemens, and Canon (Formerly Toshiba), respectively. Sensitivity for GE, Philips, Siemens, and Canon was 91.8, 84.4, 96.7, and 94.1%, respectively. Specificity was 98.1, 97.7, 90.0, and 100%, respectively.

TABLE 5 | Performance metrics and distribution for ICH cases based on different scanning parameters.

Scanning parameter		Positive ICH	Negative ICH	Sensitivity (%)	Specificity (%)
Scanner model	GE Healthcare ($n = 203$)	97	106	91.8	98.1
	Philips ($n = 461$)	64	397	84.4	97.7
	Siemens (n =90)	60	30	96.7	90
	Canon (Formerly Toshiba) ($n = 60$)	34	26	94.1	100
Detector rows	$4 <$ NDR ≤ 8 ($n = 2$)	2	0	100	–
	$8 <$ NDR ≤ 16 ($n = 46$)	14	32	78.6	100
	$16 <$ NDR ≤ 32 ($n = 185$)	91	94	94.5	96.8
	$32 <$ NDR ≤ 64 ($n = 518$)	111	407	91	97.5
	$64 <$ NDR ≤ 128 ($n = 6$)	5	1	100	100
	$128 <$ NDR ≤ 256 ($n = 12$)	8	4	87.5	100
	$256 <$ NDR ≤ 320 ($n = 14$)	12	2	100	100
	NA ($n = 31$)	12	19	–	–
Slice Thickness	ST < 2.5 mm ($n = 39$)	23	16	100	100
	$2.5 \leq$ ST ≤ 5 mm ($n = 775$)	232	543	90.5	97.4
kVp	kVp < 120 ($n = 8$)	6	2	100	100
	$120 \leq$ kVp ≤ 140 ($n = 806$)	249	557	91.2	97.5
	kVp >140 ($n = 0$)	0	0	–	–
mAs	mAs < 150 ($n = 23$)	12	11	100	90.9
	$150 \leq$ mAs ≤ 400 ($n = 765$)	231	534	90.9	97.8
	mAs > 400 ($n = 26$)	12	14	91.7	92.9

NA, not available.

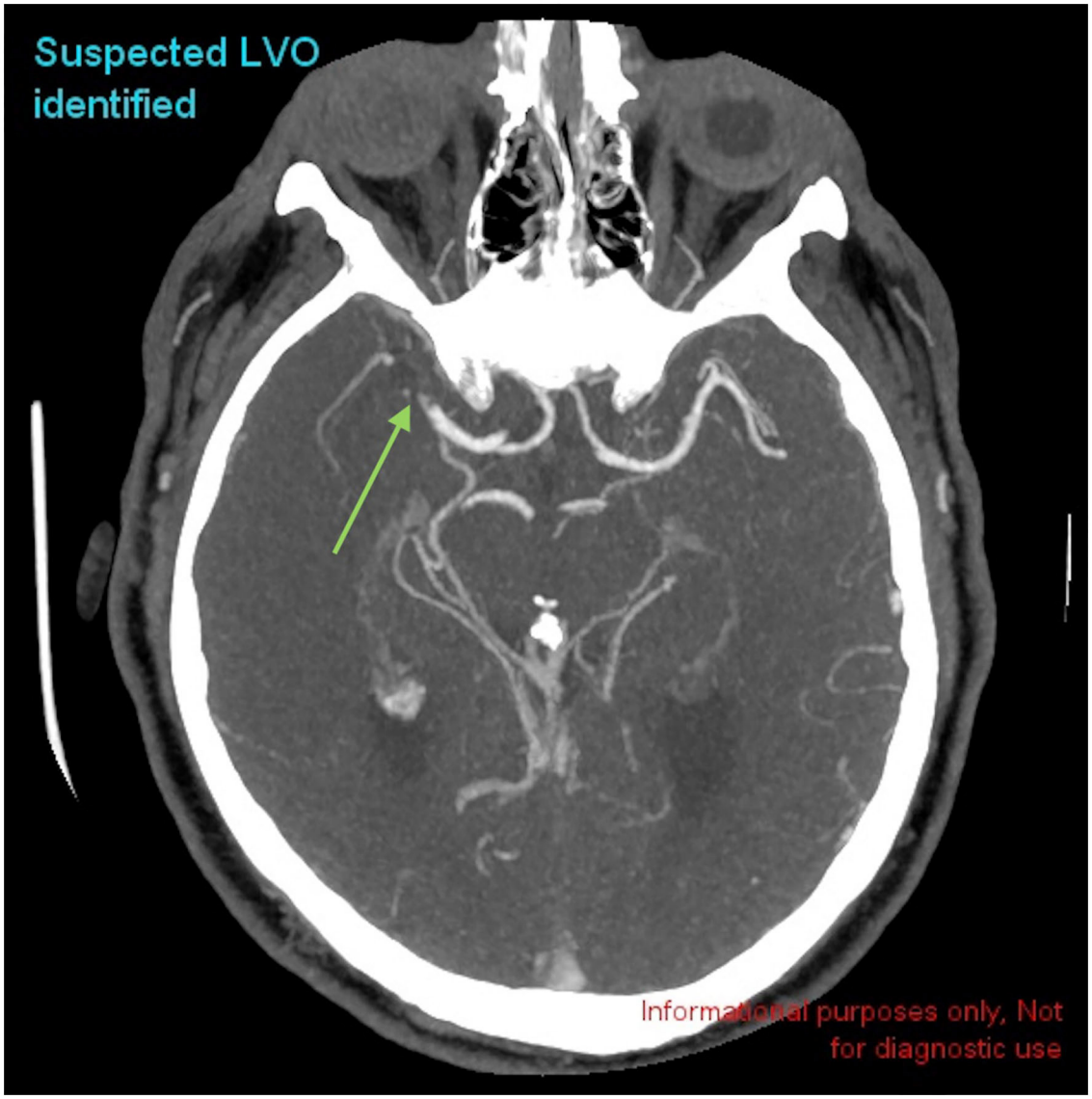


FIGURE 2 | Example of a positive LVO identified by CINA®. There is a large vessel occlusion in the distal right MCA-M1 branch on CTA (green arrow).

Scanning Parameters

Case distribution and performance metrics among the various scanning parameters can be found in **Table 5**. The number of detector rows' sensitivity ranged from 78.6% ($8 < \text{NDR} \leq 16$, $n = 46$) to 100% ($4 < \text{NDR} \leq 8$, $n = 2$; $64 < \text{NDR} \leq 128$, $n = 6$; $256 < \text{NDR} \leq 320$, $n = 14$). Specificity ranged from 96.8% ($16 < \text{NDR} \leq 32$, $n = 185$) to 100% ($8 < \text{NDR} \leq 16$; $64 < \text{NDR} \leq 128$; $128 < \text{NDR} \leq 256$, $n = 12$; $256 < \text{NDR} \leq 320$). Slice thickness sensitivity and specificity was 100% when slice thickness was <2.5 mm. Sensitivity was 90.5% and specificity was 97.4% when slice thickness was $2.5 \leq \text{ST} \leq 5$ mm. Sensitivity and specificity for kilovoltage peaks was 100% when kVp was <120. Sensitivity was 91.2% and specificity was 97.5% when kVp was $120 \leq \text{kVp} \leq 140$. Sensitivity for milliampere-seconds was 100% and specificity was 90.9% when mAs < 150. Sensitivity was 90.9% and specificity was 97.8% when $150 \leq \text{mAs} \leq 400$. Sensitivity was 91.7% and specificity was 92.9% when mAs was >400.

TABLE 6 | Performance metrics for overall cases of CINA-LVO application.

Statistical findings for LVO detection	Values [95% CI]
Sensitivity (%) [95% CI]	98.1% [94–99.5%]
Specificity (%) [95% CI]	98.2% [95.1–99.4%]

TABLE 7 | Performance metrics for LVO cases among different demographic parameters.

Demographic		Positive LVO	Negative LVO	Sensitivity (%)	Specificity (%)
Age	$18 \leq$ Age < 40 ($n = 26$)	5	21	83.3	100
	$40 \leq$ Age ≤ 70 ($n = 176$)	65	111	100	97.3
	Age > 70 ($n = 176$)	85	91	97.7	98.9
Sex	Male ($n = 185$)	74	111	98.7	98.2
	Female ($n = 186$)	80	106	97.5	98.1
	NA ($n = 7$)	2	5	–	–

NA, not available.

Large Vessel Occlusion

Patient Selection

406 anonymized CT angiography (CTA) cases were assessed; 93 from UCI and 313 from vRAD. 28 of these were excluded for the following reasons: 11 were not CTAs, 2 had no contrast, 7 did not have enough contrast, 2 were uninterpretable due to significant motion artifact, 2 were uninterpretable due to significant metal artifact, 3 did not have a full field of view, and 1 had an acquisition issue (z-spacing variability).

Overall Cases

LVO ground-truths (determined by two board-certified neuroradiologists) were as follows: 156 positive LVO cases and 222 negative LVO cases. There was initial disagreement on 19 cases between the two neuroradiologists. However, a consensus was eventually reached for each of these cases.

The CINA® v1.0 algorithm identified 153 true positive LVO (**Figure 2**), 4 false positive LVO, 218 true negative LVO, and 3 false negative LVO. Sensitivity was 98.1 % [95% CI, 94.0–99.5%] and specificity was 98.2% [95% CI, 95.1–99.4%]. Performance metrics can be found in **Table 6**.

Positive predictive values (PPV) and negative predictive values (NPV) were also assessed based on varying prevalence (from 10 to 50%, increments of 5%). PPV ranged from 85.8% (10% prevalence) to 98.2% (50% prevalence). NPV ranged from 99.8% (10% prevalence) to 98.1% (50% prevalence).

Demographics

Performance metrics for patient demographics can be found in **Table 7**. For age, sensitivity ranged from 83.3% ($18 \leq$ Age < 40, $n = 26$) to 100% ($40 \leq$ Age ≤ 70, $n = 176$). Specificity ranged from 97.3% ($40 \leq$ Age ≤ 70) to 100% ($18 \leq$ Age < 40). For sex, sensitivity was 97.5% for females ($n = 186$) and 98.7% for males ($n = 185$). Specificity was 98.1% for females and 98.2% for males.

Site

The distribution, sensitivity, and specificity of the LVO tool based on geographic U.S. regions are shown in **Table 8**. Sensitivity ranged from 97.4% (Southeast U.S., $n = 75$) to 100% (Continental U.S., $n = 27$). Specificity ranged from 97.3% (Southeast U.S.) to 100% (Continental U.S.).

LVO Subtypes

A subset of 55 patients were randomly selected to evaluate performance metrics of the tool in evaluating LVO subtypes (4 Distal ICA, 26 MCA-M1, 20 Proximal MCA-M2, and 3 Distal MCA-M2). Accuracy was 97.0%, sensitivity 94.3% [95% CI, 83.4–98.5%], and specificity 97.4% [95% CI, 95.1–98.7%].

TABLE 8 | Performance metrics for LVO cases by geographic distributions in the United States.

Region category	Positive LVO	Negative LVO	Sensitivity (%)	Specificity (%)
Continental ($n = 27$)	8	19	100	100
Northeast ($n = 155$)	50	105	98	98.1
Pacific ($n = 120$)	59	61	98.3	98.4
Southeast ($n = 75$)	38	37	97.4	97.3
NA ($n = 1$)	1	0	–	–

NA, not available.

Scanner Models

Case distribution among the various scanner models can be found in **Table 9**. Sensitivity for Siemens, Canon (formerly Toshiba), GE Medical Systems and Philips was 96.7, 94.1, 91.8, and 84.4%, respectively. Specificity for GE Healthcare, Philips, Siemens, Canon, and NMS was 90.0, 100, 98.1, 97.7, and 100%, respectively.

Scanning Parameters

Case distribution and performance metrics among the various scanning parameters can be found in **Table 9**. Sensitivity for the number of detector rows ranged from 96.2% ($32 <$ NDR ≤ 64, $n = 146$) to 100% ($8 <$ NDR ≤ 16, $n = 15$; $16 <$ NDR ≤ 32, $n = 63$; $256 <$ NDR ≤ 320, $n = 5$). Specificity ranged from 90.0% ($8 <$ NDR ≤ 16) to 100% ($16 <$ NDR ≤ 32; $64 <$ NDR ≤ 128, $n = 126$; $256 <$ NDR ≤ 320). Sensitivity and specificity for slice thickness $\leq$1.25 mm was 98.1 and 98.2%, respectively. Sensitivity for kilovoltage peak ranged from 98% ($100 \leq$ kVp ≤ 120, $n = 359$) to 100% (kVp < 100, $n = 8$; kVp > 120, $n = 11$). Specificity ranged from 83.3% (kVp > 120) to 100% (kVp < 100). Sensitivity for milliampere-seconds ranged from 97.8% ($100 \leq$ mAs ≤ 400, $n =$314) to 100% (mAs < 100, n=43; mAs $>$400, $n = 21$). Specificity ranged from 97.7% ($100 \leq$ mAs ≤ 400) to 100% (mAs < 100; mAs > 400).

TABLE 9 | Performance metrics for LVO cases based on different scanning parameters.

Scanning parameter		Positive LVO	Negative LVO	Sensitivity (%)	Specificity (%)
Scanner model	GE healthcare ($n = 129$)	50	79	91.8	98.1
	Philips ($n = 137$)	62	75	84.4	97.7
	Siemens ($n = 73$)	30	43	96.7	90.0
	Canon (Formerly Toshiba) ($n = 37$)	14	23	94.1	100
	NMS ($n = 2$)	0	2	–	100
Detector rows	$4 < NDR \leq 8$ ($n = 0$)	0	0	–	–
	$8 < NDR \leq 16$ ($n = 15$)	5	10	100	90
	$16 < NDR \leq 32$ ($n = 63$)	27	36	100	100
	$32 < NDR \leq 64$ ($n = 146$)	52	94	96.2	96.8
	$64 < NDR \leq 128$ ($n = 126$)	65	61	98.5	100
	$128 < NDR \leq 256$ ($n = 0$)	0	0	–	–
	$256 < NDR \leq 320$ ($n = 5$)	1	4	100	100
	NA (n=23)	6	17	–	–
Slice thickness	ST $\leq$ 1.25 mm ($n = 378$)	156	222	98.1	98.2
kVp	kVp < 100 ($n = 8$)	1	7	100	100
	$100 \leq$ kVp ≤ 120 ($n = 359$)	150	209	98	98.6
	kVp > 120 ($n = 11$)	5	6	100	83.3
mAs	mAs < 100 ($n = 43$)	11	32	100	100
	$100 \leq$ mAs ≤ 400 ($n = 314$)	138	176	97.8	97.7
	mAs > 400 ($n = 21$)	7	14	100	100

NA, not available.

DISCUSSION

This retrospective, multicenter study aimed to demonstrate the generalizability of a commercially available deep-learning based tool, CINA® v1.0, in the detection of ICH and LVO across multiple hospital settings. The algorithm performed well in the ICH cohort, with an overall accuracy of 95.6%, sensitivity of 91.4%, and specificity of 97.5%. Of the ICH subtypes, it achieved the highest sensitivity in the detection of intraventricular hemorrhage with a true positive rate of 100%, followed by epidural/subdural, intraparenchymal, and subarachnoid subtypes which all had sensitivity of at least 90%. When stratified by ICH size, it performed best for medium and large volumes with sensitivities of 100%, but demonstrated lower sensitivity in the detection of small volumes with a sensitivity of 71.8%.

The tool also performed well in the LVO cohort, with an accuracy of 98.1%, sensitivity of 98.1%, and specificity of 98.2%. The algorithm showed robust performance in detecting LVO location in a smaller subset of cases with an accuracy of 97.0%, sensitivity of 94.3%, and specificity of 97.4%.

These results corroborate previous studies analyzing the ability for deep-learning tools to detect intracranial emergencies. For example, Chilamkurthy et al. used a deep-learning algorithm to detect and classify ICH on large and diverse cohorts in India (6). Another study obtained an AUC of 0.99 in the detection of ICH via deep-learning algorithms; however, this was based off of a single institution using relatively uniform scanning parameters on two scanner models (9). Similar studies have been performed with respect to AI detection of LVO. For example, a commercially available deep learning software for LVO detection achieved an AUC of 0.86 using a cohort derived from three tertiary stroke centers (11). Our work expands on these previous studies by showing similar robust performance of deep learning tools across a diverse population regardless of scanner parameters and geographic distribution.

Ultimately, given the robust nature of deep learning tools such as CINA® v1.0, the goal of these tools is to streamline the radiologists' workflow by triaging studies to alert physicians to the most time-sensitive findings, and to act as a second set of eyes when studies are more ambiguous. Studies evaluating the effectiveness of such systems have already begun. For example, when a deep-learning tool was prospectively integrated to prioritize studies in a radiologist's workflow based on the presence of ICH, one study found that time to diagnosis was significantly reduced (14). Future studies with CINA® v1.0 could mirror this type of work and evaluate patient outcomes as influenced by the integration of deep-learning tools into a radiologist's workflow. For example, both inpatient and outpatient settings could be evaluated with regards to these neurologic emergencies and how these tools impact efficiency and ultimately clinical outcomes.

Our software had some limitations that warrant further investigation. Perhaps the greatest limitation was in the detection of small bleeds, with false negatives occurring predominantly in very small ICH (<1.5 mL). The false negatives that occurred in larger bleeds (1.5–5 mL), were often located within chronic

pathology such as old hematomas, areas of gliosis, or extra-parenchymal structures such as along the falx cerebri. On the other hand, ICH false positives predominantly occurred in the setting of significant streak or motion artifact. Similar limitations were identified in the LVO cohort with respect to imaging artifact and small size. For example, the LVO false negatives all occurred in the setting of small occlusions <1.3 mm in length (Note that LVO lengths were only retrospectively measured for the three false negative cases in order to understand why the application failed to detect them and were not measured for the remaining LVO cases). A false positive LVO case also occurred in the setting of significant streak artifact. Another false positive case misdiagnosed an area of stenosis as a complete occlusion. Two false positive cases misidentified the sphenoparietal sinus venous structure as an area of occlusion, likely secondary to its close proximity to the MCA. As a result, caution should be used when relying solely on the software in these settings. However, the limitations discussed above often occurred in settings that would likely pose similar challenges to radiologists and result in a similar distribution of false negatives and positives.

CINA® v1.0 was only trained to identify acute blood based off of hyperdense components. Thus, chronic hemorrhages cannot be identified by the algorithm unless they contain more acute hyperdense components. However, given that low density hemorrhages (e.g., chronic SDH) are often not emergencies, we believe this distinction is actually clinically useful and not necessarily a limitation in order to prevent the tool from flooding the radiologist with alerts for non-emergent cases. The tool did not differentiate between acute and non-acute LVO etiologies such as chronic ICA occlusion, and these may have been included as positive LVO cases. Lastly, the tool was not trained to evaluate occlusions in the anterior cerebral arteries or posterior circulation. While further work is needed for future tools to better identify more distal occlusions and subtle hemorrhages, the primary goal of CINA in its current state is to identify obvious findings that need to be assessed urgently for emergent triage and prioritization of the worklist. This is reflected in a separate standalone effectiveness assessment demonstrating a mean "time-to-notification" of 21.6 and 34.7 s for ICH and LVO detection, respectively.

Despite these limitations, CINA® v1.0 demonstrates robust generalizability in the detection of ICH and LVO. For example, in a study examining the geographic distribution of various cohorts evaluated by deep learning based algorithms in various medical specialties, 34 states were not represented among 56 studies (13). Our ICH data spans 44 states and 204 U.S. cities, while LVO data reflects scans from 40 states and 158 U.S. cities. To our knowledge, this is the most heterogeneous population cohort ever studied in the U.S. using a deep learning tool for ICH and LVO detection. This study demonstrates the potential for greater application of deep-learning tools across a wide variety of clinical settings.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of California, Irvine Institutional Review Board, and Western Institutional Review Board. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

JM and SE: manuscript preparation, data gathering, analysis. YC: manuscript revision, data gathering and analysis. SQ: data gathering and analysis. PC: data analysis, software programming. DC: data analysis. JS: manuscript revision, study design, data gathering and analysis. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Fogelholm R, Murros K, Rissanen A, Avikainen, S. Long term survival after primary intracerebral haemorrhage: a retrospective population based study. *J Neurol Neurosurg Psychiatry.* (2005) 76:1534–8. doi: 10.1136/jnnp.2004.055145
2. Malhotra K, Gornbein J, Saver JL. Ischemic strokes due to large-vessel occlusions contribute disproportionately to stroke-related dependence and death: a review. *Front Neurol.* (2017) 8:651. doi: 10.3389/fneur.2017.00651
3. Arima H, Anderson CS, Wang JG, Huang Y, Heeley E, Neal B, et al. Lower treatment blood pressure is associated with greatest reduction in hematoma growth after acute intracerebral hemorrhage. *Hypertension.* (2010) 56:852–8. doi: 10.1161/HYPERTENSIONAHA.110.154328
4. Saver JL, Goyal M, van der Lugt A, Menon BK, Majoie CB, Dippel DW, et al. Time to treatment with endovascular thrombectomy and outcomes from ischemic stroke: a meta-analysis. *JAMA.* (2016) 316:1279–88. doi: 10.1001/jama.2016.13647
5. Ginat DT. Analysis of head CT scans flagged by deep learning software for acute intracranial hemorrhage. *Neuroradiology.* (2020) 62:335–40. doi: 10.1007/s00234-019-02330-w
6. Chilamkurthy S, Ghosh R, Tanamala S, Biviji M, Campeau NG, Venugopal VK, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet.* (2018) 392:2388–96. doi: 10.1016/S0140-6736(18)31645-3
7. Chang PD, Kuoy E, Grinband J, Weinberg BD, Thompson M, Homo R, et al. Hybrid 3D/2D convolutional neural network for hemorrhage evaluation on head CT. *AJNR Am J Neuroradiol.* (2018) 39:1609–16. doi: 10.3174/ajnr.A5742
8. Lee H, Yune S, Mansouri M, Kim M, Tajmir SH, Guerrier CE, et al. An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. *Nat Biomed Eng.* (2019) 3:173–82. doi: 10.1038/s41551-018-0324-9
9. Kuo W, Häne C, Mukherjee P, Malik J, Yuh EL. Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning. *Proc Natl Acad Sci USA.* (2019) 116:22737–45. doi: 10.1073/pnas.1908021116
10. Chatterjee A, Somayaji Nayana R, Kabakis Ismail M. Abstract WMP16: artificial intelligence detection of cerebrovascular large vessel

occlusion - nine month, 650 patient evaluation of the diagnostic accuracy and performance of the viz.ai LVO algorithm. *Stroke.* (2018) 5**0**:AWMP16. doi: 10.1161/str.50.suppl_1.WMP16

11. Barreira C, Bouslama M, Lim J, Al-Bayati A, Saleem Y, Devlin T, et al. E-108 Aladin study: automated large artery occlusion detection in stroke imaging study – a multicenter analysis. *J Neurointervent Surg.* (2018) 10:A101–2. doi: 10.1136/neurintsurg-2018-SNIS.184
12. Yahav-Dovrat A, Saban M, Merhav G, Lankri I, Abergel E, Eran A, et al. Evaluation of artificial intelligence–powered identification of large-vessel occlusions in a comprehensive stroke center. *Am J Neuroradiol.* (2020) 42:247–54. doi: 10.3174/ajnr.A6923
13. Kaushal A, Altman R, Langlotz C. Geographic distribution of US cohorts used to train deep learning algorithms. *JAMA.* (2020) 324:1212–3. doi: 10.1001/jama.2020.12067
14. Arbabshirani MR, Fornwalt BK, Mongelluzzo GJ, Suever JD, Geise BD, Patel AA, et al. Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. *NPJ Digit Med.* (2018) 1:9. doi: 10.1038/s41746-017-0015-z

Conflict of Interest: YC and SQ are employees of Avicenna.ai. PC is co-founder of and owns stock in Avicenna.ai, has past and current research funding from and is a paid consultant for Canon Medical, has current research funding from GE, and is a paid consultant and speaker for Siemens. DC has a grant with Avicenna.ai (money paid to institution), has done consultancy for Canon Medical, has done expert testimony for Cullins & Grandy, has grants/grants pending with Canon Medical and Novocure, and has stock/stock options with Avicenna.ai.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neurology

BRIEF RESEARCH REPORT
published: 06 May 2021
doi: 10.3389/fneur.2021.663899

Prediction of Clinical Outcomes in Acute Ischaemic Stroke Patients: A Comparative Study

Deepthi Rajashekar[1,2,3]*, *Michael D. Hill*[2,3,4,5,6], *Andrew M. Demchuk*[2,5], *Mayank Goyal*[2,5], *Jens Fiehler*[7] and *Nils D. Forkert*[2,3,4,8]

[1] Biomedical Engineering Graduate Program, University of Calgary, Calgary, AB, Canada, [2] Depertment of Radiology, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada, [3] Hotchkiss Brain Institute, University of Calgary, Calgary, AB, Canada, [4] Department of Clinical Neurosciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada, [5] Department of Medicine, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada, [6] Department of Community Health Sciences, University of Calgary, Calgary, AB, Canada, [7] Department of Diagnostic and Interventional Neuroradiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, [8] Alberta Children's Hospital Research Institute, University of Calgary, Calgary, AB, Canada

OPEN ACCESS

Edited by:
Vida Abedi,
Geisinger Health System,
United States

Reviewed by:
Mohammad Adibuzzaman,
Purdue University, United States
Ghasem Farahmand,
Tehran University of Medical Sciences, Iran
Durgesh Prasad Chaudhary,
Geisinger Health System,
United States

***Correspondence:**
Deepthi Rajashekar
deepthi.rajasheka1@ucalgary.ca

Specialty section:
This article was submitted to Stroke, a section of the journal Frontiers in Neurology

Received: *03 February 2021*
Accepted: *09 April 2021*
Published: *06 May 2021*

Citation:
Rajashekar D, Hill MD, Demchuk AM, Goyal M, Fiehler J and Forkert ND (2021) Prediction of Clinical Outcomes in Acute Ischaemic Stroke Patients: A Comparative Study. Front. Neurol. 12:663899. doi: 10.3389/fneur.2021.663899

Background: Clinical stroke rehabilitation decision making relies on multi-modal data, including imaging and other clinical assessments. However, most previously described methods for predicting long-term stroke outcomes do not make use of the full multi-modal data available. The aim of this work was to develop and evaluate the benefit of nested regression models that utilise clinical assessments as well as image-based biomarkers to model 30-day NIHSS.

Method: 221 subjects were pooled from two prospective trials with follow-up MRI or CT scans, and NIHSS assessed at baseline, as well as 48-hours and 30 days after symptom onset. Three prediction models for 30-day NIHSS were developed using a support vector regression model: one clinical model based on modifiable and non-modifiable risk factors ($M_{CLINICAL}$) and two nested regression models that aggregate clinical and image-based features that differed with respect to the method used for selection of important brain regions for the modelling task. The first model used the widely accepted RreliefF (M_{RELIEF}) machine learning method for this purpose, while the second model employed a lesion-symptom mapping technique (M_{LSM}) often used in neuroscience to investigate structure-function relationships and identify eloquent regions in the brain.

Results: The two nested models achieved a similar performance while considerably outperforming the clinical model. However, M_{RELIEF} required fewer brain regions and achieved a lower mean absolute error than M_{LSM} while being less computationally expensive.

Conclusion: Aggregating clinical and imaging information leads to considerably better outcome prediction models. While lesion-symptom mapping is a useful tool to investigate structure-function relationships of the brain, it does not lead to better outcome predictions compared to a simple data-driven feature selection approach, which is less computationally expensive and easier to implement.

Keywords: support vector machine, lesion symptom mapping, NIHSS (National Institue of Health Stroke Scale), nested regression, ischemic stroke

INTRODUCTION

The prognosis of clinical and functional outcome in acute ischemic stroke patients is typically made based on multi-modal information such as demographic, clinical, laboratory, and radiological data. Theoretically, machine learning models can identify patterns in high-dimensional data that can be used to make data-driven and reproducible stroke outcome predictions in new patients and support patient management. However, despite the ability to integrate multimodal information, recent machine learning models have mostly utilized clinical data or image-based biomarkers alone (1) to predict stroke outcome. So far, the benefit of using true multi-modal data for stroke outcome prediction has not been investigated comprehensively. One of the few multi-modal predictive models of stroke outcome is described by Brugnara et al. (2). However, clinical assessments at various timepoints are used as input features without addressing the issue of feature collinearity. Furthermore, previous studies often predict the stroke outcome in a binary classification scheme (good vs. bad), which ignores the incremental, yet relevant non-linear differences in stroke severity scores.

Integration of image-based biomarkers for stroke outcome prediction is more complex than using other clinical assessments (in most cases), but has the potential to add considerable predictive power. A key aspect to consider within this context is the selection of regions-of-interest (ROIs) in the brain that are critically associated with the clinical deficit of interest since non-informative and redundant feature can downgrade the prediction accuracy considerably (3). Lesion-symptom mapping (LSM) (4) is able to identify brain regions that are important for a clinical outcome score of interest but has been used rarely for selection of brain regions for stroke outcome prediction (5). The more common ROI selection approach is to use classical feature selection methods during the training process. However, these two general approaches have never been compared to date with respect to stroke outcome prediction.

The aim of this work is to compare different setups of nested machine learning models using clinical information only and a combination of clinical and radiological features selected using lesion-symptom mapping and classical feature selection methods to predict the 30-days NIH stroke scale (NIHSS).

METHODS

Data

The datasets used in this study were pooled from the ESCAPE (6) and iKNOW (7) trials. Patients with remote hemorrhages, bilateral lesions, and severe white matter hyperintensities were excluded from this secondary analysis, and only patients with a follow-up MRI or CT scan (18-hours to one week from baseline) with complete clinical information (obtained after stroke and upto 6-hours post randomization) were included, leading to a final sample of 221 patients. The clinical outcome of interest in this study is the NIHSS assessed at 30 days after stroke symptom onset. The patient characteristics are summarized in **Table 1**. The measurable clinical and laboratory features used in the nested regression model include age, sex, modifiable and non-modifiable risk factors suggested in the evidence-based review of stroke rehabilitation (8). These include blood pressure, glucose, hematocrit, hypertension, diabetes, smoking status, hyperlipidemia, and atrial fibrillation (see **Supplementary Table 1**). Additionally, the baseline NIHSS score (pre-treatment) was also included as part of the clinical data to model stroke outcome (5, 9).

TABLE 1 | Characteristics of patients pooled ($N = 221$) from the ESCAPE[6] and iKNOW[7] datasets.

Variable	ESCAPE ($N = 143$)	iKNOW ($N = 78$)	Dataset ($N = 221$)
Median Age (IQR)	68 (19.5)	70.5 (15)	69 (19)
Sex—Females	75	31	106
Treatment—Alteplase	66	46	112
Median Onset to randomization time (IQR)	160 min (149)	126.5 min (118.05)	152 min (137)
Median Baseline NIHSS (IQR)	16 (7)	12 (10)	15 (8)

All lesions were manually delineated by an expert observer using the ITKSNAP tool. Each image sequence was skull stripped and non-linearly registered to the common FLAIR-NCCT (10) atlas of the elderly using the ANTs toolkit. The grey matter (GM) and white matter (WM) parcellations from the probabilistic BNA atlas (11) and the JHU atlas (12), respectively, were fused and transformed to the FLAIR-NCCT atlas. All image-based features were computed in the FLAIR-NCCT atlas space.

Model Design

Nested regression models were developed to predict the 30-days NIHSS outcome based on clinical data and image-based biomarkers. Here, the first model predicts the 48-hours NIHSS using imaging features alone whereas the result of this model is then used together with clinical features to predict the 30-days NIHSS.

$$NIHSS_{30-days} \sim (Features_{Clinical} + (NIHSS_{48-hours} \sim Features_{Imaging}))$$

For both models, epsilon-regression was used implemented using in a radial kernel support vector regression (SVR) framework. Using follow-up imaging acquired between 18-hours and 5-days from symptom onset to identify regions-of-interest (ROIs) that maximally correlate with a long-term assessment might introduce confounding effects and bias the results. Therefore, the ROIs included in the predictive models were identified with respect to the 48-hours NIHSS to ensure that the identified structure-function relationships are related to the primary stroke-induced deficits alone. This also ensures that the identified ROIs are not selected because of post-secondary comorbidities (not directly related to the primary stroke) developed either in-hospital or post-discharge. The two approaches for ROI selection are: (i) the LSM method using Brunner-Munzel test (13) and (ii) a widely accepted machine learning-based feature selection method that accounts for collinearity known as RreliefF (14).

The LSM method was implemented using the LESYMAP package (15) using the default parameters employing a p-value threshold at 0.05, discarding voxels not injured in at least 10% of the sample data, and using false discovery rate (FDR, the rate of Type 1 errors) to correct for multiple comparisons. For ease of comparison, brain regions that were not affected in at least 10% of the sample data were also removed prior to the RreliefF feature selection. The RreliefF feature selector was also employed using default parameters from the Fselector package (16) with the sample size of 10 and a neighbor count set to five. The result of the LSM is a statistical map of clusters of significant voxels that survive the FDR correction with non-zero voxel weights. Regions in the BNA-JHU parcellation that were assigned non-zero voxel weights by the LSM analysis were included as ROIs in the proposed regression analyses.

For each brain region identified by LSM as being important in the training set, the relative lesion overlap was computed and used as image-based features. Moreover, in case of WM tracts, the cross-sectional width of the tract spared after the lesion was also calculated and used as additional features (17). Therefore, the final set of image-derived input features used in this study include GM overlap, WM overlap, and WM tract integrity for all the selected ROIs.

For RreliefF feature selection, the lesion overlap (GM and WM) and tract integrity (only WM) was calculated for each atlas region and used for feature selection based on the training set.

Model Evaluation

For the sake of being able to compare brain regions selected for stroke outcome prediction qualitatively between the two models, the data was randomly split into completely independent training and test sets. This resulted in only one set of features selected for each method, which greatly enhances the interpretability and comparison of the models. Therefore, the entire dataset was partitioned into two mutually exclusive subsets for model training (80%) and testing (20%) using a stratified split that preserves the representation of stroke severity across both groups. Three models were evaluated in this framework: (i) un-nested SVR model with clinical features alone ($M_{CLINICAL}$) selected using RreliefF; (ii) nested model using clinical and imaging data with RreliefF as feature selector (M_{RELIEF}); and (iii) nested model using clinical and image data with LSM as feature selector (M_{LSM}). The resulting models were compared for predictive performance with respect to the model's mean absolute error (MAE) and coefficient of determination (R^2).

RESULTS

The overlap of all individual patient lesions in the atlas space shows that maximum incidence of stroke in this dataset occurs in the brain regions supplied by the middle cerebral artery (see **Supplementary Figure 1**). The median recovery profile of patients in this database is shown in **Supplementary Figure 2**.

The model using clinical features only resulted in a rather poor predictive performance ($R^2 = 0.13$). The optimal prediction results were achieved using age, baseline NIHSS, blood glucose and hematocrit levels, sex, presence of atrial fibrillation, hypertension, and hyperlipidemia, treatment decision (endovascular thrombectomy or tissue plasminogen activator), symptom onset to admission time, and blood pressure as features. However, the iterative feature selection procedure using RreliefF did not select presence of diabetes and smoking status, which are usually considered important predictors. Only the clinical features selected in this model were included in the two nested models to enable a direct comparison.

Compared to the simple predictive model using clinical features only, the two nested (M_{RELIEF} and M_{LSM}) models performed better and resulted in comparable R^2 and MAEs (see **Table 2**). No statistically significant MAE differences ($p > 0.05$) were found comparing the two nested models. However, M_{RELIEF} used only 44 ROIs in comparison to the 106 ROIs selected in M_{LSM} (see **Figure 1**). The plots of the predicted and ground truth scores for both models are shown in **Supplementary Figure 3**.

TABLE 2 | Model performances for each setup.

Model	MAE	RMSE	R^2	*p*-value
$M_{CLINICAL}$	4.33	5.53	0.13	0.0184
M_{RELIEF}	3.55	4.34	0.43	1.89e-06
M_{LSM}	3.50	4.54	0.40	6.22e-06

MAE, mean absolute error; RMSE, Root mean squared error; R^2, coefficient of determination; $M_{CLINICAL}$, model with clinical features alone; M_{RELIEF}, ROIs selected by RreliefF and nested with clinical features; M_{LSM}, ROIs selected using lesion-symptom mapping and nested with clinical features.

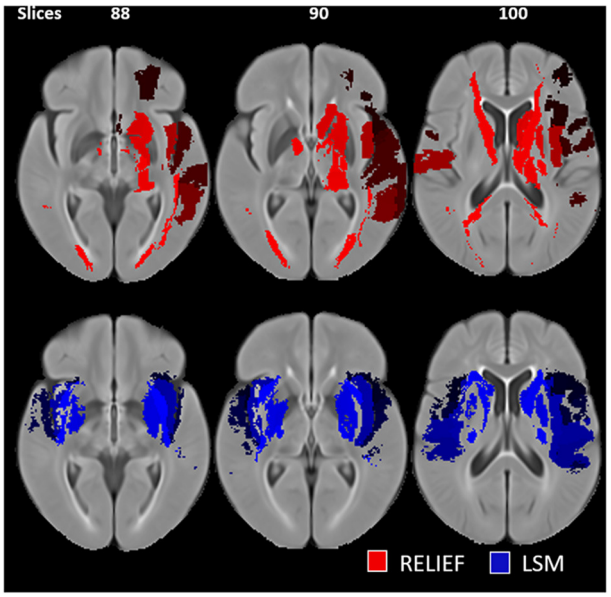


FIGURE 1 | Selected regions of interest (ROIs) for the RreliefF-based (red) and LSM-based (blue) feature selection. The LSM-based ROIs are hemispherically asymmetrical and include regions outside of the subcortical nuclei.

DISCUSSION

This study demonstrates that conventional machine learning feature selection methods (M_{RELIEF}) can identify important brain regions for stroke outcome prediction as well as the conventional lesion-symptom mapping methods (M_{LSM}).

The advantage of the M_{RELIEF} model over the M_{LSM} model are two-fold. First, the M_{RELIEF} model is simpler since it uses <50% of features compared to the M_{LSM} model and results in similar predictive performance. Second, the M_{RELIEF} setup does not require extensive LSM computations to derive structure-function relationships and identify eloquent brain regions. Specifically, despite using a fewer number of regions, the ROIs chosen by the M_{RELIEF} model are largely in the left hemisphere and include regions that correspond to the dominance of left-hemispheric functions assessed by NIHSS.

Importantly, using LSM for ROI selection has additional limitations that the RreliefF feature selection overcomes. First, LSM analyses suffer from low statistical power due to the corrections for multiple comparisons and do not account for violating assumptions of normality in the outcome score. Second, the LSM analysis results in individual voxel weights, which are not really needed to compute region-level inferences of critical brain regions that are associated with a deficit. While LSM is a powerful tool to investigate the neural correlates of stroke induced clinical deficit, its usefulness to select ROIs for stroke outcome prediction tasks seems rather limited. For these reasons, and by applying the Occam's razor principle in model selection, traditional feature selection methods seem to be better suited for future research in stroke outcome prediction.

The proposed framework has a design advantage in comparison to the existing prognostic models of stroke outcome. A recent review on predictive models of stroke outcome (18) reports that: (i) the target outcome of the predictive model is usually a categorized version of functional outcome[1]; (ii) the variables used to model this score include prognostic parameters[2], stroke risk factors, and baseline stroke severity measured by the NIHSS scale. An obvious limitation is that classification models predicting binarized functional outcome likely ignore the gradation of stroke severity, which is relevant information for stroke prognostication. Furthermore, the functional outcomes, prognostic parameters, and the baseline severity measures may be strongly correlated resulting in inflated classification accuracies. In the proposed work, both of these limitations (loss of relevant information and collinearity) are addressed by employing the nested regression model. For instance, since the 48-hours NIHSS is highly correlated with the 30-day NIHSS, it might bias the regression model. Therefore, having a nested model that utilizes the short-term outcome to derive image-based ROIs that in turn predict the long-term outcomes seems to be a promising way to reduce the affects of collinearity. Furthermore, it is important to note that the results of different studies describing predictive models are not comparable because of different sample sizes, different evaluation methods, different assessment time points, and different imaging time points. For this reason, the predictive model using clinical data only was included in this study as a means of baseline comparison.

One of the limitations of the proposed work is that the findings are population-specific and are likely to change with the stroke cohort used (type of stroke and sample size), choice of parcellation atlas, LSM technique, and/or training scheme employed. This study is also exploratory in the sense that, subject to availability, the clinical descriptors included are a subset of all potential stroke risk factors reported in the literature. The power calculations for using LSM in predictive analysis has not been explored in this study. Additionally, the burden of preprocessing each patient scan for registration, lesion segmentation, and feature computation is extensive. State-of-the-art deep learning methods have the potential to use 3D MRI or CT scans (without lesion definitions) and do not demand handcrafted image-based features and might not even need manual lesion segmentations. Furthermore, the results of this study can be considered a relevant first step toward building a computer-aided prognosis support tool using explainable machine learning methods. However, the predictive accuracy of the models generated in this study need to be further improved using additional datasets and should be evaluated prospectively using a completely independent dataset.

An important recommendation for future work is to model stroke outcomes using ordinal regression models, which can account for the relative ordering between two values in the NIHSS scale. However, ordinal regression models are more complex and typically require the definition of interval thresholds, which can either be derived from the training data or based on domain knowledge. That said, the results described in this paper will generally hold true for ordinal regression models as well. Confirmatory research in this direction may also benefit from investigating the utility of convolutional neural networks without requiring lesion segmentation to predict long-term stroke outcome as an ordinal regression problem.

CONCLUSIONS

In summary, this study shows that combining clinical and imaging data leads to better stroke outcome predictions compared to using clinical data alone. While lesion-symptom mapping is a powerful neuroscience tool to investigate structure-function relationships in stroke patients, these methods do not appear to have an additional benefit for selecting brain regions important for stroke outcome prediction compared to rather simple and data-driven feature selection methods.

DATA AVAILABILITY STATEMENT

The acquisition of the datasets for the two trials was approved by the respective local ethics board at each site contributing to the two trials. All datasets used in this secondary study were made available after complete anonymization. Requests

[1]Examples include: the modified Ranking Scale (mRS), NIHSS, Barthel Index, etc.
[2]Examples include: Preadmission Comorbidities, Level of Consciousness, Age, and Neurological Deficit (PLAN); Stroke Prognostication Using Age and National Institutes of Health Stroke Scale (SPAN); Totaled Health Risks in Vascular Events (THRIVE), etc.

to access these datasets should be directed to Nils Forkert, nils.forkert@ucalgary.ca.

AUTHOR CONTRIBUTIONS

DR conceptualized, conducted experiments, and drafted the paper. MH contributed data, validated the design of experiments, and critically reviewed the paper. JF and MG contributed data, and reviewed the paper. AD contributed data. NF validated the design of experiments, and critically reviewed the paper. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the Heart and Stroke Foundation of Canada Grant in aid (G-17-0018368), the Canada Research Chairs program, the River Fund at Calgary Foundation, and the Hotchkiss Brain Institute.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fneur.2021.663899/full#supplementary-material

REFERENCES

1. Price CJ, Hope, TM, Seghier, ML. Ten problems and solutions when predicting individual outcome from lesion site after stroke. *Neuroimage.* (2017) 145:200–8. doi: 10.1016/j.neuroimage.2016.08.006
2. Gianluca B, Neuberger U, Mahmutoglu MA, Foltyn M, Herweh C, Nage S, et al. Multimodal predictive modeling of endovascular treatment outcome for acute ischemic stroke using machine-learning. *Stroke.* (2020) 51:3541–51. doi: 10.1161/STROKEAHA.120.030287
3. Vercio LL, Amador K, Bannister JJ, Crites S, Gutierrez A, MacDonald ME, Moore J, et al. Supervised machine learning tools: a tutorial for clinicians. *J. Neural Eng.* (2020) 17:062001. doi: 10.1088/1741-2552/abbff2
4. de Haan B, Karnath H-O. A hitchhiker's guide to lesion-behaviour mapping. *Neuropsychologia.* (2018) 115:5–16. doi: 10.1016/j.neuropsychologia.2017.10.021
5. 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:e0129569. doi: 10.1371/journal.pone.0129569
6. Demchuk AM, Goyal M, Menon BK, Eesa M, Ryckborst KJ, Kamal N, et al. Endovascular treatment for small core and anterior circulation proximal occlusion with emphasis on minimizing CT to recanalization times (ESCAPE) trial: methodology. *Int J Stroke.* (2015) 10:429–38. doi: 10.1111/ijs.12424
7. Cheng B, Forkert ND,Zavaglia M, Hilgetag CC, Golsari A, Siemonsen S, et al. Influence of stroke infarct location on functional outcome measured by the modified rankin scale. *Stroke.* (2014) 45:1695–702. doi: 10.1161/STROKEAHA.114.005152
8. Cotoi A, Batey C, Hussein N, Janzen S, Teasell R. Rehabilitation of younger patients post stroke. In: *Evidence Review Handbook*. ESRBR (2018). p. 53. Available online at: http://www.ebrsr.com/evidence-review/21-rehabilitation-younger-patients-post-stroke
9. Choi JC, Kim BJ, Han M-K, Lee SJ, Kang K, Park J-M, et al. Utility of items of baseline national institutes of health stroke scale as predictors of functional outcomes at three months after mild ischemic stroke. *J Stroke Cerebrovasc Dis.* (2017) 26:1306–13. doi: 10.1016/j.jstrokecerebrovasdis.2017.01.027
10. Rajashekar D, Wilms M, MacDonald ME, Ehrhardt J, Mouches P, Frayne R, et al. High-resolution T2-FLAIR and non-contrast CT brain atlas of the elderly. *Sci Data.* (2020) 7:1–7. doi: 10.1038/s41597-020-0379-9
11. Jiang, T. Brainnetome: a new -ome to understand the brain and its disorders. *Neuroimage.* (2013) 80:263–72. doi: 10.1016/j.neuroimage.2013.04.002
12. Oishi K, Faria AV, van Zijl PCM, Mori, S. *MRI Atlas of Human White Matter*. Elsevier Science & Technology (2010).
13. Rorden C, Karnath H-O, Bonilha L. Improving lesion-symptom mapping. *J Cogn Neurosci.* (2007) 19:1081–8. doi: 10.1162/jocn.2007.19.7.1081
14. Robnik-Šikonja M, Kononenko I. An adaptation of Relief for attribute estimation in regression. In *Machine Learning: Proceedings of the Fourteenth International Conference (ICML'97).* San Francisco, CA (1997) 5:296–304.
15. LESYMAP. *LESYMAP*. Available online at: https://dorianps.github.io/LESYMAP/
16. *FSelector: Selecting Attributes version 0.33 from CRAN*. Available online at: https://rdrr.io/cran/FSelector/
17. Rajashekar D, Mouchès P, Fiehler J, Menon BK, Goyal M, Demchuk AM, et al. Structural integrity of white matter tracts as a predictor of acute ischemic stroke outcome. *Int J Stroke.* (2020) 15:965–72. doi: 10.1177/1747493020915251
18. Gao MMY, Wang J, Saposnik G. The art and science of stroke outcome prognostication. *Stroke.* (2020) 51:1358–60. doi: 10.1161/STROKEAHA.120.028980

Conflict of Interest: JF (all unrelated): Research support: EU, BMBF, BMWi, DFG, Acandis, Medtronic, Microvention, Stryker, Consultancy: Acandis, Cerenovus, Medtronic, Microvention, Penumbra, Phenox, Stryker, Stock: Tegus, Executive functions: University Medical Center Hamburg-Eppendorf, Eppdata GmbH. MH reports grants from Covidien (Medronic LLC), during the conduct of the ESCAPE study; personal fees from Merck, non-financial support from Hoffmann-La Roche Canada Ltd, grants from Covidien (Medtronic), grants from Boehringer-Ingleheim, grants from Stryker Inc., grants from Medtronic LLC, grants from NoNO Inc., outside the submitted work; In addition, MH has a patent Systems and Methods for Assisting in Decision-Making and Triaging for Acute Stroke Patients pending to US Patent office Number: 62/086,077 and owns stock in Calgary Scientific Incorporated, a company that focuses on medical imaging software, is a director of the Canadian Federation of Neurological Sciences, a not-for-profit group and has received grant support from Alberta Innovates Health Solutions, CIHR, Heart & Stroke Foundation of Canada, National Institutes of Neurological Disorders and Stroke.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neurology

ORIGINAL RESEARCH
published: 07 May 2021
doi: 10.3389/fneur.2021.668360

Risk Factors of Hypoperfusion on MRI of Ischemic Stroke Patients Within 7 Days of Onset

Jingjing Xiao[1†], *Huazheng Liang*[1,2†], *Yue Wang*[1], *Shaoshi Wang*[1], *Yi Wang*[1,3] *and Yong Bi*[1*]

[1] Department of Neurology, Shanghai Fourth People's Hospital Affiliated to Tongji University School of Medicine, Shanghai, China, [2] Department of Neurology, Translational Research Institute of Brain and Brain-Like Intelligence, Shanghai Fourth People's Hospital Affiliated to Tongji University School of Medicine, Shanghai, China, [3] College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, China

Objective: Hypoperfusion is an important factor determining the prognosis of ischemic stroke patients. The present study aimed to investigate possible predictors of hypoperfusion on MRI of ischemic stroke patients within 7 days of stroke onset.

Methods: Ischemic stroke patients, admitted to the comprehensive Stroke Center of Shanghai Fourth People's Hospital affiliated to Tongji University within 7 days of onset between January 2016 and June 2017, were recruited to the present study. Magnetic resonance imaging (MRI), including both diffusion-weighted imaging (DWI) and perfusion-weighted imaging (PWI), was performed within 7 days of the symptom onset. Time to maximum of the residue function (T_{max}) maps were automatically evaluated using the RAPID software. The volume of hypoperfusion was measured outside the infarct area based on ADC $< 620 \times 10^{-6}$ mm^2/s. The 90 d mRS score was assessed through either clinic visits or telephone calls. Multivariate step-wise analysis was used to assess the correlation between MR findings and clinical variables, including the demographic information, cardio-metabolic characteristics, and functional outcomes.

Results: Among 635 patients admitted due to acute ischemic stroke within 7 days of onset, 241 met the inclusion criteria. Hypoperfusion volume of 38 ml was the best cut-off value for predicting poor prognosis of patients with cerebral infarction (90 d-mRS score $\geq$ 2). The incidences of MR perfusion $T_{max} >$ 4–6 s maps with a volume of 0–38 mL or >38 mL were 51.9% (125/241) and 48.1% (116/241), respectively. Prior stroke and vascular stenosis (≥70%) were associated with MR hypoperfusion. Multivariate step-wise analysis showed that prior stroke and vascular stenosis (≥70%) were risk factors of T_{max} > 4–6 s maps, and the odds ratios (OR) were 3.418 (adjusted OR 95% CI: 1.537–7.600), and 2.265 (adjusted OR, 95% CI: 1.199–4.278), respectively.

Conclusion: Our results suggest that prior stroke and vascular stenosis (≥70%) are strong predictors of hypoperfusion in patients with acute ischemic stroke within 7 days of stroke onset.

Keywords: ischemic stroke, magnetic resonance imaging, time to maximum of the residue function, risk factors, correlation analysis

OPEN ACCESS

Edited by:
Ramin Zand, Geisinger Health System, United States

Reviewed by:
Seyedmehdi Payabvash, Yale University, United States
Norbert Nighoghossian, Université Claude Bernard Lyon 1, France
Aaryani Tipirneni-Sajja, University of Memphis, United States

***Correspondence:**
Yong Bi
drbiyong@126.com

*These authors have contributed equally to this work

Specialty section:
This article was submitted to Stroke, a section of the journal Frontiers in Neurology

Received: *16 February 2021*
Accepted: *13 April 2021*
Published: *07 May 2021*

Citation:
Xiao J, Liang H, Wang Y, Wang S, Wang Y and Bi Y (2021) Risk Factors of Hypoperfusion on MRI of Ischemic Stroke Patients Within 7 Days of Onset. Front. Neurol. 12:668360. doi: 10.3389/fneur.2021.668360

INTRODUCTION

Intravenous thrombolysis and endovascular therapy are effective methods for the treatment of acute ischemic stroke (1, 2). However, due to time window restrictions, only a small number of people receive timely treatment, and over 70% of stroke patients still have disabilities (modified Rankin classification, mRS2-6) due to the presence of hypoperfused tissues (3, 4). Quantitative assessment of hemodynamic indices of acute stroke patients will facilitate the discovery of potential predictors of hypoperfusion, which will reveal new targets for early and effective intervention. Currently, a number of factors have been reported to influence functional outcomes of acute ischemic stroke patients, including blood glucose, blood pressure, history of atrial fibrillation, baseline NIHSS, volume of core infarction, blood perfusion, and vascular lesions (5–12). Previous studies have shown that abnormal brain perfusion is closely related to stroke recurrence and functional outcome, but there are few studies on risk factors impacting brain perfusion.

T_{max} is a widely used parameter of magnetic resonance perfusion for patients with acute ischemic stroke and has been used in clinical trials (13, 14). Different T_{max} thresholds reflect different degrees of hypoperfused volumes, with a high threshold reflecting a low degree of hypoperfusion (15). Changes of T_{max} may reflect the microvascular integrity of collaterals and the perfusion status of brain tissue (16). In view of the fact that perfusion imaging is closely related to the status of collateral circulation, the cerebral perfusion parameters on MRI may be a good biomarker of collateral circulation. Therefore, it is reasonable to use T_{max} to evaluate the status of tissue hypoperfusion and facilitate decision-making on the choice of treatments for patients with AIS (17–19).

It is well-known that the penumbra is the area surrounding the ischemic core, which has a high risk of progressing to infarct. $T_{max} > 6$ s can accurately define the penumbra (20). Albers et al. screened patients with salvageable penumbra for endovascular thrombectomy using $T_{max} > 6$ s with the assistance of the RAPID software (17). Time to peak contrast concentration (TTP) or time at which the deconvolved residue function reaches its maximum value (T_{max}) is generally used to evaluate hypoperfusion status. Compared with TTP, T_{max} has the advantage of reducing dependence on bolus shape and cardiac output (21). Therefore, T_{max} seems to be more appropriate in evaluating tissue lesions with hypoperfusion. It has been reported that $T_{max} > 6$ s or $T_{max} > 4$ s is more accurate than $T_{max} > 2$ s in predicting the salvageable penumbra or stroke progression. Difference between the volumes of $T_{max} > 4$ s and $T_{max} > 6$ s seems to be the best biomarker in identifying severe hypoperfusion (22). Studies have shown benefit from prolonged reperfusion therapy with increased likelihood of good prognosis through evaluating the ischemic penumbra with the perfusion parameter T_{max} (17, 19). However, few studies have reported risk factors of low perfusion in Chinese populations. Therefore, the present study aimed to quantitatively evaluate the hypoperfusion status of AIS patients and to explore the potential predictors of hypoperfusion on MRI.

MATERIALS AND METHODS

Subjects

Acute ischemic stroke patients, admitted to the comprehensive Stroke Center of Shanghai Fourth People's Hospital affiliated to Tongji University within 7 days of onset between January 2016 and June 2017, were recruited to the present study.

The inclusion criteria were: (a) Patients who were admitted within 7 days of onset and evaluated by two stroke neurologists (23, 24); (b) MR images including both DWI and PWI were available at the time of hospitalization; (c) T_{max} maps were assessed using the RAPID software (iSchemaView USA, Version 4.9) (25) independently.

The exclusion criteria: patients did not have their perfusion status assessed using DWI and PWI within 7 days of stroke onset.

Demographic data, clinical variables, risk factors, neurologic deficits, time between MRI scan and stroke onset were documented for each patient. The 90 d mRS was evaluated by experienced neurologists.

MRI Parameters

MRI scans were collected using a 1.5-T Avanto scanner (Siemens, Erlangen, Germany). The imaging protocol for this study included diffusion-weighted imaging (DWI), perfusion weighted imaging (PWI), apparent diffusion coefficient (ADC), fluid-attenuated inversion recovery (FLAIR), and magnetic resonance angiography (MRA). Imaging parameters were listed below, DWI: 19 slices, 192 × 192 matrices; slice thickness = 5.5 mm; TR/TE, 3,600/102 ms; field of view = 230 mm^2, b = 0 and 1,000 s/mm^2; EPI factor = 192; bandwidth = 964 Hz/pixel. FLAIR: 18 slices, 256 × 190 matrices; slice thickness = 5.5 mm; TR/TE, 4,000/92 ms; field of view = 230 mm^2; TI = 1,532.6 ms; bandwidth = 190 Hz/Px; flip angle = 150°. Dynamic susceptibility contrast PWI (DSC-PWI): 19 slices, 128 × 128 matrices; slice thickness = 5 mm; TR/TE, 1,590/32 ms; measurements = 50; field of view = 230 mm^2; band width = 1,346 Hz/pixel; flip angle = 90°. A Gd-DTPA contrast agent (gadopentetate dimeglumine injection; Shanghai Pharmaceutical Corporation, Shanghai, China) was injected intravenously (0.2 mmol/kg body weight) at a rate of 4 mL/s after a bolus with 30 ml normal saline. Three-dimensional time-of-flight MRA of the internal carotid artery (ICA) and intracranial circulation: 241 × 256 matrices, slice thickness = 0.7 mm; TR/TE, 25/7 ms; field of view = 180 mm^2; Bandwidth = 100 Hz/PX; flip angle = 25°.

