

Machine learning in data analysis for stroke/endovascular therapy

Edited by

Benjamin Yim, Shahram Majidi, Daniel Donoho
and Andrew Gauden

Published in

Frontiers in Neurology



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-3187-7
DOI 10.3389/978-2-8325-3187-7

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 data analysis for stroke/endovascular therapy

Topic editors

Benjamin Yim — John Muir Medical Center, United States

Shahram Majidi — Mount Sinai Hospital, United States

Daniel Donoho — Children's National Hospital, United States

Andrew Gauden — Stanford University, United States

Citation

Yim, B., Majidi, S., Donoho, D., Gauden, A., eds. (2023). *Machine learning in data analysis for stroke/endovascular therapy*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-3187-7

Table of contents

- 05 **Editorial: Machine learning in data analysis for stroke/endovascular therapy**
Ari Ettleson, Benjamin Yim and Daniel A. Donoho
- 08 **Aberrant interhemispheric functional reciprocities of the default mode network and motor network in subcortical ischemic stroke patients with motor impairment: A longitudinal study**
Yongxin Li, Zeyun Yu, Xuan Zhou, Ping Wu and Jiaxu Chen
- 23 **Frequency-tuned electromagnetic field therapy improves post-stroke motor function: A pilot randomized controlled trial**
Batsheva Weisinger, Dharam P. Pandey, Jeffrey L. Saver, Arielle Hochberg, Adina Bitton, Glen M. Doniger, Assaf Lifshitz, Ofir Vardi, Esther Shohami, Yaron Segal, Shira Reznik Balter, Yael Djemal Kay, Ariela Alter, Atul Prasad and Natan M. Bornstein
- 35 **Automated quantification of atrophy and acute ischemic volume for outcome prediction in endovascular thrombectomy**
Balázs Kis, Ain A. Neuhaus, George Harston, Olivier Joly, Davide Carone, Stephen Gerry, Zoltán Chadaide, András Pánczél, Eszter Czifrus, Viktória Csike, Ágnes Surányi, István Szikora and Loránd Erőss
- 42 **A deep learning approach to predict collateral flow in stroke patients using radiomic features from perfusion images**
Giles Tetteh, Fernando Navarro, Raphael Meier, Johannes Kaesmacher, Johannes C. Paetzold, Jan S. Kirschke, Claus Zimmer, Roland Wiest and Bjoern H. Menze
- 54 **Machine learning segmentation of core and penumbra from acute stroke CT perfusion data**
Freda Werdiger, Mark W. Parsons, Milanka Visser, Christopher Levi, Neil Spratt, Tim Kleinig, Longting Lin and Andrew Bivard
- 63 **The predictors of death within 1 year in acute ischemic stroke patients based on machine learning**
Kai Wang, Longyuan Gu, Wencai Liu, Chan Xu, Chengliang Yin, Haiyan Liu, Liangqun Rong, Wenle Li and Xiu'e Wei
- 73 **Machine learning-based prediction of cerebral hemorrhage in patients with hemodialysis: A multicenter, retrospective study**
Fengda Li, Anmin Chen, Zeyi Li, Longyuan Gu, Qiyang Pan, Pan Wang, Yuechao Fan and Jinhong Feng
- 84 **Prediction of subjective cognitive decline after corpus callosum infarction by an interpretable machine learning-derived early warning strategy**
Yawen Xu, Xu Sun, Yanqun Liu, Yuxin Huang, Meng Liang, Rui Sun, Ge Yin, Chenrui Song, Qichao Ding, Bingying Du and Xiaoying Bi

- 95 **Machine learning prediction of motor function in chronic stroke patients: a systematic review and meta-analysis**
Qinglin Li, Lei Chi, Weiyang Zhao, Lei Wu, Chuanxu Jiao, Xue Zheng, Kaiyue Zhang and Xiaoning Li
- 111 **OEDL: an optimized ensemble deep learning method for the prediction of acute ischemic stroke prognoses using union features**
Wei Ye, Xicheng Chen, Pengpeng Li, Yongjun Tao, Zhenyan Wang, Chengcheng Gao, Jian Cheng, Fang Li, Dali Yi, Zeliang Wei, Dong Yi and Yazhou Wu



OPEN ACCESS

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

*CORRESPONDENCE
Ari Ettleson
✉ ariettleson@gwu.edu

RECEIVED 26 July 2023
ACCEPTED 16 August 2023
PUBLISHED 24 August 2023

CITATION
Ettleson A, Yim B and Donoho DA (2023)
Editorial: Machine learning in data analysis for
stroke/endovascular therapy.
Front. Neurol. 14:1267655.
doi: 10.3389/fneur.2023.1267655

COPYRIGHT
© 2023 Ettleson, Yim and Donoho. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License](#)
(CC BY). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted which
does not comply with these terms.

Editorial: Machine learning in data analysis for stroke/endovascular therapy

Ari Ettleson^{1*}, Benjamin Yim² and Daniel A. Donoho^{1,3}

¹The George Washington University School of Medicine and Health Sciences, Washington, DC, United States, ²East Bay Brain and Spine Medical Group, Walnut Creek, CA, United States, ³Division of Neurosurgery, Center for Neuroscience, Children's National Medical Center, Washington, DC, United States

KEYWORDS

stroke, neurointerventional surgery, machine learning, computer vision, endovascular treatment

Editorial on the Research Topic

[Machine learning in data analysis for stroke/endovascular therapy](#)

Introduction

Despite current methods of treatment and prevention, ischemic stroke causes more than 7 million deaths each year worldwide (1). Because of its prevalence, stroke research necessitates large data sets with numerous variables. Analysis of massive multivariable datasets has historically been unfeasible. Machine learning (ML) offers a paradigm-shifting opportunity to integrate several modes of data from larger cohorts, driving stroke research forward through its capacity to unravel complex relationships within intricate datasets.

In this Research Topic of *Frontiers in Neurology*, we called for original work on the theme of ML in endovascular therapy and stroke to collect novel approaches that may yield discovery.

Residual disability after stroke significantly diminishes stroke survivor quality of life, so novel and innovative technologies may be able to improve care. A pilot study by [Weisinger et al.](#) found that frequency-tuned electromagnetic field therapy can improve stroke motor function. These field therapies might be the basis of future products available in the clinic or new avenues of research to better understand the relationship between post-ischemia injury and functional connectivity.

The articles featured in this issue demonstrate the transformative potential of ML through a consistent approach: training ML models on a subset of retrospective clinical data and harnessing their predictive power. [Li F. et al.](#), for example, trained five algorithms to predict cerebral hemorrhage. The algorithm with the best performance revealed LDL, HDL, CRP, and Hgb as the strongest predictors. [Wang et al.](#)'s study trained six different ML models on a subset of data from patients hospitalized with acute ischemic stroke. Each algorithm identified significant predictors of death at 1-year post-stroke. The most successful algorithm was then used to build an ML network calculator—essentially a smart calculator—to identify high-risk patients. Models like these are life-saving alternatives to calculators like the NIH stroke scale or CHA2DS2-VASc scores.

Advanced clinical decision-making tools can be valuable in managing rare stroke subtypes, such as corpus callosum (CC) infarction. ML thrives on extracting patterns and insights from limited datasets, rendering its application even more significant given the inherent data scarcity in such cases. Xu et al.'s prospective analysis of CC infarction does exactly that. Even from a small cohort ($N = 213$) of CC infarctions, they were able to identify a logistic regression model that predicted subjective cognitive decline. Predicting cognitive decline may be useful in both prognostication and the development of specific rehabilitation paradigms. Li Q. et al.'s article, also featured in this issue, similarly uses ML as an assessment tool for post-stroke patients. Their meta-analysis included over 70,000 patients and demonstrated the ability of ML models to predict poor motor function after stroke.

Ye et al.'s article goes beyond these "smart calculators" by integrating clinical data with radiomic features. Their innovative approach improved prognosis prediction over models with only clinical or radiologic data. Approaches that integrate automated image analysis with clinical data have the potential to facilitate clinical decision-making about complex, high-risk stroke patients.

Beyond ML analysis of retrospective datasets to isolate predictors of stroke outcome, this issue also exemplifies the potential for image-based research, particularly advanced imaging modalities such as CT and MR. This approach, demonstrated by Kis et al.'s work, extends to studies by Werdiger et al. and Tetteh et al.'s, and can uncover insights into stroke therapy that cannot be seen without ML. Werdiger et al.'s study trained ML models on CT perfusion images to identify areas of ischemia. Clinically available software that identifies ischemic tissue relies on only one input, whereas here, multiple inputs led to higher accuracy. Tetteh et al.'s study—which also trained ML models on perfusion studies—used MR perfusion data, instead of CT. Their approach performed comparably to expert graders at identifying the quality of collateral circulation during stroke. Li Y. et al.'s work included a similar structure, except this time in the form of serial fMRI scans. Images were introduced to a support vector machine, somewhat of a supervised ML model, and researchers were able to compare outputs from stroke patients, healthy patients, and their respective follow-ups. Their analysis allowed them to "see" improvements in functional neural homotopy.

In addition to original research on stroke, this Research Topic also called for projects that leveraged ML techniques to investigate endovascular therapies. Risk reduction in endovascular surgery is particularly ripe for discovery by ML models. Endovascular thrombectomy (EVT), commonly performed to treat ischemic stroke, carries a significant risk of bleeding and thrombosis, making selection criteria paramount to risk mitigation. Kis et al., whose study is featured in this issue, combined automated analysis of pre-EVT CT images with clinical data to predict outcomes and improve prognostication after EVT. These findings demonstrate the need to integrate software-based analysis with clinical data when determining the futility of high-risk procedures like EVT.

As the landscape of stroke care continues to evolve, capturing intraoperative data becomes imperative to the modernization process. This primarily involves capturing the data in the form of surgical video and the application of computer vision techniques to analyze it. Computer vision, a type of task ML algorithms can be trained to complete, can analyze video-guided surgical procedures

to reveal patterns and make nuanced predictions. Although endovascular therapy is not guided by high-resolution cameras, surgeons heavily rely on fluoroscopy. Other fields—laparoscopy and endoscopy—have already begun to capture, transmit, and analyze surgical video datasets using ML. Endovascular surgery lags behind.

Visual datasets—regardless of surgical field—are ripe for clinical discovery and generally under-exploited. Fluoroscopic runs are typically used for medical/diagnostic purposes. We (two practicing neurosurgeons, one of whom is a dual-trained cerebrovascular expert, and a medical student) have performed hundreds of angiograms; rarely do neurointerventionalists store angiographic data for systematic or quantitative analysis. Fluoroscopic images from endovascular procedures remain an untapped area of analysis. Existing research on barium swallow has demonstrated ML's ability to extract valuable information from videofluoroscopic datasets (2). ML analysis of fluoroscopic video from endovascular procedures is feasible, but to analyze the data we must first hit record.

In this editorial, we argue for the potential applications of computer vision in cerebrovascular care. In doing so, we described the current state of ML in stroke care, as well as its shortcomings. Using current examples of computer vision in surgical analysis, we will argue for the storage and analysis of fluoroscopic videos and images.

Areas with potential

The first area with potential for application of ML is skill assessment. After fellowship, neurointerventional surgeons are expected to self-regulate their improvement from advanced to master. After fellowship, however, their overseers are no longer present to provide consistent, constructive feedback. Attending surgeons have little to base their performance on. Comparing complication or reoperation rates with national averages are the extent of their limited options for self-assessment.

ML models have begun to perform accurate surgical skill assessments in fields like urology and skull base surgery. A model trained on videos of radical prostatectomy was able to accurately assess a surgeon's ability to perform a specific step within the procedure (3). In this study, measurements were verified by corroborating them with expert grading of the same videos. The model was even able to mitigate bias created by small differences in video quality between institutions included in the study.

ML models have also correctly predicted a surgeon's ability to control blood loss during a lifelike surgical hemorrhage simulator (4). After being shown a brief video clip of the surgeons operating to stop bleeding, these models were better able to predict a surgeon's performance than expert reviewers. Imagine this same study structure, but replace the endoscopic video with recorded fluoroscopic runs of thrombectomies. Perhaps an ML system could view a proceduralist's ability to perform the beginning of a procedure and predict whether they would accomplish successful revascularization. Alternatively, an ML system might understand a local practitioner's prior skillset and be able to better predict whether they should first attempt a blood vessel-opening procedure

in their local setting, perhaps with a tele-coach, or transfer the patient to a center of excellence.

Video data analysis can simplify complex surgical decision-making. Internal carotid artery stenosis, estimated to cause almost 10% of all ischemic strokes, can be treated with a multitude of procedures including endarterectomy, transfemoral carotid artery stenting (TF-CAS), or emerging, perhaps safer, technologies like transcatheter artery revascularization (TCAR) (5). At present, treatment allocation depends upon the judgment of the practitioner with few guidelines and highly variable utilization practices. AI models trained on clinical variables and fluoroscopic video might help surgeons to make informed decisions.

Call to action: building video data repository for stroke and cerebrovascular care

The quantity of periprocedural cerebrovascular data left on the table is staggering. Consider the aforementioned removal of stroke-inducing plaques from the carotid artery, whether by surgical endarterectomy or percutaneous angioplasty and stenting, performed more than 100,000 times each year (6). Those procedures generate petabytes of angiographic and intraoperative visual data. Other fields that rely heavily on high-resolution cameras to guide procedures are already creating large, publicly available surgical video repositories for research and innovation. A large multimodal data repository that includes pre-, intra-, and post-procedural images as well as labeled clinical data could shed new light on the challenges of cerebrovascular and stroke care.

References

1. GBD 2019 Stroke Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the global burden of disease study 2019. *Lancet Neurol.* (2021) 20:795–820. doi: 10.1016/S1474-4422(21)00252-0
2. Kim JK, Choo YJ, Choi GS, Shin H, Chang MC, Park D. Deep learning analysis to automatically detect the presence of penetration or aspiration in videofluoroscopic swallowing study. *J Korean Med Sci.* (2022) 37:e42. doi: 10.3346/jkms.2022.37.e42
3. Kiyasseh D, Laca J, Haque TF, Miles BJ, Wagner C, Donoho DA, et al. A multi-institutional study using artificial intelligence to provide reliable and fair feedback to surgeons. *Commun Med.* (2023) 3:42. doi: 10.1038/s43856-023-00263-3
4. Pangal DJ, Kugener G, Zhu Y, Sinha A, Unadkat V, Cote DJ, et al. Expert surgeons and deep learning models can predict the outcome of surgical hemorrhage from 1 min of video. *Sci Rep.* (2022) 12:8137. doi: 10.1038/s41598-022-11549-2
5. Flaherty ML, Kissela B, Khoury JC, Alwell K, Moormaw CJ, Woo D, et al. Carotid artery stenosis as a cause of stroke. *Neuroepidemiology.* (2013) 40:36–41. doi: 10.1159/000341410
6. Maini BS, Mullins TF III, Catlin J, O'Mara P. Carotid endarterectomy: a ten-year analysis of outcome and cost of treatment. *J Vasc Surg.* (1990) 12:732–40. doi: 10.1067/mva.1990.25015

Author contributions

AE: Conceptualization, Writing—original draft. BY: Supervision, Writing—review and editing. DD: Supervision, Writing—review and editing.

Funding

DD work was supported by NIH grant K23EB034110.

Conflict of interest

DD has unvalued equity in Dotphraise.ai and is a director and founder of the non-profit Surgeon's Data Science Collective. BY is a consultant for Imperative Care as well as a Principal Investigator for MEMBRANE trial by Cerenovus, Keep Registry by Kaneka, and The Imperative Trial.

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.



OPEN ACCESS

EDITED BY
Benjamin Yim,
Stanford University, United States

REVIEWED BY
Kai Yuan,
The Chinese University of Hong
Kong, China
Yanzhe Ning,
Capital Medical University, China

*CORRESPONDENCE
Yongxin Li
yxin-li@163.com
Ping Wu
wuping_3358@163.com

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

RECEIVED 18 July 2022
ACCEPTED 15 September 2022
PUBLISHED 04 October 2022

CITATION
Li Y, Yu Z, Zhou X, Wu P and Chen J
(2022) Aberrant interhemispheric
functional reciprocities of the default
mode network and motor network in
subcortical ischemic stroke patients
with motor impairment: A longitudinal
study. *Front. Neurol.* 13:996621.
doi: 10.3389/fneur.2022.996621

COPYRIGHT
© 2022 Li, Yu, Zhou, Wu and Chen.
This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Aberrant interhemispheric functional reciprocities of the default mode network and motor network in subcortical ischemic stroke patients with motor impairment: A longitudinal study

Yongxin Li^{1*}, Zeyun Yu², Xuan Zhou¹, Ping Wu^{2*} and
Jiaxu Chen¹

¹School of Traditional Chinese Medicine, Formula-Pattern Research Center, Jinan University, Guangzhou, China, ²Acupuncture and Tuina School/Tird Teaching Hospital, Chengdu University of Traditional Chinese Medicine, Chengdu, China

Purpose: The purpose of the present study was to explore the longitudinal changes in functional homotopy in the default mode network (DMN) and motor network and its relationships with clinical characteristics in patients with stroke.

Methods: Resting-state functional magnetic resonance imaging was performed in stroke patients with subcortical ischemic lesions and healthy controls. The voxel-mirrored homotopic connectivity (VMHC) method was used to examine the differences in functional homotopy in patients with stroke between the two time points. Support vector machine (SVM) and correlation analyses were also applied to investigate whether the detected significant changes in VMHC were the specific feature in patients with stroke.

Results: The patients with stroke had significantly lower VMHC in the DMN and motor-related regions than the controls, including in the precuneus, parahippocampus, precentral gyrus, supplementary motor area, and middle frontal gyrus. Longitudinal analysis revealed that the impaired VMHC of the superior precuneus showed a significant increase at the second time point, which was no longer significantly different from the controls. Between the two time points, the changes in VMHC in the superior precuneus were significantly correlated with the changes in clinical scores. SVM analysis revealed that the VMHC of the superior precuneus could be used to correctly identify the patients with stroke from the controls with a statistically significant accuracy of 81.25% ($P \leq 0.003$).

Conclusions: Our findings indicated that the increased VMHC in the superior precuneus could be regarded as the neuroimaging manifestation of

functional recovery. The significant correlation and the discriminative power in classification results might provide novel evidence to understand the neural mechanisms responsible for brain reorganization after stroke.

KEYWORDS

subcortical stroke, resting-state functional magnetic resonance imaging, functional homotopy, superior precuneus, machine learning

Introduction

Subcortical ischemic stroke is a common acute cerebrovascular disease, with approximately half of the patients exhibiting residual disabilities (1). Because of cerebral thrombosis or cerebral embolism in the brain, brain dysfunction always occurs within the default mode network (DMN) and motor-related areas, such as the precuneus, parahippocampus, supplementary motor area (SMA), primary motor cortex (M1), and premotor cortex (2–6). The DMN is one of the most widely studied functional brain networks that support aspects of cognition. Patients' cognitive recovery after stroke has been reported to be associated with functional connectivity (FC) impairments within the DMN (4, 7). The precuneus is a key region of the DMN. Recently, several studies have reported that patients with ischemic stroke exhibited decreased connectivity in the precuneus (8–10). Because of the important role of the precuneus in a wide spectrum of higher-order brain functions, such as motor function, decreased connectivity of the precuneus could inhibit the activity in the motor cortex and impact patients' motor function. It was reported that ~50% of patients with strokes showed impaired hand motor function in the chronic phase, which always caused negative effects on the quality of life (11). Additionally, brain function recovery after stroke is always difficult to evaluate (12). Mounting evidence has suggested that the recovery of impaired behavioral function is accompanied by brain reorganization after effective treatment (13–17). Identifying stroke recovery processes by neuroimaging will benefit the selection of suitable treatments for stroke.

Neuroimaging has become the most useful tool available to detect the functional and structural organization of the human brain. Previous neuroimaging studies have extensively explored the changes in the DMN and motor-related networks after stroke (4, 18–21). These studies have shown system-wide network disturbances and brain functional reorganization in patients with stroke. Additionally, analyses of functional and structural connectivity were also used to investigate the effects of interventions on the brain network following stroke (13, 15, 22). For example, a previous study found that patients with stroke showed a significant decrease in FC between the bilateral M1 after stroke, and the disrupted FC was restored with antiplatelet therapy (22). Patients with

chronic stroke completed a rehabilitation protocol and showed significantly enhanced FC between the posterior cingulate cortex (PCC)/precuneus and M1 in contrast to the prerehabilitation condition (16). The long-term training effect following robot-hand training was also detected in patients after stroke (23). A recent study on patients with stroke found that brain-computer interface technology could significantly improve interhemispheric FC in the motor network and motor outcomes poststroke (24). These previous studies implied that brain function recovery in patients with stroke is attributed to brain reorganization (25).

A model called functional homotopy was developed based on resting-state functional magnetic resonance imaging (MRI) to directly quantify interhemispheric functional reciprocities (19, 26). In this model, the brain shows a high degree of synchrony in patterns of spontaneous activity between homotopic interhemispheric counterparts (geometrically corresponding). Voxel-mirrored homotopic connectivity (VMHC) is a validated method that can be used to quantify functional homotopy between two hemispheres (27). VMHC is one of the most noticeable characteristics of the brain's essential functional architecture, which has an important influence on cognition and behavior by interhemispheric communication. Recently, as a conspicuous indicator of disturbed functional specialization of the brain, VMHC has been chosen in some diseases to explore the alterations of functional homotopy due to intervention or recovery (28–31). This method has also been used in stroke studies to investigate the changes in functional connections between cerebral hemispheres (10, 17, 28, 31). After receiving scalp acupuncture, patients with acute ischemic stroke showed a significant increase in VMHC values in the bilateral BA6 and BA8 (17). Scalp acupuncture can specifically strengthen the functional activities of the brain regions related to motor coordination in stroke. A recent study revealed that synchrony between the DMN and the sensorimotor network can facilitate motor recovery after stroke rehabilitation (16). Although the connectivity between the DMN and motor-related regions has been investigated in the above previous stroke studies, the longitudinal changes of the interhemispheric connection have neither been considered in stroke studies nor have the relevance between the changes in interhemispheric connection in these networks and the changes in clinical scores.

Thus, the purpose of the present study was to elucidate the neurological mechanism of the DMN and the motor network organization process in stroke, which has attracted intense attention from clinical and basic researchers. The VMHC method can be used to quantify interhemispheric connectivity. The longitudinal changes in functional homotopy in these networks provide some insights into the mechanism underlying recovery processing in patients with stroke. In order to achieve these purposes, we detected the VMHC changes in patients with subcortical ischemic stroke with motor impairment, hypothesizing that patients with ischemic stroke would show reduced interhemispheric functional reciprocities. Given the evidence for interhemispheric functional and structural pathway dysfunction in patients with stroke (17, 32–35), we expected the DMN and motor-related areas to be particularly affected. In patients with stroke, the abnormal VMHC partially recovered after treatment, and the affected VMHC between the bilateral hemispheres may be related to behavioral performance. Furthermore, we used a machine learning approach to assess whether these group differences could be used to accurately distinguish patients with stroke from healthy controls.

Methods and materials

Participants

Nineteen first-ever patients with stroke (eight women; mean age \pm SD: 64.7 ± 12.4 years; range from 37 to 81 years old) with unimanual motion defects due to subcortical ischemic lesions were recruited from the Department of Neurology of the First Affiliated Hospital of the Chengdu University of Traditional Chinese Medicine, China. Most of the lesions were located in the basal ganglia and nearby regions. The demographic data and clinical characteristics of patients with stroke are provided in Table 1. The inclusion criteria for patients were as follows: (1) pure unilateral motor hemiparesis at least 20 days after the first onset ischemic stroke incident (confirmation and location of stroke by MRI); (2) no other white matter pathology as proven by structural MRI; (3) absence of additional psychiatric or neurologic disorders; (4) no neglect, aphasia or dementia; and (5) no other previous experimental therapy before participating in this research. Thirteen age-matched healthy subjects (eight females; mean age \pm SD: 62.1 ± 10.8 years) served as controls. Control subjects did not exhibit any history of medical disorders and were not taking regular medication. All participants gave written informed consent before the study. This study was approved by the Ethics Committee of Chengdu University of Traditional Chinese Medicine (no. 2011KL-002), and this method was carried out in consideration of the approved guidelines.

Treatment and behavioral evaluation of recovery

Antiplatelet therapy was administered to all patients with stroke (10 mg Erigeron breviscapus injection, 75 mg clopidogrel once each day taken orally). Then, citicoline (0.5 g, daily) was injected intravenously to improve the clinical effects. Drug therapy was conducted for 1 month (30 days) for each patient. The functional MRI data of each patient were collected before and after the treatment at \sim 1-month intervals. The clinical evaluations of recovery, such as the Fugl-Meyer Assessment (FMA) and the Neurological Deficit Scores (NDS), were carried out on the same days when we collected the functional MRI scans of the patients. The FMA is one of the most established and commonly used motor outcome measures in stroke rehabilitative trials (36). This scale includes a total of 50 items, such as assessments of the tendon reflex; joint movement of the flexor and extensor of the shoulder; elbow, wrist, knee, and hip joints; the movement of small joints with the stability of the wrist and the ankle; and coordination ability and speed. The highest total score is 100 points, and the higher the scores are, the milder the impairments in motor function. As a trial outcome measure in patients with stroke, the NDS was performed to the severity of neurological functional deficits and to assess the severity of strokes.

Data acquisition

All fMRI data were acquired with a 3-T Siemens scanner (MAGNETOM Trio Tim, Siemens, Erlangen, Germany) at the West China Hospital MRI Center. We used a gradient echo planar imaging sequence to acquire fMRI data: 30 interleaved axial slices, slice thickness = 5 mm, repetition time (TR) = 2,000 ms, echo time (TE) = 30 ms, flip angle = 90° , field of view = $240 \text{ mm} \times 240 \text{ mm}$, matrix = 64×64 , and 180 volumes. The three-dimensional T1-weighted structural MRI was collected using a spin-echo planar image sequence with the following parameters: slices = 176; TR/TE = 1,900 / 2.26 ms, flip angle = 9° ; field of view = $256 \times 256 \text{ mm}$; voxel size = $1 \times 1 \times 1 \text{ mm}^3$. During the scanning, the head of each participant was fixed using foam cushions to limit head movement. All participants were instructed to remain awake, relax with their eyes closed, remain motionless, and try not to think about anything.

The lesions were manually segmented using MRIcron software on the T1-weighted MRI images. We generated a lesion mask for each patient and normalized the results to the Montreal Neurologic Institute (MNI) space. The lesion overlay map is displayed in Figure 1. The lesions of the patients were mainly located in the left basal ganglia.

TABLE 1 Demographical and clinical data.

Patient number	Sex	Age (years)	Affected hand	Pathogeny	Lesion duration (days)	FMA pre	FMA post	NDS pre	NDS post
1	F	69	Right	Left pons-centrum semiovale	135	90	96	13	5
2	M	73	Right	Left basal ganglia and centrum semiovale	132	80	84	26	23
3	F	43	Right	Left caudate nucleus and periventricular	25	92	96	15	6
4	F	71	Left	Right basal ganglia	32	88	95	21	10
5	F	75	Right	Left occipital gyrus and thalamus	22	74	90	32	17
6	M	73	Right	Left basal ganglia-centrum semiovale	26	84	91	24	18
7	M	81	Right	Left basal ganglia-internal capsule	22	83	92	23	12
8	M	49	Right	Left basal ganglia and centrum semiovale	56	85	93	24	16
9	M	67	Right	Left basal ganglia and periventricular	33	80	93	26	12
10	M	78	Right	Left basal ganglia-posterior horn of lateral ventricle	21	84	95	19	9
11	F	70	Right	Left centrum semiovale-periventricular	23	85	90	24	2
12	F	74	Right	Left periventricular	148	90	96	20	12
13	M	37	Right	Left thalamus-lenticular nucleus	148	88	94	24	14
14	M	79	Left	Right basal ganglia	50	85	92	26	15
15	F	60	Right	Left basal ganglia and radial area	23	80	85	28	20
16	M	73	Right	Left capsula externa-periventricular	56	77	90	31	20
17	F	54	Right	Left basal ganglia	45	80	85	26	23
18	F	54	Right	Left basal ganglia	23	86	92	22	14
19	M	63	Right	Left thalamus	21	82	85	25	18

F, female; FMA, Fugl-Meyer Motor Assessment; L, left; M, male; MBI, Modified Barthel Index; R, right.

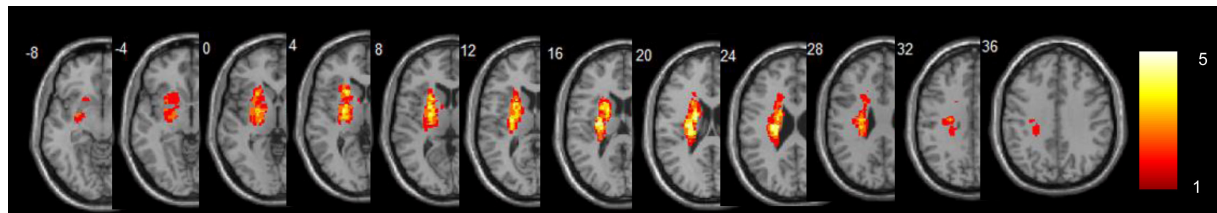


FIGURE 1

Lesion overlay map displaying regions of lesion overlap between participants. The n -value denotes the number of patients with a lesion in each voxel. L, left.

Imaging processing and statistical analysis

Behavioral data analysis

A paired t -test was conducted to examine whether the patients with stroke had actually improved based on the clinical scores from pretreatment to posttreatment.

Resting state functional connectivity

Data processing was conducted using the statistical parametric mapping (SPM12, London, UK, <http://www.fil.ion.ucl.ac.uk/spm>) package. The first 10 volumes of each subject were discarded in consideration of magnetization equilibrium effects and the adaptation of the subjects to the circumstances. Preprocessing comprised the following steps: (1) slice time correction; (2) head motion correction; (3) coregistration of the individual T1-weighted images to functional images; (4) segmentation of the T1-weighted images (gray matter, white matter, and cerebrospinal fluid) and spatial normalization to the MNI space by using a 12-parameter nonlinear transformation; (5) application of the transformation parameters to the functional images; (6) resampling to a voxel size of $3 \times 3 \times 3 \text{ mm}^3$; (7) smoothing (full width at half maximum = 6 mm); (8) temporal bandpass filtering (0.01–0.08 Hz); (9) linear and quadratic detrending; (10) regression of several nuisance covariates (white matter, cerebrospinal fluid, global signal, and the Friston 24 head motion parameters).

In the present study, we mainly focused on the DMN and motor network, which is remote from the primary lesion. We selected a user-defined mask including the DMN and motor network for the following analysis. This mask was produced from the cortical parcellation maps of the Yeo2011 resting state network (https://surfer.nmr.mgh.harvard.edu/fswiki/CorticalParcellation_Yeo2011). The components of the DMN and motor network in this template were extracted and combined to produce the user-defined mask. Then, this mask was averaged with its left-right mirrored version to generate a symmetrical mask for the following VMHC analysis (refer to [Supplementary Figure 1](#) for the symmetrical mask).

Voxel-mirrored homotopic connectivity

The VMHC was calculated using the DPARSF (<http://resting-fmri.sourceforge.net>) toolbox (37). For each subject, the Pearson correlation between each pair of mirrored voxel time series was computed to regard homotopic connectivity (27). Fisher z -transformation was then performed on the correlation values to improve normality and generate VMHC maps. Two-sample t -test analysis was performed to compare the VMHC map differences between the patients and the controls in the mask. Paired t -test analysis was performed to assess the longitudinal changes in patients between the two time points. The threshold of the resulting statistical map was a combination of $p < 0.005$ for a single voxel and a minimum cluster size of 297 mm^3 , which corresponds to a corrected threshold of $p < 0.05$ (AlphaSim correction). During the statistical analyses, age and sex were modeled as covariates. Finally, the regions that showed significant differences between groups were selected as our regions of interest, and the mean VMHC values were extracted from these regions for the following analyses. Treatment effects on the VMHC of these interesting regions were calculated among these groups.

Correlation between the neuroimaging index and clinical scores

To assess the clinical relevance of altered homotopic FC, partial correlations were calculated between the change in VMHC and the change in each clinical score between the two time points in patients with stroke. The FMA and NDS scores were included as analyzed clinical variables. The threshold was set at $p < 0.05$. The correlation results were also corrected by the Bonferroni method based on the number of regions selected as regions of interest. During the partial correlations, analyses were controlled for age, sex, disease duration, and lesion sizes.

Support vector machine analysis

To detect the above group differences specific to the patient group, a support vector machine analysis (SVM) was performed

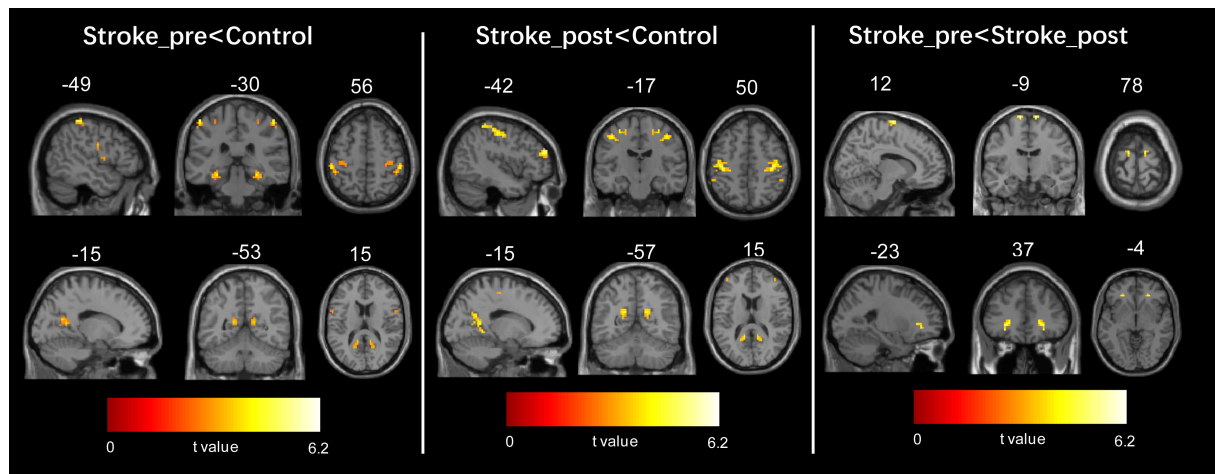


FIGURE 2

Regions showing significant changes in voxel-mirrored homotopic connectivity (VMHC) between each pair of the three groups. The threshold of the resulting statistical map was a combination of $p < 0.005$ for a single voxel and a minimum cluster size of 297 mm^3 .

on neuroimaging data to extract patterns and clarify the patterns in patients with stroke and healthy controls (38, 39). The SVM was implemented using the Pattern Recognition for Neuroimaging Toolbox (PRoNTTo) software (<http://www.mlnl.cs.ucl.ac.uk/pronto>). Individual resting-state fMRI data were treated as points located in a high dimensional space defined by the VMHC values in the preprocessed images. The regions showing significant correlation were selected as masks. The mask was applied to each image to identify the VMHC values as a feature in the model. A binary SVM machine was used as the classifier in the present study (i.e., patients with stroke (pretreatment) vs. healthy controls). Due to our limited number of samples, a “leave-one-out” method was used during the cross-validation step (38). Much of machine learning theory rests on the assumption that the data are independently and identically distributed. For functional neuroimaging, the data are within-run correlations and hemodynamic effects, which often do not meet the above assumption. To apply machine learning theory to functional neuroimaging data, permutation testing is needed, which enables us to obtain p -values for the performance metrics. When the SVM algorithm was established, permutation tests (repeated 1,000 times) were used to evaluate the performance of the SVM model. Finally, we obtained the corresponding accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve.

Results

Behavioral data

No significant differences were observed in sex ($\chi^2 = 1.17$, $p = 0.28$) or age ($t = 0.24$, $p = 0.81$) between the patients and the

healthy controls. With a 1-month interval, the FMA score was significantly increased (pair T: $t = -10.077$, $p < 0.001$) from 84.3 ± 4.1 (first time point) to 91.5 ± 4.1 (second time point). The NDS scores also showed a significant decrease (pair T: $t = 13.0$, $p < 0.001$) from 23.3 ± 4.3 (first time point) to 14.2 ± 5.2 (second time point).

VMHC: Group differences

The patients with stroke at the first timepoint showed a significantly lower VMHC in some regions than the healthy controls, such as in the parahippocampus, precuneus, precentral gyrus, and paracentral lobule/SMA (refer to Figure 2 left panel and Table 2). No areas showed a significant increase in VMHC in patients with stroke at the first time point.

The patients with stroke at the second time point still showed a significantly lower VMHC in some regions than the healthy controls, such as the precuneus, precentral gyrus, middle frontal gyrus (MFG), and parahippocampus (refer to Figure 2 middle panel and Table 2). No areas exhibited a significant increase in VMHC in patients with stroke at the second time point.

In patients with stroke, a paired t -test to assess the difference between the two time points revealed that the VMHC values in the SMA and MFG significantly increased from the first to the second time point (refer to Figure 2 right panel and Table 2). No areas showed a significant decrease in VMHC in patients with stroke from the first to the second time point.

In the present study, we focused on the regions belonging to the DMN and motor network. Thus, the significant results from the above t -test, including the results regarding the bilateral precuneus, parahippocampus, MFG, SMA, and precentral gyrus,

TABLE 2 Significant group differences in VMHC.

Cluster location	Statistical values			Peak (MNI)		
	Cluster size	<i>t</i> -value	<i>x</i>	<i>y</i>	<i>z</i>	
Control > Patient_pre						
Hippocampus/Fusiform	25	4.52	±45	−27		−18
Parahippocampus	21	4.44	±27	−30		−15
Middle temporal gyrus	43	3.75	±57	−51		−6
Middle temporal gyrus	58	4.47	±51	−66		12
Precentral gyrus	124	4.36	±60	3		24
Postcentral gyrus		4.07	±51	−9		27
Precuneus	30	3.92	±12	−51		12
Postcentral gyrus	27	3.40	±42	−21		48
Superior parietal lobe	36	6.58	±51	−33		60
Paracentral lobule/SMA	11	3.86	±12	−12		78
Superior precuneus	18	5.30	±12	−66		42
Control > Patient_post						
Parahippocampus/Fusiform	20	4.01	±33	−54		−18
Precuneus	64	4.17	±12	−63		24
Inferior frontal gyrus	55	4.46	±57	21		12
Middle frontal gyrus	37	4.65	±42	48		24
Inferior parietal lobe	32	3.84	±66	−27		30
Superior parietal lobe	208	5.37	±51	−33		60
Postcentral gyrus		4.46	±39	−24		51
Precentral gyrus		4.46	±30	−15		60
Precentral gyrus	13	3.55	±48	−6		42
Patient_pre < Patient_post						
SMA	14	5.51	±12	−9		78
Middle frontal gyrus	15	5.48	±23	37		−4

The MNI coordinates and *t*-values for the local maxima of the centers of the voxel clusters. The threshold for significant clusters reported here was set at $p < 0.005$ and a cluster size of 297 mm³.

VMHC, voxel-mirrored homotopic connectivity; MNI, Montreal Neurological Institute, SMA, supplementary motor area.

were selected as regions of interest in the following analyses. If the above regions of interest were identified in at least two comparison results, we selected the overlap regions as our areas of interest. The mean VMHC values were extracted from each region of interest. Although the disrupted interhemispheric communication measured by VMHC in patients was increased in these interesting regions from the first to the second timepoint, the VMHC in these regions except for the superior precuneus of the patients after treatment, was still significantly lower than that of the controls (Figure 3). Regarding the interesting regions of the superior precuneus, a significant increase in VMHC values was observed in patients with stroke from the first to the second time point. The mean VMHC values in the superior precuneus of the patients at the second time point showed no significant difference from the values in the controls. This result is consistent with the VMHC changes in voxels, as listed in Table 2. The superior precuneus no longer showed a significant difference between the controls and the patients at

the second time point. By combining the data regarding the consistent changes in the superior precuneus shown in Table 2 and Figure 3, we selected this region as the key region in the following correlation analysis and SVM analysis.