Post-processing

Estimates of hypoperfusion on PWI were calculated using the RAPID software (iSchemaView USA, Version 4.9), which is an automated imaging post-processing system. ADC < 620×10^{-6} mm^2/s was adopted to define the infarct core (26). Volumes of $T_{max} > 4$ s and >6 s were used to determine hypoperfusion in ischemic stroke patients. The volume of hypoperfusion was measured outside the infarct area, based on ADC < 620×10^{-6} mm^2/s. Measurement of vascular stenosis was performed on Magnetic Resonance Angiography by two independent radiologists using the North American Symptomatic Carotid Endarterectomy Trial (NASCET) method (27). The extent of

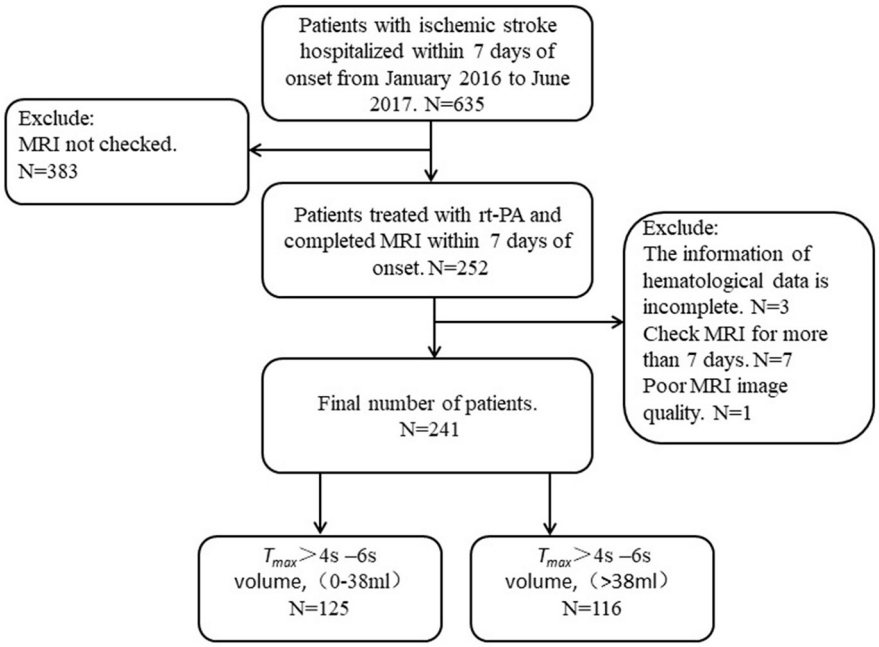


FIGURE 1 | Flowchart of patient recruitment.

reduction of the internal arterial diameter was then graded. If the measurements of two radiologists were inconsistent, repetition will be required before reaching a conclusion.

Clinical Outcomes

The primary outcome was assessed using the 90-day modified Rankin Scale (mRS). mRS ranges from 0 (asymptomatic) to 6 (death). Excellent functional outcomes are defined by a 90 d-mRS score ≤ 1, and poor functional outcomes are defined by a 90 d-mRS score ≥ 2. Three months after stroke onset, mRS scores were collected through clinic visits or telephone calls.

Ethics

Ethical approval for this study was obtained from the Human Research Ethics Committee of Shanghai Fourth People's Hospital Affiliated to Tongji University School of Medicine. Written informed consent was obtained from all subjects.

Statistical Analysis

Data analysis was performed using IBM SPSS (version 22.0) for Windows (SPSS Inc., Chicago, IL, USA). Continuous parameters were presented as mean ± standard deviation (SD) or median with interquartile range (IQR); categorical variables were summarized as independent proportions. Baseline information of patients with or without MRI perfusion abnormalities was compared using either *t*-test or Mann-Whitney U-test for continuous variables and χ^2 or Fisher's exact test for categorical variables. In this study, the cutoff value was 38 ml for hypoperfusion. Logistic regression analysis was used to identify independent predictors of $T_{max}>$ 4–6 s maps. Multivariate step-wise regression modeling was used to correlate $T_{max} >$ 4–6 s maps with potential risk factors with their *P*-values < 0.01. All correlation data were presented as odds ratios (OR) with their corresponding 95% confidence intervals (CI) and *P*-values. Statistical significance was considered when $P < 0.05$.

RESULTS

In the present study, 635 patients with acute ischemic stroke were admitted within 7 days of onset, but only 241 (73 women, 168 men; median age: 67 years) had technically adequate DWI and PWI scans. Among those 394 patients who were excluded, 383 of them did not have MRI, 3 could not provide the information of hematological data, 10 had their MRI more than 7 days after stroke onset, 1 had poor MRI image quality. Among these 241 patients, 107 patients had excellent functional outcomes and 134 poor functional outcomes. Among these patients, 125 had $T_{max} >$ 4–6 s volume in the range of 0–38 ml and the other 116 had a volume > 38 ml (**Figure 1**, **Table 1**).

Baseline Characteristics

Baseline characteristics of patients included in the present study were shown in **Tables 1**, **2**. The median (IQR) age of these patients was 67 (61–79) years and their median Admission NIHSS score was 3 (IQR: 1–9). Perfusion status was evaluated after a median (IQR) delay of 3 (1–6) days from the symptom onset. The median ADC volume was 0 (IQR: 0–7.5) ml. The median fasting blood glucose (FBG) was 6.0 (IQR: 5.2–8.2) mmol/L. A history of hypertension was present in 68.5% (165/241) of patients, diabetes mellitus in 34.4% (83/241), atrial fibrillation in 14.5% (35/241), prior stroke in17%

TABLE 1 | Comparison of demographic and clinical characteristics among ischemic stroke patients within 7 days of onset based on T_{max} >4 s −6 s volume 38 ml.

Characteristics	Total ($n = 241$)	T_{max} > 4–6 s volume, (0–38 ml) ($n = 125$)	T_{max} > 4–6 s volume, (>38 ml) ($n = 116$)	*P*-value
Age, y, median (IQR)	67 (61–79)	66 (59–75)	70 (62–80)	0.009
Male, *n* (%)	168 (69.7%)	86 (68.8%)	82 (70.7%)	0.75
Medical history, *n* (%)				
Hypertension	165 (68.5%)	84 (67.2%)	81 (69.8%)	0.661
Diabetes mellitus	83 (34.4%)	44 (35.2%)	39 (33.6%)	0.797
Atrial fibrillation	35 (14.5%)	9 (7.2%)	26 (22.4%)	0.001
Smoking	80 (33.2%)	39 (31.2%)	41 (35.3%)	0.495
Previous ischemic stroke	41 (17%)	12 (9.6%)	29 (25.0%)	0.001
Cardo-metabolic				
FBG, mmol/L, median (IQR)	6.0 (5.2–8.2)	5.9 (5.2–7.8)	6.1 (5.4–8.3)	0.310
TC, mmol/L, median (IQR)	3.72 (1.89–4.89)	3.73 (1.95–5.02)	3.68 (1.80–4.62)	0.385
LDL, mmol/L, median (IQR)	2.63(1.98–3.34)	2.65(2.0–3.40)	2.62(1.94–3.31)	0.466
TG, mmol/L, median (IQR)	2.32 (1.3–4.22)	2.35 (1.37–4.23)	2.24 (1.27–4.13)	0.408
HDL, mmol/L, median (IQR)	1.12 (0.88–1.34)	1.14 (0.91–1.33)	1.07 (0.85–1.39)	0.802
Hcy, umol/L, median (IQR)	13.6 (11.3–16.6)	12.4(10.9–15.85)	14.25 (12.0–17.0)	0.011
SBP at baseline, mm Hg, median (IQR)	140 (130–160)	150 (135–160)	140 (130–153)	0.093
DBP at baseline, mm Hg, median (IQR)	80 (78–90)	80 (79–90)	80 (77–89)	0.404
Admission NIHSS score, median (IQR)	3 (1–9)	3 (1–6)	6 (2–10)	<0.001
Treatment				0.782
Alteplase treatment	36 (14.9%)	19 (15.2%)	17 (14.7%)	
Bridge-EVT	2 (0.8%)	2 (1.6%)	0	
Direct-EVT	12 (5.0%)	1 (0.8%)	11 (9.5%)	
Standard medical therapy	191 (79.3%)	103 (82.4%)	88 (75.9%)	
Symptom onset to the MR perfusion, d, median (IQR)	3 (1–6)	4 (2–6)	3 (1–6)	0.006
Vascular stenosis (≥70%), *n* (%)	154 (63.9%)	62 (49.6%)	92 (79.3%)	<0.001
CD4/CD8 (≥1.7), *n* (%)	99 (41.1%)	58 (46.4%)	41 (35.3%)	0.081
Secondary bleeding, *n* (%)	12 (9%)	6 (11.5%)	6 (7.4%)	0.616
	7 (2.9%)	2 (1.6%)	5 (4.3%)	0.266
Clinical outcomes (90 d mRS)	2 (0–3)	1 (0–3)	2 (1–4)	<0.001
Excellent functional outcome (0–1)	107 (44.4%)	73 (58.4%)	34 (29.3%)	
Poor functional outcome (2–6)	134 (55.6%)	52 (41.6%)	82 (70.7%)	
ADC < 620×10^{-6} mm^2/s volume, ml, median (IQR)	0 (0–7.5)	0 (0–0)	0 (0–24.5)	<0.001

Measurement data conformed to normal distribution variables expressed as mean ± standard deviation; non-normal distribution variables were expressed as median (25–75%). The categorical variables are expressed in terms of frequency (percentage).

IQR, interquartile range; INR, International standardization ratio; LDL, Low-density lipoprotein; HDL, Low-density lipoprotein; NIHSS, National Institutes of Health Stroke Scale; SBP, systolic blood pressure; DBP, diastolic blood pressure; TC, total cholestrol; Hcy, homocysteine; FBG, fasting blood-glucose; IVT, intravenous thrombolysis; Direct-EVT, direct endovascular thrombectomy; Bridge-EVT, bridge endovascular thrombectomy; mRS, modified Rankin Scale; T_{max} > 4 s, Time to maximum of the residue function > 4 s; T_{max} > 6 s, Time to maximum of the residue function > 6 s; ADC, apparent diffusion coefficient; T_{max} > 4–6 s volume, Volume difference between T_{max} > 4 s and T_{max} > 6 s.

Significant difference at $\alpha < 0.05$.

(41/241) and smoking in 33.2% (80/241). The patients were divided into four groups: intravenous thrombolysis (36/241), Bridge endovascular thrombectomy (2/241), Direct endovascular thrombectomy (12/241), and standard medical therapy alone (191/241) (**Table 1**).

Comparison of Demographic and Clinical Characteristics Among Ischemic Stroke Patients Within 7 Days of Onset Based on T_{max} > 4–6 s Volume 38 ml

Demographic characteristics, possible risk factors associated with the T_{max} > 4–6 s map, and comparisons of these variables between groups with volume = 0–38 ml and >38 ml were presented in **Table 1**. Hypoperfusion volume of 38 ml was the best cut-off value for predicting poor prognosis of patients with cerebral infarction (90 d-mRS score ≥ 2) as shown by the ROC (AUC: 0.67, 95% CI: 0.603–0.738, sensitivity: 0.612, specificity 0.673, $p < 0.001$) (**Figures 2**, **3**). In univariate analyses, factors associated with hypoperfusion were: age [66 (59–75) vs. 70 (62–80), $p = 0.009$], atrial fibrillation (7.2 vs. 22.4%, $p = 0.001$), previous ischemic stroke (9.6 vs. 25.0%, $p = 0.001$), homocysteine [12.4 (10.9–15.85) vs. 14.25 (12.0–17.0), $p = 0.011$], admission NIHSS score [3 (1–6) vs. 6 (2–10), $p < 0.001$], symptom onset to the MR perfusion [4 (2–6) vs. 3 (1–6), $p = 0.006$], vascular stenosis (≥70%) (49.6 vs. 79.3%, $p < 0.001$), ADC volume [0 (0–0) vs. 0 (0–24.5), $p < 0.001$]. Other variables were not statistically different (**Table 1**).

TABLE 2 | Imaging characteristics of ischemic stroke patients with excellent and poor functional outcomes.

Characteristics	Total ($n = 241$)	Excellent functional outcome group (90 d mRS 0–1) ($n = 107$)	Poor functional outcome group (90 d mRS 2–6) ($n = 134$)	*P*-value
Vascular stenosis (≥70%), *n* (%)	154 (63.9%)	63 (58.9%)	91 (67.9%)	0.147
ADC < 620 × 10^{-6} mm^2/s volume, ml, median (IQR)	0 (0–8)	0 (0–0)	0 (0–12.3)	<0.001
T_{max} > 4 s volume, ml, median(IQR)	35 (8–105)	19 (5–63)	71 (14.0–166.3)	<0.001
T_{max} > 6 s volume, ml, median(IQR)	0 (0–13)	0 (0–3)	0 (0–32.5)	<0.001
T_{max} > 4–6 s volume, ml, median(IQR)	35 (8–88)	18 (4–57)	48 (14–113)	<0.001
T_{max} > 4 s –ADC volume, ml, median (IQR)	34 (6–95)	18 (0–63)	54 (10.8–146.0)	<0.001
T_{max} > 6 s –ADC volume, ml, median (IQR)	0 (0–3.5)	0 (0–0)	0 (0–13.5)	0.004

Measurement data conformed to normal distribution variables expressed as mean ± standard deviation; non-normal distribution variables were expressed as median (25–75%). The categorical variables are expressed in terms of frequency (percentage).
IQR, interquartile range; mRS, modified Rankin Scale; ADC, apparent diffusion coefficient; T_{max} > 4 s, Time to maximum of the residue function > 4 s; T_{max} > 6 s, Time to maximum of the residue function > 6 s; T_{max} > 4–6 s volume, Volume difference between T_{max} > 4 s and T_{max} > 6 s; T_{max} > 4 s–ADC volume, Volume difference between T_{max} > 4 s and ADC; T_{max} > 6 s–ADC volume, Volume difference between T_{max} > 6 s and ADC.
Significant difference at $\alpha < 0.05$.

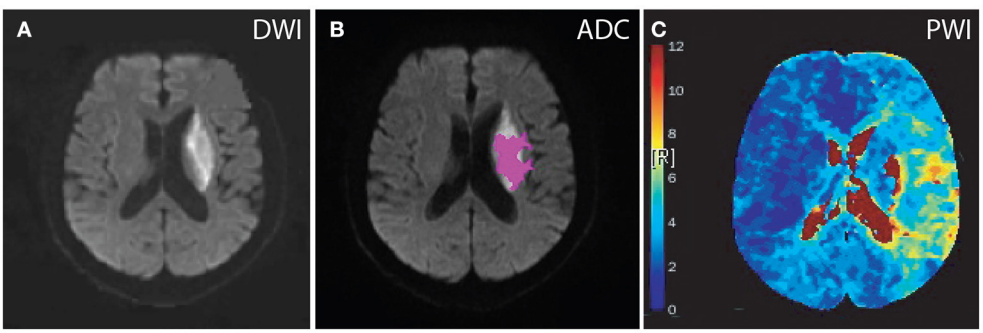


FIGURE 2 | Diffusion and perfusion abnormalities of a patient treated with intravenous tPA. A 63-year-old male who presented with dysarthria and weakness in the right arm, hypertension for more than 3 years, vascular stenosis (≥70%), no previous ischemic stroke, MRI scan was completed 6 days after stroke onset. DWI, diffusion-weighted imaging; PWI, perfusion-weighted imaging; ADC, apparent diffusion coefficient; T_{max}, Time to maximum of the residue function. T_{max} color scale: 4 s < T_{max} ≤ 6 s (blue); 6 s < T_{max} ≤ 8 s (green); 8 s < T_{max} ≤ 10 s (yellow); 10 s < T_{max} (red). **(A)** DWI, lesion volume was 22 ml. **(B)** ADC, lesion volume was 9 ml. **(C)** PWI, lesion volumes according to T_{max} delay were as follows: T_{max} > 4 s, 132 ml; T_{max} > 4 s–ADC volume, 123 ml; T_{max} > 6 s, 22 ml; T_{max} > 6 s–ADC volume, 13 ml; T_{max} > 8 s, 4 ml; and T_{max} > 10 s, 3 ml.

Excellent functional outcomes (mRS 0–1) were present in 44.4% (107/241) of patients, among whom the median (IQR) T_{max} > 4–6 s volume, T_{max} > 6 s volume, and T_{max} > 4 s volume were 18 (4–57) ml, 0 (0–3) ml, and 19 (5–63) ml, respectively. MRA scans with adequate quality of both the carotid and intracranial vessels were available in the present study. MRA showed that vascular stenosis (>70%) was detected in 154 (63.9%) patients (**Table 2**).

Prediction of MRI Perfusion Abnormality

In the present study, we divided patients with hypoperfusion into two groups based on the volume of T_{max} > 4–6 s: T_{max} > 4–6 s = 0–38 ml and T_{max} > 4–6 s > 38 ml. The incidences of MR perfusion T_{max} > 4–6 s = 0–38 ml and T_{max} > 4–6 s > 38 ml were 51.9% (125/241) and 48.1% (116/241), respectively (**Table 1**).

In the univariate binary logistic regression analysis, age ($P = 0.010$, OR = 1.031, 95% CI: 1.007–1.056), atrial fibrillation ($P = 0.001$, OR = 3.723, 95% CI: 1.662–8.340), previous ischemic stroke ($P = 0.002$, OR = 3.139, 95% CI: 1.515–6.504), admission NIHSS score ($P < 0.001$, OR = 1.091, 95% CI: 1.040–1.145), symptom onset to MR perfusion ($P = 0.018$, OR = 0.872, 95% CI: 0.779–0.977), vascular stenosis (≥70%) ($P < 0.001$, OR = 3.895, 95% CI: 2.203–6.887), ADC volume ($P < 0.001$, OR = 1.044, 95% CI: 1.019–1.070) were independently associated with MR perfusion abnormality in ischemic stroke patients within 7 days of onset (**Table 3**).

Multivariate step-wise regression modeling was performed for predictors with *P*-values < 0.01 and the multivariate logistic regression analysis performed to show the correlation between previous ischemic stroke, stenosis (≥70%), and MR perfusion abnormalities. Patients with previous ischemic stroke and stenosis were likely to develop hypoperfusion on PWI maps. The adjusted odds ratios were 3.418 (95% CI: 1.537–7.600, $P = 0.001$) and 2.265 (95% CI: 1.199–4.278, $P = 0.012$), respectively. Other variables were not significantly associated with hypoperfusion (**Table 3**).

DISCUSSION

In the present study, risk factors of hypoperfusion on MRI were analyzed for AIS patients admitted within 7 days of onset. It was found that the perfusion parameter T_{max} > 4–6 s volume was related to clinical prognosis. Patients with previous stroke and vascular stenosis (≥70%) were more likely to have hypoperfusion, and these two were independent risk factors of low perfusion as shown by T_{max} > 4–6 s map > 38 ml.

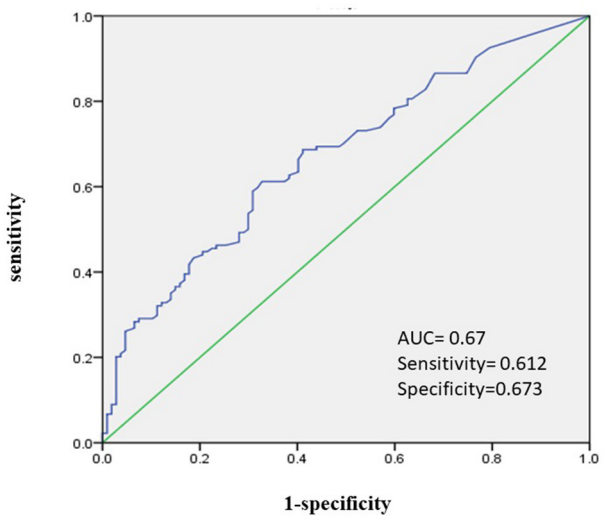


FIGURE 3 | Area under the ROC curve predicts hypoperfusion based on 90 d-mRS.

T_{max} > 4–6 s Map and 90 d mRS

It is well-known that DWI and ADC maps were closely related to the final infarct volume and were important predictors of clinical prognosis (9, 28). The present study used non-invasive multimode magnetic resonance imaging to quantitatively evaluate T_{max} > 4–6 s map. It was found that the tissue perfusion status of T_{max} > 4–6 s was closely related to 90 d mRS. The greater the T_{max} > 4–6 s map was, the worse the 90 d mRS was. Previous studies have shown that perfusion imaging was closely related to the status of collateral circulation, and the cerebral perfusion parameter T_{max} was a good biomarker of collateral volume (29). Therefore, it is reasonable to use the T_{max} > 4–6 s map to evaluate tissue hypoperfusion.

Collateral status could be used to predict the prognosis of patients with acute ischemic stroke as the key determinant (30, 31). Perfusion status changes temporally and spatially. Findings from perfusion imaging reflect collateral status or response to treatment for those without large vessel occlusion within 6 h or for those with large vessel occlusion within 24 h. These are closely correlated with clinical prognosis. In this study, the cutoff value 38 ml of T_{max} > 4–6 s map was used to define hypoperfusion on MRI with an aim to find risk factors that were related to hypoperfusion, and to provide a reasonable direction for accurate control of these risk factors. This would improve clinical prognosis and reduce the occurrence of recurrent ischemic events.

Previous Ischemic Stroke and T_{max} > 4–6 s Maps

This study found that hypoperfusion was associated with recurrent stroke and persistent deterioration of neurological functions. It might be related to a certain proportion of vascular stenosis and distal hypoperfusion in patients with previous strokes, which are likely to recur. This is consistent with the result of a previous studies (32). Previous studies have shown that stroke patients have a recurrence rate of 17% within 1 year. In symptomatic intracranial atherosclerotic stenosis (ICAS) patients, the more severe the baseline hypoperfusion was, the higher the risk of stroke recurrence was (33–35). In this study, 41 patients had a history of stroke, among whom 29 had their T_{max} > 4–6 s volume > 38 ml, accounting for 25% of this type

TABLE 3 | Factors independently associated with T_{max} > 4–6 s > 38 mL in ischemic stroke patients within 7 days of onset.

	OR (95% CI)	*P*-value	Adjusted OR (95% CI)	*P*-value
Age	1.031 (1.007–1.056)	0.010	1.037 (1.009–1.066)	0.010
Atrial fibrillation	3.723 (1.662–8.340)	0.001	1.404 (0.545–3.614)	0.482
Previous ischemic stroke	3.139 (1.515–6.504)	0.002	3.418 (1.537–7.600)	0.001
Admission NIHSS score	1.091 (1.040–1.145)	<0.001	1.038 (0.983–1.096)	0.181
symptom onset to the MR perfusion	0.872 (0.779–0.977)	0.018	0.905 (0.793–1.034)	0.142
Vascular stenosis (≥70%)	3.895 (2.203–6.887)	<0.001	2.265 (1.199–4.278)	0.012
ADC < 620 × 10^{-6} mm^2/s volume	1.044 (1.019–1.070)	<0.001	1.033 (1.009–1.058)	0.008
HCY	1.007 (0.986–1.029)	0.516	N	N

CI, confidence interval; OR, odds ratio; NIHSS, National Institutes of Health Stroke Scale; SBP, systolic blood pressure; ADC, apparent diffusion coefficient, Hcy, homocysteine.

of patients. Among these 41 patients, 30 had vascular stenosis $\geq$ 70%, accounting for 73.2% of these patients. Previous strokes are a risk factor of hypoperfusion. The possible reason is that the majority of patients with previous strokes have vascular stenosis, which leads to distal hypoperfusion, and consequently recurrence of ischemic stroke.

Vascular Stenosis and T_{max} > 4–6 s Map

Spencer and Reid first proposed the relationship between cerebral artery stenosis and cerebral blood flow, which predicted a decrease in blood flow when stenosis was >70% (36, 37). Cerebral artery stenosis or occlusion can trigger serious hemodynamic disorders. However, it has been shown that the severity of vascular stenosis does not necessarily affect the status of distal blood flow (38). In this study, it was found that vascular stenosis was related to the T_{max} > 4–6 s map. 63.9% of the patients had evidence of ipsilateral proximal artery stenosis or occlusion on MRA, among whom the proportion of tissue hypoperfusion was 60.4% (93/153). Therefore, tissue hypoperfusion may be more likely to occur in patients with ipsilateral proximal artery stenosis or occlusion on MRA, which is consistent with previous studies (39, 40). This might be due to changes in the cerebral vascular structure and function resulted from intracranial atherosclerosis. Vascular stenosis affects the hemodynamic status, mainly through the decrease of cerebrovascular reserve. In the presence of insufficient collaterals, the decrease in pressure may lead to hypoperfusion.

The volume of hypoperfusion at different thresholds of T_{max} non-invasively reflected the status of collaterals, which is consistent with findings of previous studies. Therefore, it is likely that perfusion parameters on MRI will be a good biomarker for volumes of collateral blood flow (29). The benefit of the present study is to non-invasively assess hypoperfusion volumes at the early stage of ischemic stroke, which reflects the collateral status. It will provide evidence for early intervention to halt stroke progression, prevent stroke recurrence, and to improve clinical prognosis.

The present study has a number of limitations. Firstly, it is a retrospective study in which all subjects were recruited from a local hospital, which may result in selection bias. In addition, it has a relatively small sample size and the conclusion from this study may not be extrapolated to all ischemic stroke patients. Therefore, prospective studies with a large sample size are required to confirm our findings. Secondly, it is a cross-sectional study and can not pinpoint the direct causality between hypoperfusion and the risk factors of ischemic stroke patients within 7 days of onset. A longitudinal design can help to investigate the direct causal relationship between risk factors and MR hypoperfusion in future studies. Thirdly, T_{max} > 4–6 s and volume of hypoperfusion > 38 ml were used to determine hypoperfusion,which is based on the cutoff value of 38 ml in our analysis, whether this method has better accuracy and applicability needs to be verified by future prospective, large-scaled studies. Fourthly, The present study did not repeat MR scan between the therapeutic window and 90 days to evaluate the perfusion status, which might result in treatment bias.

In conclusion, hypoperfusion could be found in ischemic stroke patients within 7 days of onset when PWI was examined, which is related to clinical prognosis. Patients with previous ischemic strokes and vascular stenosis are more likely to have severe hypoperfusion and poor functional outcomes. Accurate control of risk factors may effectively improve functional outcomes. However, larger prospective studies are needed to confirm these findings.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Human Research Ethics Committee of Shanghai Fourth People's Hospital Affiliated to Tongji University School of Medicine. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YB conceived and designed this study. JX and HL collected data and performed statistical analysis and drafted the manuscript. YueW helped with data collection and data analysis. YiW and YB revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The present study was supported by a grant from Shanghai Municipal Commission of Health and Family Planning awarded to YB (No. 201840244), a grant from Commission of Health and Family Planning, Hongkou District, awarded to JX (1803-04), the Talent Promotion Project of Shanghai Fourth People's Hospital in 2019 (SY-XKZT-2019-3003).

ACKNOWLEDGMENTS

We thank Dr. Yu Luo and his team in helping acquire MR images.

REFERENCES

1. The National Institute of Neurological Disorders, and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. *N Engl J Med*. (1995) 333:1581–7. doi: 10.1056/NEJM199512143332401
2. Goyal M, Demchuk AM, Menon BK, Eesa M, Rempel JL, Thornton J, et al. Randomized assessment of rapid endovascular treatment of ischemic stroke. *N Engl J Med*. (2015) 372:1019–30. doi: 10.1056/NEJMoa1414905
3. Bhatia R, Hill MD, Shobha N, Menon B, Bal S, Kochar P, et al. Low rates of acute recanalization with intravenous recombinant tissue plasminogen

activator in ischemic stroke: real-world experience and a call for action. *Stroke.* (2010) 41:2254–8. doi: 10.1161/STROKEAHA.110.592535
4. Saqqur M, Uchino K, Demchuk AM, Molina CA, Garami Z, Calleja S, et al. Site of arterial occlusion identified by transcranial Doppler predicts the response to intravenous thrombolysis for stroke. *Stroke.* (2007) 38:948–54. doi: 10.1161/01.STR.0000257304.21967.ba
5. Huang J, Zhang X, Li J, Tang L, Jiao X, Lv X. Impact of glucose fluctuation on acute cerebral infarction in type 2 diabetes. *Can J Neurol Sci.* (2014) 41:486–92. doi: 10.1017/S0317167100018539
6. Stead LG, Gilmore RM, Vedula KC, Weaver AL, Decker WW, Brown RD Jr. Impact of acute blood pressure variability on ischemic stroke outcome. *Neurology*. (2006) 66:1878–81. doi: 10.1212/01.wnl.0000219628.78513.b5
7. Tu HT, Campbell BC, Christensen S, Desmond PM, De Silva DA, Parsons MW, et al. Worse stroke outcome in atrial fibrillation is explained by more severe hypoperfusion, infarct growth and hemorrhagic transformation. *Int J Stroke.* (2015) 10:534–40. doi: 10.1111/ijs.12007
8. Ali SF, Siddiqui K, Ay H, Singhal A, Viswanathan A, Rost N, et al. Baseline predictors of poor outcome in patients too good to treat with intravenous thrombolysis. *Stroke.* (2016) 47:2986–92. doi: 10.1161/STROKEAHA.116.014871
9. Wheeler HM, Mlynash M, Inoue M, Tipirneni A, Liggins J, Zaharchuk G, et al. Early diffusion-weighted imaging and perfusion-weighted imaging lesion volumes forecast final infarct size in DEFUSE 2. *Stroke.* (2013) 44:681–5. doi: 10.1161/STROKEAHA.111.000135
10. Tomsick T. TIMI, TIBI, TICI: I came, I saw, I got confused. *AJNR Am J Neuroradiol*. (2007) 28:382–4.
11. Pan JW, Yu XR, Zhou SY, Wang JH, Zhang J, Geng DY, et al. Computed tomography perfusion and computed tomography angiography for prediction of clinical outcomes in ischemic stroke patients after thrombolysis. *Neural Regen Res.* (2017) 12:103–8. doi: 10.4103/1673-5374.198994
12. Soares BP, Tong E, Hom J, Cheng SC, Bredno J, Boussel L, et al. Reperfusion is a more accurate predictor of follow-up infarct volume than recanalization: a proof of concept using CT in acute ischemic stroke patients. *Stroke.* (2010) 41:e34–40. doi: 10.1161/STROKEAHA.109.568766
13. Inoue M, Mlynash M, Straka M, Kemp S, Jovin TG, Tipirneni A, et al. Clinical outcomes strongly associated with the degree of reperfusion achieved in target mismatch patients: pooled data from the diffusion and perfusion imaging evaluation for understanding stroke evolution studies. *Stroke.* (2013) 44:1885–90. doi: 10.1161/STROKEAHA.111.000371
14. De Silva DA, Brekenfeld C, Ebinger M, Christensen S, Barber PA, Butcher KS, et al. The benefits of intravenous thrombolysis relate to the site of baseline arterial occlusion in the echoplanar imaging thrombolytic evaluation trial (EPITHET). *Stroke.* (2010) 41:295–9. doi: 10.1161/STROKEAHA.109.562827
15. Lansberg MG, Lee J, Christensen S, Straka M, De Silva DA, Mlynash M, et al. RAPID automated patient selectionfor reperfusion therapy a pooled analysis of the echoplanar imaging thrombolytic evaluation trial (EPITHET) and the diffusion and perfusion imaging evaluation for understanding stroke evolution (DEFUSE) study. *Stroke.* (2011) 42:1608–14. doi: 10.1161/STROKEAHA.110.609008
16. Calamante F, Christensen S, Desmond PM, Ostergaard L, Davis SM, Connelly A. The physiological significance of the time-to-maximum (Tmax) parameter in perfusion MRI. *Stroke.* (2010) 41:1169–74. doi: 10.1161/STROKEAHA.110.580670
17. Albers GW, Marks MP, Kemp S, Christensen S, Tsai JP, Ortega-Gutierrez S, et al. Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging. *N Engl J Med.* (2018) 378:708–18. doi: 10.1056/NEJMoa1713973
18. Nogueira RG, Jadhav AP, Haussen DC, Bonafe A, Budzik RF, Bhuva P, et al. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. *N Engl J Med.* (2018) 378:11–21. doi: 10.1056/NEJMoa1706442
19. Ma H, Campbell BCV, Parsons MW, Churilov L, Levi CR, Hsu C, et al. Thrombolysis guided by perfusion imaging up to 9 hours after onset of stroke. *N Engl J Med*. (2019) 380:1795–803. doi: 10.1056/NEJMoa1813046
20. Olivot JM, Mlynash M, Thijs VN, Purushotham A, Kemp S, Lansberg MG, et al. Relationships between cerebral perfusion and reversibility of acute diffusion lesions in DEFUSE: insights from RADAR. *Stroke.* (2009) 40:1692–7. doi: 10.1161/STROKEAHA.108.538082
21. Lansberg MG, Straka M, Kemp S, Mlynash M, Wechsler LR, Jovin TG, et al. Magnetic resonance imaging profile and response to endovascular reperfusion: results of the DEFUSE 2 prospective cohort study. *Lancet Neurol.* (2012) 11:860–7. doi: 10.1016/S1474-4422(12)70203-X
22. Olivot JM, Mlynash M, Thijs VN, Kemp S, Lansberg MG, Wechsler L, et al. Optimal Tmax threshold for predicting penumbral tissue in acute stroke. *Stroke.* (2009) 40:469–75. doi: 10.1161/STROKEAHA.108.526954
23. Powers WJ, Zivin J. Magnetic resonance imaging in acute stroke not ready for prime time. *Neurology.* (1998) 50:842–3. doi: 10.1212/WNL.50.4.842
24. Tourdias T, Renou P, Sibon I, Asselineau J, Bracoud L, Dumoulin M, et al. Final cerebral infarct volume is predictable by MR imaging at 1 week. *AJNR Am J Neuroradiol.* (2011) 32:352–8. doi: 10.3174/ajnr.A2271
25. Straka M, Albers GW, Bammer R. Real-time diffusionperfusion mismatch analysis in acute stroke. *J Magn Reson Imaging.* (2010) 32:1024–37. doi: 10.1002/jmri.22338
26. Purushotham A, Campbell BC, Straka M, Mlynash M, Olivot JM, Bammer R, et al. Apparent diffusion coefficient threshold for delineation of ischemic core. *Int J Stroke.* (2015) 10:348–53. doi: 10.1111/ijs.12068
27. North American Symptomatic Carotid Endarterectomy Trial Collaborators, Barnett HJM, Taylor DW, Haynes RB, Sackett DL, Peerless SJ, et al. Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. *N Engl J Med.* (1991) 325:445–3. doi: 10.1056/NEJM199108153250701
28. Warach S, Gaa J, Siewert B, Wielopolski P, Edelman RR. Acute human stroke studied by whole brain echo planar diffusion-weighted magnetic resonance imaging. *Ann Neurol.* (1995) 37:231–41. doi: 10.1002/ana.410370214
29. Cortijo E, Calleja AI, Garciá-Bermejo P, Mulero P, Pérez-Fernández S, Reyes J, et al. Relative cerebral blood volume as a marker of durable tissue-at-risk viability in hyperacute ischemic stroke. *Stroke.* (2014) 45:113–8. doi: 10.1161/STROKEAHA.113.003340
30. Vagal A, Aviv R, Sucharew H, Reddy M, Hou Q, Michel P, et al. Collateral clock is more important than time clock for tissue fate. *Stroke.* (2018) 49:2102–7. doi: 10.1161/STROKEAHA.118.021484
31. Arenillas JF, Cortijo E, García-Bermejo P, Levy EI, Jahan R, Liebeskind D, et al. Relative cerebral blood volume is associated with collateral status and infarct growth in stroke patients in SWIFT PRIME. *J Cereb Blood Flow Metab.* (2018) 38:1839–47. doi: 10.1177/0271678X17740293
32. De Havenon A, Khatri P, Prabhakaran S,Yeatts SD, Peterson C, Sacchetti D, et al. Hypoperfusion distal to anterior circulation intracranial atherosclerosis is associated with recurrent stroke. *J Neuroimaging.* (2020) 30:468–70. doi: 10.1111/jon.12710
33. Wang Y, Zhao X, Liu L, Soo YO, Pu Y, Pan Y, et al. Prevalence and outcomes of symptomatic intracranial large artery stenoses and occlusions in China: the Chinese Intracranial Atherosclerosis (CICAS) Study. *Stroke.* (2014) 45:663–9. doi: 10.1161/STROKEAHA.113.003508
34. Kasner SE, Chimowitz MI, Lynn MJ, Howlett-Smith H, Stern BJ, Hertzberg VS, et al. Predictors of ischemic stroke in the territory of a symptomatic intracranial arterialstenosis. *Circulation*. (2006) 113:555–63. doi: 10.1161/CIRCULATIONAHA.105.578229
35. Lyu J, Ma N, Tian C, Xu F, Shao H, Zhou X, et al. Perfusion and plaque evaluation to predict recurrent stroke in symptomatic middle cerebral artery stenosis. *Stroke Vasc Neurol.* (2019) 4:129–34. doi: 10.1136/svn-2018-000228
36. Spencer MP, Reid JM. Quantitation of carotid stenosis with continuouswave (C-W) Doppler ultrasound. *Stroke.* (1979) 10:326–30. doi: 10.1161/01.STR.10.3.326
37. Carrera E, Lee LK, Giannopoulos S, Marshall RS. Cerebrovascular reactivityand cerebral autoregulation in normal subjects. *J Neurol Sci*. (2009) 285:191–4. doi: 10.1016/j.jns.2009.06.041
38. Powers WJ, Press GA, Grubb RL Jr, Gado M, Raichle ME. The effect of hemodynamically significant carotid artery disease on the hemodynamic status of the cerebral circulation. *Ann Intern Med.* (1987) 106:27–35. doi: 10.7326/0003-4819-106-1-27
39. Ge X, Zhao H, Zhou Z, Li X, Sun B, Wu H, et al. Association of fractional flow on 3D-TOF-MRA with cerebral perfusion in patients with MCA stenosis. *AJNR Am J Neuroradiol.* (2019) 40:1124–31. doi: 10.3174/ajnr.A6095
40. Lu SS, Ge S, Su CQ, Xie J, Mao J, Shi HB, et al. MRI of plaque characteristics and relationship with downstream perfusion and

cerebral infarction in patients with symptomatic middle cerebral artery stenosis. *J Magn Reson Imaging*. (2018) 48:66–73. doi: 10.1002/jmri.25879

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neurology

SYSTEMATIC REVIEW
published: 25 May 2021
doi: 10.3389/fneur.2021.641023

The Effectiveness of Music Therapy on Hand Function in Patients With Stroke: A Systematic Review of Randomized Controlled Trials

Wen-Hao Huang[1], Zu-Lin Dou[1], Hui-Min Jin[2], Ying Cui[3], Xin Li[1] and Qing Zeng[4]*

[1] Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China, [2] Shanghai Sunshine Rehabilitation Center, Shanghai, China, [3] Shandong University of Traditional Chinese Medicine, Jinan, China, [4] Department of Rehabilitation Medicine, Zhujiang Hospital of Southern Medical University, Guangzhou, China

Objective: This study aims to evaluate the efficacy of music-supported therapy for stroke patients' hand function.

Methods: The databases used included Cumulative Index to Nursing and Allied Health Literature (CINAHL), MEDLINE, PubMed, Embase, Music Index, and Google Scholar. Studies published between January 2010 and August 2020 were included. The searching key terms included "music-supported therapy," "music therapy," "hand function," "hand dysfunction," "stroke," "ischemic," and "hemorrhagic." Randomized controlled trials or controlled trials involving adults who have hand function problems caused by stroke are included in this study. The methodological quality and risk of bias of the included studies were rated by two independent assessors under the guidance of Cochrane collaboration's risk of bias tool.

Results: Twelve studies that met the inclusion criteria were included in this study. Totally, the data included 598 stroke patients (345 male, 253 female) with recruited time from 1.7 months to 3 years, and the mean age of the participants were 61.09 years old. Based on the Cochrane risk of bias tool, study quality ranged from three to seven out of seven points. Compared with the control group, outcomes including hand strength, range of joint motion, dexterity of hands, arm function, and quality of life were significantly superior with music-supported therapy. Five studies reported improved dexterity of hands, and one study reported the improvement of range of motion and strength of patients' hands, which supported the therapy has positive effects on patients' hand function and improving their quality of life after the therapy. The therapy ranged over a period of 4–8 weeks, with an average duration of 30 min/session and an average of three times per week.

Conclusion: Based on the results, music-supported therapy could be a useful treatment for improving hand function and activities of daily living in patients with stroke, especially for patients within 6 months after stroke. However, the low certainty of evidence downgrades our confidence to practice in hospital. More and more randomized controlled trials and larger sample sizes are required for a deeper review.

Keywords: music supported therapy, hand function, stroke—diagnosis, systematic review, randomized controlled trial

OPEN ACCESS

Edited by:
Vida Abedi,
Geisinger Health System,
United States

Reviewed by:
Ingrid Poulsen,
Rigshospitalet, Denmark
Shima Shahjouei,
Geisinger Health System,
United States

****Correspondence:***
Zu-Lin Dou
douzul@163.com

Specialty section:
This article was submitted to Stroke, a section of the journal Frontiers in Neurology

Received: *13 December 2020*
Accepted: *06 April 2021*
Published: *25 May 2021*

Citation:
Huang W-H, Dou Z-L, Jin H-M, Cui Y, Li X and Zeng Q (2021) The Effectiveness of Music Therapy on Hand Function in Patients With Stroke: A Systematic Review of Randomized Controlled Trials. Front. Neurol. 12:641023. doi: 10.3389/fneur.2021.641023

INTRODUCTION

Stroke is believed to affect more than two million patients annually in China and is one of the most common causes of hand function impairment in middle-aged as well as elderly people. The stroke symptoms may include numbness and weakness in the affected arms and cause a loss of coordination and dexterity (1, 2). Although most of the function can be restored with rehabilitation, ranging up to 79%, the recovery of functional problems of the hand left after stroke is not as satisfactory (3). It has been estimated that ∼67% of stroke survivors are still unable to use the affected hand 4 years after the onset of stroke (4, 5). Therefore, rediscovering the potential of hand function and improving the quality of life is of great value to stroke patients.

The most commonly used conventional treatment for hand function problems include constraint-induced movement therapy (CIMT), mirror therapy, virtual reality, and music-supported therapy (MST) (6–8). MST for hand function is usually achieved by playing the instructions. The movement patients conduct during playing the piano or grasping drumsticks can facilitate the coordination of hands, strengthen the power of grasp of the impaired hand (9). The aim of the MST is to improve the function of the upper limbs and to provide appropriate stimulation through real-time auditory feedback. Studies have shown that after a 4-week MST program, the hand mobility, fluency, and speed of stroke patients can improve during the test. Besides, the sensory stimulation brought by music can induce functional recovery in damaged hemispheres.

Through the combination of music and movement, MST uses continuous movement and sensory input to enable the patient's central nervous system to re-establish new synaptic connections to the greatest extent possible, thereby creating new neuromotor pathways. Functional magnetic resonance imaging (fMRI) shows that the blood flow of the damaged area of the brain increases when receiving stimulation from MST, which can help repair the cerebral cortex caused by cerebral hemorrhage or cerebral infarction (10). Especially, when patients with high muscle tension caused by stroke, MST can relieve high muscle tension and increase the ability of fingers to move freely. Brain plasticity is associated with treatment-induced recovery, which helps the patient to repair after the brain is damaged (11). When stroke patients participate in MST, they need to process information from multiple senses at the same time, including auditory, visual, and sensorimotor information, which is transmitted from the auditory system to the premotor cortex (PMC), thereby adjusting the top-down output (12). However, a major current focus in MST is to evaluate how does MST works and how does MST helps patients with gait problems (13). Few researchers have addressed the problem of MST improving the hand function of stroke patients.

Up to now, the effect of MST on the recovery of hand function during rehabilitation has not been gone through systematically yet. Therefore, we decided to undertake a systematic review to find out evidence that can support that MST has ideal curative effect in the recovery of impaired hand in stroke patients. The review sheds new light on the therapy for helping patients more effectively and increasing the ability of motor control, especially the hands so that they can finish the daily life task by themselves. One of the main challenges is that we need to search the randomized controlled trials (RCTs) based on MST on stroke patients, which are the gold standard for effectiveness research (14). Our systematic literature review solves the PICOS question, "Does MST can help stroke patients improve their hand function and increase the quality of life?" The answers may provide new thinking for occupational therapy and determine the effectiveness of the MST.

METHODS

This systematic review has been reported according to the Preferred Reporting Items for Systematic Review (PRISMA) statement. A protocol for this review was not registered prospectively.

Search Strategy

A systematic literature search of the following electronic databases was performed: Cumulative Index to Nursing and Allied Health Literature (CINAHL), MEDLINE, PubMed, Embase, Music Index, and Google Scholar. Databases were searched using a combination of the following keywords considering diagnose, therapy, and outcome. Diagnosis includes stroke, ischemic, and hemorrhagic. Exercise therapy includes music, rhythm, music therapy, MST, music movement therapy, neurologic music therapy (NMT), and neuroscience. The outcome includes neurologic assessment, physical and cognitive disability, strength and dexterity of the hands, gross mobility of hands, and the assessment of quality of life. MEDLINE was searched using MeSH headings which included dystonic disorders, neurofeedback/behavior therapy/exercise therapy, and recovery of function/treatment outcome. Additional articles were identified from reference lists of retrieved articles. The search was conducted in August 2020.

Eligibility Criteria

Studies were included if they met the following PICOS criteria:

(1) Population: The study population included adults (18 years old or older) diagnosed with ischemic or hemorrhagic stroke which was needed to be confirmed by computer tomography or magnetic resonance imaging and by diagnostic guidelines updated by the American Heart Association/American Stroke Association. The patients presenting with the Fugl–Meyer Assessment—upper extremity (FMA-UE) results less than 46 points.
(2) Intervention: Interventions had to be sound based, including music listening or listening to rhythmic sequences (MLI or RAS, which could be performed by various instruments, e.g., metronome, synthesizer). At least one group of participants had to perform a task in this condition.
(3) Controls: A similar motor act had to be performed without listening to music or rhythmic sequences (control intervention).

(4) Outcome: Outcome measures had to assess hand function in a biomechanical manner (e.g., fine motor and gross mobility, strength and dexterity of the hands, functional movements of hands, muscle activity, or muscle-related assessment).
(5) Study design: All designs should be randomized controlled trials.
(6) Time: Articles published between January 2010 and August 2020.

Studies were excluded from a review when studies were not written in English and if less than half of the participants were musicians. In addition, altered auditory and sensory feedback strategies and studies assessing the combined effects of neuromuscular re-education and transcranial direct current stimulation were also excluded from the review.

Study Selection

The screening procedure was performed by two independent researchers. To collect potentially relevant studies, eligibility was screened based on title and abstract based on the provided inclusion and exclusion criteria described above. Full texts were retrieved and evaluated based on the same eligibility criteria. Afterward, full texts were gathered and evaluated on the previously set inclusion criteria. Reference lists were manually screened to identify additional relevant studies. Results between the two reviewers were compared, in situations where two reviewers were unable to come to an agreement, we took the original articles back together to solve the problem. See **Figure 1**.

Data Extraction and Analysis

Data were extracted and documented using an extraction form developed to identify relevant information. Details recorded from each reference included the author's background and discipline, participants, intervention (follow-up), control/comparison, outcome measures, and results. Data and information were extracted by one reviewer and checked for accuracy by a second reviewer. The patients' characteristics and intervention detail of each study were summarized in **Table 1**, reflecting the heterogeneity of the included studies.

The risk of bias assessment was based on the handbook of Cochrane (5.1 version). To assess bias, two reviewers independently followed the steps to choose low risk and high risk; if a question was impossible to answer because the original article did not specify it or it was unclear, we chose "unclear." Reviewers assessed selection (random sequence generation and allocation concealment), performance (blinding of participants and personnel), detection (blinding of outcome assessors), attrition (incomplete outcome data), reporting (selective reporting), and other sources of bias. If there is a disagreement between two reviewers, a third reviewer will solve it. See **Figures 2**, **3**.

RESULTS

Study Characteristics

In order to identify the review, a total of 269 articles were retrieved in this search. Two hundred twelve articles were

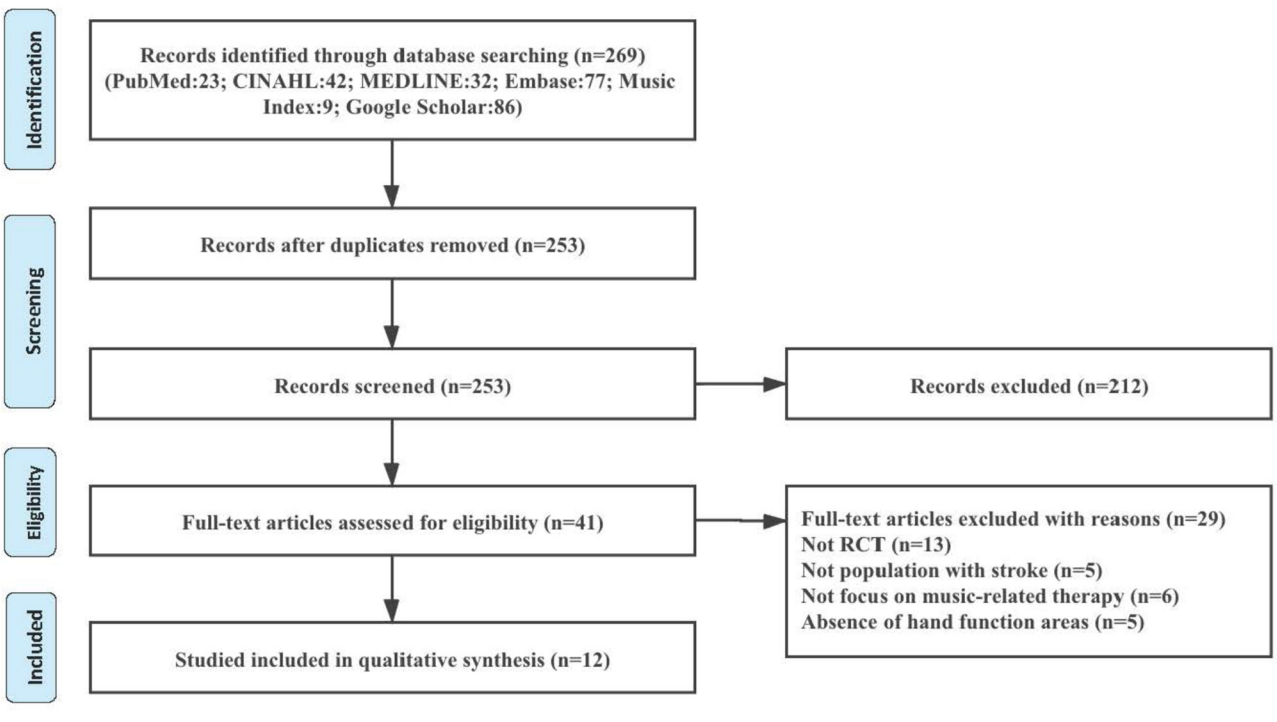


FIGURE 1 | A summary of the PRISMA flow of the study selection process.

TABLE 1 | Summary of the included studies and the detail of intervention and measurement.

Reference	Participants	Intervention (follow-up)	Control/comparison	Outcome measures	Results
Raglio et al. (15)	$N = 38$; age range: 54–89	The standard of care and relational active music therapy approach ($n = 19$) Repeat exercise: 20 sessions lasting 30 min each, 3-weekly	Only standard of care, including physiotherapy and occupational therapy ($n = 19$)	Measured at baseline and at the end of treatment Neurologic: It-NIHSS Physical and cognitive disability: FIM Strength/dexterity of the hands: the Grip-Pinch Dynamometric Test and the Nine-hole Peg Test Gross mobility: TUG Psychological traits and quality of life: HADS and MQOL-It Video-recorded sessions: MTRS	Neurologic: no significant difference between groups Physical and cognitive disability: improved both in experimental and control groups ($p = 0.001$) Strength/dexterity of the hands: the amount of left paretic patients ($n = 5$) improved more than control groups ($n = 2$) Gross mobility: improved both in experimental and control groups ($p = 0.032$) Psychological traits and quality of life: a decrease of anxiety and depression and a significant positive trend
Street et al. (16)	$N = 11$; age range: 53–67	Play acoustic musical instruments and/or iPads with touch screen musical instruments ($n = 6$) Repeat exercise: 20–30 min a session, twice weekly for 6 weeks	Received no intervention ($n = 5$)	Measured at baseline and at 6-, 9-, 15-, and 18-week follow-up Arm function: the action research arm test and the 9-hole peg test	Arm function: no significant difference between groups
Street et al. (9)	$N = 14$; age range: 18–90	Therapeutic instrumental music performance therapy ($n = 7$) Repeat exercise: twice weekly for 6 weeks	Received no intervention ($n = 7$)	Measured at baseline and at 6-, 9-, 15-, and 18-week follow-up Arm function: the action research arm test and the 9-hole peg test Electroencephalography Recording	Arm function: no significant difference between groups Electroencephalography recording: no significant difference between groups
Grau-Sanchez et al. (17)	$N = 39$; age range: 54–92	The regular therapy and extra sessions to play a keyboard and an electronic drum set ($n = 20$) Repeat exercise: 20 individual sessions, 30 min each, 5 sessions/week for 4 weeks	Extra time for exercises for the upper extremity based on the regular therapy ($n = 20$)	Measured at baseline, after the intervention, and at a 3-month follow-up Functional movements: the action research arm test Motor outcomes: FMA and grip strength Fine dexterity: the 9-hole peg test and BBT Activities of daily living: CAHAI Working memory and attention: the digit span subtest from the Wechsler Adult Intelligence Scale III, response inhibition by the Stroop task and processing speed and mental flexibility by the trail-making test Verbal memory: RAVLT and the story recall from the Rivermead behavioral memory test Mood outcomes: the Profile of Mood States, the Beck Depression Inventory Scale, the Positive and Negative Affect Scale, and the Apathy Evaluation Scale. QoL outcomes: the Stroke-Specific QoL Scale and health-related QoL with the health survey questionnaire SF36.	Functional movements: significantly improved functional performance score in the MST group compared with CT group (mean ± SD, standard treatment with exercise, 9.8 ± 7.9, vs. exercise, 6.7 ± 7.9; $p < 0.001$) Motor outcomes: no significant difference between groups Fine dexterity: no significant difference between groups Activities of daily living: no significant difference between groups The cognitive outcomes: no significant difference between groups QoL outcomes: significantly improved in the MST group from baseline to posttreatment compared with CT group (MST group of $t_{(18)} = -2.23$, $p < 0.05$, $d = 0.54$ vs. CT group of no improvements) Mood outcomes: no significant differences between groups in the change scores

(Continued)

TABLE 1 | Continued

Reference	Participants	Intervention (follow-up)	Control/comparison	Outcome measures	Results
Jun et al. (18)	$N = 30$; age range: 54–93	Received music and movement therapy ($n = 15$). Repeat exercise: 1 h/session, 3 times/week for 8 weeks	Received routine care ($n = 15$)	Measured at baseline and at 8-week follow-up Physical functions: range of joint motion Muscle strength: Medical Research Council scale Activities of daily living: K-MBI Mood state: the Korean version of the Profile of Mood States Brief instrument Depression: CES-D	Physical functions: no significant difference between groups Muscle strength: no significant difference between groups Activities of daily living: no significant difference between groups Mood state: the score of experimental group members improved when compared with that of the control group ($t = 1.818$, $p = 0.040$) Depression: no significant difference between groups
Van Vugt et al. (19)	$N = 34$; age range: 30–75	Received its sounds after a random delay sampled from a flat distribution between 100 and 600 ms when the patients play the piano ($n = 19$) Repeat exercise: 10 sessions of half an hour	Received the its sounds immediately when the patients play the piano ($n = 15$)	Measured at baseline, after the intervention Fine motor control: the 9-hole peg test Finger tapping measurements: a triaxial accelerometer (ADXL 335) Mood measurements: POMS	Fine motor control: significantly improved fine motor score in the jitter group compared with normal group (mean ± SD, the average improvement of jitter group, 14 ± 53.6 vs. normal, 3.8 ± 17.9; $p < 0.001$) Tapping speed: no significant difference between groups Tapping variability: no significant difference between groups Mood measurements: no significant difference between groups
Fotakopoulos and Kotlia (20)	$N = 65$; age range: 71–79	A music group (MG) (daily listening to experiential/traditional music) Repeat exercise: 6 months at a frequency of four training sessions/week, of 45 min each session	A control group (CG) with no experiential/traditional music therapy (standard care only)	Measured at baseline, after the intervention Cognitive deficits: mMt Performance in activities of daily living: BI CT perfusion: CBF	Cognitive deficits: significantly improved cognitive score in the recovery group compared with no-recovery group (mean ± SD, the recovery group, 26.38 ± 1 vs. no-recovery group, 24.33 ± 2; $p < 0.001$) Performance in activities of daily living: significantly improved ADL score in the recovery group compared with no-recovery group (mean ± SD, the recovery group, 81.92 ± 2 vs. no-recovery group, 76.53 ± 7; $p = 0.007$) CT perfusion: significantly improved in CBF in affected area in the recovery group compared with no-recovery group (mean ± SD, the recovery group, 29.16 ± 4 vs. no-recovery group, 12.27 ± 11; $p < 0.001$)
Bunketorp-Käll et al. (21)	$N = 123$; age range: 56–70.4	Rhythm-and-music therapy (n =41) Horse-riding therapy ($n = 41$) Repeat exercise: 2 times a week for 12 weeks	Control group continue with their regular activities and usual care such as outpatient physiotherapy, occupational therapy, or speech therapy ($n = 41$)	Outcome measures were reported at 0 and 6 months postintervention Hand strength: Grippit	Hand strength: significant differences in the mean changes in right-sided maximum and left-sided final grip force Rhythm-and-music group significantly improved their right-sided maximum grip force(16.41 [95% CI, 5.65–27.17]) and left-sided final grip force (17.26 [95% CI, 6.19–28.33]) compared with controls (−1.29 [95% CI, −7.99 to 5.41]) (0.55 [95% CI, −7.07 to 8.17]; $p = 0.015$ and 0.042, respectively); The left-sided improvements were sustained at the 6-month follow-up ($p = 0.011$).

(Continued)

TABLE 1 | Continued

Reference	Participants	Intervention (follow-up)	Control/comparison	Outcome measures	Results
Tong et al. (12)	$N = 33$; age range: 34–64.9	Audible music group (MG) includes conventional rehabilitation treatments and extra sessions of audible musical instrument training ($n = 15$) Repeat exercise: 20 extra sessions over 4 weeks	Mute music group (CG) includes conventional rehabilitation treatments and extra sessions of "mute" musical instrument training ($n = 18$) Repeat exercise: 20 extra sessions over 4 weeks	Measured at baseline, after the intervention Motor function: WMFT, FMA	Motor functions of upper limbs: significant improvements Significant differences in the WMFT were found between the two groups (WMFT-quality: $p = 0.025$; WMFT-time: $p = 0.037$) but not in the FMA ($p = 0.448$). Subjects in MG demonstrated greater improvement than those in CG.
Schneider et al. (22)	$N = 77$; age range: 41.2–68	Music-supported therapy in addition to conventional therapy ($n = 32$). Repeat exercise: 30 min each unit in duration, totally 27.4 units, over 3 weeks	Conventional treatment only ($n = 30$), Without specific additional selection criteria ($n = 15$) standard therapies (physical therapy and individual occupational therapy) 30 min each unit in duration. TG: 28.0 units over 3 weeks CG: 27.2 units over 3 weeks	Measured at baseline, 3-week intervention Motor functions: BBT, the 9-hole peg test, action research arm test, arm paresis score Motor test/parameter: frequency (FREQ), Number of inversions of velocity profiles, Average maximum angular velocity in °/s	BBT, the 9-hole peg test, action research arm test, and arm paresis score: significant improvements in groups TG and MG. Conventional physiotherapy in CG did not produce an improvement, differences between MG, CG, and TG were highly significant, $F(2, 66) = 6.66$, $p = 0.002$. BBT: MG increased the number of cubes grasped by around 10/min. Differences between MG, CG, and TG were highly significant, $F_{(2,74)} = 57.08$, $p < 0.001$. FREQ: Increase in MG but not TG and CG
Fujioka et al. (23)	$N = 29$; age range: 54.3–64.2	Music-supported therapy used an electronic keyboard and a series of eight electronic drum pads ($n = 14$). Repeat exercise: 30 h of training over 10 weeks	Conventional physical training ($n = 14$) Repeat exercise: 30 h of training over 10 weeks	Measured at baseline, after 5 weeks, after 10 weeks, and 3 months after training completion. Arm and hand subsections of the CMSA Impairment Inventory, action research arm test, BBT	CMSA: Both showed only minor changes over the time course of treatment, hand score was improved at the post 2 time point compared with pre [$t_{(27)} = -2.27$, $p = 0.031$]. A tendency for such improvement was found for the MST group [$t_{(13)} = -1.88$, $p = 0.082$]. The improvement in the GRASP group was not significant. Action research arm test: in the MST group, the decrease between pre and post 2 time points approached significance [$t_{(13)} = 2.10$, $p = 0.056$]. BBT: not to analyze, as eight participants were unable to perform the test at any time point using their affected hand.
Bunketorp-Käll et al. (24)	$N = 123$; age range: 56–70.4	Rhythm-and-music therapy ($n = 41$) Horse-riding therapy ($n = 41$) Repeat exercise: 2 times a week for 12 weeks.	Control group continue with their regular activities and usual care such as outpatient physiotherapy, occupational therapy, or speech therapy ($n = 41$).	Measured at baseline, after the intervention Motor function: Modified Motor Assessment Scale.	Modified Motor Assessment Scale: The MST group did not produce any immediate gains. 6 months 31 post-intervention, the MST group performed better with respect to time; −0.75 s [95% CI, −1.36 to −0.14]; ($p = 0.035$)

It-NIHSS, the National Institutes of Health Stroke Scale; FIM, the Functional Independence Measure; HADS, the Hospital Anxiety and Depression Scale; MQOL-It, the Italian version of McGill Quality-of-Life Questionnaire; TUG, the Timed Up and Go Test; MTRS, the Music Therapy Rating Scale; mMt, the mini mental test; BI, the Barthel Index; CBF, cerebral blood flow; CMSA, the Chedoke–McMaster Stroke; CAHAI, the Chedoke Arm and Hand Activity Inventory; RAVLT, the Rey auditory verbal learning test; K-MBI, Korean-modified Barthel index; CES-D, The Center for Epidemiologic Studies Depression Scale; POMS, the Profile of Mood States; WMFT, Wolf motor function test; FMA, Fugl–Meyer assessment; BBT, Box and Block test.

considered for screening, and 41 full-text articles were excluded because the study did not focus on the stroke patients ($n = 5$), the type of study did not meet our inclusion criteria ($n = 13$), the abstract of study did not focus on the music-supported therapy ($n = 6$), and hand function areas ($n = 5$). Twelve articles were included in the systematic review. See **Table 1**.

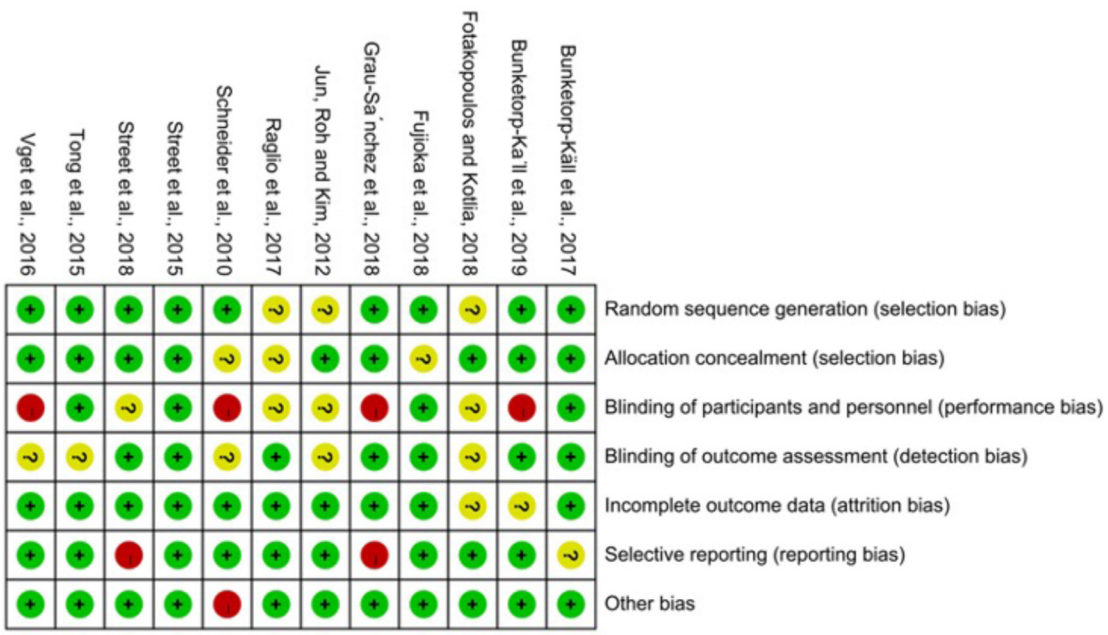


FIGURE 2 | Risk of bias summary.

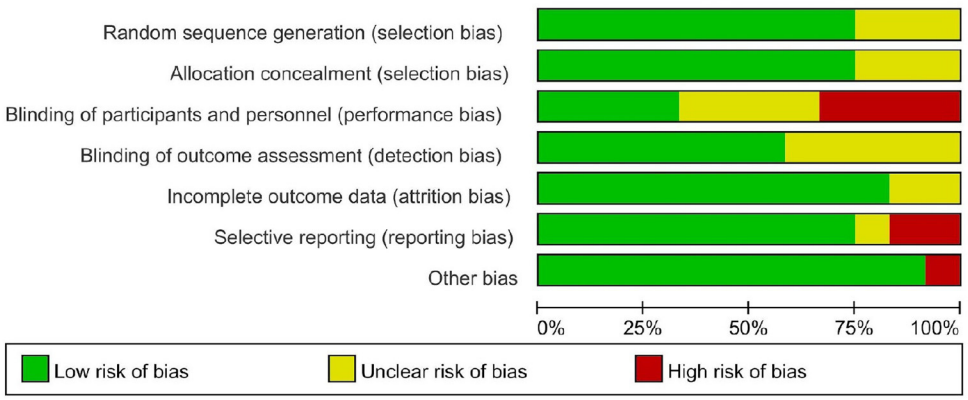


FIGURE 3 | Risk of bias graph.

Baseline of Patients

Totally, the data involved 598 stroke patients of mixed gender population. The included studies had the gender distribution as follows: 253 females and 345 males. The mean age of the participants was 61.09 years old (SD = 11.43), ranging from 48.6 to 76.02 years old (9, 12, 16, 17, 19, 21–24). All subjects suffered from stroke from 1.7 months to 3 years before exposed to the MST intervention. Four articles did not provide the time of poststroke of the population in their articles. Based on the data of demographics, the percentage of ischemic stroke was 74.53% (199 participants) and the percentage of hemorrhagic stroke was 25.47% (68 participants).

Risk of Bias of Included Studies

The items' random sequence generation, allocation concealment, incomplete outcome data, and other sources of bias were assessed as low risk of bias in most of the included studies (9, 12, 16, 17, 19, 21–24). Blinding of participants and personnel scored high risk of bias or unclear risk in half of the included studies, which is inherent to the intervention (15–20, 22, 24). There are some studies that the blinding of outcome assessment was not

described clearly so that they were scored unclear in the detection bias (12, 19, 20, 22).

Intervention Characteristics

Study objectives varied greatly; music was used to influence grip strength (15, 17, 18, 21), range of joint motion (18), dexterity of the hands (9, 15–17, 19, 22), and arm function (9, 12, 16, 17, 22–24), and demonstrated the changes of activities of daily living (17, 18, 20) and quality of life (15, 17, 23). To determine effectiveness, musical interventions were compared with blank control group (9, 16), standard of care, regular activities (15, 18, 20, 21, 23), conventional rehabilitation treatments (12, 17, 22, 24), and mute music training (19, 24). Conventional rehabilitation treatments such as passive mobilization, stretch and progressive resistance exercises, and task-specific training (17, 22), mute music training such as using a sponge or custom-made pad on the musical instrument to inhibit or delay patients from hearing sounds during the training (19, 24).

Nine trials required participants played rhythmical-melodic musical instruments or digital music equipment, such as xylophones, glockenspiels, drums, bongos, ethnic percussion, piano, iPads with touch screen musical instruments (9, 12, 15–19, 22, 23). The methods included free interactions between patients and music therapists and sang a song while playing musical instruments. Three trials employed prepared music and daily listening to experiential/traditional music (20, 21, 24); two of three trials involved structured R-MT combining listening to music while performing coordinated rhythmic sequences and cognitively (21, 24). Participants were trained to begin with motions of the unaffected side and then the affected upper extremity following a modular therapy regime with a stepwise increase of complexity (17, 18).

Sessions were offered one-to-one with individual participants at home or clinical setting (9, 12, 16, 17, 19, 22, 23) or to small groups (15, 18, 20, 21, 24). Within these sessions, there was consistent treatment "dosage," lasting a single session of 30 or 45 min each. The duration of interventions was variability ranging from 20 sessions over 3 weeks (15) to four sessions weekly over 6 months (20). Delivery of musical interventions was predominantly provided by experienced music therapists or licensed therapists.

Qualitative Synthesis: Outcome

The included articles show that MST has a better effect on stroke patients in the acute and recovery phases when compared with the group that received conventional rehabilitation. The effect on stroke patients includes many aspects. See **Table 2**.

Hand Strength

One trial showed significant differences in the mean changes in both sided maximum and one-sided final grip force, as measured with Grippit. Some improvements were also sustained at the 6-month follow-up (21). In the grip-pinch test, one trial showed that the strength of the nondominant hand significantly increased (15). Two trials showed no significant differences were observed for the group comparisons after treatment or at follow-up (17, 18).

Range of Joint Motion

Only one trial showed that the ROM (shoulder, elbow joint, and hip joint flexion) on the affected side of subjects in the experimental group was increased following the music therapy, whereas the ROM of these joints in the control group either decreased or remained the same. There was a significant increase in shoulder flexion and elbow joint flexion (18).

Dexterity of Hands

Five trials with nine-hole peg test examined fine motor skills of hands improved gradually (9, 15, 16, 19, 22). One trial showed no significant differences were observed for the group comparisons after treatment or at follow-up; however, the within-group analyses revealed that both groups improved after and at follow-up using nie-hole peg test and box and blocks test (17). In the test of finger tapping measurements, there was no significant difference between groups in areas of tapping speed and variability (19).

Arm Function

Significant improvements in motor functions of upper limbs after 4 weeks of treatment; however, we only found differences between the two groups in Wolf motor function test (WMFT-quality: $p = 0.025$; WMFT-time: $p = 0.037$), but found no differences in Fugl–Meyer assessment ($p = 0.448$) (12). Grau-Sanchez's (17) study showed that in 39 included patients, no significant difference between MST groups and control groups was shown after a 4-week treatment. Four trials showed no significant difference between groups in arm function through the action research arm test (9, 16, 17, 23). One of the four trials found the within-group analyses revealed that both groups improved after and at follow-up (17). Another trial also indicated minor changes by the CMSA arm and hand impairment scale (23). However, MST patients in one trial showed a substantial improvement over time compared with other groups of patients in the action research arm test and arm paresis score (22). One trial also did not produce any immediate gains with the Modified Motor Assessment Scale (M-MAS) (24).

Activities of Daily Living

No significant differences in effect using Chedoke Arm and Hand Activity Inventory (CAHAI) or Korean-modified Barthel index (K-MBI) (18, 19). Only one trial showed significantly improved ADL score in the recovery group compared with the no-recovery group (20).

Quality of Life

One trial showed a significant positive trend in quality of life through the Italian version of McGill Quality-of-Life Questionnaire (MQOL-It), but no clinical differences between groups were found (15). There was a significant improvement in the MST group from baseline to posttreatment compared with the conventional treatment groups among other trials (17). Negative effect of affective functions and quality of life was significantly reduced after intervention (23).

TABLE 2 | The result of outcome measures in the included studies.

Outcome measures	Measurements	Results
Muscle strength	Grippit	Improvements were shown at the final of intervention and 6-month follow-up (21)
	Grip-pinch test	Strength of nondominant hand significantly increased (15)
	Medical Research Council scale and grip strength	No significant difference between groups (17, 18)
Range of joint motion	Measuring the ROM of shoulder, elbow joint, and hip joint flexion	Significant increase in shoulder flexion and elbow joint flexion (18)
Dexterity of hands	9-Hole peg test	Improved gradually (9, 15, 16, 19, 22)
	9-Hole peg test and box and blocks test	Both groups improved but no significant differences (17)
	Test of finger tapping measurements	No significant difference between groups (19)
Arm function	Wolf motor function test	Significant differences between the 2 groups (12)
	The action research arm test and arm paresis score	Significant differences between the 2 groups (22)
	The action research arm test	No significant difference between groups (9, 16, 17, 23)
	The Modified Motor Assessment Scale (M-MAS)	No significant difference between groups (24)
Activities of daily living	The Barthel Index	Significant differences between the two groups (20)
	Chedoke Arm and Hand Activity Inventory (CAHAI) or Korean-modified Barthel index (K-MBI)	No significant difference between groups (18, 19)
Quality of life	The Stroke-Specific QoL Scale and health-related QoL with the health survey questionnaire SF36	Significant differences between the 2 groups (17)
	Italian version of McGill Quality-of-Life Questionnaire (MQOL-It)	No significant difference between groups (15)

DISCUSSION

The purpose of this systematic review was to evaluate the effectiveness of MST on hand function improvement in stroke patients. The result shows that MST can be useful in improving hand function and the quality of life in stroke patients.

The subjects of studies we focus are symptoms of unilateral hemiparesis in ischemic and hemorrhagic stroke patients. Through a systematic review, we found that MST can improve hand function in patients. Especially for patients within 6 months after stroke, MST can significantly improve their hand function with different aspects such as dexterity of hands.

However, according to the included studies, only a few studies have clear results indicating that the grasping ability and the dexterity of the finger are improved (9, 15, 16, 19, 22). For example, one study stated that the hand function was improved for both the MST group and the control group, but there were no significant differences between the two groups when compared (15). MST did not show superiority improvement when compared with conventional therapy. At this point, we conclude that it may be caused by the following reasons. Firstly, low-intensity music-supported therapy is not known to cause an effect on the improvement of a patient's ability. Therefore, according to the results of a systematic review, training for at least 30 min a day and five times a week is the suitable intensity that is proven as effective. Secondly, some assessment scales, such as FMA, are not sensitive enough to assess the difference in the area of hand function before and after treatment, while assessment scales like WMFT can. The reasons may be because the movement of MST is similar to the movement of evaluation so that it causes some bias that impacts the result of the assessment. Therefore, when we select the assessment scale, we should choose an appropriate scale that can have a good sensitivity to identify the differences. Finally, several studies stated that for their research, the sample size was not large enough to obtain a significantly different result, which suggests that we should do a larger experiment to show the effect of music-supported therapy in the future.

Among the included studies, MST has been described to improve hand function. Compared with traditional treatment, the patient mainly improved the ability of priming, timing, trajectory, and muscle force requirements for the movements of the upper limbs (16), and because of the characteristics of MST, such as, noninvasive, low cost, convenient, and effective intervention, it has been widely used in practice and accepted by Chinese occupational therapists. When playing a musical instrument, the hands follow the beat of the music and make corresponding movements, which is a kind of stimulation to the damaged part of the brain (25). The musical stimulation can promote the extensive activation of the patient's central nervous system' functional network, increase blood flow in the brain area, and accept more stimulation from the movement (26, 27). Among the included articles, there are descriptions of the use of violin, piano, and drumming instruments for treatment. When the patient's hand grasps or strikes the instruments, the patient's grasping ability can be trained purposefully. When the stroke patients train with the rhythm of music, it can help muscle contraction to become more active, so that the sense of participation, rhythm, and speed in the exercise will be more effective (28).

MST can have an improved performance in activities of daily living and enhance the quality of life (17). Furthermore, MST decreases a patient's depression and helps deal with the emotional stress caused by sudden and severe neurological diseases (19). MST helps the patient facilitate their emotions and share their feelings. At the same time, music also plays an important role in motivating patients and stimulating their inner motivation. MST

increases patients' motivation due to happiness and intrinsic motivation. We can help them acquire a new skill as well as a hobby so that they can actively participate in MST and enjoy the fun of playing.

LIMITATIONS

The systematic review has some limitations that should be given attention. First, the sample size in the study is not big enough. Studies with larger sample sizes are needed in the coming future. Second, when conducting a systematic search of the literatures, only studies written in English were searched. It is possible that we missed articles of high significance written in other languages. Third, there was a lack of good way of randomization in the included studies, especially with the blinding of patients. It is clear that patients had knowledge of whether they were receiving MST or not, so there was an inability to avoid a placebo effect as a result of receiving MST. Higher-quality articles are needed to provide ideas on how to avoid the placebo effect, and it will be applied for future research development.

CONCLUSION

The review included data from 12 randomized controlled trials to explore the effectiveness of MST on the hand function for patients after stroke. Based on current evidence, this study demonstrated that MST can improve hand function and enhance a patient's quality of life. The rhythm and auditory feedback play a vital part in the treatment of MST. However, more well-described randomized controlled trials are required to prove its efficacy.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

W-HH, Z-LD, H-MJ, YC, XL, and QZ worked together to complete the manuscript. Z-LD and H-MJ assisted W-HH in document retrieval and screening. YC provided statistical assistance and support. XL and QZ provided opinions on grammar and rhetoric. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Key R&D Program of Guangdong Province, China under grant 2018B030339001.

REFERENCES

1. Luo Z, Zhou Y, He H, Lin S, Zhu R, Liu Z, et al. Synergistic effect of combined mirror therapy on upper extremity in patients with stroke: a systematic review and meta-analysis. *Front Neurol.* (2020) 11:155. doi: 10.3389/fneur.2020.00155
2. Wu S, Wu B, Liu M, Chen Z, Wang W, Anderson CS, et al. Stroke in China: advances and challenges in epidemiology, prevention, and management. *Lancet Neurol.* (2019) 18:394–405. doi: 10.1016/S1474-4422(18)30500-3
3. Behrendt F, Rizza JC, Blum F, Suica Z, Schuster-Amft C. German version of the Chedoke McMaster arm and hand activity inventory (CAHAI-G): intra-rater reliability and responsiveness. *Health Qual Life Outcomes.* (2020) 18:1–9. doi: 10.1186/s12955-020-01499-6
4. Samuelkamaleshkumar S, Reethajanetsureka S, Pauljebaraj P, Benshamir B, Padankatti S, David J. Mirror therapy enhances motor performance in the paretic upper limb after stroke: a pilot randomized controlled trial. *Arch Phys Med Rehabil.* (2014) 95:2000–5. doi: 10.1016/j.apmr.2014.06.020
5. Lim KB, Lee HJ, Yoo J, Yun HJ, Hwang HJ. Efficacy of mirror therapy containing functional tasks in poststroke patients. *Ann Rehabil Med.* (2016) 40:629–36. doi: 10.5535/arm.2016.40.4.629
6. Kwakkel G, Veerbeek JM, van Wegen EE, Wolf SL. Constraint-induced movement therapy after stroke. *Lancet Neurol.* (2015) 14:224–34. doi: 10.1016/S1474-4422(14)70160-7
7. Park JY, Chang M, Kim KM, Kim HJ. The effect of mirror therapy on upper-extremity function and activities of daily living in stroke patients. *J Phys Ther Sci.* (2015) 27:1681–3. doi: 10.1589/jpts.27.1681
8. Zhang Y, Cai J, Zhang Y, Ren T, Zhao M, Zhao Q. Improvement in stroke-induced motor dysfunction by music-supported therapy: a systematic review and meta-analysis. *Sci Rep.* (2016) 6:38521. doi: 10.1038/srep38521
9. Street AJ, Magee WL, Odell-Miller H, Bateman A, Fachner JC. Home-based neurologic music therapy for upper limb rehabilitation with stroke patients at community rehabilitation stage-a feasibility study protocol. *Front Hum Neurosci.* (2015) 9:480. doi: 10.3389/fnhum.2015.00480
10. Särkämö T, Soto D. Music listening after stroke: beneficial effects and potential neural mechanisms. *Ann N Y Acad Sci.* (2012) 1252:266–81. doi: 10.1111/j.1749-6632.2011.06405.x
11. Cramer SC, Sur M, Dobkin BH, O'Brien C, Sanger TD, Trojanowski JQ, et al. Harnessing neuroplasticity for clinical applications. *Brain.* (2011) 134:1591–609. doi: 10.1093/brain/awr039
12. Tong Y, Forreider B, Sun X, Geng X, Zhang W, Du H, et al. Music-supported therapy (MST) in improving post-stroke patients' upper-limb motor function: a randomised controlled pilot study. *Neurol Res.* (2015) 37:434–40. doi: 10.1179/1743132815Y.0000000034
13. Ghai S, Ghai I. Effects of (music-based) rhythmic auditory cueing training on gait and posture post-stroke: a systematic review and dose-response meta-analysis. *Sci Rep.* (2019) 9:2183. doi: 10.1038/s41598-019-38723-3
14. Hariton E, Locascio JJ. Randomised controlled trials-The gold standard for effectiveness research. *BJOG.* (2018) 125:1716. doi: 10.1111/1471-0528.15199
15. Raglio A, Zaliani A, Baiardi P, Bossi D, Sguazzin C, Capodaglio E, et al. Active music therapy approach for stroke patients in the post-acute rehabilitation. *Neurol Sci.* (2017) 38:893–7. doi: 10.1007/s10072-017-2827-7
16. Street AJ, Magee WL, Bateman A, Parker M, Odell-Miller H, Fachner J. Home-based neurologic music therapy for arm hemiparesis following stroke: results from a pilot, feasibility randomized controlled trial. *Clin Rehabil.* (2018) 32:18–28. doi: 10.1177/0269215517717060
17. Grau-Sanchez J, Ramos-Escobar N, Sierpowska J, Rueda N, Susana R, Rifà X, et al. Music-supported therapy in the rehabilitation of subacute stroke patients: a randomized controlled trial. *Ann N Y Acad Sci.* (2018) 61:e191. doi: 10.1016/j.rehab.2018.05.438
18. Jun EM, Roh YH, Kim MJ. The effect of music-movement therapy on physical and psychological states of stroke patients. *J Clin Nurs.* (2013) 22:22–31. doi: 10.1111/j.1365-2702.2012.04243.x
19. Van Vugt FT, Kafczyk T, Kuhn W, Rollnik JD, Tillmann B, Altenmüller E. The role of auditory feedback in music-supported stroke rehabilitation: a single-blinded randomised controlled intervention. *Restor Neurol Neurosci.* (2016) 34:297–311. doi: 10.3233/RNN-150588

20. Fotakopoulos G, Kotlia P. The value of exercise rehabilitation program accompanied by experiential music for recovery of cognitive and motor skills in stroke patients. *J Stroke Cerebrovasc Dis*. (2018) 27:2932–9. doi: 10.1016/j.jstrokecerebrovasdis.2018.06.025
21. Bunketorp-Käll L, Lundgren-Nilsson Å, Samuelsson H, Pekny T, Blomvé K, Pekna M, et al. Long-term improvements after multimodal rehabilitation in late phase after stroke: a randomized controlled trial. *Stroke*. (2017) 48:1916–24. doi: 10.1161/STROKEAHA.116.016433
22. Schneider S, Müünte T, Rodriguez-Fornells A, Sailer M, Altenmüüller E. Music-supported training is more efficient than functional motor training for recovery of fine motor skills in stroke patients. *Music Percept*. (2010) 27:271–80. doi: 10.1525/mp.2010.27.4.271
23. Fujioka T, Dawson DR, Wright R, Honjo K, Chen JL, Chen JJ, et al. The effects of music-supported therapy on motor, cognitive, and psychosocial functions in chronic stroke. *Ann N Y Acad Sci*. (2018) 1423:264–74. doi: 10.1111/nyas.13706
24. Bunketorp-Käll L, Pekna M, Pekny M, Blomstrand C, Nilsson M. Effects of horse-riding therapy and rhythm-and music based therapy on functional mobility in late phase after stroke. *NeuroRehabilitation*. (2019) 45:483–92. doi: 10.3233/NRE-192905
25. Neugebauer CT, Serghiou M, Herndon DN, Suman OE. Effects of a 12-week rehabilitation program with music and exercise groups on range of motion in young children with severe burns. *J Burn Care Res*. (2008) 29:939–48. doi: 10.1097/BCR.0b013e31818b9e0e
26. Jäncke L. Music, memory and emotion. *J Biol*. (2008) 7:1–5. doi: 10.1186/jbiol82
27. Karmonik C, Brandt A, Anderson JR, Brooks F, Lytle J, Silverman E, et al. Music listening modulates functional connectivity and information flow in the human brain. *Brain Connect*. (2016) 6:632–41. doi: 10.1089/brain.2016.0428
28. Biasutti M, Mangiacotti A. Assessing a cognitive music training for older participants: a randomised controlled trial. *Int J Geriatr Psychiatry*. (2018) 33:271–8. doi: 10.1002/gps.4721

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neurology

ORIGINAL RESEARCH
published: 12 July 2021
doi: 10.3389/fneur.2021.691057

Diagnosis of Acute Central Dizziness With Simple Clinical Information Using Machine Learning

Bum Joon Kim[1†], Su-Kyeong Jang[2†], Yong-Hwan Kim[2,3], Eun-Jae Lee[1], Jun Young Chang[1], Sun U. Kwon[1], Jong S. Kim[1] and Dong-Wha Kang[1]*

[1] Department of Neurology, Asan Medical Center, University of Ulsan, College of Medicine, Seoul, South Korea, [2] Asan Medical Center, Asan Institute for Life Sciences, Seoul, South Korea, [3] Nunaps Inc., Seoul, South Korea

Background: Acute dizziness is a common symptom among patients visiting emergency medical centers. Extensive neurological examinations aimed at delineating the cause of dizziness often require experience and specialized training. We tried to diagnose central dizziness by machine learning using only basic clinical information.

Methods: Patients were enrolled who had visited an emergency medical center with acute dizziness and underwent diffusion-weighted imaging. The enrolled patients were dichotomized as either having central (with a corresponding central lesion) or non-central dizziness. We obtained patient demographics, risk factors, vital signs, and presentation (non-whirling type dizziness or vertigo). Various machine learning algorithms were used to predict central dizziness. The area under the receiver operating characteristic curve (AUROC) was measured to evaluate diagnostic accuracy. The SHapley Additive exPlanations (SHAP) value was used to explain the importance of each factor.

Results: Of the 4,481 visits, 414 (9.2%) were determined as central dizziness. Central dizziness patients were more often older and male and had more risk factors and higher systolic blood pressure. They also presented more frequently with non-whirling type dizziness (79 vs. 54.4%) than non-central dizziness. Catboost model showed the highest AUROC (0.738) with a 94.4% sensitivity and 31.9% specificity in the test set ($n = 1,317$). The SHAP value was highest for previous stroke presence (mean; 0.74), followed by male (0.33), presentation as non-whirling type dizziness (0.30), and age (0.25).

Conclusions: Machine learning is feasible for classifying central dizziness using demographics, risk factors, vital signs, and clinical dizziness presentation, which are obtainable at the triage.

Keywords: dizziness, vertigo, machine learning, stroke, vertebrobasilar insufficiency

OPEN ACCESS

Edited by:
Ramin Zand,
Geisinger Health System,
United States

Reviewed by:
Anita Ante Arsovska,
Saints Cyril and Methodius University of Skopje, North Macedonia
Prasanna Tadi,
Geisinger Wyoming Valley Medical Center, United States
Ming Wang,
Penn State Milton S. Hershey Medical Center, United States

***Correspondence:**
Dong-Wha Kang
dwkang@amc.seoul.kr

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Stroke, a section of the journal Frontiers in Neurology

Received: *05 April 2021*
Accepted: *04 June 2021*
Published: *12 July 2021*

Citation:
Kim BJ, Jang S-K, Kim Y-H, Lee E-J, Chang JY, Kwon SU, Kim JS and Kang D-W (2021) Diagnosis of Acute Central Dizziness With Simple Clinical Information Using Machine Learning. Front. Neurol. 12:691057. doi: 10.3389/fneur.2021.691057

INTRODUCTION

Acute dizziness and vertigo are common symptoms presented by patients admitted to emergency medical centers (EMCs) (1). Though dizziness is usually attributable to benign etiologies originating from peripheral causes, 5% of acute dizziness may be caused by cerebrovascular issues (2). Acute dizziness and vertigo are the most common presenting symptoms of vertebra-basilar ischemia (3), which shows a stepwise deterioration of poor prognosis when the diagnosis is inappropriately delayed (4).

Unfortunately, because there is no standard test or biomarker that can be used for the confirmatory diagnosis of central dizziness, verification of the etiology remains challenging. Many efforts have been made to distinguish central dizziness from peripheral dizziness, especially those utilizing extensive neurological examinations (5). However, to some extent, misdiagnosis stems from an overreliance on negative neurological examinations (2). Approximately 11% of medial posterior-inferior cerebellar artery infarction patients were shown to present with isolated vertigo, abnormal ocular manifestations, and imbalance (6). However, interpretation of the oculomotor findings often requires further examinations with experienced and specialized neuro-ophthalmology staff. While acute dizziness and vertigo are very commonly observed clinical symptoms, which most physicians, not only specialists, may encounter daily, misdiagnosis can lead to devastating results (7).

Therefore, clinicians require a simple and widely applicable method with high sensitivity that can significantly reduce misdiagnosis of central dizziness. Machine learning (ML) has previously been used and has shown an acceptable performance in predicting the characteristics and prognosis of ischemic stroke (8–11). Several studies have shown that ML can be used to analyze nysagmogram or postulography videos to diagnose the causes of dizziness, which still needs equipment to measure the nystagmus or posture (12, 13). Here, we used ML techniques to diagnose isolated acute dizziness patients visiting EMCs. As such, we used only simple clinical information to delineate the central causes of dizziness from peripheral causes. Additionally, we aimed to examine the feature importance of the ML model and understand its behavior.

MATERIALS AND METHODS

Participants

Patients visiting the EMC of the Asan Medical Center with acute dizziness or vertigo were consecutively checked with diffusion-weighted imaging (DWI) to exclude central dizziness. In the present study, we have retrospectively recruited patients who visited the EMC presenting with acute dizziness or vertigo between January 2010 and December 2013 and received DWI before being discharged from the EMC.

We excluded patients refusing DWI or with contraindications for magnetic resonance imaging (MRI; i.e., pacemaker). In addition, patients were excluded who presented with symptoms indicative of nausea, even though the chief complaint was dizziness or vertigo, and were diagnosed with gastrointestinal disorders. Patients who presented with non-specific dizziness and were diagnosed with general weakness due to poor medical conditions, such as systemic infection or cancer, were also excluded from the final analysis. The Institutional Review Board of the Asan Medical Center approved this study. Informed consent was waived because of the retrospective design.

Classification of the Cause of Dizziness

The cause of acute dizziness and vertigo was categorized based on the final diagnosis upon discharge from the EMC. The final diagnosis was based on extensive evaluation with neuroimaging and comprehensive evaluation by neurologists, otorhinolaryngologists, and emergency medicine physicians. All patients included received DWI. Additional neuroimaging procedures were performed for patients who were suspicious of vertebrobasilar insufficiency after consultation to the neurologists. Computed tomography angiography (CTA) or magnetic resonance angiography (MRA) were performed based on the clinician's decision.

The cause of dizziness was initially categorized as one of the following: (1) central, (2) peripheral, (3) psychogenic, (4) cardio-circulatory, and (5) non-specific. A diagnosis of central dizziness was dependent upon the identification of a focal structural lesion from the corresponding area. Central dizziness included patients with DWI lesions presenting acute ischemic stroke or a significant stenosis (more than 50%) at the vertebrobasilar system. Peripheral dizziness included patients diagnosed as benign paroxysmal positional vertigo, Meniere's disease, vestibule-neuritis, and other vestibulopathies. Patients diagnosed with depression, anxiety disorder, or hyperventilation were categorized as psychogenic dizziness. Cardio-circulatory dizziness included patients diagnosed as syncope or presyncope due to cardiac problems, such as symptomatic arrhythmias, causing insufficient cerebral circulation. Dizziness with an unclear etiology but excluded from being categorized as central causing dizziness by neuroimaging was determined to be non-specific dizziness. All causes of dizziness, except central dizziness, were regarded as non-central dizziness.

Development and Evaluation of Model

Predictors included demographics (e.g., age and sex), previous medical history (e.g., history of hypertension, diabetes, hyperlipidemia, stroke, or coronary artery disease), systolic and diastolic blood pressure (BP), and heart rate. In the current study, we built classification models using various ML algorithms, including the radial basis function kernel support vector machine (SVM), random forest (RF), Catboost, and conventional logistic regression (LR).

The data was split by order of admission date (i.e., temporal validation) into a training set and a test set. Within a training set, multiple hyperparameters were tuned using a five-fold cross-validation. The loss function was negative a log-likelihood with class weights. Area under the curve of the receiver operating characteristic curve (AUROC) was measured to validate performance in a test set. Sensitivity and specificity were also calculated on a test set using threshold at which sensitivity on the training set was 99% or 99.9%, since not missing patients with central dizziness is a more critical factor not missing those with non-central dizziness.

In addition, to understand the reasoning behind certain ML model predictions, we used a Tree Explainer of SHapley Additive exPlanations (SHAP) value (https://github.com/skjang54/Asan_Central-Dizziness-and-Machine-Learning/) (14). The SHAP value (log-odds unit) identifies the degree of impact a predictor has on a prediction. A positive SHAP indicates that the feature drives an increase in the probability of central dizziness (response variable), and a negative SHAP implies that the feature decreases

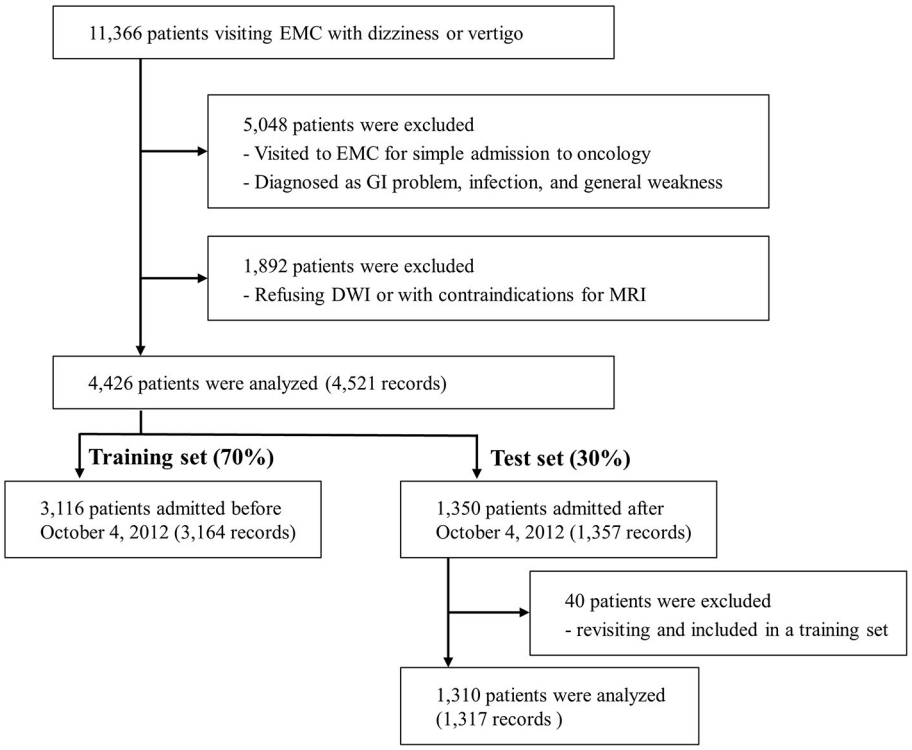


FIGURE 1 | Flow diagram. EMC, emergency medical center; GI, gastrointestinal; DWI, Diffusion-weighted imaging; MRI, Magnetic resonance imaging.

TABLE 1 | Baseline characteristics of the records of the enrolled patients with central and non-central dizziness.

	Central dizziness† (*n* = 414)	Non-central dizziness† (*n* = 4,067)	*p*-value‡
Age (y)	66.1 ± 11.8	61.3 ± 11.5	<0.001
Sex (male)	245 (59.2)	1592 (39.1)	<0.001
Hypertension	198 (47.8)	1531 (37.6)	<0.001
Diabetes	100 (24.2)	531 (13.1)	<0.001
Hyperlipidemia	133 (32.1)	1,368 (33.6)	0.571
Current smoking	41 (9.9)	265 (6.5)	0.012
History of previous coronary artery disease	77 (18.6)	495 (12.2)	<0.001
History of previous stroke	100 (24.2)	313 (7.7)	<0.001
Systolic blood pressure (mmHg)	148.3 ± 23.6	145.7 ± 22.0	0.032
Diastolic blood pressure (mmHg)	87.7 ± 15.8	88.1 ± 14.0	0.584
Heart rate (beat/min)	75.2 ± 15.1	74.6 ± 14.0	0.425
Presentation of dizziness			<0.001
Non-whirling type dizziness	327 (79.0)	2,211 (54.4)	
Vertigo	87 (21.0)	1,856 (45.6)	

†*Values represented as frequency (percentage) or mean ± SD.*
‡*P-values were calculated using t-test for continuous variables and χ^2-test for categorical variables.*

the probability. This approach provides local explanations by illustrating the attribution of a feature within a single patient. Also, the mean of the absolute values of SHAP explains the importance of each feature across the population (global explanation). The SHAP values of Catboost and RF models were computed by Tree Explainer, with SVM as the Kernel Explainer and LR as the Linear Explainer (SHAP version 0.34.0).

RESULTS

Of the 11,366 patients who visited the EMC with dizziness or vertigo, 4,426 patients were included in the final analysis (**Figure 1**). All patients had DWI data and 989 patients had additional angiographic evaluation (CTA, $n = 81$ and MRA, $n = 908$). Of these final patients, 3,116 patients were included in

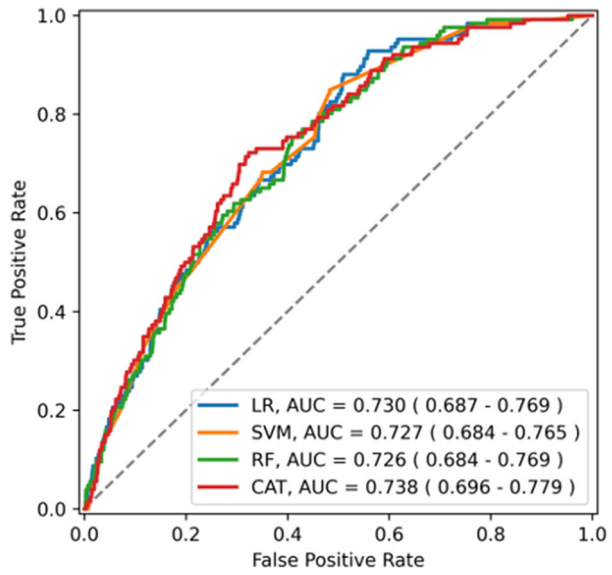


FIGURE 2 | The receiver operating characteristics curve of the model for central dizziness. LR, logistic regression; SVM, support vector machine; RF, random forest; CAT, Catboost; AUC, area under the receiver operating characteristic curve.

the training set, and the remaining 1,310 patients were included in the test set. Within the period of the training set, 46 patients visited twice, and one patient visited three times. Only the first visit was used for the test set. The ratio of central to non-central causes of dizziness was not significantly different in the training set than the test set (**Supplementary Table 1**). Among 4,481 records, 414 (9.2%) were diagnosed with central dizziness. There was no missing data among clinical variables.

Characteristics of Patients With Central Dizziness

As shown in **Table 1**, patients with central dizziness were older (66.1 ± 11.8 vs. 61.3 ± 11.5 years old; $p < 0.001$) and more often male than those with non-central dizziness (59.2 vs. 39.1%; $p < 0.001$). Patients with central dizziness showed a higher prevalence of hypertension (47.8 vs. 37.6%; $p < 0.001$), diabetes (24.2 vs. 13.1%; $p < 0.001$), current smoking (9.9 vs. 6.5%; $p = 0.012$), previous coronary artery disease (18.6 vs. 12.2%; $p < 0.001$), and history of stroke (24.2 vs. 7.7%; $p < 0.001$). Systolic BP was higher in patients with central dizziness than in those with non-central dizziness (148.3 ± 23.6 vs. 145.7 ± 22.0 mmHg; $p = 0.032$). The clinical presentation of non-whirling-type dizziness was observed more often in patients with central dizziness than in those with non-central dizziness (79.0 vs. 54.4%; $p < 0.001$).

Machine Learning Predicting Central Dizziness

The ability of the ML algorithms to discriminate between central and non-central dizziness is shown in **Figure 2**. In the ROC analysis, the models achieved an AUROC of 0.730 (0.690–0.771) in LR, 0.727 (0.687–0.767) in SVM, 0.726 (0.686–0.768) in RF, and 0.738 (0.693–0.780) in Catboost, suggesting moderate predictive accuracy with highest performance by Catboost but without statistically significant difference among models. (Catboost vs. LR: $p = 0.443$; Catboost vs. SVM: $p = 0.263$; Catboost vs. RF: $p = 0.099$). Sensitivity and specificity were also similar among the models, evaluated in the independent test set (**Table 2**). At a classification threshold of 99% sensitivity in the training set, the models showed a sensitivity of 97.6–99.2% and a specificity of 10.5–18.6%. At a threshold of 99.9% sensitivity, the models showed a sensitivity of 99.2–100% and a specificity of 4.6–8.1%.

TABLE 2 | Sensitivity and specificity of the classification models in the independent test set.

	Sensitivity	Specificity
A. Threshold at which sensitivity is 99% in training set		
Logistic regression	0.992 (0.976–1.000)	0.107 (0.088–0.125)
Support vector machine	0.984 (0.960–1.000)	0.157 (0.136–0.177)
Random forest	0.992 (0.976–1.000)	0.125 (0.108–0.144)
Catboost	0.976 (0.944–1.000)	0.167 (0.146–0.190)
B. Threshold at which sensitivity is 99.9% in training set		
Logistic regression	0.992 (0.968–1.000)	0.068 (0.055–0.083)
Support vector machine	0.992 (0.976–1.000)	0.116 (0.099–0.134)
Random forest	0.992 (0.976–1.000)	0.060 (0.046–0.073)
Catboost	1.000 (1.000–1.000)	0.046 (0.034–0.060)

Factors Predicting Central Dizziness

Overall feature attributions of the Catboost model were compared with those of the LR as shown in **Figure 3**. The mean Catboost SHAP value was highest for the presence of a previous stroke history (0.74 ± 0.12), followed by male (0.33 ± 0.04), presentation as non-whirling-type dizziness (0.30 ± 0.02), and age (0.25 ± 0.18). The mean of the absolute values of SHAP was highest for the presentation of dizziness (0.31 ± 0.02), followed by sex (0.27 ± 0.06), age (0.25 ± 0.18), and history of previous stroke (0.14 ± 0.20) in Catboost (**Figure 3A**). These four features—presentation of non-whirling-type dizziness, sex, and age—were also factors with relatively strong impacts on other algorithms, including LR (**Figure 3B** and **Supplementary Figure 1**).

The difference in the way to use features between LR and Catboost can be seen in **Figure 4**. The plots represent how a single feature affects the classifier according to their values. As shown in **Figure 4**, lower than normal systolic BP contributed to a greater negative prediction for central dizziness, whereas higher BP, between 125 and 150 mmHg, showed a neutral SHAP value. Systolic BP over 150 mmHg showed a positive SHAP value, indicating the model considers patients having the range of systolic BP at high risk for central dizziness (**Figure 4A**). Additionally, diastolic BP under 75 or higher than 125 mmHg increased the risk for central dizziness (**Figure 4C**; U-shape

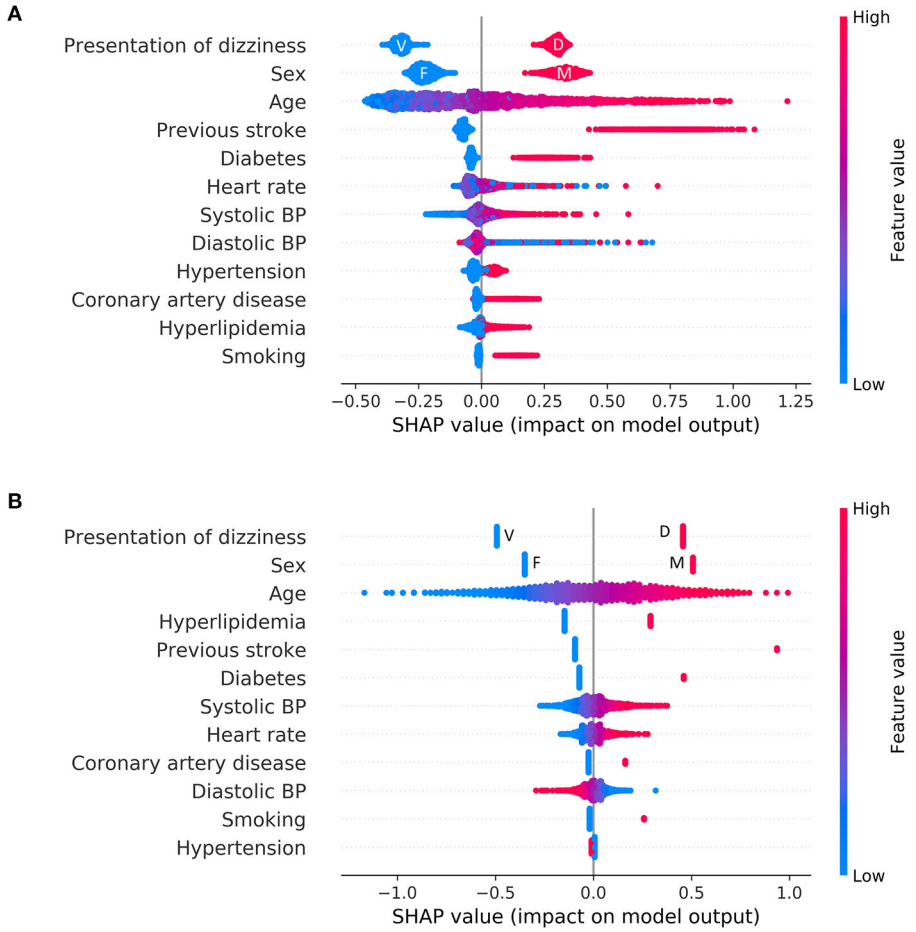


FIGURE 3 | Feature attributions of all features. Summary plot of SHAP values for all features: **(A)** Logistic regression **(B)** Catboost. Each point represents the feature attribution on the log-odds scale for one patient in the training set. Continuous variables were colored by feature value. In the case of categorical variables, except for representation of dizziness and sex, the red color denotes a positive (i.e., a patient has the feature), and the blue color denotes negative. F, female; M, male; V, vertigo; D, dizziness.

curve). Similarly, heart rates lower than 60 and higher than 80 bpm showed positive SHAP values (**Figure 4E**). However, the LR model considered that as the value of feature increases [systolic BP (B), and heart rate (D)] or decreases [diastolic BP (F)], the feature contributes to increasing the log-odds of central dizziness linearly. The RF seemed to have a similar pattern as Catboost, and the SVM appeared to have a smooth curve because of the radial basis function kernel (**Supplementary Figure 2**).

DISCUSSION

In the current study, 9% of patients visiting the EMC showed central dizziness. The ML algorithms designed to predict central dizziness using simple clinical data obtained from triage showed moderate predictive accuracy. The presentation type of dizziness (non-whirling-type dizziness), age (older), sex (male), and history of stroke (present) were shown to be important factors for predicting central dizziness in the Catboost model.

Previous studies have differentiated central dizziness from peripheral causes of vertigo by extensive neuro-ophthalmological examination, including the head-thrust test, gaze-evoked nystagmus, and skew deviation (15). These three tests were proven to be even more sensitive than neuroimaging (15). However, these tests are typically difficult to apply for non-neuro-otology specialists. Alternatively, many scoring systems have been applied to diagnose central dizziness. The modified ABCD2 (age, BP, clinical presentation, and diabetes) score showed an AUROC curve of 0.79 (16). Another scoring system that used eight items, the TriAGE+ score, had an AUROC of 0.82 (17). However, these scoring systems are complicated and still require neurological tests, such as testing for cerebellar dysfunction or focal weakness. Among patients without weakness, the AUROC of the ABCD2 score was lowered to 0.63 (17). Considering that ∼10% of patients with cerebellar infarction present with isolated vertigo (6), it may be inappropriate to differentiate central dizziness based on these scoring systems. Our ML-based diagnosis of central vertigo

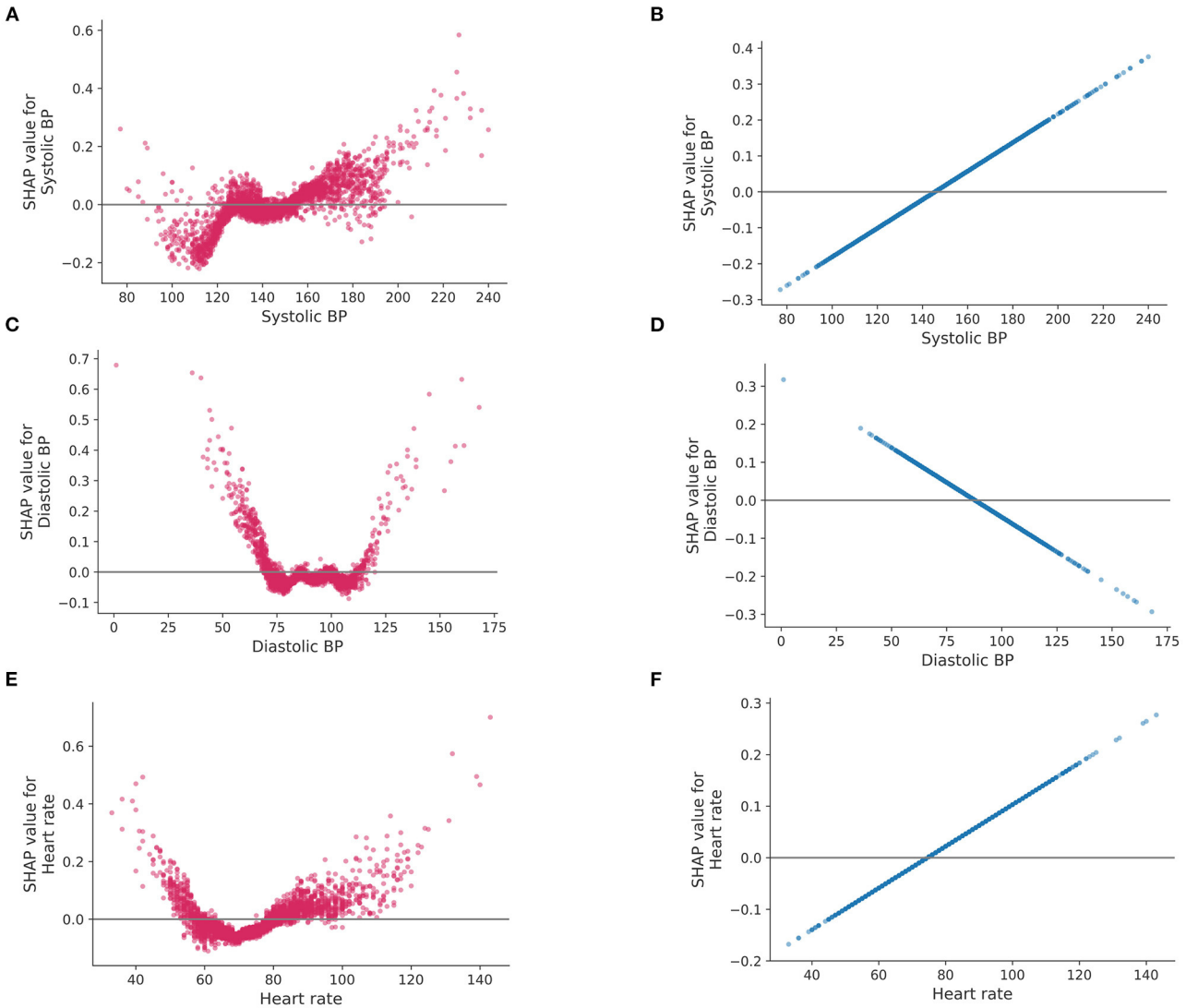


FIGURE 4 | SHAP values for a single feature. SHAP values (impact of each feature on the model) of logistic regression and Catboost are represented on the left and right side, respectively. Systolic blood pressure **(A,B)**, diastolic blood pressure **(C,D)** and heart rate **(E,F)**. BP, blood pressure.

based specifically on simple clinical information, in the absence of neurological information, showed strong predictive power in classifying central dizziness reaching accuracies close to previous scoring systems.

ML has recently shown promising results in various medical fields, some of which have been validated in real-world settings (18). In ML, algorithms are designed to identify important features and/or complicated relationships between these features in an attempt to predict or classify response variables; this is in contrast to rule-based algorithms that use features defined manually. In the current study, Catboost had the highest AUROC among the ML models (0.738) and had 16.7% of specificity at 97.6% of sensitivity and 4.6% of specificity at 100% of sensitivity. These results indicate that, if the model is used to differentiate between central dichotomously in practice, 2.4% of patients with central dizziness would be misdiagnosed, while 16.7% of those with non-central dizziness do not require further neuroimaging services.

We further uncovered important predictors of central dizziness and, moreover, how these variables impact the decision making of the Catboost model using SHAP. Based on our results, previous stroke history, sex (males), presentation of dizziness, and age (older) were the most important factors for classifying central dizziness (**Supplementary Figure 1**). However, according to SHAP, other factors, such as heart rate and systolic and diastolic BP, were still proven to be useful for classifying central dizziness. However, these factors displayed some differences, with higher systolic and diastolic BP showing positive SHAP values. However, lower systolic BP was correlated with lower SHAP values, whereas lower diastolic BP was associated with higher SHAP values. Heart rate results were similar to that of diastolic BP. Patients with low or high heart rates may have a change

of subclinical arrhythmias that may increase the risk of central dizziness. SHAP value patterns exhibit non-linear characteristics using the Catboost model, which is similar to previously observed effects of blood pressure or heart rate on stroke risk. Meanwhile, a linear model, such as LR, showed linearly increased or decreased SHAP values as the variables increase (**Figure 4**).