Brain–behavior relationships: Correlations between VMHC and FMA or MBI

To assess whether the VMHC values in the superior precuneus could reflect the motor ability and neurological deficit of patients with stroke, we further explored the relationship between the change in the behavioral scale and the change in the mean VMHC in this region from the first to the second time point. The same correlation processes were performed on the other interesting regions, including the parahippocampus, MFG,

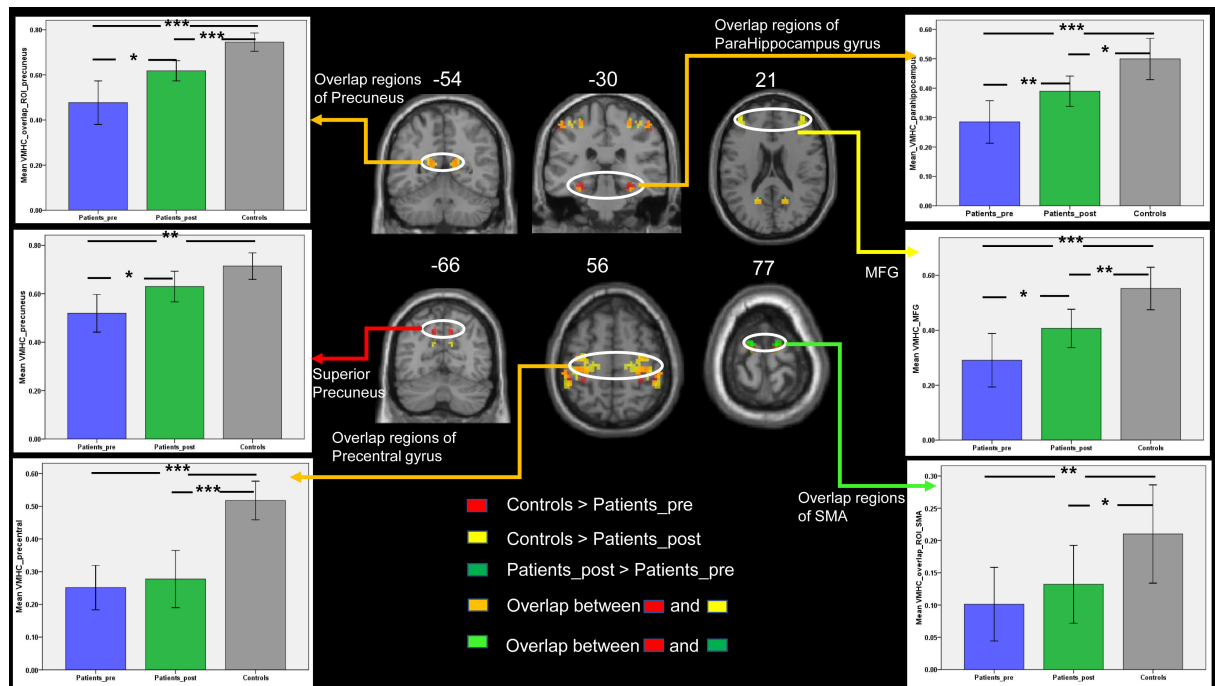


FIGURE 3

The comparisons of VMHC between the three groups in the DMN and motor-related regions of interest. Bars represent the mean and error bars represent the standard deviations. SMA, supplementary motor area; MFG, middle frontal gyrus. Patients_pre = the patient group in the pretreatment stage. Patients_post = the patient group in the posttreatment stage. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

precuneus, precentral gyrus, and SMA. The same correlation process was performed six times (six interesting regions) for each behavioral score. The corrected p -value of Bonferroni correction on the correlation result was set at 0.0083 (0.05/6). A partial correlation between the changes in VMHC values in the superior precuneus and the change in FMA scores showed a significant correlation in the patient group ($r = 0.55$, $p = 0.03$, refer to Figure 4A). Moreover, the changes in VMHC values in the superior precuneus showed a significant partial correlation with the change in NDS scores ($r = 0.72$, $p = 0.002$, refer to Figure 4B). The changes in VMHC values in the SMA also showed a significant partial correlation with the change of NDS scores ($r = -0.53$, $p = 0.04$, refer to Figure 4C). Analyses were controlled for age, sex, disease duration, and lesion sizes. No other significant correlations were detected in the present study. Only the correlation between the change in VMHC values in the superior precuneus and the change in NDS scores could withstand the Bonferroni correction ($p < 0.0083$).

SVM classification

The result of the SVM classification between 19 patients with stroke and 13 healthy controls based on the feature of VMHC in the superior precuneus derived from resting-state fMRI can be seen in Figure 5. This SVM classifier achieved an accuracy of 81.25% ($p < 0.003$). The model exhibited

a sensitivity of 84.21% and specificity of 76.92%. The area under the receiver operating characteristic curve (AUC) value was 0.84. We also used the binary SVM machine based on the VMHC features in the superior precuneus to classify patients with stroke (posttreatment) and healthy controls. The analysis of SVM classification achieved an accuracy of 31.25%, which was statistically significant at $p < 0.953$ (refer to Supplementary Figure 2).

Discussion

The present study investigated the longitudinal changes in VMHC in the DMN and motor-related regions of patients with subcortical stroke between two time points with a 1-month interval. Patients with stroke showed a significantly lower VMHC in the precuneus, parahippocampus, MFG, SMA, and precentral gyrus than the healthy controls. Although the impaired VMHC in these regions was enhanced after 1 month of treatment, the VMHC values in all but the superior precuneus were still significantly lower than those in the healthy controls. The decreased VMHC values in the superior precuneus of the patients were enhanced significantly at the second time point and showed no significant difference from those in the healthy controls. Furthermore, between the two time points, the changes in VMHC in the superior precuneus region were significantly related to the changes in clinical scores in patients with stroke.

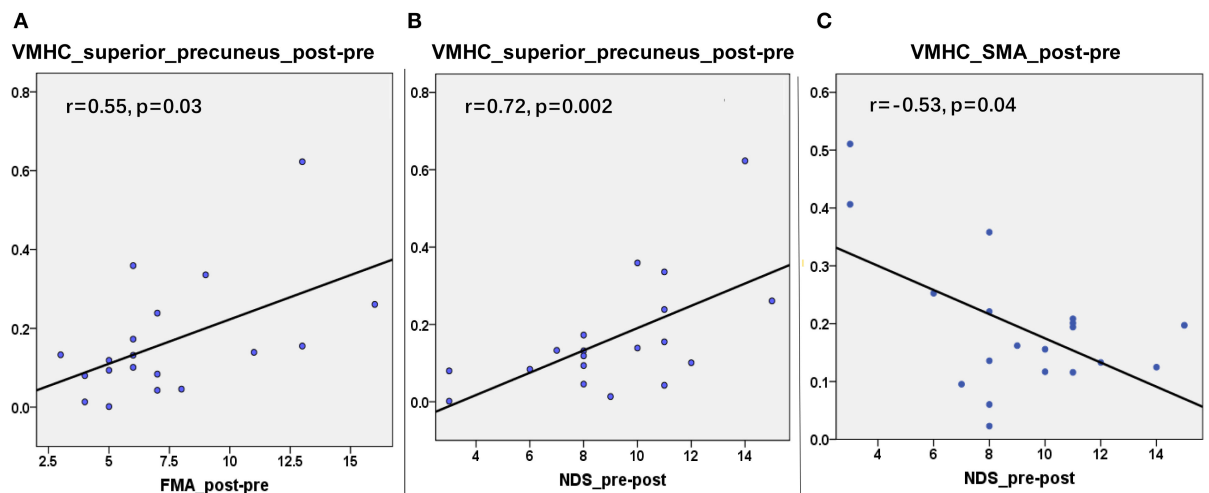


FIGURE 4

The correlations between the changes in VMHC and the changes in clinical variables from the first timepoint and the second timepoint in the superior precuneus of stroke patients. (A) The correlation ($r = 0.55$, $p = 0.03$) between the changes of VMHC in superior precuneus and the changes of FMA. (B) The correlation ($r = 0.72$, $p = 0.002$) between the changes of VMHC in superior precuneus and the changes of NDS. (C) The correlation ($r = -0.53$, $p = 0.04$) between the changes of VMHC in SMA and the changes of NDS.

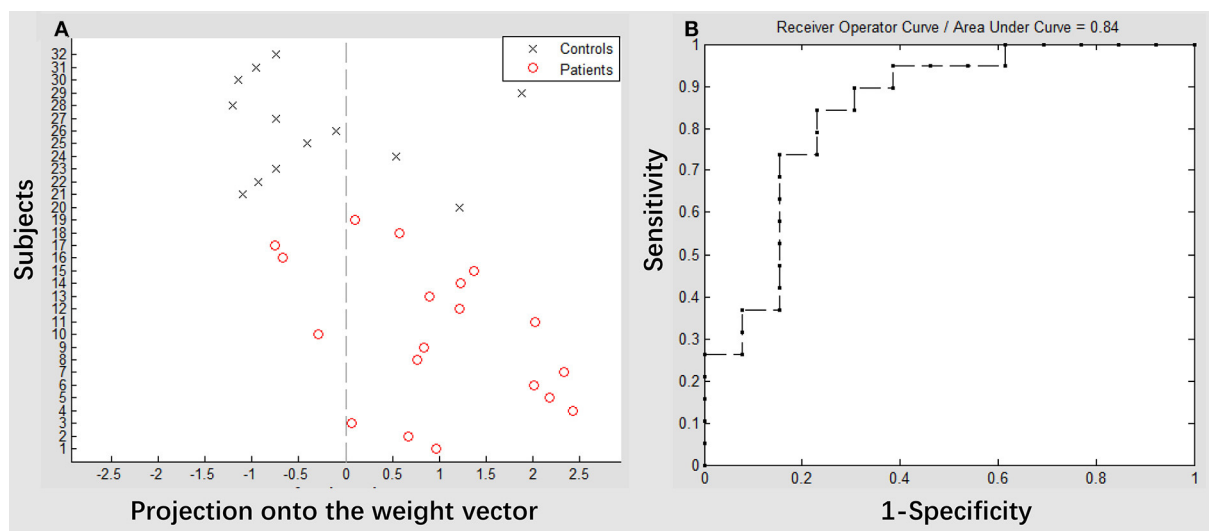


FIGURE 5

Classification plot (A) and receiver operating characteristic (ROC) curve (B) for the comparison between groups. Voxel-mirrored homotopic connectivity maps of 19 patients with stroke (pretreatment) and 13 healthy controls were used for classification, which yielded an accuracy of 81.25% (84.21% sensitivity, 76.92% specificity), which was statistically significant at $p < 0.003$.

The SVM analysis results demonstrated that the VMHC values in the superior precuneus exhibited discriminative power in the classification of patients with stroke from healthy controls. The recovery of functional homotopy and the discriminative power in the classification of the superior precuneus belonging to the DMN may provide novel evidence for understanding the neural mechanisms responsible for brain reorganization in stroke.

Significant changes in VMHC in patients with stroke

In the present study, patients with stroke showed a significant decrease in VMHC in the precuneus, parahippocampus, and MFG, which are part of the DMN. The DMN is one of the most widely studied functional brain networks at rest (40). This network shows reduced activity

during tasks and presents higher levels of activity than other networks during rest periods. The DMN plays an important role in “resting” brain activity, which is involved in emotional control, self-consciousness, and sustaining attention (41). Previous resting-state functional studies in stroke showed that DMN disruption is a common observation (20, 42, 43). Patients with stroke showed lower activity in the DMN and lower functional connectivity between the precuneus and other regions than healthy controls (7, 9, 44). The present study found decreased functional homotopy in the DMN in patients with stroke, which is similar to these previous fMRI studies. Reductions in DMN activity and connectivity can disrupt the cognitive processes mediated by the associated network. A previous study showed that the coordination between two hemispheres plays a significant role in human behavior (45). The decreased VMHC in the DMN that was observed in the present study indicated that disrupted interhemispheric connectivity may have some correlations with the occurrence of cognitive impairment in patients with stroke. The precuneus is the posterior region of the superior parietal lobe. This region plays an important role in the implementation of higher-order brain functions. Disruption of the precuneus can be considered an early sign of cognitive impairment (8). The precuneus has also been proposed as a structural and functional hub in the brain connectome. A previous study also used the VMHC method in patients with stroke and observed that patients showed significantly decreased VMHC values in the bilateral precuneus and precentral gyrus (10). Thus, the decreased functional homotopy between the bilateral precuneus observed in the present study was consistent with the results of a previous study, and this result can be regarded as the neuroimaging manifestation of cognitive disturbance in stroke.

Numerous studies on stroke have also reported disruptions in FC in the cortical motor-related network after stroke. For instance, previous studies have demonstrated that stroke can cause significant disturbances in the effective connectivity of motor areas (5, 14, 18). FC between the bilateral primary sensorimotor cortex was significantly decreased in patients with stroke (17, 46). Using diffusion tensor imaging (DTI), patients with stroke also demonstrated a decrease in the interhemispheric fiber connections between the left and right motor cortex (33, 47, 48). The DTI findings reflect anatomical disconnection, which may underlie the impaired interhemispheric resting-state FC. In the present study, patients with stroke showed significantly lower VMHC in the SMA and precentral gyrus than healthy controls. This result is consistent with previous neuroimaging studies on stroke showing that cortical motor connectivity can reflect poststroke sensorimotor signal processing (5, 49, 50). The SMA has strong anatomical connections with the areas of the central nervous system, including the thalamus, spinal cord, dorsal premotor cortex, and contralateral hemisphere regions (51, 52). In stroke patients with subcortical injuries, the functional ability of the brain enabled by exchanges and

cooperation between the two sides of the hemisphere is affected, and motor function is also influenced (49). Thus, it is easy to understand the result that the intrinsic neural interhemispheric coupling of bilateral SMAs was significantly reduced in the patient group. The precentral gyrus and anterior paracentral lobules together form the first somatic movement area, which is the main functional area governing the contralateral somatic movement. A previous DTI study tracing the fibers originating from the precentral gyrus in chronic stroke found that the integrity of all motor tracts showed a descending trend in patients (53). The motor tract damage from the precentral gyrus of patients with stroke can explain the significant decrease of VMHC in the precentral gyrus. Additionally, a consistent result of previous VMHC studies on stroke is that patients with stroke showed a significant decrease in VMHC values between the bilateral precentral gyrus (10, 28, 31). This result observed in the present study is consistent with those of these previous studies. However, the precentral gyrus is the main functional area of contralateral limb movement. The somatomotor center is located in the precentral gyrus, which is the higher center that controls the movement of the body. This region is responsible for regulating and connecting the lower motor center located in the brainstem and spinal cord. Therefore, limb movement disorders after stroke may result from disrupted connectivity of the precentral gyrus.

Longitudinal changes in VMHC between two time points with a 1-month interval

The most substantial discovery observed in the present research was the longitudinal changes in VMHC in the DMN and motor-related regions of patients with stroke. The functional couplings between the bilateral DMN and motor-related regions were improved. These results are consistent with previous studies showing that the intervention effects of brain networks in patients with stroke were mainly expressed as restorations in the connectivity pattern of interhemispheric interactions and the recovery of cognitive functions (13, 18, 23, 54). Stroke has been reported to be associated with FC impairment within DMN. Significant decreases in FC among the medial frontal cortex, PCC, and precuneus were found in patients with stroke (9, 42). A previous study showed that significantly reduced FC could be detected in the left precuneus, right SMA, and right superior frontal gyrus in patients with basal ganglia stroke (6). In the present study, most patients with stroke had a lesion in the basal ganglia. Thus, aberrant VMHC between the bilateral hubs within the DMN observed in our results can provide information for imaging diagnosis and early intervention. Additionally, the decreased VMHC in the regions of the DMN was enhanced significantly in all patients with stroke after 1 month of

treatment. The explanation of this result should combine the functional role of these regions of the DMN in stroke. The precuneus is the intermediate connector between the two hemispheres. Previous studies have revealed that the FC of the PCC/precuneus was increased significantly after a period of rehabilitation treatments or acupuncture treatment (16, 55). Thus, the increased VMHC between the bilateral hemisphere regions within the DMN observed in our study could be interpreted as the effects of the enhancement of brain function recovery.

In addition to the DMN, the VMHC in the SMA and precentral gyrus showed an increase after treatment. All of these regions are involved in motor function. A previous study reported an increase in coupling between the SMA, primary motor cortex, and premotor cortex following rehabilitation (56). These regions may play a critical role in motor recovery. In a resting-state fMRI study of patients with stroke with 3 weeks of upper limb rehabilitation therapy, Jame (57) found a stronger influence of the ipsilesional dorsal premotor cortex on its contralesional homolog. A previous study found that patients with stroke showed a significant increase in VMHC values between the bilateral premotor area and SMA after scalp acupuncture treatment (17). Enhanced interhemispheric communication was associated with improvements in motor performance. One of our previous multimodal neuroimaging studies also revealed that the impaired interhemisphere FC was restored after treatment (5). Patients with stroke also exhibited recovery of the impaired structural connectivity of the corpus callosum following the treatment intervention. In the present study, all patients were treated with 1-month antiplatelet therapy. Patients' clinical scores of FMA and NDS were changed significantly after treatment. The present neuroimaging findings were inline with the behavioral results. Enhanced interhemispheric communication of the SMA and precentral gyrus after treatment in patients with stroke may be the neuroimaging expression of their motor function recovery. Such increased interhemispheric coupling in patients with stroke might be induced by growth-related neurobiological processes enabling the formation of new synapses (58). Based on these results, we can see that the recovery of motor function may depend on reorganization processes within both hemispheres leading to enhanced interhemispheric connectivity (25).

Although the levels of VMHC in the DMN and motor-related regions were increased after antiplatelet therapy in patients with stroke, they were still significantly lower than those in healthy controls. A previous study with a longitudinal design revealed that the neural activity and FC of motor-related areas in patients with stroke returned to normal levels over 1 year after stroke onset (3). The recovery process of the inter-region coupling in patients with subcortical stroke demonstrated a dynamic change following long-term observation (46). The restored VMHC results observed in the present study implied

that antiplatelet therapy is efficient in the rehabilitation process of patients with stroke. However, the treatment cycle of 1 month is too short to achieve full recovery of functional connections between the bilateral hemispheres in patients with stroke. Regarding the superior precuneus, this region has reciprocal cortico-cortical connections with the adjacent areas of the posteromedial cortex. This interconnection is bilateral and bridges the homologous components of the two hemispheres (59). Because this region plays a central role in a wide spectrum of higher-order brain functions, increased activation of the precuneus could stimulate the motor cortex, boosting neuroplasticity for motor functional recovery. As a result, motor function in patients with stroke was enhanced and the severity of neurological functional deficits was decreased after 1 month of treatment in the present study. The superior precuneus may be more sensitive to the recovery of brain functions in patients with stroke. Thus, the VMHC in this region was enhanced significantly after treatment and showed no significant difference from that in the healthy controls. The synchrony of spontaneous activity in the bilateral superior precuneus may reflect the treatment effect. Future studies should be designed with multiple time points to determine whether the impaired brain functional connectivity after stroke can be restored to the normal level with treatment intervention.

Correlations between VMHC and clinical scores after treatment

The primary motor impairments after stroke are spasticity and spastic paresis, which impose substantial challenges to treatment and patient care (60). Manual function dexterity can be described as the ability to perform precise and coordinated hand and finger movements. Accurate evaluation of clinical manual motor function is, therefore, essential for assessing the recovery of motor ability in patients with stroke after treatment. As commonly used clinical recovery assessment scales, the FMA and NDS enable a good overall assessment of arm motor function, the ability to perform simple grasping tasks, and the severity of neurological functional deficits. In the present study, the patients with high changes in FMA or NDS scores between the two time points exhibited motor ability and hand control function that were well-preserved. The severity of stroke damage was reduced after 1 month of treatment. The neurological functional deficits showed recovery. Correspondingly, the patients with high increases in VMHC values in the superior precuneus exhibited enhanced interhemispheric connectivity. The disrupted interhemispheric coupling of the DMN was restored. The precuneus has been suggested to be associated with cognitive impairment in stroke. Significantly enhanced neural interhemispheric

coupling in the superior precuneus may indicate that the cognitive impairment of patients was reduced at the second time point. These findings are in accordance with the results of previous studies showing that activation in the precuneus can reflect an increase in functional coupling between bilateral somatosensory areas in patients with stroke (61). The significantly decreased FC in the PCC/precuneus was related to the Montreal Cognitive Assessment scores 10 days after the stroke (8). Spontaneous activity and FC of the PCC/precuneus were correlated with cognitive decline in patients with stroke (16, 42). Precuneus interhemispheric white matter integrity had a positive correlation with haptic performance in normal control participants (62). Combined with these previous studies, the significant correlation results observed in the present study confirmed that the level of coupling between the bilateral superior precuneus can be regarded as a neuroimaging biomarker to reflect the level of recovery of brain function in stroke patients with intervention.

Functional homotopy in the superior precuneus for the classification

The machine learning results also confirmed that the functional homotopy of the superior precuneus plays an important role in the discrimination of patients with stroke from healthy controls. This finding is important because it demonstrates the specific VMHC changes that occur in patients with stroke. Significantly decreased VMHC values in the important hub of the DMN are generally regarded as a reflection of injury to the brain functional network. A previous study on unilateral ischemic stroke demonstrated that the interhemispheric balance returned to healthy control levels in stroke patients with the successful recovery of dexterous hand function (63). The patient with poor recovery exhibited cerebral blood flow that was lateralized to the contralesional hemisphere, including the regions of SMA, paralimbic anterior cingulate cortex, and superior precuneus. The role of the precuneus cortex in sensorimotor transformations may have some association with goal-directed movements. Another previous study found that precuneus interhemispheric tract integrity can be regarded as a strong predictor of haptic performance (62). Precuneus interhemispheric tracts can be regarded as an appropriate target for piloting rehabilitation to improve poststroke haptic performance. During sensory discrimination, patients with stroke exhibited significantly different cortical activation than controls only in the precuneus (61). The activation in the precuneus may reflect an increase in functional coupling of bilateral somatosensory areas. In the present study, using functional homotopy values in the superior precuneus as a feature achieved high accuracy in the discrimination of

stroke participants from healthy controls. Additionally, the correlation results revealed there was a significant relationship between the changes in behavioral performance and the VMHC values in the superior precuneus. Consistent with previous reports and our correlation results, this machine learning result again confirmed the important role of the precuneus in the regulation of brain function in patients with stroke. The VMHC values in superior precuneus can also be considered neuroimaging-based biomarkers for cortical organization in stroke.

The current study has some potential limitations. First, the number of subjects was small, and it is necessary to investigate this topic in a larger sample size. In the present study, we attempted to reduce the effect of a simple size on the results by controlling for the lesion location. This attempt can partially mitigate the influence of sample size. Second, the medication was taken by patients with stroke without placebo control, which may affect the credibility of the present results. Future studies should consider this factor and improve the experimental design.

Conclusion

In conclusion, the results of the current study demonstrated that restored VMHC of the superior precuneus in patients with stroke can be detected following 1 month of antiplatelet therapy. Importantly, the significant correlations between the VMHC values and the clinical scores in patients with stroke showed that functional homotopy in the superior precuneus can reflect the recovery level of stroke patients' brain function with intervention. The present results have the potential to elucidate the recovery mechanisms of stroke.

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 the Ethics Committee of Chengdu University of Traditional Chinese Medicine. The patients/participants provided their written informed consent to participate in this study. 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

Conceived and designed the experiments: YL and PW. Performed the experiments: ZY and PW. Analyzed the data: YL and XZ. Responsible for patient management and conceptualized the study: JC and PW. Wrote and revised the paper: YL. All authors contributed to the article and approved the submitted version.

Funding

This research received funding from the National Natural Science Foundation of China (81072864 and 81601483) as well as support from the Medical Science and Technology Research Foundation of Guangdong Province (A2021076). This study was also supported by the Guang Zhou Science and Technology Project (202201011812), the Administration of Traditional Chinese Medicine of Guangdong Province (20221099), the Key-Area Research and Development Program of Guangdong Province (No. 2020B1111100001), and Huang Zhendong Research Fund for Traditional Chinese Medicine of Jinan University.

Acknowledgments

The authors are grateful to the patients who participated in the study. The authors also thank doctor Qie-zhu Wu in the West China Hospital MRI Center for his cooperation in data collection and analysis.

References

- Gottesman RF, Hillis AE. Predictors and assessment of cognitive dysfunction resulting from ischaemic stroke. *Lancet Neurol.* (2010) 9:895–905. doi: 10.1016/S1474-4422(10)70164-2
- Rehme AK, Eickhoff SB, Rottschy C, Fink GR, Grefkes C. Activation likelihood estimation meta-analysis of motor-related neural activity after stroke. *Neuroimage.* (2012) 59:2771–82. doi: 10.1016/j.neuroimage.2011.10.023
- Xu H, Qin W, Chen H, Jiang L, Li K, Yu C, et al. Contribution of the resting-state functional connectivity of the contralesional primary sensorimotor cortex to motor recovery after subcortical stroke. *PLoS One.* (2014) 9:e84729. doi: 10.1371/journal.pone.0084729
- Dacosta-Aguayo R, Graña M, Iturria-Medina Y, Fernández-Andújar M, López-Cancio E, Cáceres C, et al. Impairment of functional integration of the default mode network correlates with cognitive outcome at three months after stroke. *Hum Brain Mapp.* (2015) 36:577–90. doi: 10.1002/hbm.22648
- Li Y, Wang D, Zhang H, Wang Y, Wu P, Zhang H, et al. Changes of brain connectivity in the primary motor cortex after subcortical stroke: a multimodal magnetic resonance imaging study. *Medicine.* (2016) 95:2579. doi: 10.1097/MD.0000000000002579
- Chen HY, Shi MY, Zhang H, Zhang YD, Geng W, Jiang L, et al. Different patterns of functional connectivity alterations within the default-mode network and sensorimotor network in basal ganglia and pontine stroke. *Medical Science Monitor.* (2019) 25:9585–93. doi: 10.12659/MSM.918185
- Tuladhar AM, Snaphaan L, Shumskaya E, Rijpkema M, Fernandez G, Norris DG, et al. Default mode network connectivity in stroke patients. *PLoS ONE.* (2013) 8:e66556. doi: 10.1371/journal.pone.0066556
- Ding X, Li CY, Wang QS, Du FZ, Ke ZW, Peng F, et al. Patterns in default-mode network connectivity for determining outcomes in cognitive function in acute stroke patients. *Neuroscience.* (2014) 277:637–46. doi: 10.1016/j.neuroscience.2014.07.060
- Zhu T, Li L, Song Y, Han Y, Zhou C, Zhou D, et al. Altered functional connectivity within default mode network in patients with transient ischemic attack: a resting-state functional magnetic resonance imaging study. *Cerebrovasc Dis.* (2019) 48:61–9. doi: 10.1159/000502884
- Yao G, Li J, Liu S, Wang J, Cao X, Li X, et al. Alterations of functional connectivity in stroke patients with basal Ganglia damage and cognitive impairment. *Front Neurol.* (2020) 11:980. doi: 10.3389/fneur.2020.00980
- Gorelick PB, Scuteri A, Black SE, DeCarli C, Greenberg SM, Iadecola C, et al. Vascular contributions to cognitive impairment and dementia: a statement for healthcare professionals from the american heart association/american stroke association. *Stroke.* (2011) 42:2672–713. doi: 10.1161/STR.0b013e3182299496
- Bowden MG, Clark DJ, Kautz SA. Evaluation of abnormal synergy patterns poststroke: relationship of the Fugl-Meyer assessment to hemiparetic locomotion. *Neurorehabil Neural Repair.* (2010) 24:328–37. doi: 10.1177/1545968309343215

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2022.996621/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

The symmetrical mask of the default mode network (DMN) and motor network for the voxel-mirrored homotopic connectivity (VMHC) analysis.

SUPPLEMENTARY FIGURE 2

Classification plot (A) and receiver operating characteristic (ROC) curve (B) for the comparison between 19 patients with stroke (posttreatment) and 13 healthy controls. The analysis of SVM classification based on the feature of VMHC in the superior precuneus achieved an accuracy of 31.25% (42.86% sensitivity, 9.09% specificity), statistically significant at $p < 0.953$.

13. Bajaj S, Butler AJ, Drake D, Dhamala M. Functional organization and restoration of the brain motor-execution network after stroke and rehabilitation. *Front Hum Neurosci.* (2015) 9:173. doi: 10.3389/fnhum.2015.00173
14. Li Y, Wang Y, Zhang H, Wu P, Huang W. The effect of acupuncture on the motor function and white matter microstructure in ischemic stroke patients. *Evid-Based Complement Alternat Med.* (2015) 2015:164792. doi: 10.1155/2015/164792
15. Zheng X, Sun L, Yin D, Jia J, Zhao Z, Jiang Y, et al. The plasticity of intrinsic functional connectivity patterns associated with rehabilitation intervention in chronic stroke patients. *Neuroradiology.* (2016) 58:417–27. doi: 10.1007/s00234-016-1647-4
16. Wu CWW, Lin SHN, Hsu LM, Yeh SC, Guu SF, Lee SH, et al. Synchrony between default-mode and sensorimotor networks facilitates motor function in stroke rehabilitation: a pilot fMRI study. *Front Neurosci.* (2020) 14. doi: 10.3389/fnins.2020.00548
17. Liu H, Jiang Y, Wang N, Yan H, Chen L, Gao J, et al. Scalp acupuncture enhances local brain regions functional activities and functional connections between cerebral hemispheres in acute ischemic stroke patients. *Anat Rec.* (2021) 304:2538–51. doi: 10.1002/ar.24746
18. Grefkes C, Fink GR. Reorganization of cerebral networks after stroke: new insights from neuroimaging with connectivity approaches. *Brain.* (2011) 134:1264–76. doi: 10.1093/brain/awr033
19. Carter AR, Shulman GL, Corbetta M. Why use a connectivity-based approach to study stroke and recovery of function? *Neuroimage.* (2012) 62:2271–80. doi: 10.1016/j.neuroimage.2012.02.070
20. Wu P, Zeng F, Yx L, Bl Y, Lh Q, Qin W, et al. Changes of resting cerebral activities in subacute ischemic stroke patients. *Neural Regenerat Res.* (2015) 10:760. doi: 10.4103/1673-5374.156977
21. Griffis JC, Metcalf NV, Corbetta M, Shulman GL. Damage to the shortest structural paths between brain regions is associated with disruptions of resting-state functional connectivity after stroke. *Neuroimage.* (2020) 210:116589. doi: 10.1016/j.neuroimage.2020.116589
22. Li YX, Wang Y, Liao CX, Huang WH, Wu P. Longitudinal brain functional connectivity changes of the cortical motor-related network in subcortical stroke patients with acupuncture treatment. *Neural Plast.* (2017) 2017:5816263. doi: 10.1155/2017/5816263
23. Wang X, Wong WW, Sun R, Chu WCW, Tong KY. Differentiated effects of robot hand training with and without neural guidance on neuroplasticity patterns in chronic stroke. *Front Neurol.* (2018) 9:810. doi: 10.3389/fneur.2018.00810
24. Sinha AM, Nair VA, Prabhakaran V. Brain-computer interface training with functional electrical stimulation: facilitating changes in interhemispheric functional connectivity and motor outcomes post-stroke. *Front Neurosci.* (2021) 15:670953–670953. doi: 10.3389/fnins.2021.670953
25. Paul T, Hensel L, Rehme AK, Tscherpel C, Eickhoff SB, Fink GR, et al. Early motor network connectivity after stroke: an interplay of general reorganization and state-specific compensation. *Hum Brain Mapp.* (2021) 42:5230–43. doi: 10.1002/hbm.25612
26. Stark DE, Margulies DS, Shehzad ZE, Reiss P, Kelly AM, Uddin LQ, et al. Regional variation in interhemispheric coordination of intrinsic hemodynamic fluctuations. *J Neurosci.* (2008) 28:13754–64. doi: 10.1523/JNEUROSCI.4544-08.2008
27. Zuo XN, Kelly C, Di Martino A, Mennes M, Margulies DS, Bangaru S, et al. Growing together and growing apart: regional and sex differences in the lifespan developmental trajectories of functional homotopy. *J Neurosci.* (2010) 30:15034–43. doi: 10.1523/JNEUROSCI.2612-10.2010
28. Tang C, Zhao Z, Chen C, Zheng X, Sun F, Zhang X, et al. Decreased functional connectivity of homotopic brain regions in chronic stroke patients: a resting state fMRI study. *PLoS ONE.* (2016) 11:e0152875. doi: 10.1371/journal.pon.e0152875
29. Wei J, Wei S, Yang R, Yang L, Yin Q, Li H, et al. Voxel-mirrored homotopic connectivity of resting-state functional magnetic resonance imaging in blepharospasm. *Front Psychol.* (2018) 9:1620. doi: 10.3389/fpsyg.2018.01620
30. Dong ZZ, Zhu FY, Shi WQ, Shu YQ, Chen LL, Yuan Q, et al. Abnormalities of interhemispheric functional connectivity in individuals with acute eye pain: a resting-state fMRI study. *Int J Ophthalmol.* (2019) 12:634–9. doi: 10.18240/ijo.2019.04.18
31. Chen J, Sun D, Shi Y, Jin W, Wang Y, Xi Q, et al. Altered static and dynamic voxel-mirrored homotopic connectivity in subacute stroke patients: a resting-state fMRI study. *Brain Imaging Behav.* (2021) 15:389–400. doi: 10.1007/s11682-020-00266-x
32. Bhadelia RA, Price LL, Tedesco KL, Scott T, Qiu WQ, Patz S, et al. Diffusion tensor imaging, white matter lesions, the corpus callosum, and gait in the elderly. *Stroke.* (2009) 40:3816–20. doi: 10.1161/STROKEAHA.109.564765
33. Li Y, Wu P, Liang F, Huang W. The microstructural status of the corpus callosum is associated with the degree of motor function and neurological deficit in stroke patients. *PLoS ONE.* (2015) 10:e0122615. doi: 10.1371/journal.pone.0122615
34. Liu J, Qin W, Zhang J, Zhang X, Yu C. Enhanced interhemispheric functional connectivity compensates for anatomical connection damages in subcortical stroke. *Stroke.* (2015) 46:1045–51. doi: 10.1161/STROKEAHA.114.007044
35. Kalinsky BT, Berrios Barillas R, Schmit BD. Structurofunctional resting-state networks correlate with motor function in chronic stroke. *Neuroimage Clin.* (2017) 16:610–23. doi: 10.1016/j.nicl.2017.07.002
36. Gladstone DJ, Danells CJ, Black SE. The fugl-meyer assessment of motor recovery after stroke: a critical review of its measurement properties. *Neurorehabil Neural Repair.* (2002) 16:232–40. doi: 10.1177/154596802401105171
37. Chao-Gan Y, Yu-Feng Z. DPARSF: a MATLAB toolbox for “Pipeline” data analysis of resting-state fMRI. *Front Syst Neurosci.* (2010) 4:13. doi: 10.3389/fnsys.2010.00013
38. Li F, Huang X, Tang W, Yang Y, Li B, Kemp GJ, et al. Multivariate pattern analysis of DTI reveals differential white matter in individuals with obsessive-compulsive disorder. *Hum Brain Mapp.* (2014) 35:2643–51. doi: 10.1002/hbm.22357
39. Wu MJ, Mwambi B, Bauer IE, Passos IC, Sanches M, Zunta-Soares GB, et al. Identification and individualized prediction of clinical phenotypes in bipolar disorders using neurocognitive data, neuroimaging scans and machine learning. *NeuroImage Part B.* (2017) 145:254–64. doi: 10.1016/j.neuroimage.2016.02.016
40. Raichle ME, Macleod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL, et al. A default mode of brain function. *Proc Natl Acad Sci U S A.* (2001) 98:676–82. doi: 10.1073/pnas.98.2.676
41. Yuan C, Zhu H, Ren Z, Yuan M, Gao M, Zhang Y, et al. Precuneus-related regional and network functional deficits in social anxiety disorder: a resting-state functional MRI study. *Compr Psychiatry.* (2018) 82:22–9. doi: 10.1016/j.comppsy.2017.12.002
42. Liu J, Qin W, Wang H, Zhang J, Xue R, Zhang X, et al. Altered spontaneous activity in the default-mode network and cognitive decline in chronic subcortical stroke. *J Neurol Sci.* (2014) 347:193–8. doi: 10.1016/j.jns.2014.08.049
43. Wang C, Qin W, Zhang J, Tian T, Li Y, Meng L, et al. Altered functional organization within and between resting-state networks in chronic subcortical infarction. *J Cereb Blood Flow Metab.* (2014) 34:597–605. doi: 10.1038/jcbfm.2013.238
44. Zhao Z, Wu J, Fan M, Yin D, Tang C, Gong J, et al. Altered intra- and inter-network functional coupling of resting-state networks associated with motor dysfunction in stroke. *Hum Brain Mapp.* (2018) 39:3388–97. doi: 10.1002/hbm.24183
45. Luo C, Guo X, Song W, Zhao B, Cao B, Yang J, et al. Decreased resting-state interhemispheric functional connectivity in Parkinson's disease. *Biomed Res Int.* (2015) 2015:692684. doi: 10.1155/2015/692684
46. Park CH, Chang WH, Ohn SH, Kim ST, Bang OY, Pascual-Leone A, et al. Longitudinal changes of resting-state functional connectivity during motor recovery after stroke. *Stroke.* (2011) 42:1357–62. doi: 10.1161/STROKEAHA.110.596155
47. Wang LE, Tittgemeyer M, Imperati D, Diekhoff S, Ameli M, Fink GR, et al. Degeneration of corpus callosum and recovery of motor function after stroke: a multimodal magnetic resonance imaging study. *Hum Brain Mapp.* (2012) 33:2941–56. doi: 10.1002/hbm.21417
48. Li Y, Yu Z, Wu P, Chen J. The disrupted topological properties of structural networks showed recovery in ischemic stroke patients: a longitudinal design study. *BMC Neurosci.* (2021) 22:47. doi: 10.1186/s12868-021-00652-1
49. Zhang Y, Liu HL, Wang L, Yang J, Yan RB, Zhang JN, et al. Relationship between functional connectivity and motor function assessment in stroke patients with hemiplegia: a resting-state functional MRI study. *Neuroradiology.* (2016) 58:503–11. doi: 10.1007/s00234-016-1646-5
50. Lam TK, Dawson DR, Honjo K, Ross B, Binns MA, Stuss DT, et al. Neural coupling between contralateral motor and frontoparietal networks correlates with motor ability in individuals with chronic stroke. *J Neurol Sci.* (2018) 384:21–9. doi: 10.1016/j.jns.2017.11.007
51. Kato H, Izumiyama M. Activation of brain sensorimotor network by somatosensory input in patients with Hemiparetic stroke: a functional MRI study. *Novel Front Adv Neuroimag.* (2013) 10:51693. doi: 10.5772/51693
52. Dum RP, Levinthal DJ, Strick PL. Motor, cognitive, and affective areas of the cerebral cortex influence the adrenal medulla. *Proc Natl Acad Sci U S A.* (2016) 113:9922–7. doi: 10.1073/pnas.1605044113
53. Lindenberg R, Renga V, Zhu LL, Betzler F, Alsop D, Schlaug G, et al. Structural integrity of corticospinal motor fibers predicts motor impairment in chronic stroke. *Neurology.* (2010) 74:280–7. doi: 10.1212/WNL.0b013e3181ccc6d9

54. Ray AM, Figueiredo TDC, Lopez-Larraz E, Birbaumer N, Ramos-Murguialday A. Brain oscillatory activity as a biomarker of motor recovery in chronic stroke. *Hum Brain Mapp.* (2020) 41:1296–308. doi: 10.1002/hbm.24876
55. Zhang Y, Li K, Ren Y, Cui F, Xie Z, Shin JY, et al. Acupuncture modulates the functional connectivity of the default mode network in stroke patients. *Evidence-Based Complement Alternat Med.* (2014) 2014:765413. doi: 10.1155/2014/765413
56. Rehme AK, Eickhoff SB, Wang LE, Fink GR, Grefkes C. Dynamic causal modeling of cortical activity from the acute to the chronic stage after stroke. *Neuroimage.* (2011) 55:1147–58. doi: 10.1016/j.neuroimage.2011.01.014
57. Andrew James G, Lu ZL, VanMeter JW, Sathian K, Hu XP, Butler AJ. Changes in resting state effective connectivity in the motor network following rehabilitation of upper extremity poststroke paresis. *Topics Stroke Rehabil.* (2009) 16:270–81. doi: 10.1310/tsr1604-270
58. Rehme AK, Grefkes C. Cerebral network disorders after stroke: evidence from imaging-based connectivity analyses of active and resting brain states in humans. *J Physiol.* (2013) 591:17–31. doi: 10.1113/jphysiol.2012.243469
59. Cavanna AE, Trimble MR. The precuneus: a review of its functional anatomy and behavioural correlates. *Brain.* (2006) 129:564–83. doi: 10.1093/brain/awl004
60. Li S. Spasticity, motor recovery, and neural plasticity after stroke. *Front Neurol.* (2017) 8:120. doi: 10.3389/fneur.2017.00120
61. Borstad A, Schmalbrock P, Choi S, Nichols-Larsen DS. Neural correlates supporting sensory discrimination after left hemisphere stroke. *Brain Res.* (2012) 1460:78–87. doi: 10.1016/j.brainres.2012.03.060
62. Borstad AL, Choi S, Schmalbrock P, Nichols-Larsen DS. Frontoparietal white matter integrity predicts haptic performance in chronic stroke. *NeuroImage: Clin.* (2016) 10:129–39. doi: 10.1016/j.nicl.2015.11.007
63. Wiest R, Abela E, Missimer J, Schroth G, Hess CW, Sturzenegger M, et al. Interhemispheric cerebral blood flow balance during recovery of motor hand function after ischemic stroke—a longitudinal MRI study using arterial spin labeling perfusion. *PLoS ONE.* (2014) 9:e106327. doi: 10.1371/journal.pone.0106327



OPEN ACCESS

EDITED BY

Benjamin Yim,
Stanford University, United States

REVIEWED BY

Yang Song,
Óbuda University, Hungary
Santosh A. Helekar,
Houston Methodist Research Institute,
United States

*CORRESPONDENCE

Natan M. Bornstein
natanb@szmc.org.il

SPECIALTY SECTION

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

RECEIVED 27 July 2022

ACCEPTED 05 October 2022

PUBLISHED 14 November 2022


CITATION

Weisinger B, Pandey DP, Saver JL,
Hochberg A, Bitton A, Doniger GM,
Lifshitz A, Vardi O, Shohami E, Segal Y,
Reznik Balter S, Djemal Kay Y, Alter A,
Prasad A and Bornstein NM (2022)
Frequency-tuned electromagnetic
field therapy improves post-stroke
motor function: A pilot randomized
controlled trial.
Front. Neurol. 13:1004677.
doi: 10.3389/fneur.2022.1004677

COPYRIGHT

© 2022 Weisinger, Pandey, Saver,
Hochberg, Bitton, Doniger, Lifshitz,
Vardi, Shohami, Segal, Reznik Balter,
Djemal Kay, Alter, Prasad and
Bornstein. This is an open-access
article distributed under the terms of
the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution
or reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Frequency-tuned electromagnetic field therapy improves post-stroke motor function: A pilot randomized controlled trial

Batsheva Weisinger¹, Dharam P. Pandey², Jeffrey L. Saver³,
Arielle Hochberg¹, Adina Bitton¹, Glen M. Doniger¹,
Assaf Lifshitz¹, Ofir Vardi¹, Esther Shohami^{1,4}, Yaron Segal ¹,
Shira Reznik Balter¹, Yael Djemal Kay¹, Ariela Alter¹,
Atul Prasad⁵ and Natan M. Bornstein^{6*}

¹BrainQ Technologies, Ltd., Jerusalem, Israel, ²Manipal Hospital Physiotherapy and Rehabilitation, New Delhi, India, ³Department of Neurology, UCLA Comprehensive Stroke and Vascular Neurology Program, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, United States, ⁴Hebrew University of Jerusalem, Jerusalem, Israel, ⁵Department of Neurology, B. L. Kapur Super Specialty Hospital (BLK), National Capital Territory of Delhi, New Delhi, India, ⁶Brain Division, Shaare Zedek Medical Center, Jerusalem, Israel

Background and purpose: Impaired upper extremity (UE) motor function is a common disability after ischemic stroke. Exposure to extremely low frequency and low intensity electromagnetic fields (ELF-EMF) in a frequency-specific manner (Electromagnetic Network Targeting Field therapy; ENTF therapy) is a non-invasive method available to a wide range of patients that may enhance neuroplasticity, potentially facilitating motor recovery. This study seeks to quantify the benefit of the ENTF therapy on UE motor function in a subacute ischemic stroke population.