There are some noteworthy limitations to this study. First, this was a single-center study and may have a limitation in generalizability. Though we included a large number of patients, consecutively to reduce the bias, external validation may be needed to strengthen our results. Second, the clinical information was based on a conservative format that evaluated dizziness based on the presentation of dizziness and risk factors (19). A more updated algorithm that delineates dizziness using timing and triggers may show different results. However, using the presentation of dizziness and risk factors is still a widely accepted clinical approach for differentiating central dizziness. Finally, since ML has an advantage in processing large data, combining findings with video nystagmography or video oculography may enhance the predictive power of the algorithms. However, here we simply used clinical information for the ML input in an attempt to diagnose central dizziness, a strategy that may be more applicable for the non-neuro-otology specialists. Furthermore, as we only used simple data obtainable from the triage, there was no missing clinical data throughout the study.

Our results show that ML models for predicting central dizziness are feasible and require only simple clinical data and the presentation of dizziness. This tool for diagnosing central dizziness may be extremely helpful for non-neuro-otology specialists in determining the priorities of urgent patients and differentiating central dizziness from non-central dizziness in clinical practice.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board of the Asan Medical Center. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

BK contributed by the conceptualization, data curation, and writing the draft preparation. S-KJ contributed by data curation, formal analysis, and draft preparation. Y-HK contributed by formal analysis, visualization, and critical revision of the manuscript. E-JL contributed by conceptualization and critical revision of the manuscript. JC contributed by data curation and revision of the manuscript. SK and JK contributed by supervision and critical revision of the manuscript. D-WK contributed by the conceptualization, supervision, and writing the draft preparation and critical revision of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by grants from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare (Grant Number: HI18C2383) and the Brain Convergence Research Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. 2020M3E5D2A01084576).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fneur.2021.691057/full#supplementary-material

REFERENCES

1. Cappello M, di Blasi U, di Piazza L, Ducato G, Ferrara A, Franco S, et al. Dizziness and vertigo in a department of emergency medicine. *Eur J Emerg Med*. (1995) 2:201–11. doi: 10.1097/00063110-199512000-00006
2. Newman-Toker DE, Edlow JA. TiTrATE: a novel, evidence-based approach to diagnosing acute dizziness and vertigo. *Neurol Clin*. (2015) 33:577. doi: 10.1016/j.ncl.2015.04.011
3. Choi KD, Lee H, Kim JS. Vertigo in brainstem and cerebellar strokes. *Curr Opin Neurol*. (2013) 26:90–5. doi: 10.1097/WCO.0b013e32835c5edd
4. Kim JT, Park MS, Choi KH, Kim BJ, Han MK, Park TH, et al. Clinical outcomes of posterior versus anterior circulation infarction with low national institutes of health stroke scale scores. *Stroke*. (2017) 48:55–62. doi: 10.1161/STROKEAHA.116.013432
5. Lee H. Isolated vascular vertigo. *J Stroke*. (2014) 16:124–30. doi: 10.5853/jos.2014.16.3.124
6. Lee H, Sohn SI, Cho YW, Lee SR, Ahn BH, Park BR, et al. Cerebellar infarction presenting isolated vertigo: Frequency and vascular topographical patterns. *Neurology*. (2006) 67:1178–83. doi: 10.1212/01.wnl.0000238500.02302.b4
7. Savitz SI, Caplan LR, Edlow JA. Pitfalls in the diagnosis of cerebellar infarction. *Acad Emerg Med*. (2007) 14:63–8. doi: 10.1111/j.1553-2712.2007.tb00373.x
8. Lee EJ, Kim YH, Kim N, Kang DW. Deep into the brain: artificial intelligence in stroke imaging. *J Stroke*. (2017) 19:277–85. doi: 10.5853/jos.2017.02054
9. Heo J, Yoon JG, Park H, Kim YD, Nam HS, Heo JH. Machine learning-based model for prediction of outcomes in acute stroke. *Stroke*. (2019) 50:1263–5. doi: 10.1161/STROKEAHA.118.024293
10. Lee H, Lee EJ, Ham S, Lee HB, Lee JS, Kwon SU, et al. Machine learning approach to identify stroke within 4.5 hours. *Stroke*. (2020) 51:860–6. doi: 10.1161/STROKEAHA.119.027611
11. Jang SK, Chang JY, Lee JS, Lee EJ, Kim YH, Han JH, et al. Reliability and clinical utility of machine learning to predict stroke prognosis: comparison with logistic regression. *J Stroke*. (2020) 22:403–6. doi: 10.5853/jos.2020.02537
12. Lim EC, Park JH, Jeon HJ, Kim HJ, Lee HJ, Song CG, et al. Developing a diagnostic decision support system for benign paroxysmal

positional vertigo using a deep-learning model. *J Clin Med*. (2019) 8:633. doi: 10.3390/jcm8050633
13. Kamogashira T, Fujimoto C, Kinoshita M, Kikkawa Y, Yamasoba T, Iwasaki S. Prediction of vestibular dysfunction by applying machine learning algorithms to postural instability. *Front Neurol*. (2020) 11:7. doi: 10.3389/fneur.2020.00007
14. Lundberg S, Lee SI. A unified approach to interpreting model predictions. In: *31st Conference on Neural Information Processing Systems (NIPS 2017)*. Long Beach, CA (2017).
15. Kattah JC, Talkad AV, Wang DZ, Hsieh YH, Newman-Toker DE. Hints to diagnose stroke in the acute vestibular syndrome: three-step bedside oculomotor examination more sensitive than early mri diffusion-weighted imaging. *Stroke*. (2009) 40:3504–10. doi: 10.1161/STROKEAHA.109.551234
16. Navi BB, Kamel H, Shah MP, Grossman AW, Wong C, Poisson SN, et al. Application of the abcd2 score to identify cerebrovascular causes of dizziness in the emergency department. *Stroke*. (2012) 43:1484–9. doi: 10.1161/STROKEAHA.111.646414
17. Kuroda R, Nakada T, Ojima T, Serizawa M, Imai N, Yagi N, et al. The triage+ score for vertigo or dizziness: a diagnostic model for stroke in the emergency department. *J Stroke Cerebrovasc Dis*. (2017) 26:1144–53. doi: 10.1016/j.jstrokecerebrovasdis.2017.01.009
18. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. (2019) 25:44–56. doi: 10.1038/s41591-018-0300-7
19. Muncie HL, Sirmans SM, James E. Dizziness: approach to evaluation and management. *Am Fam Phys*. (2017) 95:154–62. Available online at: https://www.aafp.org/afp/2017/0201/p154.pdf

Conflict of Interest: Y-HK was employed by company Nunaps Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neurology

ORIGINAL RESEARCH
published: 13 July 2021
doi: 10.3389/fneur.2021.649521

Prediction of 30-Day Readmission After Stroke Using Machine Learning and Natural Language Processing

*Christina M. Lineback[1], Ravi Garg[2], Elissa Oh[2], Andrew M. Naidech[1,2], Jane L. Holl[2] and Shyam Prabhakaran[3]**

[1] Department of Neurology, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States, [2] Department of Neurology, Biological Sciences, Division and Center for Healthcare Delivery Science and Innovation, University of Chicago, Chicago, IL, United States, [3] Department of Neurology, University of Chicago, Chicago, IL, United States

Background and Purpose: This study aims to determine whether machine learning (ML) and natural language processing (NLP) from electronic health records (EHR) improve the prediction of 30-day readmission after stroke.

Methods: Among index stroke admissions between 2011 and 2016 at an academic medical center, we abstracted discrete data from the EHR on demographics, risk factors, medications, hospital complications, and discharge destination and unstructured textual data from clinician notes. Readmission was defined as any unplanned hospital admission within 30 days of discharge. We developed models to predict two separate outcomes, as follows: (1) 30-day all-cause readmission and (2) 30-day stroke readmission. We compared the performance of logistic regression with advanced ML algorithms. We used several NLP methods to generate additional features from unstructured textual reports. We evaluated the performance of prediction models using a five-fold validation and tested the best model in a held-out test dataset. Areas under the curve (AUCs) were used to compare discrimination of each model.

Results: In a held-out test dataset, advanced ML methods along with NLP features out performed logistic regression for all-cause readmission (AUC, 0.64 vs. 0.58; $p < 0.001$) and stroke readmission prediction (AUC, 0.62 vs. 0.52; $p < 0.001$).

Conclusion: NLP-enhanced machine learning models potentially advance our ability to predict readmission after stroke. However, further improvement is necessary before being implemented in clinical practice given the weak discrimination.

Keywords: stroke, readmission, machine learning, natural language processing, bioinformatics

OPEN ACCESS

Edited by:
Thanh G. Phan,
Monash Health, Australia

Reviewed by:
Seana Gall,
University of Tasmania, Australia
Aladine Elsamadicy,
Yale University, United States

***Correspondence:**
Shyam Prabhakaran
shyam1@neurology.bsd.uchicago.edu

Specialty section:
This article was submitted to Stroke, a section of the journal Frontiers in Neurology

Received: *04 January 2021*
Accepted: *04 June 2021*
Published: *13 July 2021*

Citation:
Lineback CM, Garg R, Oh E, Naidech AM, Holl JL and Prabhakaran S (2021) Prediction of 30-Day Readmission After Stroke Using Machine Learning and Natural Language Processing. Front. Neurol. 12:649521. doi: 10.3389/fneur.2021.649521

INTRODUCTION

Nearly 800,000 patients experience a stroke each year in the USA (1). The cost of initial admissions for stroke averages US$20,000 while readmissions cost on average US$10,000 (1–3). Reduction in readmission is, thus, an important target to reduce healthcare costs and improve patient care. However, several studies have demonstrated that available prediction models for readmission perform modestly (4, 5). A better understanding of the causes leading to readmission and better prediction tools may allow hospital systems to better allocate resources to the patients who are most at risk for readmission (6, 7).

Prior efforts to stratify risk of readmission have utilized basic statistical models, such as logistic regression, with modest results (AUC range: 0.53–0.67) (5, 7, 8). However, these studies do not report results on a separate held out dataset thereby not addressing the generalizability of these results. Also, since these methods are trained and validated on the same datasets, the results are highly prone to be inflated due to overfitting. Furthermore, logistic regression base models are incapable of properly weighing the interactions between the complex variables in additive analyses (4, 9).

Machine learning (10) (ML) has emerged as a new statistical approach to overcome the limitation of non-linearity and improve predictive analysis in healthcare. Advanced ML methods have shown to be superior for predicting readmission in patients with heart failure (11). Furthermore, natural language processing (NLP) methods can be utilized to automatically extract much of the rich but difficult-to-access medical information that is often buried in unstructured text notes within electronic health records (EHR). There has been widespread interest to use ML in conjunction with NLP to build clinical tools for cohort construction, clinical trials, and clinical decision support (9, 12). There has been, however, no study to use NLP of clinical notes and ML to predict readmissions after stroke. We, therefore, sought to evaluate advanced ML algorithms that incorporate NLP features of textual data in the EHR to improve prediction of 30-day readmission after stroke. We also seek to evaluate our models on a separate held out dataset in order to test the generalizability of our results.

METHODS

Cohort

Using the Northwestern Medicine Enterprise Data Warehouse (NM-EDW), a database that collects and integrates data from the EHR at Northwestern Medicine Healthcare (NMHC) system practice settings, we identified stroke patients hospitalized at Northwestern Memorial Hospital between January 1, 2011 and December 31, 2015. Inclusion criteria were age >18 years old. We defined stroke by ICD-9 codes 430–436 for hemorrhagic and ischemic stroke, excluding 432.x, and 433.x0, and 435.x for transient ischemic attack or asymptomatic cerebrovascular conditions. We excluded patients who expired during index hospitalization and those with psychiatric admissions due to privacy restrictions on access to this type of data in the EDW.

Data Extraction

We obtained discrete structured variables and unstructured free-form text-based clinical notes from the EHR (Cerner, Kansas City, MO) pertaining to the index stroke hospitalization for all patients meeting study criteria from the EDW. The EDW currently contains clinical data on nearly 6.2 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.

For discrete variables, we recorded demographics (age, sex, race, ethnicity, insurance status, marriage status, smoking status), comorbidities based on ICD-9/10 codes (prior stroke, prior transient ischemic attack (TIA), hypertension, diabetes, coronary artery disease, hyper/dyslipidemia, atrial fibrillation, chronic obstructive pulmonary disease, hypothyroidism, dementia, end stage renal disease, cancer, valvular heart disease, congestive heart disease, prior coronary stent or bypass), prior healthcare utilization (number of ED visits and number of hospitalizations in the preceding year), stroke type (hemorrhagic vs. ischemic), length of stay, index hospital stay complications (pneumonia, mechanical ventilation, and percutaneous gastrostomy tube placement), discharge disposition, and discharge medications (e.g., anticoagulants). For non-discrete variables (e.g., text), a data analyst extracted the notes from the EDW. We included only a small appropriate subset of report types to identify potential predictors of readmission: admission, progress, consultation, and discharge notes. We pre-processed them to make it usable for machine learning and combined the raw text data with the discrete data, linking by a common identifier.

Feature Selection

A feature is an individual measurable property or characteristic of a phenomenon being observed. We built different feature sets for our predictive models. First, we compiled discrete features, some of which were used previously in studies of readmission after stroke (**Table 1**). We then extracted these features from the structured data, when available, in the EDW. These 35 discrete features formed the first feature set. We ranked each feature based on its importance using feature importance methods. Specifically, we used xgboost in order to find out the importance of each feature.

Next, we constructed three different types of NLP features from the unstructured clinical notes. To do that, we first pre-processed the notes to remove language abnormalities and make it usable for feature extraction. Specifically, we lowercased the text, removed punctuations, and stop words and non-alphanumeric words. We aggregated all the reports for each patient and then created a large corpus of all the aggregated reports from all the patients. We then created a token dictionary of all the unique important terms from the corpus.

TABLE 1 | List of discrete features extracted from enterprise data warehouse.

Demographics	Age, gender, race, ethnicity, marital status, and insurance status
Risk factors	Hypertension, diabetes mellitus, atrial fibrillation, prior stroke, coronary artery disease, congestive heart failure, valvular heart disease, coronary artery bypass graft/stent, end-stage renal disease, hypothyroidism, dementia, cancer, chronic lung disease, and smoking status
Index stroke encounter characteristics	Primary stroke type, initial NIHSS score, initial GCS score, in-hospital pneumonia, medications (e.g., anticoagulants) at discharge, percutaneous endoscopic gastrostomy, mechanical ventilation, intensive care unit stay, and discharge destination
Other baseline factors	Miles from residence to hospital, frequency of hospital admissions in preceding year, and frequency of stroke admissions in preceding year

We experimented with unigrams, bigrams, trigrams, and noun phrases; however, we found the combination of unigrams and bigrams to work best. An n-gram is a set of occurring words within a given window (for example, $n = 1$ it is unigram, $n = 2$ it is bigram, $n = 3$ it is trigram, and so on).

For our first set of NLP features, using the token dictionary, we transformed the corpus to a patient-token matrix in which each token (unigram or bigram) is represented by term-frequency-inverse document frequency (tf-idf). Next, we used logistic regression with "l1" penalty (LASSO) to reduce the large dimensionality of features (13). The LASSO method puts a constraint on the sum of the parameter coefficient and applies shrinking (regularization) to penalize the coefficient of non-essential features to zero. We filtered all the non-zero coefficient features and used them as our second set of features.

For second set of features, on the patient-token matrix, we applied principal component analysis (PCA) (14) and constructed a graph of the variance by cumulative number of principal components. This graph provided us with the most effective number of principal components that explained the most variance in the data set. We then selected these principal components to form our third set of features.

For final set of features, we ran word2vec (15) on the text corpus to learn word vectors for each token in our dictionary. We used genism (16) package and continuous bag of words approach with standard parameters for running word2vec algorithm. Next, to construct a patient vector, we summed all the individual token

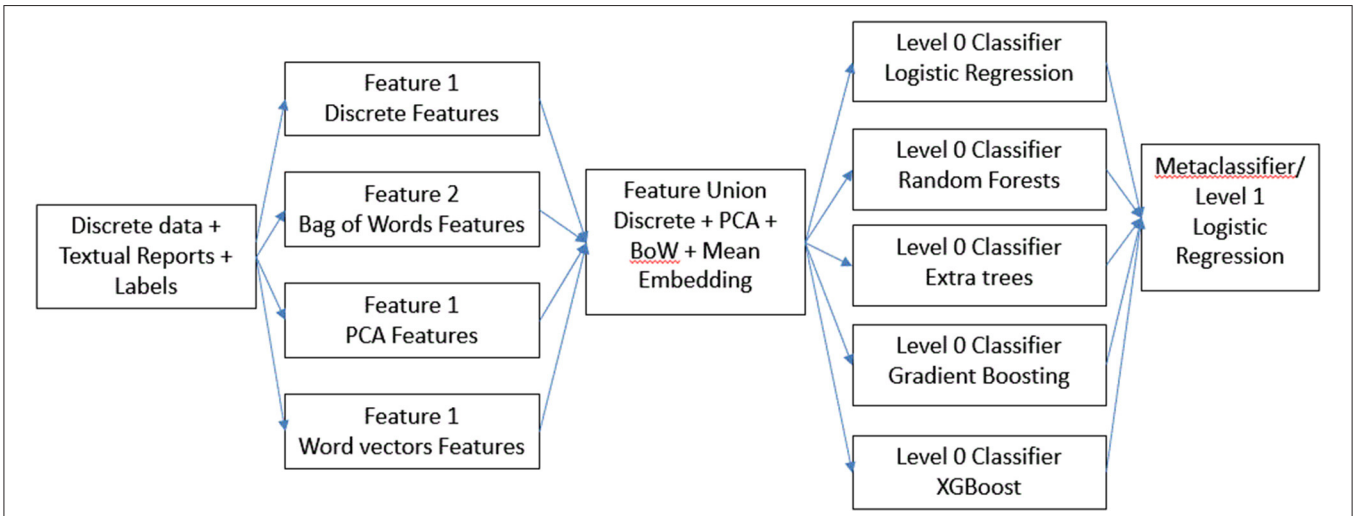


FIGURE 1 | Description of feature ensemble method.

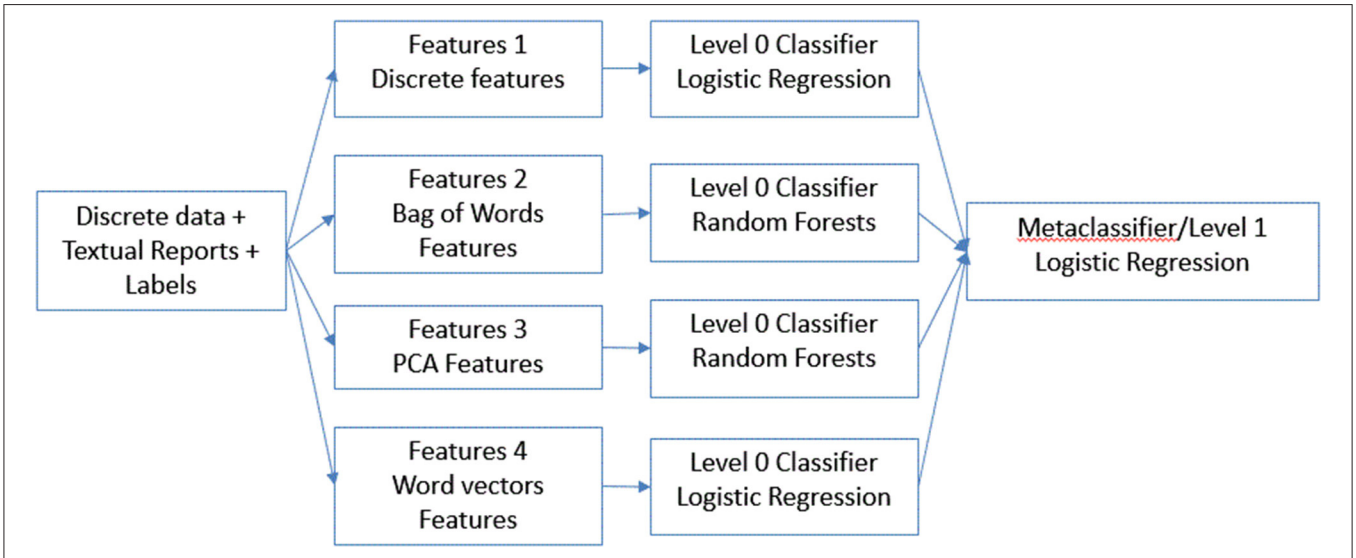


FIGURE 2 | Description of classifier ensemble method.

TABLE 2 | Baseline characteristics of the training cohort ($n = 2,305$) and testing cohort ($n = 550$).

Characteristic	Training cohort	Testing cohort	*P*-value
Mean age in years (SD)	64.4 (16.4)	64.8 (15.1)	0.90
Male sex [*n* (%)]	1,156 (50.2)	297 (54)	0.11
Race [*n* (%)]			
White	1,156 (50.2)	284 (51.6)	0.09
Black	613 (26.6)	138 (25)	
Asian	78 (3.4)	13 (2.4)	
American Indian or Alaskan Native	4 (0.2)	4 (0.7)	
Native Hawaiian/Pacific Islander	4 (0.2)	3 (0.5)	
Declined, missing, or unknown	233 (10.1)	63 (11.4)	
Other	217 (9.41)	45 (8.1)	
Hispanic [*n* (%)]	164 (7.1)	63 (11.4)	<0.01
Marital status [*n* (%)]			
Married	1,001 (43.4)	265 (48.1)	0.02
Widowed	253 (11.0)	45 (8.1)	
Single	759 (32.9)	157 (28.5)	
Divorced	142 (6.2)	33 (6)	
Separated	8 (0.3)	1 (0.2)	
Unknown, other, or missing	142 (6.2)	49 (8.9)	
Insurance status [*n* (%)]			
Private	833 (36.1)	173 (31.5)	<0.01
Medicare	1,060 (46.0)	278 (50.5)	
Medicaid	182 (7.9)	63 (11.5)	
Other or self-pay	230 (10.0)	36 (6.5)	
Primary index stroke diagnosis [*n* (%)]			
Ischemic stroke	1,825 (79.1)	416 (75.6)	<0.01
Intracerebral hemorrhage	257 (11.1)	94 (17)	
Subarachnoid hemorrhage	223 (9.7)	40 (7.3)	
Hypertension [*n* (%)]	1,853 (78.8)	466 (84.7)	0.01
Diabetes mellitus [*n* (%)]	629 (27.3)	179 (32.6)	0.13
Atrial fibrillation [*n* (%)]	430 (18.7)	111 (20.2)	0.42
Coronary artery disease [*n* (%)]	189 (8.2)	30 (5.5)	0.03
Congestive heart failure [*n* (%)]	232 (10.1)	67 (12.2)	0.15
Valvular heart disease [*n* (%)]	42 (1.8)	36 (6.5)	<0.01
Prior stroke [*n* (%)]	218 (9.5)	57 (10.3)	0.57
Chronic lung disease [*n* (%)]	236 (10.2)	48 (8.7)	0.29
Dementia [*n* (%)]	149 (6.5)	37 (6.7)	0.87
Cancer [*n* (%)]	180 (7.8)	45 (8.2)	0.75
End-stage renal disease [*n* (%)]	39 (1.7)	13 (2.3)	0.34
Hypothyroidism [*n* (%)]	270 (11.7)	56 (10.2)	0.32
Smoking [*n* (%)]			
Current	363 (15.7)	76 (13.8)	0.03
Former	595 (25.8)	115 (20.9)	
Non-smoker	1,224 (53.1)	328 (59.6)	
Missing or other	123 (5.3)	31 (5.6)	
Any prior hospitalization [*n* (%)]	1,428 (61.0)	324 (58.9)	0.37
Median initial NIHSS score (IQR)	2 (0–6)	2 (0–6)	0.09
Median initial GCS (IQR)	15 (14–15)	15 (14–15)	0.10
Missing [*n* (%)]	83 (3.6)	22 (4)	0.65
Intensive care unit stay [*n* (%)]	1,166 (50.6)	306 (55.64)	0.04
Inhospital pneumonia [*n* (%)]	108 (4.7)	24 (4.4)	0.76
Mechanical ventilation [*n* (%)]	226 (9.8)	49 (8.9)	0.52
Gastrostomy [*n* (%)]	153 (6.6)	35 (6.3)	0.80
Discharge destination [*n* (%)]			
Home	1,659 (72.0)	350 (63.6)	<0.01
Acute inpatient rehabilitation	429 (18.6)	148 (26.9)	
Skilled nursing facility or long-term facility	153 (6.6)	33 (6)	
Other hospital or against medical advice	64 (2.8)	19 (3.45)	
Any unplanned readmission within 30 days [*n* (%)]	337 (14.6)	62 (11.5)	0.04
Stroke readmission within 30 days [*n* (%)]	124 (5.4)	24 (4.5)	0.33

vectors for each token present in each patient's report. Doing this, each patient is then represented by a single vector, which formed our fourth and final set of features.

Definition of Outcomes

Readmission was defined as any unplanned inpatient hospitalization for any cause after index stroke hospitalization discharge. We excluded planned or scheduled readmissions, emergency department visits without admission, and observation visits. Using the date of index stroke hospital discharge and date of readmission, we identified unplanned readmissions occurring within 30 days of hospital discharge.

Predictive Models

We developed models to predict two separate outcomes: (1) 30-day all-cause readmission and (2) 30-day stroke readmission. For each of these outcomes, we trained different predictive models and compared them with each other. In addition, we also used different types of features for each of predictive models as discussed above. Thus, our study not only evaluates the performance of different predictive algorithms but also the added value of different types of features. We trained a number of different base predictive models as well as several hierarchical predictive models to enhance predictive performance. The base models included logistic regression (17), naïve Bayes (18), support vector machines (19), random forests (18), gradient boosting machines (20), and finally extreme gradient boosting (XGBoost) (21). We trained each of these models for each of the feature types and compared the performance across multiple models.

For our first hierarchical model (**Figure 1**), we combined all the features in the dataset to form a "super" feature set and then trained each of the base models on top of it. In addition, we combined the results from each of these base models and using those as features, we trained another meta-classifier model. We experimented with logistic regression as well as XGBoost for

meta-classifier, but we found logistic regression to perform better. We designated this model a feature ensemble model.

Next, for our final model (**Figure 2**), instead of combining all the features, we concatenated results from the best performing model on individual features. We used the predictions from each of these models as features to train a meta-classifier. This technique is known as stacking (22) wherein outputs from base predictive models are combined to form a feature set which is then used to train another level 2 classifier. We designated this method a classifier ensemble model.

Validation and Evaluation

To avoid over-fitting, we performed five-fold cross-validation (23). Cross-validation, 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 results from the folds can then be averaged (or otherwise combined) to produce a single estimation. We also performed hyper-parameter tuning for our base model within each fold using "hyperopt" python package (24).

In order to test true generalizability of our results, we obtained another dataset spanning from January 1, 2016 to December 31, 2016. We pre-processed it the same way as we did for training data we used for 5-fold cross validation. Next, we trained the best performing models for both outcomes on all the training data and performed the trained model in the test dataset to generate final predictions. We also bootstrapped the test dataset over 50 iterations to generate confidence intervals.

To evaluate the performance of each model, we estimated area under the curve or AUCs from receiver operating characteristic curve analysis. We also compared the best performing model with the baseline logistic regression model of discrete variables alone. p-values < 0.05 were considered significant in all analyses.

Interpretability of NLP Features

To evaluate which NLP-based features were helpful in the prediction model, we ranked the bag of words features according to the feature importance given by the model.

Standard Protocol Approvals, Registrations, and Patient Consents

This study was approved by the Institutional Review Board of Northwestern University. Informed consent was waived for this retrospective data analysis.

Data Availability

All data not presented in this paper will be made available in a trusted data repository or shared at the request of other investigators for purposes of replicating procedures and results.

RESULTS

After pre-processing and combining various data files, we had 2,305 patients for training and 550 patients for testing. The mean age for training cohort and testing cohort was 64.4 and 64.8 years, respectively. The training and testing datasets were similar except the testing set contained more Hispanic,

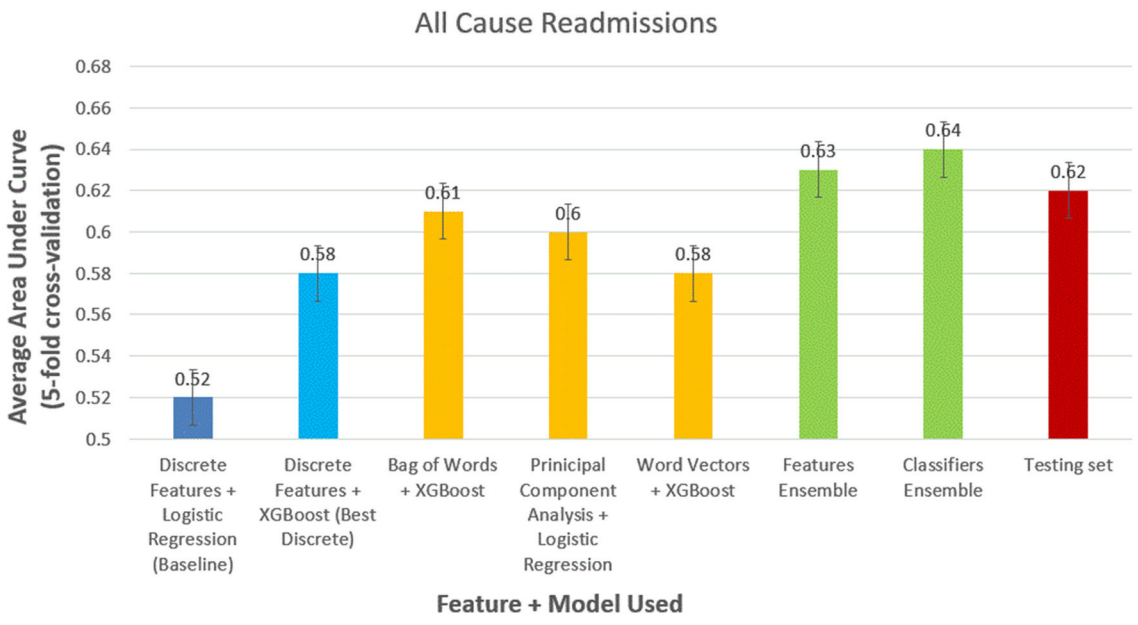


FIGURE 3 | Comparison of models to predict 30-day all-cause readmissions.

government-insured, married, hypertensive, cardiac disease, and intracerebral hemorrhage patients with more ICU days; the testing set also contained more patients who required acute inpatient rehabilitation at discharge (**Table 2**).

In training cohort, there were 337 patients (14.6%) with all-cause readmission within 30 days and 124 patients (5.4%) with stroke readmission within 30 days. In testing cohort, there were 62 patients (11.3%) with all-cause readmission within 30 days and 24 patients (4.4%) with stroke readmission within 30 days. We collected ∼28,500 different patient reports for the training data set and 6,606 reports for the test dataset. We extracted 35 discrete features, 250 principal components features, 400 word-vector features, and 200 bag of words features for all patients in both cohorts.

For all-cause readmission (**Figure 3**), a model using logistic regression using discrete features had AUC of 0.58 (95% CI, 0.57–0.59). In comparison, XGBoost outperformed logistic regression using the same discrete features with an AUC of 0.62 (95% CI, 0.61–0.63). Using NLP-based features, we obtained similar results with XGBoost performing best with bag of words features (AUC, 0.61; 95% CI, 0.60–0.62), logistic regression performing best with PCA features scoring (AUC, 0.61; 95% CI, 0.59–0.62), and XGBoost performing best with word-vector-based features (AUC, 0.60; 95% CI, 0.59–0.61). Ensemble model performed best with feature ensemble method (AUC, 0.64; 95% CI, 0.62–0.66) and classifier ensemble method (AUC, 0.65; 95% CI, 0.62–0.66). We performed the trained classifier ensemble model in the test dataset with bootstrapping over 50 iterations, which resulted in an AUC of 0.64 (95% CI, 0.63–0.65).

We obtained similar results for 30-day stroke readmissions (**Figure 4**). Logistic regression with discrete features formed modest baseline with AUC of 0.52 (95% CI, 0.51–0.53). XGBoost outperformed logistic regression using discrete features alone with AUC of 0.58 (95% CI, 0.56–0.59). The models using the best NLP-based features produced AUCs of 0.61 (95% CI, 0.59–0.63), 0.60 (95% CI, 0.59–0.62), and 0.58 (95% CI, 0.57–0.59) for bag of words features, PCA features, and word-vector features, respectively. Ensemble methods were again the best performing models with AUCs of 0.63 (95% CI, 0.6–0.65) and 0.64 (95% CI, 0.62–0.66) for feature ensemble model and classifier ensemble models, respectively. Performed on the test set, we obtained an AUC of 0.62 (95% CI, 0.61–0.63) using classifier ensemble.

Some of the NLP features that were ranked higher in importance by the model were as follows: "stenosis," "encephalomalacia," "craniectomy," "encephalomalacia," "mild calcified atherosclerotic," "hypoattenuation white matter," and "chiari ii malformation."

DISCUSSION

Given the burden of readmission on the patient and the healthcare system, improving prediction of readmissions with a goal of preventing them is of major importance. A prior study estimated that the cost to Medicare of unplanned rehospitalizations in 2004 was $17.4 billion (25). Readmission to the hospital within 30 days after stroke is also associated with 1-year mortality and serves as a quality metric across specialties under the guidance of the Affordable Care Act (3, 26).

Currently, clinician judgment and simple mathematical models are able to only modestly predict readmission after stroke. In our study, the baseline model that used logistic regression and discrete variables resulted in poor discrimination of 30-day readmission, a result that is consistent with prior studies (5, 7, 8). While NLP-enhanced ML models advance conventional

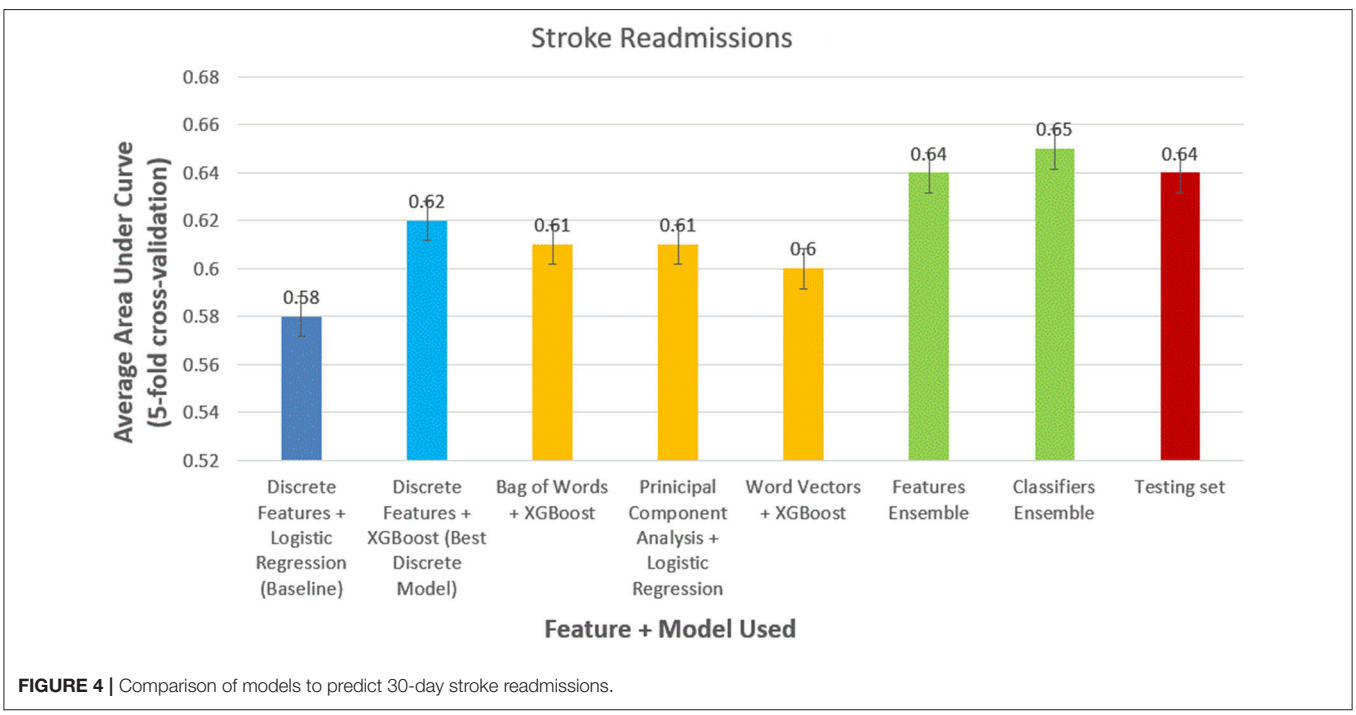


FIGURE 4 | Comparison of models to predict 30-day stroke readmissions.

approaches, further improvement is necessary before these predictive models can be implemented in practice given the weak discrimination. Our finding is similar to another study using machine learning in readmission after heart failure (11).

Given the challenges in accurate prediction of 30-day readmission even using modern machine learning approaches, grading and penalizing hospitals on this metric may not be justifiable. Indeed, hospitals may be forced to "game" the system by increasing observation status visits and avoid penalties at the cost of increasing mortality as a recent study in heart failure patients found (27). Therefore, the penalties facing hospitals seem misguided until such a time when readmission prediction is more robust.

Machine learning is able to weigh the interactions between complex variables in additive analysis to produce better prediction models. In addition, the use of NLP in medicine may be revolutionary. Untangling the complex data within clinical notes and other non-discrete and unstructured data could be valuable in tackling a myriad of research questions. Our advanced models could further ongoing machine learning efforts across specialties to better identify patients for clinical trials, radiologic findings in neurologic emergencies, dermatologic-related malignancies, automatic infectious disease prediction in the emergency room, and outcomes in psychiatric admissions (28–32).

The strengths of our study include a five-fold cross-validation technique to avoid overfitting. The internal validity of our results was further tested by obtaining a second dataset not used in the derivation and validation steps. We also bootstrapped the test dataset over 50 iterations to generate confidence intervals. Our study, however, has limitations. ML algorithms are also limited by the data that are fed into them such that data that are not commonly reflected in the EHR, such as psychosocial factors, post-discharge care coordination, detailed social support post-hospitalization, and post-stroke rehabilitation care are not accounted for in our study. Prior studies suggest including post-acute care data improve prediction of readmission (5, 33). Healthcare systems across the country are heterogeneous, and the variables we used may be non-uniformly available at other hospitals. External validation of our results is necessary. An additional limitation of a single-center cohort is the potential for incomplete follow-up (e.g., care fragmentation leading to admission at another hospital in the region) resulting in an underestimation of readmission rates. However, a recent Chicago multihospital study noted a low rate of care fragmentation (34). There are several differences between the two datasets: the training dataset as it was later chronologically noted changes in the health system and stroke program. These differences may result in error in trained model validation. However, it does provide some measure of external validation as the model performed well. Nevertheless, formal external validation of the model is recommended. In addition, these algorithms require large volume, structured pools of data. Approximately 80% of EHR data is composed of provider notes. Our use of NLP provided a tool for deconstructing these language blocks; however, sufficient time is required to design and train these programs (9). Lastly, these programs lack the clinical insight that is essential for unsupervised implementation, and with any "black box" program, results must be interpreted cautiously (11).

SUMMARY

In summary, we demonstrated a modest added utility of NLP-enhanced ML algorithms to improve prediction of 30-day readmission after stroke hospitalization compared with conventional statistical approaches using discrete predictors alone. While these results are encouraging, further work is required. Given the challenges in predicting readmission after stroke even using the most advanced techniques, the current penalties applied to hospitals for unplanned readmissions should be reevaluated.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the code and data contains Protected Health Information (PHI). Requests to access the datasets should be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

Statistical analysis was done by RG. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Benjamin EJ, Virani SS, Callaway CW, Chamberlain AM, Chang AR, Cheng S, et al. Heart disease and stroke statistics-2018 update: a report from the American Heart Association. *Circulation*. (2018) 137:e67–492. doi: 10.1161/CIR.0000000000000573
2. Wang G, Zhang Z, Ayala C, Dunet DO, Fang J, George MG. Costs of hospitalization for stroke patients aged 18-64 years in the United States. *J Stroke Cerebrovasc Dis*. (2014) 23:861–8. doi: 10.1016/j.jstrokecerebrovasdis.2013.07.017
3. Kind AJ, Smith MA, Liou JI, Pandhi N, Frytak JR, Finch MD. The price of bouncing back: one-year mortality and payments for acute stroke patients with 30-day bounce-backs. *J Am Geriatr Soc*. (2008) 56:999–1005. doi: 10.1111/j.1532-5415.2008.01693.x
4. Kansagara D, Englander H, Salanitro A, Kagen D, Theobald C, Freeman M, et al. Risk prediction models for hospital readmission: a systematic review. *JAMA*. (2011) 306:1688–98. doi: 10.1001/jama.2011.1515
5. Fehnel CR, Lee Y, Wendell LC, Thompson BB, Potter NS, Mor V. Post-acute care data for predicting readmission after ischemic stroke: a nationwide cohort analysis using the minimum data set. *J Am Heart Assoc*. (2015) 4:e002145. doi: 10.1161/JAHA.115.002145
6. Burke JF, Skolarus LE, Adelman EE, Reeves MJ, Brown DL. Influence of hospital-level practices on readmission after ischemic stroke. *Neurology*. (2014) 82:2196–204. doi: 10.1212/WNL.0000000000000514
7. Lichtman JH, Leifheit-Limson EC, Jones SB, Wang Y, Goldstein LB. Preventable readmissions within 30 days of ischemic stroke among Medicare beneficiaries. *Stroke*. (2013) 44:3429–35. doi: 10.1161/STROKEAHA.113.003165

8. Fonarow GC, Smith EE, Reeves MJ, Pan W, Olson D, Hernandez AF, et al. Hospital-level variation in mortality and rehospitalization for medicare beneficiaries with acute ischemic stroke. *Stroke*. (2011) 42:159–66. doi: 10.1161/STROKEAHA.110.601831
9. Artetxe A, Beristain A, Grana M. Predictive models for hospital readmission risk: a systematic review of methods. *Comput Method Programs Biomed*. (2018) 164:149–64. doi: 10.1016/j.cmpb.2018.06.006
10. Nasrabadi NM. Pattern recognition and machine learning. *J Electon Imaging*. (2007) 16:049901. doi: 10.1117/1.2819119
11. Golas SB, Shibahara T, Agboola S, Otaki H, Sato J, Nakae T, et al. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. *BMC Med Inform Decis Mak*. (2018) 18:44. doi: 10.1186/s12911-018-0620-z
12. Shameer K, Johnson Kw, Yahi A, Miotto R, Li LI, Ricks D, et al. Predictive modeling of hospital readmission rates using electronic medical record-wide machine learning: a case-study using mount sinai heart failure cohort. *Biocomputing*. (2017) 22:276–87. doi: 10.1142/9789813207813_0027
13. Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learning Res*. (2003) 3:1157–82.
14. Abdi H, Williams LJ. Principal component analysis. *Wiley Interdisc Rev Comput Stat*. (2010) 2:433–59. doi: 10.1002/wics.101
15. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. In: *Advances in Neural Information Processing Systems*. Lake Tahoe, NV (2013).
16. Rehurek R, Sojka P. Gensim-Statistical Semantics in Python. 2011.
17. Hosmer DW, Jr., Lemeshow S, Sturdivant RX. *Applied Logistic Regression*. Hoboken, NJ: John Wiley & Sons (2013).
18. Liaw A, Wiener M. Classification and regression by random forest. *R News*. (2002) 2:18–22.
19. Cortes C, Vapnik V. Support-vector networks. *Machine Learn*. (1995) 20:273–97. doi: 10.1007/BF00994018
20. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat*. (2001) 29:1189–232. doi: 10.1214/aos/1013203451
21. 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*. San Francisco, CA (2016).
22. Wolpert DH. Stacked generalization. *Neural Netw*. (1992) 5:241–59. doi: 10.1016/S0893-6080(05)80023-1
23. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Ijcai*. San Francisco, CA (1995).
24. Bergstra J, Yamins D, Cox DD. Hyperopt: a python library for optimizing the hyperparameters of machine learning algorithms. In: *Proceedings of the 12th Python in Science Conference*. Austin, TX (2013).
25. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. *N Engl J Med*. (2009) 360:1418–28. doi: 10.1056/NEJMsa0803563
26. Centers for Medicare & Medicaid Services. *Centers for Medicare & Medicaid Services: Readmissions Reduction Program*. Available online at: http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html (accessed 2019).
27. Gupta A, Allen LA, Bhatt DL, Cox M, DeVore AD, Heidenreich PA, et al. Association of the Hospital Readmissions Reduction Program Implementation With Readmission and Mortality Outcomes in Heart Failure. *JAMA Cardiol*. (2018) 3:44–53. doi: 10.1001/jamacardio.2017.4265
28. Miotto R, Weng C. Case-based reasoning using electronic health records efficiently identifies eligible patients for clinical trials. *J Am Med Inform Assoc*. (2015) 22:e141–50. doi: 10.1093/jamia/ocu050
29. Titano JJ, Badgeley M, Schefflein J, Pain M, Su A, Cai M, et al. Automated deep-neural-network surveillance of cranial images for acute neurologic events. *Nat Med*. (2018) 24:1337–41. doi: 10.1038/s41591-018-0147-y
30. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. (2017) 542:115–8. doi: 10.1038/nature21056
31. Rumshisky A, Ghassemi M, Naumann T, Szolovits P, Castro VM, McCoy TH, et al. Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. *Transl Psychiatry*. (2016) 6:e921. doi: 10.1038/tp.2015.182
32. Tou H, Yao L, Wei Z, Zhuang X, Zhang BJBB. Automatic infection detection based on electronic medical records. *BMC Bioinform*. (2018) 19:117. doi: 10.1186/s12859-018-2101-x
33. Slocum C, Gerrard P, Black-Schaffer R, Goldstein R, Singhal A, DiVita MA, et al. Functional status predicts acute care readmissions from inpatient rehabilitation in the stroke population. *PLoS ONE*. (2015) 10:e0142180. doi: 10.1371/journal.pone.0142180
34. Galanter WL, Applebaum A, Boddipalli V, Kho A, Lin M, Meltzer D, et al. Migration of patients between five urban teaching hospitals in Chicago. *J Med Syst*. (2013) 37:9930. doi: 10.1007/s10916-013-9930-y

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neurology

ORIGINAL RESEARCH
published: 27 September 2021
doi: 10.3389/fneur.2021.670379

Application of Machine Learning Techniques to Identify Data Reliability and Factors Affecting Outcome After Stroke Using Electronic Administrative Records

*Santu Rana[1], Wei Luo[2], Truyen Tran[1], Svetha Venkatesh[1], Paul Talman[3], Thanh Phan[4], Dinh Phung[5] and Benjamin Clissold[3,4,6]**

[1] Applied Artificial Intelligence Institute (A2I2), Deakin University, Geelong, VIC, Australia, [2] School of Information Technology, Deakin University, Burwood, VIC, Australia, [3] Neurosciences Department, University Hospital Geelong, Geelong, VIC, Australia, [4] Stroke and Ageing Research Group, Department of Medicine, Monash University, Melbourne, VIC, Australia, [5] Department of Science and AI, Monash University, Clayton, VIC, Australia, [6] Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, VIC, Australia

OPEN ACCESS

Edited by:
David S. Liebeskind,
University of California, Los Angeles,
United States

Reviewed by:
Steffen Tiedt,
LMU Munich University Hospital, Germany
Sofia Vallila Rohter,
MGH Institute of Health Professions,
United States

***Correspondence:**
Benjamin Clissold
Benjamin.Clissold1@monash.edu

Specialty section:
This article was submitted to Stroke, a section of the journal Frontiers in Neurology

Received: 21 February 2021
Accepted: 30 August 2021
Published: 27 September 2021

Citation:
Rana S, Luo W, Tran T, Venkatesh S, Talman P, Phan T, Phung D and Clissold B (2021) Application of Machine Learning Techniques to Identify Data Reliability and Factors Affecting Outcome After Stroke Using Electronic Administrative Records. Front. Neurol. 12:670379. doi: 10.3389/fneur.2021.670379

Aim: To use available electronic administrative records to identify data reliability, predict discharge destination, and identify risk factors associated with specific outcomes following hospital admission with stroke, compared to stroke specific clinical factors, using machine learning techniques.

Method: The study included 2,531 patients having at least one admission with a confirmed diagnosis of stroke, collected from a regional hospital in Australia within 2009–2013. Using machine learning (penalized regression with Lasso) techniques, patients having their index admission between June 2009 and July 2012 were used to derive predictive models, and patients having their index admission between July 2012 and June 2013 were used for validation. Three different stroke types [intracerebral hemorrhage (ICH), ischemic stroke, transient ischemic attack (TIA)] were considered and five different comparison outcome settings were considered. Our electronic administrative record based predictive model was compared with a predictive model composed of "baseline" clinical features, more specific for stroke, such as age, gender, smoking habits, co-morbidities (high cholesterol, hypertension, atrial fibrillation, and ischemic heart disease), types of imaging done (CT scan, MRI, etc.), and occurrence of in-hospital pneumonia. Risk factors associated with likelihood of negative outcomes were identified.

Results: The data was highly reliable at predicting discharge to rehabilitation and all other outcomes vs. death for ICH (AUC 0.85 and 0.825, respectively), all discharge outcomes except home vs. rehabilitation for ischemic stroke, and discharge home vs. others and home vs. rehabilitation for TIA (AUC 0.948 and 0.873, respectively). Electronic health record data appeared to provide improved prediction of outcomes over stroke specific clinical factors from the machine learning models. Common risk factors associated with a negative impact on expected outcomes appeared clinically intuitive, and included older age groups,

prior ventilatory support, urinary incontinence, need for imaging, and need for allied health input.

Conclusion: Electronic administrative records from this cohort produced reliable outcome prediction and identified clinically appropriate factors negatively impacting most outcome variables following hospital admission with stroke. This presents a means of future identification of modifiable factors associated with patient discharge destination. This may potentially aid in patient selection for certain interventions and aid in better patient and clinician education regarding expected discharge outcomes.

Keywords: electronic records, stroke outcomes, machine learning, discharge destinations, stroke mortality

INTRODUCTION

The use of electronic administrative records has become widespread in many settings in recent years. This includes the primary care setting and hospital environment (1). Administrative data in the Australian setting may be in the form of mandatory hospital collected data relating to every hospital episode of care, with the data reported to state health departments, in order to inform health care delivery, resourcing, and financial allocation (2). Administrative datasets include primary and secondary diagnosis codes, coding related to comorbidities, discharge destination, and other demographic data. The ability to harness this data to improve patient care, predict outcomes, and identify risk factors for recurrent disease and readmission means that this has become an important area for research and health metrics (3). The heterogeneity of the data and data systems themselves mean that close collaboration between clinicians and analysts is required. Identifying the type of data available and applying this to appropriate clinical questions not yet answered makes this exciting future area of endeavor. This also increases the importance of accurate data collection. Even more vital is the capture of disease specific factors.

Despite the apparent decrease in stroke incidence, in an aging population, stroke survival, and prevalence is increasing (4, 5). This dramatically increases the societal burden of care. Importantly, stroke outcomes are significantly affected by timely hyperacute therapies such as thrombolysis and endovascular clot retrieval for ischemic stroke (6–8), admission to a specialized stroke unit setting (9), appropriate imaging and secondary prevention therapies (10), dysphagia screening, and early mobilization (11). These interventions directly impact the need for rehabilitation or other discharge outcomes, including the potential need for long-term high-level care, and mortality (12). Understanding the factors contributing to functional outcome after stroke provides a potential target for clinicians to alter their management of patients (13). It is important to clarify if these strategies are routinely implemented through available data and audit processes, which may be best performed by disease specific quality clinical registries (14). Whilst the interventions above are well-proven to influence outcomes and also result in a reduction in hospital length of stay and readmission (15), there may be other novel factors during the admission process that have not been previously captured or studied. Analysis of available administrative data may identify process, structural, and outcome measures not previously recognized.

It is important to acknowledge the limitations of administrative datasets. Functional outcome data for stroke from administrative data may not be well-documented at any stage in the collection process. Stroke severity such as the NIHSS score may not be routinely captured or mandated and is known to directly impact outcomes (12, 15). Standard functional scoring such as the modified Rankin score or Barthel index may not be well-recorded and are not mandated in the electronic data. At best, in some cases, we may only be able to use proxy markers of function, such as in-hospital mortality, or discharge destination. Whilst these surrogate outcomes are well-captured from administrative data, they may not illustrate functional status comprehensively and in particular relation to stroke outcomes, do not inform around the 3- or 12-month clinical status, often used to assess the benefits of interventions in stroke patients. However, the systematic methods used, relatively complete capture of admitted patient data and system wide data collection in administrative datasets make these compelling sources to utilize.

Using machine learning techniques to answer health related questions presents a unique and powerful option for improving diagnosis, treatment, and outcome measures. There are also opportunities for identifying predictive factors impacting patient outcomes. Knowledge regarding patient and other factors associated with certain outcomes may allow future application of measures that influence patient care.

AIMS

We sought to use data from existing electronically collected administrative records to identify risk factors associated with specific outcomes for patients with stroke (both ischemic and hemorrhagic) admitted to a large regional hospital, in Victoria, Australia. In addition, we sought to evaluate the utility of using a large array of available electronic health record data from a cohort of patients, when compared to a cohort of patients with available stroke specific clinical factors, to predict discharge outcomes following hospital admission with stroke, using machine learning techniques.

METHODS

Study Setting

Barwon Health is a large regional tertiary hospital, located in Geelong, approximately 1 h to the west of Melbourne, the second most populous city in Australia. This health service provides public hospital care to the population of Geelong and surrounding regional areas. The hospital includes a comprehensive neurology service, including acute stroke thrombolysis, dedicated specialized and geographically located stroke unit, and high-level imaging facilities available for acute stroke investigation. The benefits of evaluating this patient cohort include that the majority of patients with stroke are admitted to the public hospital, via the emergency department, rather than local private hospitals. Nearly all cases were likely to be captured for this region as a result. Stroke units in Australia do not currently require formal stroke unit certification, however, designated stroke units are required to adhere to a number of key elements defined in the national stroke services framework (16).

We obtained a comprehensive selection of data fields from the routinely collected electronic administrative data from Barwon Health, for the period 2003–2014. Administrative data refers to both coding and demographic data and is reportable to the state Department of Health and Human Services (2, 17). We analyzed data based on all patients with an admission diagnosis of stroke, using ICD 10 coding nomenclature. Due to the lack of stroke specific data on functional outcomes after the incident event, surrogate outcomes of discharge destination, and in-hospital mortality were thought to be the most appropriate markers of outcome. Comparisons were made between patient admission source i.e., from home, rehabilitation, nursing home, other hospital, and discharge destination, including death in hospital. The comparisons were performed in order of perceived severity of the outcome. Patient admission source is a defined variable collected for all hospital admitted episodes, as opposed to their discharge destination. By ascertaining relevant factors contributing positively or negatively to our defined outcomes, we hoped to be able to understand novel patient, investigation, and management factors associated with our outcomes. Prior ethics approval had been provided for all data use and analysis between Barwon Health and Deakin University in an institutional agreement.

Dataset

The patient cohort consisted of 2,531 patients with confirmed diagnosis of Stroke or TIA admitted between July 2009 and June 2013. A stroke admission was defined by ICD-10 codes G46, I60-69, G450-453, and G458-459 in the discharge diagnoses (either primary or secondary). For each patient, the index admission was defined as the first stroke admission of the patient starting from 1st January 2009. Patient records available from Barwon Health admissions prior to the index admission were available and were used to construct independent variables. Available data from index admissions and prior admissions included all data reportable to the state Department of Health and Human Services as part of mandatory hospital reporting (2, 17). Our dataset was not able to capture admission data outside of Barwon Health admissions i.e., was not linked to private hospital admissions or admissions to other public institutions. The outcome considered was the discharge destination (home, rehabilitation, or nursing home) if the patient is alive, or death if the patient had died during hospitalization.

Data Analysis

We considered all available administrative hospital data including static information (age, gender, occupation, insurance types), and time-stamped events associated with emergency visits, hospitalizations, radiological tests, length-of-stay, emergency attendance time, primary and secondary diagnoses, and procedures. The use of cerebral imaging such as CT and MRI in stroke evaluation is an important process measure in helping to accurately diagnose and manage patients and was felt important to include in the analysis. Medication usage data was not available from our dataset. Age was coded as a binary variable (i.e., the age variable or not) in one of 10-year intervals, in line with other stroke community and cohort studies (18, 19). Occupation was a binary of value 1 if it was either pensioner, retired, or home duties and 0 otherwise. Time-stamped events were aggregated over two periods of time prior to the index admission: 0–12 months and beyond 12 months. This resulted in a total of 1,303 features. Models were built to analyse the factors associated with different outcomes [e.g., in-hospital death vs. others (i.e., Discharge to home, Rehabilitation, Nursing home), Discharge to home vs. others] using penalized logistic regression with Lasso (20).

We split the data in time (external validation) with data from July 2009 to June 2012 for derivation of predictive models and July 2012 to June 2013 as validation. Confidence intervals were computed based on 100 bootstrapped derivation cohorts from the original derivation cohorts using sampling with replacement.

Five different comparison settings for each of the three sub-cohorts of stroke [intracerebral hemorrhage (ICH), ischemic stroke, transient ischemic attack (TIA)] are considered, by evaluating factors likely to be associated with the defined outcomes, vs. other outcomes.

- Discharge to home vs. others (rehabilitation, nursing home, in-hospital death) out of all patients
- Discharge to rehabilitation vs. home for patients either discharged to home or rehabilitation
- Discharge to nursing home vs. rehabilitation for patients either discharged at nursing home or rehabilitation
- Discharge to nursing home vs. death
- In hospital death vs. discharge to all other places (home, rehabilitation, nursing home)

Where there were small sample sizes, data were collapsed together for the purposes of comparison.

All data processing was performed off-line using a commercial software package (MATLAB, Statistics Toolbox, The MathWorks Inc., 1994–2014). Prediction accuracy is expressed as the area under the receiver operating characteristic curve (AUC). Missing data were imputed.

Two feature sets were constructed:

1. Features constructed from the electronic administrative record which included all available detailed diagnosis, procedure, and administrative data. This included stroke and TIA related diagnostic codes (I60–I69, G45) relating to primary diagnosis, secondary (comorbidity) diagnostic codes, and all available procedure codes relating to patient admissions. The number of variables was 1,303 (some examples of the types of features included can be seen in the data items listed in the **Appendix Figures**).
2. Features constructed from more stroke specific clinical data such as age, gender, smoking habits, co-morbidities (hyperlipidaemia, diabetes, hypertension, atrial fibrillation, and ischemic heart disease), types of imaging done (CT scan, MRI, etc.,—an important stroke management process marker), and occurrence of in-hospital pneumonia. Specific stroke risk factors such as alcohol use, anticoagulant use, and obesity are not included in the routine data collection.

RESULTS

We derived prediction results for three subcohorts of stroke patients (ICH, ischemic stroke, and TIA) in five different settings, as outlined above. All results presented are based on the validation cohort, unless otherwise specified. Patient characteristics and discharge destinations are summarized in **Tables 1**, **2**.

The percentage of stroke type found in our cohort is similar to other cohorts. The occurrence of "Not specified" diagnostic codes highlights a key problem in using administrative datasets and is identified as a limitation in other cohort studies (21).

The percentage of patients with specified comorbidities is again similar to other cohort studies (4, 22), although the percentage with IHD was lower. In relation to imaging, 100% of patients underwent imaging with CT scan of the brain, as is standard clinic practice in patients with suspected stroke or TIA, in order to ascertain presence of infarction or hemorrhage, as well as other causes of potential stroke mimics. The majority of patients had a length of stay of between 1 and 5 days, in keeping with findings from local acute stroke audits.

We sought to identify specific predictive factors from our analysis associated with the outcomes we have studied. These factors were items from our administrative data, presented in the figures below as both positively and negatively weighted variables. **Table 6** below summarizes factors found to negatively impact the outcome presented. For example, for patients with ICH, patients were less likely to be discharged home vs. to all other discharge destinations (rehabilitation, nursing home, or die in hospital) in older age groups (80–90 years old), had had prior ventilatory support, a history of urinary incontinence, or diagnosis of SAH.

Figures in the **Appendix** below identify all factors from the administrative dataset that both positively and negatively impact the outcomes being studied and represent weights of the linear model.

TABLE 1 | Patient characteristics.

No. of patients	2,531	%
Males	1,346	53.2
Females	1,185	46.8
Mean age	72.9 (18.4–99.8)	
<50 years		1.5
50–59 years		9.6
60–69 years		24.5
70–79 years		24.2
80–89 years		32.2
90–99 years		8.0
Stroke type		
Transient ischemic attack		25.1
Intracerebral hemorrhage		14.6
Ischemic stroke		37.7
Aneurysm		0.9
Not specified		21.8
Comorbidities		
Hypertension		52.5
Atrial fibrillation		15.4
Hyperlipidaemia		6.7
Smoking		13.8
Ischemic heart disease		8.4
Imaging		
CT brain		100
X ray chest		91.8
US carotid doppler		43.8
MRI brain		36.8
Length of stay	1–5 days	99.1
	>5 days	0.9

TABLE 2 | Discharge destination.

Discharge destinations	**%**
Home	58.9
Rehabilitation	24.5
Nursing home	5.1
In hospital death	9.4

DISCUSSION

Our goal was to compare the utilization of an electronic health record model constructed using a general set of coding data and demographic data, with a model based on a specifically selected set of clinically recognized features, in identifying data reliability, predict discharge destination, and identify risk factors associated with specific outcomes following hospital admission with stroke. Analysis using the electronic health record data provided better prediction of outcome and use of stroke specific factors did not appear to improve the model's reliability. When comparing the data from **Tables 4**, **5**, our data was highly reliable in predicting outcomes in patients with ICH of discharge to rehabilitation vs. nursing home or death, as well as all other discharge

TABLE 3 | Percentage of patients that fit the model in the derivation cohort under five different prediction settings for three sub-cohorts of stroke.

	Intracerebral hemorrhage	Ischemic stroke	TIA
Home vs. others	16.5% (357)	47.1% (830)	87.5% (659)
Home vs. rehab	24.6% (240)	57.8% (677)	93.8% (531)
Rehab vs. nursing home or death (nursing home and death collapsed due to small sample size)	60.7% (298)	65.2% (437)	
Nursing home vs. death	17.8% (117)	35.3% (153)	
Others vs. death	26.9% (357)	11.9% (830)	

Where there are missing outcomes in the table, this denotes scenarios where derivation is difficult due to the lack of sufficient number of patients. The numbers in parentheses denote total patient numbers in the derivation category for that pair of outcomes.

outcomes vs. death. In ischemic stroke, the data was reliable at predicting discharge home vs. other outcomes, discharge to rehabilitation vs. nursing home or death, discharge to nursing home vs. death, and all other outcomes vs. death. For TIA, the data proved reliable in predicting discharge home and to home vs. rehabilitation.

There are several problems in using electronic administrative records data to identify risk factors and predict outcomes. The amount of electronic data collection contained in these datasets is copious, and there is significant risk in misinterpreting data if it is not disease specific. The complexities of interactions between patient demographic, diagnostic, imaging, procedural, and outcome data may be difficult to interpret. If there is a small group of well-known risk factors, which have been expertly evaluated or have a sound scientific or peer reviewed connection with the research question or patient group, this may be applied in the analysis. Another method may be to examine a larger group of risk factors and determine their statistical significance and predictive power, and hence refine these to the patient population, using regression methods. However, this method again may not be disease specific. The risk factors used in any analysis may be too limited for the data available, and too much data may make the results noisy or uninterpretable. There are inherent differences in risk factors, measures of severity, and specific management strategies for ischemic stroke/TIA and hemorrhagic stroke, which may be useful to capture in any comprehensive medical record.

The use of logistic regression with Lasso is a common linear classifier method that is also suitable for feature selection. The models obtained are likely to be more parsimonious than logistic regression alone. Our aim was to contribute to understanding about the utility of using electronic health record data for clinical prediction, rather than use of different machine learning methods.

Although we understand risk factors such as age, gender, and co-morbidities well in terms of their likely effect on outcomes in stroke patients, the highly detailed data collected by the hospital data warehouse, both for reporting, planning, and financial purposes, means there are likely to be novel but useful predictive factors identified from analyses like this one. Of interest from our list of identified predictive factors for discharge destination were the findings of prior factors in patient histories including prior ventilatory support, imaging factors, respiratory and urinary tract conditions, and allied health input. These novel past history and other elements may indicate new and innovative areas to focus on, guiding clinically, and patient relevant insights and exploration.

Note that factors for Nursing Home vs. Rehabilitation and Death vs. Others for patients with TIA are not presented since the predictive models are unstable (as seen by the lack of valid data in **Table 3**).

The burden of stroke is significant, and recurrent events may add significantly to pre-existing disability, with further acute healthcare, career, and economic impact. Being able to better identify factors associated with poorer outcome can help clinicians intensify efforts in certain areas. Predictive measures can be factored into clinical care paradigms in situations where the data is reliable and serve as an additional tool.

Many of the identified factors from the model felt to influence the outcomes in question appear clinically intuitive. Older age group, the need for allied health and complications of illness such as pneumonitis the clinician understands have a substantial impact on good outcomes in patients with stroke and other diseases. However, understanding these specific factors may help us to better define which patients require more attention or intervention, and supports the strength of the dataset. Some of these factors are not modifiable but can help us in prognostication and better informing patients and families.

One of the limitations of this study was the lack of an available functional outcome measure in the electronic data, leading to the use of "surrogate" markers of function on discharge from the acute event. The use of clinically important scores such as the modified Rankin score and NIHSS (23) in most stroke outcome studies is not possible using the current dataset and highlights the important areas of deficit in clinically relevant/disease specific measures from administrative data. The lack of important imaging data such as stroke infarct volume, and stroke specific treatments, is also a barrier.

CONCLUSION

The electronic administrative record data for our stroke cohort appeared reliable in outcome prediction for most patients and for different stroke types, when based on discharge destination. Risk factors having a negative impact on the defined discharge destinations provide useful and intuitive patient factors which could allow therapeutic intervention and a clearer understanding of which patients are more likely to have better clinical outcomes following an index stroke. In

TABLE 4 | AUC of prediction for three different sub-cohorts of stroke at five different settings.

	Intracerebral hemorrhage	Ischemic stroke	TIA
Home vs. others	0.604 (0.404–0.791)	0.803 (0.746–0.891)	0.948 (0.901–0.955)
Home vs. rehab	0.600 (0.418–0.783)	0.752 (0.683–0.820)	0.873 (0.749–0.996)
Rehab vs. nursing home or death	0.850 (0.737–0.963)	0.818 (0.736–0.801)	
Nursing home vs. death	0.550 (0.245–0.855)	0.902 (0.777–1.00)	
Others vs. death	0.825 (0.698–0.952)	0.881 (0.804–0.959)	

The features used are constructed from the electronic administrative record. 95% CI for reported AUC is presented in the respective parenthesis. Results with missing values implies invalid CI associated with unstable models, generally resulted from lack of sufficient data.

TABLE 5 | AUC of prediction for three different sub-cohorts of stroke at five different settings.

	Intracerebral hemorrhage	Ischemic stroke	TIA
Home vs. others	0.459 (0.285–0.634)	0.702 (0.634–0.769)	0.794 (0.585–1.00)
Home vs. rehab	0.296 (0.131–0.462)	0.636 (0.558–0.714)	0.729 (0.283–0.996)
Rehab vs. nursing home or death	0.504 (0.346–0.661)	0.767 (0.674–0.860)	
Nursing home vs. death	0.625 (0.369–0.881)	0.778 (0.586–0.970)	
Others vs. death	0.583 (0.424–0.742)	0.808 (0.718–0.899)	

The features used are stroke specific clinical data. 95% CI for reported AUC is presented in the respective parenthesis. Results with missing values implies invalid CI associated with unstable models, generally resulted from lack of sufficient data.

TABLE 6 | Selected predictive factors associated with the prediction models.

	Discharge home vs. other outcomes	Discharge home vs. to rehabilitation	Discharge to rehabilitation vs. nursing home or death	Discharge to nursing home vs. death	All other discharge outcomes vs. death
ICH	Older age group (80–90), prior ventilatory support, urinary incontinence, SAH	SAH, prior ventilatory support, prior CT imaging, urinary incontinence, older age group	Prior admission from emergency to the ward, prior CT brain/cervical spine, older age group, prior ventilatory support	Prior ventilatory support, age 70–80, male gender, SAH	Ventilatory support, age >90, prior CT brain/cervical spine, age 80–90, past admission from emergency to ward
Ischemic stroke	Urinary retention, hemiplegia, age group 80–90, allied health input as inpatient, chest X-ray, pneumonitis	Urinary retention, inpatient allied health involvement, hemiplegia, older age group (80–90)	Older age group (>90), pneumonitis, other intestinal disorders, and restlessness/agitation	Other medical care (Z51)*, prior ventilatory support, pneumonitis, unspecified threat to breathing, chest X-ray, and hemiplegia	Other medical care, pneumonitis, chest X-ray, and unspecified threat to breathing
TIA	Older age group (>90), cerebral infarction diagnosis, disorientation, prior allied health care and diagnosis of syncope/collapse	Older age group (>90), diagnosis of cerebral infarction, syncope/collapse, and prior allied health involvement	N/A*	N/A*	N/A*

Factors are those having a negative impact on the outcome in question.

**The other medical care (Z51) diagnosis is very broad—includes radiotherapy session, chemotherapy session, blood transfusion without reported diagnosis, preparatory care for subsequent treatment, palliative care, desensitization to allergens, other specified medical care, medical care unspecified.*

future, the availability of more stroke specific clinical factors in the dataset, including better clinical outcome variables, will likely aid in improving the validity of our data for analysis and prediction.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the raw data outputs are no longer available due to changes in University and health service agreements. Requests to access the datasets should be directed to benc73@hotmail.com.

AUTHOR CONTRIBUTIONS

All authors contributed to conception and design of the study. SR, WL, TT, DP, and BC organized the database. SR, WL, TT, DP, and SV performed the statistical analysis. BC wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

REFERENCES

1. Dregan A, Toschke MA, Wolfe CD, Rudd A, Ashworth M, Gulliford MC. Utility of electronic patient records in primary care for stroke secondary prevention trials. *BMC Public Health.* (2011) 11:86. doi: 10.1186/1471-2458-11-86
2. *Victorian Admitted Episodes Dataset: Department of Health and Human Services, State Government of Victoria.* (2015). Available online at: https://www2.health.vic.gov.au/hospitals-and-health-services/data-reporting/health-data-standards-systems/data-collections/vaed (accessed June, 2015).
3. Frisher M, Short D, Bashford J. Determining patient characteristics for decision analysis support systems using anonymized electronic patient records. *Health Informatics J.* (2010) 16:49–57. doi: 10.1177/1460458209353559
4. Clissold BB, Sundararajan V, Cameron P, McNeil J. Stroke incidence in Victoria, Australia—emerging improvements. *Front Neurol.* (2017) 8:180. doi: 10.3389/fneur.2017.00180
5. Jamrozik K, Broadhurst RJ, Lai N, Hankey GJ, Burvill PW, Anderson CS. Trends in the incidence, severity, and short-term outcome of stroke in perth, Western Australia. *Stroke.* (1999) 30:2105–11. doi: 10.1161/01.STR.30.10.2105
6. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. *N Engl J Med.* (1995). 333:1581–7. doi: 10.1056/NEJM199512143332401
7. Goyal M, Menon BK, van Zwam WH, Dippel DW, Mitchell PJ, Demchuk AM, et al. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *Lancet.* (2016) 387:1723–31. doi: 10.1016/S0140-6736(16)00163-X
8. Hacke W, Kaste M, Bluhmki E, Brozman M, Davalos A, Guidetti D. et al. Thrombolysis with alteplase 3 to 45 hours after acute ischemic stroke. *N Engl J Med.* (2008) 359:1317–29. doi: 10.1056/NEJMoa0804656
9. Stroke Unit Trialists' Collaboration. Organised inpatient (stroke unit) care for stroke. *Cochrane Database Syst Rev.* (2013) 2013:CD000197. doi: 10.1002/14651858.CD000197.pub3
10. Hankey GJ. Preventable stroke and stroke prevention. *J Thromb Haemost.* (2005) 3:1638–45. doi: 10.1111/j.1538-7836.2005.01427.x
11. Fjaertoft H, Indredavik B, Magnussen J, Johnsen R. Early supported discharge for stroke patients improves clinical outcome. Does it also reduce use of health services and costs? One-year follow-up of a randomized controlled trial. *Cerebrovasc Dis.* (2005) 19:376–83. doi: 10.1159/000085543
12. Elwood D, Rashbaum I, Bonder J, Pantel A, Berliner J, Yoon S, et al. Length of stay in rehabilitation is associated with admission neurologic deficit and discharge destination. *PM R.* (2009) 1:147–51. doi: 10.1016/j.pmrj.2008.10.010
13. Frank M, Conzelmann M, Engelter S. Prediction of discharge destination after neurological rehabilitation in stroke patients. *Eur Neurol.* (2010) 63:227–33. doi: 10.1159/000279491
14. *Registry ASC*. AUSCR (2016). Available online at: www.auscr.com.au (accessed June, 2013).
15. Ruuskanen EI, Laihosalo M, Kettunen J, Losoi H, Nurmi L, Koivisto AM, et al. Predictors of discharge tohome after thrombolytic treatment in right hemisphere infarct patients. *J Cent Nerv Syst Dis.* (2010) 2:73–9. doi: 10.4137/JCNSD.S6411
16. *National Stroke Services Frameworks: Stroke Foundation.* (2017). Available online at: https://strokefoundation.org.au/what-we-do/treatment-programs/clinical-guidelines/national-stroke-services-frameworks (accessed June, 2017).
17. *Victorian Emergency Minimum Dataset: Department of Health and Human Services, State Government of Victoria.* (2015). Available online at: https://www2.health.vic.gov.au/hospitals-and-health-services/data-reporting/health-data-standards-systems/data-collections/vemd (accessed June, 2015).
18. Thrift AG, Dewey HM, Macdonell RA, McNeil JJ, Donnan GA. Stroke incidence on the east coast of Australia: the North East Melbourne Stroke Incidence Study (NEMESIS). *Stroke.* (2000) 31:2087–92. doi: 10.1161/01.STR.31.9.2087
19. Leyden JM, Kleinig TJ, Newbury J, Castle S, Cranefield J, Anderson CS, et al. Adelaide stroke incidence study: declining stroke rates but many preventable cardioembolic strokes. *Stroke.* (2013) 44:1226–31. doi: 10.1161/STROKEAHA.113.675140
20. Tibshirani R. Regression shrinkage and selection via the lasso. *J Roy Stat Soc B Methodol.* (1996) 58:267–88. doi: 10.1111/j.2517-6161.1996.tb02080.x
21. Hall R, Mondor L, Porter J, Fang J, Kapral MK. Accuracy of administrative data for the coding of acute stroke and TIAs. *Canad J Neurol Sci.* (2016) 43:765–73. doi: 10.1017/cjn.2016.278
22. Islam MS, Anderson CS, Hankey GJ, Hardie K, Carter K, Broadhurst R, et al. Trends in incidence and outcome of stroke in Perth, Western Australia during 1989 to 2001: the Perth Community Stroke Study. *Stroke.* (2008) 39:776–82. doi: 10.1161/STROKEAHA.107.493643
23. Schlegel DJ, Tanne D, Demchuk AM, Levine SR, Kasner SE. Prediction of hospital disposition after thrombolysis for acute ischemic stroke using the National Institutes of Health Stroke Scale. *Arch Neurol.* (2004) 61:1061–4. doi: 10.1001/archneur.61.7.1061

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX

Factors for Discharge to Nursing Home vs Rehabilitation and Death vs All Other Discharge Destinations for Patients With TIA Are Not Presented as the Predictive Model Is Unstable.

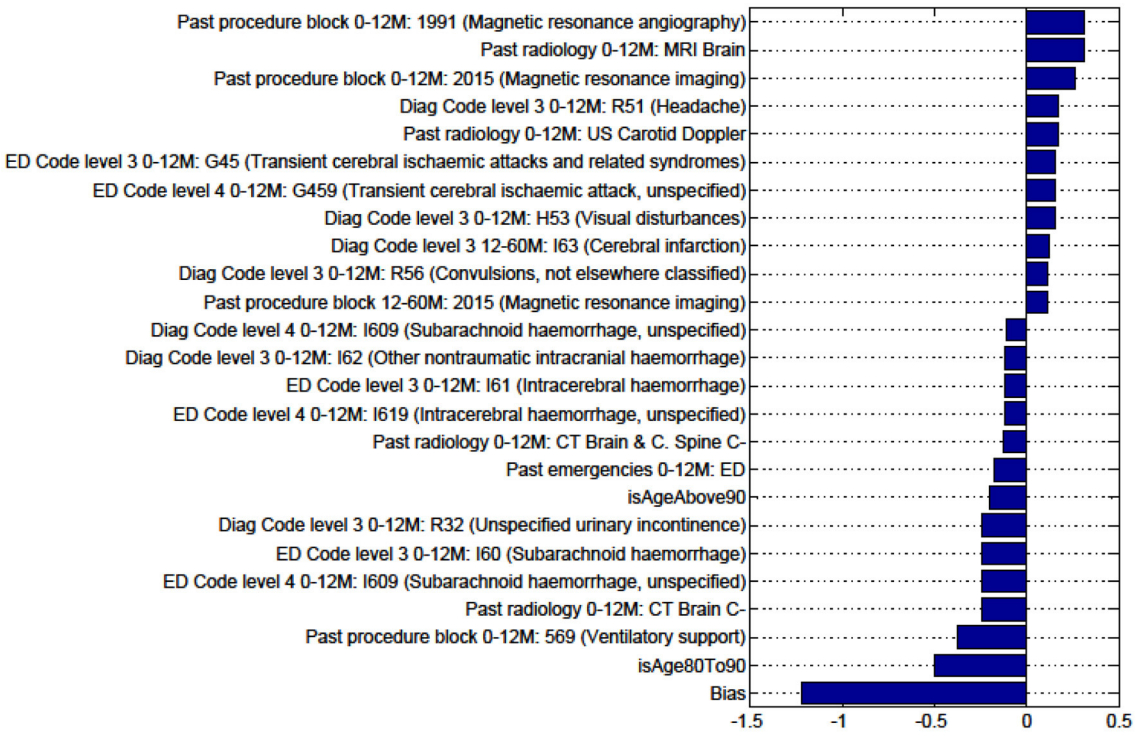


FIGURE A1 | Factors for prediction of Discharge Home vs all Other discharge destinations for the sub-cohort of Intracerebral haemorrhage stroke patients.

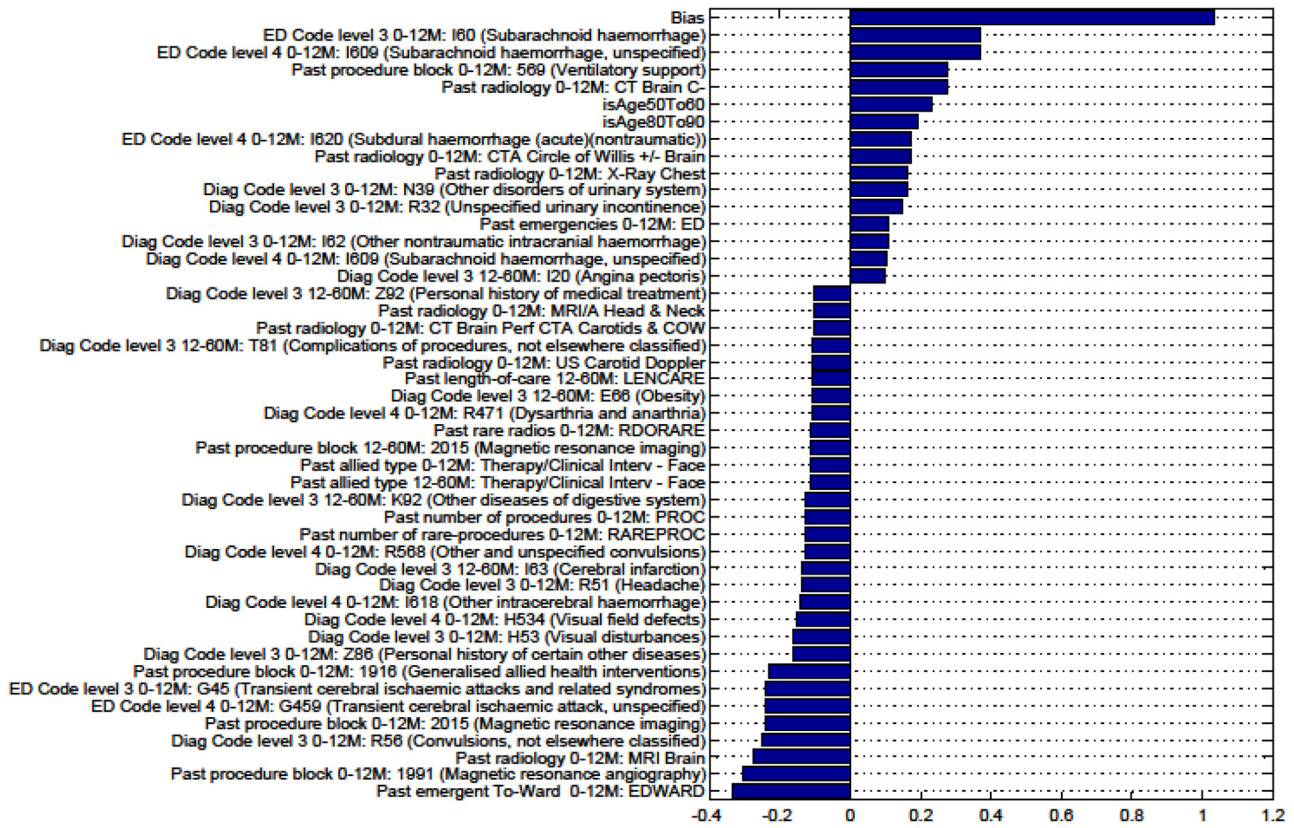


FIGURE A2 | Factors for prediction of Discharge to Rehabilitation vs Home for the sub-cohort of intracerebral hemorrhage stroke patients.

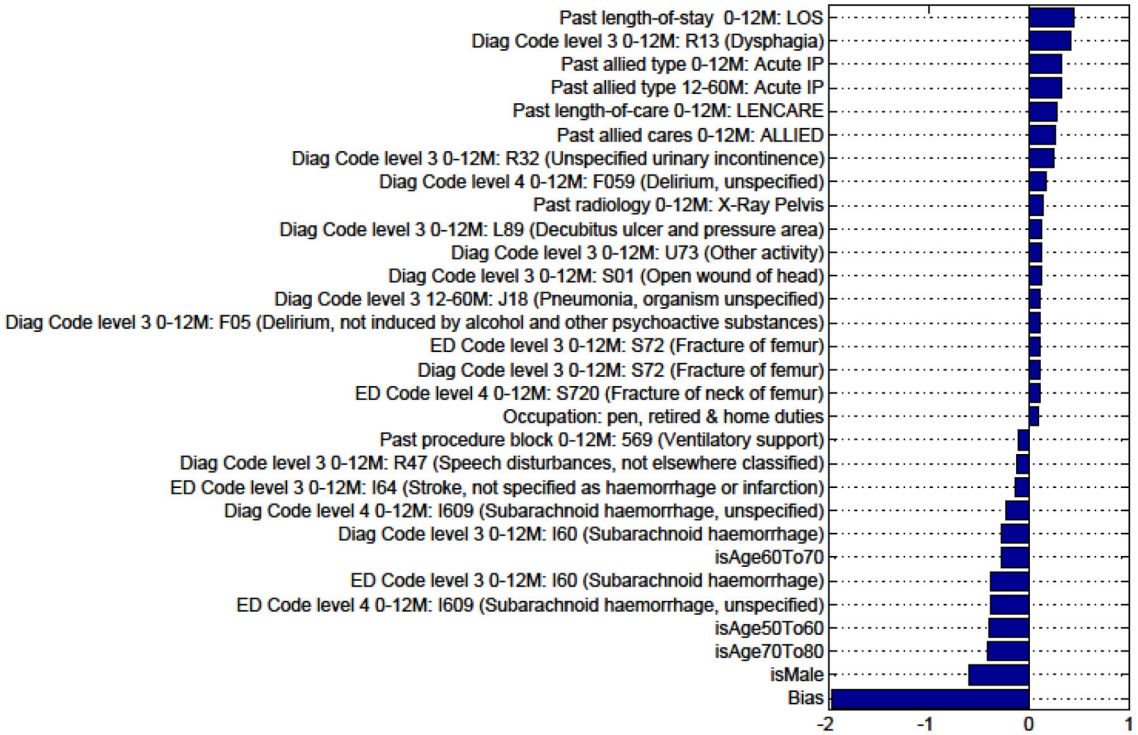


FIGURE A3 | Factors for prediction of Discharge to Nursing Home vs Rehabilitation for the sub-cohort of intracerebral hemorrhage stroke patients.

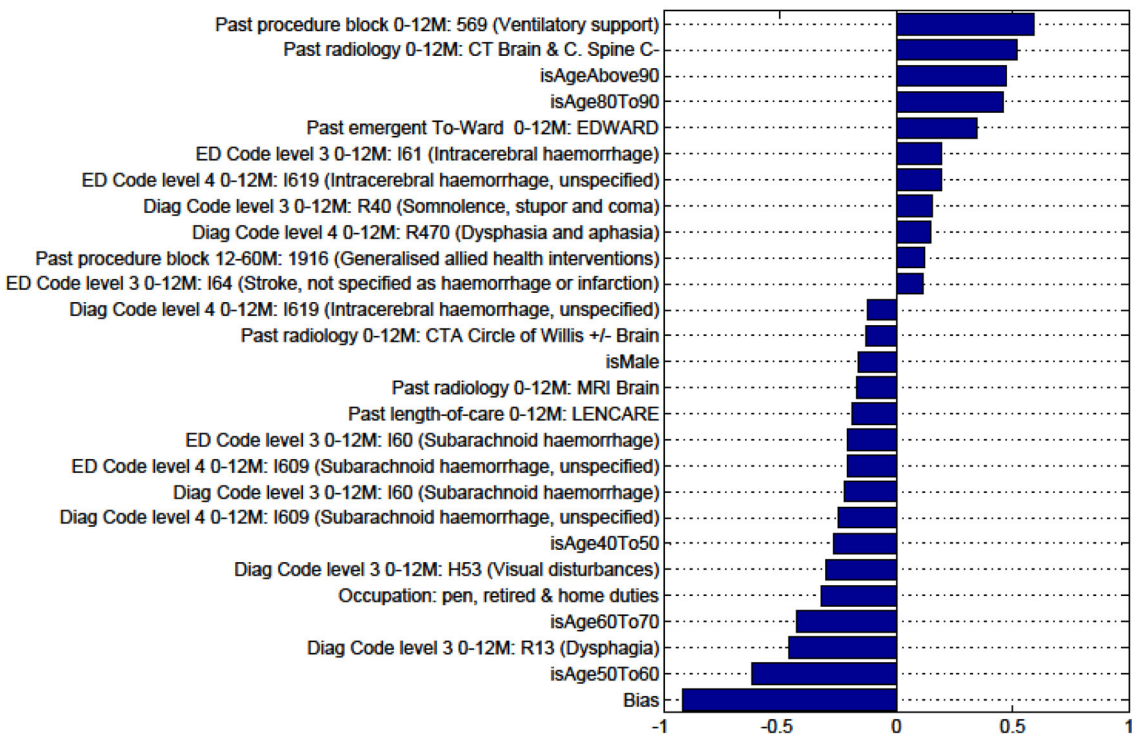


FIGURE A4 | Factors for prediction of Death vs all Other discharge destinations for the sub-cohort of intracerebral hemorrhage stroke patients.

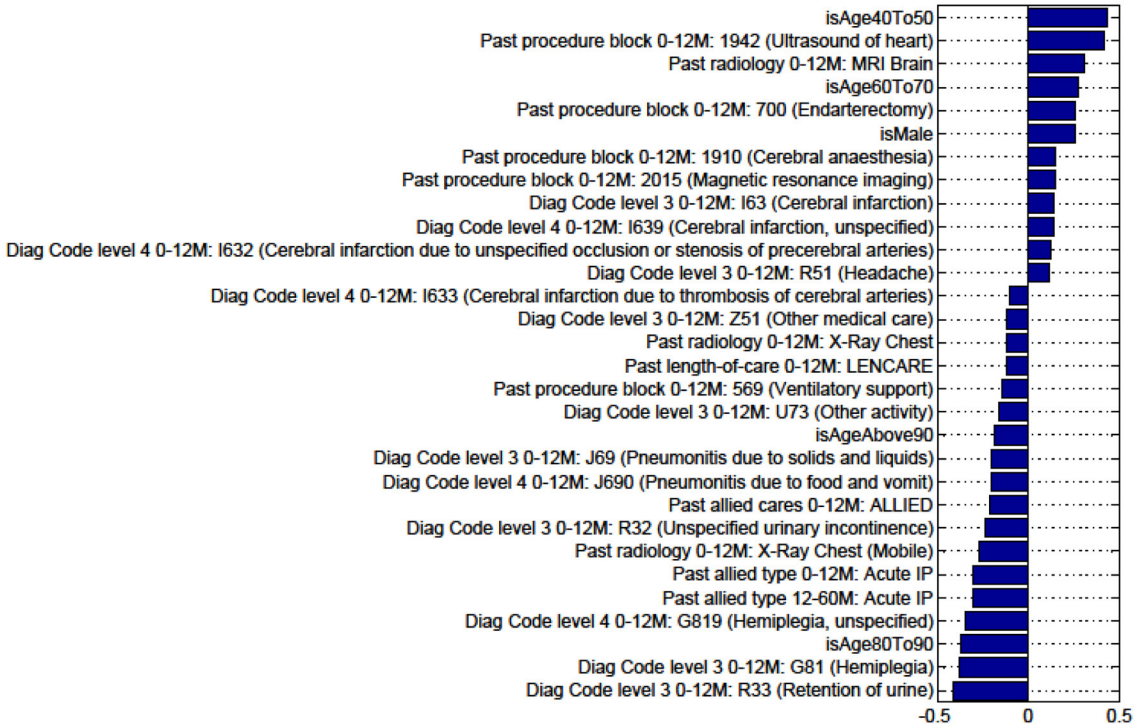


FIGURE A5 | Factors for prediction of Discharge Home vs all Other discharge destinations for the sub-cohort of ischaemic stroke patients.

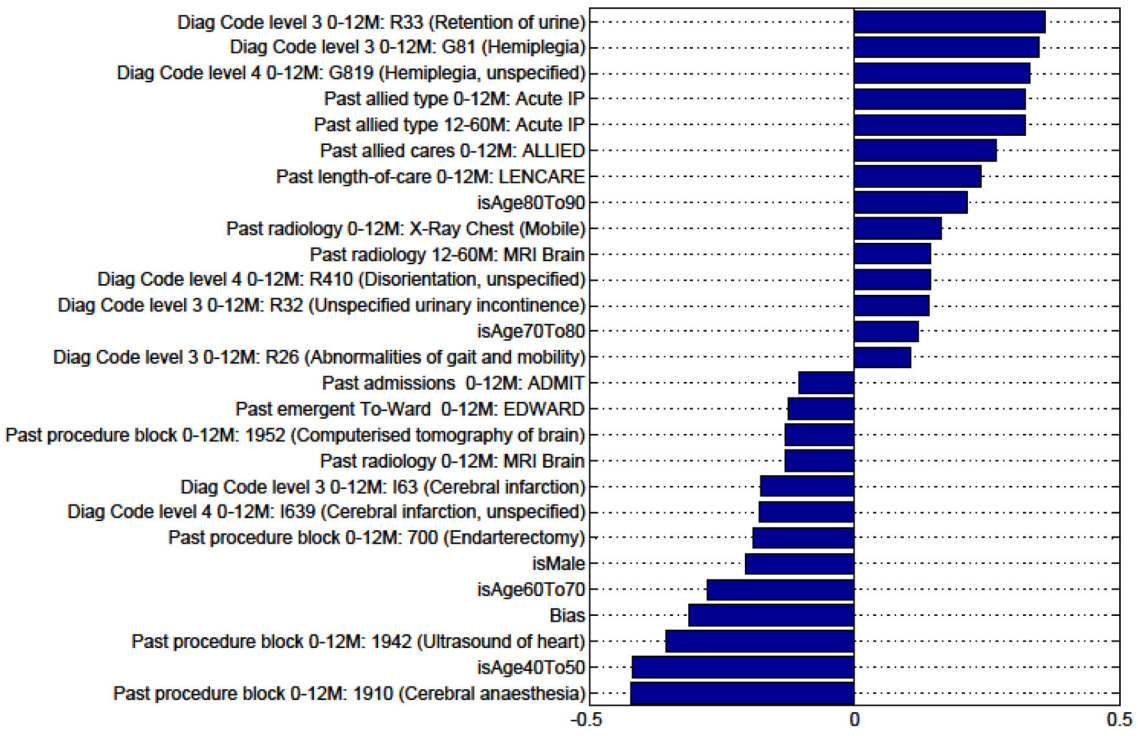


FIGURE A6 | Factors for prediction of Discharge to Rehabilitation vs Home for the sub-cohort of ischemic stroke patients.

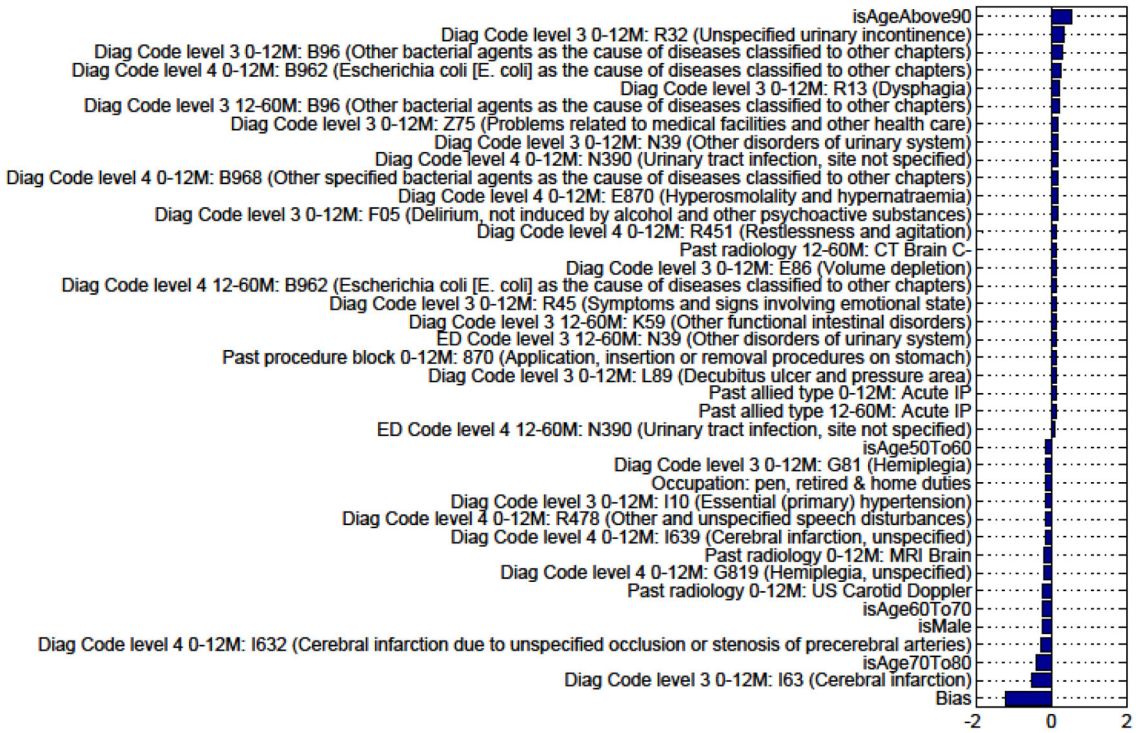


FIGURE A7 | Factors for prediction of Discharge to Nursing Home vs Rehabilitation for the sub-cohort of ischemic stroke patients.

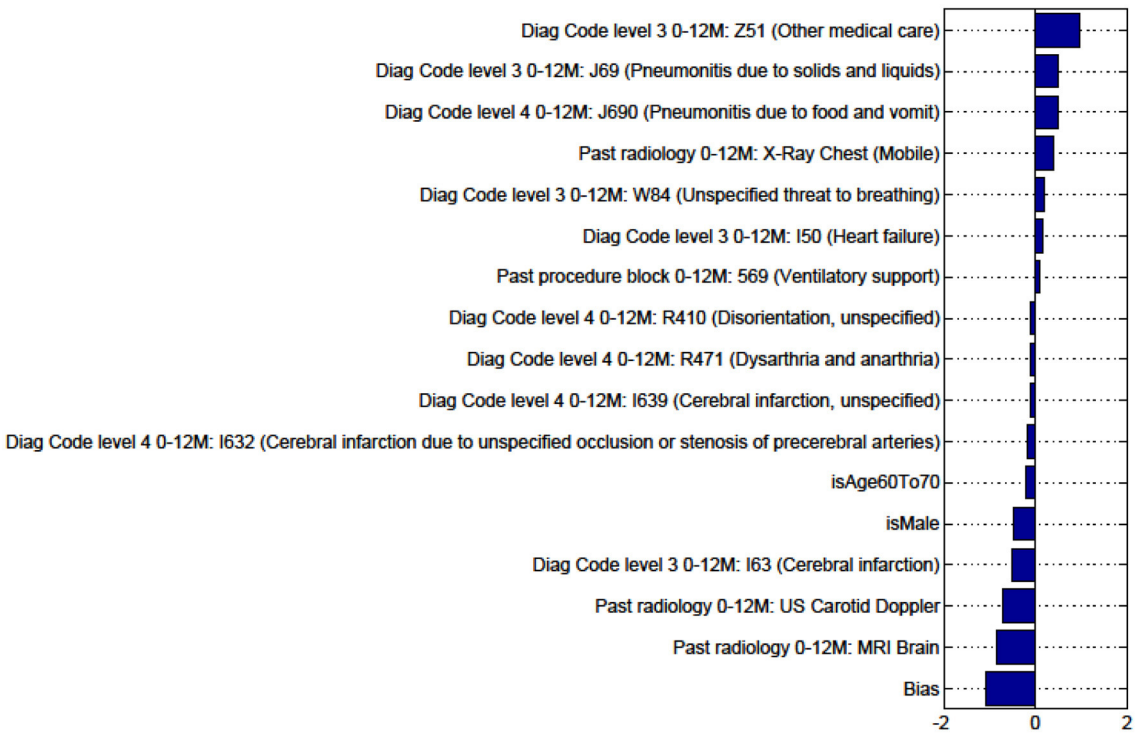


FIGURE A8 | Factors for prediction of Death vs Discharge to all Other discharge destinations for the sub-cohort of ischemic stroke patients.

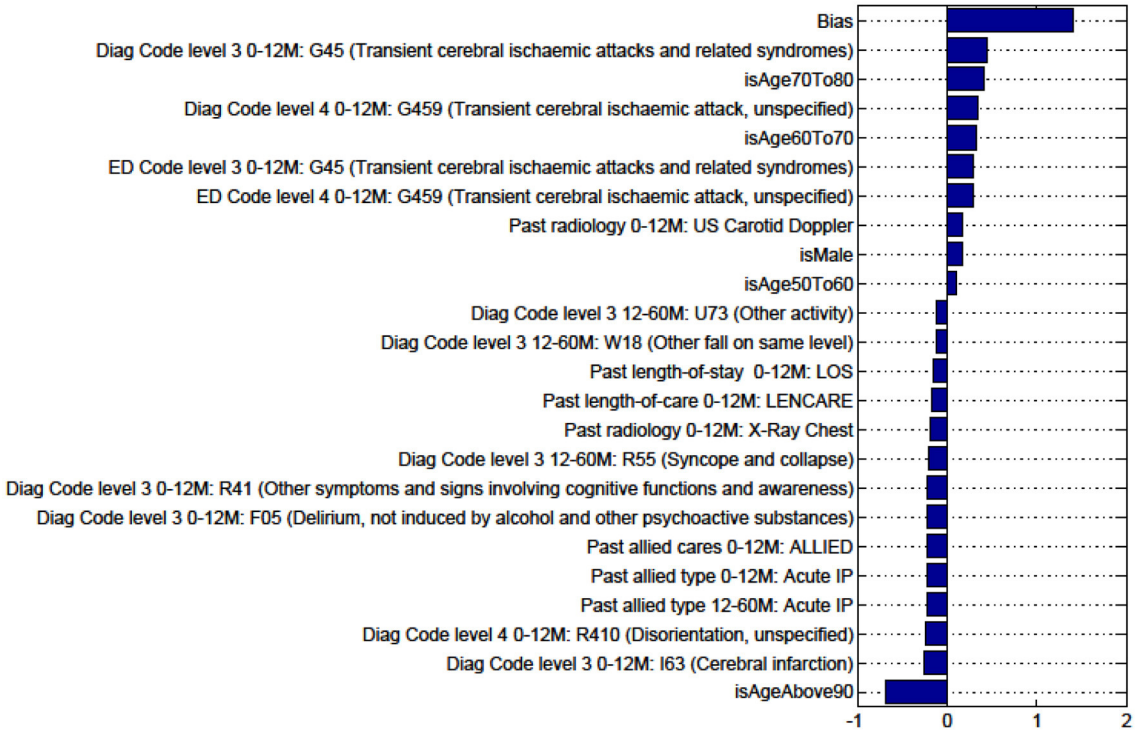


FIGURE A9 | Factors for prediction of Discharge Home vs all Other discharge destinations for the sub-cohort with TIA.

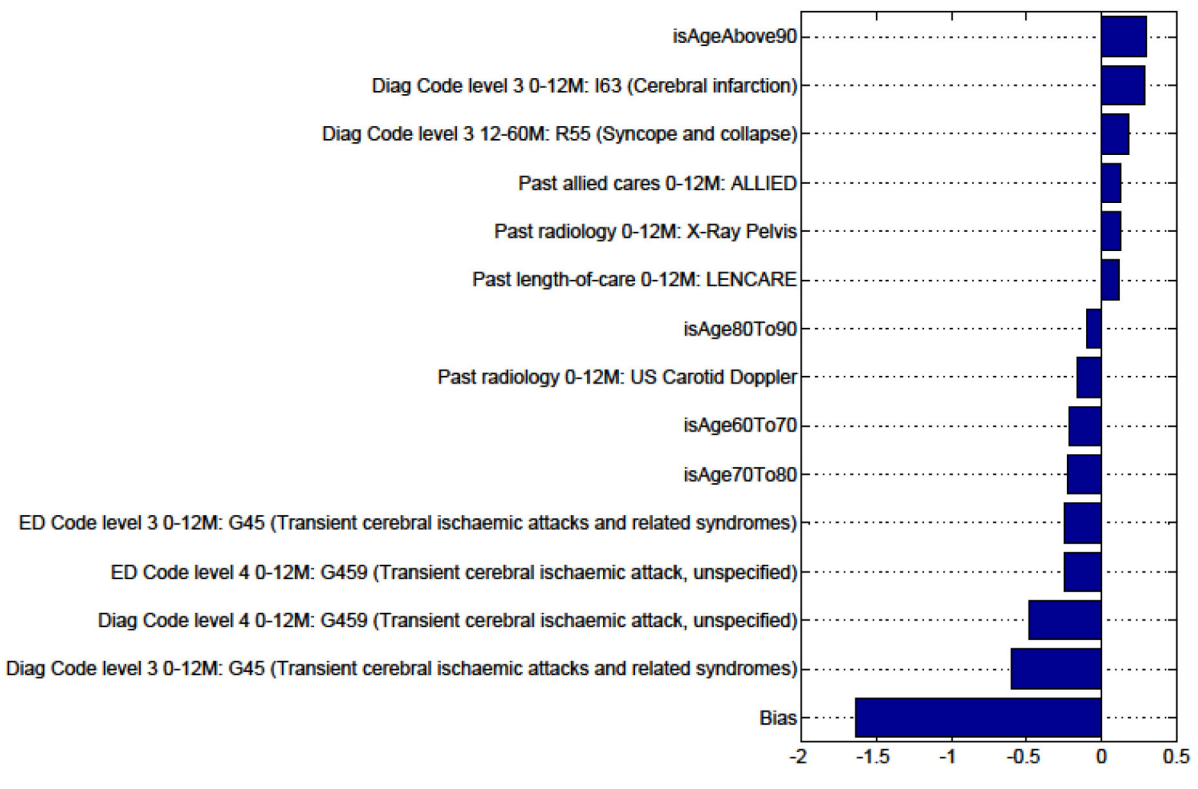


FIGURE A10 | Factors for prediction of Discharge to Rehabilitation vs Home for the sub-cohort with TIA.

frontiers in Neurology

ORIGINAL RESEARCH
published: 02 December 2021
doi: 10.3389/fneur.2021.749599

Machine Learning-Based Model for Predicting Incidence and Severity of Acute Ischemic Stroke in Anterior Circulation Large Vessel Occlusion

Junzhao Cui[1†], Jingyi Yang[2†], Kun Zhang[1], Guodong Xu[3], Ruijie Zhao[4], Xipeng Li[4], Luji Liu[1], Yipu Zhu[1], Lixia Zhou[5], Ping Yu[1], Lei Xu[1], Tong Li[1], Jing Tian[1], Pandi Zhao[1], Si Yuan[1], Qisong Wang[1], Li Guo[1] and Xiaoyun Liu[1,6]*

[1] Department of Neurology, The Second Hospital of Hebei Medical University, Shijiazhuang, China, [2] Department of Information Center, The Second Hospital of Hebei Medical University, Shijiazhuang, China, [3] Department of Neurology, Hebei Province People's Hospital, Shijiazhuang, China, [4] Department of Neurology, Xingtai People's Hospital, Xingtai, China, [5] Department of Medical Iconography, The Second Hospital of Hebei Medical University, Shijiazhuang, China, [6] Neuroscience Research Center, Medicine and Health Institute, Hebei Medical University, Shijiazhuang, China

OPEN ACCESS

Edited by:
Jiang Li,
Geisinger Medical Center,
United States

Reviewed by:
Durgesh Prasad Chaudhary,
Geisinger Health System,
United States
Akram Mohammed,
University of Tennessee Health Science Center (UTHSC),
United States

***Correspondence:**
Xiaoyun Liu
audrey-l@163.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Stroke, a section of the journal Frontiers in Neurology

Received: 29 July 2021
Accepted: 29 October 2021
Published: 02 December 2021

Citation:
Cui J, Yang J, Zhang K, Xu G, Zhao R, Li X, Liu L, Zhu Y, Zhou L, Yu P, Xu L, Li T, Tian J, Zhao P, Yuan S, Wang Q, Guo L and Liu X (2021) Machine Learning-Based Model for Predicting Incidence and Severity of Acute Ischemic Stroke in Anterior Circulation Large Vessel Occlusion. Front. Neurol. 12:749599. doi: 10.3389/fneur.2021.749599

Objectives: Patients with anterior circulation large vessel occlusion are at high risk of acute ischemic stroke, which could be disabling or fatal. In this study, we applied machine learning to develop and validate two prediction models for acute ischemic stroke (Model 1) and severity of neurological impairment (Model 2), both caused by anterior circulation large vessel occlusion (AC-LVO), based on medical history and neuroimaging data of patients on admission.

Methods: A total of 1,100 patients with AC- LVO from the Second Hospital of Hebei Medical University in North China were enrolled, of which 713 patients presented with acute ischemic stroke (AIS) related to AC- LVO and 387 presented with the non-acute ischemic cerebrovascular event. Among patients with the non-acute ischemic cerebrovascular events, 173 with prior stroke or TIA were excluded. Finally, 927 patients with AC-LVO were entered into the derivation cohort. In the external validation cohort, 150 patients with AC-LVO from the Hebei Province People's Hospital, including 99 patients with AIS related to AC- LVO and 51 asymptomatic AC-LVO patients, were retrospectively reviewed. We developed four machine learning models [logistic regression (LR), regularized LR (RLR), support vector machine (SVM), and random forest (RF)], whose performance was internally validated using 5-fold cross-validation. The performance of each machine learning model for the area under the receiver operating characteristic curve (ROC-AUC) was compared and the variables of each algorithm were ranked.

Results: In model 1, among the included patients with AC-LVO, 713 (76.9%) and 99 (66%) suffered an acute ischemic stroke in the derivation and external validation cohorts, respectively. The ROC-AUC of LR, RLR and SVM were significantly higher than that of the RF in the external validation cohorts [0.66 (95% CI 0.57–0.74) for LR, 0.66 (95% CI 0.57–0.74) for RLR, 0.55 (95% CI 0.45–0.64) for RF and 0.67 (95% CI 0.58–0.76) for SVM]. In model 2, 254 (53.9%) and 31 (37.8%) patients suffered disabling ischemic stroke

in the derivation and external validation cohorts, respectively. There was no difference in AUC among the four machine learning algorithms in the external validation cohorts.

Conclusions: Machine learning methods with multiple clinical variables have the ability to predict acute ischemic stroke and the severity of neurological impairment in patients with AC-LVO.

Keywords: anterior circulation large vessel occlusion, acute ischemic stroke, machine learning, prediction model, neurological impairment

INTRODUCTION

Acute ischemic stroke caused by large vessel occlusion accounts for more than 40% of cases, ∼80% of which occurs in the anterior circulation (1). Compared to non-large vessel occlusion (LVO) acute ischemic stroke (AIS), patients with anterior circulation large vessel occlusion (AC-LVO) stroke are considered to be at greater risk of mortality or disability before endovascular treatment (2). They tend to improve significantly after mechanical thrombectomy (3, 4). Previously reported prediction models for AC-LVO stroke such as prehospital scales (Prehospital Acute Stroke Severity scale, PASS; Cincinnati Prehospital Stroke Severity Scale, CPSSS; stroke Vision Aphasia Neglect, VAN; Rapid Arterial Occlusion Evaluation scale RACE and Field Assessment Stroke Tri-age for Emergency Destination, FAST-ED) (5–9) that are based on NIHSS, and the recently proposed model by Philipp Hendrix et al., which combines past medical history and neurologic examination (10), have focused on the identification of large vessel occlusion in patients with AIS. The main clinical purpose of the prediction scores is to identify which patients with AIS have LVO so that they can be referred to capable centers for endovascular treatment (EVT). However, accurate prediction of AIS in patients with AC-LVO remains a challenge.

Anterior circulation-LVO stroke can be further divided based on pathogenesis and severity of clinical consequences, into non-disabling and disabling stroke with the latter frequently resulting in post-stroke dependence. Nevertheless, no previous studies have predicted the risk of disabling ischemic stroke in patients with AC-LVO, which may be useful in treatment decisions and prevention.

In this study, we developed and validated two models based on machine learning algorithms with clinical variables, to predict acute ischemic stroke (Model 1) and severity of neurological impairment (Model 2) in patients with AC-LVO.

METHODS

Patient Cohorts

A total of 1,100 patients with AC- LVO admitted between June 2016 and April 2018 at the Second Hospital of Hebei Medical University, North China, were registered in the derivation cohort; 927 of them who presented with AIS related with AC-LVO and asymptomatic AC-LVO were retrospectively reviewed. In addition, 471 patients with first-ever ischemic stroke (including disabling and non-disabling stroke) were selected. For the external validation, we collected data of patients with AC-LVO from Hebei Province People's Hospital, China between September 2016 and April 2021.

Anterior circulation-LVO was defined as complete occlusion of at least one intracranial internal carotid artery (ICA) or middle cerebral artery (MCA) visualized on computed tomography angiography (CTA) or magnetic resonance angiography (MRA). ICA occlusion refers to the complete occlusion of the C1–C7 segment of the internal carotid artery based on CTA or MRA. MCA occlusion refers to the occlusion of the MCA involving at least the M1 segment (for more details please see in **Supplementary Figure I**). Asymptomatic AC-LVO was defined as the absence of a transient ischemic attack (TIA), amaurosis fugax, and ischemic stroke attributed to anterior circulation large vessel (11, 12). In accordance with previous studies, disabling and non-disabling ischemic strokes were defined by the initial clinician as National Institutes of Health Stroke Scale (NIHSS) > 5 and ≤ 5 on admission, respectively (13).

Data Collection and Variable Selection

Patient characteristics that were collected on admission for the development of Models 1 and 2 include (1) demographic data of the patients such as the age, sex, body mass index (BMI), current smoking and drinking status, comorbidity (hypertension, coronary atherosclerotic heart disease, atrial fibrillation, diabetes mellitus, and hyperlipidemia), history of transient ischemic attack (TIA); (2) clinical variables such as serum apolipoprotein B (Apo B) and homocysteine on arrival; (3) imaging variables such as occluded vessels (unilateral MCA, unilateral ICA, and multiple arteries), posterior circulation large vessel severe stenosis ($\geq$ 70%) /occlusion, anterior cerebral artery (ACA) occlusion, and Alberta Stroke Program Early CT Score (ASPECTS). Data on 14 variables were included in Model 1, and on 12 in Model 2. Specifically, normal blood flow status of the vertebrobasilar arteries *via* the posterior communicating artery plays a major role in primary collateral compensation after anterior circulation large vessel occlusion. Therefore, posterior circulation large vessel stenosis/occlusion was introduced into Model 1. Posterior circulation large vessel refers to the intracranial vertebral artery, basilar artery, or segment P1 of the posterior cerebral artery.

Data Pre-processing

Processing of the data was performed using Python. First, records containing outliers, which were identified by boxplot, were excluded. Furthermore, the median imputation method was used to impute missing values in derivation cohorts. Finally, the

categorical variables were converted into numerical values with dummy encoding, and the continuous features were standardized by removing the mean and scaling to unit variance.

Prediction Models With Machine Learning

Machine learning is a discipline that constructs models base on data, which is a part of artificial intelligence. Machine learning extracts the characteristics and abstracts the model of the data, discovers the information in the data, and then analyzes and predicts it. First, an algorithm and some parameters of the model which were supplied with training data were selected arbitrarily. During training procedures, the model automatically adjusts some trainable parameters stage by stage to achieve better performance optimization. After the training, all the model parameters are fixed. Importantly, the true effectiveness of the model was evaluated using test data that were completely separate from the training data.

We selected logistic regression without regularization (LR), regularized logistic regression (RLR), random forest (RF), and support vector machine (SVM) as machine learning algorithms that are commonly used.