Methods: In a randomized, sham-controlled, double-blind trial, ischemic stroke patients in the subacute phase with moderately to severely impaired UE function were randomly allocated to active or sham treatment with a novel, non-invasive, brain computer interface-based, extremely low frequency and low intensity ENTF therapy (1–100 Hz, <1 G). Participants received 40 min of active ENTF or sham treatment 5 days/week for 8 weeks; ~three out of the five treatments were accompanied by 10 min of concurrent physical/occupational therapy. Primary efficacy outcome was improvement on the Fugl-Meyer Assessment – Upper Extremity (FMA-UE) from baseline to end of treatment (8 weeks).

Results: In the per protocol set (13 ENTF and 8 sham participants), mean age was 54.7 years (± 15.0), 19% were female, baseline FMA-UE score was 23.7 (± 11.0), and median time from stroke onset to first stimulation was 11 days (interquartile range (IQR) 8–15). Greater improvement on the FMA-UE from baseline to week 4 was seen with ENTF compared to sham stimulation, 23.2 ± 14.1 vs. 9.6 ± 9.0 , $p = 0.007$; baseline to week 8 improvement was 31.5 ± 10.7 vs. 23.1 ± 14.1 . Similar favorable effects at week 8 were observed for other UE and global disability assessments, including the Action Research Arm Test

(Pinch, 13.4 ± 5.6 vs. 5.3 ± 6.5 , $p = 0.008$), Box and Blocks Test (affected hand, 22.5 ± 12.4 vs. 8.5 ± 8.6 , $p < 0.0001$), and modified Rankin Scale (-2.5 ± 0.7 vs. -1.3 ± 0.7 , $p = 0.0005$). No treatment-related adverse events were reported.

Conclusions: ENTF stimulation in subacute ischemic stroke patients was associated with improved UE motor function and reduced overall disability, and results support its safe use in the indicated population. These results should be confirmed in larger multicenter studies.

Clinical trial registration: <https://clinicaltrials.gov/ct2/show/NCT04039178>, identifier: NCT04039178.

KEYWORDS

ischemic stroke, ELF-EMF, ENTF, neurostimulation, NIBS, magnetic field therapy, upper extremity motor function, neurorecovery

Introduction

Stroke is a leading cause of adult disability worldwide (1–3). While early reperfusion interventions improve outcomes (4), they are delivered to a small proportion of patients, leaving many stroke survivors with residual disabilities, impairments and dependency on others. This is accompanied by a large economic burden on both a personal and societal level, as a result of direct medical cost, as well as indirect costs due to underemployment and premature death (5).

Beyond the acute phase, standard of care focuses on rehabilitation through a coordinated effort of medical, social, educational, and vocational approaches to retrain an individual with newly acquired disabilities (6). Effective rehabilitation programs employ highly intensive and repetitive physical therapy to enhance neurologic recovery (7), possibly via direct influence on functional reorganization in the brain (i.e., plasticity). However, there is considerable variability among facilities in the implementation of therapeutic approaches that maximize functional recovery (8–11). Further, despite receiving standard rehabilitation care, many patients are left with lifelong disabilities and impairments, never returning to their pre-stroke ability level. For such people, one of the most common and persistent disabling symptoms is hemiparesis and upper limb motor impairment (12–14).

In the subacute post-stroke phase, and in response to both the initial, primary injury and the ensuing secondary injury cascade, the central nervous system (CNS) attempts to repair and reorganize itself via the secretion of survival-promoting agents (such as growth factors and anti-inflammatory cytokines), recovering damaged networks, and sprouting collateral synaptic connections to restore motor and cognitive functions (15–18). In preclinical models, neuroplasticity can be altered by external factors, including pharmacologic agents, electrical stimulation, and environmental stimulation (18).

Non-invasive brain stimulation (NIBS) techniques have demonstrated the capacity to enhance neuroplasticity in preclinical models, and have shown evidence suggestive of recovery in clinical trials (19–21). NIBS methods include repetitive transcranial magnetic stimulation (rTMS; (22)), transcranial direct current stimulation (tDCS; (23)), vagal nerve stimulation (24), and extremely low frequency, low intensity electromagnetic fields (ELF-EMF). While many such methods have been used with relative success, the limited applicability and strict usability requirements are such that none have yet to qualify as a standard of care treatment. ELF-EMF is a promising noninvasive therapeutic technique for post-stroke care, shown in preclinical studies to exert a beneficial effect on many of the cellular processes that modulate damage and recovery post-stroke, including calcium signaling, oxidative stress, and inflammatory response (25, 26). In a randomized clinical trial, ELF-EMF in the subacute post-stroke period was associated with increased enzymatic antioxidant activity, reduced oxidative stress, and improved performance on standardized assessments of activities of daily living, cognition, and mood (27, 28). Accordingly, and due to its relative safety and wide range of applicability, ELF-EMF treatment in the subacute phase may be a viable post-stroke therapy.

A novel ELF-EMF technique is a non-invasive, brain computer interface-based (BCI-based), low frequency, low intensity, frequency-tuned EMF therapy (Electromagnetic Network Targeting Field therapy; ENTF therapy), designed to expose impaired neuronal networks to oscillating fields similar to those of the CNS, in an effort to promote network reorganization post-injury. The human brain is organized into complex functional networks (29); healthy activity in the CNS results from the synchronization of thousands of individual neurons in the form of sophisticated and organized oscillations. These synchronous global oscillations within specific frequency bands represent functionally connected neural networks, which generate electrical activity measurable with electrophysiological

techniques, and are a fundamental part of the functionality of the brain (30, 31). These oscillatory patterns are correlated with cognitive states, motor functions, and electrophysiological activity both within and beyond the CNS. Changes in these patterns have been observed following ischemic stroke, and after other nervous system disorders like traumatic brain injury (31–33).

Neural network dynamics are sensitive to endogenous (34, 35) and exogenous electric and magnetic fields at specific frequencies (35–37), and oscillating ELF-EMF fields are hypothesized to promote the return of synchronization and network reorganization within the targeted networks. The motivation behind the present study (BQ3) is the possibility that ELF-EMF exposures can specifically target these impaired networks by exposing such networks to oscillating fields similar to those that characterize a healthy CNS, in order to promote network reorganization post-injury. The treatment protocol is based upon the most prominent frequencies of these motor-related oscillations, extracted using advanced machine learning algorithms from electrophysiological recordings of large populations of healthy and impaired individuals performing motor tasks (35).

In a preclinical rodent stroke model, oscillating ELF-EMF stimulation was associated with decreased edema, increased white matter integrity, and evidence of neural regeneration (38). Additionally, initial data from ongoing preclinical collaborative studies (unpublished) using this technique indicate changes in measures of oxidative stress, inflammation, and cell death. Overall, data suggest that such treatment targets cellular pathways comprising functional neural networks, promotes neural plasticity, and modulates the secondary injury cascade, all of which aid clinical recovery.

Accordingly, a pilot randomized, double-blind, sham-controlled trial of a BCI-based, low frequency, low intensity, frequency-tuned ENTF therapy to improve upper extremity motor function and reduce disability in subacute post-stroke patients was designed and executed. Greater improvement in upper extremity motor function is expected in individuals who receive ENTF treatment, as compared to those in the sham control group.

Materials and methods

Study design and participants

This study, the BQ3 trial (NCT04039178), was a prospective, randomized, double-blind, sham-controlled study. See [Supplementary Table S1](#) for full study entry criteria.

The trial was conducted at the BLK Super Specialty Hospital, New Delhi, India, a multi-specialty private hospital accredited by the Joint Commission International. Study operations were overseen by JSS Medical Research, an international, full-service

contract research organization. The hospital institutional review board provided ethics approval, and written informed consent was obtained from all participants.

The main inclusion criteria for this study were: patients 4–21 days post-ischemic stroke with first stroke or no prior upper extremity impairment, right hand dominant, with a Fugl-Meyer Assessment – Upper Extremity (FMA-UE) score between 10 and 45. Patients were also screened for their ability to participate in the treatment procedures based on their ability to be seated for 70 consecutive minutes, and follow a three-step command. Patients who were not medically stable, with a physiological, neurological, or psychiatric history that might confound study measures, or contraindications for MRI scanning were not considered for this study.

Randomization and blinding

The study was planned for 50 participants, the first four of whom would be assigned directly to the treatment group (run-in phase). The remaining 46 participants were to be randomly assigned to active ENTF or sham stimulation (randomized phase) in a 1:1 ratio (block randomization; SAS-generated), by an individual not otherwise associated with the study. After determining group allocation, the individual keyed in group assignment to the device (required only once per participant). Participants and study staff were not aware of group assignment. The device does not produce any perceptible light, sound, or sensation during the ELF-EMF activity; sham stimulation consists of the same general treatment flow, but with the wave generator inactive during the session, and as governed by the group assignment saved for each participant within the device. Thus, experience with the BQ device during sessions was the same irrespective of group assignment, allowing for proper blinding.

Materials

Treatment was administered with a proprietary BCI-based stimulation device (BQ 1.0; BrainQ Technologies Ltd., Jerusalem, Israel; product manual available upon request), exposing the entire brain and the cervical and upper thoracic portion of the spinal cord to the ENTF. The device technology uses machine learning algorithms (Python, 3.6) to identify high resolution spectral patterns that characterize motor functions within EEG measurements recorded during functional motor tasks. For this custom-made algorithm, EEG data from healthy and unhealthy individuals was collected while executing discreet motor tasks. A novel normalization technique was used to reduce inter-subject variability. Machine learning models were used to differentiate between healthy and unhealthy data traces. The explanatory features used by these models to generate

their decisions were then used to inform a non-invasive and frequency-specific, extremely low frequency (1–100 Hz), low intensity (<1 Gauss) electromagnetic field treatment applied to a participant's CNS, delivered via a magnetic coil. The device emitted ELF-EMF only for participants in the ENTF group but not for those in the sham group.

Procedure

ENTF or sham therapy was provided 5 days a week for ~8 weeks, for a total of 40 treatment sessions. During each treatment session, participants received 40 minutes of treatment with the BQ device (active or sham). During ~3 weekly sessions, concurrent with the ENTF or sham therapy, participants performed 10 minutes of upper extremity physical therapy/occupational therapy-based exercises (e.g., gripping a ball, reaching) with the guidance of a therapist. Separate from the treatment sessions, participants also received ~1 h of physical therapy per day throughout their participation as part of the hospital's standard clinical regimen.

Outcome measures

The primary clinical efficacy outcome was change in upper extremity motor function from baseline to end of treatment (week 8), measured with the FMA-UE, a performance-based impairment index designed to assess motor function, balance, sensation and joint function in patients with post-stroke hemiparesis (39, 40). FMA-UE was assessed throughout the course of treatment at baseline, week 4, week 8 and week 12. However, due to early trial closure because of the COVID-19 pandemic, follow-up assessments at week 12 were not completed for many participants (available data for <80%), so analyses included only changes at week 4 and week 8.

Secondary clinical efficacy outcomes were: Action Research Arm Test [ARAT, coordination, dexterity, and function (41)]; Box & Blocks Test [BBT, gross manual dexterity (42)]; Fugl-Meyer Assessment – Lower Extremity (FMA-LE) (39, 40); modified Rankin Scale (mRS) of global disability (43); National Institutes of Health Stroke Scale [NIHSS, stroke-related neurological deficit (44)]; Patient-Reported Outcome Measurement Information System Global 10 [PROMIS-10, patient-reported assessment of global health and quality of life (45)]. Notably, some prespecified outcome measures were not analyzed due to <80% valid data. These include: cognitive measures [Trail Making Test (46); Montreal Cognitive Assessment (47)], as they were administered in English, which was not most participants' primary language; imaging (MRI), because of variability in scan parameters due to use of multiple scanners; and blood biomarkers, as some growth factors were out of the detection range for many subjects (28, 48). EEG

was collected as an additional, exploratory endpoint, and was analyzed separately from the clinical results; these results are reported elsewhere (49).

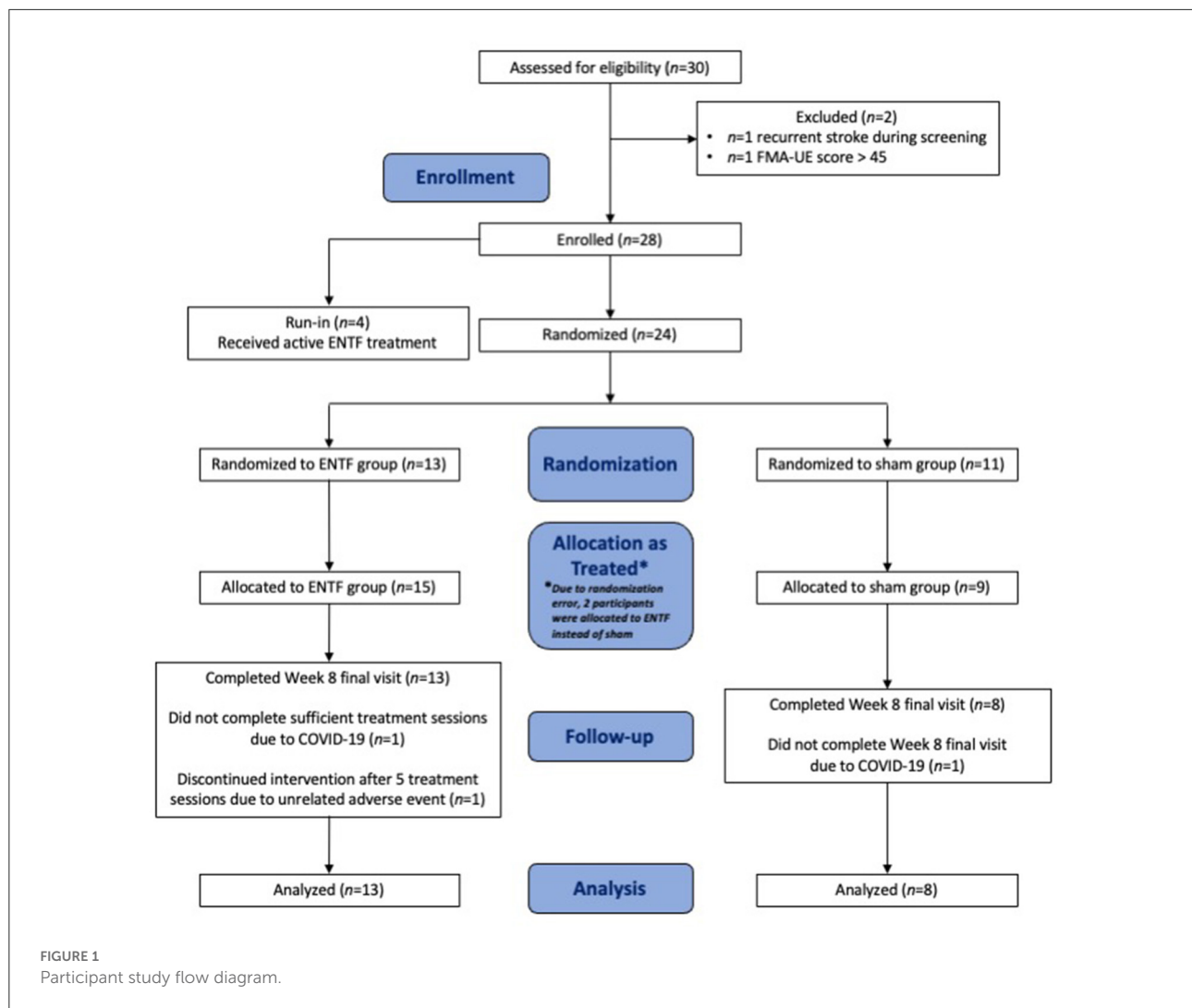
The primary safety outcome was adverse events during the trial period.

Statistical analyses

Statistical analyses were conducted with SAS V9.4 (SAS Institute, Cary NC, USA). For behavioral outcomes, continuous variables were generally summarized by mean and standard deviation (SD), and categorical variables by percentage. Changes from baseline in continuous outcomes were analyzed by analysis of variance (ANOVA) or repeated measures ANOVA (SAS PROC MIXED), with each outcome modeled as a function of treatment group; if >1 post-baseline visit, each outcome was also modeled as a function of visit, as well as the treatment group*visit interaction. Baseline value for the respective outcome was entered as a covariate to adjust for variation in baseline score between the groups. LSmeans (model estimated means) per group and the differences between the groups were estimated from the models, along with respective levels of significance. If >1 post-baseline visit, then the treatment group*visit interaction term was the parameter of interest, and the LSmeans per group, as well as the differences between the groups at each visit, were estimated from the models, along with respective levels of significance. Continuous demographic and baseline data were compared between the groups with a Wilcoxon two-sample test. Categorical variables were compared using Fisher's exact test when the response was binary, and with a Cochran Armitage trend test when the response was ranked (e.g., mRS). Nominal two-sided *p* values are presented without *post-hoc* testing to correct for multiple comparisons given that this was a pilot study.

As a pilot study, sample size justification was formulated based on the literature regarding the study's primary outcome, as well as the treatment characteristics of the target population (50). The sample size calculation assumed 80% power, a 5-point difference between groups on the FMA-UE, and a SD of 6, resulting in a study with 46 participants. The study was designed to include 46 randomized participants, preceded by 4 treatment-only run-in participants, for a total sample size of 50.

The prespecified primary analysis set for efficacy evaluation was the per protocol (PP) set, defined as all participants who were randomized and participated in 80% of the treatment visits, were not absent for more than 5 consecutive visits, and completed the week 8 outcome evaluation; the PP set consisted of 8 participants allocated to sham stimulation, and 13 to ENTF stimulation. Participant characteristics were evaluated in the intention to treat (ITT) analysis set, defined as all randomized participants who completed at least one treatment session (sham or active); the ITT set consisted of 9 sham and 15 ENTF



participants. The as treated (AT) safety set, defined as all participants (including the $n=4$ treatment-only run-ins) who received at least one treatment session, consisted of 9 sham and 19 ENTF participants.

Results

The study was conducted between first enrollment on December 11, 2018 and final study visit on March 21, 2020. As emergence of the COVID-19 pandemic precluded continuation of study operations, the study was discontinued early, and data was analyzed after enrollment of 4 run-in and 24 randomized participants. Participant study flow diagram is shown in Figure 1.

Participant baseline characteristics in the PP set are shown in Table 1. Participant characteristics were generally well balanced across the treatment groups. Participant baseline characteristics

in the ITT and the AT safety sets were similar to the PP set (Supplementary Tables S2, S3).

Efficacy outcomes

For FMA-UE (primary clinical efficacy outcome measure), performance score improvements were greater in the ENTF compared to the sham group throughout treatment, both at week 4 ($p = 0.007$) and week 8 (Table 2; Figure 2). In terms of absolute numerical values, a ceiling effect was noted at week 8. In the active stimulation group, 77% of participants attained scores in the top 10% of the scale, including 46% receiving the highest score (66/66). By comparison, only 38% of the sham group attained scores in the top 10% and no participants attained the highest score.

TABLE 1 Participant demographics*.

	Sham Group (<i>n</i> = 8)	ENTF Group (<i>n</i> = 13)	Total (<i>n</i> = 21)
Age, yrs, mean (\pm SD)	55.3 (\pm 10.1)	54.3 (\pm 17.8)	54.7 (\pm 15.0)
Sex, female (%)	25%	15%	19%
Race-Ethnicity, South-Asian (%)	100%	100%	100%
Hand dominance, right (%)	100%	100%	100%
Affected hand, right (%)	63%	38%	48%
Time from stroke onset to first treatment, days, median (IQR)	14.0 (10.8–16.0)	9.0 (7.0–14.0)	11.0 (8.0–15.0)
FMA-UE Baseline, mean (\pm SD)	18.8 (\pm 8.7)	26.8 (\pm 11.5)	23.7 (\pm 11.0)
mRS Baseline, mean (\pm SD)	3.4 (\pm 0.7)	3.6 (\pm 0.5)	3.5 (\pm 0.6)

*No significant differences were noted between groups at baseline.

Score changes for secondary efficacy outcomes related to motor function generally reflect greater improvement for ENTf compared with sham control, with limited exception (Table 2). Specifically, greater improvement for ENTf was found for the ARAT total score (Figure 3A) and three out of four of the ARAT subscales, most notably, the Pinch subscale ($p = 0.008$, Figure 3B). Significantly greater improvement was also seen on the BBT in the affected hand (week 6: $p = 0.02$; week 8: $p < 0.0001$, Figure 3C); the non-affected hand also showed greater improvement in the ENTf group. Significantly greater improvement was also seen in the FMA-LE scores changes (Table 2).

In addition, both mRS and NIHSS scores showed greater improvement for the ENTf group. On the mRS, the ENTf group showed significantly greater reduction in degree of disability between baseline and week 8; -2.5 ± 0.7 vs. -1.3 ± 0.5 , $p = 0.0005$. Notably, 92% of participants in the ENTf group improved by at least two points compared to only 25% in the sham group (Figure 4A). By week 8, 77% of participants in the ENTf group vs. 25% in the sham group had an mRS score of 1 or 0, indicative of little to no residual disability (Figure 4B). In contrast, patient-reported efficacy outcomes related to generic health-related quality of life (PROMIS-10) did not show a difference in degree of improvement between treatment groups. A descriptive table of raw scores for each of the analyzed clinical outcome measures at all assessment time points is included in Supplementary Table S4.

Safety outcomes

Two adverse events (AEs) were reported, neither related to the ENTf treatment. There were no device-related infections or unexpected device-related adverse events. Additionally, there were no complaints of discomfort during ENTf treatment.

Discussion

In this double-blind, randomized, sham-controlled trial, frequency-tuned ELF-EMF stimulation with the BQ device, or ENTf therapy, was associated with enhanced recovery of upper extremity function when initiated in the subacute period and continued for 2 months as compared to sham control. Beneficial results were evident not only for the primary outcome measure (FMA-UE), but also for several other measures of upper extremity function, including the BBT (manual dexterity) and the ARAT (coordination, dexterity, and function). Moreover, ELF-EMF treatment was associated with a greater reduction of global disability in daily activities (mRS). In addition, there was no evidence of safety concerns, and there were no participant complaints of discomfort during treatment.

There was an indication of beneficial effect on lower extremity function in addition to upper extremity function, suggesting a general enhancement of motor function beyond the upper extremity. In contrast, ENTf stimulation was not associated with benefit on a generic measure of mental and physical health-related quality of life. Indeed as the intervention in this study was specifically designed to target motor impairment, it cannot be assumed that generic physical and mental health measures would show differential benefits between the two groups. Still, it should be noted that these were subjective patient-reported ratings.

The magnitude of benefit of ENTf treatment was robust and clinically meaningful across multiple metrics of upper extremity motor function, and especially the FMA-UE (51, 52). Further, the reduction of global disability as assessed by the mRS supports a strong positive effect on overall functioning. The substantial difference in outcomes between active and sham-treated groups was not related to unusually poor performance in the control arm. The degree of improvement on the FMA-UE in the control group was typical of those in prior natural history studies (51, 52). Similarly, in terms of overall stroke disability, the degree of improvement on the mRS in the control group was similar

TABLE 2 Efficacy outcome measures*.

	Sham group (<i>n</i> = 8)	ENTF group (<i>n</i> = 13)	Significance**
Primary outcome measure			
FMA-UE Week 4	9.6 ± 9.0	23.2 ± 14.1	0.007
FMA-UE Week 8	23.1 ± 14.1	31.5 ± 10.7	0.06
Secondary outcome measures			
mRS Week 8	−1.3 ± 0.5	−2.5 ± 0.7	0.0005
ARAT Grasp, Week 8	8.1 ± 7.6	9.1 ± 6.8	0.15
ARAT Grip, Week 8	5.4 ± 4.1	6.8 ± 3.7	0.13
ARAT Pinch, Week 8	5.3 ± 6.5	13.4 ± 5.6	0.008
ARAT gross movement, Week 8	3.8 ± 2.1	2.8 ± 1.9	0.50
ARAT total score, Week 8	22.5 ± 17.1	32.1 ± 14.2	0.09
BBT (Affected hand), Week 2	0.5 ± 0.8	1.3 ± 1.7	0.07
BBT (Affected hand), Week 4	1.4 ± 1.5	6.8 ± 5.5	0.08
BBT (Affected hand), Week 6	3.3 ± 3.6	10.9 ± 7.1	0.02
BBT (Affected hand), Week 8	8.5 ± 8.6	22.5 ± 12.4	<0.0001
BBT (Non-affected hand), Week 2	2.9 ± 0.6	1.8 ± 2.0	0.04
BBT (Non-affected hand), Week 4	4.4 ± 2.1	9.0 ± 5.0	0.08
BBT (Non-affected hand), Week 6	7.8 ± 2.1	12.4 ± 5.4	0.08
BBT (Non-affected hand), Week 8	9.4 ± 3.9	18.4 ± 7.4	0.0003
FMA-LE, Week 8	9.9 ± 6.5	13.8 ± 7.8	0.03
NIHSS, Week 8	−4.8 ± 3.2	−6.6 ± 3.4	0.03
PROMIS-10	9.0 ± 5.2	11.0 ± 6.8	0.33
global physical health, Week 8			
PROMIS-10	13.0 ± 7.2	14.7 ± 6.1	0.48
global mental health, Week 8			

*All values represent change from baseline (mean ± SD).

**From LSmean adjusted means.

to that in control groups in prior large trials and observational studies (53–57).

EEG recordings (exploratory) of study participants reflected a pattern of brain activity indicative of recovery exclusively in the ENTf group (49). More specifically, the EEG results are consistent with improvement in movement inhibition or motor learning (58) as well as increased signal complexity, a characteristic of healthy brain activity (59). In effect, the EEG data provide evidence for a biomarker of recovery putatively linked to plasticity (59) in the ENTf group but not in the sham group.

At the neuronal level, given the continued degradation of neurons in the days and weeks following a stroke, as well as the secondary injury cascade whereby cells adjacent to the site of injury continue to degrade (60), a non-invasive treatment that targets affected cells and networks during this critical time period and prevents further degradation has great clinical utility, addressing a gap in subacute care options. The most challenging question regarding the effect of ELF-EMF on (neuronal) tissue is in identifying the transduction mechanism by which the applied field and the biological

medium interact, achieving such effects. Although the exact mechanism remains unknown, two different steps have to be taken into account when contemplating the mechanism of action of ELF-EMF: 1) the initial interaction between the external magnetic field (MF) and the biological system and 2) the cascade of biological events leading to the physiological/behavioral effect seen in this as well as other studies, both human and animal (61, 62).

Regarding the initial interaction step, there are two plausible transduction mechanisms: electric currents inducing minor changes in the conductive tissues (unlikely for intensities <1 G such as used in this study), and possible direct action of the MF on endogenous magnetoreception (63). As for the cascade of biological events that follow, it has been shown that ELF-EMF effects are likely to involve a number of cellular targets such as changes in intracellular Ca²⁺ signaling (64–69), elements of the oxidative stress cascade (70, 71), nitric oxide (72, 73), G-protein receptor coupling (74, 75), and the inflammatory response (76, 77), to name a few. Some of these cellular targets have been identified and described in the studies conducted previously by members of this group, as well as by others in

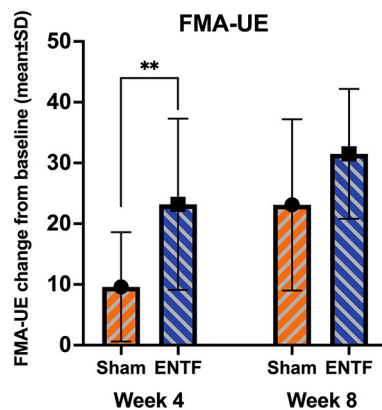


FIGURE 2

Fugl-Meyer Assessment – Upper Extremity (FMA-UE) score changes from baseline to week 4 and week 8 (range: 0–66; 66 is greatest mobility/optimal recovery). FMA-UE absolute score change from baseline to week 4 (sample mean, error bars correspond to SD; significance based on difference in LSmeans) was significantly greater for the ENTf group than sham group (23.2 ± 14.1 vs. 9.6 ± 9.0 ; $p = 0.007$; $** < 0.01$). Absolute score change from baseline to week 8 was also greater, though not significantly so, for the ENTf group than sham group (31.5 ± 10.7 vs. 23.1 ± 10.7 ; $p = 0.06$).

the field, revealing a candidate for the cascade of events that may ultimately give rise to the observed effects on the cellular, network and behavioral levels.

For example, in human neuroblastoma and rat pituitary cells, ELF-EMF exposure increases proliferation and inhibits programmed cell death by up-regulating the expression of voltage-gated Ca^{2+} (Ca_v) channels [5–1,000 μT and frequencies of 1–100 Hz, (65)]. Additionally, it has been shown that ELF-EMF increases generation and metabolism of nitric oxide (NO) in poststroke patients, promoting cellular processes that support neuroplasticity, and thus may enhance post-stroke rehabilitation (27). Furthermore, ELF-EMF exposure (50 Hz, 1 mT, 1 to 7 h/day for 7 days) significantly enhanced neurogenesis in the dentate gyrus (DG) of adult mice, as demonstrated by increased numbers of cells double-labeled for BrdU and doublecortin (78).

Converging evidence has been obtained from within our own rodent stroke study. While no adverse effects (e.g., abnormal changes in body weight) were observed, results indicate that daily exposure to ENTf treatment (7.86 Hz, 1 G) over 8 weeks post-injury significantly improved the Neurological Severity Score (NSS) in the treatment group. Importantly, a significant increase in the number of BrdU positive cells was found in the dentate gyrus, in addition to the restoration of biomarkers indicative of healthy cortical tissue in the injured parietal cortex of ENTf-treated mice. These results further support the hypothesis that ENTf treatment may promote neurogenesis (38). Additionally, in rats with spinal cord injury, diffusion tensor imaging (DTI) revealed that

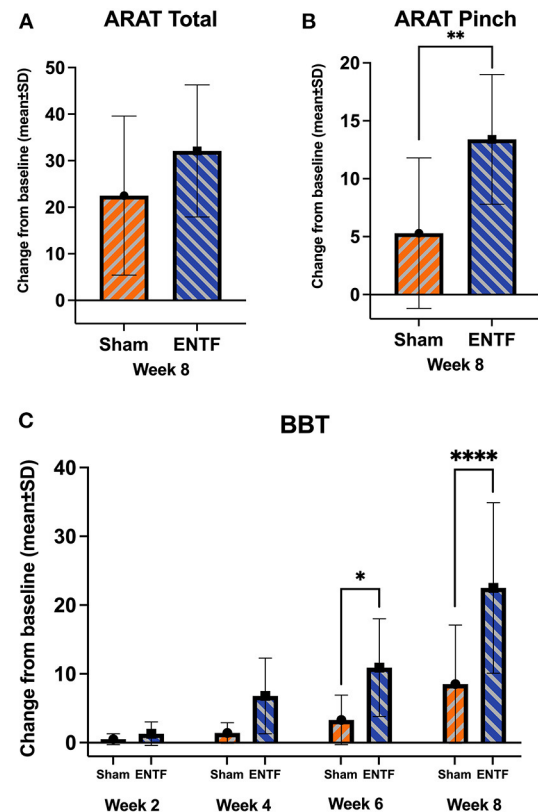


FIGURE 3

Evolution of upper extremity function secondary efficacy outcomes (sample mean; error bars correspond to SD; significance based on difference in LSmeans). (A) Action Research Arm Test (ARAT) total score from baseline to week 8 (absolute score change: ENTf group 32.1 ± 14.2 vs. sham group 22.5 ± 17.1 , $p = 0.09$). (B) ARAT Pinch subscale score from baseline to week 8 (absolute score change: ENTf 13.4 ± 5.6 vs. sham 5.3 ± 6.5 , $p = 0.008$; $** p < 0.01$). (C) Box and Blocks Test (BBT) affected hand scores from baseline through weeks 2, 4, 6 (absolute change score: ENTf 10.9 ± 7.1 vs. sham 3.3 ± 3.6 , $p = 0.02$; $* p < 0.05$) and 8 (absolute change score: ENTf 22.5 ± 12.4 vs. sham 8.5 ± 8.6 , $p < 0.0001$; $**** p < 0.0001$).

those receiving ENTf treatment showed evidence of structural neuroplasticity, compared to the spinal cord degradation observed in non-treated rats (61).

The substantial number of published studies clearly demonstrate effects from cellular to physiological, and consequently behavioral, suggesting a robust mechanism of action mediating the effect of ELF-EMF on the brain. The present results extend these findings to a clinical post-stroke population, and demonstrate the effectiveness of ENTf treatment in accelerating recovery in the subacute phase post-stroke.

Importantly, the present results provide useful data on the safety and feasibility of ENTf treatment as there were no

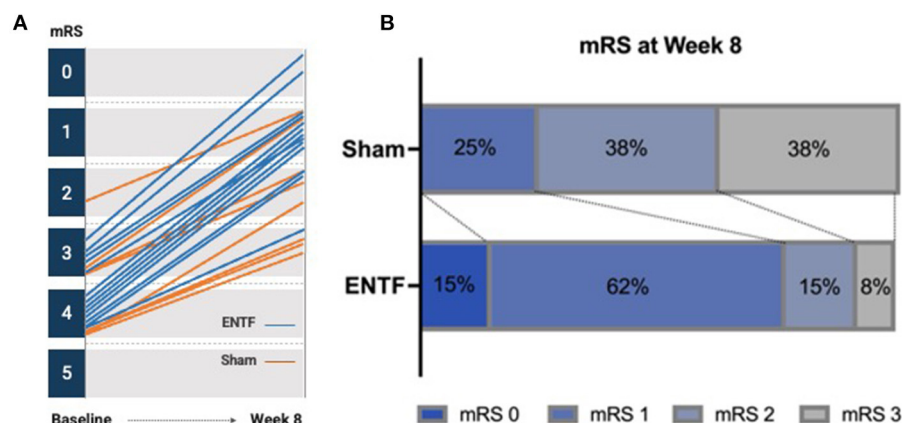


FIGURE 4
Change in modified Rankin Scale (mRS) from baseline to week 8. **(A)** Individual participant trajectories from baseline to week 8, yielding greater group reduction in disability severity in the ENTF compared to sham group, -2.5 ± 0.7 vs. -1.3 ± 0.5 , $p = 0.0005$. **(B)** Distribution of final mRS scores at week 8, with more favorable outcomes evident in ENTF vs. sham group.

safety concerns or complaints about comfort. Indeed, a non-invasive, user-friendly device with a favorable safety profile may be ideally suited for use after a patient returns home. Further, in the wake of the COVID-19 pandemic, efficacious treatment options that minimize in-hospital exposure are valuable for an older, vulnerable, post-stroke population. The ability to integrate such treatment into a care plan that is patient-centered and addresses the normally fragmented treatment pathway remains an important target of future studies.