Logistic regression, a classic classification algorithm in machine learning, was regularized using a combination of L1 and L2 loss in this study. Here, the target was determined by Y:

$$\begin{aligned} Y = \{&\text{“disabling ischemic stroke,”} \\ &\text{“non-disabling ischemic stroke”}\} \end{aligned}$$

$$Z = W^{T}X + b$$

$$y = \frac{1}{1 + e^{-Z}} = \frac{1}{1 + e^{-W^{T}X+b}}$$

We selected binary cross-entropy loss as the cost function, where y is the ground truth, y^ is the predicted score of the model, and R represents the regularization. The loss functions L1 and L2 are defined as follows:

$$J(w, b) = \frac{1}{m}\sum_{i=1}^{m} L\left(\widehat{y^{(i)}}, y^{(i)}\right)$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left(-y^{(i)}\log\widehat{y^{(i)}} - \left(1 - y^{(i)}\right)\log\left(1 - \widehat{y^{(i)}}\right)\right) + \frac{\lambda}{2}R$$

$$R_{L1} = \sum_{j=1}^{m}\left|W_j\right| ....R_{L2} = \sum_{j=1}^{m}\left[W_j^2\right]$$

In the training process of the model, standardization of numerical variables was carried out to accelerate the convergence process and speed of the model.

Random forest is an extended variant of bagging, which uses a decision tree as the base learner and introduces the selection of random attributes in the training process of the decision tree (14). The main parameters that can affect the model performance in RF include the number of trees in the forest, maximum depth of the tree, minimum number of samples required to split an internal node, minimum number of samples required to be at a leaf node, and function to measure the quality of a split. In this study, the values in the dataset were discretized, and the parameters were optimized with a grid search during the training process.

An SVM classifies data by calculating the maximum-margin hyperplane, which adds a regularization term in the solving process to optimize the structural risk. The strength of SVM is that it can process complex datasets with many variables or dimensions (15). The validity of SVM depends mainly on the selection of the kernel function, parameters of the kernel, and soft margin parameter C. Otherwise, in this study, each combination of parameter selections was checked using cross-validation, and only parameters with optimal accuracy were selected.

Moreover, LR, RLR, RF, and SVM can estimate the contribution of each feature to the model by calculating the absolute value of the standardized regression coefficient, information gain / Gini coefficient, and weight coefficient.

Model Derivation and Internal Validation

In this study, for model derivation, we adopted 5-fold cross-validation, which is a standard way of optimizing the model with inner test data and has been used in a previous study (16). During modeling, the grid search algorithm which is a greedy algorithm was combined to tune and optimize the model hyperparameters. For each group of hyperparameters, we selected 5-fold cross-validation to determine the optimal ones, after which we calculated the means of sensitivity, specificity, accuracy, and AUC to evaluate the performance of each model (**Figure 1**).

The derivation and validation models were conducted using Python 3.6. The model algorithms, cross-validation, and grid search were based on the Scikit-Learn library of Python in the PyCharm. Matplotlib 3.3.3, NumPy 1.19.5, pandas version 1.1.5, and Scikit-Learn toolkit version 0.21.0 were used to train the machine learning models.

External Validation

After internal training and testing, the performance of the model was evaluated using external validation data. Subsequently, the AUCs were compared among machine learning algorithms.

Statistical Analysis

Clinical variables are presented as mean ± *SD* or median with interquartile range, depending on the distribution of the variables. To compare the group differences, continuous variables were compared using the Student's *t*-test or Mann-Whitney *U* test, and categorical variables were compared with the $\chi 2$ test or Fisher's exact test. These two prediction models were discriminated against using AUC. Calculation of AUC, sensitivity, specificity, precision, negative predictive value (NPV), and accuracy criteria were performed with R statistical software version 3.0.2. The area under the precision-recall curve (PRC) and F1-score were calculated with MedCalc. For the derivation and validation cohorts, a comparison of AUC among the machine learning methods was performed using the DeLong test with Bonferroni correction. Two-sided $P < 0.05$ were considered statistically significant.

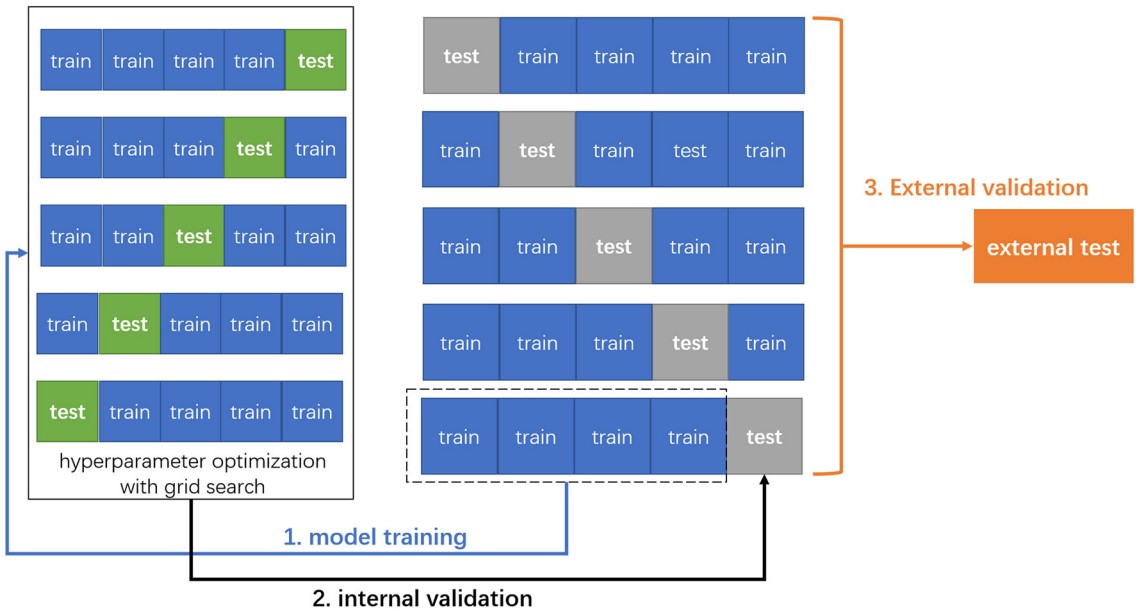


FIGURE 1 | The schema of 5-fold cross-validation and external validation. First, the original data were randomly divided into five patterns without duplication, one of which was used as the test set, and the remaining four as the training set for model training. Next, we adapted grid search with 5-fold cross-validation to optimize the hyperparameters for each machine learning model. Finally, the trained models were externally validated on the external test data.

RESULTS

Baseline Characteristics

Figures 2A,B illustrate the flow diagram of the enrolled patients. For the derivation cohort, 1,100 patients with AC-LVO were hospitalized at the study institution. After excluding 173 patients with prior stroke or TIA as non-acute ischemic cerebrovascular events, 713 patients with AIS related with AC-LAO and 214 with asymptomatic AC-LAO were finally included in the analysis for model 1. Among the 214 patients with asymptomatic AC-LVO, 119 (56%) were hospitalized for head discomfort such as heaviness of the head and fullness in the head. The other reasons for hospitalization in patients with asymptomatic AC-LVO included coronary artery disease, subarachnoid hemorrhage, migraine, cerebral large artery disease detected by routine physical examination, unruptured intracranial aneurysms, diabetic peripheral vascular disease, central nervous system infection, lower extremity atherosclerotic occlusive disease, intracranial space-occupying lesions, epilepsy, Parkinson's disease, cerebral atrophy, subclavian artery steal blood syndrome, cough syncope, and cardiac syncope. The general screening of large artery disease was performed with transcranial Doppler and carotid artery ultrasound in these patients. Further computed tomography angiography (CTA) or magnetic resonance angiography (MRA) examinations were conducted and AC-LVO was identified. Among the 713 patients with AIS, 242 with prior stroke were excluded, and 471 patients with first-ever ischemic stroke (254 with disabling and 217 with non-disabling strokes) were included in the analysis for Model 2. For the external validation cohort, 150 eligible patients with AC-LVO were included in Model 1. Of the 99 patients with AIS, 82 who presented with the first episode were included in the analysis for model 2. The baseline characteristics of the included patients are presented in **Tables 1**, **2**, and **Supplementary Tables I–IV**.

Comparison Between the Models in the Derivation Cohort

The performance metrics of each approach for Models 1 and 2 in the derivation cohort are shown in **Tables 3**, **4**, respectively. The receiver operating characteristic (ROC) curve (indicating the predictive performance of our LR/RLR/RF/SVM model) for each algorithm in the two models and the comparison among these machine learning algorithms are shown in **Figures 4A,C**. In model 1, the AUCs of RF and SVM were significantly higher than those of the LR and RLR, when using the DeLong test with Bonferroni correction (RF vs. LR, $P < 0.0001$; RF vs. RLR, $P < 0.0001$; SVM vs. LR, $P < 0.0001$; SVM vs. RLR, $P < 0.0001$; **Figure 3A**). Similar results were obtained for accuracy and F1-score. In model 2, while the differences in AUCs among the four machine learning algorithms were not significant (**Figure 3C**), the RF showed the most perfect classification accuracy (71.8%) compared to that of the other machine learning approaches.

Comparison Between the Models in the External Validation Cohort

The ROC curves for Models 1 and 2 in the external validation cohort are shown in **Figures 4B,D**. In Model 1, RF exhibited the worst performance among the machine learning models (**Table 5**). The AUCs in LR, RLR and SVM were significantly

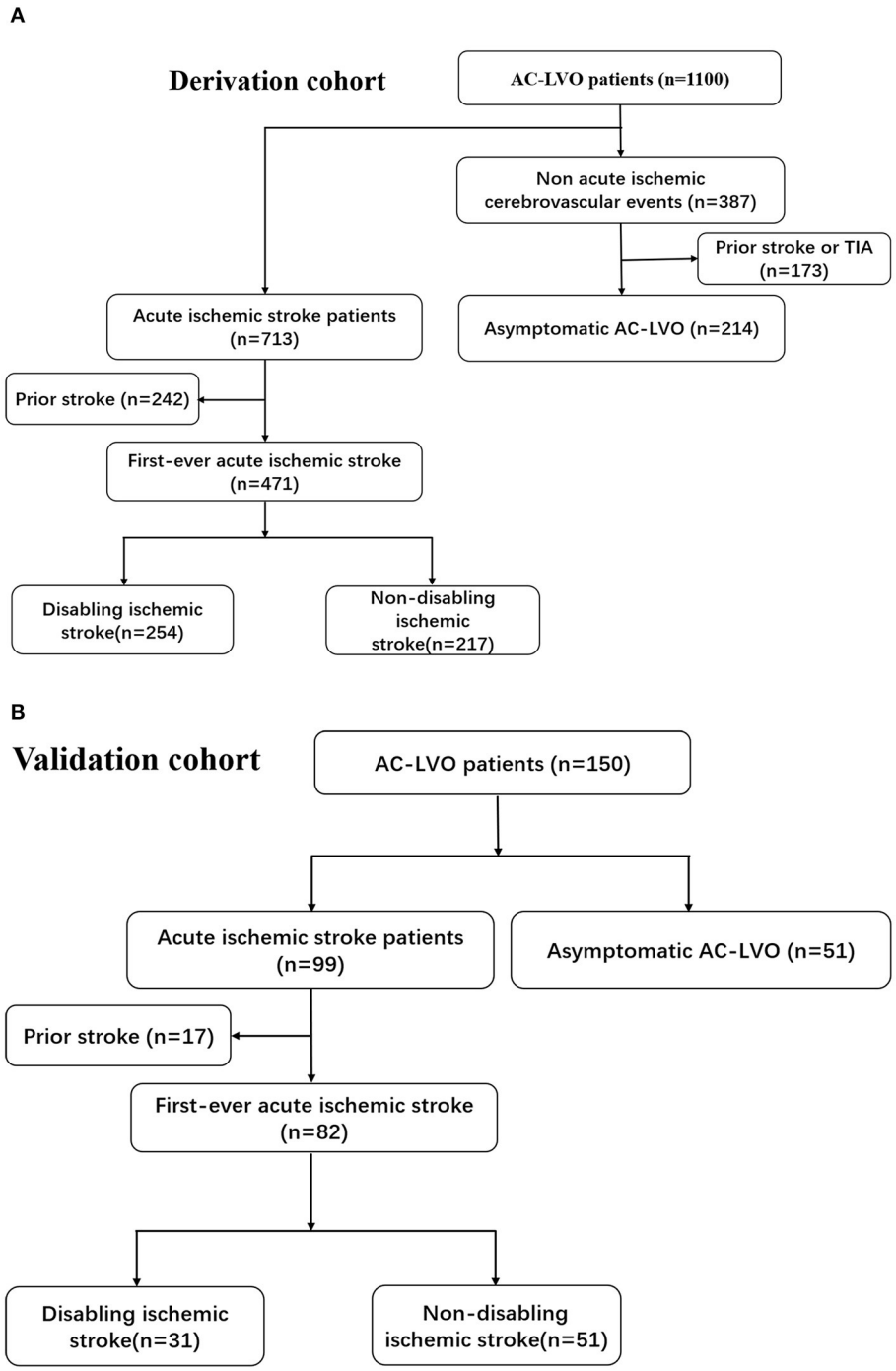


FIGURE 2 | The flow diagram of the patients included in this study is shown in **(A,B)**.

higher than that in RF, when using the Delong test with Bonferroni correction (LR vs. RF, $P = 0.0048$; RLR vs. RF, $P = 0.0048$, SVM vs. RF, $P = 0.0006$; **Figure 3B**). In Model 2, there was no difference in AUCs among the four machine learning algorithms (**Figure 3D**). The AUC of each algorithm was as follows: LR 0.68 (95% CI 0.56–0.8), RLR 0.76 (95% CI 0.66–0.87), RF 0.71 (95% CI 0.59–0.83) and SVM 0.77 (95% CI 0.66–0.87) (**Table 6**).

Important Variables of the Machine Learning Models

After calculating the importance of each feature, the top five selected variables of Models 1 and 2 were ranked by their discriminative performance (**Figures 5**, **6**). For LR and RLR, the absolute value of the standardized regression coefficient was calculated in both models. For RF, the important features for information gain and Gini coefficient were ranked in Models 1

TABLE 1 | Baseline characteristics of patients with anterior circulation large vessel occlusion.

	Derivation cohort (n = 927)	Validation cohort (n = 150)
Mean ± SD age, y	59.3 ± 12.4	62.0 ± 10.9
Males *n* (%)	626 (67.5)	108 (72.0)
Current smoking *n* (%)	268 (28.9)	57 (38.0)
Drinking *n* (%)	267 (28.8)	26 (17.3)
Hypertension *n* (%)	602 (64.9)	107 (71.3)
Coronary atherosclerotic heart disease *n* (%)	143 (15.4)	22 (14.7)
Atrial fibrillation *n* (%)	52 (5.6)	11 (7.3)
Diabetes *n* (%)	225 (24.3)	52 (34.7)
Hyperlipidaemia *n* (%)	292 (31.5)	14 (9.3)
Occluded vessels *n* (%) Unilateral MCA	442 (47.7)	51 (34.0)
Unilateral ICA	261 (28.2)	69 (46.0)
Multiple artery	224 (24.2)	30 (20.0)
BMI≥24 *n* (%)	795 (85.8)	108 (72.0)
Posterior circulation large vessel severe stenosis /occlusion n (%)	176 (19.0)	9 (6.0)
ApoB (g/L)	1.0 ± 0.3	0.7 ± 0.2
Homocysteine (μmol/L)	19.0 ± 11.2	15.9 ± 9.6
Acute ischemic stroke *n* (%)	713 (76.9)	99 (66.0)

ApoB, apolipoprotein B; BMI, body mass index; ICA, internal carotid artery; IQR, interquartile range; MCA, middle cerebral artery; SD, standard deviation.

TABLE 2 | Baseline characteristics of patients with first-ever acute ischemic stroke (AIS) caused by anterior circulation large vessel occlusion.

	Derivation cohort (n = 471)	Validation cohort (n = 82)
Mean ± SD age, y	58.6 ± 12.7	61.0 ± 13.0
Males *n* (%)	321 (68.2)	59 (72.0)
Current smoking *n* (%)	147 (31.2)	35 (42.7)
Hypertension *n* (%)	290 (61.6)	51 (62.2)
Diabetes *n* (%)	105 (22.3)	25 (30.5)
Coronary atherosclerotic heart disease *n* (%)	69 (14.6)	7 (8.5)
Prior TIA *n* (%)	30 (6.4)	0 (0)
Hyperlipidaemia *n* (%)	158 (33.5)	8 (9.8)
BMI ≥24 *n* (%)	410 (87.0)	60 (73.2)
Median ASPECTS (IQR)	6 (3–8)	8 (7–9)
Occluded vessels *n* (%) Unilateral MCA	236 (50.1)	31 (37.8)
Unilateral ICA	133 (28.2)	36 (43.9)
Multiple artery	102 (21.7)	15 (18.3)
Anterior cerebral artery occlusion n (%)	87 (18.5)	3 (3.7)
Disabling ischemic stroke (NIHSS > 5) *n* (%)	254 (53.9)	31 (37.8)

ASPECTS, Alberta Stroke Program Early CT Score; BMI, body mass index; ICA, internal carotid artery; IQR, interquartile range; MCA, middle cerebral artery; NIHSS, National Institutes of Health Stroke Scale; SD, standard deviation; TIA, transient ischemic attack.

and 2 respectively. For SVM, the absolute value of the weight was used to rank the variables only in model 2 due to the introduction of the kernel function in Model 1. The absolute values of the important metrics for the features were normalized, ensuring the comparability in feature importance ranking. In Model 1, homocysteine, occluded vessels and BMI appeared together in the top five rankings of all machine learning algorithms. In addition, coronary atherosclerotic heart disease was an important feature in both LR and RLR. Age and Apo B appeared to be important variables in RF. In Model 2, ASPECT, age and BMI were common variables for all machine learning algorithms. Prior TIA was included in LR, RLR, and RF. Hypertension, current smoking, and gender appeared in RLR, RF, and SVM, respectively. Furthermore, occluded vessels coexisted in LR and SVM.

DISCUSSION

This study demonstrated that the use of a machine learning approach can predict the risk of AIS and severity of ischemic stroke in AC-LVO from clinical data. To the best of our knowledge, this is the first report on an attempt to predict AIS and severity of neurological impairment in patients with AC-LVO using the machine learning approach. The machine learning algorithm can eliminate linearity and has various ways of overcoming the imperfections of the polyfactorial models such as overfitting of models and collinearity of variables, which may lead to a series of problems when it comes to variable selection (17). In the two prediction models in this study, 14 and 12 common variables were collected, respectively, bypassing the traditional method of variable selection.

Contrary to the findings in the derivation cohort of model 1 that RF showed significantly better predictive performance than LR and RLR, in the validation cohort, RF had the worst performance among the machine learning models. The decision trees of RF forced interactions between the features, which might make the result rather inferior if the majority of the features have no or very weak interactions. Therefore, we suspect that the RF was not able to carry on an accurate classified forecast owing to extremely weak interactions between the variables in our dataset. Moreover, the small data sets with 150 cases in the validation cohort may be another reason for the poor performance of the RF. In model 2, although the LR showed a predictive property similar to those of the other three algorithms both in the validation cohort and derivation cohort, the RLR exhibited a higher AUC compared with LR in the validation cohort, this was as a result of the poor generalization performance of LR compared with other algorithms. Accordingly, LR with L2 regularization was implemented in this study to avoid overfitting and improve the generalization performance and robustness of the model; thus, a more optimal result was obtained with an AUC of 0.76.

As shown in **Figures 5**, **6**, the important features were not entirely consistent in the machine learning algorithms in

TABLE 3 | Scores for each algorithm of model 1 in derivation cohort.

Model	AUC (95% CI)	PRC	Sensitivity	Specificity	Precision	NPV	Accuracy	$F1_{max}$
LR	0.68 (0.64–0.72)	0.88	57.9	71.5	87.1	33.8	61.1	0.87
RLR	0.68 (0.64–0.72)	0.88	57.9	71.5	87.1	33.8	61.1	0.87
RF	0.80 (0.77–0.83)	0.93	69.1	79.9	92.0	43.7	71.6	0.89
SVM	0.77 (0.74–0.81)	0.92	76.9	68.2	89.0	46.9	74.9	0.88

AUC, receiver operator characteristic area under the curve; $F1_{max}$, the maximum F1 score; LR, logistic regression without regularization; PRC, area under the precision-recall curve; RF, random forest; RLR, regularized logistic regression; SVM, support vector machine; NPV, negative predictive value.

TABLE 4 | Scores for each algorithm of model 2 in derivation cohort.

Model	AUC (95% CI)	PRC	Sensitivity	Specificity	Precision	NPV	Accuracy	$F1_{max}$
LR	0.78 (0.74–0.81)	0.78	63.0	77.9	76.9	64.3	69.9	0.76
RLR	0.75 (0.71–0.79)	0.77	61.0	78.3	76.7	63.2	69.0	0.74
RF	0.77 (0.73–0.81)	0.78	67.7	76.5	77.1	66.9	71.8	0.75
SVM	0.76 (0.71–0.80)	0.78	63.4	77.0	76.3	64.2	69.6	0.74

AUC, receiver operator characteristic area under the curve; $F1_{max}$, the maximum F1 score; LR, logistic regression without regularization; PRC, area under the precision-recall curve; RF, random forest; RLR, regularized logistic regression; SVM, support vector machine; NPV, negative predictive value.

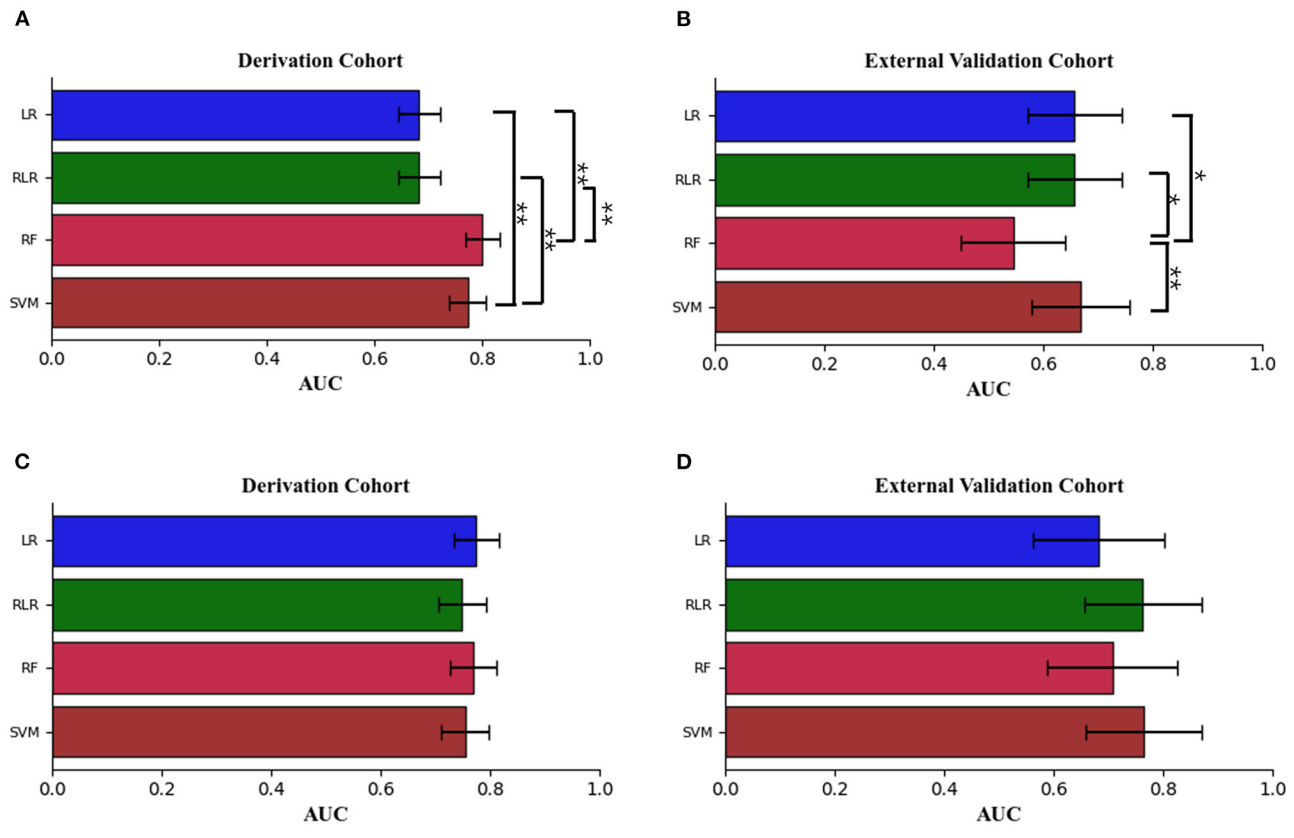


FIGURE 3 | The means ± 95% CI of the receiver operating characteristic area under the curve (AUC) for models 1 and 2 are displayed as bar graphs using the derivation cohort data **(A,C)**, and the validation cohort data **(B,D)**. For the derivation cohort data, there were significant differences between random forest [RF], support vector machine [SVM] and logistic regression without regularization [LR], regularized logistic regression [RLR] in model 1. For the external validation cohort data, there were significant differences between the random forest [RF] and the other three machine learning methods in model 1. For the derivation and external validation cohort data, the Delong test with Bonferroni correction was used. LR indicates logistic regression without regularization; RF, random forest; RLR, regularized logistic regression; and SVM, support vector machine. $^{*}P < 0.01$, $^{**}P < 0.001$.

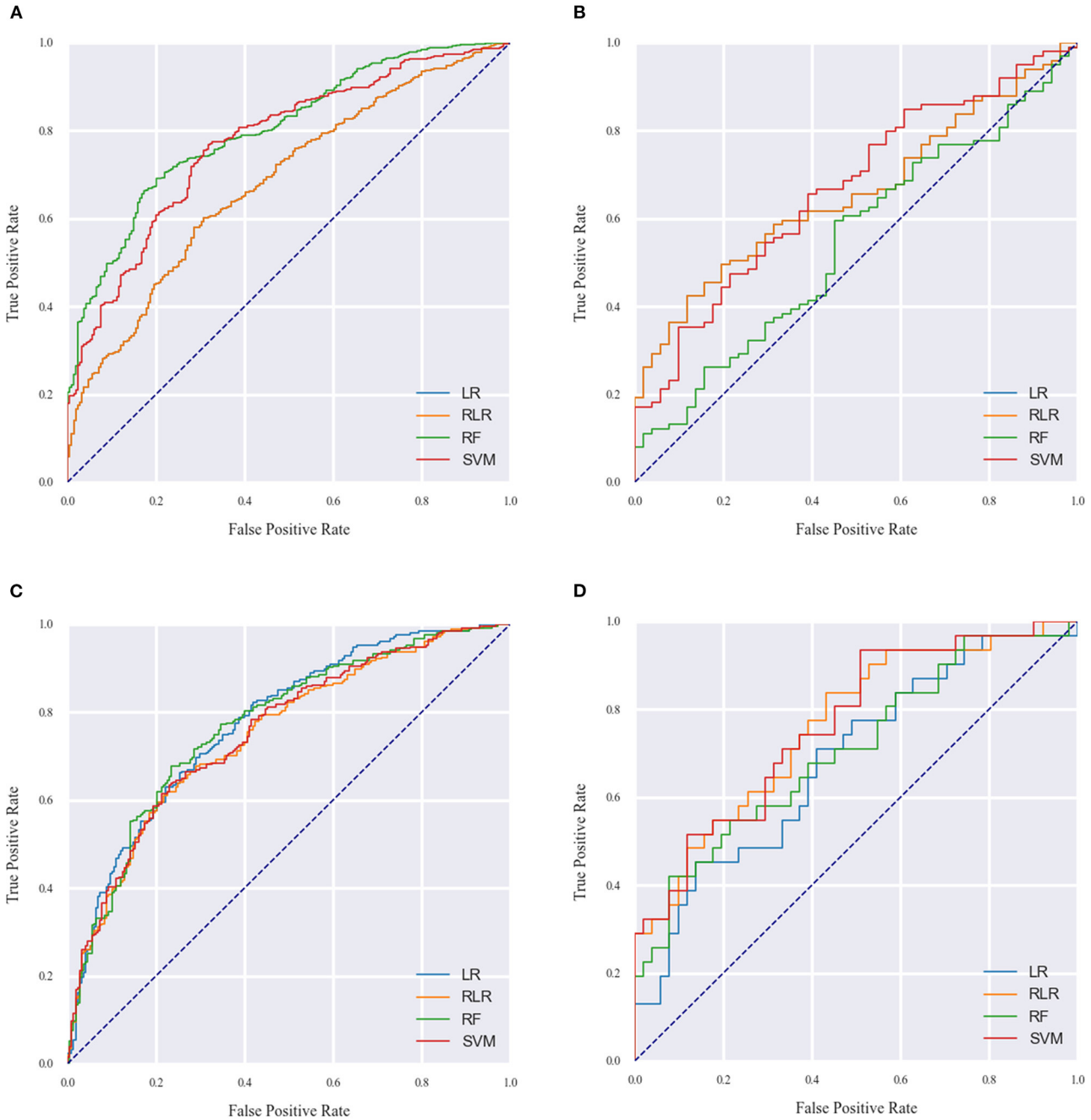


FIGURE 4 | The AUC of the machine learning models for model 1 **(A,B)** and model 2 **(C,D)** on the derivation and external validation cohort data. LR indicates logistic regression without regularization; RF, random forest; RLR, regularized logistic regression, SVM, support vector machine.

model 1 and model 2. As important variables of model 1, homocysteine, BMI, and occluded vessels (unilateral MCA) appeared in all three algorithms, and atrial fibrillation and coronary atherosclerotic heart disease were detected in both LR and RLR. Elevated blood homocysteine concentration increases the risk of ischemic stroke by inducing oxidative damage to vascular endothelial cells and enhancing platelet adhesion to endothelial cells, especially in large vessel strokes (18–22). The results of our study are in accordance with the aforementioned studies, suggesting that elevated homocysteine levels may be a significant marker for predicting ischemic stroke in AC-LVO.

TABLE 5 | Scores for each algorithm of model 1 in external validation cohort.

Model	AUC (95% CI)	PRC	Sensitivity	Specificity	Precision	NPV	Accuracy	$F1_{max}$
LR	0.66 (0.57–0.74)	0.82	42.4	88.2	87.5	44.1	58.0	0.80
RLR	0.66 (0.57–0.74)	0.82	42.4	88.2	87.5	44.1	58.0	0.80
RF	0.55 (0.45–0.64)	0.72	59.6	54.9	72.0	41.2	58.0	0.80
SVM	0.67 (0.58–0.76)	0.81	65.7	60.8	76.5	47.7	64.0	0.80

AUC, receiver operator characteristic area under the curve; $F1_{max}$, the maximum F1 score; LR, logistic regression without regularization; PRC, area under the precision-recall curve; RF, random forest; RLR, regularized logistic regression; SVM, support vector machine; NPV, negative predictive value.

TABLE 6 | Scores for each algorithm of model 2 in external validation cohort.

Model	AUC (95% CI)	PRC	Sensitivity	Specificity	Precision	NPV	Accuracy	$F1_{max}$
LR	0.68 (0.56–0.80)	0.60	45.2	86.3	66.7	72.1	70.7	0.60
RLR	0.76 (0.66–0.87)	0.71	83.9	56.9	54.2	85.3	67.1	0.66
RF	0.71 (0.59–0.83)	0.65	41.9	92.2	76.5	72.3	73.2	0.61
SVM	0.77 (0.66–0.87)	0.71	93.5	49.0	52.7	92.6	65.9	0.67

AUC, receiver operator characteristic area under the curve; $F1_{max}$, the maximum F1 score; LR, logistic regression without regularization; PRC, area under the precision-recall curve; RF, random forest; RLR, regularized logistic regression; SVM, support vector machine; NPV, negative predictive value.

Regarding the association between BMI and ischemic stroke, a previous meta-analysis revealed a J-shaped dose-response relationship between being overweight or obese and an increased risk of incident ischemic stroke (23). However, few studies have focused on the relationship between BMI and risks of ischemic stroke subtype (24, 25). Our study showed a robust positive association between overweight/obesity and AC-LVO AIS. Possible explanations for our findings include insulin resistance, endothelial dysfunction, and inflammation, which have been considered to influence the relationship between obesity and atherosclerosis (26). Moreover, our findings further revealed that a high BMI ($\geq$ 24 kg/m^2) shows a greater predisposes to disabling than non-disabling ischemic stroke with AC-LVO, emphasizing the importance of weight control and aerobic fitness.

The compensation of the collateral pathway in MCA occlusion mainly depends on the pia meningeal branch from the anterior cerebral artery and the posterior cerebral artery with worse compensatory ability than the circle of Willis, which means it would result in hemodynamic failure and is more prone to decompensation (27). Our study delves deeper into this field and demonstrates that unilateral MCA occlusion plays a crucial role in the occurrence of ischemic stroke. Furthermore, we found that stroke severity at admission was greater in the multiple AC-LAO patients than in unilateral MCA occlusion or unilateral ICA occlusion patients. This is consistent with a previously published study of patients with AC-LVO AIS, which showed that high NIHSS was associated with multiple AC-LAO (28).

Cardioembolism might be responsible for large vessel occlusion, in which atrial fibrillation accounts for ∼50% (29, 30). Atrial fibrillation is strongly associated with a high occurrence rate of LVO, suggesting that it may be a potential risk factor for LVO (31). Otherwise, large emboli that block intracranial vessels usually originate from the left atrial appendage in patients with symptomatic carotid stenosis or atrial fibrillation (32). Similarly, in our analysis, atrial fibrillation showed a robust association with AC-LVO AIS, further suggesting that knowledge of the potential complications of atrial fibrillation is likely to motivate both patient and clinician to comply with standard treatment.

Large-artery atherosclerotic stroke is associated with a high risk of coronary atherosclerotic heart disease (33). Nevertheless, our results indicate that coronary atherosclerotic heart disease is associated with a low risk of AIS in AC-LVO patients. One explanation for this finding might be that coronary atherosclerosis is significantly correlated with stenosis of the extracranial carotid; therefore, the development of intracranial anterior circulation large vessel occlusion may be independent of coronary atherosclerotic heart disease (34). Furthermore, antiplatelet and statin therapy in coronary atherosclerotic heart disease may reduce the risk of ischemic stroke in AC-LVO.

Apolipoprotein B is the primary apolipoprotein component of chylomicrons and low-density lipoproteins (35). In this study, we found that elevated serum levels of Apo B were associated with an increased risk of ischemic stroke in AC-LVO. Additionally, a Mendelian randomization study reported a positive correlation of Apo B with large artery stroke and small vessel stroke (36). Therefore, we advocate Apo B as a marker of routine serum lipid examination.

In our study, age emerged as an important predictor in both models, as well as in a previously developed model for predicting the clinical outcome of AIS with LVO (17). In general, our results indicate that the prevalence of ischemic stroke and disability increases with age in patients with AC-LVO. In addition, our data also suggested that ASPECT was the common element included in all machine learning methods. Studies have demonstrated that diffusion-weighted imaging (DWI) ASPECTS which represents infarct volume,

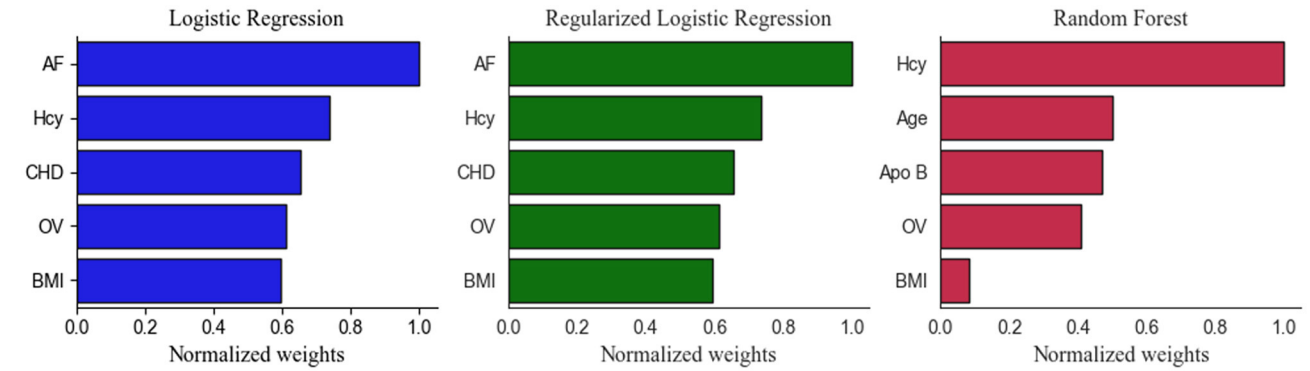


FIGURE 5 | Top five Important Features in the Model 1. Apo B indicates apolipoprotein B; AF, atrial fibrillation; BMI, body mass index; CHD, coronary atherosclerotic heart disease; Hcy, homocysteine; LR, logistic regression without regularization; OV, occluded vessels; RF, random forest; RLR, regularized logistic regression.

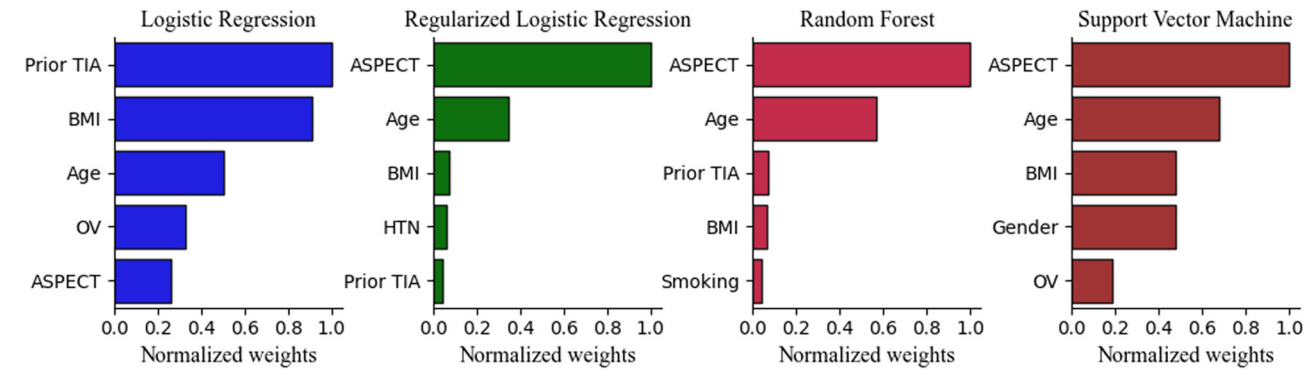


FIGURE 6 | Top 5 Important Features in the Model 2. ASPECT indicates Alberta Stroke Program Early CT Score; BMI, body mass index; HTN, hypertension; LR, logistic regression without regularization; OV, occluded vessels; RF, random forest; RLR, regularized logistic regression; SVM, support vector machine; TIA, transient ischemic attack.

is a significant independent predictor of functional outcome in AC-LVO strokes (37). Correspondingly, patients presenting with ASPECTS $\geq$7 are correlated with favorable outcomes following intravascular or thrombolytic therapy (38, 39). Our study further supports the association between ASPECT and the severity of neurological defects in first-ever ischemic stroke with AC-LVO. Consequently, a lower score of ASPECTS suggests less preserved brain parenchyma and predicts severe neurological impairment in patients with first-ever AC-LVO ischemic strokes.

It is well established that TIA increases the risk of ischemic stroke. In the present study, we found that prior TIA decreased ischemic stroke severity at admission, which is similar to the results of Marc Gotkine et al. showing that previous TIA was independently associated with lower severity of the ischemic stroke and a better short-term outcome (40). Prior TIA may have a neuroprotective effect on the subsequent ischemic stroke.

The chief strength of this study is the development and external validation of a new scoring tool, which predicts the risk of ischemic stroke and the severity of ischemia in AC-LVO based on machine learning approaches. Nevertheless, this study has several limitations. Foremost, few neuroimaging features were taken into consideration, excluding others such as the collateral flow status which might improve the predictive performance of the models. Further evaluation of the level of collateral circulation is necessary. Second, the sample size for this study was small which might have been due to the stringent inclusion criteria for patients with AC-LVO. As a result, the performance advantages of machine learning models may not have been fully realized. Finally, this was a retrospective study; the performance of the model needs to be tested in a prospective population in future studies.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Research Ethics Committee of the Second Hospital of Hebei Medical University on June 29, 2021 (approval No. 2021-R435). Written informed consent for participation was not

required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

JC and JY drafted and revised the manuscript, participated in the study conception and design, performed the statistical analyses, and analyzed and interpreted the data. XLiu participated in the conception and design of the study, data interpretation, and made a major contribution to manuscript revision. JY assisted in designing the machine learning model and in data analysis. KZ, GX, RZ, XLi, LL, YZ, LZ, PY, LX, TL, JT, PZ, SY, QW, and LG participated in the design of the study and contributed to manuscript revision. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by a grant from XLiu from the National Natural Science Foundation of China (81571160).

ACKNOWLEDGMENTS

We wish to acknowledge the contributions of patients and their family members who provide clinical information.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fneur.2021.749599/full#supplementary-material

REFERENCES

1. Rennert RC, Wali AR, Steinberg JA, Santiago-Dieppa DR, Olson SE, Pannell JS et al. Epidemiology, natural history, and clinical presentation of large vessel ischemic stroke. *Neurosurgery*. (2019) 85:S4–S8. doi: 10.1093/neuros/nyz042
2. Smith WS, Lev MH, English JD, Camargo EC, Chou M, Johnston SC et.al. Significance of large vessel intracranial occlusion causing acute ischemic stroke and TIA. *Stroke*. (2009) 40:3834–40. doi: 10.1161/STROKEAHA.109.561787
3. Goyal M, Menon BK, van Zwam WH, Dippel DW, Mitchell PJ, Demchuk AM et al. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *Lancet*. (2016) 387:1723–31. doi: 10.1016/S0140-6736[16]00163-X
4. Goyal M, Demchuk AM, Menon BK, Eesa M, Rempel JL, Thornton J et al. Randomized assessment of rapid endovascular treatment of ischemic stroke. *N Engl J Med*. (2015) 372:1019–30. doi: 10.1056/NEJMoa1414905
5. Hastrup S, Damgaard D, Johnsen SP, Andersen G. Prehospital acute stroke severity scale to predict large artery occlusion: design and comparison with other scales. *Stroke*. (2016) 47:1772–6. doi: 10.1161/STROKEAHA.115.012482
6. Katz BS, McMullan JT, Sucharew H, Adeoye O, Broderick JP. Design and validation of a prehospital scale to predict stroke severity: Cincinnati Prehospital Stroke Severity Scale. *Stroke*. (2015) 46:1508–12. doi: 10.1161/STROKEAHA.115.008804
7. Teleb MS, Ver Hage A, Carter J, Jayaraman MV, McTaggart RA. Stroke vision, aphasia, neglect (VAN) assessment-a novel emergent large vessel occlusion screening tool: pilot study and comparison with current clinical severity indices. *J Neurointerv Surg*. (2017) 9:122–6. doi: 10.1136/neurintsurg-2015-012131
8. Pérez de. la Ossa N, Carrera D, Gorchs M, Querol M, Millán M, Gomis M et al. Design and validation of a prehospital stroke scale to predict large arterial occlusion: the rapid arterial occlusion evaluation scale. *Stroke*. (2014) 45:87–91. doi: 10.1161/STROKEAHA.113.003071
9. Lima FO, Silva GS, Furie KL, Frankel MR, Lev MH, Camargo ÉC et al. Field assessment stroke triage for emergency destination: a simple and accurate prehospital scale to detect large vessel occlusion strokes. *Stroke*. (2016) 47:1997–2002. doi: 10.1161/STROKEAHA.116.013301
10. Hendrix P, Sofoluke N, Adams MD, Kunaprayoon S, Z and R, Kolinovsky AN et al. Risk factors for acute ischemic stroke caused by anterior large vessel occlusion. *Stroke*. (2019) 50:1074–80. doi: 10.1161/STROKEAHA.118.023917
11. Zhao W, Meng R, Ma C, Hou B, Jiao L, Zhu F et al. Safety and efficacy of remote ischemic preconditioning in patients with severe carotid artery stenosis before carotid artery stenting: a proof-of-concept, randomized controlled trial. *Circulation*. (2017) 135:1325–35. doi: 10.1161/CIRCULATIONAHA.116.024807
12. Miyazawa N, Hashizume K, Uchida M, Nukui H. Long-term follow-up of asymptomatic patients with major artery occlusion: rate of symptomatic change and evaluation of cerebral hemodynamics. *AJNR Am J Neuroradiol*. (2001) 22:243–7.
13. Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K et al. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: a guideline for healthcare professionals from the American heart association/American stroke association. *Stroke*. (2019) 50:e344–418. doi: 10.1161/STR.0000000000000211
14. Podgorelec V, Kokol P, Stiglic B, Rozman I. Decision trees: an overview and their use in medicine. *J Med Syst*. (2002) 26:445–63. doi: 10.1023/A:1016409317640
15. Loosli G, Canu S, Ong CS. Learning SVM in Krein Spaces. *IEEE Trans Pattern Anal Mach Intell*. (2016) 38:1204–16. doi: 10.1109/TPAMI.2015.2477830
16. Nishi H, Oishi N, Ishii A, Ono I, Ogura T, Sunohara T et al. Deep learning-derived high-level neuroimaging features predict clinical outcomes for large vessel occlusion. *Stroke*. (2020) 51:1484–92. doi: 10.1161/STROKEAHA.119.028101
17. Nishi H, Oishi N, Ishii A, Ono I, Ogura T, Sunohara T et al. Predicting clinical outcomes of large vessel occlusion before mechanical thrombectomy using machine learning. *Stroke*. (2019) 50:2379–88. doi: 10.1161/STROKEAHA.119.025411
18. Stamler JS, Osborne JA, Jaraki O, Rabbani LE, Mullins M, Singel D et al. Adverse vascular effects of homocysteine are modulated by endothelium-derived relaxing factor and related oxides of nitrogen. *J Clin Invest*. (1993) 91:308–18. doi: 10.1172/JCI116187
19. Okamura T, Kitamura A, Moriyama Y, Imano H, Sato S, Terao A et al. Plasma level of homocysteine is correlated to extracranial carotid-artery atherosclerosis in non-hypertensive Japanese. *J Cardiovasc Risk*. (1999) 6:371–7. doi: 10.1177/204748739900600603
20. Dardik R, Varon D, Tamarin I, Zivelin A, Salomon O, Shenkman B et al. Homocysteine and oxidized low density lipoprotein enhanced platelet adhesion to endothelial cells under flow conditions: distinct mechanisms of thrombogenic modulation. *Thromb Haemost*. (2000) 83:338–44. doi: 10.1055/s-0037-1613809
21. Tsai JC, Perrella MA, Yoshizumi M, Hsieh CM, Haber E, Schlegel R et al. Promotion of vascular smooth muscle cell growth by homocysteine: a link to atherosclerosis. *Proc Natl Acad Sci U S A*. (1994) 91:6369–73. doi: 10.1073/pnas.91.14.6369
22. Shi Z, Guan Y, Huo YR, Liu S, Zhang M, Lu H et al. Elevated total homocysteine levels in acute ischemic stroke are associated with long-term mortality. *Stroke*. (2015) 46:2419–25. doi: 10.1161/STROKEAHA.115.009136

23. Liu X, Zhang D, Liu Y, Sun X, Hou Y, Wang B. et al. A J-shaped relation of BMI and stroke: Systematic review and dose-response meta-analysis of 443 million participants. *Nutr Metab Cardiovasc Dis.* (2018) 28:1092–9. doi: 10.1016/j.numecd.2018.07.004
24. Li Y, Yatsuya H, Iso H, Yamagishi K, Saito I, Kokubo Y et al. Body mass index and risks of incident ischemic stroke subtypes: the japan public health center-based prospective (JPHC) study. *J Epidemiol.* (2019) 29:325–33. doi: 10.2188/jea.JE20170298
25. Yatsuya H, Yamagishi K, North KE, Brancati FL, Stevens J, Folsom AR, et al. Study investigators. Associations of obesity measures with subtypes of ischemic stroke in the ARIC Study. *J Epidemiol.* (2010) 20:347–54. doi: 10.2188/jea.JE20090186
26. Lovren F, Teoh H, Verma S. Obesity and atherosclerosis: mechanistic insights. *Can J Cardiol.* (2015) 31:177–83. doi: 10.1016/j.cjca.2014.11.031
27. Ogasawara K, Ogawa A, Yoshimoto T. Cerebrovascular reactivity to acetazolamide and outcome in patients with symptomatic internal carotid or middle cerebral artery occlusion: a xenon-133 single-photon emission computed tomography study. *Stroke.* (2002) 33:1857–62. doi: 10.1161/01.STR.0000019511.81583.A8
28. Zhang K, Li T, Tian J, Li P, Fu B, Yang X et al. Subtypes of anterior circulation large artery occlusions with acute brain ischemic stroke. *Sci Rep.* (2020) 10:3442. doi: 10.1038/s41598-020-60399-3
29. Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Borden WB et al. Heart disease and stroke statistics—2013 update: a report from the American Heart Association. *Circulation*. (2013) 127:e6–e245.
30. Freeman WD, Aguilar MI. Stroke prevention in atrial fibrillation and other major cardiac sources of embolism. *Neurol Clin.* (2008) 26:1129–60. doi: 10.1016/j.ncl.2008.07.001
31. Inoue M, Noda R, Yamaguchi S, Tamai Y, Miyahara M, Yanagisawa S et al. Specific factors to predict large-vessel occlusion in acute stroke patients. *J Stroke Cerebrovasc Dis.* (2018) 27:886–91. doi: 10.1016/j.jstrokecerebrovasdis.2017.10.021
32. Pagola J, Juega J, Francisco-Pascual J, Bustamante A, Penalba A, Pala E et al. Large vessel occlusion is independently associated with atrial fibrillation detection. *Eur J Neurol.* (2020) 27:1618–24. doi: 10.1111/ene.14281
33. Adams RJ, Chimowitz MI, Alpert JS, Awad IA, Cerqueria MD, Fayad P et al. American Heart Association/American Stroke Association. Coronary risk evaluation in patients with transient ischemic attack and ischemic stroke: a scientific statement for healthcare professionals from the Stroke Council and the Council on Clinical Cardiology of the American Heart Association/American Stroke Association. *Stroke*. (2003) 34:2310–22. doi: 10.1161/01.STR.0000090125.28466.E2
34. Conforto AB, Leite Cda C, Nomura CH, Bor-Seng-Shu E, Santos RD. Is there a consistent association between coronary heart disease and ischemic stroke caused by intracranial atherosclerosis? *Arq Neuropsiquiatr.* (2013) 71:320–6. doi: 10.1590/0004-282X20130028
35. Au A, Griffiths LR, Irene L, Kooi CW, Wei LK. The impact of APOA5, APOB, APOC3 and ABCA1 gene polymorphisms on ischemic stroke: Evidence from a meta-analysis. *Atherosclerosis.* (2017) 265:60–70. doi: 10.1016/j.atherosclerosis.2017.08.003
36. Yuan S, Tang B, Zheng J, Larsson SC. Circulating Lipoprotein Lipids, Apolipoproteins and Ischemic Stroke. *Ann Neurol.* (2020) 88:1229–36. doi: 10.1002/ana.25916
37. Rangaraju S, Streib C, Aghaebrahim A, Jadhav A, Frankel M, Jovin TG. Relationship Between Lesion Topology and Clinical Outcome in Anterior Circulation Large Vessel Occlusions. *Stroke.* (2015) 46:1787–92. doi: 10.1161/STROKEAHA.115.009908
38. Demchuk AM, Hill MD, Barber PA, Silver B, Patel SC, Levine SR. NINDS rtPA Stroke Study Group, NIH. Importance of early ischemic computed tomography changes using ASPECTS in NINDS rtPA Stroke Study. *Stroke.* (2005) 36:2110–5. doi: 10.1161/01.STR.0000181116.15426.58
39. Hill MD, Rowley HA, Adler F, Eliasziw M, Furlan A, Higashida RT et al. Selection of acute ischemic stroke patients for intra-arterial thrombolysis with pro-urokinase by using ASPECTS. *Stroke.* (2003) 34:1925–31. doi: 10.1161/01.STR.0000082483.37127.D0
40. Wang WW, Chen DZ, Zhao M, Yang XF, Gong DR. Prior transient ischemic attacks may have a neuroprotective effect in patients with ischemic stroke. *Arch Med Sci.* (2017) 13:1057–61. doi: 10.5114/aoms.2016.63744

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neurology

MINI REVIEW
published: 06 December 2021
doi: 10.3389/fneur.2021.734345

Machine Learning in Action: Stroke Diagnosis and Outcome Prediction

Shraddha Mainali[1]*, *Marin E. Darsie*[2,3†] *and Keaton S. Smetana*[4†]

[1] Department of Neurology, Virginia Commonwealth University, Richmond, VA, United States, [2] Department of Emergency Medicine, University of Wisconsin Hospitals and Clinics, Madison, WI, United States, [3] Department of Neurological Surgery, University of Wisconsin Hospitals and Clinics, Madison, WI, United States, [4] Department of Pharmacy, The Ohio State University Wexner Medical Center, Columbus, OH, United States

The application of machine learning has rapidly evolved in medicine over the past decade. In stroke, commercially available machine learning algorithms have already been incorporated into clinical application for rapid diagnosis. The creation and advancement of deep learning techniques have greatly improved clinical utilization of machine learning tools and new algorithms continue to emerge with improved accuracy in stroke diagnosis and outcome prediction. Although imaging-based feature recognition and segmentation have significantly facilitated rapid stroke diagnosis and triaging, stroke prognostication is dependent on a multitude of patient specific as well as clinical factors and hence accurate outcome prediction remains challenging. Despite its vital role in stroke diagnosis and prognostication, it is important to recognize that machine learning output is only as good as the input data and the appropriateness of algorithm applied to any specific data set. Additionally, many studies on machine learning tend to be limited by small sample size and hence concerted efforts to collate data could improve evaluation of future machine learning tools in stroke. In the present state, machine learning technology serves as a helpful and efficient tool for rapid clinical decision making while oversight from clinical experts is still required to address specific aspects not accounted for in an automated algorithm. This article provides an overview of machine learning technology and a tabulated review of pertinent machine learning studies related to stroke diagnosis and outcome prediction.

Keywords: machine learning, artificial intelligence, deep learning, stroke diagnosis, stroke prognosis, stroke outcome prediction, machine learning in medical imaging, machine learning in medicine

OPEN ACCESS

Edited by:
Jiang Li,
Geisinger Medical Center,
United States

Reviewed by:
Feifei Ma,
Vall d'Hebron Research Institute (VHIR), Spain
Harshawardhan Deshpande,
National Institute on Drug Abuse (NIDA), United States

***Correspondence:**
Shraddha Mainali
shraddha.mainali@vcuhealth.org

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Stroke, a section of the journal Frontiers in Neurology

Received: *01 July 2021*
Accepted: *28 October 2021*
Published: *06 December 2021*

Citation:
Mainali S, Darsie ME and Smetana KS (2021) Machine Learning in Action: Stroke Diagnosis and Outcome Prediction. Front. Neurol. 12:734345. doi: 10.3389/fneur.2021.734345

INTRODUCTION

The term machine learning (ML) was coined by Arthur Samuel in 1959 (1). He investigated two machine learning procedures using the game of checkers and concluded that computers can be programmed quickly to play a better game of checkers than the person who wrote the program. Simply put, machine learning can be defined as a subfield of artificial intelligence (AI) that uses computerized algorithms to automatically improve performance through iterative learning process or experience (i.e., data acquisition) (2). Of late, the field of ML has vastly evolved with the development of various computerized algorithms for pattern recognition and data assimilation to improve predictions, decisions, perceptions, and actions across various fields and serves as an extension to the traditional statistical approaches. In our day-to-day life, a relatable example of

ML is the application of spam filters to the 319 billion emails sent and received daily worldwide, of which, nearly 50% can be classified as spam (3). Use of ML technology has made this process efficient and manageable. The ML technology utilizes various methods for automated data analysis including linear and logistic regression models as well as other methods such as the support vector machines (SVM), random forests (RF), classification trees and discriminant analysis that allow combination of features (data points) in a non-linear manner with flexible decision boundaries. The advent of neural networks and deep learning (DL) technology has transformed the field of ML with automatic and efficient feature identification and processing within a covert analytic network, without the need for a priori feature selection. Notably, performance of DL is known to improve with access to larger datasets, whereas classic ML methods tend to plateau at relatively lower performance levels. Hence, in this era of big data where clinicians are constantly inundated with plethora of clinical information, use of DL technology has significnalty enhanced our ability to assimilate the vast amount of clinical data to make expeditious clinical decision.

Stroke is a leading cause of death, disability, and cognitive impairment in the United States (4). According to the 2013 policy statement from the American Heart Association, an estimated 4% of US adults will suffer from a stroke by 2030, accounting for total annual stroke-related medical cost of $240.67 billion by 2030 (5). For ischemic stroke, acute management is highly dependent on prompt diagnosis. According to the current ischemic stroke guidelines, patients are eligible for intravenous thrombolysis up to 4.5 h from symptom onset and endovascular thrombectomy without advanced imaging within 6 h of symptom onset (6–8). For patients presenting between 6 and 24 h of symptom onset (or last known well time), advanced imaging is recommended to assess salvageable penumbra for decisions regarding endovascular therapy (9–11). Similarly for hemorrhagic stroke, timely diagnosis utilizing imaging technology to evaluate the type and etiology of hemorrhage is important in guiding acute treatment decisions. Prompt diagnosis with emergent treatment decision and accurate prognostication is hence the cornerstone of acute stroke management. Over the recent years, a multitude of ML methodologies have been applied to stroke for various purposes, including diagnosis of stroke (12, 13), prediction of stroke symptom onset (14, 15), assessment of stroke severity (16, 17), characterization of clot composition (18), analysis of cerebral edema (19), prediction of hematoma expansion (20), and outcome prediction (21–23). In particular, there has been a rapid increase in the trend of ML application for *imaging-based* stroke diagnosis and outcome prediction. The Ischemic Stroke Lesion Segmentation Challenge (ISLES: http://www.isles-challenge.org/) provides a global competing platform encouraging teams across the world to develop advanced tools for stroke lesion analysis using ML. In this platform, competitors train their algorithms on a standardized dataset and eventually generate benchmarks for algorithm performance.

Deciding which type of ML to use on a specific dataset depends on factors such as the size of dataset, need for supervision, ability to learn, and the generalizability of the model (24). DL technology such as the deep neural networks has significantly improved the ability for image segmentation, automated featurization (e.g., conversion of raw signal into clinically useful parameter), and multimodal prognostication in stroke; and it is increasingly utilized in stroke-based applications (25–27). For example, DL algorithms can be applied to extract meaningful imaging features for image processing in an increasing order of hierarchical complexity to make predictions, such as the final infarct volume (27). Some commonly used ML types with their respective algorithms and practical examples are outlined in **Figures 1–3**. In the healthcare setting, supervised and unsupervised algorithms are both commonly used. In this review, we will specifically focus on ML strategies for stroke diagnosis and outcome prediction. **Table 1** provides an overview of pertinent studies with use of ML in stroke diagnosis (Section A) and outcome prediction (Section B). A glossary of machine learning terms with brief description is separately provided in **Supplementary Table 1**.

METHODS

We searched PubMed, Google Scholar, Web of Science, and IEEE Xplore® for relevant articles using various combination of the following key words: "machine learning," "artificial intelligence," "stroke," "ischemic stroke," "hemorrhagic stroke," "diagnosis," "prognosis," "outcome," "big data," and "outcome prediction." Resulting abstracts were screened by all authors and articles were hand-picked for full review based on relevance and scientific integrity. Final article list was reviewed and approved by all authors.

Machine Learning in Stroke Diagnosis

The time-sensitive nature of stroke care underpins the need for accurate and rapid tools to assist in stroke diagnosis. Over the recent years, the science of brain imaging has vastly advanced with the availability of a myriad of AI based diagnostic imaging algorithms (77). Machine learning is particularly useful in diagnosis of acute stroke with large vessel occlusion (LVO). Various automated methods for detection of stroke core and penumbra size as well as mismatch quantification and detection of vascular thrombi have recently been developed (77). Over the past decade, 13 different companies have developed automated and semi-automated commercially available software for acute stroke diagnostics (Aidoc®, Apollo Medical Imaging Technology®, Brainomix®, inferVISION®, RAPID®, JLK Inspection®, Max-Q AI®, Nico.lab®, Olea Medical®, Qure.ai®, Viz.ai®, and Zebra Medical Vision®) (78). The RapidAI® and Viz.ai® technology have been approved under the medical device category of computer-assisted triage by the United States Food and Drug Administration (FDA). The RAPID MRI® (Rapid processing of Perfusion and Diffusion) software allows for an unsupervised, fully-automated processing of perfusion and diffusion data to identify those who may benefit from thrombectomy based on the mismatch ratio (79). Such commercial platforms available for automatic detection of ischemic stroke and LVO have facilitated rapid treatment decisions. When compared to manual

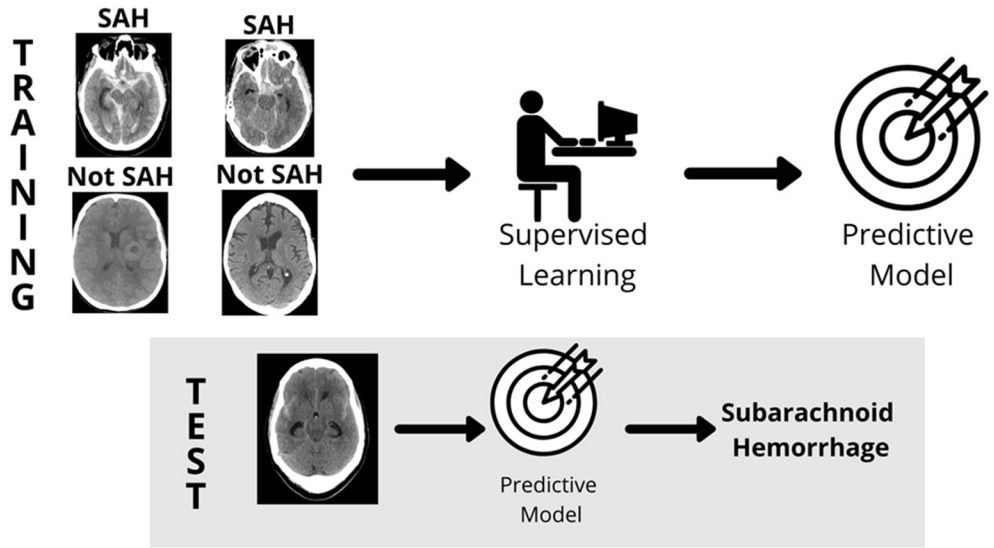


FIGURE 1 | Supervised learning. In supervised learning, a model is built by labeling images [Subarachnoid Hemorrhage (SAH) and Not Subarachnoid Hemorrhage (Not SAH)], a predictive model is created, and then tested for accuracy in reading unlabeled images (gray box). Source: WesternDigital BLOG.

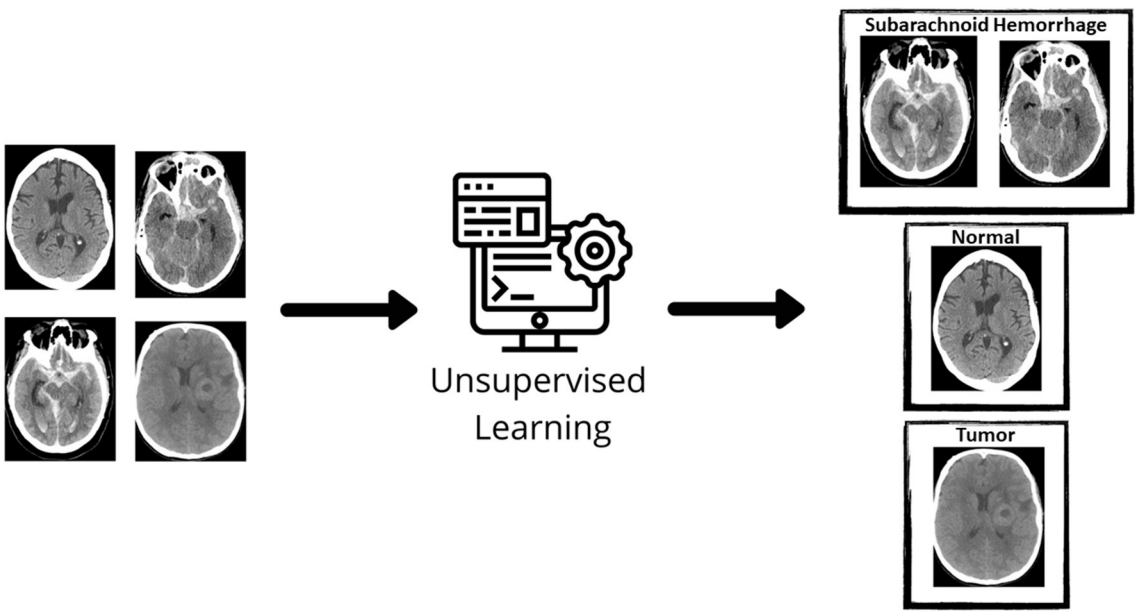


FIGURE 2 | Unsupervised learning. In unsupervised learning, the machine learning algorithm discovers structures within given data. The initial data is not labeled and a clustering algorithm groups unlabeled data together. Source: WesternDigital BLOG.

segmentation of lesion volume and mismatch identification from patients enrolled in DEFUSE 2, the RAPID results were found to be well-correlated (r^2 = 0.99 and 0.96 for diffusion and perfusion weighted imaging, respectively) with 100% sensitivity and 91% specificity for mismatch identification (80). Since 2008, the RapidAI® platform has expanded to include other products (Rapid® ICH, ASPECTS, CTA, LVO, CTP, MRI, Angio, and Aneurysm) that assist across the entire spectrum of stroke. Viz LVO® was the first FDA-cleared software to detect and alert clinicians of LVO *via* the "Viz Platform" (81). In a recent single center study with 1,167 CTAs analyzed, Viz LVO® was found to have a sensitivity of 0.81 and a negative predictive value of 0.99 with an accuracy of 0.94 (82).

Other areas of stroke diagnostics that have seen an increase in attention over the past decade are the identification of intracerebral hemorrhage (ICH) and patients at risk for delayed cerebral

FIGURE 3 | Created from the following referenes: Dey (28) Zhou (29) Geron (30).

ischemia in the setting of aneurysmal subarachnoid hemorrhage (aSAH). While most studies tend to have good accuracy in detecting an ICH there is more variability in subclassification and measurements of hematoma volume. A summary of recent publications on ML in stroke diagnosis is presented in **Table 1** (Section A).

Machine Learning in Stroke Outcome Prediction

Despite recent advances in stroke care, it remains the second leading cause of death and disability world-wide (4, 83). Although acute stroke diagnosis and determination of the time of stroke onset are the initial steps of comprehensive stroke management, clinicians are also often charged with the task of determining stroke outcomes. These outcomes range from discrete radiological outcomes (e.g., final infarct volume, the likelihood of hemorrhagic transformation, etc.), the likelihood of morbidity (e.g., stroke-associated pneumonia) and mortality, and various measures of functional independence (e.g., mRS score, Barthel Index score, cognitive, and language function, etc.).

Prognostication after an acute brain injury is notoriously challenging, particularly within the first 24–48 h (84). However, a clinician may be called upon to provide estimates of a patient's short-term and long-term mortality and degree of functional dependence to assist with decision-making regarding the intensity of care (e.g., use of thrombolytics or endovascular treatment, intubation, code status, etc.) (60, 64, 66, 67, 69, 70, 72–76). Like all medical emergencies, it is incumbent upon the stroke clinician to ensure that all care provided is concordant with an individual patient's goals (85). For example, a surrogate decision-maker may decline to reverse a patient's longstanding "do not intubate" order to facilitate mechanical thrombectomy if the clinician predicts the patient has a high likelihood of functional dependence or short-term mortality. Hence, accuracy in outcome prediction is critical in guiding management of our patients.

Determining a patient's likelihood of developing symptomatic intracranial hemorrhage (sICH) is of obvious, immediate value in acute stroke management in determining candidacy for thrombolytic therapy or endovascular treatment. Historically, clinician-based prognostication tools to predict the risk of symptomatic intracranial hemorrhage after IV thrombolysis, such as the SEDAN (Sugar, Early Infarct signs, Dense cerebral artery sign, Age, and NIHSS) and HAT (Hemorrhage After Thrombolysis) scores have been used to predict the risk of symptomatic intracranial hemorrhage after IV thrombolysis (23). Advances in ML and DL have allowed for the development of more accurate models which outperform the traditional SEDAN and HAT scores (23, 54, 55). Similarly, the ability to predict final infarct volume and the likelihood of the development of malignant cerebral edema have important treatment implications and remain a significant focus of ML in stroke (26, 51–53).

In patients with intracerebral hemorrhage (ICH), the ICH-score is one of the most widely used clinical prediction scores (85–88). Although ML technology for outcome prediction has rapidly advanced for ischemic stroke, recent ML studies predicting functional outcomes after ICH have also demonstrated high-discriminating power (63, 89). A recent study by Sennfält et al. tracked long-term functional dependence and mortality after an acute ischemic stroke of more than 20,000 Swedish patients (90). The 30-day mortality rate was 11.1%. At 5 years,

TABLE 1 | Studies utilizing machine learning for stroke diagnosis and prediction.

References	Study objective	ML-based approaches	Validation method	Sample size	Feature	Optimal results	Optimal ML approach	Clinical implications	Limitations
Section A: stroke diagnosis									
Ischemic stroke									
Garca-Terriza et al. (31)	Stroke type diagnosis and mortality	RF	10-fold cross validation resampling	• 119 • (AIS 105, ICH 14)	• Type of stroke • Mortality • Non-invasive variables (cardiac and pulmonary)	• *Accuracy* • Subtype - 92% • Mortality - 96%	-	May predict the type of stroke a patient is at risk for and outcomes	Data obtained after event to for prediction models but do not include usual risk factors for consideration
Sung et al. (32)	Ischemic stroke phenotype*	Various models (C4.5, CART, KNN, RF, SVM, LR, with aggregation algorithms	10-fold cross validation	4,640	Clinical notes with preprocessing and MetaMap to identify medical entities +/- NIHSS	• *Accuracy; kappa* • NIHSS + text • (0.489–0.583; 0.272–0.399) • NIHSS • (0.465–0.533; 0.254–0.344) • Text • (0.465–0.533; 0.170–0.328)	-	Clinical text plus validated scoring tools might aid in phenotyping of stroke	• Phenotype based on OCSP definitions,* • Difficult delineating certain phenotypes, • Unclear who were the authors of the clinic notes
Giri et al. (33)	Ischemic stroke diagnosis by EEG	1D CNN vs. various models (NB, Classification Tree, ANN, RF, kNN, LR)	Leave-one-out cross-validation	• 32 – AIS • 30 – Controls	15-min EEG with 24 chosen features	• Accuracy - 0.86 • F-Score 0.861	Leave-one-out scenario of 1D CNN	In areas with limited access to CT imaging may help diagnosis AIS	Time to apply EEG electrodes may result in delays of care
Lee et al. (14)	Identify patients within 4.5-h thrombolysis window	LR, RF, SVM	• 85% training • 15% test	355	MRI features	• Sensitivity 75.8% • Specificity 82.6% • AUC 85.1%	RF	Improved sensitivity than human readings in identifying stroke patients within thrombolysis window	Assessed only dichotomized visibility of signals in the lesion territory
Ho et al. (15)	Classifying onset time from imaging	LR, RF, GBRT, SVM, SMR	10-fold cross validation on training data with optimal hyperparameters	104	MRI	• Sensitivity 78.8% • AUC 76.5%	LR with deep autoencoder features	Improved stroke onset detection compared to DWI-FLAIR	Trained on MRI only
Takahashi et al. (34)	Detection for MCA dot sign in unenhanced CT	SVM	Not described	297 images	Unenhanced CT	Sensitivity 97.5%	SVM	Accurately detect hyperdense MCA dot sign	Data from 7 patients
Chen et al. (35)	Automatically segment stroke lesions in DWI	CNN	Train / Test	741 subjects	DWI	Dice score 0.67	CNN	Segment stroke lesions automatically	Improved Dice scores on larger lesions
Bouts et al. (36)	Depict ischemic tissue that can recover after reperfusion	GLM, GAM, SVM, Adaptive boosting, RF	Generalized cross validation with unbiased risk estimator scoring	19 rats	MRI	Dice Score 0.79	GLM	MRI-based algorithms could estimate extent of salvageable tissue	Varying efficacy in differentiating between areas irreversibly damaged vs. salvaged after reperfusion

(Continued)

TABLE 1 | Continued

References	Study objective	ML-based approaches	Validation method	Sample size	Feature	Optimal results	Optimal ML approach	Clinical implications	Limitations
Chen et al. (37)	Quantify cerebral edema following infarction *via* CSF quantification	RF with geodesic active contour segmentation	• 10-fold cross validation • Train / Test	38 subjects	CT Imaging	• Baseline Dice Score 0.76 • 6-h Dice score 0.73	RF with geodesic active contour segmentation	Efficiently and accurately measure evolution of cerebral edema	
Colak et al. (38)	Stroke Prediction	MLP ANN and SVM with radial basis function kernel	Train / Test	297 subjects (130 sick and 167 healthy)	9 predictors (CAD, DM, HTN, CVA history, AF, smoking, carotid Doppler findings, cholesterol, CRP	• Accuracy 85.9% • AUC 0.93	ANN	Ability to screen patients at risk for stroke based on comorbidities	Factors used to predict model are known to be risk factors for stroke
Maier et al. (39)	Classify lesion segmentation	KNN, GNB, GLM, RF, CNN	Leave-one-out cross-validation	37 subjects	MRI	• RF: • Precision 82% • Recall 62% • CNN: • Precision 77% • Recall 64%	• RF • CNN	Future work may be able to segment lesions	No methods achieved results in the range of the human observer agreement
Öman et al. (40)	Detection of ischemic stroke	3D CNN	Train / Test	60 subjects	CT Angiography	• Sensitivity 93% • Specificity 82% • AUC 0.93 • Dice 0.61	3D CNN	Lesion can be detected with CNN	Contralateral hemisphere data may reduce false positive findings
Chen et al. (41)	Prehospital detection of large vessel occlusion	ANN	10-fold cross validation	600 subjects	Baseline demographics, medical history, NIHSS, risk factors	• Youden index 0.640 • Sensitivity 0.807 • Specificity 0.833 • Accuracy 0.822	ANN	Known patient risk factors may help in predicting large vessel occlusion	Cohort included stroke patients and not those with mimics or hemorrhagic stroke
Hemorrhagic stroke									
Dhar et al. (42)	Hemorrhage and perihematomal edema (PHE) quantification	CNN	• 10-fold cross validation • Train / Test	124	24-h CT head scans	• Dice score • 0.9 – hemorrhage • 0.54 - PHE	-	Rapid and consistent measurements of supratentorial ICH	-IVH not delineated from ICH
Arab et al. (43)	Hematoma segmentation and volume quantification	CNN with deep supervision based on reader labeling	Train / Test	55	64 axial slices of 128 × 128 voxels	• Dice score • 0.84 ± 0.06 • Precision • 0.85 ± 0.07 • Recall • 0.83 ± 0.07 • F-Score 0.84	CNN with deep supervision	Fast and reliable quantification of hematoma volume	• False positives observed with calcifications • False negatives observed with blood close to bone
Ko et al. (44)	ICH detection	CNN and long-short term memory	Train / Test	5,244,234	Pre-processed CTH to balance subtypes and window settings	• Classification accuracy • 92 – 93%	-	Identification of ICH and subtypes	-Preprocessing of data required to attain accuracy
Irene et al. (45)	ICH segmentation and volume approximation	Dynamic Graph CNN	• 4-fold cross validation • Train / Test	27	CTH	• Accuracy 96.4% • Precision 0.93 • Recall 0.98 • *F*-Score 0.96	SVM method with radial basis function kernel	Identification of ICH and blood volume prediction	Small dataset

(Continued)

TABLE 1 | Continued

References	Study objective	ML-based approaches	Validation method	Sample size	Feature	Optimal results	Optimal ML approach	Clinical implications	Limitations
Arbabshinrani et al. (13)	Diagnose ICH and prioritize radiology worklists	Deep CNN	• Training (75%) • Cross validation (5%) • Testing (20%)	46,573 studies	Preprocessing of CTH images	• ROC 0.846 • Specificity 0.8 • Sensitivity 0.73	-	Assist in upgrading image reads to "stat" from "routine"	Did not identify location of ICH
Sage et al. (46)	ICH subtype detection	Double-branch CNN of SVM, RF	Concatenation of double-branch features and classification	9,997 subjects	372,556 images (11,454 CT scans)	• Accuracy range • SVM • 76.9 – 96% • RF • 74.3 – 96.7%	-	Identify and classify ICH	EDH performed the worst in SVM and RF possibly due to under representation in data
Ye et al. (47)	ICH subtype detection	3D joint CNN – recurrent NN	• Training (80%) • Validation (10%) • Testing (10%)	2,836 subjects	76,621 slices from non-contrast head CT scans	• AUC for +/- ICH • 0.98 • AUC range for subtypes • 0.89 – 0.96	-	Identify and classify ICH	SAH classification may have been more difficult due to blended ICH examples
Chang et al. (48)	ICH detection and volume measurements	Hybrid 3D/2D CNN	5-fold cross validation	10,841 Scans	Non-contrast CTH	• *ICH detection* • Accuracy 0.97 • Sensitivity 0.951 • Specificity 0.073 • Volume • Dice score 0.772–0.931	-	Identification of ICH and blood volume prediction	Generalization needs to be confirmed in other institutions
Subarachnoid hemorrhage									
Capoglu et al. (49)	Vasospasm prediction	Sparse dictionary learning and covariance-based features	Not described	20	3D brain angiograms	ROC 0.93	-	Proof of concept to predict those who might have vasospasm	Small dataset
Ramos et al. (22)	DCI Prediction	LogReg, SVM, RF, MLP	Monte-Carlo cross-validation with 100 random splits (75% training / 25% test) and 5-fold cross-validation	317	Non-contrast CT image data and 48 clinical variables	• ROC 0.74 • Specificity 0.67 • Sensitivity 0.75	RF with clinical variables and image features	ML improved prediction of DCI especially when image features included (aneurysm height / width)	Manual extraction of features from medical images is time-consuming
Tanioka et al. (50)	DCI prediction	RF	Leave-one-out cross-validation	95	Clinical variables and matricellular proteins (MCP) on days 1 – 3	• Accuracy • 93.9% - clinical variables • 87.2% - MCP only • 95.2% - clinical variables + MCP	-	MCP might play a role in predicting DCI but further data needed	Other biomarkers not assessed

(Continued)

TABLE 1 | Continued

References	Study objective	ML-based approaches	Validation method	Sample size	Feature	Optimal results	Optimal ML approach	Clinical implications	Limitations
Miscellaneous									
Ni et al. (12)	Stroke Case Detection	LR, SVM-P, SVM-R, RF, ANN	Two iterations of 10-fold cross validation	8,131	Medical record information compared to ICD codes	• Accuracy 88.6% • Precision 93.8% • Recall 92.8% • *F* Score 93.3% • AUC 89.8% • AUC-PR 97.5%	RF	Detection of stroke diagnosis through EHR data that was miscoded	Accurate ICD codes limit utility of the algorithm
Park et al. (16)	Autonomously grade NIHSS and MRC scores through wearable sensors	• SVM • Ensemble	5-fold cross validation searched by Bayes optimization in 30 trials	240	Wearable sensors	• NIHSS: • Accuracy 83.3% • AUC 0.912 • MRC: • Accuracy 76.7% AUC 0.87	SVM	Automatic grading in real time of proximal weakness	Requires sensors to be applied

Section B: stroke outcome prediction

References	Study objective	ML-based approach	Validation method	Sample size	Feature	Optimal results	Best predictors	Clinical implications	Limitations
Radiological outcomes									
Nielsen et al. (26)	Prediction of final infarct volume	CNN_{deep}	85% training/15% testing	222	MRI images	AUC 0.88 ± 0.12	-	Facilitates treatment selection	No external validation, retrospective
Giacalone et al. (51)	Prediction of final infarct volume	SVM	K-fold cross-validation	4	MRI images	95% accuracy	-	" "	Small sample size, Retrospective
Grosser et al. (52)	Prediction of final infarct volume	XGBoost	Leave-one-out cross-validation	99	MRI images	AUC 0.893 ± 0.085	Spatial lesion probability	" "	Retrospective, Limited generalizability (patient data is from 2006 to 2009)
Foroushani et al. (53)	Prediction of malignant cerebral edema	LR	10-fold cross-validation	361	Serial, quantitative CT images	AUC 0.96	Reduction in CSF volume	" "	No external validation
Bentley et al. (23)	Prediction of sICH	SVM	K-fold cross-validation	116	Unenhanced CT images	AUC 0.744	Baseline NIHSS, CT evidence of acute ischemia	" "	Image processing took ∼30 min; Small number of sICH cases
Yu et al. (54)	Prediction of HT	SR-KDA	Leave-one-out cross-validation	155	MRI images	83.7 ± 2.6% accuracy	-	" "	Single-center, Retrospective
Scalzo et al. (55)	Prediction of HT	SR-KDA	10-fold cross-validation	263	MRI images	88% accuracy	-	" "	Retrospective, current limitations in measuring BBB permeability
van Os et al. (56)	Prediction of reperfusion after EVT (mTICI <2b vs. ≥2b)	LR (using backward elimination)	Nested cross-validation, consisting of an outer and an inner cross-validation loop	1,383	EHR data, CT/CTA images	AUC 0.57	-	" "	Retrospective; Only moderate predictive value, LR outperformed machine-learning

(Continued)

TABLE 1 | Continued

References	Study objective	ML-based approaches	Validation method	Sample size	Feature	Optimal results	Optimal ML approach	Clinical implications	Limitations
Hilbert et al. (57)	Prediction of reperfusion after EVT (mTICI <2b vs. ≥2b)	RFNN-ResNet-AE fine-tuned	4-fold cross-validation	1301	CTA images	Average AUC 0.65	-	" "	Retrospective; Only moderate predictive value
Rondina et al. (58)	Comparison of imaging approaches (lesion load per ROI vs. pattern of voxel) to predict post stroke motor impairment	GPR	10-fold cross-validation	50	Post stroke MRI	Best prediction was obtained using motor ROI and CST (derived from probabilistic tractography) R = 0.83, RMSE = 0.68	Patterns of voxels representing lesion probability produced better results	Informs appropriate methodology for predicting long term motor outcomes from early post-stroke MRI.	Small sample size, no external validation
Discrete morbidity and mortality clinical outcomes									
Matsumoto et al. (59)	Prediction of all-cause, in-hospital mortality	LASSO	10-fold cross-validation	4,232	EHR data	AUC 0.88	-	Facilitates GOC decision making	Retrospective, Single-center, Limited generalizability (ETV used in only 1.5% of patients), Low rate (3.5%) of in-hospital mortality
Scrutinio et al. (60)	Prediction of 3-yr mortality after severe stroke	SMOTE RF	10-fold cross-validation	1,207	EHR data	AUC 0.928	Age	Facilitates GOC decision making	No external validation
Ge et al. (61)	Prediction of SAP at 7 and 14 d	Attention-augmented GRU	10-fold cross-validation	13,930	EHR data	• 7 d: AUC 0.928 • 14 d: AUC 0.905	PPI use	Facilitates early detection and targeted application of prophylaxis interventions	Single-center, No external validation
Li et al. (62)	Prediction of SAP at 7 d	XGBoost	5-fold cross-validation	3,160	EHR data	AUC 0.841	Age, Baseline NIHSS, FBG, sex, Premorbid mRS score, & History of AF	" "	Single-center, No external validation
Wang et al. (63)	Predicting functional outcome (mRS) at 1st and 6th months	RF	10-fold cross-validation	333	Demographics, labs, CT brain	• 1 month outcome: AUC 0.899; • 6 months outcome AUC: 0.917	• 1 month outcome= 26 attributes; • 6 months outcome: 22 attributes	Use of ML to predict functional outcome after ICH is feasible, and RF model provides the best predictive performance	Small sample size, excluded large hematomas, did not evaluate hematoma or edema expansion, no external validation
Functional outcomes									
Heo et al. (64)	Prediction of mRS score (0–2 vs. 3–6) at 90 d	Deep neural network	67% training/ 33% testing	2,604	EHR data	AUC 0.888	-	Informs patient expectations, Facilitates GOC decision making	Single-center, No external validation

(Continued)

TABLE 1 | Continued

References	Study objective	ML-based approaches	Validation method	Sample size	Feature	Optimal results	Optimal ML approach	Clinical implications	Limitations
Lin et al. (65)	Prediction of mRS score (0–2 vs. 3–6) at 90 d	SVM	10-fold cross-validation	35,798	Registry data	f1-score 87.9 ± 0.2% (92.9 ± 0.1%, with follow-up data)	mRS score at 30 d, toilet use degree of dependence	" "	More severe strokes accounted for most prediction errors
Brugnara et al. (66)	Prediction of mRS score (0–2 vs. 3–6) at 90 d	"Machine-learning models with gradient boosting classifiers"	Not specified	246	Clinical data, radiological data (CT, CTA, CTP, and angiographic images)	AUC 0.856	NIHSS score at 24 h, Premorbid mRS score, Final infarct volume on CT	" "	Single center, No external validation, Retrospective
Forkert et al. (67)	Prediction of mRS score at 90 d	SVM (specifically the Extended Problem-specific model)	Leave-one-out cross-validation	68	Clinical data, MRI images	• mRS score ± 1: 82.4% accuracy • mRS score 0–2 vs. 3–6: 85.4% accuracy	• L-hemisphere strokes: lesion-based *t*-score sum • Rt-hemisphere strokes: Lesion volume	" "	No external validation, Retrospective
Monteiro et al. (68)	Prediction of mRS score (0–2 vs. 3–6) at 90 d	RF	10-fold cross-validation	425	Clinical data, CT or MRI images	AUC 0.936 ± 0.34	Baseline NIHSS score, Baseline NIHSS score on subsection 2 (Best gaze, horizontal EOMs)	" "	Single center, No external validation, Retrospective, Performed worse than non-imaging model
Jang et al. (69)	Prediction of mRS score (>1 vs. >2) at 90 d	XGBoost	3-fold cross-validation and a random search strategy	6,731	Registry data	• mRS >1: AUC 0.84 • mRS >2: AUC 0.87		" "	Treatment-related factors were not included, No external validation
Hope et al. (70)	Prediction of speech production scores	GPR	Leave-one-out cross-validation	270	Clinical data, Assessments, MRI images	R^2 0.59	Time post-stroke, Lesion site	Informs patient expectations	Post-stroke imaging obtained over a wide range of times (<1 month to +30 y), No external validation, Retrospective
Lopes et al. (71)	Prediction of cognitive functions at 3 y after minor stroke	Ridge Regression	3-step nested leave-one-out cross-validation, consisting of inner, middle, and outer loops	72	Clinical data, Assessments, functional MRI images	R^2 values for attention, memory, visuospatial functions, and language functions: 0.73, 0.67, 0.55, 0.48	-	" "	Limited generalizability (mean NIHSS on admission was 1.5 ± 2.2), Retrospective
Sale et al. (72)	Prediction of change in BI score and FIM score during inpatient rehab	SVM	Nested 5-fold cross-validation	55	Clinical biomarker data, Assessments	Discharge cognitive FIM score: MADP 17.55%, RMSE 4.28	Cognitive FIM score upon admission	Informs patient expectations, Facilitates GOC decision making	Small sample size, included hemorrhagic stroke patients
Iwamoto et al. (73)	Prediction of ADL dependence after inpatient rehab	CART method	Not specified	994	Clinical data, Assessments	AUC 0.83	FIM transfer score (≤4 or >4)	" "	Single center, Retrospective
Lin et al. (74)	Prediction of BI score (<60, 60–90, >90) upon discharge from inpatient rehab	LR, RF	5-fold cross-validation	313	Clinical data, Assessments	LR: AUC 0.796, RF: AUC 0.792	BI, IADL, and BBT scores on admission	" "	Limited generalizability due to aggressive rehab strategy, No external validation

(Continued)

TABLE 1 | Continued

References	Study objective	ML-based approaches	Validation method	Sample size	Feature	Optimal results	Optimal ML approach	Clinical implications	Limitations
Tozlu et al. (75)	Prediction of post-intervention UE motor impairment in chronic stroke	Elastic net	Nested 10-fold cross-validation with outer and inner loops	102	Clinical data, Assessments	Median R^2 0.91	Pre-intervention UE-FMA, difference in MT between the affected and unaffected hemispheres	Informs patient expectations, Increases rehabilitation efficiency	Retrospective, No external validation
Stinear et al. (76)	Predicts potential for UE recovery	Cluster analyses	Not applicable	40	Clinical assessments ± neurophysiological assessments and MRI images	Partial η^2 0.811	-	" "	Small sample size, Single center, No external validation

Section A and B

ADL, Activities of daily living; AE, Auto-encoders; AF, Atrial fibrillation; AIS, Acute ischemic stroke; ANN, Artificial neural network; AUC, area under the receiver operating characteristic curve; BBB, blood-brain barrier; BBT, Berg balance test; BI, Barthel Index; CART, Classification and regression tree; CNN, convolutional neural network; CSF, cerebral spinal fluid; CST, Corticospinal tract; CT, computed tomography; CTA, Computerized tomography angiography; CTP, Computerized tomography perfusion; CXR, Chest radiograph; D, days; DCI, delayed cerebral ischemia; DTI, Diffusion Tensor Imaging; DWI, diffusion weighted image; EDH, epidural hematoma; EEG, electroencephalogram; EHR, electronic health record; EOMs, Extra-ocular movements; EVT, endovascular treatment; FBG, Fasting blood glucose; FIM, Functional independence measure; GAM, generalized additive model; GBRT, gradient boosted regression tree; GLM, generalized linear model; GOC, Goals-of-care; GRU, gated recurrent unit; GPR, Gaussian Process model Regression; H, hours; HT, hemorrhagic transformation; IADL, Instrumental activities of daily living scale; ICH, Intracerebral hemorrhage; IVH, intraventricular hemorrhage; KNN, K nearest neighbor; L, Left; LASSO, Least absolute shrinkage and selection operator regression; LR, logistic regression; MADP, Mean absolute percentage deviation; MCA, middle cerebral artery; MCP, matricellular proteins; Min, minutes; MLP, multilayer perceptron; MRC, medical research council; MRI, magnetic resonance imaging; mRS, Modified Rankin Score; MT, motor threshold; NB, naïve bayes; NIHSS, National Institutes of Health Stroke Scale; PHE, perihematomal edema; PPI, Proton pump inhibitor; RF, Random forest; RFNN, Structured Receptive Field Neural Networks; RMSE, Root mean square error; ROI, region of interest; Rt, Right; SAP, Stroke-associated pneumonia; sICH, symptomatic intracranial hemorrhage; SMOTE, synthetic minority oversampling technique; SMR, stepwise multilinear regression; SR-KDA, Kernel Spectral Regression for Discriminant Analysis; SVM, support vector model; SVM-P, support vector machine with polynomial; SVM-R, support vector machine with radial basis function; UE, Upper extremity; UE-FMA, Upper extremity Fugl-Meyer Assessment; XGBoost, Extreme gradient boosting; Yr, year.

NB: List of ML terms with definitions is provided in ***Supplementary Table 1****.*

Section B

Many of the listed studies utilize a variety of machine learning (ML)-based approaches. The approach listed on the table is the approach with the optimal result from each individual study.

** Phenotype based on Oxfordshire Community Stroke Project (OCSP) (total anterior circulation infarcts, lacunar infarcts, partial anterior circulation infarcts, posterior circulation infarcts).*

70.6% of ischemic stroke patients were functionally dependent (defined as mRS score of ≥3) or had died (5-year mortality rate of 50.6%). These sobering outcomes partially account for the development of many stroke prognostic models over the years, which frequently serve as benchmarks in stroke research. Recently, Matsumoto et al. compared the performance of six existing stroke prognostic models for predicting poor functional outcomes and in-hospital mortality with linear regression or decision tree ensemble models (59). The novel prediction models performed slightly better than the conventional models in predicting poor functional outcomes (AUC 0.88–0.94 vs. AUC 0.70–0.92) but were equivalent or marginally worse in predicting in-hospital death (AUC 0.84–0.88 vs. AUC 0.87–0.88). Many such stroke prediction models have emerged over the recent years. An overview of ML based automated algorithms for stroke outcome prediction is provided in **Table 1** (Section B).

DISCUSSION

In recent years, some DL algorithms have approached human levels of performance in object recognition (91). One of the greatest strengths of ML is its ability to endlessly process data and tirelessly perform an iterative task. Further, creation of a ML model can be performed much faster (i.e., in a matter of 5–6 days compared with 5–6 months or even years) than traditional computer-aided detection and diagnosis (CAD) (92). which makes ML an attractive field for computer experts and scientists. Several ML tools are currently in use including the FDA-approved ML algorithms previously discussed for rapid stroke diagnosis which have significantly enhanced the workflow of acute ischemic stroke patients.

Despite the prolific advent of new and improved ML algorithms with increasing clinical applications, it is important to recognize that computer-based algorithms are only as good as the data used to train the models. For a reliable algorithm, it is important to develop well-defined training, validation, and testing sets. Testing should be done on a diverse set of data points reflective of a real-world scenario. Overfitting can be an issue in ML algorithms when the model is trained on a group of highly-selected, specific features, which when tested on a larger dataset with varied features, fails to perform adequately. Similarly, underfitting can occur when a model is oversimplified with generalized feature selection in the training set which then becomes unable to capture the relevant features within a complex pattern of a larger or more diverse testing set. The aphorism "garbage in, garbage out" remains true as the use of inadequate or unvalidated data points (e.g., unverified clinical reports from electronic health record) in the training set can lead to poor performance of the ML algorithm in the testing set. Hence, it is important to note that the algorithmic decision-making tools do not guarantee accurate and unbiased interpretation compared to established logistic regression models (56, 59, 93). Comparisons to well-established models should be standard when developing new ML algorithms given the high cost associated with ML (e.g., the time required to collect data, train the model, perform internal and external validations, cost of reliable and secure data storage, etc.) (94). Specifically, as it relates to diagnostics there are a myriad of considerations that must be taken into account. Not only should the algorithm provide accurate information quickly, but it should have the ability to integrate into the electornic health record (EHR) to improve end user experience and efficiency in workflow. Programs such as RAPID®, Viz.ai®, and Brainomix® have started to successfully integrate into the EHR, which has helped expedite acute stroke diagnosis and triage process. One of the major technical challenges of ML include the ability to develop an algorithm with a "reasonable" detection rate of pathology without an excessive rate of false-positives. For example, there are notable discrepancies among various ML studies for ICH diagnosis, with varying accuracy depending on the type of ICH (e.g., spontaneous ICH, SDH, aSAH, or IVH). Overfitting and underfitting of the model could lead to poor applicability and therefore, image preprocessing with meticulous feature selection is necessary. Furthermore, the "black-box" nature of ML precludes the clinicians from identifying and addressing biases within the algorithms (95, 96). Hence, proper external validation is necessary to ensure generalizability of the algorithm in diverse clinical scenarios.

For stroke prediction, most existing ML algorithms utilize dichotomized outcomes. Functional outcome is frequently defined as "good" when mRS score is 0–2 and "poor" when mRS score is 3–6 by convention and IS studies often measure mRS score at 90 days after stroke (64–69, 97). However, the medical community is increasingly embracing patient-centered outcomes. People are starting to recognize the need for longitudinal patient follow-up given potential for functional improvement beyond conventional norms of 90 days (98). Once patient-centered outcomes are clinically validated (e.g., MRS cutoff of 0–2 vs. 3–6, 0–3 vs. 4–6, or 0–4 vs. 5–6), new ML algorithms incorporating such outcomes would be increasingly helpful to the clinicians. The use of high-yield, ML programs using patient-centered outcomes could ease the commonplace but challenging discussions of the anticipated quality of life and the risk of long-term dependency or death before deciding on a patient's goals-of-care. It is however important to apply caution while using ML algorithms for outcome prediction as patient demographics and clinical practice continue to evolve and updates to the ML algorithms would be necessary to remain applicable to evolving patient populations and clinical standards. Additionally, developers often retrieve data from existing datasets (e.g., clinical trial data) with its inherent biases including selection bias, observer bias and other confounders (e.g., withdrawal of life supporting therapy may be more common in older patients with large hemispheric stroke compared to younger patients, which could confound outcome prediction in older patients compared to younger ones).

Overall, compared to other diseases such as Alzheimer's disease, there is a relative paucity of large, high-quality datasets within stroke. Some limitations that have stymied the development of large, open-access stroke registries include the need for data-sharing agreements, patient privacy concerns, high costs of data storage and security, arbitration of quality control of the input data, etc. (95). Cohesive and collaborative efforts across hospital systems, regions, and nations with data acquisition and

harmonization is needed to improve future ML-based programs in stroke. With adoption of EHR systems, healthcare data is rapidly accumulating with an estimated over 35 zettabytes of existing healthcare data! (99). Adoption of AI and ML algorithms allow us to efficiently process the plethora of information that surround us every day. Nonetheless, as we continue to adapt to this evolving landscape of medical practice surrounding big data, clinicians need to remain aware of the limitations of this modern day "black box" magic.

CONCLUSION

The emerging ML technology has rapidly integrated into multiple fields of medicine including stroke. Deep learning has significantly enhanced practical applications of ML and some newer algorithms are known to have comparable accuracy to humans. However, the diagnosis and prognosis of a disease, including stroke, is highly intricate and depends on various clinical and personal factors. The development of optimal ML programs requires comprehensive data collection and assimilation to improve diagnostic and prognostic accuracy. Given the "black box" or cryptic nature of these algorithms, it is extremely important for the end-user (i.e., clinicians) to understand the intended use and limitations of any ML algorithm to avoid inaccurate data interpretation. Although ML algorithms have improved stroke systems of care, blind dependence on such computerized technology may lead to misdiagnosis or inaccurate prediction of prognostic trajectories. At the current state, ML tools are best used as "aids" for clinical decision making while still requiring oversight to address relevant clinical aspects that are overlooked by the algorithm.

AUTHOR CONTRIBUTIONS

SM: substantial contributions including conception and design of the work, literature review, interpretation and summarization of data, drafting the complete manuscript, revising it critically for important intellectual content, and final approval of the manuscript to be published. MD and KS: contribution including conception and design of the work, literature review, interpretation and summarization of the data, drafting of critical portion of the manuscript, critical revision for important intellectual content, and final approval of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This article was supported by the Virginia Commonwealth University, Department of Neurology.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fneur.2021.734345/full#supplementary-material

REFERENCES

1. Samuel L. Some studies in machine learning using the game of checkers. *IBM J Res Dev.* (1959) 3:210–29. doi: 10.1147/rd.33.0210
2. Mitchell TM. *Machine Learning.* New York: Mcgraw-hill (1997).
3. Tankovska H. *E-mail Usage in the United States—Statistics and Facts. In, Statista.* Hamburg: Statista (2021).
4. GBD 2016 Neurology Collaborators. Global, regional, and national burden of neurological disorders:1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* (2019) 18:459–80. doi: 10.1016/S1474-4422(18)30499-X
5. Ovbiagele B, Goldstein LB, Higashida RT, Howard VJ, Johnston SC, Khavjou OA, et al. Forecasting the future of stroke in the United States: a policy statement from the American Heart Association and American Stroke Association. *Stroke.* (2013) 44:2361–75. doi: 10.1161/STR.0b013e31829734f2
6. Hacke W, Kaste M, Bluhmki E, Brozman M, Dávalos A, Guidetti D, et al. Thrombolysis with alteplase 3 to 45 hours after acute ischemic stroke. *N Engl J Med.* (2008) 359:1317–29. doi: 10.1056/NEJMoa0804656
7. Bluhmki E, Chamorro A, Dávalos A, Machnig T, Sauce C, Wahlgren N, et al. Stroke treatment with alteplase given 30-45 h after onset of acute ischaemic stroke (ECASS III): additional outcomes and subgroup analysis of a randomised controlled trial. *Lancet Neurol.* (2009) 8:1095–102. doi: 10.1016/S1474-4422(09)70264-9
8. Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K, et al. 2018 Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke.* (2018) 49:e46–110. doi: 10.1161/STR.0000000000000158
9. Nogueira RG, Jadhav AP, Haussen DC, Bonafe A, Budzik RF, Bhuva P, et al. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. *N Engl J Med.* (2018) 378:11–21. doi: 10.1056/NEJMoa1706442
10. Campbell A, Bruce CV, Peter Mitchell J, Timothy Kleinig J, Helen Dewey M, Churilov L, et al. Endovascular therapy for ischemic stroke with perfusion-imaging selection. *N Engl J Med.* (2015) 372:1009–18. doi: 10.1056/NEJMoa1414792
11. Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K, et al. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke.* (2019) 50:e344–418. doi: 10.1161/STR.0000000000000211
12. Ni Y, Alwell K, Moomaw CJ, Woo D, Adeoye O, Flaherty ML, et al. Towards phenotyping stroke: leveraging data from a large-scale epidemiological study to detect stroke diagnosis. *PLoS ONE.* (2018) 13:e0192586. doi: 10.1371/journal.pone.0192586
13. Arbabshirani MR, Fornwalt BK, Mongelluzzo GJ, Suever JD, Geise BD, Patel AA, et al. Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. *NPJ Digit Med.* (2018) 1:9. doi: 10.1038/s41746-017-0015-z
14. Lee H, Lee EJ, Ham S, Lee HB, Lee JS, Kwon SU, et al. Machine learning approach to identify stroke within 45 hours. *Stroke.* (2020) 51:860–6. doi: 10.1161/STROKEAHA.119.027611
15. Ho KC, Speier W, Zhang H, Scalzo F, El-Saden S, Arnold CW, et al. Machine learning approach for classifying ischemic stroke onset time from imaging. *IEEE Trans Med Imaging.* (2019) 38:1666–76. doi: 10.1109/TMI.2019.2901445
16. Park E, Lee K, Han T, Nam HS. Automatic grading of stroke symptoms for rapid assessment using optimized machine learning and 4-limb kinematics: clinical validation study. *J Med Internet Res.* (2020) 22:e20641. doi: 10.2196/20641
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) 20:8. doi: 10.1186/s12911-019-1010-x
18. Chung JW, Kim YC, Cha J, Choi EH, Kim BM, Seo WK, et al. Characterization of clot composition in acute cerebral infarct using machine learning techniques. *Ann Clin Transl Neurol.* (2019) 6:739–47. doi: 10.1002/acn3.751
19. 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) 9:687. doi: 10.3389/fneur.2018.00687
20. Liu J, Xu H, Chen Q, Zhang T, Sheng W, Huang Q, et al. Prediction of hematoma expansion in spontaneous intracerebral hemorrhage using support vector machine. *EBioMedicine.* (2019) 43:454–9. doi: 10.1016/j.ebiom.2019.04.040
21. Dengler NF, Madai VI, Unteroberdörster M, Zihni E, Brune SC, Hilbert A, et al. Outcome prediction in aneurysmal subarachnoid hemorrhage: a comparison of machine learning methods and established clinico-radiological scores. *Neurosurg Rev.* (2021) 44:2837–46. doi: 10.1007/s10143-020-01453-6
22. Ramos LA, van der Steen WE, Sales Barros R, Majoie C, van den Berg R, Verbaan D, et al. Machine learning improves prediction of delayed cerebral ischemia in patients with subarachnoid hemorrhage. *J Neurointerv Surg.* (2019) 11:497–502. doi: 10.1136/neurintsurg-2018-014258
23. Bentley P, Ganesalingam J, Carlton Jones AL, Mahady K, Epton S, Rinne P, et al. Prediction of stroke thrombolysis outcome using CT brain machine learning. *Neuroimage Clin.* (2014) 4:635–40. doi: 10.1016/j.nicl.2014.02.003
24. Géron A. *Hands-On Machine Learning With Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. Sebastopol, CA: O'Reilly Media (2019).
25. Feng R, Badgeley M, Mocco J, Oermann EK. Deep learning guided stroke management: a review of clinical applications. *J Neurointerv Surg.* (2018) 10:358–62. doi: 10.1136/neurintsurg-2017-013355
26. Nielsen A, Hansen MB, Tietze A, Mouridsen K. Prediction of tissue outcome and assessment of treatment effect in acute ischemic stroke using deep learning. *Stroke.* (2018) 49:1394–401. doi: 10.1161/STROKEAHA.117.019740
27. Yu Y, Xie Y, Thamm T, Gong E, Ouyang J, Huang C, et al. Use of deep learning to predict final ischemic stroke lesions from initial magnetic resonance imaging. *J Am Med Assoc Netw Open.* (2020) 3:e200772. doi: 10.1001/jamanetworkopen.2020.0772
28. Dey A. Machine learning algorithms: a review. *Int J Comp Sci Inform Technol.* (2016) 7:1174–9. Available online at: https://ijcsit.com/docs/Volume%207/vol7issue3/ijcsit2016070332.pdf
29. Zhou ZH. Ensemble learning. *Encycl Biometr.* (2009) 2009:270–3. doi: 10.1007/978-0-387-73003-5_293
30. Géron A. *Hands-on Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. Sebastopol, CA: O'Reilly Media, Inc. (2017).
31. Garcia-Terriza L, Risco-Martin JL, Ayala JL, Rosello GR, Camarasaltas JM. IEEE: comparison of different machine learning approaches to model stroke subtype classification and risk prediction. In: *2019 Spring Simulation Conference*. Tuscan, AZ: Springsim. (2019). doi: 10.23919/SpringSim.2019.8732846
32. Sung SF, Lin CY, Hu YH. EMR-based phenotyping of ischemic stroke using supervised machine learning and text mining techniques. *IEEE J Biomed Health Informat.* (2020) 24:2922–31. doi: 10.1109/JBHI.2020.2976931
33. Giri EP, Fanany MI, Arymurthy AM, Wijaya SK. Ischemic stroke identification based on EEG and EOG using ID convolutional neural network and batch normalization. In: *2016 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2016*. Malang (2017). doi: 10.1109/ICACSIS.2016.7872780
34. Takahashi N, Lee Y, Tsai DY, Matsuyama E, Kinoshita T, Ishii K. An automated detection method for the MCA dot sign of acute stroke in unenhanced CT. *Radiol Phys Technol.* (2014) 7:79–88. doi: 10.1007/s12194-013-0234-1
35. Chen L, Bentley P, Rueckert D. Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks. *Neuroimage Clin.* (2017) 15:633–43. doi: 10.1016/j.nicl.2017.06.016
36. Bouts MJ, Tiebosch IA, van der Toorn A, Viergever MA, Wu O, Dijkhuizen RM. Early identification of potentially salvageable tissue with MRI-based predictive algorithms after experimental ischemic stroke. *J Cereb Blood Flow Metab.* (2013) 33:1075–82. doi: 10.1038/jcbfm.2013.51
37. Chen Y, Dhar R, Heitsch L, Ford A, Fernandez-Cadenas I, Carrera C, 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) 12:673–80. doi: 10.1016/j.nicl.2016.09.018
38. Colak A, Karaman E, Turtay MG. Application of knowledge discovery process on the prediction of stroke. *Comput Methods Programs Biomed.* (2015) 119:181–5. doi: 10.1016/j.cmpb.2015.03.002
39. Maier O, Schröder C, Forkert ND, Martinetz T, Handels H. Classifiers for ischemic stroke lesion segmentation: a comparison study. *PLoS ONE.* (2015) 10:e0145118. doi: 10.1371/journal.pone.0145118
40. Öman O, Mäkelä T, Salli E, Savolainen S, Kangasniemi M. 3D convolutional neural networks applied to CT angiography in the detection of acute ischemic stroke. *Eur Radiol Exp.* (2019) 3:8. doi: 10.1186/s41747-019-0085-6
41. Chen Z, Zhang R, Xu F, Gong X, Shi F, Zhang M, et al. Novel prehospital prediction model of large vessel occlusion using artificial neural network. *Front Aging Neurosci.* (2018) 10:181. doi: 10.3389/fnagi.2018.00222
42. Dhar R, Falcone GJ, Chen YS, Hamzehloo A, Kirsch EP, Noche RB, et al. Deep learning for automated measurement of hemorrhage and perihematomal edema in supratentorial intracerebral hemorrhage. *Stroke.* (2020) 51:648–51. doi: 10.1161/STROKEAHA.119.027657
43. Arab A, Chinda B, Medvedev G, Siu W, Guo H, Gu T, et al. A fast and fully-automated deep-learning approach for accurate hemorrhage segmentation and volume quantification in non-contrast whole-head CT. *Sci Rep.* (2020) 10:7. doi: 10.1038/s41598-020-76459-7
44. Ko A, Chung H, Lee H, Lee J. IEEE: feasible study on intracranial hemorrhage detection and classification using a CNN-LSTM network. In: *42nd Annual International Conference of the IEEE-Engineering-in-Medicine-and-Biology-Society (EMBC)*. Montreal (2020). doi: 10.1109/EMBC44109.2020.9176162
45. Irene K, Masum MA, Yunus RE, Jatmiko W. Segmentation and approximation of blood volume in intracranial hemorrhage patients based on computed tomography scan images using deep learning method. In: *2020 International Workshop on Big Data and Information Security, IWBIS 2020*. Depok (2020). doi: 10.1109/IWBIS50925.2020.9255593
46. Sage A, Badura P. Intracranial hemorrhage detection in head CT using double-branch convolutional neural network, support vector machine, and random forest. *Appl Sci.* (2020) 10:577. doi: 10.3390/app10217577
47. Ye A, Gao F, Yin YB, Guo DF, Zhao PF, Lu Y, et al. Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network. *Eur Radiol.* (2019) 29:6191–201. doi: 10.1007/s00330-019-06163-2
48. Chang PD, Kuoy E, Grinband J, Weinberg BD, Thompson M, Homo R, et al. Hybrid 3D/2D convolutional neural network for hemorrhage evaluation on head CT. *Am J Neuroradiol.* (2018) 39:1609–16. doi: 10.3174/ajnr.A5742
49. Capoglu S, Savarraj JP, Sheth SA, Choi HA, Giancardo L. Representation learning of 3D brain angiograms, an application for cerebral vasospasm prediction. In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. Berlin (2019). doi: 10.1109/EMBC.2019.8857815
50. Tanioka S, Ishida F, Nakano F, Kawakita F, Kanamaru H, Nakatsuka Y, et al. Machine learning analysis of matricellular proteins and clinical variables for early prediction of delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage. *Mol Neurobiol.* (2019) 56:7128–35. doi: 10.1007/s12035-019-1601-7
51. Giacalone M, Rasti P, Debs N, Frindel C, Cho TH, Grenier E, et al. Local spatio-temporal encoding of raw perfusion MRI for the prediction of final lesion in stroke. *Med Image Anal.* (2018) 50:117–26. doi: 10.1016/j.media.2018.08.008
52. Grosser M, Gellißen S, Borchert P, Sedlacik J, Nawabi J, Fiehler J, et al. Improved multi-parametric prediction of tissue outcome in acute ischemic stroke patients using spatial features. *PLoS ONE.* (2020) 15:e0228113. doi: 10.1371/journal.pone.0228113
53. Foroushani HM, Hamzehloo A, Kumar A, Chen Y, Heitsch L, Slowik A, et al. Quantitative serial CT imaging-derived features improve prediction of malignant cerebral edema after ischemic stroke. *Neurocrit Care.* (2020) 33:785–92. doi: 10.1007/s12028-020-01056-5