This study has several limitations. First, though results are robust across multiple metrics, sample size was small. In addition, there were limited long-term follow-up evaluations to assess the continued effects of the treatment on recovery. COVID-19 restrictions forced a reduction of the planned sample size and follow-up duration, thus studies with larger samples and longer-term follow-up are needed. Second, the FMA-UE and ARAT measures appeared suboptimal for moderately impaired participants due to reasonable likelihood of reaching the maximum score (79). Indeed, the trend towards meaningful improvement of the FMA-UE ($p = 0.06$) coupled with an overwhelming majority of treated participants in the top 10% of the metric at end of treatment, it is likely that the benefit to the treatment group was not fully captured in this score. In comparison, there was continued improvement on BBT which has a greater responsive range. Third, the study was conducted at a single site in participants of one ethnicity. Multicenter trials in larger, more diverse populations are desirable.

In conclusion, this study demonstrates efficacy of extremely low frequency, low intensity frequency-tuned ENTF exposure in improving upper extremity motor function and reducing disability during the subacute phase in post-ischemic stroke patients. There was clinically meaningful improvement in upper extremity motor function and overall disability reduction

as measured by multiple metrics, including FMA-UE, mRS, ARAT, BBT, and NIHSS. Given the high stroke prevalence and limited treatment options beyond the acute phase, these results represent a promising avenue for alternative treatment that non-invasively targets and rehabilitates compromised brain processes, and is applicable to a wide swath of post-stroke patients. The current findings should be extended by examining ENTF treatment in a larger sample with longer follow-up, as well as examining direct indices of plasticity and feasibility of continuing treatment at home. Additionally, future work may explore the generalizability of this approach to other functional domains (e.g., cognitive function), as well as other neurological and neurodegenerative disorders.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors upon reasonable request.

Ethics statement

The studies involving human participants were reviewed and approved by Dr. B. L. Kapur Memorial Hospital Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

Author contributions

BW, DP, AL, ES, YS, YD, AA, and NB were all involved in the initial study design and execution plan. DP and AP lead participant screening and recruitment, as well as study

execution. AL and OV were responsible for study oversight. OV was additionally responsible for data preparation and validation. BW, JS, AH, AL, ES, YS, SR, and NB contributed to data analysis and presentation. BW, AB, and GD prepared this manuscript. All authors reviewed and approved the final version of the manuscript.

Funding

This study was sponsored by BrainQ Technologies Ltd.

Acknowledgments

The authors wish to thank the following for their contributions to study execution and manuscript preparation: Nadav Bitton, MD, Natan Davidovics, PhD, Lisa Deutsch, PhD, Noam Kahlon, Elyaqim Oster, and Ana Parabucki, PhD. The authors also wish to acknowledge Gilda Avila-Rinek, Christa D. Brown, and Dashiell F. Young-Saver for access to the FAST-MAG database.

Conflict of interest

Authors BW, AH, AB, GD, AL, OV, ES, YS, SR, YD, and AA are/were employed by the study funder, BrainQ Technologies

Ltd. Authors ES and YS have ownership interest in BrainQ Technologies Ltd. Authors JS and NB serve as advisors to BrainQ Technologies Ltd. Author YS is the father of a neurologically disabled child. This study received funding from BrainQ Technologies Ltd. The funder was involved in the study design, data analysis and interpretation, and the writing of this article or the decision to submit it for publication.

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2022.1004677/full#supplementary-material>

References

- Feigin VL, Stark BA, Johnson CO, Roth GA, Bisignano C, Abady GG, et al. Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Neurol.* (2021) 20:795–820. doi: 10.1016/S1474-4422(21)00252-0
- Patel AP, Fisher JL, Nichols E, Abd-Allah F, Abdela J, Abdelalim A, et al. Global, regional, and national burden of brain and other CNS cancer, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* (2019) 18:376–93. doi: 10.1016/S1474-4422(18)30468-X
- Katan M, Luft A. Global burden of stroke. *Semin Neurol.* (2018) 38:208–11. doi: 10.1055/s-0038-1649503
- Goyal M, Menon BK, van Zwam WH, Dippel DWJ, 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
- Girotra T, Lekoubou A, Bishu KG, Oviagele B. A contemporary and comprehensive analysis of the costs of stroke in the United States. *J Neurol Sci.* (2019) 410:116643. doi: 10.1016/j.jns.2019.116643
- Winstein CJ, Stein J, Arena R, Bates B, Cherney LR, Cramer SC, et al. Guidelines for Adult Stroke Rehabilitation and Recovery. *Stroke.* (2016) 47:e98–e169. doi: 10.1161/STR.0000000000000098
- Veerbeek JM, van Wegen E, van Peppen R, van der Wees PJ, Hendriks E, Rietberg M, et al. What is the evidence for physical therapy poststroke? A systematic review and meta-analysis. *PLoS ONE.* (2014) 9:e87987. doi: 10.1371/journal.pone.0087987
- Bettger JP, Thomas L, Liang L, Xian Y, Bushnell CD, Saver JL, et al. Hospital Variation in Functional Recovery after Stroke. *Circ Cardiovasc Qual Outcomes.* (2017) 10:e002391. doi: 10.1161/CIRCOUTCOMES.115.002391
- Gage B, Morley M, Spain P, Ingber M. *Examining Post Acute Care Relationships in an Integrated Hospital System.* (2009). Available online at: <http://aspe.dhhs.gov/health/reports/09/pacihs/report.pdf> (accessed May 08, 2022).
- Hong I, Goodwin JS, Reistetter TA, Kuo YF, Mallinson T, Karmarkar A, et al. Comparison of functional status improvements among patients with stroke receiving postacute care in inpatient rehabilitation vs skilled nursing facilities. *JAMA Netw Open.* (2019) 2:e1916646. doi: 10.1001/jamanetworkopen.2019.16646
- Kane RL, Lin W, Blewett LA. Geographic variation in the use of post-acute. *Care Health Serv Res.* (2002) 37:667–82. doi: 10.1111/1475-6773.00043
- Canning CG, Ada L, Adams R, O'Dwyer NJ. Loss of strength contributes more to physical disability after stroke than loss of dexterity. *Clin Rehabil.* (2004) 18:300–8. doi: 10.1191/0269215504cr7150a
- Nadeau S, Arseneault AB, Gravel D, Bourbonnais D. Analysis of the clinical factors determining natural and maximal gait speeds in adults with a stroke. *Am J Phys Med Rehabil.* (1999) 78:123–30. doi: 10.1097/00002060-199903000-00007
- Rosamond W, Flegal K, Furie K, Go A, Greenlund K, Haase N, et al. Heart disease and stroke statistics-2008 Update: A report from the American heart association statistics committee and stroke statistics subcommittee. *Circulation.* (2008) 117:e25–146. doi: 10.1161/CIRCULATIONAHA.107.187998
- Brown CE, Li P, Boyd JD, Delaney KR, Murphy TH. Extensive turnover of dendritic spines and vascular remodeling in cortical tissues recovering from stroke. *J Neurosci.* (2007) 27:4101–9. doi: 10.1523/JNEUROSCI.4295-06.2007

16. Cassidy JM, Cramer SC. Spontaneous and therapeutic-induced mechanisms of functional recovery after stroke. *Transl Stroke Res.* (2017) 8:33–46. doi: 10.1007/s12975-016-0467-5
17. Johansen-Berg H, Rushworth MFS, Bogdanovic MD, Kischka U, Wimalaratna S, Matthews PM. The role of ipsilateral premotor cortex in hand movement after stroke. *Proc Natl Acad Sci U S A.* (2002) 99:14518–23. doi: 10.1073/pnas.222536799
18. Ward NS, Cohen LG. Mechanisms underlying recovery of motor function after stroke. *Arch Neurol.* (2004) 61:1844–8. doi: 10.1001/archneur.61.12.1844
19. di Pino G, Pellegrino G, Assenza G, Capone F, Ferreri F, Formica D, et al. Modulation of brain plasticity in stroke: a novel model for neurorehabilitation. *Nat Rev Neurol.* (2014) 10:597–608. doi: 10.1038/nrnneurol.2014.162
20. Liew SL, Santarnecchi E, Buch ER, Cohen LG. Non-invasive brain stimulation in neurorehabilitation: Local and distant effects for motor recovery. *Front Hum Neurosci.* (2014) 8:378. doi: 10.3389/fnhum.2014.00378
21. Xu Y, Hou QH, Russell SD, Bennett BC, Sellers AJ, Lin Q, et al. Neuroplasticity in post-stroke gait recovery and noninvasive brain stimulation. *Neural Regen Res.* (2015) 10:2072–80. doi: 10.4103/1673-5374.172329
22. Fan H, Song Y, Cen X, Yu P, Biró I, Gu Y. The effect of repetitive transcranial magnetic stimulation on lower-limb motor ability in stroke patients: a systematic review. *Front Hum Neurosci.* (2021) 15:508. doi: 10.3389/fnhum.2021.620573
23. Gomez Palacio Schjetnan A, Faraji J, Metz GA, Tatsuno M, Luczak A. Transcranial direct current stimulation in stroke rehabilitation: a review of recent advancements. *Stroke Res Treat.* (2013) 2013:170256. doi: 10.1155/2013/170256
24. Dawson J, Pierce D, Dixit A, Kimberley TJ, Robertson M, Tarver B, et al. Safety, feasibility, and efficacy of vagus nerve stimulation paired with upper-limb rehabilitation after ischemic stroke. *Stroke.* (2016) 47:143–50. doi: 10.1161/STROKEAHA.115.010477
25. Ben Yakir-Blumkin M, Loboda Y, Schächter L, Finberg JPM. Neuroprotective effect of weak static magnetic fields in primary neuronal cultures. *Neuroscience.* (2014) 278:313–26. doi: 10.1016/j.neuroscience.2014.08.029
26. Cichon N, Synowiec E, Miller E, Sliwinski T, Ceremuga M, Saluk-Bijak J, et al. Effect of rehabilitation with extremely low frequency electromagnetic field on molecular mechanism of apoptosis in post-stroke patients. *Brain Sci.* (2020) 10:266. doi: 10.3390/brainsci10050266
27. Cichon N, Bijak M, Miller E, Saluk J. Extremely low frequency electromagnetic field (ELF-EMF) reduces oxidative stress and improves functional and psychological status in ischemic stroke patients. *Bioelectromagnetics.* (2017) 38:386–96. doi: 10.1002/bem.22055
28. Cichon N, Bijak M, Czarny P, Miller E, Synowiec E, Sliwinski T, et al. Increase in blood levels of growth factors involved in the neuroplasticity process by using an extremely low frequency electromagnetic field in post-stroke patients. *Front Aging Neurosci.* (2018) 10:294. doi: 10.3389/fnagi.2018.00294
29. Lachaux JP, Rudrauf D, Kahane P. Intracranial EEG and human brain mapping. *J Physiol Paris.* (2003) 97:613–28. doi: 10.1016/j.jphysparis.2004.01.018
30. Light GA, Williams LE, Minow F, Sprock J, Rissling A, Sharp R, et al. Electroencephalography (EEG) and event-related potentials (ERPs) with human participants. *Curr Protoc Neurosci.* (2010) 52:6.25.1–24. doi: 10.1002/0471142301.n0625s2
31. Rabiller G, He JW, Nishijima Y, Wong A, Liu J. Perturbation of brain oscillations after ischemic stroke: a potential biomarker for post-stroke function and therapy. *Int J Mol Sci.* (2015) 16:25605–40. doi: 10.3390/ijms161025605
32. Goldfine AM, Victor JD, Conte MM, Bardin JC, Schiff ND. Determination of awareness in patients with severe brain injury using EEG power spectral analysis. *Clin Neurophysiol.* (2011) 122:2157–68. doi: 10.1016/j.clinph.2011.03.022
33. López-Larraz E, Montesano L, Gil-Agudo Á, Minguez J, Oliviero A. Evolution of EEG motor rhythms after spinal cord injury: a longitudinal study. *PLoS ONE.* (2015) 10:0131759. doi: 10.1371/journal.pone.0131759
34. Fröhlich F, McCormick DA. Endogenous electric fields may guide neocortical network activity. *Neuron.* (2010) 67:129–43. doi: 10.1016/j.neuron.2010.06.005
35. Martinez-Banaclocha M. Astroglial isopotentiality and calcium-associated biomagnetic field effects on cortical neuronal coupling. *Cells.* (2020) 9:439. doi: 10.3390/cells9020439
36. Afrasiabi A, Riaz GH, Abbasi S, Dadras A, Ghalandari B, Seidkhani H, et al. Synaptosomal acetylcholinesterase activity variation pattern in the presence of electromagnetic fields. *Int J Biol Macromol.* (2014) 65:8–15. doi: 10.1016/j.ijbiomac.2014.01.006
37. Iaccarino HF, Singer AC, Martorell AJ, Gao F, Gillingham TZ, Mathys H, et al. Gamma frequency entrainment attenuates amyloid load and modifies microglia. *Hannah.* (2016) 540:230–5. doi: 10.1038/nature20587
38. Segal Y, Segal L, Blumenfeld-Katzir T, Sasson E, Poliansky V, Loeb E, et al. The effect of electromagnetic field treatment on recovery from ischemic stroke in a rat stroke model: clinical, imaging, and pathological findings. *Stroke Res Treat.* (2016) 2016:6941946. doi: 10.1155/2016/6941946
39. Fugl-Meyer AR. Post-stroke hemiplegia: assessment of physical properties. *Scand J Rehabil Med.* (1980) 12 (Suppl 7):85–93.
40. Fugl-Meyer AR, Jaasko L, Leyman I. The post stroke hemiplegic patient. I A method for evaluation of physical performance. *Scand J Rehabil Med.* (1975) 7:13–31.
41. Lyle RC. A performance test for assessment of upper limb function in physical rehabilitation treatment and research. *Int J Rehabil Res.* (1981) 4:483–92. doi: 10.1097/00004356-198112000-00001
42. Mathiowetz V, Volland G, Kashman N, Weber K. Adult Norms for the Box and Block Test of Manual Dexterity (hand evaluation, hand, motor skills, occupational therapy, tests). *Am J Occup Therapy.* (1985) 39:386–91. doi: 10.5014/ajot.39.6.386
43. van Swieten JC, Koudstaal PJ, Visser MC, Schouten H, van Gijn J. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke.* (1988) 19:604–7. doi: 10.1161/01.STR.19.5.604
44. Ortiz GA, Sacco RL. National Institutes of Health Stroke Scale (NIHSS). In: D'Agostino RB, Sullivan L, Massaro J, editors. *Wiley Encyclopedia of Clinical Trials.* doi: 10.1002/9780471462422.eoct400
45. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol.* (2010) 63:1179–94. doi: 10.1016/j.jclinepi.2010.04.011
46. Partington JE, Leiter RG. Partington's pathways test. *Psychol Serv Center J.* (1949) 1:11–20. doi: 10.1037/t66320-000
47. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc.* (2005) 53:695–9. doi: 10.1111/j.1532-5415.2005.53221.x
48. Ren C, Kobeissy F, Alawieh A, Li N, Li N, Zibara K, et al. Assessment of serum UCH-L1 and GFAP in acute stroke patients. *Sci Rep.* (2016) 6:24588. doi: 10.1038/srep24588
49. Reznik Balter S, Doniger GM, Weisinger BS, Bornstein NM, Shohami E, Segal Y, et al. EEG Signatures of frequency-tuned non-invasive brain stimulation over the course of stroke recovery. *2022 American Clinical Neurophysiology Society (ACNS) Annual Meeting.* (2022). Available online at: https://www.researchgate.net/publication/359926904_EEG_Signatures_of_Frequency-Tuned_Non-invasive_Brain_Stimulation_over_the_Course_of_Stroke_Recovery
50. Page SJ, Fulk GD, Boyne P. Clinically Important Differences for the Upper-Extremity Fugl-Meyer Scale in People With Minimal to Moderate Impairment Due to Chronic Stroke. *Phys Ther.* (2012) 92:791–8. doi: 10.2522/ptj.20110009
51. Arya KN, Verma R, Garg RK. Estimating the Minimal Clinically Important Difference of an Upper Extremity Recovery Measure in Subacute Stroke Patients. (2015) 18:599–610. doi: 10.1310/tsr18s01-599. Available online at: <https://www.tandfonline.com/doi/abs/10.1310/tsr18s01-599>
52. Hiragami S, Inoue Y, Harada K. Minimal clinically important difference for the Fugl-Meyer assessment of the upper extremity in convalescent stroke patients with moderate to severe hemiparesis. *J Phys Ther Sci.* (2019) 31:917–21. doi: 10.1589/jpts.31.917
53. This Research is Based on the National Institute of Neurologic Disease and Stroke's Archived Clinical Research Data from the National Institute of Neurological Disorders and Stroke rt-PA Stroke Study, S. R. Levine (N01-NS-02382, N01-NS-02374, N01-NS-02377, N01-NS-02381, N01-NS-02379, N01-NS-02373, N01-NS-02378, N01-NS-02376, and N01-NS-02380) Received From the Archived Clinical Research Dataset Web Site.
54. Publicly available database of the NIH Field Administration of Stroke Therapy –Magnesium (FAST-MAG) Trial. Accessible to all upon request to the NIH-NINDS Archived Clinical Research Dataset repository. (2014). Available online at: <https://www.ninds.nih.gov/Current-Research/Research-Funded-NINDS/Clinical-Research/Archived-Clinical-Research-Datasets> (accessed May 08, 2022).
55. 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
56. Lai SM, Duncan PW. Stroke recovery profile and the modified rankin assessment. *Neuroepidemiology.* (2001) 20:26–30. doi: 10.1159/000054754
57. Taleb S, Duncan P, Cramer SC, Bahr-hosseini M, Starkman S, Avila G, et al. Abstract WMP42: Trajectory Of Global Disability During First 90 Days Poststroke:

Foundational Data For Early Rehabilitation Intervention Trials. *Stroke*. (2022) 53(Suppl_1). doi: 10.1161/str.53.suppl_1.WMP42

58. Espenhahn S, Rossiter HE, van Wijk BCM, Redman N, Rondina JM, Diedrichsen J, et al. Sensorimotor cortex beta oscillations reflect motor skill learning ability after stroke. *Brain Commun*. (2020) 2:fcaa161. doi: 10.1093/braincomms/fcaa161

59. Rubega M, Formaggio E, Molteni F, Guanziroli E, Marco R, Baracchini C, et al. EEG fractal analysis reflects brain impairment after stroke. *Entropy*. (2021) 23:592. doi: 10.3390/e23050592

60. Leker RR, Shohami E. Cerebral ischemia and trauma - Different etiologies yet similar mechanisms: neuroprotective opportunities. *Brain Res Rev*. (2002) 39:55–73. doi: 10.1016/S0165-0173(02)00157-1

61. Shupak NM, Prato FS, Thomas AW. Therapeutic uses of pulsed magnetic-field exposure: a review N. *URSI Radio Sci Bull*. (2003) 2003:9–32. doi: 10.23919/URSIRSB.2003.7909506

62. di Lazzaro V, Capone F, Apollonio F, Borea PA, Cadossi R, Fassina L, et al. A Consensus panel review of central nervous system effects of the exposure to low-intensity extremely low-frequency magnetic fields. *Brain Stimul*. (2013) 6:469–76. doi: 10.1016/j.brs.2013.01.004

63. Dufor T, Grehl S, Tang AD, Doulazmi M, Traoré M, Debray N, et al. Neural circuit repair by low-intensity magnetic stimulation requires cellular magnetoreceptors and specific stimulation patterns. *Sci Adv*. (2019) 5:1–12. doi: 10.1126/sciadv.aav9847

64. Piacentini R, Ripoli C, Mezzogori D, Azzena GB, Grassi C. Extremely low-frequency electromagnetic fields promote in vitro neurogenesis via upregulation of Cav1-channel activity. *J Cell Physiol*. (2008) 215:129–39. doi: 10.1002/jcp.21293

65. Grassi C, D'Ascenzo M, Torsello A, Martinotti G, Wolf F, Cittadini A, et al. Effects of 50 Hz electromagnetic fields on voltage-gated Ca²⁺ channels and their role in modulation of neuroendocrine cell proliferation and death. *Cell Calcium*. (2004) 35:307–15. doi: 10.1016/j.ceca.2003.09.001

66. Karabakhtsian R, Broude N, Shalts N, Kochlatyia S, Goodman R, Henderson AS. Calcium is necessary in the cell response to EM fields. *FEBS Lett*. (1994) 349:1–6. doi: 10.1016/0014-5793(94)00618-0

67. Fanelli C, Coppola S, Barone R, Colussi C, Gualandi G, Volpe P, et al. Magnetic fields increase cell survival by inhibiting apoptosis via modulation of Ca²⁺ influx. *FASEB J*. (1999) 13:95–102. doi: 10.1096/fasebj.13.1.95

68. Fixler D, Yitzhaki S, Axelrod A, Zinman T, Shainberg A. Correlation of magnetic AC field on cardiac myocyte Ca²⁺ transients at different magnetic DC levels. *Bioelectromagnetics*. (2012) 33:634–40. doi: 10.1002/bem.21729

69. Bauréus Koch CLM, Sommarin M, Persson BRR, Salford LG, Eberhardt JL. Interaction between Weak Low Frequency Magnetic Fields and Cell Membranes. *Bioelectromagnetics*. (2003) 24:395–402. doi: 10.1002/bem.10136

70. Balind SR, Selaković V, Radenović L, Prolić Z, Janać B. Extremely low frequency magnetic field (50 Hz, 05 mT) reduces oxidative stress in the brain of gerbils submitted to global cerebral ischemia. *PLoS ONE*. (2014) 9:e88921. doi: 10.1371/journal.pone.0088921

71. Park WH, Soh KS, Lee BC, Pyo MY. 4 Hz magnetic field decreases oxidative stress in mouse brain: A chemiluminescence study. *Electromagn Biol Med*. (2008) 27:165–72. doi: 10.1080/15368370802118662

72. Cichoń N, Czarny P, Bijak M, Miller E, Sliwiński T, Szemraj J, et al. Benign effect of extremely low-frequency electromagnetic field on brain plasticity assessed by nitric oxide metabolism during poststroke rehabilitation. *Oxid Med Cell Longev*. (2017) 2017:2181942. doi: 10.1155/2017/2181942

73. Cho SI, Nam YS, Chu LY, Lee JH, Bang JS, Kim HR, et al. Extremely low-frequency magnetic fields modulate nitric oxide signaling in rat brain. *Bioelectromagnetics*. (2012) 33:568–74. doi: 10.1002/bem.21715

74. Varani K, Vincenzi F, Targa M, Corciulo C, Fini M, Setti S, et al. Effect of pulsed electromagnetic field exposure on adenosine receptors in rat brain. *Bioelectromagnetics*. (2012) 33:279–87. doi: 10.1002/bem.20704

75. Capone F, Salati S, Vincenzi F, Liberti M, Aicardi G, Apollonio F, et al. Pulsed electromagnetic fields: a novel attractive therapeutic opportunity for neuroprotection after acute cerebral ischemia. *Neuromodulation*. (2021) 2021:13489. doi: 10.1111/ner.13489

76. Patruno A, Costantini E, Ferrone A, Pesce M, Francesca D, Trubiani O, et al. Short ELF-EMF Exposure Targets SIRT1/Nrf2/HO-1 Signaling in THP-1 Cells. *Int J Mol Sci*. (2020) 21:7284. doi: 10.3390/ijms21197284

77. Vincenzi F, Ravani A, Pasquini S, Merighi S, Gessi S, Setti S, et al. Pulsed electromagnetic field exposure reduces hypoxia and inflammation damage in neuron-like and microglial cells. *J Cell Physiol*. (2017) 232:1200–8. doi: 10.1002/jcp.25606

78. Cuccurazzu B, Leone L, Podda MV, Piacentini R, Riccardi E, Ripoli C, et al. Exposure to extremely low-frequency (50Hz) electromagnetic fields enhances adult hippocampal neurogenesis in C57BL/6 mice. *Exp Neurol*. (2010) 226:173–82. doi: 10.1016/j.expneurol.2010.08.022

79. Gladstone DJ, Danells CJ, Black SE. The Fugl-Meyer assessment of motor recovery after stroke: a critical review of its measurement properties. *Neurorehabil Neural Repair*. (2002) 16:232–40. doi: 10.1177/154596802401105171



OPEN ACCESS

EDITED BY

Andrew Gauden,
Stanford University, United States

REVIEWED BY

Bin Yang,
Capital Medical University, China
Alvin S. Das,
Beth Israel Deaconess Medical Center
and Harvard Medical School,
United States

*CORRESPONDENCE

Balázs Kis
blazsejdoktor@gmail.com

†These authors share first authorship

SPECIALTY SECTION

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

RECEIVED 28 September 2022

ACCEPTED 28 November 2022

PUBLISHED 15 December 2022

CITATION

Kis B, Neuhaus AA, Harston G, Joly O,
Carone D, Gerry S, Chadaide Z,
Pánczél A, Czifrus E, Csike V, Surányi Á,
Szikora I and Eröss L (2022) Automated
quantification of atrophy and acute
ischemic volume for outcome
prediction in endovascular
thrombectomy.
Front. Neurol. 13:1056532.
doi: 10.3389/fneur.2022.1056532

COPYRIGHT

© 2022 Kis, Neuhaus, Harston, Joly,
Carone, Gerry, Chadaide, Pánczél,
Czifrus, Csike, Surányi, Szikora and
Eröss. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Automated quantification of atrophy and acute ischemic volume for outcome prediction in endovascular thrombectomy

Balázs Kis^{1*†}, Ain A. Neuhaus^{2,3†}, George Harston^{2,4},
Olivier Joly⁴, Davide Carone^{2,4}, Stephen Gerry⁵,
Zoltán Chadaide⁴, András Pánczél⁶, Eszter Czifrus⁶,
Viktória Csike¹, Ágnes Surányi¹, István Szikora¹ and
Loránd Eröss¹

¹National Institute of Mental Health, Neurology, and Neurosurgery (NIMNN), Budapest, Hungary,

²Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom, ³Radcliffe Department of Medicine, University of Oxford, Oxford, United Kingdom, ⁴Brainomix Ltd., Oxford, United Kingdom, ⁵Centre for Statistics in Medicine, Nuffield Department of Orthopedics,

Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom,

⁶Faculty of Medicine, Semmelweis University, Budapest, Hungary

Background: Short- and long-term outcomes from endovascular thrombectomy (EVT) for large vessel occlusion stroke remain variable. Numerous relevant predictors have been identified, including severity of neurological deficits, age, and imaging features. The latter is typically defined as acute changes (most commonly Alberta Stroke Programme Early CT Score, ASPECTS, at presentation), but there is little information on the impact of imaging assessment of premorbid brain health as a determinant of outcome.

Aims: To examine the impact of automated measures of stroke severity and underlying brain frailty on short- and long-term outcomes in acute stroke treated with EVT.

Methods: In 215 patients with anterior circulation stroke, who subsequently underwent EVT, automated analysis of presenting non-contrast CT scans was used to determine acute ischemic volume (AIV) and e-ASPECTS as markers of stroke severity, and cerebral atrophy as a marker of brain frailty. Univariate and multivariate logistic regression were used to identify significant predictors of NIHSS improvement, modified Rankin scale (mRS) at 90 and 30 days, mortality at 90 days and symptomatic intracranial hemorrhage (sICH) following successful EVT.

Results: For long-term outcome, atrophy and presenting NIHSS were significant predictors of mRS 0–2 and death at 90 days, whereas age did not reach significance in multivariate analysis. Conversely, for short-term NIHSS improvement, AIV and age were significant predictors, unlike presenting NIHSS. The interaction between age and NIHSS was similar to the interaction of AIV and atrophy for mRS 0–2 at 90 days.

Conclusion: Combinations of automated software-based imaging analysis and clinical data can be useful for predicting short-term neurological outcome and may improve long-term prognostication in EVT. These results provide a basis for future development of predictive tools built into decision-aiding software in stroke.

KEYWORDS

stroke, endovascular thrombectomy, thrombolysis, neuroimaging, neuroradiology, artificial intelligence, machine learning

Introduction

Outcomes in endovascular thrombectomy (EVT) for emergent large vessel occlusion (ELVO) stroke remain variable: only 46% of patients treated with EVT in early time-window trials achieved functional independence (modified Rankin scale, mRS 0–2) (1). In real-world registry studies, the overall functional independence has been reported to be as low as 23% after EVT (2). As such, improved patient selection criteria are required to maximize cost-effectiveness while minimizing unnecessary procedure-related risks.

Currently, identifying suitable candidates relies on clinical features (presenting severity on the National Institutes of Health Stroke Scale, NIHSS; baseline mRS), time from stroke onset (3) and imaging findings, including infarct volume (4), Alberta Stroke Programme Early CT Score (ASPECTS) (5) and volume of ischemic tissue relative to infarct (6). However, these radiological features all represent acute changes. There is increasing evidence that imaging biomarkers unrelated to the index stroke, including atrophy as an indicator of brain frailty, have a significant impact on clinical outcome (7, 8). However, there is no consensus on the best integrative approach to use these factors for long-term prognosis and poor understanding of how they interact. Furthermore, manual quantification of atrophy using visual analog scales can be subjective (9) while automated quantification usually requires separate software, which would complicate reporting workflow in the acute setting.

In this study, we sought to use machine learning based automated image analysis of routine non-contrast CT (NCCT) brain imaging, in conjunction with clinical variables, to improve prognostication for short-term neurological improvement (changes in NIHSS), long-term functional outcomes (mRS at 30 and 90 days), mortality and symptomatic intracranial hemorrhage. Our hypothesis was that automated image analysis of routine NCCT imaging markers of both stroke severity and brain frailty would be significant predictors of short- and long-term outcome alone following successful EVT.

Methods

Study design

This retrospective study was performed at the National Institute of Mental Health, Neurology and Neurosurgery (NIMNN) in Budapest, Hungary. From 1 January 2017 to 31 December 2019, we included patients receiving endovascular thrombectomy for acute ischemic stroke. All treatment decisions were made based on clinical criteria, as indicated according to first line international (3) and local guidelines. Further inclusion criteria included: age ≥ 18 , causative middle cerebral artery M1 segment occlusion, onset to groin puncture ≤ 6 h, NIHSS score ≥ 6 at presentation, premorbid mRS state ≤ 1 , ASPECTS ≥ 6 , and successful recanalization (TICI ≥ 2 b).

Since more than 90% of the patients are transferred from primary stroke centers to NIMNN for EVT, an NCCT scan was repeated on arrival for final decision-making regarding EVT. These preprocedural scans were defined as baseline imaging and processed using e-Stroke software (version 10; Brainomix, UK), a machine learning based decision aid tool for acute ischemic stroke, which has been validated against manual analysis by neuroradiologists in previous studies (10, 11). e-Stroke was used to estimate acute ischemic volume (AIV), automated e-ASPECTS, and atrophy (defined based on CSF volume in the lateral ventricles and surrounding brain parenchyma, relative to parenchymal volume). Patients with missing or inadequate baseline imaging were excluded from analyses.

We further collected patient demographics, baseline imaging features (including radiologist ASPECTS scores), treatment times (onset to groin, groin to recanalization, NIHSS, whether the patient received intravenous thrombolysis, TICI scores, short-term neurological outcome (NIHSS improvement at discharge or 7 days), longer term functional outcome (mRS at 30 and 90 days) and whether the patient suffered symptomatic intracranial hemorrhage [defined as European Cooperative Acute Stroke Study (ECASS) parenchymal hematoma 1 and 2 categories].

TABLE 1 Demographic details of the study cohort.

Age (years)	67.6 (SD 14.1)
Female	128 (59.5%)
Presenting features	
Left-sided occlusion	103 (47.9%)
NIHSS on admission	15 (IQR 12–19)
Manual ASPECTS on admission	8 (IQR 7–8)
e-ASPECTS on admission	8 (IQR 8–9)
NCCT infarct volume on admission (mL)	17.5 (SD 15.6)
Atrophy on admission (% brain volume)	11.48 (SD 4.64)
Comorbidities	
Atrial fibrillation	91 (42.3%)
Hypertension	163 (75.8%)
Diabetes mellitus	48 (22.3%)
Ischaemic heart disease	34 (15.8%)
Peripheral arterial disease	27 (12.6%)
Prior stroke	26 (12.1%)
Procedural features	
Onset to recanalization (min)	287.2 (SD 68.3)
Onset to groin (min)	251 (SD 65.1)
Door to groin (min)	49.8 (SD 26.4)
Groin to recanalization (min)	36.7 (SD 22.2)
TICI 2B	71 (33.0%)
TICI 2C	32 (14.9%)
TICI 3	112 (52.1%)
First pass recanalization	122 (56.7%)
IV thrombolysis	130 (60.5%)

Data analysis

Data were collected into a database as part of clinical care and therefore unblinded. Patients were invited for follow-up clinic appointments where mRS was determined by the reviewing physician; for patients who did not attend appointments, trained abstractors collected mRS over the phone. Statistical analyses were performed in R version 4.1.1 (12). Primary outcomes were considered as early, defined as dichotomized NIHSS improvement at discharge (reduction of NIHSS of ≥ 4 points in patients with a presenting NIHSS ≥ 4), and late, defined as dichotomized functional outcome at 90 days (with mRS 0–2 considered as good outcome). Secondary outcomes included mRS 0–2 at 30 days, intracranial hemorrhage, and mortality at 90 days. Univariate logistic regression was performed for manually selected variables. Multivariate logistic regression was performed with our four primary predictors of interest (presenting NIHSS and AIV as clinical and radiological indicators of acute severity, and age and atrophy as clinical and radiological indicators of brain frailty), alongside variables that were significant in univariate analyses ($p < 0.05$). As AIV and e-ASPECTS were highly collinear,

TABLE 2 Univariate and multivariate predictors of mRS 0–2 at 90 days.

Predictors	mRS 0–2 at 90 days		
	Odds Ratios	CI	<i>p</i>
Univariate			
(Intercept)	13.92	3.38–63.91	<0.001
Age	0.96	0.94–0.98	<0.001
NIHSS at admission	0.89	0.84–0.94	<0.001
AIV	0.98	0.96–0.99	0.017
Atrophy	0.88	0.82–0.94	<0.001
TICI 2C/3	2.05	1.15–3.70	0.016
e-ASPECTS	1.26	1.04–1.56	0.021
IV thrombolysis	1.28	0.74–2.21	0.385
Onset to recanalisation	1	0.99–1.00	0.066
Multivariate			
(Intercept)	44.54	7.78–299.35	<0.001
Age	0.98	0.95–1.01	0.113
NIHSS at admission	0.9	0.84–0.95	0.001
AIV	0.98	0.96–1.00	0.127
Atrophy	0.91	0.83–0.99	0.038
TICI 2C/3	3	1.55–5.99	0.001

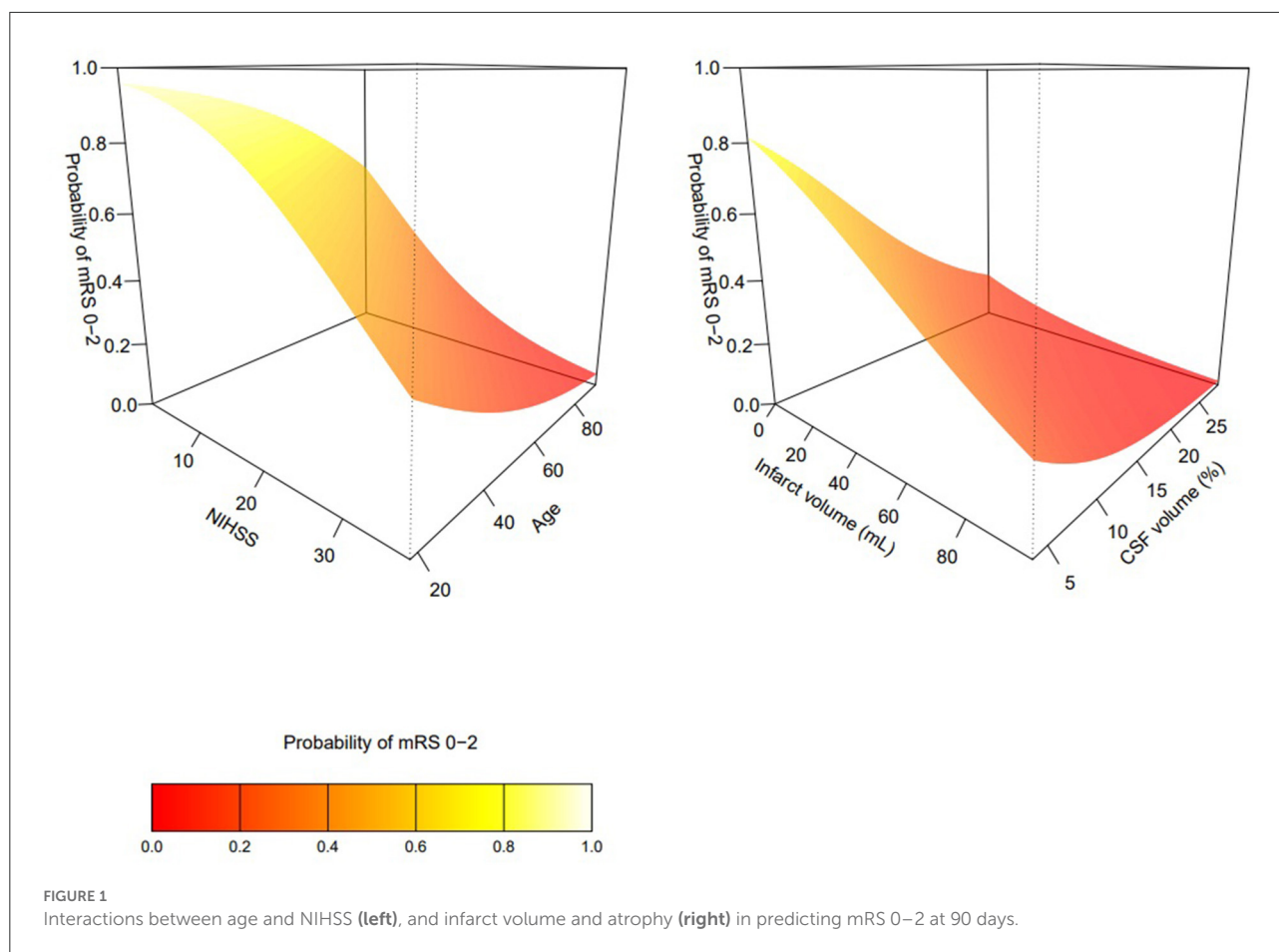
The bold values indicate the statistical significance of $p < 0.05$.

the latter was not included in multivariate models even if significant in univariate analysis. Cases with missing outcome data were excluded from analysis for that model. Descriptive analysis was undertaken to explore the interaction between stroke severity and brain frailty on outcome, using both clinical and imaging biomarkers.

Results

The study population included 215 patients; demographic details and procedural characteristics are displayed in Table 1. Overall, 104 patients (48.4%) achieved mRS 0–2 at 90 days. The significant univariate and multivariate predictors of achieving mRS 0–2 are summarized in Table 2. Briefly, univariate analyses showed that likelihood of good outcome was significantly associated with age, NIHSS on admission, NCCT AIV, e-ASPECTS, atrophy and TICI status. When adjusted in multivariate regression, however, only atrophy, NIHSS on admission and TICI 2C/3 retained statistical significance. The cumulative effect of age and NIHSS, and separately infarct volume, and atrophy on functional outcome are shown in Figure 1. The pattern of interaction between stroke severity and surrogates of baseline frailty showed a similar relationship when assessed using both clinical and imaging characteristics.

Using the same approach, we then looked at early neurological outcomes following thrombectomy. NIHSS improvement data were available in 203 patients, with six



patients excluded due to missing follow up NIHSS and six patients excluded due to baseline severity of NIHSS < 4. NIHSS improvement was achieved by 162 patients (79.8%), albeit with a wide range of absolute NIHSS values at discharge (range 0–30, IQR 2–11). The univariate and multivariate predictors for NIHSS improvement are summarized in [Table 3](#). In univariate analyses, only AIV and onset to recanalization time were associated with NIHSS improvement. After adjustment in multivariate analysis, age, AIV and onset to recanalization demonstrated statistical significance.

We performed the same analyses for mRS 0–2 at 30 days, achieved by 90 patients (41.9%). Univariate and multivariate predictors are displayed in [Supplementary Table 1](#). In univariate analysis, age, AIV, NIHSS, atrophy and TICI 2C/3 status were all associated with functional independence at 30 days, similarly to the 90-day mRS; in the adjusted analysis, only age, AIV and TICI retained significance. At 90 days, 65 patients (30.2%) had died. The univariate and multivariate analyses for mortality are in [Supplementary Table 2](#); age, admission NIHSS and atrophy were associated with death in univariate analysis, while only NIHSS and atrophy demonstrated statistical significance in multivariate analysis. We also attempted to examine post-procedural sICH, but this was limited by sample size as only 12 patients developed

parenchymal hemorrhage type 1 or 2. NIHSS on admission was the only variable to show significant association to sICH in both univariate and multivariate analysis ([Supplementary Table 3](#)).

Discussion

Imaging biomarkers automatically derived from CT imaging routinely acquired at presentation provided information on both the acute injury sustained (acute infarct volume) and on the premorbid condition of the brain (brain atrophy). Both clinical and imaging markers of stroke severity, i.e., presenting NIHSS and acute infarct volume, offer potentially useful information when predicting neurological improvement. Furthermore, there is additional information to be gained from both clinical and imaging markers of brain frailty when predicting early outcomes. The interactions between markers of stroke severity and surrogates of frailty appear similar whether clinical or radiological variables are used.

Atrophy was more strongly associated with long-term outcome compared to age. Compared to age, atrophy likely provides a better estimate of brain frailty, a key determinant

TABLE 3 Univariate and multivariate predictors of NIHSS improvement.

Predictors	NIHSS improvement		
	Odds Ratios	CI	<i>p</i>
Univariate			
Age	0.98	0.95–1.01	0.141
NIHSS at admission	1.02	0.95–1.09	0.637
AIV	0.97	0.94–0.99	0.001
Atrophy	1.01	0.94–1.10	0.792
TICI 2C/3	1.97	0.97–3.99	0.059
e-ASPECTS	1.20	0.95–1.50	0.116
IV thrombolysis	1.43	0.71–2.86	0.311
Onset to recanalisation	0.99	0.98–1.00	0.001
Multivariate			
(Intercept)	3983.64	140.65–178884.09	<0.001
Age	0.94	0.90–0.98	0.004
NIHSS at admission	1.08	1.00–1.18	0.061
AIV	0.95	0.92–0.97	<0.001
Atrophy	1.12	0.99–1.28	0.082
Onset to recanalisation	0.99	0.98–0.99	<0.001

The bold values indicate the statistical significance of $p < 0.05$.

of vulnerability to injury (13). Metrics capturing biological age, as opposed to calendar age, might also better inform of the capacity of an individual to recover following stroke as they may better represent the brain neurological reserve and the ability to compensate for an infarct of given volume following stroke.

While long term outcome is influenced by multiple factors, such as post EVT care, rehabilitation, and comorbid factors, the early neurological outcome might be argued to be more sensitive to the impact of factors related to the index stroke severity and degree of procedural success. It is unclear why age rather than atrophy demonstrated this association in these results; hypothetically, non-neurological factors such as musculoskeletal comorbidity may contribute to the potential for early neurological compensation which is more likely to be captured by age than a neurologically specific biomarker such as atrophy.

The influence of the individual biomarkers used in this study is broadly consistent with previous observations. e-ASPECTS, which is derived from the automated AIV, has been demonstrated to predict mRS at 90 days in other cohorts (14). Similarly, automated atrophy quantification strongly predicts mRS, with an odds ratio of 0.65 per 5% increase in intracranial cerebrospinal fluid volume (15). A previous machine learning study of NCCT and CTA features using e-Stroke found that age, baseline NIHSS, occlusion side, atrophy and e-ASPECTS were the best predictors for mRS at 90 days, in keeping with the results shown here (16). The clinical variables used in this study—age and presenting NIHSS—have previously been used as a prognostic score, which was also an independent predictor of

outcome in EVT (17). Other studies have also found significant predictive effects from age and NIHSS at presentation (18), and there is a combined effect between NIHSS and age in agreement with the data presented here (19). Procedurally, there is a strong correlation between treatment times and outcome, particularly in the early window (2). Excellent reperfusion (TICI 2C/3) is also highly predictive of mRS at 3 months, especially when this is achieved during the first pass (20).

There is comparatively less data on predictors of early neurological improvement which is more directly linked to the immediate impact of an intervention, and most of these focus on the earliest time windows. One study identified age, blood glucose, TICI, baseline ASPECTS and the presence of sICH as predictors for early neurological recovery (21). Others have looked at predictive factors for failure to neurologically improve, and identified variables including premorbid mRS, hyperglycemia, longer treatment times, lack of tPA bridging, and involvement of motor cortex and internal capsule in the infarct (22). Separately, there is considerable literature suggesting the utility of early neurological improvement itself as a long-term prognostic factor. In a large cohort study, absolute NIHSS at 24h was the best predictor of mRS at 90 days, although NIHSS improvement was also strongly correlated with long-term outcome (23). Notably, in our cohort AIV was a better predictor of early neurological improvement than atrophy, illustrating the differential impact of acute and chronic brain changes to short- and long-term recovery.

Our study has limitations. First, although the group sizes of those achieving mRS 0–2 and 3–6 were similar, the overall sample size was relatively small at 215 patients. In addition, the intracerebral hemorrhage analyses were underpowered as there were only 12 cases in the cohort. Second, this was a single-center retrospective cohort study. Validation in prospective cohorts is required. Third, our hypotheses and choices of variables were clinically driven, and we cannot exclude that there are other imaging parameters that may further improve performance but were not captured in this dataset.

Overall, these data support the use of automated imaging analysis for improving the prediction of neurological recovery following EVT and demonstrate the additional information that can be captured from simple NCCT imaging. The results show the interaction between markers of stroke severity and the brain frailty of the individual and suggest an opportunity to refine estimates of an individual's capacity to recover beyond that of their chronological age. If validated in prospective cohorts, this may provide a useful adjunct tool for prognostication in large vessel occlusion ischemic stroke.

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 National Institute of Mental Health, Neurology and Neurosurgery (NIMNN), Budapest, Hungary. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

BK, ZC, AP, EC, VC, ÁS, IS, and LE designed the study and collected data. BK, AN, and SG performed data analysis. GH and OJ provided essential data analysis tools. AN, GH, and DC drafted the manuscript. All authors were involved with manuscript revisions and contributed to the article and approved the submitted version.

Conflict of interest

Authors DC, GH, OJ, and ZC have share options in and are employees of Brainomix. Author SG has received consultancy fees from Brainomix.

References

- Menon BK, van Zwam WH, Dippel DWJ, Mitchell PJ, Demchuk AM, et al. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *The Lancet*. (2016) 387:1723–31. doi: 10.1016/S0140-6736(16)00163-X
- Jahan R, Saver JL, Schwamm LH, Fonarow GC, Liang L, Matsouka RA, et al. Association between time to treatment with endovascular reperfusion therapy and outcomes in patients with acute ischemic stroke treated in clinical practice. *JAMA*. (2019) 322:252–63. doi: 10.1001/jama.2019.8286
- 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
- Boers AMM, Jansen IGH, Beenen LFM, Devlin TG, San Roman L, Heo JH, et al. Association of follow-up infarct volume with functional outcome in acute ischemic stroke: a pooled analysis of seven randomized trials. *J Neurointerv Surg*. (2018) 10:1137–42. doi: 10.1136/neurintsurg-2017-013724
- Cagnazzo F, Derraz I, Dargazanli C, Lefevre PH, Gascou G, Riquelme C, et al. Mechanical thrombectomy in patients with acute ischemic stroke and ASPECTS 6: a meta-analysis. *J Neurointerv Surg*. (2020) 12:350–5. doi: 10.1136/neurintsurg-2019-015237
- 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
- Lauksio I, Lindstrom I, Khan N, Sillanpaa N, Hernesniemi J, Oksala N, et al. Brain atrophy predicts mortality after mechanical thrombectomy of proximal anterior circulation occlusion. *J Neurointerv Surg*. (2021) 13:415–20. doi: 10.1136/neurintsurg-2020-016168
- Pedraza MI, de Lera M, Bos D, Calleja AI, Cortijo E, Gomez-Vicente B, et al. Brain atrophy and the risk of futile endovascular reperfusion in acute ischemic stroke. *Stroke*. (2020) 51:1514–21. doi: 10.1161/STROKEAHA.119.028511
- Harper L, Barkhof F, Fox NC, Schott JM. Using visual rating to diagnose dementia: a critical evaluation of MRI atrophy scales. *J Neurol Neurosurg Psychiatry*. (2015) 86:1225–33. doi: 10.1136/jnnp-2014-310090
- Sundaram VK, Goldstein J, Wheelwright D, Aggarwal A, Pawha PS, Doshi A, et al. Automated ASPECTS in acute ischemic stroke: a comparative analysis with CT perfusion. *AJNR Am J Neuroradiol*. (2019) 40:2033–8. doi: 10.3174/ajnr.A6303
- Wolff L, Berkhemer OA, van Es A, van Zwam WH, Dippel DWJ, Majoie C, et al. Validation of automated alberta stroke program early ct score (ASPECTS) software for detection of early ischemic changes on non-contrast brain CT scans. *Neuroradiology*. (2021) 63:491–8. doi: 10.1007/s00234-020-02533-6
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing (2021).
- Evans NR, Todd OM, Minhas JS, Fearon P, Harston GW, Mant J, et al. Frailty and cerebrovascular disease: concepts and clinical implications for stroke medicine. *Int J Stroke*. (2022) 17:251–9. doi: 10.1177/17474930211034331
- Pfaff J, Herweh C, Schieber S, Schonenberger S, Bosel J, Ringleb PA, et al. e-ASPECTS correlates with and is predictive of outcome after mechanical thrombectomy. *AJNR Am J Neuroradiol*. (2017) 38:1594–9. doi: 10.3174/ajnr.A5236
- Diprose WK, Diprose JB, Wang MTM, Tarr GP, McFetridge A, Barber PA. Automated measurement of cerebral atrophy and outcome in endovascular thrombectomy. *Stroke*. (2019) 50:3636–8. doi: 10.1161/STROKEAHA.119.027120
- Jabal MS, Joly O, Kallmes D, Harston G, Rabinstein A, Huynh T, et al. Interpretable machine learning modeling for ischemic stroke outcome prediction. *Front Neurol*. (2022) 13:884693. doi: 10.3389/fneur.2022.884693
- Almekhlafi MA, Davalos A, Bonafe A, Chapot R, Gralla J, Pereira VM, et al. Impact of age and baseline NIHSS scores on clinical outcomes in the mechanical thrombectomy using solitaire FR in acute ischemic stroke study. *AJNR Am J Neuroradiol*. (2014) 35:1337–40. doi: 10.3174/ajnr.A3855
- Reddy ST, Friedman E, Wu TC, Arevalo O, Zhang J, Rahbar MH, et al. Rapid infarct progression in anterior circulation large vessel occlusion ischemic

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2022.1056532/full#supplementary-material>

stroke patients during inter-facility transfer. *J Stroke Cerebrovasc Dis.* (2020) 29:105308. doi: 10.1016/j.jstrokecerebrovasdis.2020.105308

19. Ospel JM, Brown S, Kappelhof M, van Zwam W, Jovin T, Roy D, et al. Comparing the prognostic impact of age and baseline national institutes of health stroke scale in acute stroke due to large vessel occlusion. *Stroke.* (2021) 52:2839–45. doi: 10.1161/STROKEAHA.120.032364

20. Yoo AJ, Soomro J, Andersson T, Saver JL, Ribo M, Bozorgchami H, et al. Benchmarking the extent and speed of reperfusion: first pass tICI 2c-3 is a preferred endovascular reperfusion endpoint. *Front Neurol.* (2021) 12:669934. doi: 10.3389/fneur.2021.669934

21. Zhang X, Peng M, Feng C, Wang H, Gong P, Jiang T, et al. Nomogram predicting early neurological improvement in ischaemic stroke patients treated with endovascular thrombectomy. *Eur J Neurol.* (2021) 28:152–60. doi: 10.1111/ene.14510

22. Weyland CS, Mokli Y, Vey JA, Kieser M, Herweh C, Schonenberger S, et al. Predictors for failure of early neurological improvement after successful thrombectomy in the anterior circulation. *Stroke.* (2021) 52:1291–8. doi: 10.1161/STROKEAHA.120.030519

23. Meyer L, Broocks G, Bechstein M, Flottmann F, Leischner H, Brekenfeld C, et al. Early clinical surrogates for outcome prediction after stroke thrombectomy in daily clinical practice. *J Neurol Neurosurg Psychiatry.* (2020) 91:1055–9. doi: 10.1136/jnnp-2020-323742



OPEN ACCESS

EDITED BY

Jean-Claude Baron,
University of Cambridge, United Kingdom

REVIEWED BY

Anna Podlasek,
University of Dundee, United Kingdom
Arvind Bambhroliya,
University of Texas Health Science Center at
Houston, United States
Jennifer Soun,
University of California, Irvine, United States

*CORRESPONDENCE

Giles Tetteh
✉ giles.tetteh@tum.de

†PRESENT ADDRESS

Johannes C. Paetzold,
Department of Computing, Imperial College
London, London, United Kingdom

SPECIALTY SECTION

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

RECEIVED 08 September 2022

ACCEPTED 02 February 2023

PUBLISHED 21 February 2023

CITATION

Tetteh G, Navarro F, Meier R, Kaesmacher J,
Paetzold JC, Kirschke JS, Zimmer C, Wiest R
and Menze BH (2023) A deep learning approach
to predict collateral flow in stroke patients
using radiomic features from perfusion images.
Front. Neurol. 14:1039693.
doi: 10.3389/fneur.2023.1039693

COPYRIGHT

© 2023 Tetteh, Navarro, Meier, Kaesmacher,
Paetzold, Kirschke, Zimmer, Wiest and Menze.
This is an open-access article distributed under
the terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

A deep learning approach to predict collateral flow in stroke patients using radiomic features from perfusion images

Giles Tetteh^{1,2*}, Fernando Navarro¹, Raphael Meier³,
Johannes Kaesmacher³, Johannes C. Paetzold^{1†}, Jan S. Kirschke²,
Claus Zimmer², Roland Wiest³ and Bjoern H. Menze^{1,4}

¹Department of Computer Science, Technische Universität München, München, Germany,

²Neuroradiology, Klinikum Rechts der Isar, Technische Universität München, München, Germany,

³Institute for Diagnostic and Interventional Neuroradiology, Inselspital University Hospital, Bern, Switzerland, ⁴Department of Quantitative Biomedicine, University of Zurich, Zurich, Switzerland

Collateral circulation results from specialized anastomotic channels which are capable of providing oxygenated blood to regions with compromised blood flow caused by arterial obstruction. The quality of collateral circulation has been established as a key factor in determining the likelihood of a favorable clinical outcome and goes a long way to determining the choice of a stroke care model. Though many imaging and grading methods exist for quantifying collateral blood flow, the actual grading is mostly done through manual inspection. This approach is associated with a number of challenges. First, it is time-consuming. Second, there is a high tendency for bias and inconsistency in the final grade assigned to a patient depending on the experience level of the clinician. We present a multi-stage deep learning approach to predict collateral flow grading in stroke patients based on radiomic features extracted from MR perfusion data. First, we formulate a region of interest detection task as a reinforcement learning problem and train a deep learning network to automatically detect the occluded region within the 3D MR perfusion volumes. Second, we extract radiomic features from the obtained region of interest through local image descriptors and denoising auto-encoders. Finally, we apply a convolutional neural network and other machine learning classifiers to the extracted radiomic features to automatically predict the collateral flow grading of the given patient volume as one of three severity classes - no flow (0), moderate flow (1), and good flow (2). Results from our experiments show an overall accuracy of 72% in the three-class prediction task. With an inter-observer agreement of 16% and a maximum intra-observer agreement of 74% in a similar experiment, our automated deep learning approach demonstrates a performance comparable to expert grading, is faster than visual inspection, and eliminates the problem of grading bias.

KEYWORDS

collateral flow, radiomics, perfusion, reinforcement learning, image descriptors, angiography, auto-encoder, deep learning

1. Introduction

Collateral circulation results from specialized anastomotic channels which are present in most tissues and capable of providing nutrient perfusion to regions with compromised blood flow due to ischemic injuries caused by ischemic stroke, coronary atherosclerosis, peripheral artery disease, and similar conditions or diseases (1). Collateral circulation

helps to sustain blood flow in the ischaemic areas in acute, subacute, or chronic phases after an ischaemic stroke or transient ischaemic attack (2). The quality of collateral circulation has been convincingly established as a key factor in determining the likelihood of successful reperfusion and favorable clinical outcome (3). It is also seen as one of the major determinants of infarct growth in the early time windows which is likely to have an impact on the chosen stroke care model that is the decision to transport or treat eligible patients immediately.

A high number of imaging methods exist to assess the structure of the cerebral collateral circulation and several grading criteria have been proposed to quantify the characteristics of collateral blood flow. However, this grading is mostly done through visual inspection of the acquired images which introduces two main challenges.

First, there are *biases and inconsistencies in the current grading approaches*: There is a high tendency of introducing bias in the final grade assigned to a patient depending on the experience level of the clinician. There are inconsistencies also in the grade assigned by a particular clinician at different times for the same patient. These inconsistencies are quantified at 16% interobserver agreement and a maximum intraobserver agreement of 74% respectively in a similar study by Ben Hassen et al. (4).

Second, *grading is time-consuming and tedious*: Aside the problem of bias prediction, it also takes the clinician several minutes to go through the patient images to first select the correct image sequence, detect the region of collateral flow and then to be able to assign a grading a period of time which otherwise could have been invested in the treatment of the patient.

In this work, we analyze several machine learning and deep learning strategies that aim toward automating the process of collateral circulation grading. We present a set of solutions focusing on two main aspects of the task at hand.

First, *the region of interest needs to be identified*. We automate the extraction of the region of interest (ROI) from the patient images using deep reinforcement learning (RL). This is necessary for achieving a fully automated system that will require no human interaction and save the clinician the time spent on performing this task.

Finally, *the region of interest needs to be processed and classified*. We consider various feature extraction schemes and classifiers suitable for the task described above. This helps to extract useful image features, both learned and hand-crafted, which are relevant to the classification task. We predict digitally subtracted angiography (DSA) based collateral flow grading from MR perfusion images in this task. This saves the time required in choosing the right DSA sequence from the multiple DSA sequences acquired and helps achieve a fully automated system.

1.1. Prior work and open challenges

1.1.1. Imaging criteria for cerebral collateral circulation

Imaging methods for assessing cerebral collateral flow can be grouped under two main classification schemes, invasive vs. non-invasive and structural vs. functional imaging. Structural

imaging methods provide information about the underlying structure of the cerebral collateral circulation network. Some of the commonly used structural imaging modalities are traditional single-phase computed tomography angiography (CTA), time-of-flight magnetic resonance angiography (TOF-MRA), and digitally subtracted angiography (DSA), among others. Other imaging modalities have been used in clinical practice and relevant research areas in accessing the structure of the cerebral collateral circulation are discussed in Liu et al. (2), McVerry et al. (5), Martinon et al. (6). DSA is the gold standard for assessing the collateral flow, however, due to the associated high cost and invasive nature, other non-invasive methods like CTA and MRA are commonly used (2).

Functional imaging methods are used to assess the function of the underlying cerebral collateral circulation. Single-photon emission CT (SPECT), MR perfusion, and positron emission tomography (PET) are examples of imaging methods that provide functional information about the cerebral collateral flow. MR perfusion imaging is often followed by a post-processing step to extract parametric information. Very common parametric information includes the *time-to-peak* (T_{max}) time taken for the blood flow to reach its peak (max) at a given region in the brain, *relative blood flow* (rBF) volume of blood flowing through a given brain tissue per unit of time, and *relative blood volume* (rBV) volume of blood in a given brain tissue relative to an internal control (e.g. normal white matter or an arterial input function). Functional imaging is sometimes combined with structural imaging either in a single scanning procedure or separate procedures and can serve as complementing information in the decision making process. Here, structural imaging is oftentimes used to map the anatomy and probe tissue microstructure.

MRI perfusion and diffusion have evolved as key biomarkers in determining collateralization of stroke patients, and a patient stratification based on these markers has been proposed repeatedly (7). At the same time, a qualitative CTA and DSA based grading are the most common approaches for evaluating collateralization (8–10).

1.1.2. Cerebral collateral flow grading

Cerebral collateral circulation plays an important role in stabilizing cerebral blood flow when the normal blood circulation system is compromised in cases of acute, subacute, or chronic ischaemic stroke. The quality of the cerebral collateral circulation system is one of the factors that determine the speed of infarct growth and the outcome of stroke treatment and reperfusion therapies. A poor collateral flow is associated with worse outcomes and faster growth of infarcts while a good collateral flow is associated with good outcomes and slower growth of infarcts in acute stroke treatment (11). Due to the important role played by cerebral collateral blood flow, various grading scales and their association with risk factors and treatment outcomes have been discussed extensively in literature.

Several grading systems have been proposed for assessing the quality of the collateral circulation network. Among these grading systems, the DSA based system proposed by the American Society of Interventional and Therapeutic Neuroradiology/Society of Interventional Radiology (ASITN/SIR) is the most widely

accepted scheme. This grading system describes the collateral flow as one of five levels of flow which are; absence of collaterals (0), slow collaterals (1), rapid collaterals (2), partial collaterals (3), and complete collaterals (4) to the periphery of the ischaemic site (2, 12). In most studies that use the ASITN/SIR scheme, the grading scale is merged into three levels—grades 0–1 (poor), 2 (moderate) and 3–4 (good collateral) flow. CTA based systems also have several grading schemes ranging from two (good, bad) to five (absent, diminished >50%, <50%, equal, more) labels (12).

The relationship between pretreatment collateral grade and vascular recanalization has been assessed for patients who received endovascular therapy for acute cerebral ischemia from two distinct study populations by Bang et al. (13). The study showed that 14.1, 25.2, and 41.5% of patients with poor, good, and excellent pretreatment collaterals respectively achieved complete revascularization. Another study by Bang et al. (14) on the relationship between MRI diffusion and perfusion lesion indices, angiographic collateral grade, and infarct growth showed that the greatest infarct growth occurred in patients with both non-recanalization and poor collaterals. Mansour (15) assessed collateral pathways in acute ischemic stroke using a new grading scale (Mansour Scale) and correlated the findings with different risk factors, clinical outcomes, and recanalization rates with endovascular management. More research (13–17) has been conducted into the relationship between the cerebral collateral circulation, its grading, and the clinical outcome of the choice of treatment of acute ischemic stroke, and they all confirm a positive association between collateral flow and the success of the outcome.

Due to the crucial role played by collateral circulation, it is a common practice in most clinical procedures to determine the quality of a patient's collateral as first-hand information toward the choice of the treatment or care model. This grading is done manually by inspecting patient scans which is time-consuming and also introduces some level of bias in the final grade assigned to a patient. Ben Hassen et al. (4) evaluated the inter- and intraobserver agreement in angiographic leptomeningeal collateral flow assessment on the ASITN/SIR scale and found an overall interobserver agreement $\kappa = 0.16 \pm 6.5 \times 10^{-3}$ among 19 observers with grades 0 and 1 being associated with the best results of $\kappa = 0.52 \pm 0.001$ and $\kappa = 0.43 \pm 0.004$ respectively. By merging the scales into two classes, poor collaterals (grade 0, 1, or 2), versus good collaterals (grade 3 or 4), the interobserver agreement increased to $\kappa = 0.27 \pm 0.014$. The same study recorded maximum intraobserver agreements of $\kappa = 0.74 \pm 0.1$ and $\kappa = 0.79 \pm 0.11$ for the ASITN/SIR and dichotomized scales respectively. McHugh (18) presented a study on interrater reliability and the kappa statistic as a measure of agreement and recommended a moderate interobserver agreement of $0.60 \leq \kappa \leq 0.79$ as a minimum requirement for medical data and study. These results are evidence of the need to automate the collateral grading process to achieve speed and consistency in the assigned grading.

Methods for automating the grading of collateral flow have not yet been properly explored in literature. Kersten-Oertel et al. (19) presented an automated technique to compute a collateral circulation score based on differences seen in mean intensities between left and right cerebral hemispheres in 4D angiography images and found a good correlation between the computed score and radiologist score ($r^2 = 0.71$) and good separation

between good and intermediate/poor groups. Grunwald et al. (20) used a machine learning approach to categorize the degree of collateral flow in 98 patients who were eligible for mechanical thrombectomy and generated an e-CTA collateral score (CTA-CS) for each patient. The experiments showed that the e-CTA generated improved the intraclass correlation coefficient between three experienced neuroradiologists from 0.58 (0.46–0.67) to 0.77 (0.66–0.85, $p = 0.003$).

1.1.3. Reinforcement learning for medical imaging

Defining the region of interest (ROI) is often the first step in most image-based radiomics pipelines. This is the case because full patient scans often include artifacts and other information which are irrelevant and can affect the final outcome of the study. Therefore, most pipelines propose a manual localization of a ROI as a preprocessing step. However, it is crucial to define the ROI in an automated and reproducible fashion in order to achieve a fully automated pipeline. We propose a reinforcement learning approach for the localization of the region of interest due to increased speed and lower training data requirements compared to other supervised learning approaches.

Reinforcement learning (RL) has become one of the most active research areas in machine learning and involves the training of a machine learning agent to make a sequence of reward-based decisions toward the achievement of a goal through interaction with the environment. The idea of RL has been long applied in the field of robotics for robot vision and navigation (21–23) before the topic became very popular in the image processing society. RL has been used in the general field of computer vision mainly for object detection (24–26), image segmentation (27, 28), and image enhancement (29–31). However, in medical imaging RL is still in the research phase with very high potential. Netto et al. (32) presented an overview of medical imaging applications applying reinforcement learning with a detailed illustration of a use case involving lung nodules classification which showed promising results. Sahba et al. (27) implemented an RL based thresholding for segmenting prostate in ultrasound images with results that showed high potential for applying RL in medical image segmentation. Alansary et al. (33) evaluated reinforcement learning agents for anatomical landmark detection by comparing fixed and multi-scale search strategies with hierarchical action steps in a coarse-to-fine manner and achieved a performance better than state-of-the-art supervised learning methods.

1.2. Main contributions

In this study, we employ parametric information (T_{max} , rBF, rBV) from MR perfusion images of patients with acute ischaemic stroke and predict the three-level DSA based grading of these patients based on this functional information. We hypothesize that the rich information on blood flow visible from MRI perfusion can be used to predict collateral flow in a similar manner to DSA. Moreover, we argue that this approach, using 3D information, may even offer a more reliable biomarker than the interpretation of DSA images. As collateralization patterns are often unstable and

may undergo significant changes in the course of minutes, a second estimate of the activation of collateral flow using MRI—in addition to the subsequent DSA—will offer better diagnostic information.

We explore machine learning and deep learning methods in collateral flow grading. We apply deep reinforcement learning, a variant of RL which combines the power of deep learning and reinforcement learning, to detect a rigid-sized cube around the occluded region in an acute ischemic stroke patient scan as an initial step toward the prediction of cerebral collateral flow grading. This step is necessary to automate the detection of the occluded region which improves the accuracy of the prediction. Reducing the time spent on this task and ensuring that the proposed methodology is fully automated.

We provide experiments on different feature extraction strategies including denoising autoencoder (DAE), histogram of oriented gradient (HOG), and local binary pattern (LBP). The extracted features are further utilized in a random forest (RF), K-nearest neighbor (KNN), support vector machine (SVM), and convolutional neural network (CNN) classifiers for the prediction of the collateral flow grading. We provide detailed experimental setup and results which will serve as a guide for further research in this direction.

2. Methodology

In this section, we will discuss the details of the steps we employed in predicting the collateral flow grading from MR perfusion data. Figure 1 shows an overview of the main steps involved in the classification process. The first step is the detection of the region of interest (ROI) using reinforcement learning. This step helps to narrow down the classification task to only the area which has been occluded from normal blood flow. The second step deals with extracting features from the ROI. Finally, we feed the extracted features to a set of classifiers to obtain the collateral flow grading for the given patient data.

2.1. Deep reinforcement learning for region of interest detection

The idea of reinforcement learning includes an artificial agent which is trained to interact with an environment through a sequence of reward-based decisions toward a specific goal. At every time step t , the agent takes into account its current state s and performs an action a in a set of actions A and receives a reward r which is a measure of how good or bad the action a is toward the achievement of the set goal. The aim of the agent, which is to find an optimal policy (set of states, actions, and rewards) that maximizes the future reward, can be formulated as a Markov Decision Process. Since Markov Decision Process involves a large number of possible decision points which are normally not fully observable, RL approximates the optimal decision function by iteratively sampling from the set of policies through a process known as Q-learning.

2.1.1. Q-learning

At time point t and state s , let $\pi = a_{i=t}^{t+T}$ be a policy that is a sequence of actions needed by the agent to move from the current state s to the target. Let Q_t be a future discounted reward function such that

$$Q_t(s, \pi) = \sum_{i=t}^{t+T} \gamma^{i-t} r_{\pi i}, \quad (1)$$

where $r_{\pi i}$ is the reward associated with the action a_i of policy π at time $t = i$, $\gamma \in [0, 1]$ is the future reward discounting factor, and T is the number of steps needed to reach the target by the chosen policy π . At any time step t the optimal policy π^* is the policy that maximizes the expected value of Q_t . This can be represented by an action-value function $Q_t^*(s)$ defined by

$$\pi^* = Q_t^*(s) = \max_{\pi} \mathbb{E}[Q_t(s, \pi)] \quad (2)$$

The optimal value function $Q_t^*(s)$ obeys the Bellman equation, stating that if the optimal value $Q_{t+1}^*(s)$ of the next state is known for all possible policies π , then the optimal behavior is to select the policy π^* that maximizes the expected value of $r_{\pi t} + Q_{t+1}(s, \pi)$ [which follows from setting $i = t$ in Equation (1)]. The action-value function can therefore be estimated recursively as

$$Q_t^*(s) = \max_{\pi} \mathbb{E}[r_{\pi t} + Q_{t+1}(s, \pi)] \quad (3)$$

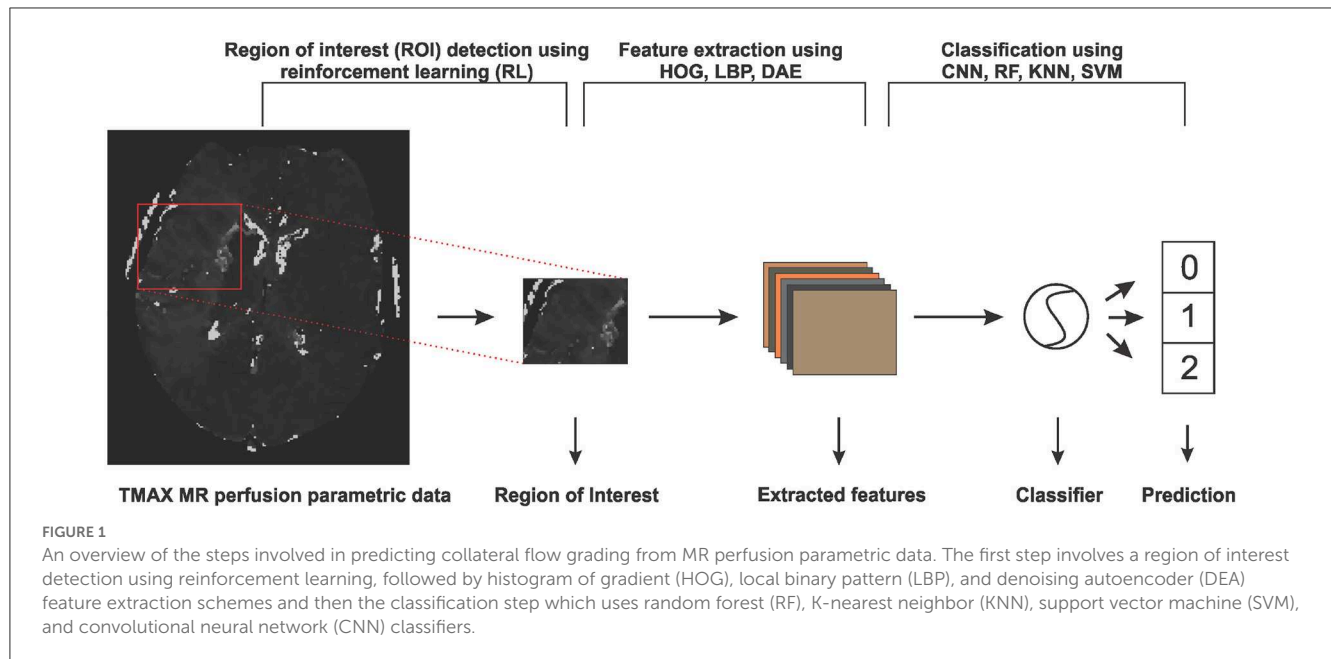
If the problem space is small enough then the set of policies and state can be fully observed and Equation (3) can be used to determine the optimal policy toward the target. However, in most cases, the problem space is too complex to explore, and hence evaluating the future reward for all possible policies is not feasible. $Q_t^*(s)$ is therefore approximated by a non-linear deep network $Q^*(s, \theta)$ with a set of parameters θ resulting in what is known as deep Q-learning (34).

2.1.2. Agent state, action definition, and reward function

Given a 3-D scan as the agent's environment, a state s is represented by (s_x, s_y, s_z) which is the top-left corner of a $(64 \times 64 \times 64)$ cube contained in the 3-D scan. We adopt an agent history approach which involves feeding the last four states visited by the agent to the network to prevent the agent from getting stuck in a loop. Since we have a fixed-sized cube as a state our agent's set of six actions $\{m_u, m_d, m_l, m_r, m_f, m_b\}$ is made up of only movements up, down, left, right, forward, and backward respectively which enables the agent to visit all possible locations within the volume. The agent's reward for taking an action a is a function of the intersection over union (IoU) of the target state s^* and the state before (s_{ab}), and after (s_{aa}) taking the action. This is given by

$$R_a(s_{aa}, s_{ab}) = \text{sign}[IoU(s_{aa}, s^*) - IoU(s_{ab}, s^*)] \quad (4)$$

where sign is the sign function that returns -1 for all values less than 0 and 1 otherwise. This leads to a binary reward ($r \in \{-1, 1\}$) scheme which represents good and bad decisions respectively. During the training stage, the agent search sequence is terminated when the IoU of the current agent's state and the target state is



greater than or equal to a predefined threshold τ . At test time the agent is terminated when a sequence of decisions leads to an oscillation [as proposed by Alansary et al. (33)], that is when the agent visits one state back and forth for a period of time.

Experiments by Alansary et al. (33) and Navarro et al. (35) show that deep reinforcement learning has superior performance in object detection as compared to classical supervised learning, especially in images with a noisy background. RL agents also require lesser training data as compared to other supervised learning methods like CNN. These proven advantages make deep reinforcement learning the right choice for our limited and noisy data.

2.2. Feature extraction and classification

Feature extraction methods are used in many machine learning tasks to either reduce the dimension of the problem or to extract information from the raw input which would otherwise not be easily extracted by the underlying classifier. In this work, we extract two main classes of features—learned features through a denoising auto-encoder (DAE), and local image descriptors made up of histogram of oriented gradients (HOG) and local binary pattern (LBP).

2.2.1. Denoising auto-encoder

An auto-encoder is an unsupervised deep learning method used for dimension reduction, feature extraction, image reconstruction or denoising and is sometimes also used as a pre-training strategy in supervised learning networks. An auto-encoder is made up of two parts: an encoder $\Phi: \mathcal{X} \rightarrow \mathcal{F}$ which maps an image $x \in X$ to $f_x \in \mathcal{F}$ in the features domain and a decoder $\Psi: \mathcal{F} \rightarrow \mathcal{X}$ which maps a feature set $f \in \mathcal{F}$ to $x_f \in X$. The full auto-encoder is therefore a composite function of the form $\Psi \circ \Phi: \mathcal{X} \rightarrow \mathcal{X}$. Let

$\hat{y} = \Psi(\Phi(x))$ for a given input image $x \in \mathcal{X}$, then the learning process of auto-encoder involves finding a pair of $\{\Phi, \Psi\}$ such that $\hat{y}_i = x_i$ for all $x_i \in \mathcal{X}$. The encoder Φ then becomes the feature extractor which is used for extracting the needed features.

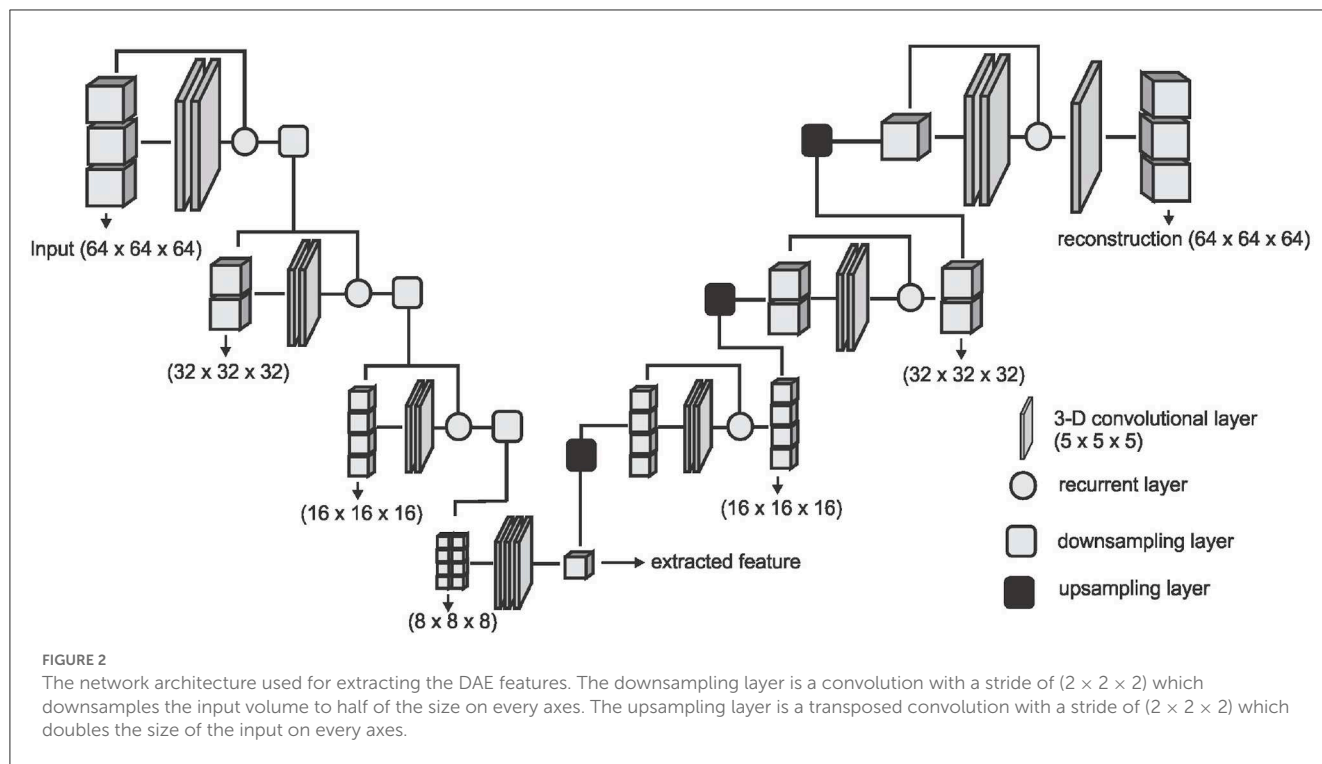
If the function Φ is invertible, then the learning process can lead to a trivial solution by just choosing Ψ to be the inverse of Φ , and $\Psi \circ \Phi$ becomes an identity function leading to what is known as identity-function risk. To prevent this, the input image x is first corrupted by adding noise before feeding it to Φ leading to a denoising auto-encoder. We therefore have

$$\hat{y} = \Psi(\Phi(\tilde{x})), \quad \tilde{x} = \Upsilon(x) \quad (5)$$

where Υ is the random image corruption function. We approximate the encoder and decoder by deep CNNs $E(x, \theta_e)$ and $D(f, \theta_d)$ parameterized by θ_e and θ_d ; respectively. Training is done through back-propagating the Mean Squared Error (MSE) of the original image x and the reconstructed image \hat{y} given by

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - x_i)^2 \quad (6)$$

where N is the number of images in the training set or training batch. We adopt the V-Net architecture proposed by Milletari et al. (36) and simplify it by removing the fine-grained feature forwarding, and reducing the depth of the network due to limitations on the amount of training data available. The downsampling layers of the VNET architecture represent the encoding part $[E(x, \theta_e)]$ of the DAE and the upsampling layers represent the decoding part $[D(f, \theta_d)]$ of the DAE. Figure 2 shows an overview of the simplified architecture used for extracting the DAE features.



2.2.2. Local image descriptors and classifiers

We consider two types of local image descriptors - histograms of oriented gradients (HOG) and a local binary pattern (LBP). Given a volume X , we extract the LBP encoding of each voxel by thresholding its $3 \times 3 \times 3$ neighborhood by the intensity value p^* of the center voxel which results in 26 long bits $b_0, b_1, b_2, \dots, b_{25}$ where $b_i = \{1, \text{if } p_i \geq p^*, 0 \text{ otherwise}\}$ and p_i is the intensity value of the i th neighbor. We then concatenate the binary encoding to a single binary number $b_0b_1b_2\dots b_{25}$ and then into a decimal value which results in 2^{25} possible binary codes. Details of the implementation until this point can be found in Heikkilä and Pietikäinen (37). We group the codes into two main classes—uniform codes which have at most two binary transitions and non-uniform codes which have more than two binary transitions. A binary transition is a switch from 0 to 1 or vice versa. For example the codes 0000, 000111, 011100, and 110110 have zero, one, two, and three transitions respectively. To handle noisy data and to reduce the feature space, we group all the non-uniform codes into one class and add it to the uniform codes resulting in 927 codes instead of 2^{25} . Finally, the histogram distribution of the individual codes is extracted as the LBP feature representation for the volume X .

We also explore the HOG feature extractor based on the method proposed in Klašer et al. (38). Given a volume X , we quantize gradient orientations over an icosahedron and merge opposite directions in one bin resulting in 10 gradient orientations. The gradient for each voxel $x_i \in X$ is obtained by convolving the $5 \times 5 \times 5$ neighborhood of the voxel by gradient filters k_x, k_y , and k_z of the same size, giving us a gradient vector $\vec{x}_i \in \mathbb{R}^3$. The gradient filters are zero everywhere except for the center columns along the respective axes $k_x(i, 3, 3) = k_y(3, i, 3) = k_z(3, 3, i) = [1, 0, -2, 0, 1]$ for $i \in \{1, 2, \dots, 5\}$. The gradient vectors \vec{x}_i are then projected to the gradient orientations and a histogram representation of

these orientations are obtained and used as the HOG feature representation of the volume X .

We run experiments with four machine learning classifiers on each of the features extracted. We implement Convolutional Neural Network (CNN), Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN) classifiers. Our CNN classifier in Figure 3 has four convolutional layers, aimed at extracting local image features, followed by two fully connected layers and a sigmoid layer for classification. Each layer is followed by a non-linear hyperbolic tangent (tanh) activation function. For classification based on the HOG, LBP, and DEA features, we remove the convolutional layers and feed the features directly to the fully connected layers and then the sigmoid layer for the classification. For the RF, SVM, and KNN classifiers we use the implementation of these classifiers from the Scikit-Learn library (39) in python.

3. Experiments and results

3.1. Patient population and image data

We test our proposed methods on parametric volumes extracted from MR perfusion data from 183 patients with acute ischemic stroke. Details of the image acquisition and preparation are already published by Pinto et al. (40). Our dataset is made up of three parametric information— T_{max} volumes which refer to the time taken for the blood flow to reach its peak, relative blood flow (rBF) volumes which refer to the volume of blood passing through a given brain tissue per unit of time, and relative blood volume (rBV) defined as the volume of blood in a given brain tissue relative to an internal control (e.g., normal white matter or an arterial input function). Each volume has a resolution of (0.9, 0.9, and 6.5 mm)

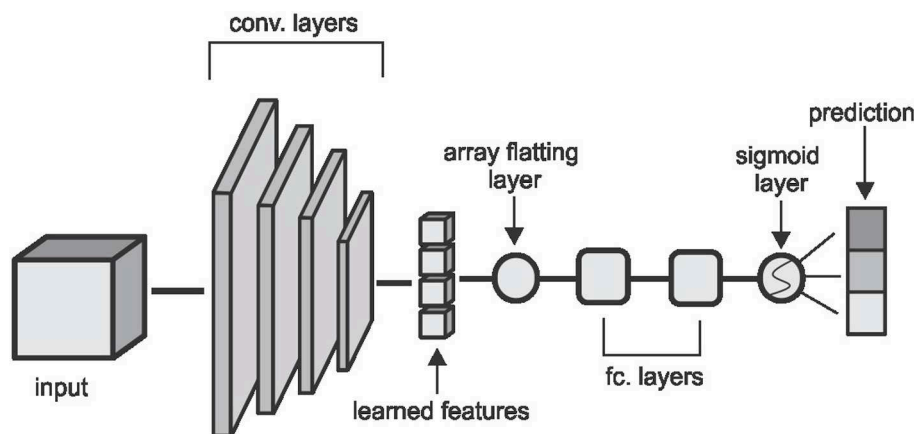


FIGURE 3
The CNN architecture used in the classification task. Convolutional layers are made up of $(5 \times 5 \times 5)$ kernels with a stride of $(2 \times 2 \times 2)$ which reduces the volume by half of the input size on each layer. The first two layers extract 2 feature cubes and the last two layers extract 4 feature cubes each. The fully connected layers have 64 and 32 hidden nodes respectively and the convolutional and fully connected layers are followed by a non-linear hyperbolic tangent (tanh) activation function.

and a dimension of (256, 256, and 19) voxels on the sagittal, coronal and axial planes respectively. Ground truth labels are obtained from a trained neuroradiologist, with over ten years of experience, who manually investigates the DSA slides of the associated patient and assigns one of three labels (0-poor, 1-medium, 2-good) to this patient. We use these labels for a 3-class prediction experiment and we also experiment on a risk-stratified nested test where we first predict good - (2) against not good (0, 1) collaterals and then separate the not good class into poor (0) and medium (1) collaterals in a cascaded approach.

3.2. Preprocessing

Our image preprocessing involves two main tasks. First, we make our datasets isotropic by applying a B-spline interpolation to the axial axis since the other two axes have the same spacing leading to volume with a resolution of 0.9 mm on each plane and a new dimension of (256, 256, and 127). This is followed by an extraction of the brain region from the skull using the brain extraction tool (BET) from the ANTS library. The brain extraction is carried out on the T_{max} volumes and the resulting mask is then applied to the rBF and rBV volumes.

3.3. Region of interest localization

After the preprocessing step we extract the occluded regions as the region of interest (ROI) using the reinforcement learning architecture described in Section 2.1. We adopt the network architecture from Alansary et al. (33) with modifications proposed in Navarro et al. (35). A stopping criterion of $\tau = 0.85$ is used during training—that is, an intersection over union (IoU) value greater than or equal to 0.85 implies that the region of interest is detected. We perform the ROI detection task on the T_{max} volumes since the occluded regions are easier to detect in these volumes. The

TABLE 1 Quantitative results from the region of interest detection task.

Type	Class	Mean	Std	Max	Min
IoU	0	0.49	0.22	0.79	0.08
	1	0.52	0.14	0.81	0.09
	2	0.42	0.21	0.81	0.04
Center points displacement (in voxels)	0	20	13	51	5
	1	17	9	52	4
	2	23	14	63	5

IoU refers to the intersection over union ratio between the prediction and the ground truth. Center point displacement is the euclidean distance between the predicted center point and the ground truth center point.

resulting cube region is then applied on the rBF and rBV volumes to extract the corresponding cubes in these volumes as well. For each volume, we select 20 starting cubes of size $(64 \times 64 \times 64)$ at random and run the agent till the stopping criterion is reached. We then aggregate the results from the 20 different runs to get the prediction of the final ROI. After getting the region we extract the mirror of the ROI (ROI+M) by reflecting the ROI on the opposite side of the brain and using it as an additional feature. This results in 6 cubes per patient (i.e., two volumes each from T_{max} , rBF, and rBV volumes). Qualitative and quantitative results from the region of interest extraction can be found in Table 1 and Figures 4, 5. From the box plots in Figure 5, it is evident that the region of interest detection was more successful in the poor collateral flow classes (class 0 and 1) than in the good collateral flow class. This can be explained by the fact that in cases of good collateral flow, there is a uniform distribution of the T_{max} value within the occluded region and its neighborhood making it hard for the RL agent to detect the ROI. From Figure 4 we observe that in most cases the ground truth does not cover the total occluded region [e.g., column (b)] and hence the predicted ROI, though does not completely overlap the ground truth, still contains other parts of the occluded region which

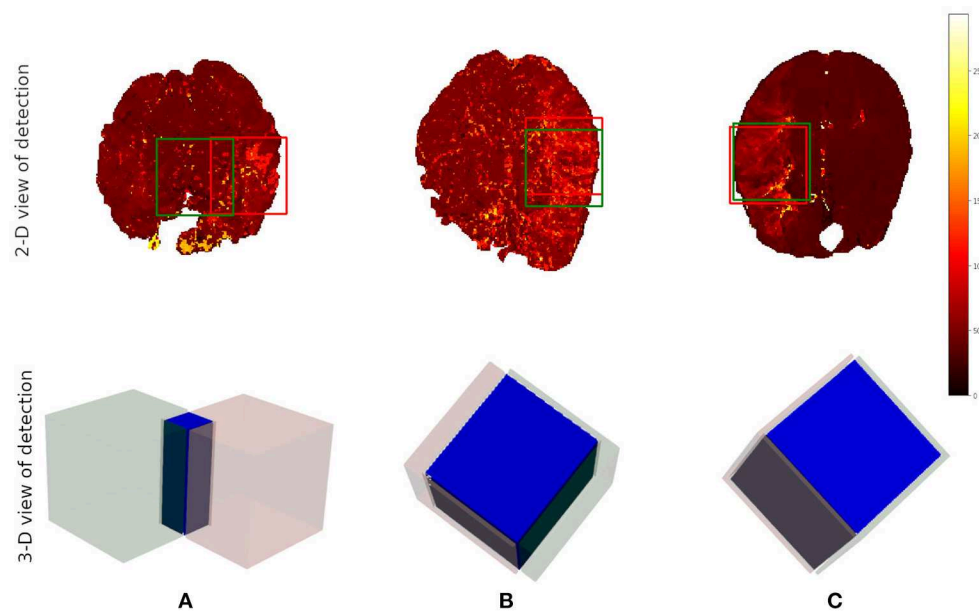


FIGURE 4

Qualitative results from the ROI detection task. The top row is the axial view of the ground truth (in red) and the prediction (in green). The bottom row is a 3-D visualization of the ground truth cube (in red), the predicted cube (in green), and the intersection between the two (in blue). Column (A) corresponds to the worst prediction in our test set while column (C) refers to the best result in terms of IoU. In column (B), we can observe that though the overlap is not perfect the prediction still contains some part of the occluded region which is not in the ground truth. This implies that though we have poor scores we still have good ROI detection which can be used for the classification task.

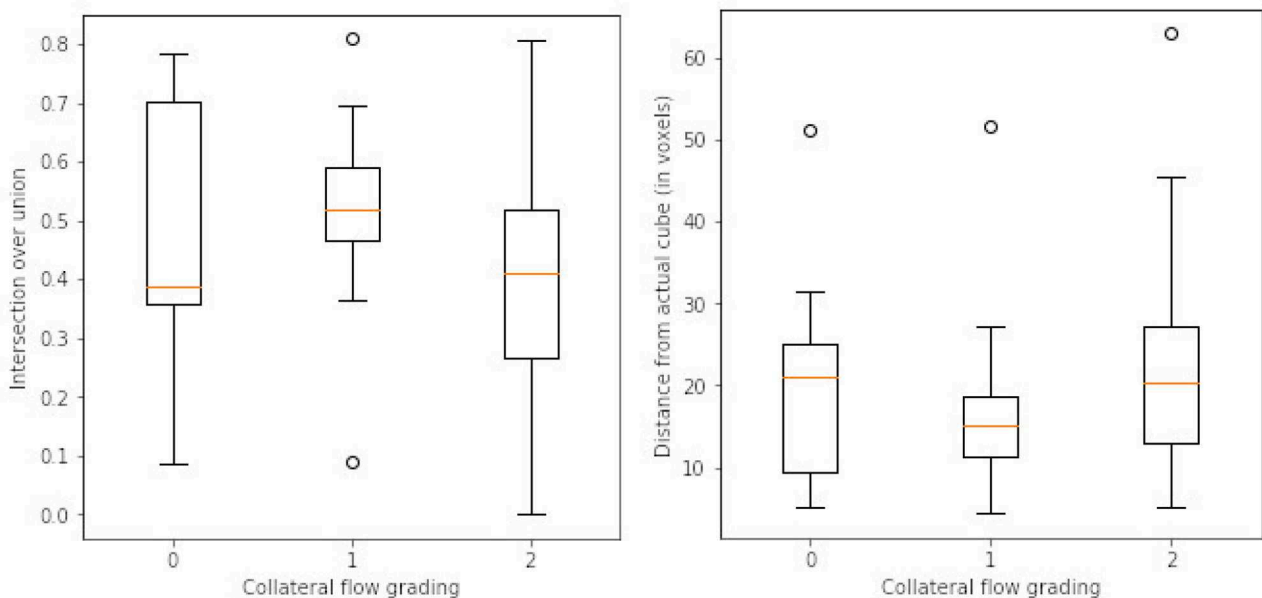


FIGURE 5

Box plots of results from ROI detection task. **Left** is the intersection over union (IoU) ratio between the prediction and the ground truth over the three classes. **Right** is the euclidean distance between the predicted center point and the ground truth center point. From the distributions, it is clear that it is easy to detect the ROI in the poor collateral flow class (class 0) compared to the good collateral flow class (class 2). This can be explained by the fact that in good collateral flow cases T_{max} shows uniform values in the whole volume.

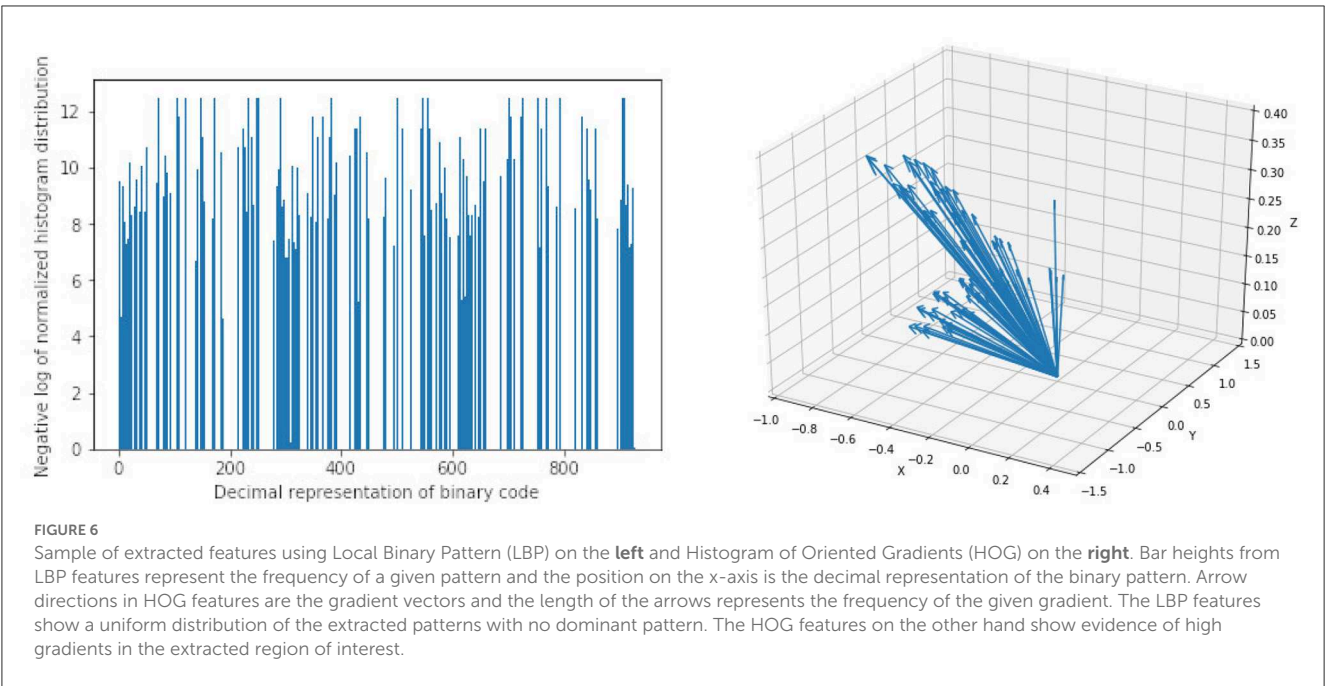


TABLE 2 Results from preliminary experiments on collateral flow grading.

Type	Method	RAW	ROI	ROI+M	DAE	HOG	LBP
Three classes	CNN+MLP	0.51(±0.04)	0.63(±0.06)	0.65(±0.03)	0.50(±0.07)	0.38(±0.13)	0.25(±0.14)
	RF	0.51(±0.02)	0.65(±0.04)	0.67(±0.05)	0.66(±0.04)	0.69(±0.02)	0.60(±0.05)
	KNN	0.48(±0.10)	0.54(±0.02)	0.58(±0.05)	0.55(±0.06)	0.59(±0.02)	0.43(±0.04)
	SVM	0.56(±0.04)	0.66(±0.05)	0.70(± 0.03)	0.70(±0.04)	0.53(±0.02)	0.25(±0.15)
Cascaded (two step)	CNN+MLP	0.55(±0.01)	0.72(± 0.05)	0.70(±0.04)	0.66(±0.05)	0.54(±0.02)	0.21(±0.13)
	RF	0.47(±0.07)	0.67(±0.03)	0.64(±0.03)	0.65(±0.04)	0.70(±0.03)	0.56(±0.07)
	KNN	0.44(±0.04)	0.55(±0.06)	0.56(±0.07)	0.52(±0.08)	0.60(±0.05)	0.48(±0.08)
	SVM	0.38(±0.02)	0.51(±0.04)	0.46(±0.04)	0.51(±0.04)	0.46(±0.04)	0.10(±0.00)

RAW features refer to the full-sized three parametric volumes (T_{max} , rBF, and rBV) after skullstripping. ROI refers to the corresponding cubes extracted from the parametric volumes based on the manually annotated ROI and ROI+M is the ROI combined with its mirror cube on the opposite side of the brain. Other features (DAE, HOG, and LBP) are all extracted from the ROI cubes. Scores represent mean accuracy over the 5-fold cross-validation experiments with their corresponding standard deviation in parenthesis. Values in bold refer to the feature-classifier combination with the highest accuracy under each experiment type.

is not captured in the ground truth and it is therefore sufficiently accurate for the classification task.

3.4. Classification

3.4.1. Feature representations

In total three sets of features (DAE, HOG, and LBP) are extracted in addition to the actual extracted cube (ROI) and its mirror cube (ROI+M). We learn features automatically through an unsupervised denoising auto-encoder. The network takes the extracted ROI cubes from the T_{max} , rBV, and rBF volumes as three input channels and produces a single channel feature set of size ($8 \times 8 \times 8$). We normalize the cubes individually into the range [0, 1] before feeding them to the network.

For HOG features we extract 10 features each for the three parametric volumes and concatenate them into a vector of length 30 for the classification task. Figure 6 shows a sample of the extracted

HOG features for a patient for the three input channels. Finally, LBP features are extracted using the method described in Section 2.2. Here we combine all the three channels and run the histogram over the three channels which results in a 927 feature vector as explained in Section 2.2. Figure 6 shows a sample of the extracted LBP features from our dataset.

3.4.2. Classifier training

We handle the collateral flow classification through two main approaches—a three-class multi-label classification task where we predict three labels in one step, and a two-step cascaded approach where we predict a binary label of classes (0, 1) against 2 in the first step and separate the class 0 from 1 in the second step. We implement our CNN architecture using the Keras library (41) in python with TensorFlow as the backend. Random forest, support vector machine, and K-nearest neighbor classifiers were

implemented using the scikit-learn library (39) in python. We set up our experiments as follows:

CNN classifier: For the CNN classifier, we use a weighted categorical cross-entropy with a weight of $\frac{1}{|k|}$ for each class k in the training set. A stochastic gradient descent optimizer with a learning rate of 0.001, decay of $1e^{-6}$, and momentum of 0.9 is used to fine-tune the network parameters at 20 epochs.

K-Nearest neighbor classifier: We conduct preliminary a experiment with a grid search to know which parameters will work best. For our final experiment, we use $k = 3$ neighbors with uniform weights, a leaf size of 30, and the Minkowski metric.

Random forest classifier: After the initial grid search experiment, we implement the classifier with 200 estimators, and the Gini impurity function is used to measure the quality of a split.

Support vector machine classifier: We use a regularization parameter $C = 10$, a third-degree polynomial kernel, a balanced class weight, and a tolerance of $1e^{-3}$ for the stopping criterion.

3.4.3. Classification results

We test different combinations of the feature sets extracted in the previous experiments and classifiers discussed in a preliminary experiment and present the results in Table 2. Due to limitations in the size of the dataset, we adopt a 5-fold cross-validation approach in a preliminary experiment instead of a training-validation-test splitting approach and report the average scores over the accuracy in the individual validations. In the preliminary experiments (results in Table 2), we use the manually annotated ROI and not the ROI predicted from the proposed reinforcement learning. We later, in a follow-up experiment, compare the performance of the proposed CNN on manually annotated ROI and the predicted ROI (results in Table 3).

The results in Table 2 show that the region of interest extraction step helps improve the results in all classification methods. This can be verified by comparing the performance from the full image (RAW column) with the performance of the region of interest (ROI column) in Table 2. Also by adding the mirror of the occluded region to the extracted ROI (ROI+M) we achieve improved results in most of the classifiers with performance falling in classifiers like KNN and SVM due to the increase in the dimension of data introduced by the mirror of the ROI. The cascaded method shows higher accuracy in almost all the classifier-feature combinations when compared to the direct three-class prediction. This can be explained by the distribution of classes in the dataset. That is, for the cascaded approach we have fairly balanced data when we combine poor and moderate flow against good collateral flow which is not the case with the direct three-class multi-label prediction approach. It, therefore, suggests that in cases where we have highly imbalanced class distributions a multi-label classification might perform poorly. The overall performance of CNN is better than the other machine learning classifiers and can be explained by the fact that the convolutional layers of the CNN architecture extract features while paying attention to the class of the input data. This makes the feature extraction process more efficient than the other feature extraction schemes which have no knowledge of the underlying label of the input data at the time of extracting the features. Again CNN with only ROI data performs slightly better

TABLE 3 Results from the experiment on collateral flow grading using only ROI data on our proposed cascaded CNN.

Input data	Binary	Three classes
Manual ROI	0.84	0.74
Automated ROI	0.80	0.72

Manual ROI refers to the ground truth ROI and automated ROI refers to the predicted ROI from our proposed Reinforcement Learning approach. Binary refers to the result from the first binary classification (i.e., {0,1} vs. 2) and three classes is the three-class classification based on the cascaded networks.

than with the mirror of the ROI (72 vs. 70% in Table 2) and this can also be explained by the fact that the CNN used in our experiments is fairly shallow and hence could not handle the additional feature dimensions introduced by the mirrored images.

Based on the results of the preliminary experiment, we further probe into the training of the proposed CNN classifier with the ROI data. In this experiment, we split the data into training and testing sets. The test set is made up of 50 volumes randomly selected with reference to the ratio of class count in the entire dataset. We make use of both the manually annotated ROI and the automated ROI from our proposed Reinforcement Learning approach during training. We finally evaluate the trained models on the automated ROI and compare it with the same network trained and evaluated solely on the manually annotated ROI data. Table 3 shows the result of this experiment.

The results in Table 3 from our follow-up experiment show that the automated ROI from the proposed Reinforcement Learning approach is comparable to the manually detected ROI in terms of predicting collateral flow (2% drop in accuracy which represents one out of the 50 patients in the test set). This is crucial in automating the whole collateral flow prediction workflow in a clinical setting.

4. Summary and conclusion

In this work, we present a deep learning approach toward grading collateral flow in ischemic stroke patients based on parametric information extracted from MR perfusion data. We start by extracting regions of interest using deep reinforcement learning. We then learn denoising auto-encoder features and modern implementation of 3-D HOG and LBP features. We proceed to the actual classification task using a combination of the extracted features and CNN, random forest, K-nearest neighbor, and support vector machine classifiers.

Our experiments show that the rich information on blood flow visible from MRI perfusion can be used to predict collateral flow in a similar manner to DSA images which are invasive in nature. Region of interest detection with reinforcement learning is successful to an acceptable level and can be used as a guide to estimate the region in the brain which requires more attention. It is evident that high class imbalance can be a major challenge in the collateral flow grading task and many similar works. We however show that for datasets with high class imbalance, a two-step cascaded classification approach performs better than a one-time multi-label classification method. It is also evident from our results that a direct CNN classifier is able to extract relevant features

from the region of interest and has an advantage over classical machine learning classifiers like RF, KNN, and SVM that depend on handcrafted features like HOG and LBP.

Collateral flow grading is an essential clinical procedure in the treatment of ischemic stroke patients. We have presented a framework for automating the process in clinical setup and have achieved promising results given our limited dataset. For the proposed framework to be clinically useful there is the need for further tests with possibly more data from multiple stroke centers. The grading can also be customized for specific patient groups for example providing information about age group, gender, and other biographical and historical information of patients as an additional feature can help improve the result of the framework.

Data availability statement

Data is currently private, pending approval from the source of the data. Requests to access these datasets should be directed to giles.tetteh@tum.de.

Author contributions

GT and BM contributed to the conception and design of the study. GT, FN, and JP contributed to the experiments and computer codes. RM, JK, and RW organized the database. GT wrote the first draft of the manuscript. BM, JP, and JK contributed to the final manuscript. BM, CZ, JK, and RW handled the administrative tasks. All authors

contributed to manuscript revision, read, and approved the submitted version.

Funding

GT and BM acknowledge support by the Deutsche Forschungsgemeinschaft (DFG) (ME 3511/5-1).

Acknowledgments

BM acknowledges support from the Helmut-Horten-Foundation.

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Faber J, Chilian W, Deindl E, van Royen N, Simons M. A brief etymology of the collateral circulation. *Arterioscler Thromb Vasc Biol.* (2014) 9:1854–9. doi: 10.1161/ATVBAHA.114.303929
2. Liu L, Go T, Leng X, Pu Y, Huang LA, Xu A, et al. Guidelines for evaluation and management of cerebral collateral circulation in ischaemic stroke 2017. *Stroke Vasc Neurol.* (2018) 3:svn-2017. doi: 10.1136/svn-2017-000135
3. Ginsberg M. The cerebral collateral circulation: Relevance to pathophysiology and treatment of stroke. *Neuropharmacology.* (2017) 134:280–92. doi: 10.1016/j.neuropharm.2017.08.003
4. Ben Hassen W, Malley C, Boulouis G, Clarençon F, Bartolini B, Bourcier R, et al. Inter- and intraobserver reliability for angiographic leptomeningeal collateral flow assessment by the American Society of Interventional and Therapeutic Neuroradiology/Society of Interventional Radiology (ASITN/SIR) scale. *J Neurointerv Surg.* (2018) 11:338–41. doi: 10.1136/neurintsurg-2018-014185
5. McVerry F, Liebeskind D, Muir K. Systematic review of methods for assessing leptomeningeal collateral flow. *AJNR Am J Neuroradiol.* (2011) 33:576–82. doi: 10.3174/ajnr.A2794
6. Martinon E, Lefevre P, Thouant P, Osseby G, Ricolfi F, Chavent A. Collateral circulation in acute stroke: Assessing methods and impact: A literature review. *J Neuroradiol.* (2014) 41:97–107. doi: 10.1016/j.neurad.2014.02.001
7. Sheth SA, Liebeskind DS. Imaging evaluation of collaterals in the brain: physiology and clinical translation. *Curr Radiol Rep.* (2014) 29:1–9. doi: 10.1007/s40134-013-0029-5
8. Jansen IGH, Berkhemer OA, Yoo AJ, Vos JA, Lycklama A, Nijeholt GJ, et al. Comparison of CTA- and DSA-based collateral flow assessment in patients with anterior circulation stroke. *AJNR Am J Neuroradiol.* (2016) 37:2037–42. doi: 10.3174/ajnr.A4878
9. Muehlen I, Kloska SP, Göllitz P, Hölter P, Breuer L, Ditt H, et al. Noninvasive collateral flow velocity imaging in acute ischemic stroke: intraindividual comparison of 4D-CT angiography with digital subtraction angiography. In: *RöFo-Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*. Georg Thieme Verlag KG. (2019). Vol. 191, p. 827–835. doi: 10.1055/a-0825-6660
10. Kauw F, Dankbaar JW, Martin BW, Ding VY, Boothroyd DB, van Ommen F, et al. Collateral status in ischemic stroke: a comparison of computed tomography angiography, computed tomography perfusion, and digital subtraction angiography. *J Comput Assist Tomogr.* (2020) 44:984–92. doi: 10.1097/RCT.00000000000001090
11. Jung S, Wiest R, Gralla J, McKinley R, Mattie H, Liebeskind D. Relevance of the cerebral collateral circulation in ischaemic stroke: Time is brain, but collaterals set the pace. *Swiss Med Wkly.* (2017) 147:w14538. doi: 10.4414/sm.w.2017.14538
12. Maija R. *Collateral blood supply in acute stroke.* (2015). doi: 10.1594/ecr2015/C-1460
13. Bang OY, Saver J, Kim S, Kim GM, Chung CS, Ovbiagele B, et al. Collateral flow predicts response to endovascular therapy for acute ischemic stroke. *Stroke.* (2011) 42:693–9. doi: 10.1161/STROKEAHA.110.595256
14. Bang OY, Saver J, Buck B, Alger JR, Starkman S, Ovbiagele B, et al. Impact of collateral flow on tissue fate in acute ischaemic stroke. *J Neurol Neurosurg Psychiatry.* (2008) 79:625–9. doi: 10.1136/jnnp.2007.132100
15. Mansour O. Assessment of collateral pathways in acute ischemic cerebrovascular stroke using a mansour grading scale; a new scale, a pilot study. *Internet J Internal Med.* (2013) 3:1–10.
16. Maas M, Lev M, Ay H, Singhal A, Greer D, Smith W, et al. Collateral vessels on CT angiography predict outcome in acute ischemic stroke. *Stroke.* (2009) 40:3001–5. doi: 10.1161/STROKEAHA.109.552513

17. Tan I, Demchuk A, Hopyan J, Zhang L, Gladstone D, Wong K, et al. CT angiography clot burden score and collateral score: correlation with clinical and radiologic outcomes in acute middle cerebral artery infarct. *AJNR Am J Neuroradiol*. (2009) 30:525–31. doi: 10.3174/ajnr.A1408
18. McHugh ML. Interrater reliability: the kappa statistic. *Croatian Soc Med Biochem Lab Med*. (2012) 22:276–82. doi: 10.11613/BM.2012.031
19. Kersten-Oertel M, Alamer A, Fonov V, Lo BWY, Tampieri D, Collin DL. Towards a computed collateral circulation score in ischemic stroke: evaluating automated scoring of collaterals in acute stroke on computed tomography scans. *Cerebrov Dis*. (2019) 47:1–6. doi: 10.1159/000500076
20. Grunwald IQ, Kulikovski J, Reith W, Gerry S, Namias R, Politi M, et al. Collateral automation for triage in stroke: evaluating automated scoring of collaterals in acute stroke on computed tomography scans. *Cerebrov Dis*. (2019) 47:1–6. doi: 10.1159/000500076
21. Pauli J. *Learning-Based Robot Vision, Principles and Applications*. Berlin: Springer Science & Business Media. (2001). doi: 10.1007/3-540-45124-2
22. DeSouza G, Kak A. Vision for mobile robot navigation: a survey. *IEEE Trans Pattern Anal Mach Intell*. (2002) 24:237–67. doi: 10.1109/34.982903
23. Bonin-Font F, Ortiz A, Oliver G. Visual navigation for mobile robots: a survey. *J Intell Robot Syst*. (2008) 53:263–96. doi: 10.1007/s10846-008-9235-4
24. Peng J, Bhanu B. Delayed reinforcement learning for closed-loop object recognition. In: *Proceedings of 13th International Conference on Pattern Recognition*. IEEE. (1996). p. 310–314.
25. Peng J, Bhanu B. Closed-loop object recognition using reinforcement learning. *IEEE Trans Pattern Anal Mach Intell*. (1998) 20:139–54. doi: 10.1109/34.659932
26. Taylor G, Wolf C. Reinforcement learning for parameter control of text detection in images from video sequences. In: *Proceedings 2004 International Conference on Information and Communication Technologies: From Theory to Applications*. (2004). p. 517–518.
27. Sahba F, Tizhoosh HR, Salama MM. A reinforcement learning framework for medical image segmentation. In: *The 2006 IEEE international joint conference on neural network proceedings*. IEEE (2006). p. 511–517. doi: 10.1109/IJCNN.2006.246725
28. Sahba F, Tizhoosh HR, Salama MM. Application of opposition-based reinforcement learning in image segmentation. In: *2007 IEEE Symposium on Computational Intelligence in Image and Signal Processing*. IEEE (2007). p. 246–251. doi: 10.1109/CIISP.2007.369176
29. Shokri M, Tizhoosh HR. Using reinforcement learning for image thresholding. In: *CCECE 2003-Canadian Conference on Electrical and Computer Engineering*. IEEE (2003). p. 1231–1234.
30. Sahba F, Tizhoosh HR, Salama MM. Using reinforcement learning for filter fusion in image enhancement. In: *IASTED International Conference on Computational Intelligence*. (2005). p. 262–266.
31. Tizhoosh H, Taylor G. Reinforced contrast adaptation. *Int J Image Graphics*. (2006) 6:377–92. doi: 10.1142/S0219467806002379
32. Netto SM, Leite VR, Silva AC, de Paiva AC, de Almeida Neto A. Application on reinforcement learning for diagnosis based on medical image. In: *Reinforcement Learning*. IntechOpen (2008) 379.
33. Alansary A, Oktay O, Li Y, Folgoc L, Hou B, Vaillant G, et al. Evaluating reinforcement learning agents for anatomical landmark detection. *Med Image Anal*. (2019) 53:156–64. doi: 10.1016/j.media.2019.02.007
34. Mnih V, Kavukcuoglu K, Silver D, Rusu A, Veness J, Bellemare M, et al. Human-level control through deep reinforcement learning. *Nature*. (2015) 518:229–33. doi: 10.1038/nature14236
35. Navarro F, Sekuboyina A, Waldmannstetter D, Peekan JC, Combs SE, Menze BH. Deep Reinforcement Learning for Organ Localization in CT. *Proc Mach Learn Res*. (2020) 121:544–54.
36. Milletari F, Navab N, Ahmadi SA. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: *2016 Fourth International Conference on 3D vision (3DV)*. IEEE (2016). p. 565–571. doi: 10.1109/3DV.2016.79
37. Heikkilä M, Pietikäinen M. A texture-based method for modeling the background and detection moving objects. *IEEE Trans Pattern Anal Mach Intell*. (2006) 28:657–62. doi: 10.1109/TPAMI.2006.68
38. Klaser A, Marszaäek M, Schmid C. A spatio-temporal descriptor based on 3d-gradients. In: *BMVC 2008-19th British Machine Vision Conference*. British Machine Vision Association (2008). p. 275–1. doi: 10.5244/C.22.99
39. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in python. *J Mach Learn Res*. (2011) 12:2825–30.
40. Pinto A, Amorim J, Hakim A, Alves V, Reyes M, Silva CA. Prediction of stroke lesion at 90-day follow-up by fusing raw DSC-MRI with parametric maps using deep learning. *IEEE Access*. (2021) 9:26260–70. doi: 10.1109/ACCESS.2021.3058297
41. Chollet F. Keras: GitHub (2015). Available online at: <https://github.com/fchollet/keras> (accessed August 31, 2022).



OPEN ACCESS

EDITED BY

Daniel Donoho,
Children's National Hospital, United States

REVIEWED BY

Matthias Wilms,
University of Calgary, Canada
Jesse Klostranec,
McGill University Health Centre, Canada
Alex Buioite Stella,
University of Trieste, Italy

*CORRESPONDENCE

Freda Werdiger
✉ freda.werdiger@unimelb.edu.au

SPECIALTY SECTION

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

RECEIVED 20 November 2022

ACCEPTED 02 February 2023

PUBLISHED 23 February 2023

CITATION

Werdiger F, Parsons MW, Visser M, Levi C,
Spratt N, Kleinig T, Lin L and Bivard A (2023)
Machine learning segmentation of core and
penumbra from acute stroke CT perfusion data.
Front. Neurol. 14:1098562.
doi: 10.3389/fneur.2023.1098562

COPYRIGHT

© 2023 Werdiger, Parsons, Visser, Levi, Spratt,
Kleinig, Lin and Bivard. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).
The use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Machine learning segmentation of core and penumbra from acute stroke CT perfusion data

Freda Werdiger^{1,2*}, Mark W. Parsons^{3,4,5}, Milanka Visser^{1,2},
Christopher Levi^{6,7}, Neil Spratt^{6,7}, Tim Kleinig⁸, Longting Lin^{6,7} and
Andrew Bivard^{1,2}

¹Melbourne Brain Centre, Department of Neurology, The Royal Melbourne Hospital, Melbourne, VIC, Australia, ²Department of Medicine, University of Melbourne, Melbourne, VIC, Australia, ³Southwestern Sydney Clinical School, University of New South Wales, Sydney, NSW, Australia, ⁴Department of Neurology, Liverpool Hospital, Liverpool, NSW, Australia, ⁵Ingham Institute for Applied Medical Research, Liverpool, NSW, Australia, ⁶Hunter Medical Research Institution, University of Newcastle, Newcastle, NSW, Australia, ⁷Department of Neurology, John Hunter Hospital, University of Newcastle, Newcastle, NSW, Australia, ⁸Department of Neurology, Royal Adelaide Hospital, Adelaide, SA, Australia

Introduction: Computed tomography perfusion (CTP) imaging is widely used in cases of suspected acute ischemic stroke to positively identify ischemia and assess suitability for treatment through identification of reversible and irreversible tissue injury. Traditionally, this has been done *via* setting single perfusion thresholds on two or four CTP parameter maps. We present an alternative model for the estimation of tissue fate using multiple perfusion measures simultaneously.

Methods: We used machine learning (ML) models based on four different algorithms, combining four CTP measures (cerebral blood flow, cerebral blood volume, mean transit time and delay time) plus 3D-neighborhood (patch) analysis to predict the acute ischemic core and perfusion lesion volumes. The model was developed using 86 patient images, and then tested further on 22 images.

Results: XGBoost was the highest-performing algorithm. With standard threshold-based core and penumbra measures as the reference, the model demonstrated moderate agreement in segmenting core and penumbra on test images. Dice similarity coefficients for core and penumbra were 0.38 ± 0.26 and 0.50 ± 0.21 , respectively, demonstrating moderate agreement. Skull-related image artefacts contributed to lower accuracy.

Discussion: Further development may enable us to move beyond the current overly simplistic core and penumbra definitions using single thresholds where a single error or artefact may lead to substantial error.

KEYWORDS

acute ischemic stroke, CT perfusion imaging, machine learning, ischemic core, penumbra

1. Introduction

Rapid diagnosis of acute ischemic stroke is of vital importance and is confirmed by computed tomography (CT) or magnetic resonance (MR) imaging. Historically improved patient outcomes were obtained by early reperfusion treatment, with significant effort and resources being provided to improve both stroke detection and clinical workflows to facilitate faster treatment (1–3). Recently, clinical trials have demonstrated that patients with a favorable perfusion imaging profile benefit from treatment up to 9 h from symptom onset/mid-point of wake-up with thrombolysis and up to 24 h with thrombectomy (4–7). Perfusion imaging allows estimation of salvageable brain tissue (penumbra) and tissue

already infarcted or destined for infarction irrespective of reperfusion (ischemic core) (4, 7–11). Patient outcomes have been shown to be strongly related to the estimated volume of ischemic core at baseline (12, 13). As a result, CT perfusion (CTP) is increasingly being used in clinical practice around the world, with several software providing automated estimates of salvageable and ischemic core derived through various mathematical models (hemodynamic maps) (14, 15).

The hemodynamic maps generated by CTP are obtained by tracking a contrast medium as it flows into and out of the brain. The data is then processed using one of several different algorithms (14, 15). The estimation of salvageable tissue and ischemic core is then performed by applying a single threshold to one or two maps (9, 16, 17). However, there is significant variation between algorithms used when estimating tissue perfusion, and single-value thresholds have been shown to both under and overestimate the size of the infarct core and penumbra (18, 19). This may be partly due to the misclassification of image voxels as core or penumbra that results from single-value thresholding of core and penumbra. More sophisticated methods of processing CTP maps are required that can, for example, delineate artifactual signals from those caused by perfusion deficit.

The currently used perfusion thresholds have been validated to some degree and have shown success in selecting patients for treatment through clinical trials (6). However, a predictive model that uses all available perfusion data and spatial context of voxels may provide a more nuanced representation of the pathophysiology of evolving ischemic stroke, improving the accuracy of the images and the robustness of the output. Furthermore, shifting from a rigid single threshold model to a trained Machine Learning (ML) model is highly advantageous as the ML model may continue to improve performance with the addition of data.

There are many studies that develop and test ML and Deep Learning (DL) models for lesion segmentation and there have been great advances in developing applications of ML and DL to healthcare in general [e.g., (20, 21)]. However, there are challenges in widespread deployment such as lack of standardized methods to evaluate performance. Furthermore, the inner mathematical processes of ML and DL are often difficult to understand, and their outputs difficult to interpret. These issues of “explainability” and “interpretability” lead to ML being approached as a “black box” problem, without understanding of internal mechanisms. This has hampered implementation into medical practice. It is therefore essential to integrate ML in small, explainable steps rather than large, black-box overhauls that will result in issues of reliability (22). In this study we investigate if single-value thresholds for measurement of ischemic core and penumbra can be replaced with a ML-based method. We also outline challenges that must be addressed for successful integration into acute stroke assessment protocols.

2. Materials and methods

We developed an early ML model that is trained to delineate both ischemic core and penumbra from surrounding tissue using acute CTP data. We used retrospective data from an acute ischemic stroke patient cohort to develop models based on four ML algorithms (Logistic regression, Random Forest, XGBoost and

Support Vector Machine). We tested performance of the model on an additional set of new, unseen patient data.

2.1. Data acquisition

We analyzed CTP images from the International Stroke Perfusion Imaging Registry (INSPIRE), which is a database of acute stroke perfusion imaging and associated clinical information. For this study we used consecutive patients presenting with acute ischemic stroke who had whole brain CTP and who were recruited into INSPIRE between 2010 and 2017 at the John Hunter Hospital, Newcastle, Australia. For standardization, only one site was used at this stage. As is routine in INSPIRE, patients all underwent baseline multimodal CT imaging with non-contrast CT, CTA, and CTP. Written informed consent was obtained from all participants, and the INSPIRE study was approved by the site’s ethics committee (23).

To obtain the perfusion images, a total of 19 acquisitions occurred over 60 s. The CTP data were processed by commercial software MISTar (Apollo Medical Imaging Technology, Melbourne, VIC, Australia). CTP parameters were generated by applying the mathematical algorithm of singular value decomposition with delay and dispersion correction (24). The following four CTP parameters were generated: cerebral blood flow (CBF), cerebral blood volume (CBV), mean transit time (MTT), and delay time (DT). The penumbra and core volumes were defined with dual thresholds: DT at the threshold of 3 s for total ischemic lesion volume and CBF at the threshold setting of 30% for acute core volume (8, 16, 25). After single-value thresholding, core/penumbra areas were limited to a single lesion and artifactual or erroneous regions were removed. The resulting map was used as the ground truth (GT). Core/penumbra were reviewed by experts to ensure they were accurate.

To develop the model, we used 86 acute ischemic stroke patients with a large vessel occlusion (LVO): M1 segment of the middle cerebral artery (MCA) or internal carotid artery (ICA). To provide additional testing and external validation, 25 patients were used, with both LVO and non-LVO occlusions. This was done to observe whether a model trained only on lesions resulting from an occlusion of large vessel will perform as well when testing on a variety of occlusion sites. Each patient in the test set underwent follow-up MR diffusion-weighted imaging (DWI) between 24 and 72 h after onset. The volume (mL) of the infarct core, as estimated by MR-DWI, was recorded and used for external validation. On follow-up imaging, all patients had a thrombolysis in cerebral infarction (TICI) score of at least 2b, indicating relatively complete reperfusion of initially hypoperfused regions. In these cases, the volume of the acute CTP core should more closely match that of the follow-up infarct core and could therefore be used to validate the predictions.

2.2. Creating labeled data

2.2.1. Class labels

The four hemodynamic maps (hereafter referred to as *features*) and core-penumbra segmentation maps (hereafter referred to as *lesion map*) were used in the development of the algorithm. The

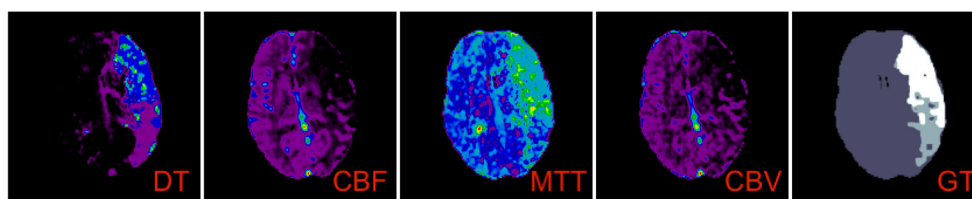


FIGURE 1

Feature maps and lesion map corresponding to the M1 test image. A single axial slice is shown with corresponding perfusion data for delay time (DT), cerebral blood flow (CBF), mean transit time (MTT) and cerebral blood volume (CBV). The corresponding class labels which make up the lesion map, used as ground truth (GT) in the algorithm, is shown on the far right.

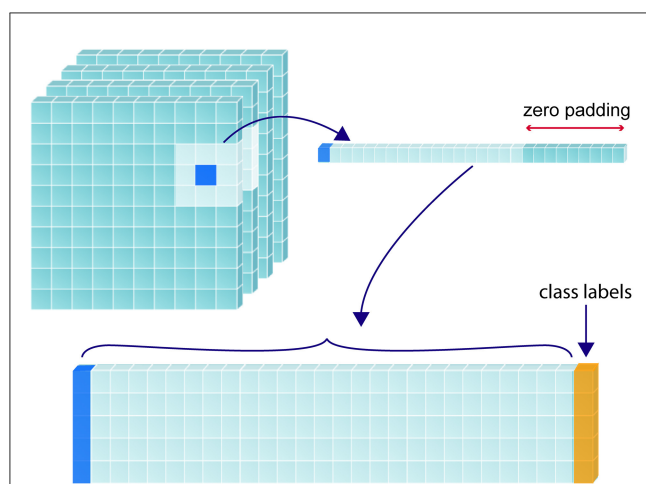


FIGURE 2

Construction of training matrix through sampling and patch extraction. For a given randomly selected sample (shown in dark blue), its corresponding perfusion map value/s and the values corresponding to its 26 immediate neighbors are collapsed into a 1-dimensional array, with the corresponding class label (yellow) added at the furthestmost right position. If multiple perfusion maps are used, the 27 values from each map are recursively added to extend the 1-D array to the left of the class label. The 1-D array for each label are stacked to form a 2-D training matrix.

lesion map, together with the spatial coordinates of the mean baseline image from the CTP acquisition, was used to create a 3-D array of tissue class labels, where each voxel was one of four values: 0—background; 1—non-ischemic brain tissue; 2—penumbra; 3—core). Figure 1 shows the features alongside their class label array for a single patient.

2.2.2. Under-sampling

For this early model, we avoided the issue of class imbalance by sampling the same number of voxels from each class in each image. We processed all lesion maps in the training data, counting the number of voxels belonging to each class. The smallest core volume contained 708 voxels and the smallest penumbra volume contained 8,436 voxels, and two images in the group had a penumbra but no core. We then randomly sampled 300 voxels from each class in each image. For the two images with no core, 300 extra healthy

tissue samples were randomly taken from the image, ensuring 1,200 voxels were sampled from each feature channel.

2.2.3. Patch analysis

To predict the tissue status of a single sample (i.e., voxel of interest), we included the feature values associated with the coordinates of that voxel as well as the values associated with every direct neighboring voxel (26 in total), creating a patch-wise analysis. This was done to include spatial context in the determination of sample tissue status. Zero padding was used for samples that lay around the edges of the image. Figure 2 demonstrates this process for a single voxel of interest, where a 1-D array is created from the sample and its neighbors. Each sample resides in a single row of the training matrix, alongside its class label. All feature channels are concatenated along the same row.

2.3. Machine learning models

The sampled training data was further split into training (60%) and validation (40%) cohorts. Optimization and training were performed on the training data and evaluation was performed on the validation data. All data was standardized to $[-1, 1]$ using the Standard Scalar function in Scikit-Learn in Python (v 0.0) (26).

We used Scikit-learn to optimize four models, based on logistic regression (LR), random forest (RF), XGBoost (XGB) and support vector machine (SVM), respectively. Except for SVM, a randomized search was initially performed, to estimate the best hyperparameters for each algorithm, after which a grid search was performed to narrow down the best hyperparameters. The chosen range for each hyperparameter was determined based on recommendations in Scikit-learn documentation. For each unique parameter combination, three-fold cross validation was performed.

2.4. Impact of added features

For this early model, we wished to determine whether performance was enhanced by including all four CTP maps vs. CBF and Delay Time alone. In particular, we wish to learn whether using four maps reduced the presence of artifactual perfusion lesions. Therefore, each model was trained twice; first with data only from

CBF and Delay Time and then on data from CBF, Delay Time, MTT and CBV.

2.5. Performance evaluation

All the data used to train and optimize the model comprised random samples from images. However, the model will ultimately be used to process whole patient images and provide a prediction that can be displayed as an image. Therefore, we used an additional 25 whole brain patient images to further test the model's performance as it would be applied in a clinical scenario, and to provide a visualization of the model's accuracy. The images were processed as follows: from each voxel in the image, a 3D neighborhood patch was extracted and added to a matrix as in [Figure 2](#). Each 3D patch from the image was forwarded through the model, and the resulting predictions were accumulated in a common space, preserving their spatial location and allowing the image to be reconstructed.

2.5.1. Quantitative performance evaluation

The predictive model was trained using random samples, evenly distributed among the classes. For the test images, however, classes were severely imbalanced. Using receiver-operating characteristics (ROC) or average accuracy would favor the majority class and it is the minority classes that are of interest in this case. Furthermore, the area under the ROC curve (AUC) metric rewards positively predicted background pixels. Therefore, it is not a fair representation of the accuracy of a brain lesion segmentation, whereby background pixels constitute much of the image. For this reason, it was more appropriate to choose a metric more in line with perceptual quality, which reflects both size and localization agreement.

The Dice similarity coefficient (DSC) is a measure of spatial overlap for two regions (A, B), and is given by $DSC(A, B) = 2(A \cap B) / (A + B)$, where \cap is the intersection. It can be seen as the percentage overlap between A and B. A perfect intersection between A and B will give a DSC of 1, and if there is no intersection between the two regions, the score is 0. DSC is sensitive to both size and location differences and is a highly intuitive manner of expressing similarity between two regions. We calculated the DSC between the ground truth and predicted images for the core and penumbra regions separately. After (27), DSC can be separated into a similar manner to the Kappa coefficient for agreement, into the following six categories (28, 29): 0, "No Agreement"; 0–0.2, "Slight agreement"; 0.2–0.4, "Fair agreement"; 0.4–0.6, "Moderate agreement"; 0.6–0.8, "Substantial agreement"; "0.8–1"; "Almost perfect agreement".

The Jaccard Index (JI), also known as the Intersection of Union (IoU), like the DSC, ranges from 0 (no agreement) to 1 (perfect agreement). The JI is mathematically represented by $IoU(A, B) = A \cap B / A \cup B$, where \cup is the union. The relationship between JI and DSC can therefore be described as $JI = DSC / (2 - DSC)$. The DSC tends to be higher as it counts the true positive classifications twice in both the numerator and denominator of its equation, while the JI gives a greater penalty for bad classifications. Therefore,

providing an average score over a set of classification will lead the average DSC and average JI to diverge from one another. The two metrics will always be positively correlated, however, we found it worthwhile to analyse the distinction as both are used throughout literature to evaluate segmentation tasks. The DSC and JI values for each the core and penumbra were calculated for all 25 images, and the differences between them were evaluated using paired *t*-tests.

Finally, lesion volume, one of the most important predictors of outcome after ischemic stroke, was calculated for the additional test images. The volumes of the core and penumbra were calculated for each of the ground truth and the predicted lesion by counting the number of voxels assigned to each area (30). Using pixel information encoded in the image, the absolute volume in milliliters could be calculated. As an external validation, the predicted core volume was compared with the follow-up (24–72h) infarct core derived from MR-DWI imaging and reviewed by the expert stroke neurologist (MP).

2.5.2. Qualitative performance evaluation

We identified eleven images within the cohort affected by artifacts relating to the skull. In brain CT imaging, beam hardening from the dense skull region or, to a lesser degree, contrast-enhanced arteries, may result in a characteristic "streaking" artifact (31). When the skull, a highly attenuating region is adjacent to less attenuating tissue, such as soft tissue, and there is limited CT resolution, partial volume averaging may also occur. Here, the image intensity of affected voxels is a mixture, or an average, or the intensity of both these regions (32). [Figure 3](#) shows an example of the partial volume artifact in Subject 3. Upon CTP processing, such voxels near the edge of the brain shows increased Delay Time. However, these artifacts are common and, if the image is otherwise of good quality, artifactual perfusion lesions are easy to identify to the trained eye. Therefore, we did not exclude these cases from the study and instead prefer to investigate the impact of artifact on model performance. We qualitatively compared the ability of the algorithms to make a correct prediction around those areas, based on both the inclusion of all four CTP maps and the additional spatial information provided by the 3D patches.

3. Results

For the training set, 55 patients had an occlusion of the M1 segment of the middle cerebral artery (MCA), and 31 had an occlusion of the internal carotid artery (ICA). Forty-three patients were female (50%), and the median onset age was 74 (IQR 63–82). The median baseline NIHSS (National Institutes of Health Stroke Scale) was 17 (IQR 14–20). Of these, 70 patients had a known time of onset; the median time between onset and CT imaging was 121 min (IQR 95–157). One patient had a wake-up stroke, and 15 patients had an unknown time of onset. Seventy-six patients received intravenous (IV) thrombolysis, one received intraarterial (IA) thrombectomy, two received both, five received no treatment and two patients did not have any treatment documented.

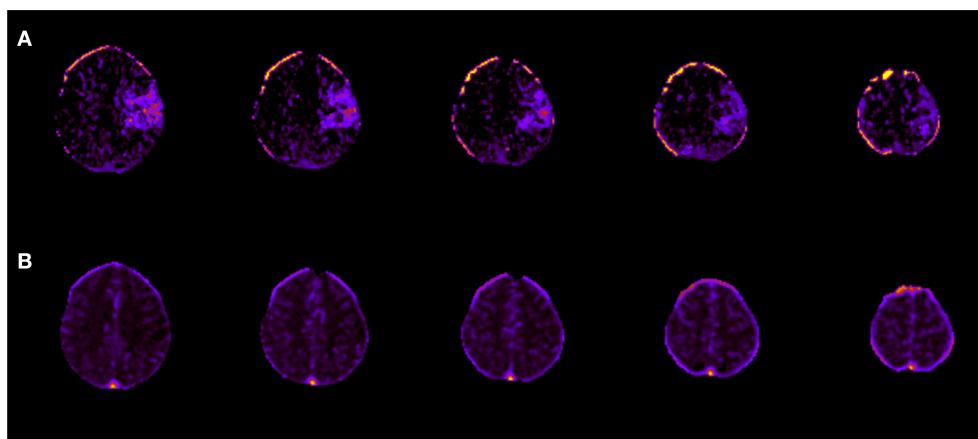


FIGURE 3

Skull artifacts. For subject 3, skull artifacts can be seen in their (A) Delay time and (B) CBF maps near the top of the skull.

TABLE 1 Class representations across the training and validation cohorts.

	Background	Non-ischemic brain	Core	Penumbra
Train	15,485	15,870	15,141	15,424
Validation	10,315	10,530	10,059	10,376
Total	25,800	26,400	25,200	25,800

Samples were split into these categories using Scikit-Learn (26).

This was done to ensure the model did not bias any class.

Three patients were discarded from the test set due to considerable infarct growth. For the remaining patients in the test set, 16 patients had an M2-MCA occlusion, four had an M3-MCA occlusion, and one each with an occlusion of the anterior cerebral artery (ACA) and ICA. Thirteen patients (59%) were female, and the median onset age was 79 (IQR 74–83). The median baseline NIHSS was 11 (IQR 6–16). In total, 20 patients had a known time of onset; the median time between onset and CT imaging for these patients was 110 min (IQR 96–168). The remaining patients had an unknown time of onset. Of all the patients in the test set, 20 received IV treatment, one received IA treatment and one received no treatment. Sixteen were given a TICI 3 score, and 6 were given a TICI 2b score. The median day of DWI image after stroke onset was 1 (IQR 1–2, min-max 0–12). The median size of the follow-up DWI core was as 10 mL (IQR 6–33). A Pearson correlation test shown a strong correlation ($p < 0.005$, two tailed) between the data used for the ground truth core measurement and the expert assessed MR-DWI measurements for core volume.

For model development, a total of 103,200 patch samples was used. Table 1 shows the class instances for the train and validation groups used to develop the model.

Table 2 shows details of optimizing each model. Each model was trained using six computer processing units (CPU) in parallel. For SVM, only a random search for the two-map model was carried due to the excessive training times (>22 h), and only

polynomial and linear kernels were tested, with the polynomial kernel outperforming the linear kernel. Table 3 shows results for each model on the under-sampled data. XGBoost was the highest performing algorithm, and there was an improvement in performance when all four CTP maps were included.

The performance of the best performing model (shown in bold in Table 3) was tested on the remaining 22 images in the test set. The results are shown in Supplementary Table A1. Figure 4 shows axial slices of lesion predictions (overlayed on non-contrast CT image slices) using the model based on all four CTP maps for a selection of datasets (subjects 7, 8, and 1 with reference to Supplementary Table A1).

For all 22 patients, the mean DSC values for core and penumbra were 0.39 (SD 0.26) and 0.50 (SD 0.22), respectively, and the mean JI values for core and penumbra were 0.28 (SD 0.23) and 0.36 (SD 0.20), respectively. For both core and penumbra, JI and DSC were significantly different across the dataset (core: paired t -test, $p < 0.0001$; penumbra: paired t -test, $p < 0.0001$).

To explore the difference between performance on core and penumbra, a volume analysis was performed. Each similarity measured varied significantly with volume: A Pearson's correlation for DSC variation with volume showed ($r = 0.56$, $p = 0.0065$) for penumbra and ($r = 0.71$, $p = 0.0002$) for core. For JI a Pearson's correlation calculation showed ($r = 0.61$, $p = 0.0028$) for penumbra and ($r = 0.72$, $p < 0.0002$) for core.

Out of the 22 testing images, 16 lesions were due to an occlusion of the M2 segment of the MCA. The DSC scores for core and penumbra averaged to 0.34 (SD 0.23) and 0.50 ± 0.20 , respectively. The mean volume of core and penumbra for M2 lesions was 9.89 mL (SD 8.17) and 38.34 mL (SD 22.5), respectively, lower as compared with the entire testing set.

There was no significant correlation between the XGB-predicted core and the 24 h DWI infarct core (Pearson's r ; $r = 0.18$, $p = 0.41$). However, visual inspection confirmed that artifacts due to the skull were present in half the cases ($n = 11$) and led to overestimation of perfusion regions. When considering the test cases with no obvious skull artifacts, there was a significant correlation between the predicted core and the follow-up DWI

TABLE 2 Details of model training.

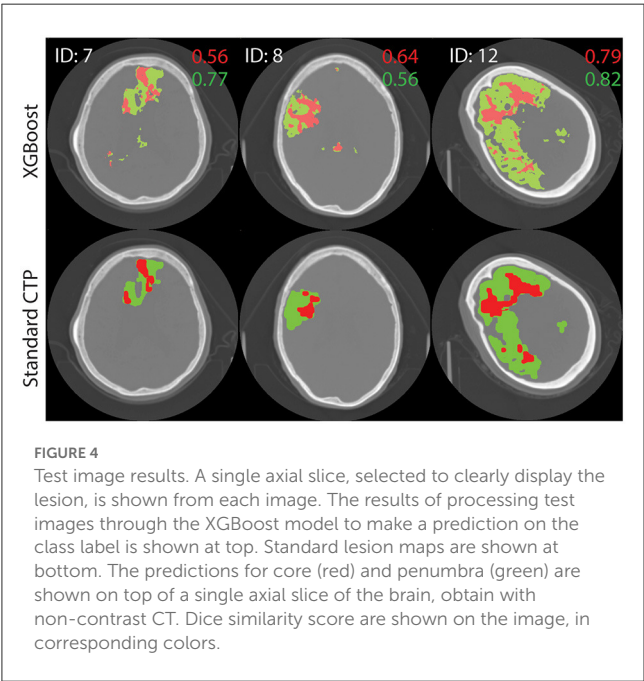
Algorithm	#Parameters optimized	#Candidates in random search	Time taken (2 map, 4 map)	#Candidates in grid search	Time taken (2 map, 4 map)
LR	6	28	5 min, 33 min	3	20 min, 53 min
RF	7	80	1 h 53 min, 3 h 1min	81	4 h 26 min, 6 58 min
XGB	5	10	30 min, 57 min	27	1 h 38 min, 2 h 53 min
SVM	3	30	22 h 33 min, N/A	N/A	N/A

Four different algorithms were used to train models: Logistic Regression (LR), Random Forest (RF), XGBoost (XGB) and Support Vector Machine (SVM). Each algorithm has different hyperparameters, and the number of different hyperparameters that were optimized here is shown (“#Parameters optimized”). Except for SVM, the parameters were optimized by first running a random search, training and testing models with a number of different random hyperparameter models (“#Candidates in random search”). The best performing combination was used to create the range for a more refined grid search. The number of candidates tested in the grid search was determined by the number of parameters that were optimized and the possible values for each parameter (e.g., whether values were discrete or continuous). The time that was taken to run all the different combinations is included. Six CPUs were used in parallel.

TABLE 3 Results of models on validation data.

	ROC-AUC	DSC (core)	DSC (pen)	Jl (core)	Jl (pen)
LR	0.9757	0.8438	0.7874	0.7298	0.6494
	0.9776	0.848	0.7907	0.736	0.6538
RF	0.9825	0.8553	0.8172	0.7471	0.6908
	0.9841	0.8611	0.8269	0.7561	0.7048
XGB	0.983	0.8552	0.8185	0.7470	0.6927
	0.9844	0.8610	0.8275	0.7559	0.7057
SVM	0.9799	0.8467	0.8081	0.7341	0.678

Eight models in total were optimized. Three algorithms (Logistic Regression/LR, Random Forest/RF, XGBoost/XGB) were trained twice, one on Cerebral Blood Flow (CBF) and Delay Time (DT) data (top), and once on data from CBF, DT, Cerebral Blood Flow and Mean Transit Time (bottom). Support Vector Machine (SVM) was only trained for two maps due to excessive training times. Results on the validation data are shown for each model. The highest performance across all categories was obtained for XGB, trained on all four CTP maps.



core (Pearson’s r ; $r = 0.82$, $p = 0.0018$). Figure 5 shows a comparison of results from each algorithm for the subject shown in Figure 3. This subject had significant CTP artifacts due to the skull. While LR could not distinguish actual from artifactual

perfusion lesions, in this case all the other algorithms were able to.

4. Discussion

This study proposes a machine learning algorithm using the entire perfusion map datasets as an alternative to measuring the penumbra and ischemic core using binary thresholds with CTP data. Models based on four different well-known ML algorithms were tested. Accuracy was tested both quantitatively, using similarity measurements, and qualitatively, by using visual inspected to determine which algorithm was better at prediction on artifactual CTP hyperintensities. Simple neighborhood analysis was used to make a prediction on a single voxel; all surrounding voxels were considered. Our model may easily be expanded to include additional input channels, such as non-contrast CT, or relevant clinical information such as time-from-onset, blood pressure, clinical severity measurements and age.

Out of the four algorithms tested, XGBoost performed best in the quantitatively analysis, achieving good accuracy in mimicking the CTP perfusion lesions derived by the clinically used software MISTar. There was an improvement in performance when all four CTP maps were used compared to only CBF and DT for this early model. Future versions of the model will continue to use all four CTP maps to make a prediction.

Ideally, an automated CTP algorithm should differentiate between genuine and artifactual hypoperfusion patterns, just as an

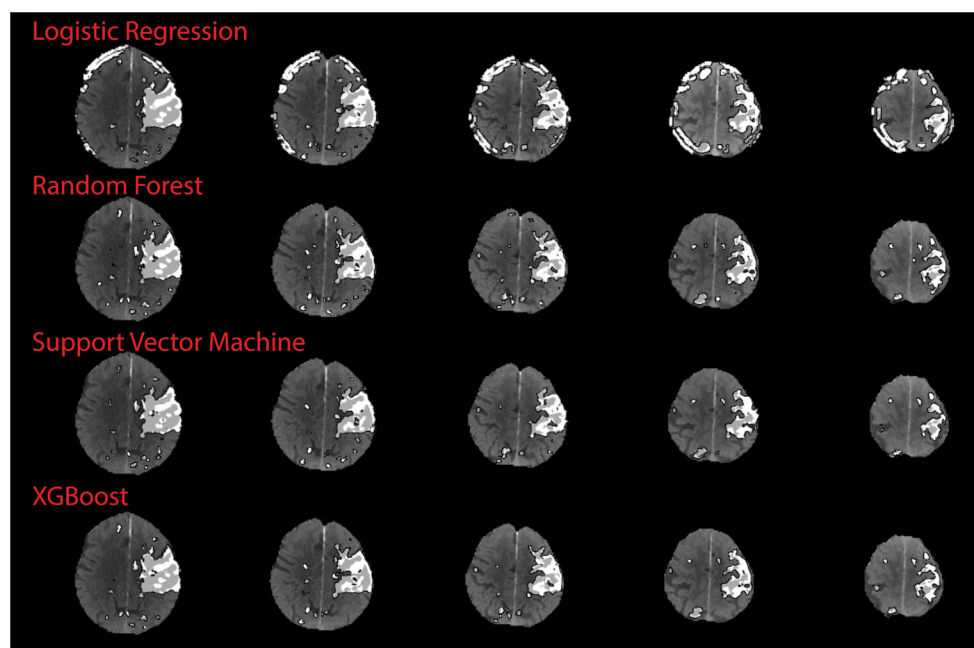


FIGURE 5

Qualitative results on skull artifact. Predictions made on subject 3 (shown in Figure 3) for each algorithm. Linear regression performed the poorest in terms of identifying artifact as areas of perfusion, as demonstration by subject 3 axial slices.

experienced stroke physician should be able to determine whether the pattern is topographically consistent with stroke phenotype (33, 34). For the qualitative study, SVM, XGB and RF improved on the ability of the LR algorithm to distinguish real from artifactual CTP hyperintensities. This is because LR is the only algorithm based on linear first-order interactions between variables, whereas the other three are more sophisticated, and able to model non-linear and higher interactions. This is shown clearly in Figure 5, although in other cases the ML model still derived artifact in making a prediction, leading to an overall worse correlation with the DWI infarct core for images with obvious artifact. As CT artifacts are difficult to avoid altogether in a clinical setting, this is a useful insight. Further development is required to ensure future versions of this model to not derive artifactual perfusion lesions.

For the testing images, DSC and JI scores were shown to vary significantly even though they are both commonly used similarity metrics. In addition, both metrics varied significantly with volume. Therefore, the DSC or JI score for a large lesion may not represent the same accuracy as for a small lesion, even though (27) has proposed otherwise (35, 36). For example, the large core in Figure 4 (ID = 12, M1) receives an almost perfect DSC value, while the smaller cores received lower DSC scores; these differences may have resulted merely from size differences. The same behavior was seen with JI. An average DSC or JI score that is a result of the summing over results from lesions of different sizes will not be an accurate representation of the overall performance of a model. We propose a weighted mean DSC/JI to account for size variation before these scores can be fully interpretable. Further studies will explore the application of a weighted mean. In lieu of a robust and subjective

model performance metric, benchmark data [ISLES 2018 (37)] will be used in future studies to report performance.

The most significant limitation to this study is that, as a first step, we have used the CTP core and penumbra estimations derived by MISTar as the ground truth, even though these only approximate the ground truth. The gold standard in the determination of acute ischemic tissue is an expertly segmented MR-DWI lesion, either with (core) or without reperfusion (penumbra) (9, 38). Without a perfect ground truth, it remains difficult to interpret model performance in an objective fashion. For example, as MISTar maps are based on a simple thresholding method, a meaningful comparison of this ML method to a thresholding method against MISTar maps is challenging. Although MISTar and other software (39) CTP core estimates have been shown to be a fair approximation of DWI lesions previously, there are certainly ongoing issues (18), one of which is that the reference standard for core is imperfect (16). Nonetheless, future studies will adopt manually segmented DWI images as ground truth and therefore be able to provide performance metrics that are more robust and interpretable. In addition, this model uses derived perfusion parameters rather than raw CTP time series images, risking a loss of valuable information contained in the raw images which may be lost in the derivation process. The model uses a simple approach over more advanced approaches that have been tested in the literature, such as those based on Deep Learning. With DL, features may be automatically extracted from images, both locally and globally, to make predictions with efficiency (40, 41). Future models will adopt DL, however, the current analysis using more explainable algorithms, was a necessary first step.

With this study we have shown that a Machine Learning method is capable of mimicking common use perfusion lesion measurements to a high accuracy. With the increasing prevalence of CTP assessments for treatment selection of ischemic stroke patients, particularly in the extended time window, it is vital that the measurements be accurate and representative of the underlying pathophysiology. There is significant scope for the current single threshold methods to overestimate the ischemic core depending on onset to reperfusion speed, and other factors. The proposed model may prove more accurate with further development than the currently used single threshold maps and can consider physiologically relevant information such as blood pressure, cardiac output and fluid status which would influence contrast flow and hence perfusion measures on the CTP. Imaging metadata such as time may also influence accuracy, as “ghost cores” have been noted in the hyperacute phase (42).

While the model is simple in its current form, we were able to demonstrate salient points about CTP-based predictions of stroke infarct. We have demonstrated that similarity indices such as DSC and JI have some difficulty in interpretations and further development of performance metrics is required. We have also demonstrated that non-linear algorithms are more adept at making predictions on common CT artifacts that linear model such as logistic regression. Further studies will use manually segmented DWI volumes as ground truth, as well as digest raw CTP data rather than post-processed CTP maps for Deep Learning predictions. Benchmark datasets will be used to measure performance. In addition, the role of clinical data and imaging metadata will be explored in making predictions.

5. Conclusion

We have described a Machine Learning model for the delineation of ischemic tissue from CTP data which is based on the XGBoost algorithm combined with 3D neighborhood analysis. The model is trained on lesion segmentations derived by clinically used software and can derive perfusion lesions to high accuracy. The model improves on clinically available software in that it is able to use multiple input channels but is currently limited by the lack of validation against gold standard lesion segmentations. Nonetheless, the model allowed us to demonstrate useful insight

into CTP-based prediction of stroke infarct which will be used to make future developments.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: De-identified data from the current study are available for qualified investigators upon reasonable request to the corresponding author. Requests to access these datasets should be directed to freda.werdiger@unimelb.edu.au.

Author contributions

FW performed conception of the work, data analysis and interpretation, and drafting of the article. AB, MP, MV, TK, CL, NS, and LL provided critical revision of the article and final approval of the version to be published. All authors contributed to the article and approved the submitted version.

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2023.1098562/full#supplementary-material>

References

- Goyal M, Menon BK, van Zwam WH, Dippel DWJ, 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
- Jovin TG, Chamorro A, Cobo E, de Miquel MA, Molina CA, Rovira A, et al. Thrombectomy within 8 hours after symptom onset in ischemic stroke. *N Engl J Med*. (2015) 372:2296–306. doi: 10.1056/NEJMoa1503780
- Saver JL, Goyal M, van der Lugt A, Menon BK, Majoie CBLM, Dippel DW, et al. Time to treatment with endovascular thrombectomy and outcomes from ischemic stroke: a meta-analysis. *JAMA*. (2016) 316:1279. doi: 10.1001/jama.2016.13647
- 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
- Campbell BCV, Ma H, Ringleb PA, Parsons MW, Churilov L, Bendzus M, et al. Extending thrombolysis to 4.5–9 h and wake-up stroke using perfusion imaging: a systematic review and meta-analysis of individual patient data. *Lancet*. (2019) 394:139–47. doi: 10.1016/S0140-6736(19)31053-0
- 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

7. 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
8. Campbell BCV, Christensen S, Levi CR, Desmond PM, Donnan GA, Davis SM, et al. Cerebral blood flow is the optimal CT perfusion parameter for assessing infarct core. *Stroke.* (2011) 42:3435–40. doi: 10.1161/STROKEAHA.111.618355
9. Bivard A, Levi C, Spratt N, Parsons M. Perfusion CT in acute stroke: a comprehensive analysis of infarct and penumbra. *Radiology.* (2013) 267:543–50. doi: 10.1148/radiol.12120971
10. Campbell BCV, Mitchell PJ, Kleinig TJ, Dewey HM, Churilov L, Yassi N, et al. Endovascular therapy for ischemic stroke with perfusion-imaging selection. *N Engl J Med.* (2015) 372:1009–18. doi: 10.1056/NEJMoa1414792
11. Bivard A, Spratt N, Levi C, Parsons M. Perfusion computer tomography: imaging and clinical validation in acute ischaemic stroke. *Brain.* (2011) 134:3408–16. doi: 10.1093/brain/awr257
12. Albers GW, Goyal M, Jahan R, Bonafe A, Diener H-C, Levy EI, et al. Ischemic core and hypoperfusion volumes predict infarct size in SWIFT PRIME: CT perfusion volumes. *Ann Neurol.* (2016) 79:76–89. doi: 10.1002/ana.24543
13. Lansberg MG, Christensen S, Kemp S, Mlynash M, Mishra N, Federau C, et al. Computed tomographic perfusion to Predict Response to Recanalization in ischemic stroke: results of the CRISP study. *Ann Neurol.* (2017) 81:849–56. doi: 10.1002/ana.24953
14. Kudo K, Sasaki M, Yamada K, Momoshima S, Utsunomiya H, Shirato H, et al. Differences in CT perfusion maps generated by different commercial software: quantitative analysis by using identical source data of acute stroke patients. *Radiology.* (2010) 254:200–9. doi: 10.1148/radiol.254082000
15. Zussman BM, Boghosian G, Gorniak RJ, Olszewski ME, Read KM, Siddiqui KM, et al. The relative effect of vendor variability in CT perfusion results: a method comparison study. *Am J Roentgenol.* (2011) 197:468–73. doi: 10.2214/AJR.10.6058
16. Lin L, Bivard A, Krishnamurthy V, Levi CR, Parsons MW. Whole-brain CT perfusion to quantify acute ischemic penumbra and core. *Radiology.* (2016) 279:876–87. doi: 10.1148/radiol.2015150319
17. McVerry F, Dani KA, MacDougall NJJ, MacLeod MJ, Wardlaw J, Muir KW. Derivation and evaluation of thresholds for core and tissue at risk of infarction using CT perfusion: CT perfusion thresholds for core and tissue at risk. *J Neuroimaging.* (2014) 24:562–8. doi: 10.1111/jon.12134
18. Parsons MW. Automated measurement of computed tomography acute ischemic core in stroke: does the emperor have no clothes? *Stroke.* (2021) 52:642–4. doi: 10.1161/STROKEAHA.120.032998
19. Bivard A, Churilov L, Ma H, Levi C, Campbell B, Yassi N, et al. Does variability in automated perfusion software outputs for acute ischemic stroke matter? Reanalysis of EXTEND perfusion imaging CNS. *Neurosci Ther.* (2022) 28:139–44. doi: 10.1111/cns.13756
20. Anand VK, Khened M, Alex V, Krishnamurthy G. Fully automatic segmentation for ischemic stroke using CT perfusion maps. In: Crimi A, Bakas S, Kuijff H, Keyvan F, Reyes M, van Walsum T, editors. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Lecture Notes in Computer Science. Cham: Springer International Publishing. (2019). p. 328–334. doi: 10.1007/978-3-030-11723-8_33
21. Zhang J, Shi F, Chen L, Xue Z, Zhang L, Qian D. Ischemic Stroke Segmentation from CT Perfusion Scans Using Cluster-Representation Learning. In: Kia SM, Mohy-ud-Din H, Abdulkadir A, Bass C, Habes M, Rondina JM, Tax C, Wang H, Wolfers T, Rathore S, et al., editors. *Machine Learning in Clinical Neuroimaging and Radiogenomics in Neuro-oncology*. Cham: Springer International Publishing (2020). p. 67–76
22. Gupta R, Krishnam SP, Schaefer PW, Lev MH, Gonzalez RG. An east coast perspective on artificial intelligence and machine learning. *Neuroimaging Clin N Am.* (2020) 30:467–78. doi: 10.1016/j.nic.2020.08.002
23. Gao L, Tan E, Moodie M, Parsons M, Spratt NJ, Levi C, et al. Reduced impact of endovascular thrombectomy on disability in real-world practice, relative to randomized controlled trial evidence in Australia. *Front Neurol.* (2020) 11:593238. doi: 10.3389/fneur.2020.593238
24. Lin L, Bivard A, Kleinig T, Spratt NJ, Levi CR, Yang Q, et al. Correction for delay and dispersion results in more accurate cerebral blood flow ischemic core measurement in acute stroke. *Stroke.* (2018) 49:924–30. doi: 10.1161/STROKEAHA.117.019562
25. Bivard A, Levi C, Krishnamurthy V, McElduff P, Miteff F, Spratt NJ, et al. Perfusion computed tomography to assist decision making for stroke thrombolysis. *Brain.* (2015) 138:1919–31. doi: 10.1093/brain/awv071
26. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. *J Mach Learn Res.* (2011) 12:2825–30. doi: 10.48550/arXiv.1201.0490
27. Zijdenbos AP, Dawant BM, Margolin RA, Palmer AC. Morphometric analysis of white matter lesions in MR images: method and validation. *IEEE Trans Med Imaging.* (1994) 13:716–24. doi: 10.1109/42.363096
28. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* (1977) 33:159. doi: 10.2307/2529310
29. Pajula J, Kauppi J-P, Tohka J. Inter-subject correlation in fMRI: method validation against stimulus-model based analysis. *PLoS ONE.* (2012) 8:e41196. doi: 10.1371/journal.pone.0041196
30. Hope TMH, 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
31. Barrett JF, Keat N. Artifacts in CT: Recognition and avoidance. *RadioGraphics.* (2004) 24:1679–91. doi: 10.1148/rg.246045065
32. Coolens C, Mohseni H, Dhodi S, Ma S, Keller H, Jaffray DA. Quantification accuracy for dynamic contrast enhanced (DCE) CT imaging: phantom and quality assurance framework. *Eur J Radiol.* (2018) 106:192–8. doi: 10.1016/j.ejrad.2018.08.003
33. Kauw F, Heit JJ, Martin BW, van Ommen F, Kappelle LJ, Velthuis BK, et al. Computed tomography perfusion data for acute ischemic stroke evaluation using rapid software: pitfalls of automated postprocessing. *J Comput Assist Tomogr.* (2020) 44:75–7. doi: 10.1097/RCT.0000000000000946
34. Siegler JE, Olsen A, Pulst-Korenberg J, Cristancho D, Rosenberg J, Raab L, et al. Multicenter volumetric assessment of artifactual hypoperfusion patterns using automated CT perfusion imaging. *J Neuroimaging.* (2019) 29:573–9. doi: 10.1111/jon.12641
35. Cheng D-C, Chi J-H, Yang S-N, Liu S-H. Organ contouring for lung cancer patients with a seed generation scheme and random walks. *Sensors.* (2020) 20:4823. doi: 10.3390/s20174823
36. Ghaffari M, Sanchez L, Xu G, Alaraj A, Zhou XJ, Charbel FT, et al. Validation of parametric mesh generation for subject-specific cerebroarterial trees using modified Hausdorff distance metrics. *Comput Biol Med.* (2018) 100:209–20. doi: 10.1016/j.combiomed.2018.07.004
37. Hakim A, Christensen S, Winzeck S, Lansberg MG, Parsons MW, Lucas C, et al. Predicting infarct core from computed tomography perfusion in acute ischemia with machine learning: lessons from the ISLES challenge. *Stroke.* (2021) 52:2328–37. doi: 10.1161/STROKEAHA.120.030696
38. Bivard A, Kleinig T, Miteff F, Butcher K, Lin L, Levi C, et al. Ischemic core thresholds change with time to reperfusion: a case control study. *Ann Neurol.* (2017) 82:995–1003. doi: 10.1002/ana.25109
39. Cereda CW, Christensen S, Campbell BC, Mishra NK, Mlynash M, Levi C, et al. A benchmarking tool to evaluate computer tomography perfusion infarct core predictions against a DWI standard. *J Cereb Blood Flow Metab.* (2016) 36:1780–9. doi: 10.1177/0271678X15610586
40. Robben D, Boers AMM, Marquering HA, Langezaal LLCM, Roos YBWEM, van Oostenbrugge RJ, et al. Prediction of final infarct volume from native CT perfusion and treatment parameters using deep learning. *Med Image Anal.* (2020) 59:101589. doi: 10.1016/j.media.2019.101589
41. Amador K, Wilms M, Winder A, Fiehler J, Forkert ND. Predicting treatment-specific lesion outcomes in acute ischemic stroke from 4D CT perfusion imaging using spatio-temporal convolutional neural networks. *Med Image Anal.* (2022) 82:102610. doi: 10.1016/j.media.2022.102610
42. Boned S, Padroni M, Rubiera M, Tomasello A, Coscojuela P, Romero N, et al. Admission CT perfusion may overestimate initial infarct core: the ghost infarct core concept. *J NeuroInterventional Surg.* (2017) 9:66–9. doi: 10.1136/neurintsurg-2016-012494



OPEN ACCESS

EDITED BY

Daniel Donoho,
Children's National Hospital, United States

REVIEWED BY

Kais Gadhumi,
Duke University, United States
Xiaohuan Xia,
Tongji Hospital Affiliated to Tongji
University, China

*CORRESPONDENCE

Wenle Li
✉ drlee0910@163.com
Liangqun Rong
✉ rongliangqun@163.com
Xiu'e Wei
✉ wxeqq@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 20 November 2022

ACCEPTED 02 February 2023

PUBLISHED 23 February 2023


CITATION

Wang K, Gu L, Liu W, Xu C, Yin C, Liu H, Rong L,
Li W and Wei X (2023) The predictors of death
within 1 year in acute ischemic stroke patients
based on machine learning.
Front. Neurol. 14:1092534.
doi: 10.3389/fneur.2023.1092534

COPYRIGHT

© 2023 Wang, Gu, Liu, Xu, Yin, Liu, Rong, Li and
Wei. This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

The predictors of death within 1 year in acute ischemic stroke patients based on machine learning

Kai Wang^{1,2†}, Longyuan Gu^{3†}, Wencai Liu^{4†}, Chan Xu⁵,
Chengliang Yin⁶, Haiyan Liu^{1,2}, Liangqun Rong^{1,2*}, Wenle Li ^{2,7*}
and Xiu'e Wei^{1,2*}

¹Department of Neurology, The Second Affiliated Hospital of Xuzhou Medical University, Xuzhou, Jiangsu, China, ²Key Laboratory of Neurological Diseases, The Second Affiliated Hospital of Xuzhou Medical University, Xuzhou, Jiangsu, China, ³Department of Neurosurgery, The Affiliated Hospital of Xuzhou Medical University, Xuzhou, Jiangsu, China, ⁴Department of Orthopaedic Surgery, The First Affiliated Hospital of Nanchang University, Nanchang, China, ⁵Department of Dermatology, Xianyang Central Hospital, Xianyang, China, ⁶Faculty of Medicine, Macau University of Science and Technology, Taipa, Macao SAR, China, ⁷The State Key Laboratory of Molecular Vaccinology and Molecular Diagnostics and Center for Molecular Imaging and Translational Medicine, School of Public Health, Xiamen University, Xiamen, China

Objective: To explore the predictors of death in acute ischemic stroke (AIS) patients within 1 year based on machine learning (ML) algorithms.

Methods: This study retrospectively analyzed the clinical data of patients hospitalized and diagnosed with AIS in the Second Affiliated Hospital of Xuzhou Medical University between August 2017 and July 2019. The patients were randomly divided into training and validation sets at a ratio of 7:3, and the clinical characteristic variables of the patients were screened using univariate and multivariate logistics regression. Six ML algorithms, including logistic regression (LR), gradient boosting machine (GBM), extreme gradient boosting (XGB), random forest (RF), decision tree (DT), and naive Bayes classifier (NBC), were applied to develop models to predict death in AIS patients within 1 year. During training, a 10-fold cross-validation approach was used to validate the training set internally, and the models were interpreted using important ranking and the SHapley Additive exPlanations (SHAP) principle. The validation set was used to externally validate the models. Ultimately, the highest-performing model was selected to build a web-based calculator.

Results: Multivariate logistic regression analysis revealed that C-reactive protein (CRP), homocysteine (HCY) levels, stroke severity (SS), and the number of stroke lesions (NOS) were independent risk factors for death within 1 year in patients with AIS. The area under the curve value of the XGB model was 0.846, which was the highest among the six ML algorithms. Therefore, we built an ML network calculator (<https://mlmedicine-de-stroke-de-stroke-m5pijk.streamlitapp.com/>) based on XGB to predict death in AIS patients within 1 year.

Conclusions: The network calculator based on the XGB model developed in this study can help clinicians make more personalized and rational clinical decisions.

KEYWORDS

ischemic stroke, biomarkers, machine learning, prediction model, web calculator

1. Introduction

Acute ischemic stroke (AIS) is a disease caused by the occlusion of cerebral arteries, accompanied by brain tissue infarction and neuronal cell damage, causing severe trauma to the body. AIS is the leading cause of disability in adults and the primary cause of human death worldwide (1, 2). In 2019, there were 7,630,800 cases of AIS globally, an 87.55% increase compared to the previous 30 years. The high morbidity, mortality, and disability rates associated with AIS impose a severe economic burden on society and families (3). Several factors may have a significant impact on the pathogenesis and prognosis of patients with AIS, including the immune inflammatory response during AIS development, with the involvement of different pathways and sources of activated inflammatory factors, and is an important regulator of stroke progression, post-stroke damage, cerebral function repair and death (4–6). Approximately 10% of AIS patients, representing a type of morbidity, experience a fatal event within 1 year (7). There is an urgent need to identify the early and effective predictors of death 1 year after the onset of AIS. The construction of a model of death prediction in stroke patients within 1 year could provide clinicians with a reliable tool to assess the condition of their patients. However, there are few reports in this area.

ML-assisted clinical decision-making and analysis have been widely used in clinical settings (8–11), especially in the screening phase of big data feature variables (12, 13). The superior performance demonstrated by ML algorithms in medical big data makes it possible to obtain better predictive tools than traditional statistical models under certain conditions. However, few studies have been conducted to screen the risk factors of death in AIS patients within 1 year using ML algorithms.

Therefore, this study aimed to develop and validate an interpretable ML model that used clinically relevant variables to predict death within 1 year in AIS patients and construct an easy-to-use web calculator as a convenient and practical protective tool for clinical practitioners to provide valid information for AIS patients.

2. Materials and methods

2.1. Subjects

Patients who were hospitalized in the Department of Neurology of the Second Affiliated Hospital of Xuzhou Medical University and diagnosed with AIS between August 2017 and July 2019 were retrospectively analyzed. A total of 677 patients with AIS were included in this study, 32 of whom died after admission and during follow-up. The study was approved by the Ethics Committee of the Second Affiliated Hospital of Xuzhou Medical University [ethics number: [2020] 081603], and all patients provided written informed consent.

2.2. Inclusion and exclusion criteria

The inclusion criteria were a diagnosis of AIS in accordance with the World Health Organization criteria, and the time between

onset and hospital admission did not exceed 24 h. The exclusion criteria were: (1) incomplete clinical data, (2) those with severely abnormal organ function, (3) inadequate ancillary investigations, (4) follow-up of <1 year, and (5) Patients who discontinued treatment for various reasons according to their relatives.

2.3. Methods

2.3.1. Observational variables

In this study, clinical data were collected from the enrolled patients, including demographics (age and sex); vascular risk factors (hypertension, diabetes mellitus, and ischemic heart disease); baseline blood pressure [systolic blood pressure (SBP) and diastolic blood pressure (DBP)]; Trial of Org 10 172 in Acute Stroke Treatment (TOAST) [large-artery atherosclerosis, cardioembolism, small-vessel occlusion, acute stroke of other determined etiology, stroke of undetermined etiology]; stroke severity (SS) [defined as mild stroke according to the National Institutes of Health Stroke Scale (NIHSS) scores of ≤ 8 , moderate-to-severe stroke according to NIHSS scores of ≥ 9 ; all assessments completed on admission]; magnetic resonance imaging (MRI) features [stroke distribution (SD; anterior circulation, posterior circulation, and anterior/posterior circulation), side of hemisphere (SOH; left, right, and bilateral), number of stroke lesions (NOSs; single and multiple stroke lesions), site of stroke lesions (SOSs; cortical, cortico-subcortical, subcortical, brainstem, and cerebellum)]; laboratory tests [total cholesterol, triglycerides, low-density lipoprotein (LDL), fasting blood glucose (FBG), homocysteine (HCY), uric acid (UA), fibrinogen (FIB), myoglobin (MB), C-reactive protein (CRP), D-dimer brain natriuretic peptide (BNP), HBALC, neuron-specific enolase (NSE), and S-100 β levels], treatment regimen [intravenous thrombolysis, arterial thrombolysis, antiplatelet, anticoagulation, statin, and proton pump inhibitor therapy (PPI)]; and stroke comorbidities [dysphagia and stroke-associated pneumonia (SAP)].

2.3.2. Statistical methods

This study used R version 4.0.5 software for data processing and statistical analyses. Continuous variables are expressed as the median or interquartile range (IQR) while categorical variables are presented as frequencies (percentage, %). The continuous variables were compared by independent samples *t*-tests and the categorical variables were compared using χ^2 -tests. Understanding the relationship between the independent and dependent variables was clinically meaningful and *P*-values of < 0.05 were considered statistically significant (two-sided).

2.3.3. Modeling of machine learning algorithms

Univariate and multivariate logistic regression analyses were used to assess the risk factors of death within 1 year in the training group study population. The odds ratio (OR) and 95% confidence interval (CI) were calculated, with an OR of > 1 indicating that the variable was a risk factor, and *P* < 0.05 considered to indicate a statistically significant difference. Then, the factors that were significant in both univariate and multivariate logistic regression

TABLE 1 Baseline table of whether stroke patients died within 1 year.

Characteristics		Overall (N = 677)	No (N = 645)	Yes (N = 32)	P-value
Age, n (%)	≤60	383 (56.6)	362 (56.1)	21 (65.6)	0.381
	>60	294 (43.4)	283 (43.9)	11 (34.4)	
Gender, n (%)	Female	279 (41.2)	263 (40.8)	16 (50.0)	0.395
	Male	398 (58.8)	382 (59.2)	16 (50.0)	
SD, n (%)	Anterior circulation	270 (39.9)	258 (40.0)	12 (37.5)	0.082
	Posterior circulation	252 (37.2)	235 (36.4)	17 (53.1)	
	Anterior/posterior circulation	155 (22.9)	152 (23.6)	3 (9.4)	
SOH, n (%)	Left	283 (41.8)	271 (42.0)	12 (37.5)	0.87
	Right	270 (39.9)	256 (39.7)	14 (43.8)	
	Bilateral	124 (18.3)	118 (18.3)	6 (18.8)	
SOS, n (%)	Cortex	155 (22.9)	149 (23.1)	6 (18.8)	0.95
	Cortex-subcortex	155 (22.9)	147 (22.8)	8 (25.0)	
	Subcortex	186 (27.5)	176 (27.3)	10 (31.2)	
	Brainstem	104 (15.4)	100 (15.5)	4 (12.5)	
	Cerebellum	77 (11.4)	73 (11.3)	4 (12.5)	
NOS, n (%)	Single stroke lesion	470 (69.4)	453 (70.2)	17 (53.1)	0.064
	Multiple stroke lesions	207 (30.6)	192 (29.8)	15 (46.9)	
Thrombolysis, n (%)	No	473 (69.9)	448 (69.5)	25 (78.1)	0.398
	Yes	204 (30.1)	197 (30.5)	7 (21.9)	
Thrombectomy, n (%)	No	644 (95.1)	614 (95.2)	30 (93.8)	0.665
	Yes	33 (4.9)	31 (4.8)	2 (6.2)	
Antiplatelet, n (%)	No	122 (18.0)	117 (18.1)	5 (15.6)	0.9
	Yes	555 (82.0)	528 (81.9)	27 (84.4)	
Anticoagulation, n (%)	No	576 (85.1)	553 (85.7)	23 (71.9)	0.041
	Yes	101 (14.9)	92 (14.3)	9 (28.1)	
Statin, n (%)	No	103 (15.2)	98 (15.2)	5 (15.6)	1
	Yes	574 (84.8)	547 (84.8)	27 (84.4)	
PPI, n (%)	No	535 (79.0)	519 (80.5)	16 (50.0)	<0.001
	Yes	142 (21.0)	126 (19.5)	16 (50.0)	
SS, n (%)	No	385 (56.9)	380 (58.9)	5 (15.6)	<0.001
	Yes	292 (43.1)	265 (41.1)	27 (84.4)	
SAP, n (%)	No	512 (75.6)	494 (76.6)	18 (56.2)	0.016
	Yes	165 (24.4)	151 (23.4)	14 (43.8)	
SBP, median [Q1, Q3]		143.0 [132.0, 156.0]	143.0 [132.0, 156.0]	144.0 [134.8, 156.2]	0.678
DBP, median [Q1, Q3]		87.0 [74.0, 97.0]	87.0 [74.0, 97.0]	87.0 [74.0, 98.2]	0.492
Cholesterol, median [Q1, Q3]		5.3 [4.4, 6.2]	5.3 [4.4, 6.2]	5.4 [4.7, 6.0]	0.815
Triglyceride, median [Q1, Q3]		2.2 [1.9, 2.4]	2.2 [1.9, 2.4]	2.1 [1.9, 2.4]	0.801
LDL, median [Q1, Q3]		4.8 [4.3, 4.9]	4.8 [4.3, 4.9]	4.7 [4.5, 4.8]	0.69
FBG, median [Q1, Q3]		5.3 [4.6, 5.8]	5.2 [4.6, 5.8]	5.7 [5.2, 6.2]	0.003
HBALC, median [Q1, Q3]		5.6 [5.3, 5.9]	5.6 [5.3, 5.9]	5.8 [5.5, 6.1]	0.023
HCY, median [Q1, Q3]		15.8 [12.7, 19.4]	15.5 [12.6, 19.1]	20.4 [17.8, 22.7]	<0.001
UA, median [Q1, Q3]		349.8 [309.8, 408.1]	350.1 [310.8, 407.6]	335.5 [290.7, 435.4]	0.573

(Continued)

TABLE 1 (Continued)

Characteristics		Overall (N = 677)	No (N = 645)	Yes (N = 32)	P-value
MB, median [Q1, Q3]		97.7 [75.1, 147.8]	97.0 [74.9, 144.7]	106.6 [78.9, 236.5]	0.078
CRP, median [Q1, Q3]		12.6 [7.7, 17.6]	11.9 [7.5, 17.1]	20.5 [17.3, 25.0]	<0.001
FIB, median [Q1, Q3]		4.3 [4.0, 4.8]	4.4 [4.0, 4.8]	4.2 [3.8, 4.6]	0.134
D-dimer, median [Q1, Q3]		174.0 [133.0, 221.0]	174.0 [133.0, 221.0]	171.5 [132.0, 216.5]	0.844
BNP, median [Q1, Q3]		93.0 [73.0, 162.0]	93.0 [73.0, 162.0]	121.5 [77.0, 177.2]	0.25
NSE, median [Q1, Q3]		16.2 [12.7, 18.6]	16.2 [12.7, 18.6]	17.6 [12.5, 19.4]	0.197
S100β, median [Q1, Q3]		275.0 [224.0, 290.0]	275.0 [223.0, 289.0]	278.0 [248.2, 311.2]	0.111

Overall: All patients, No, Patients who did not die, Yes, Patients who died.
SD, stroke distribution; SOH, side of hemisphere; NOS, number of stroke lesions; SOS, site of stroke lesions; LDL, low-density lipoprotein; FBG, fasting blood glucose; HCY, homocysteine; UA, uric acid; FIB, fibrinogen; MB, myoglobin; CRP, C-reactive protein; BNP, brain natriuretic peptide; NSE, neuron-specific enolase; PPI, proton pump inhibitor therapy; SAP, stroke-associated pneumonia.

were included and subjected to stepwise regression analysis. The factors selected by stepwise regression were used as input variables to construct ML models.

The ML algorithm process was based on Python (V3.7) software and the scikit-learn (version 0.24) library. First, the original dataset was randomly divided into training and test sets at a ratio of 7:3. Then, six machine algorithms [logistic regression (LR), gradient boosting machine (GBM), extreme gradient boosting (XGB), random forest (RF), decision tree (DT), and naive Bayes classifier (NBC)] were used to analyze the data and construct the model. To validate the predictive power of the model, the 10-fold cross-validation method was used for internal validation against the training group. The random search method was used to adjust the hyperparameters of the models.

In the test group, the area under the receiver operating characteristic curve (ROC-AUC), classification accuracy, recall, specificity, and F1 score were used to evaluate the prediction models. We also plotted the prediction recall curve (PRC) as a complementary metric to evaluate the model performance.

2.3.4. Interpretation of the model and importance of features

To illustrate the risk factors of death within 1 year in AIS patients, Shapley Additive explanation (SHAP) analysis was used to interpret the predictive models ranked in terms of feature importance. SHAP analysis is a tool proposed by Lloyd Shapley in game theory to explain the output of machine learning models. The core idea is to calculate the marginal contribution of a variable feature when it is added to the model, and then to interpret the global and local levels of the “black box model” in an additive explanatory model (14, 15). That is, it can assign predictive values to each feature and evaluate and visualize the contribution of each feature to the outcome of the machine learning model (16). Ultimately, a web-based calculator based on the best-performing model was created for inputting patient data to facilitate the clinicians’ assessment of death within 1 year in AIS patients.

3. Results

3.1. Baseline patient data characteristics

In this study, clinical information was collected on 677 AIS patients, of whom 645 survived and 32 died of AIS (Table 1). In the observed population, 383 patients (56.6%) were aged < 60 years and 294 patients (43.4%) were aged ≥ 60 years, 398 (58.8%) were male and 279 (41.2%) were female. AIS lesions occurred in 270 (39.9%) patients in the anterior circulation, 252 (37.2%) in the posterior circulation, and 155 (22.9%) in both anterior and posterior circulations. The distribution of lesions in the left and right hemispheres was approximately equal, with 283 (41.8%) in the left and 270 (39.9%) in the right, and a relatively small number [124 (18.3%)] in the bilateral cerebral hemispheres. The location of the lesions was mainly subcortical in 186 patients (27.5%), cortical and cortico-subcortical in 155 patients (22.9%), the brainstem in 104 patients (15.4%), and the cerebellum in 77 patients (11.4%). Four hundred and seventy patients (69.4%) had a single AIS lesion, while multiple lesions were found in only 207 patients (30.6%). Two hundred and four patients (30.1%) received intravenous thrombolytic therapy, and 473 (69.9%) did not. Six hundred and forty-four patients (95.1%) did not receive arterial thrombolytic therapy, and 33 (4.9%) did. Five hundred and fifty-five patients (82.0%) received antiplatelet therapy, and 122 (18.0%) did not. Five hundred and seventy-six patients (85.1%) did not receive anticoagulation therapy, and 101 (14.9%) did. The majority of the patients (574, 84.8%) received statin therapy and 103 (15.2%) did not.

The median systolic and diastolic blood pressure was 143 mmHg (IQR 132.0,156.0) and 87 mmHg (IQR 74.0, 97.0), respectively. Total cholesterol, triglycerides, HDL, blood HCY, blood UA, and median FIB, MB, ultrasensitive CRP, D-dimer BNP, atrial natriuretic peptide, NSE, and S-100β were 5.3 mmol/L [4.4, 6.2], 2.2 mmol/L [1.9, 2.4], 4.8 mmol/L [4.3, 4.9], 15.7 μmol/L [12.4, 19.1], 349.8 μmol/L [309.8, 408.1], 4.3 g/L [4.0, 4.8], 97.7 ng/mL [75.1, 147.8], 12.2 mg/L [7.2, 18.1], 174.0 ng/mL [133.0, 221.0], 93.0 ng/mL [73.0, 162.0], 16.2 ng/mL [12.7, 18.6], and 275.0 ng/mL [224.0, 290.0], respectively.

TABLE 2 Univariate and multivariate logistic regression analysis of 1-year death in AIS patients.

Characteristics	Univariate logistic analysis			Multivariate logistic analysis		
	OR	95% CI	P-value	OR	95% CI	P-value
Age:						
≤60	Ref.	Ref.	Ref.			
>60	0.67	(0.31–1.40)	0.297			
Gender:						
Female		Ref.	Ref.			
Male	0.69	(0.33–1.42)	0.309			
SD:						
Anterior circulation	Ref.	Ref.	Ref.			
Posterior circulation	1.55	(0.72–3.41)	0.26			
Anterior/posterior circulation	0.44	(0.09–1.44)	0.186			
SOH:						
Left		Ref.	Ref.			
Right	1.23	(0.55–2.78)	0.607			
Bilateral	1.16	(0.39–3.11)	0.774			
SOS:						
Cortex		Ref.	Ref.			
Cortex-subcortex	1.34	(0.45–4.26)	0.599			
Subcortex	1.4	(0.50–4.27)	0.53			
Brainstem	1.01	(0.24–3.72)	0.993			
Cerebellum	1.38	(0.33–5.11)	0.642			
NOS:						
Single stroke lesion	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Multiple stroke lesions	2.08	(1.00–4.28)	0.049	3.44	(1.41–8.36)	0.007
Thrombolysis:						
No	Ref.	Ref.	Ref.			
Yes	0.65	(0.25–1.45)	0.305			
Thrombectomy:						
No	Ref.	Ref.	Ref.			
Yes	1.41	(0.20–5.00)	0.671			
Antiplatelet:						
No	Ref.	Ref.	Ref.			
Yes	1.17	(0.47–3.56)	0.755			
Anticoagulation:						
No	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Yes	2.37	(1.00–5.16)	0.049	0.97	(0.35–2.67)	0.951
Statin:						
No		Ref.	Ref.			
Yes	0.94	(0.38–2.89)	0.91			
PPI:						
No	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Yes	4.11	(1.98–8.54)	<0.001	1.65	(0.62–4.38)	0.317

(Continued)

TABLE 2 (Continued)

Characteristics	Univariate logistic analysis			Multivariate logistic analysis		
	OR	95% CI	P-value	OR	95% CI	P-value
SS:						
No	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Yes	7.53	(3.09–22.8)	<0.001	3.12	(1.03–9.83)	0.046
SAP:						
No	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Yes	2.55	(1.21–5.25)	0.015	0.98	(0.36–2.68)	0.971
SBP	1.01	(0.99–1.03)	0.242			
DBP	1.01	(0.99–1.04)	0.303			
Cholesterol	1	(0.75–1.32)	0.988			
Triglyceride	0.77	(0.27–2.16)	0.616			
LDL	1.08	(0.66–1.77)	0.769			
FBG	1.53	(1.09–2.14)	0.015	2.15	(0.89–5.21)	0.088
HBALC	2.34	(1.01–5.43)	0.048	0.56	(0.07–4.54)	0.59
HCY	1.32	(1.20–1.45)	<0.001	1.29	(1.16–1.45)	<0.001
UA	1	(0.99–1.00)	0.732			
MB	1.01	(1.00–1.01)	0.003	1	(0.99–1)	0.746
CRP	1.17	(1.11–1.22)	<0.001	1.15	(1.07–1.23)	<0.001
FIB	0.62	(0.35–1.10)	0.106			
D-dimer	1	(0.99–1.01)	0.761			
BNP	1	(1.00–1.01)	0.157			
NSE	1.06	(0.96–1.17)	0.214			
S100β	1.01	(1.00–1.02)	0.082			

3.2. Univariate and multivariate regression analysis of death within 1 year in AIS patients

In the univariate regression analysis of death within 1 year in AIS patients (Table 2), there was a statistically significant difference ($P < 0.05$) in the overall population for death within 1 year according to NOS, FBG, HBALC, MB, and CRP levels, as well as anticoagulation therapy, PPI treatment, and SS.

All parameters that were statistically different in the univariate analysis above were included in the multivariate logistic regression analysis. The results suggested that NOS (OR = 3.44, 95% CI: 1.41 – 8.36, $P = 0.007$), HCY (OR = 1.29, 95% CI: 1.16 – 1.45, $P < 0.001$), CRP (OR = 1.15, 95% CI: 1.07 – 1.23, $P < 0.001$), and SS (OR = 3.12, 95% CI: 1.03 – 9.83, $P = 0.046$) were independent predictors of death within 1 year in AIS patients.

3.3. Machine learning model building and validation

To compare the predictive performance of the six ML algorithm models, this study performed 10-fold cross-validation within the

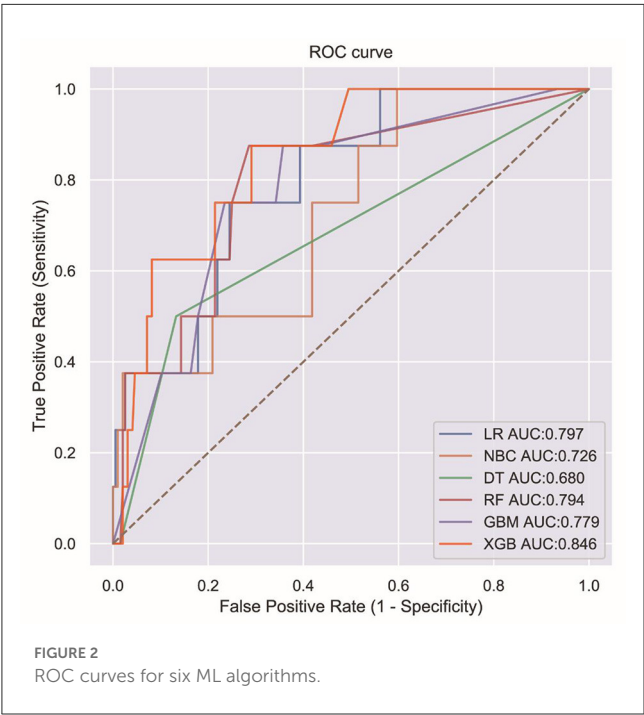
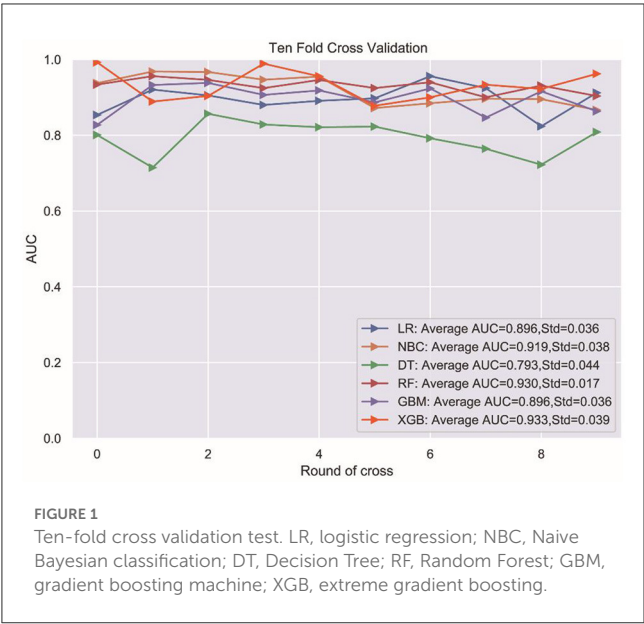
training group. The results are shown in Figure 1. Figure 2 shows the ROC curves of the predictive performance differences of the six ML algorithm models after external validation, and Figure 3 shows the result of radar plot analysis, which is a blanket, clear, intuitive, and easy-to-judge analysis and is suitable for comprehensive evaluation as it can show the AUC value, accuracy, recall, and F1 value of the models in multiple dimensions (Figure 3, Table 3) to more clearly reflect the performance of the models. The PRC curves of the mortality prediction model are shown in Supplementary Figure 1.

The results suggest that the XGB model performed best in predicting death within 1 year in AIS patients after a comprehensive evaluation. The remaining models were ranked in descending order according to their predictive performance.

In summary, we finally adopted the XGB model as the preferred predictive model.

3.4. Relative importance of variables in ML algorithms

A SHAP interpretability study was used to analyze the results of the ML models. Generally, the higher the SHAP value of a feature,



the higher the probability of the occurrence of the target event. In SHAP analysis, red represents the eigenvalues with positive impact on the model and blue represents the eigenvalues with negative impact on the model (17). The results of the study suggest that SS was the most important variable, followed by CRP, HCY, and NOS in descending order of importance, as shown in Figure 4.

3.5. The web calculator

A web-based calculator based on the XGB model was developed in this study. By entering the clinical characteristic variables of a patient with AIS, clinicians could predict their risk of death

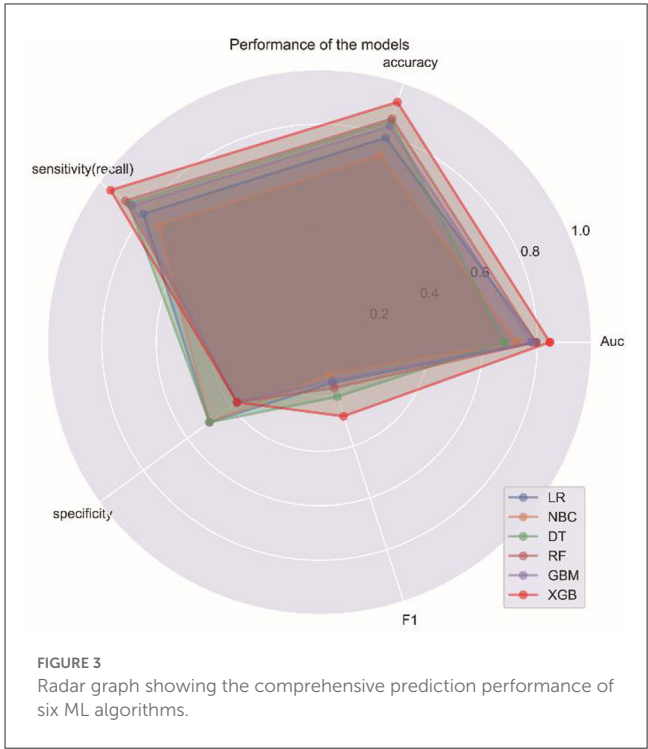


TABLE 3 The result of specific performance of six ML algorithm models.

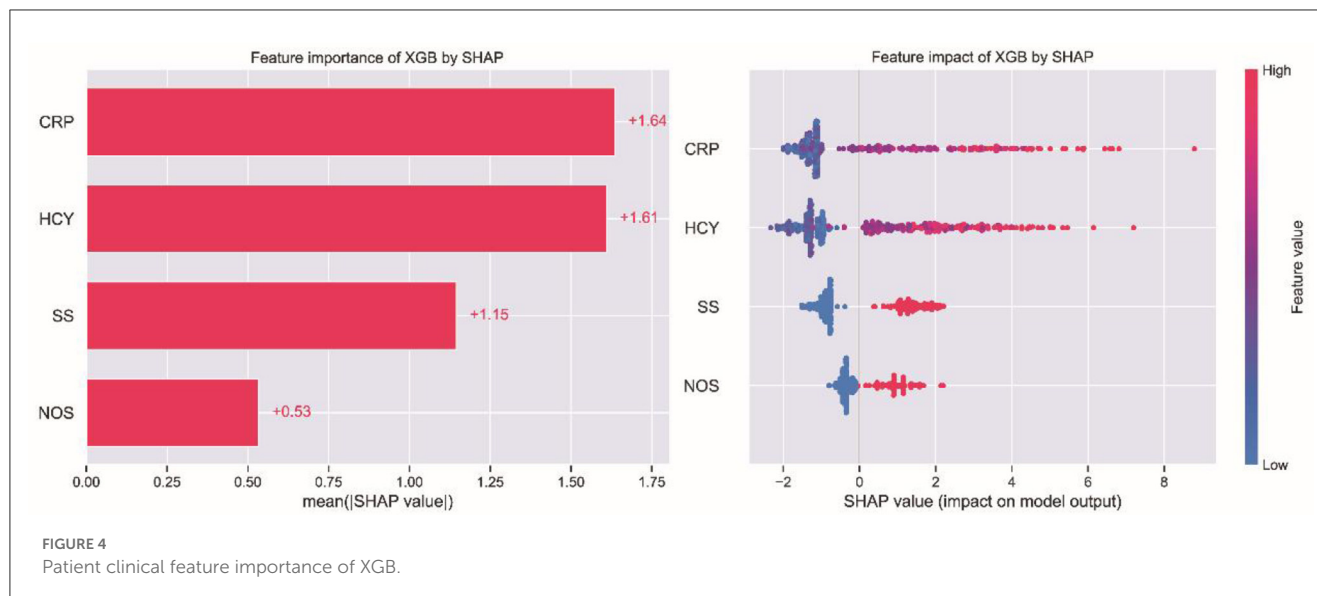
Scoring	LR	NBC	DT	RF	GBM	XGB
Auc	0.797	0.726	0.680	0.794	0.779	0.846
Accuracy	0.789	0.721	0.853	0.863	0.833	0.926
Sensitivity (recall)	0.801	0.730	0.867	0.883	0.852	0.949
Specificity	0.500	0.500	0.500	0.375	0.375	0.375
F1	0.157	0.123	0.211	0.176	0.150	0.286

within 1 year (<https://mlmedicine-de-stroke-de-stroke-m5pijk.streamlitapp.com/>; Figure 5).

4. Discussion

In this study, we retrospectively analyzed the clinical data of AIS patients and developed a web-based calculator with ML algorithms to predict the risk of death within 1 year. The accuracy and rationality of the model were validated by 10-fold cross-validation, allowing the model to be used for clinical practice to help clinicians make more rational treatment decisions.

ML is an emerging field of medicine that has demonstrated an extraordinary ability to handle large, complicated, and disparate data, and is the future of biomedical research, personalized medicine, and computer-aided diagnosis. It holds the promise of significantly advancing global healthcare (18, 19). Unlike traditional predictive models, ML is very good at discovering complex structures in selected variables in high-dimensional data and can easily combine a large number of variables (20, 21). ML has been reported to improve the predictive accuracy of long-term prognoses for AIS patients (8, 10).



In this study, six ML methods were used to analyze and construct a model of death prediction within 1 year in AIS patients, and the performance of the six ML algorithms was compared to each other. The XGB algorithm performed best (Figure 1), with a better AUC value than the other five algorithmic models, and the highest accuracy, sensitivity, and F1 score. Therefore, the XGB algorithm model was finally chosen.

ML models are often considered to be a black box where it is difficult to explain the predictive performance, and it becomes extremely important to study the interpretability of machine learning models. Therefore, this study attempted to introduce SHAP analysis, a new method for interpreting various black-box ML models that have been previously validated based on their interpretability performance. It can achieve both local and global interpretability and has a solid theoretical foundation compared to other methods (22). The SHAP analysis used in this study could interpret the model prediction results well, and its intuitive visualization is more easily accepted. This study further built a web-based calculator to estimate the probability of death within 1 year in AIS patients to make better use of the model.

AIS is characterized by a high morbidity rate, which increases the economic burden on society and families (23). It is significant to explore the factors influencing the risk of death within 1 year for patients. In this study, the mortality rate of AIS patients within 1 year was only 4.7% (32/677), which was significantly lower than the 10% reported in previous studies (7), probably because of the exclusion of those whose families discontinued treatment for various reasons. Previously, an 8-point scoring system was constructed to predict the risk of death within 7 days of hospitalization (24). Factors influencing death within 6 months of stroke onset were also reported, with variables such as the Barthel index and platelet/lymphocyte ratio screened by LASSO regression and multiple logistic regression (25). A 30-year stroke burden predictive model was established (26). In contrast, unlike many previous studies, this study innovatively used machine learning algorithms to screen variables and, to our knowledge, was the first to develop a predictive model using machine learning algorithms to assess the probability of death within 1 year in patients with AIS.

Machine Learning Application for Predicting One-year Death

Side of hemisphere: HCY (μmol/L): CRP (mg/L):

Stroke severity:

Probability: 0.4%

FIGURE 5
The web-based calculator for predicting 1-year death in AIS patients.

There is a growing body of research on the relationship between serum inflammatory biomarkers and AIS. A number of studies showed that AIS could induce an inflammatory response, which plays a major role in late ischemic damage to the brain parenchyma, and that inflammatory responses caused by various clinical factors could lead to an increase in inflammatory factors (27, 28). It is also an inflammatory factor that can indirectly indicate the presence of pathogenic microorganisms in patients when it is upraised, which can help the physician in the diagnosis and treatment. In this study, we concluded that CRP levels were the most important predictor of death within 1 year in AIS patients. Elevated CRP levels were previously reported to reflect the severity of AIS, correlate with stroke subtype and risk stratification (27, 28), and be an independent predictor of long-term mortality after ischemic stroke (29). Elevated CRP levels can lead to increased mortality after stroke, which may be related to inflammation-induced endothelial cell dysfunction and platelet activation (30). HCY is a sulfur-containing non-essential amino acid produced by metabolism *in vivo* as a derivative of methionine cycle demethylation. It is

also an inflammatory substance that induces the activation of nuclear factor (NF)- κ B, which is a transcription factor common to inflammation and the immune response. Elevated levels of HCY are associated with a variety of diseases, which may lead to endothelial dysfunction, neurotoxicity, and the upregulation of thrombogenic factors. At the same time, monitoring HCY levels may provide a good indication of the development of related diseases (31). Previous studies also showed that elevated HCY levels were associated with AIS dysfunction and recurrent