54. 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) 65:2058–65. doi: 10.1109/TBME.2017.2783241
55. Scalzo F, Alger JR, Hu X, Saver JL, Dani KA, Muir KW, et al. Multi-center prediction of hemorrhagic transformation in acute ischemic stroke using permeability imaging features. *Magn Reson Imaging.* (2013) 31:961–9. doi: 10.1016/j.mri.2013.03.013
56. van Os HJA, Ramos LA, Hilbert A, van Leeuwen M, van Walderveen MAA, Kruyt ND, et al. Predicting outcome of endovascular treatment for acute ischemic stroke: potential value of machine learning algorithms. *Front Neurol.* (2018) 9:784. doi: 10.3389/fneur.2018.00784
57. Hilbert A, Ramos LA, van Os HJA, Olabarriaga SD, Tolhuisen ML, Wermer MJH, et al. Data-efficient deep learning of radiological image data for outcome prediction after endovascular treatment of patients with acute ischemic stroke. *Comput Biol Med.* (2019) 115:103516. doi: 10.1016/j.compbiomed.2019.103516
58. ARondina M, Filippone M, Girolami M, Ward NS. Decoding post-stroke motor function from structural brain imaging. *Neuroimage Clin.* (2016) 12:372–80. doi: 10.1016/j.nicl.2016.07.014
59. Matsumoto K, Nohara Y, Soejima H, Yonehara T, Nakashima N, Kamouchi M. Stroke prognostic scores and data-driven prediction of clinical outcomes after acute ischemic stroke. *Stroke.* (2020) 51:1477–83. doi: 10.1161/STROKEAHA.119.027300
60. Scrutinio D, Ricciardi C, Donisi L, Losavio E, Battista P, Guida P, et al. Machine learning to predict mortality after rehabilitation among patients with severe stroke. *Sci Rep.* (2020) 10:20127. doi: 10.1038/s41598-020-77243-3
61. Ge Y, Wang Q, Wang L, Wu H, Peng C, Wang J, et al. Predicting post-stroke pneumonia using deep neural network approaches. *Int J Med Inform.* (2019) 132:103986. doi: 10.1016/j.ijmedinf.2019.103986
62. Li X, Wu M, Sun C, Zhao Z, Wang F, Zheng X, et al. Using machine learning to predict stroke-associated pneumonia in Chinese acute ischaemic stroke patients. *Eur J Neurol.* (2020) 27:1656–63. doi: 10.1111/ene.14295
63. Wang HL, Hsu WY, Lee MH, Weng HH, Chang SW, Yang JT, et al. Automatic machine-learning-based outcome prediction in patients with primary intracerebral hemorrhage. *Front Neurol.* (2019) 10:910. doi: 10.3389/fneur.2019.00910
64. Heo J, Yoon JG, Park H, Kim YD, Nam HS, Heo JH. Machine learning-based model for prediction of outcomes in acute stroke. *Stroke.* (2019) 50:1263–5. doi: 10.1161/STROKEAHA.118.024293
65. Lin J, Jiang A, Ling M, Mo Y, Li M, Zhao J. Prediction of neurologic deterioration based on support vector machine algorithms and serum osmolarity equations. *Brain Behav.* (2018) 8:e01023. doi: 10.1002/brb3.1023
66. Brugnara G, Neuberger U, Mahmutoglu MA, Foltyn M, Herweh C, Nagel S, et al. Multimodal predictive modeling of endovascular treatment outcome for acute ischemic stroke using machine-learning. *Stroke.* (2020) 51:3541–51. doi: 10.1161/STROKEAHA.120.030287
67. 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:e0129569. doi: 10.1371/journal.pone.0129569
68. Monteiro M, Fonseca AC, Freitas AT, Pinho EMT, Francisco AP, Ferro JM, et al. Using machine learning to improve the prediction of functional outcome in ischemic stroke patients. *IEEE/ACM Trans Comput Biol Bioinform.* (2018) 15:1953–9. doi: 10.1109/TCBB.2018.2811471
69. Jang SK, Chang JY, Lee JS, Lee EJ, Kim YH, Han JH, et al. Reliability and clinical utility of machine learning to predict stroke prognosis: comparison with logistic regression. *J Stroke.* (2020) 22:403–6. doi: 10.5853/jos.2020.02537
70. Hope TM, Seghier ML, Leff AP, Price CJ. Predicting outcome and recovery after stroke with lesions extracted from MRI images. *Neuroimage Clin.* (2013) 2:424–33. doi: 10.1016/j.nicl.2013.03.005
71. Lopes R, Bournonville C, Kuchcinski G, Dondaine T, Mendyk AM, Viard R, et al. Prediction of long-term cognitive functions after minor stroke, using functional connectivity. *Neurology.* (2021) 96:e1167–79. doi: 10.1212/WNL.0000000000011452
72. Sale P, Ferriero G, Ciabattoni L, Cortese AM, Ferracuti F, Romeo L, et al. Predicting motor and cognitive improvement through machine learning algorithm in human subject that underwent a rehabilitation treatment in the early stage of stroke. *J Stroke Cerebrovasc Dis.* (2018) 27:2962–72. doi: 10.1016/j.jstrokecerebrovasdis.2018.06.021
73. Iwamoto Y, Imura T, Tanaka R, Imada N, Inagawa T, Araki H, et al. Development and validation of machine learning-based prediction for dependence in the activities of daily living after stroke inpatient rehabilitation: a decision-tree analysis. *J Stroke Cerebrovasc Dis.* (2020) 29:105332. doi: 10.1016/j.jstrokecerebrovasdis.2020.105332
74. Lin WY, Chen CH, Tseng YJ, Tsai YT, Chang CY, Wang HY, et al. Predicting post-stroke activities of daily living through a machine learning-based approach on initiating rehabilitation. *Int J Med Inform.* (2018) 111:159–64. doi: 10.1016/j.ijmedinf.2018.01.002
75. Tozlu A, Edwards D, Boes A, Labar D, Tsagaris KZ, Silverstein J, et al. Machine learning methods predict individual upper-limb motor impairment following therapy in chronic stroke. *Neurorehabil Neural Repair.* (2020) 34:428–39. doi: 10.1177/1545968320909796
76. Stinear CM, Barber PA, Petoe M, Anwar S, Byblow WD. The PREP algorithm predicts potential for upper limb recovery after stroke. *Brain.* (2012) 135:2527–35. doi: 10.1093/brain/aws146
77. Murray NM, Unberath M, Hager GD, Hui FK. Artificial intelligence to diagnose ischemic stroke and identify large vessel occlusions: a systematic review. *J Neurointerv Surg.* (2020) 12:156–64. doi: 10.1136/neurintsurg-2019-015135
78. Mokli Y, Pfaff J, Dos Santos DP, Herweh C, Nagel S. Computer-aided imaging analysis in acute ischemic stroke - background and clinical applications. *Neurol Res Pract.* (2019) 1:23. doi: 10.1186/s42466-019-0028-y
79. Straka MG, Albers W, Bammer R. Real-time diffusion-perfusion mismatch analysis in acute stroke *J Magn Reson Imaging.* (2010) 32:1024–37. doi: 10.1002/jmri.22338
80. Lansberg MG, Straka M, Kemp S, Mlynash M, Wechsler LR, Jovin TG, et al. MRI profile and response to endovascular reperfusion after stroke (DEFUSE 2): a prospective cohort study. *Lancet Neurol.* (2012) 11:860–7. doi: 10.1016/S1474-4422(12)70203-X
81. Justin P. FDA approves stroke-detecting AI software. *Nat Biotechnol.* (2018) 36:290. doi: 10.1038/nbt0418-290
82. Yahav-Dovrat A, Saban M, Merhav G, Lankri I, Abergel E, Eran A, et al. Evaluation of artificial intelligence-powered identification of large-vessel occlusions in a comprehensive stroke center. *Am J Neuroradiol.* (2021) 42:247–54. doi: 10.3174/ajnr.A6923
83. GBD 2016 Causes of Death Collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death:1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet.* (2017) 390:1151–210. doi: 10.1016/S0140-6736(17)32152-9
84. Stevens RD, Sutter R. Prognosis in severe brain injury. *Crit Care Med.* (2013) 41:1104–23. doi: 10.1097/CCM.0b013e318287ee79
85. Holloway RG, Arnold RM, Creutzfeldt CJ, Lewis EF, Lutz BJ, McCann RM, et al. Palliative and end-of-life care in stroke: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke.* (2014) 45:1887–916. doi: 10.1161/STR.0000000000000015
86. Morgenstern LB, Zahuranec DB, Sánchez BN, Becker KJ, Geraghty M, Hughes R, et al. Full medical support for intracerebral hemorrhage. *Neurology.* (2015) 84:1739–44. doi: 10.1212/WNL.0000000000001525
87. AHwang Y, Dell CA, Sparks MJ, Watson TD, Langefeld CD, Comeau ME, et al. Clinician judgment vs. formal scales for predicting intracerebral hemorrhage outcomes. *Neurology.* (2016) 86:126–33. doi: 10.1212/WNL.0000000000002266
88. Hemphill JC 3rd, Bonovich DC, Besmertis L, Manley GT, Johnston SC. The ICH score: a simple, reliable grading scale for intracerebral hemorrhage. *Stroke.* (2001) 32:891–7. doi: 10.1161/01.STR.32.4.891
89. Nawabi J, Kniep H, Elsayed S, Friedrich C, Sporns P, Rusche T, et al. Imaging-based outcome prediction of acute intracerebral hemorrhage. *Transl Stroke Res.* (2021). doi: 10.1007/s12975-021-00891-8
90. Sennfält S, Pihlsgård M, Petersson J, Norrving B, Ullberg T. Long-term outcome after ischemic stroke in relation to comorbidity - an observational study from the Swedish Stroke Register (Riksstroke). *Eur Stroke J.* (2020) 5:36–46. doi: 10.1177/2396987319883154

91. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis.* (2015) 115:211–52. doi: 10.1007/s11263-015-0816-y
92. Chan S, Siegel EL. Will machine learning end the viability of radiology as a thriving medical specialty? *Br J Radiol.* (2019) 92:20180416. doi: 10.1259/bjr.20180416
93. ABeam L, Kohane IS. Big data and machine learning in health care. *J Am Med Assoc.* (2018) 319:1317–8. doi: 10.1001/jama.2017.18391
94. SLiebeskind S, Scalzo Editorial F. Machine learning and decision support in stroke. *Front Neurol.* (2020) 11:486. doi: 10.3389/fneur.2020.00486
95. Vayena A, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. *PLoS Med.* (2018) 15:e1002689. doi: 10.1371/journal.pmed.1002689
96. Obermeyer Z, Emanuel EJ. Predicting the future - big data, machine learning, and clinical medicine. *N Engl J Med.* (2016) 375:1216–9. doi: 10.1056/NEJMp1606181
97. Marra A, Ely EW, Pandharipande PP, Patel MB. The ABCDEF bundle in critical care. *Crit Care Clin.* (2017) 33:225–43. doi: 10.1016/j.ccc.2016.12.005
98. Cramer SC, Le V, Saver JL, Dodakian L, See J, Augsburger R, et al. Intense arm rehabilitation therapy improves the modified rankin scale score: association between gains in impairment and function. *Neurology.* (2021) 96:e1812–22. doi: 10.1212/WNL.0000000000011667
99. Gui H, Zheng R, Ma C, Fan H, Xu L. An architecture for healthcare big data management and analysis. In: *International Conference on Health Information Science*. Cham: Springer (2016). p. 154–60. doi: 10.1007/978-3-319-48335-1_17

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neurology

ORIGINAL RESEARCH
published: 23 December 2021
doi: 10.3389/fneur.2021.761092

Prediction-Driven Decision Support for Patients With Mild Stroke: A Model Based on Machine Learning Algorithms

Xinping Lin[1,2†], Shiteng Lin[1,2†], XiaoLi Cui[3†], Daizun Zou[1,2], FuPing Jiang[4], JunShan Zhou[5], NiHong Chen[5], Zhihong Zhao[6], Juan Zhang[3*] and Jianjun Zou[2,7*]*

[1] School of Basic Medicine and Clinical Pharmacy, China Pharmaceutical University, Nanjing, China, [2] Department of Clinical Pharmacology, Nanjing First Hospital, Nanjing Medical University, Nanjing, China, [3] Department of Neurology, Nanjing Yuhua Hospital, Yuhua Branch of Nanjing First Hospital, Nanjing Medical University, Nanjing, China, [4] Department of Geriatrics, Nanjing First Hospital, Nanjing Medical University, Nanjing, China, [5] Department of Neurology, Nanjing First Hospital, Nanjing Medical University, Nanjing, China, [6] Department of Neurology, The First Affiliated Hospital (People's Hospital of Hunan Province), Hunan Normal University, Changsha, China, [7] Department of Clinical Pharmacology, Nanjing First Hospital, China Pharmaceutical University, Nanjing, China

OPEN ACCESS

Edited by:
Ramin Zand, Pennsylvania State University, United States

Reviewed by:
Vida Abedi, The Pennsylvania State University, United States
Craig S. Anderson, University of New South Wales, Australia

***Correspondence:**
Jianjun Zou
zoujianjun100@126.com
Juan Zhang
zhangjuan_1@126.com
Zhihong Zhao
zhaozhihong51818@hotmail.com

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Stroke, a section of the journal Frontiers in Neurology

Received: *19 August 2021*
Accepted: *22 November 2021*
Published: *23 December 2021*

Citation:
Lin X, Lin S, Cui X, Zou D, Jiang F, Zhou J, Chen N, Zhao Z, Zhang J and Zou J (2021) Prediction-Driven Decision Support for Patients With Mild Stroke: A Model Based on Machine Learning Algorithms. Front. Neurol. 12:761092. doi: 10.3389/fneur.2021.761092

Background and Purpose: Treatment for mild stroke remains an open question. We aim to develop a decision support tool based on machine learning (ML) algorithms, called DAMS (Disability After Mild Stroke), to identify mild stroke patients who would be at high risk of post-stroke disability (PSD) if they only received medical therapy and, more importantly, to aid neurologists in making individual clinical decisions in emergency contexts.

Methods: Ischemic stroke patients were prospectively recorded in the National Advanced Stroke Center of Nanjing First Hospital (China) between July 2016 and September 2020. The exclusion criteria were patients who received thrombolytic therapy, age < 18 years, lack of 3-month modified Rankin Scale (mRS), disabled before the index stroke, with an admission National Institute of Health stroke scale (NIHSS) > 5. The primary outcome was PSD, corresponding to 3-month mRS ≥ 2. We developed five ML models and assessed the area under curve (AUC) of receiver operating characteristic, calibration curve, and decision curve analysis. The optimal ML model was selected to be DAMS. In addition, SHapley Additive exPlanations (SHAP) approach was introduced to rank the feature importance. Finally, rapid-DAMS (R-DAMS) was constructed for a more urgent situation based on DAMS.

Results: A total of 1,905 mild stroke patients were enrolled in this study, and patients with PSD accounted for 23.4% (447). There was no difference in AUCs between the five models (ranged from 0.691 to 0.823). Although there was similar discriminative performance between ML models, the support vector machine model exhibited higher net benefit and better calibration (Brier score, 0.159, calibration slope, 0.935, calibration intercept, 0.035). Therefore, this model was selected for DAMS. In addition, SHAP

approach showed that the most crucial feature was NIHSS on admission. Finally, R-DAMS was constructed and there was similar discriminative performance between R-DAMS and DAMS, but the former performed worse on calibration.

Conclusions: DAMS and R-DAMS, as prediction-driven decision support tools, were designed to aid clinical decision-making for mild stroke patients in emergency contexts. In addition, even within a narrow range of baseline scores, NIHSS on admission is the strongest feature that contributed to the prediction.

Keywords: mild stroke, machine learning, post-stroke disability, decision support tool, predictive model

INTRODUCTION

Around half of patients with ischemic stroke have mild neurological symptoms (1), usually with the expectation that such patients will come back to their pre-stroke activities regardless of the treatment. However, over one-third of mild stroke patients present with some degree of post-stroke disability (PSD) (2–4), which may be the result of inadequate acute treatments, early stroke recurrence, serious complications, or other reasons (1, 5). For the acute treatment of mild stroke patients, the guidelines from the American Heart Association/American Stroke Association (AHA/ASA) (6) distinguish disabling from non-disabling stroke and recommend intravenous (IV) alteplase only for the former. Nonetheless, the more certain, but not definitive, concept of "disabling stroke" is subjective and requires interpretation by individual neurologists. On the other hand, there is a trade-off between the benefits of IV alteplase and the risk of symptomatic intracranial hemorrhage (sICH). Therefore, decisions on how to treat mild stroke patients should be made on an individual basis.

3-month modified Rankin Scale (mRS), a valuable instrument for testing therapeutic interventions (7, 8), was used to assess the levels of PSD (5, 8). For mild stroke patients who only received medical therapy but had PSD, such therapy is not enough. Therefore, mild stroke patients who would be at high risk of PSD if they only received medical therapy should be early identified in emergency contexts, and some aggressive treatments, such as IV alteplase or close monitoring preventing worsening, should be taken in time. Unexpectedly, neurologists' overall accuracy for identifying those patients was staggeringly low (16.9%) (9). Each day that such a problem continues to exist means that uncounted mild stroke patients are being left with preventable disability.

However, none of the previously published risk models which were developed to predict the function outcome after stroke are fit to solve this problem. For example, the Totaled Health Risks in Vascular Events (THRIVE) score and the Houston Intra-Arterial Therapy (HIAT) score assign 0 points for National Institute of Health stroke scale (NIHSS) ≤ 5, losing the predictive power of NIHSS in mild stroke patients (10, 11). NIHSS on admission has been proven to be a strong predictor of PSD (5). Thus, despite convenient clinical applicability, these models cannot accurately identify mild stroke patients at high risk of PSD. Such models remain inadequate.

With the increased clinical data gathered for each patient, modern medical decision-making demands accurate, novel, and prediction-driven decision support. Machine learning (ML) algorithm, as a burgeoning statistical approach, is well-suited for that mission. Numerous studies with a considerable number of patients have shown great potential for ML approaches to predict recurrence (12), swallowing recovery (13), or aphasia (14) in patients with stroke. However, a model based on ML algorithms, focusing on the more debatable area of treating MS, has not yet been established.

Here, our goal was to develop and validate a prediction-driven decision support tool based on ML algorithms, called DAMS (Disability After Mild Stroke), to early identify mild stroke patients who would be at high risk of PSD if they only received medical therapy, and more importantly, to assist neurologists to make individual clinical decisions for mild stroke patients.

MATERIALS AND METHODS

Study Population

The study population involved the sequential ischemic stroke patients within 12 h of symptoms onset recorded in the National Advanced Stroke Center of Nanjing First Hospital (China) between July 2016 and September 2020. The exclusion criteria were patients who received thrombolytic therapy, age < 18 years, lack of 3-month mRS, who were disabled before the stroke (premorbid mRS score ≥ 2), with an admission NIHSS > 5. The primary outcome was PSD, corresponding to 3-month mRS ≥ 2.

Based on the Helsinki declaration, this study was allowed by the ethics committee of Nanjing First Hospital (document number: KY20130424-01), and informed consent of all patients was obtained.

Patient Clinical and Demographic Variables

Data used for prediction were routinely gathered and stored in the electronic health record. Demographic variables included age, education level, and education. Laboratory data included fasting blood glucose (FBG), systolic blood pressure (SBP), and platelet count. The quality of laboratory data was validated throughout the study period by regular internal quality control procedures and participation in an External Quality Assessment scheme. Comorbidities were diagnosed by experienced clinicians and identified according to *International Statistical Classification of Diseases and Related Health Problems, 10th Revision [ICD-10]* codes, including hypertension, diabetes mellitus, and atrial fibrillation. Clinical symptoms included language disorder, facial paralysis, hemiplegia, and dizziness. Medication use

history was recorded on admission. Based on the clinical characteristics, imaging, and laboratory examination, ischemic stroke etiology was classified by a trained physician by Trial of Org 10172 in Acute Stroke Treatment (TOAST) criteria (15). NIHSS on admission and 3-month mRS were evaluated by certified assessors during telephone questionnaires or face-to-face interviews with the patients, their relatives, or general practitioners. Data must have been recorded and available in the electronic health record before prediction to be included.

Statistical Analysis

The continuous variable data was presented as the median value and interquartile range, using Mann-Whitney U test for clinical and demographic comparison between two groups. Univariate tests were conducted using Pearson's chi-square test or Fisher's exact test for categorical data which were indicated as the number of events (fraction of the total). All tests were two-sided and p-values <0.05 were considered statistically significant. The above statistics and descriptions were implemented with SPSS version 25.0 (IBM Corporation, Armonk, NY, USA).

ML Algorithms

Before introducing the ML prediction model with the demographic and clinical variables mentioned above, missing values were first filled following the k-nearest neighbor algorithm (16). In addition, patients who missed more than one data would be excluded. The continuous data were standardized by z-score normalization (17), and the categorical data were converted by one-hot encoding (18). To select the ML algorithm that exhibits the best predictive ability, five ML classifiers, logistic regression (LR), support vector machine (SVM), random forest classifier (RFC), extreme gradient boosting (XGB), and deep neural network (DNN), were implemented for model construction to predict PSD in mild stroke patients.

Feature Selection

Superfluous and extraneous factors may lead to model overfitting and affect the predictive power of the model, respectively. Thus, a feature selection process was carried out in the study. All variables with significant difference ($p < 0.05$) in the univariate analysis were subjected to the least absolute selection and shrinkage operator (LASSO) algorithm, which is available for software python (version 3.7; https://www.python.org/). LASSO algorithm implements variable selection and regularization to improve the prediction accuracy and interpretability of the model (19). Finally, variables with non-zero coefficients determined by LASSO method were incorporated for building ML models. The feature selection algorithm was carried out with Python Scikit-learn environment (version 0.23.2).

Model Development

Supervised ML algorithms mentioned above with binary classification (PSD and non-PSD) were applied to establish predictive models. The study population was randomly divided into the training set (80%) for developing models and the testing set (20%) for assessing the models' performance. In the training step, 10-fold cross-validation was implemented, dividing and generating ten different derivation and inner validation subsets, which improved the generalizability and avoided overfitting. Grid search algorithm was adapted to tune model hyper-parameters to achieve the highest area under curve (AUC) of receiver operating characteristic (ROC).

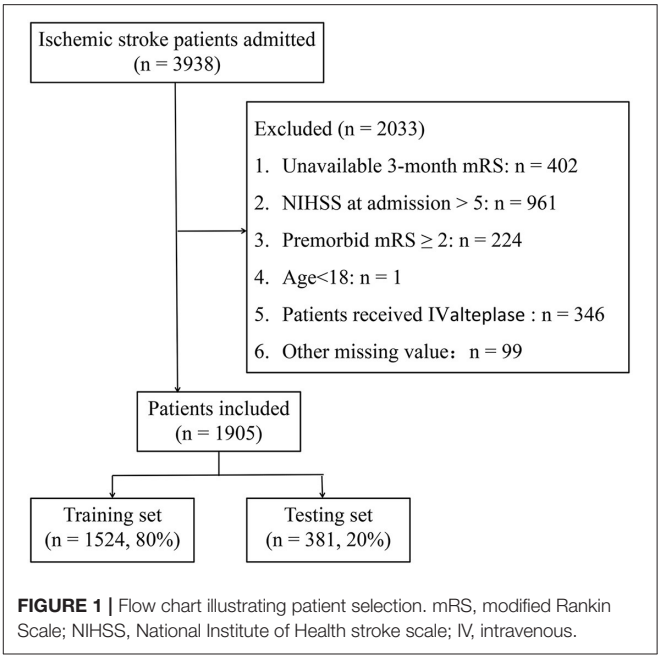


FIGURE 1 | Flow chart illustrating patient selection. mRS, modified Rankin Scale; NIHSS, National Institute of Health stroke scale; IV, intravenous.

Model Evaluation

Upon obtaining the models, the predictive performance was assessed on a testing set according to scores of AUC of ROC, drawn by sensitivity and 1-specificity across a series of cut-off points. Discrimination of the ML model on the testing set was evaluated by AUC. Delong test was carried out to compare the ROC curves in different models. Calibration of the ML model on the testing set was evaluated by calculating Brier score, calibration slope, and calibration intercept. The difference between the estimated and observed risk for PSD was calculated by Brier score, and the model with calibration slope = 1 and calibration intercept = 0 indicated perfect calibration. In addition, the null model Brier score was calculated to compare the relative gain of the algorithms to this benchmark (20). Decision curve analysis was introduced to evaluate the clinical utility (weighted average of true positives and false positives) by calculating the net benefits in the range of threshold probabilities. To evaluate the dominance of the ML models in terms of predictive performance, we also implemented THRIVE and HIAT score on the testing set (10, 11). Finally, the optimal model was selected for DAMS.

Feature Importance

ML models were accused of being "black boxes," which means that the development and validation processes of ML models are uninterpretable. In order to rank features in ML models, we introduced the SHapley Additive exPlanations (SHAP) approach. The SHAP approach has a high potential for rationalization of

TABLE 1 | Demographic and clinical data of the patients.

	Total (n = 1905)	Non-PSD (n = 1458)	PSD (n = 447)	*p*-value
Demographic				
Age years median (IQR)	65 (58–73)	63 (56–71)	70 (63–79)	<0.001*
Male sex n (%)	1337 (70.2%)	1046 (71.7%)	291 (65.1%)	0.007*
BMI kg/m median (IQR)	24.57 (22.49–26.67)	24.62 (22.49–26.67)	24.39 (22.22–26.67)	0.089
Education n (%)				0.018*
0–6	739 (38.8%)	540 (37.0%)	199 (45.5%)	
6–9	617 (32.4%)	487 (33.4%)	130 (29.1%)	
9–12	392 (20.6%)	314 (21.5%)	78 (17.4%)	
>12	157 (8.2%)	117 (8.0%)	40 (8.9%)	
Risk factors of n (%)				
Hypertension	1322 (69.4%)	977 (67.0%)	345 (77.2%)	<0.001*
Diabetes mellitus	557 (29.2%)	394 (27.0%)	163 (36.5%)	<0.001*
Dyslipidemia	55 (2.9%)	44 (3.0%)	11 (2.5%)	0.538
Coronary artery disease	158 (8.3%)	106 (7.3%)	52 (11.6%)	0.003*
Atrial fibrillation	90 (4.7%)	62 (4.3%)	28 (6.3%)	0.079
Previous TIA	8 (0.4%)	7 (0.5%)	1 (0.5%)	0.689
Previous ischemic stroke	196 (10.3%)	136 (9.3%)	60 (13.4%)	0.013*
Previous hemorrhagic stroke	46 (2.4%)	27 (1.9%)	19 (4.3%)	0.004*
Current smoker	877 (46.0%)	710 (48.7%)	167 (37.4%)	<0.001*
Current drink	652 (34.2%)	529 (36.3%)	123 (27.9%)	0.001*
Clinical symptoms *n* (%)				
Amaurosis	4 (0.2%)	3 (0.2%)	1 (0.2%)	1.000
Language disorder	61 (3.2%)	44 (3.0%)	17 (3.8%)	0.409
Facial paralysis	840 (44.1%)	612 (42.0%)	228 (51.0%)	0.001*
Hemiplegia	1271 (66.7%)	924 (63.4%)	347 (77.6%)	<0.001*
Dizziness	225 (11.8%)	176 (12.1%)	49 (11.0%)	0.525
Consciousness disturbance	27 (1.4%)	16 (1.1%)	11 (2.5%)	0.038*
Sensory disturbance	293 (15.4%)	230 (15.8%)	63 (14.1%)	0.389
Medication use history n (%)				
Previous antiplatelet	206 (10.8%)	144 (9.9%)	62 (13.9%)	0.017*
Previous anticoagulation	29 (1.5%)	23 (1.6%)	6 (1.3%)	0.722
Previous statin	115 (6.0%)	84 (5.8%)	31 (6.9%)	0.362
TOAST classification				<0.001*†
LAA (%)	895 (47.0%)	620 (42.5%)	275 (61.5%)	
CE (%)	124 (6.5%)	89 (6.1%)	35 (7.8%)	
SAO (%)	824 (43.3%)	697 (47.8%)	127 (28.4%)	
SOC (%)	16 (0.8%)	13 (0.9%)	3 (0.7%)	
SUC (%)	46 (2.4%)	39 (2.7%)	7 (1.6%)	
Baseline data				
Premorbid mRS=1 (%)	90 (4.7%)	52 (3.6%)	38 (8.5%)	<0.001*†
NIHSS at admission median (IQR)	2 (1–3)	2 (1–3)	3 (2–4)	<0.001*†
SBP mmHg median (IQR)	143 (130–158)	143 (130–158)	146 (130–160)	0.010*†
DBP mmHg median (IQR)	85 (80–90)	85 (80–90)	84 (80–90)	0.352
Platelet count 10/L median (IQR)	204 (164–204)	205 (165–241)	201 (158–234)	0.068
Creatinine mmol/L median (IQR)	77 (58–82)	75 (59–81)	82 (58–88)	0.022*†
FBG mmol/L median (IQR)	5.97 (4.58–6.61)	5.86 (4.55–6.40)	6.31 (4.64–7.36)	0.002*†
TC mmol/L median (IQR)	4.51 (3.77–5.16)	4.50 (3.77–5.15)	4.54 (3.73–5.19)	0.547
TG mmol/L median (IQR)	1.69 (1.02–1.95)	1.73 (1.04–1.98)	1.58 (0.99–1.80)	0.006*†
HDL mmol/L median (IQR)	1.08 (0.86–1.19)	1.07 (0.87–1.19)	1.07 (0.86–1.21)	0.917
LDL mmol/L median (IQR)	2.72 (2.12–3.26)	2.70 (2.13–3.24)	2.76 (2.11–3.34)	0.189

IQR interquartile range; BMI body mass index; TIA transient ischemic attacks; TOAST Trial of Org 10172 in Acute Stroke Treatment; LAA large artery atherosclerosis; CE cardioembolism; SAO small artery occlusion; SOC stroke of other determined cause; SUC stroke of undetermined cause; mRS modified Ranking Scale; NIHSS National Institutes of Health Stroke Scale; SBP systolic blood pressure; DBP diastolic blood pressure; FBG fasting blood glucose; TC total cholesterol; TG triglycerides; HDL high-density lipoprotein; LDL low-density lipoprotein. Data are given as n (%) or median (interquartile range).

**Variables included into the least absolute selection shrinkage operator regression (P<0.05).*

† Variables selected by the least absolute selection shrinkage operator regression.

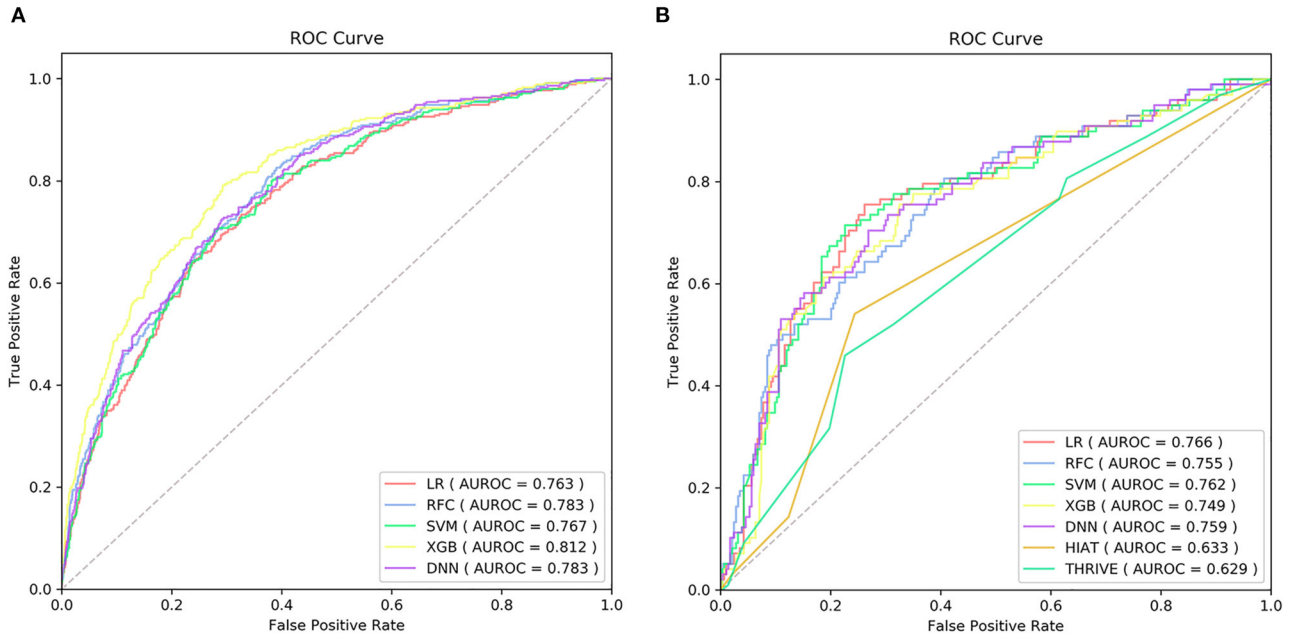


FIGURE 2 | The receiver operating characteristic curve (ROC) of the machine learning models on training set **(A)** and ROC of the machine learning models and traditional models on testing set **(B)** AUC, the area under curve; LR, logistic regression; SVM, support vector machine; RFC, random forest classifier; XGB, extreme gradient boost; DNN, deep neural network; THRIVE, Totaled Health Risks in Vascular Events; HIAT, Houston Intra-arterial Recanalization Therapy.

TABLE 2 | Discrimination and calibration of each machine learning algorithms on the testing set.

Model	AUC (95% CL)	Sensitivity %	Specificity %	Accuracy %	Intercept	Slope	Brier
LR	0.766 (0.709–0.823)	78.6	64.3	68.0	−0.129	0.805	0.221
RFC	0.755 (0.699–0.812)	80.6	59.4	64.8	−0.488	1.5533	0.228
SVM	0.762 (0.705–0.819)	74.5	71.0	71.9	0.035	0.935	0.159
XGB	0.749 (0.691–0.807)	70.4	68.2	68.8	0.068	0.879	0.165
DNN	0.759 (0.702–0.816)	74.5	67.1	69.0	−0.030	0.576	0.227

AUC, area under curve of receiver operating characteristic; CL, confidence interval; LR, logistic regression; RFC, random forest classifier; SVM, support vector machine; XGB, extreme gradient boosting; DNN, deep neural network.
Null model Brier score = 0.180.

the predictions from sophisticated ML models (21). In addition, the SHAP method indicates whether the effect of a feature on the result is positive or negative.

Rapid Prediction Model

DAMS may include some variables that take a relatively long time to obtain in emergency contexts, such as triglycerides and creatinine levels. For a more urgent situation, rapid-DAMS (R-DAMS), which excluded these variables, would be constructed based on DAMS. Then we will compare R-DAMS with DAMS in multiple dimensions such as ROC, calibration curve, and decision curve analysis.

RESULTS

Study Population

As shown in **Figure 1**, 1,905 patients met the inclusion criteria and were included in the present study. Patients with PSD account for 23.5% (447) of mild stroke patients; analogous proportions of PSD patients were established between training and testing sets (22.9 vs. 25.7%, $p > 0.05$). The median age of included patients was 65 (interquartile range: 58–73) years and 1,337 (70.2%) patients were men. The baseline statistics of both PSD and non-PSD groups were exhibited in **Table 1**. The characteristics of the patients struck a balance between the training ($n = 1{,}524$, 80%) and testing ($n = 381$, 20%) sets (**Supplementary Table 1**).

Feature Selection

Table 1 shows that 21 features were significantly different ($p < 0.05$) between patients with and without PSD with univariate analyses. Then, nine features without non-zero coefficients were excluded by LASSO regression. The final 12 variables incorporated into ML models were age, NIHSS at admission, SBP, creatinine, FBG, triglyceride, hemiplegia, hypertension, previous

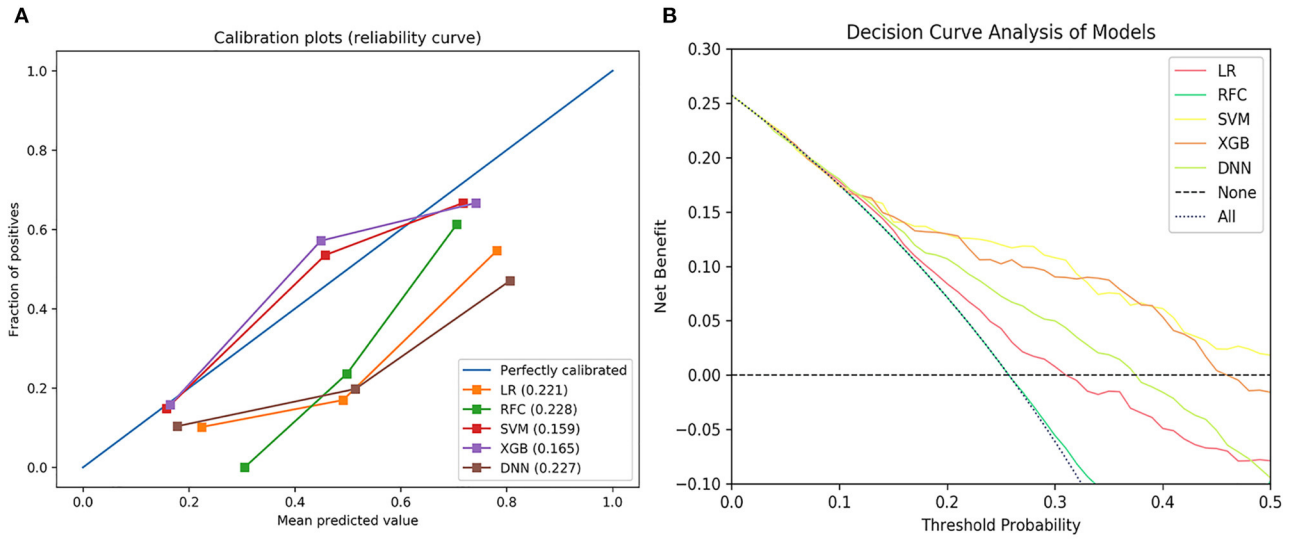


FIGURE 3 | The calibration curve of the machine learning models **(A)** and decision curve analysis of the machine learning models **(B)**. LR, logistic regression; SVM, support vector machine; RFC, random forest classifier; XGB, extreme gradient boost; DNN, deep neural network.

ischemic stroke, current drink, premorbid mRS, and TOAST classification.

Model Performance

Supplementary Table 2 exhibited the model hyper-parameters. ROCs of each model on the training set were shown in **Figure 2A**. **Table 2** shows performance metrics on the testing set, including AUC, sensitivity, Brier score, calibration slope, and calibration intercept.

As shown in **Table 2** and **Figure 2B**, the discriminative performance was observed in LR (AUC, 0.766; 95% CL, 0.709–0.823), RFC (AUC, 0.755; 95% CL, 0.699–0.812), SVM (AUC, 0.762; 95% CL, 0.705–0.819), XGB (AUC, 0.749; 95% CL, 0.691–0.807), and DNN (AUC, 0.759; 95% CL, 0.702–0.816) on the testing set, and AUCs on the testing set were 0.633 (95% CL, 0.577–0.689) and 0.629 (95% CL, 0.596–0.721) in HIAT and THRIVE score, respectively. The results of the DeLong test indicated that there was no statistical difference in the AUCs of the five ML models, but the AUCs of the five ML models was significantly better than that of HIAT and THRIVE scores (**Supplementary Table 3**).

The null model Brier score in the present study was 0.180. On the testing set, the Brier score ranged from 0.159 to 0.228. The calibration slope ranged from 0.576 to 1.553 and calibration intercept ranged from −0.488 to 0.068 (**Figure 3A** and **Table 2**). Decision curve analysis indicated that SVM and XGB models exhibited higher net benefit than other ML models as well as default strategies of treating all patients or no patients (**Figure 3B**).

There was no statistical difference in AUCs of the ML models, but the SVM model exhibited higher net benefit and calibration (Brier score, 0.159, calibration slope, 0.935, calibration intercept, 0.035). Therefore, the SVM model was selected to be DAMS.

Feature Importance

SHAP was introduced to rank the feature importance based on DAMS. **Figures 4A,B** show that the most important features were NIHSS on admission, age, and FBG. **Figure 4A** shows the individual distribution of SHAP values for single variables on DAMS. The redder the color of the sample dot, the higher the feature value of the variable for the sample. The higher the SHAP value of the abscissa, the greater the likelihood of PSD. Feature importance based on other ML models trained in the present study were provided in **Supplementary Figure 1.**

Rapid Prediction Model

DAMS included triglycerides and creatinine levels, which may take some time to obtain in an emergency context. Therefore, rapid-DAMS (R-DAMS) that excluded triglycerides and creatinine levels were constructed for more urgent situations. Then, we compared it with DAMS on a testing set using ROC, calibration curve, and decision curve analysis. As shown in **Figure 5** and **Supplementary Table 4**, there was no significant difference in AUC between R-DAMS and DAMS but the former performed slightly worse on calibration.

DISCUSSION

In this study, we demonstrated DAMS had the capacity to early identify mild stroke patients who would be at high risk of PSD if they only received medical therapy, achieving an optimal performance compared with our other ML models and previous scoring systems (THRIVE and HIAT scores). In addition, R-DAMS was developed for more urgent situations. DAMS and R-DAMS were able to generate reliable risk estimates for individuals, relying merely on data that were acquired in an emergency setting, and R-DAMS was able to do this within 4.5 h or less of symptom onset. Hitherto none of the prognosis models

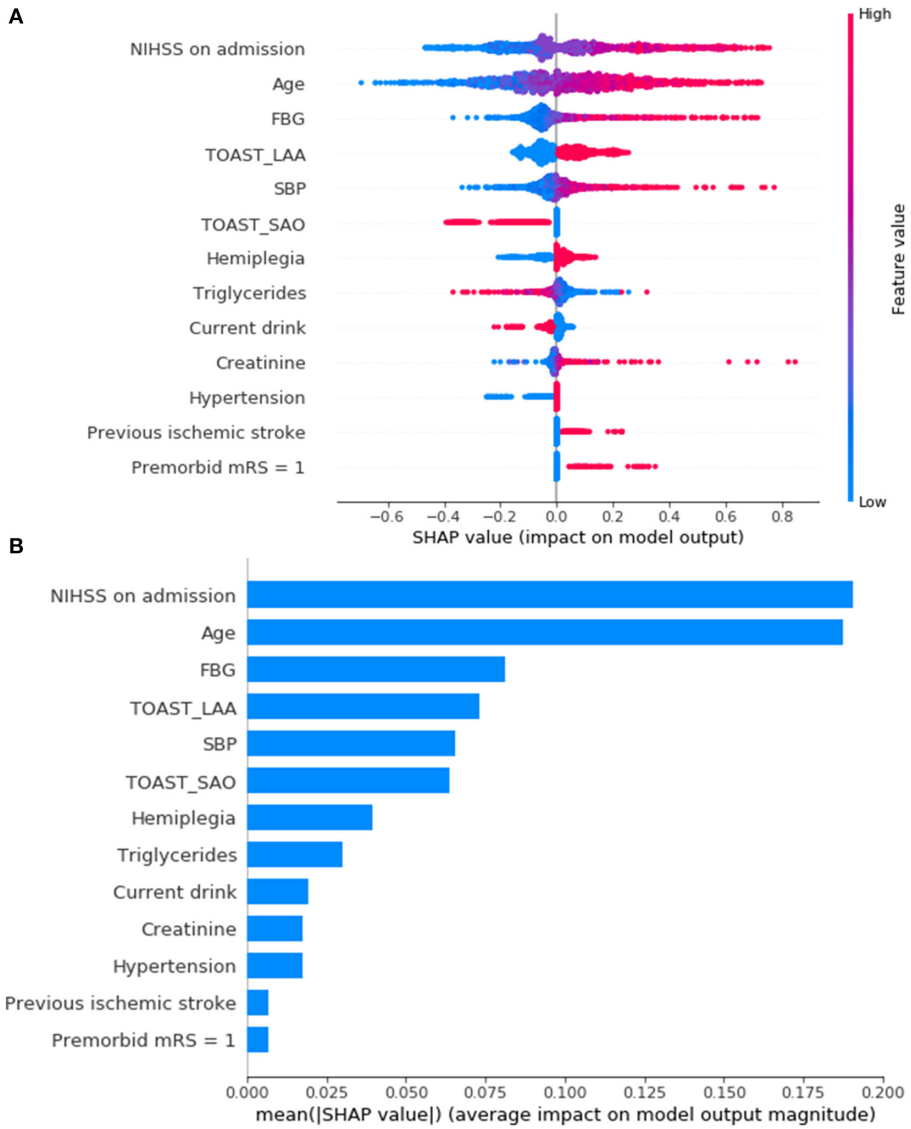


FIGURE 4 | Feature importance ranking based on Shapley Additive exPlanations (SHAP) values **(A,B)** in DAMS. **(A)** Red indicates that the value of the feature is high, and blue indicates that the value of the feature is low; the x-axis represents the SHAP values. The features are ranked according to the sum of the SHAP values for all patients. **(B)** Standard bar charts were drawn and sorted using the average absolute value of the shape values of each feature in DAMS. NIHSS, National Institutes of Health Stroke Scale; FBG, fasting blood glucose; TOAST, Trial of Org 10172 in Acute Stroke Treatment; LAA, large artery atherosclerosis; SAO, small artery occlusion; SBP, systolic blood pressure; mRS, modified Ranking Scale.

for mild stroke patients were developed for the prime objective of providing clinical decision support which targets treatment in the emergency contexts. DAMS and R-DAMS, as prediction-driven clinical decision support tools with this target in mind, are significant because neurologists faced a dilemma about the more debatable area of treating mild stroke: using IV alteplase but with the risk of sICH, or not using IV alteplase but potentially leaving the patient with brain ischemia.

In our study, the use of R-DAMS could offer neurologists effective support in the IV alteplase decision. Whether mild stroke patients will benefit from IV alteplase is still controversial. A meta-analysis reported that mild stroke patients who were treated with IV alteplase had lower odds of PSD even if the incidence of sICH increased slightly (22, 23). However, this research relied on retrospective data. The Potential of rtPA for Ischemic Strokes with Mild Symptoms (PRISMS) trial, which prospectively enrolled mild stroke patients without "clearly disabling" deficits, demonstrated no benefit for IV alteplase in this subgroup of patients (23). This trial defined a more certain, but not definitive, population for which the use of IV alteplase cannot be recommended. In line with the findings of the PRISMS trial, the AHA/ASA guidelines distinguish mild

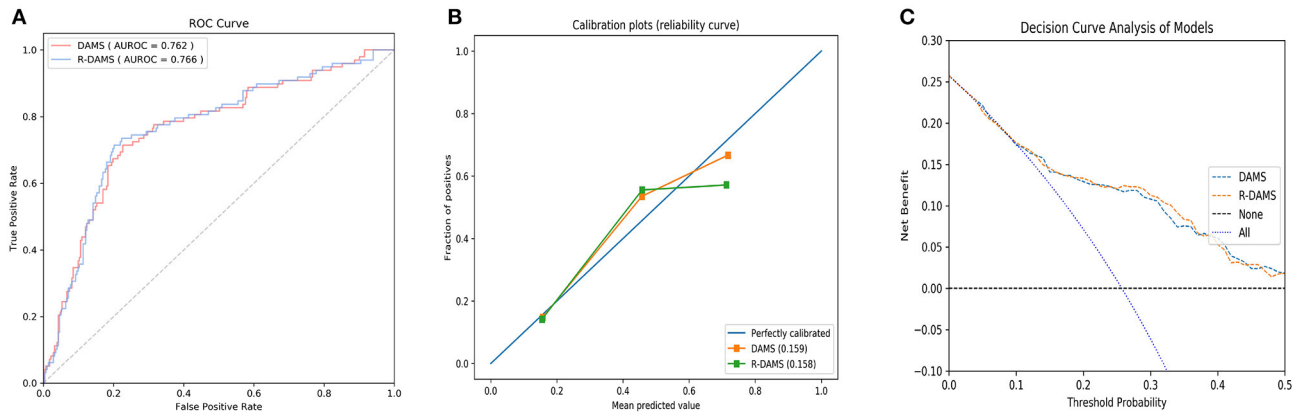


FIGURE 5 | The receiver operating characteristic curve (ROC) **(A)**, the calibration curve **(B)**, and decision curve analysis **(C)** between R-DAMS model and DAMS model. AUC, the area under curve.

disabling stroke from mild non-disabling stroke and recommend IV alteplase within 3 and 4.5 h only for the former (6). The population in our study was not categorized by whether their initial symptoms were "clearly disabling," because there are subtle differences in judgments about "clearly disabling" deficits in individual neurologists. In the present study, it should be stated explicitly that for patients who were identified to be at high risk of PSD by DAMS or R-DAMS, medical therapy alone is not enough. Thus, the two models support decision-making in the following ways: First, for mild stroke patients judged to be eligible for IV alteplase by current guidelines, R-DAMS was the best choice. The prediction generated by R-DAMS, paired with neurologists' expertise, enables them to choose the most appropriate candidates for IV alteplase. Second, for patients who are not eligible but are at high risk of PSD according to DAMS, best medical therapy alone with close monitoring may be an appropriate course of action.

On the other hand, we unlocked the potential utility of DAMS in secondary prevention. In a secondary analysis of the Acute Stroke or Transient Ischemic Attack Treated with Aspirin or Ticagrelor and Patient Outcomes (SOCRATES) trial, recurrent cerebrovascular event occurred at a significantly higher rate in patients with PSD than patients without PSD (29.0 vs. 3.7%) (5). Furthermore, as a leading cause of PSD (5, 8), a recurrent cerebrovascular event would do more irreparable harm to the patients at high risk of PSD compared with those at low risk. Therefore, effective prevention of recurrent cerebrovascular event to the patients at high risk of PSD portends a decreased risk of PSD. In the present study, DAMS could help to identify mild stroke patients at high risk of PSD, namely those who would most likely obtain substantial benefits from secondary prevention. For this patient group, a focus on evidence-based treatments for secondary prevention, and a support program to improve achievement of secondary prevention targets (e.g., blood pressure, diabetic control, cholesterol) in the long-term, might significantly reduce PSD.

With the expectation that DAMS and R-DAMS can be integrated into clinical practice, we had to acknowledge that our results represent only one step toward one component of a prediction-driven decision support tool for mild stroke patients. Some other steps need to be considered. Firstly, external validation, using data sets from different centres, should be carried out to duplicate the present results. Secondly, an impact study, quantifying whether application of DAMS and R-DAMS in clinical practice improves neurologists' decision making and subsequent patient outcome, is indispensable (24). Finally, development of simple-to-use software, providing a clear interpretation of the prediction and further treatment/prevention information based on this prediction, is required. The present results are promising but we need to emphasize that much work must be done before completely integrating DAMS and R-DAMS into clinical practice.

In the present study, several predictors of PSD have been discovered. NIHSS is a widespread assessment tool used to quantify the baseline severity in stroke patients. As shown in **Figure 4A**, even within a narrow range of baseline scores, the strongest feature that contributed to the prediction was NIHSS on admission and the higher the values of NIHSS, the more likely the chance of PSD. Noticeably, although the NIHSS has been widely favored in clinical research, some neurological deficits are measured objectively. For example, one NIHSS item, ataxia, confused hemiplegia and normal function by scoring ataxia as "normal" (0) in patients with hemiplegia (25). In the present study, patients with hemiplegia at admission are more likely to be PSD.

There are some limitations to the present study. Firstly, the mRS used to assess the levels of PSD in our study lacks sufficient detail to describe cognition and mood outcomes. A study published in 2017 in the Stroke journal demonstrates that a considerable number of patients with a good mRS outcome were incapable of socially reintegrating because of cognitive impairment and depression (26). However, the validity and reliability of the mRS was recognized by several clinical researchers (23, 27). Since the mRS is easy to use and interpret, the scale has been a valuable tool for assessing the efficacy of therapeutic interventions till now. Secondly, the lack of

external validation in our study hinders the evaluation of external generalizability. As a result, whether DAMS and R-DAMS, which have the selection bias that is inherent in any prediction model, can be used directly in other health institutions is still uncertain. To solve this problem, we provided as much detail as possible about the study cohort (**Table 1**). This information enables other institutions to judge whether their selected population matches the population here. In addition, the process of model development has been described in a precise fashion in Methods and **Supplementary Table 2**. Therefore, DAMS and R-DAMS may be transferable to other institutions. Thirdly, recurrent cerebrovascular event, a known predictor of PSD in mild stroke patients, was absent in the process by which DAMS and R-DAMS are developed (5, 8). Our models were initially designed for supporting clinical decision-making in emergency contexts, in which the data of recurrent cerebrovascular event is unavailable.

CONCLUSIONS

DAMS and R-DAMS represent one step within a larger process to early identify mild stroke patients who would be at high risk of PSD if they only received medical therapy, by assisting neurologists to make individual clinical decisions for mild stroke patients. Compared with our other ML models and previous scoring systems (THRIVE and HIAT scores), DAMS had a better performance and R-DAMS was able to operate within 4.5 h or less of symptom onset. Future work should build on these findings to transfer DAMS and R-DAMS to different centers.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

XL formed the conception and study design. XC and JZho did the data collection. NC and FJ did the data analysis. SL and DZ did the literature review and model development. XL and SL drafted the manuscript. ZZ, JZha, and JZo made significant revisions and supplied valuable improvement suggestions. The work presented in this paper was carried out in collaboration with all authors. All authors provided approval of the final version. All authors have read and agreed to the published version of the manuscript.

FUNDING

This study received the following financial support: National Natural Science Foundation of China grants 81673511, 81700398, 81970309; and Jiangsu key Research and Development Plan grant BE2017613.

ACKNOWLEDGMENTS

We gratefully acknowledge all the patients who participated in this study and to those who have provided their medical care.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fneur.2021.761092/full#supplementary-material

REFERENCES

1. Ferrari J, Krebs S, Sykora M. Intravenous thrombolysis and mechanical thrombectomy in patients with minor or rapidly improving neurological deficits. *Curr Opin Neurol.* (2019) 32:13–8. doi: 10.1097/WCO.0000000000000647
2. Ali S F, Siddiqui K, Ay H, Silverman S, Singhal A, Viswanathan A, et al. Baseline predictors of poor outcome in patients too good to treat with intravenous thrombolysis. *Stroke.* (2016) 47:2986–92. doi: 10.1161/STROKEAHA.116.014871
3. Khatri P, Conaway MR, Johnston KC. Ninety-day outcome rates of a prospective cohort of consecutive patients with mild ischemic stroke. *Stroke.* (2012) 43:560–2. doi: 10.1161/STROKEAHA.110.593897
4. Barber PA, Zhang J, Demchuk AM, Hill MD, Buchan AM. Why are stroke patients excluded from TPA therapy? An analysis of patient eligibility. *Neurology.* (2001) 56:1015–20. doi: 10.1212/WNL.56.8.1015
5. Cucchiara B, George DK, Kasner SE, Knutsson M, Denison H, Ladenvall P, et al. Disability after minor stroke and TIA: a secondary analysis of the SOCRATES trial. *Neurology.* (2019) 93:e708–16. doi: 10.1212/WNL.0000000000007936
6. Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K, et al. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke.* (2019) 50:e344–418. doi: 10.1161/STR.0000000000000211
7. Banks JL, Marotta CA. Outcomes validity and reliability of the modified Rankin scale: implications for stroke clinical trials: a literature review and synthesis. *Stroke.* (2007) 38:1091–6. doi: 10.1161/01.STR.0000258355.23810.c6
8. Coutts SB, Modi J, Patel SK, Aram H, Demchuk AM, Goyal M, et al. What causes disability after transient ischemic attack and minor stroke?: Results from the CT and MRI in the Triage of TIA and minor Cerebrovascular Events to Identify High Risk Patients (CATCH) Study. *Stroke.* (2012) 43:3018–22. doi: 10.1161/STROKEAHA.112.665141
9. Gao MM, Wang J, Saposnik G. The art and science of stroke outcome prognostication. *Stroke.* (2020) 51:1358–60. doi: 10.1161/STROKEAHA.120.028980
10. Flint AC, Cullen SP, Faigeles BS, Rao VA. Predicting long-term outcome after endovascular stroke treatment: the totaled health risks in vascular events score. *AJNR Am J Neuroradiol.* (2010) 31:1192–6. doi: 10.3174/ajnr.A2050
11. Hallevi H, Barreto AD, Liebeskind DS, Morales MM, Martin-Schild SB, Abraham AT, et al. Identifying patients at high risk for poor outcome after intra-arterial therapy for acute ischemic stroke. *Stroke.* (2009) 40:1780–5. doi: 10.1161/STROKEAHA.108.535146
12. Abedi V, Avula V, Chaudhary D, Shahjouei S, Khan A, Griessenauer CJ, et al. Prediction of long-term stroke recurrence using machine learning models. *J Clin Med.* (2021) 10:1286. doi: 10.3390/jcm10061286

13. Lee WH, Lim MH, Seo HG, Seong MY, Oh B-M, Kim S. Development of a novel prognostic model to predict 6-month swallowing recovery after ischemic stroke. *Stroke.* (2020) 51:440–8. doi: 10.1161/STROKEAHA.119.027439
14. Landrigan J-F, Zhang F, Mirman D. A data-driven approach to post-stroke aphasia classification and lesion-based prediction. *Brain.* (2021) 144:1372–83. doi: 10.1093/brain/awab010
15. Adams HP, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, 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. doi: 10.1161/01.STR.24.1.35
16. Batista GEAPA, Monard MC. An analysis of four missing data treatment methods for supervised learning. *Appl Artif Intell.* (2003) 17:519–33. doi: 10.1080/713827181
17. Shalabi LA, Shaaban Z, Kasasbeh B. Data mining: a preprocessing engine. *J Comput Sci.* (2006) 2:735–9. doi: 10.3844/jcssp.2006.735.739
18. Okada S, Ohzeki M, Taguchi S. Efficient partition of integer optimization problems with one-hot encoding. *Sci Rep.* (2019) 9:13036. doi: 10.1038/s41598-019-49539-6
19. Sun Y, Hu P, Wang J, Shen L, Xia F, Qing G, et al. Radiomic features of pretreatment MRI could identify T stage in patients with rectal cancer: preliminary findings. *J Magn Reson Imaging.* (2018) 48:1–7. doi: 10.1002/jmri.26309
20. Karhade AV, Schwab JH, Bedair HS. Development of machine learning algorithms for prediction of sustained postoperative opioid prescriptions after total hip arthroplasty. *J Arthroplasty.* (2019) 34:2272–7.e1. doi: 10.1016/j.arth.2019.06.013
21. Rodríguez-Pérez R, Bajorath J. Interpretation of compound activity predictions from complex machine learning models using local approximations and shapley values. *J Med Chem.* (2020) 63:8761–77. doi: 10.1021/acs.jmedchem.9b01101
22. Emberson J, Lees KR, Lyden P, Blackwell L, Albers G, Bluhmki E, et al. Effect of treatment delay, age, and stroke severity on the effects of intravenous thrombolysis with alteplase for acute ischaemic stroke: a meta-analysis of individual patient data from randomised trials. *Lancet.* (2014) 384:1929–35. doi: 10.1016/S0140-6736(14)60584-5
23. Khatri P, Kleindorfer DO, Devlin T, Sawyer RN, Starr M, Mejilla J, et al. Effect of alteplase vs aspirin on functional outcome for patients with acute ischemic stroke and minor nondisabling neurologic deficits: the PRISMS randomized clinical trial. *JAMA.* (2018) 320:156–66. doi: 10.1001/jama.2018.8496
24. Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ.* (2009) 338:b606. doi: 10.1136/bmj.b606
25. Heinemann AW, Harvey RL, McGuire JR, Ingberman D, Lovell L, Semik P, et al. Measurement properties of the NIH Stroke Scale during acute rehabilitation. *Stroke.* (1997) 28:1174–80. doi: 10.1161/01.STR.28.6.1174
26. Kapoor A, Lanctôt KL, Bayley M, Kiss A, Herrmann N, Murray BJ, et al. "Good Outcome" isn't good enough: cognitive impairment, depressive symptoms, and social restrictions in physically recovered stroke patients. *Stroke.* (2017) 48:1688–90. doi: 10.1161/STROKEAHA.117.016728
27. Thomalla G, Simonsen CZ, Boutitie F, Andersen G, Berthezene Y, Cheng B, et al. MRI-guided thrombolysis for stroke with unknown time of onset. *N Engl J Med.* (2018) 379:611–22. doi: 10.1056/NEJMoa1804355

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership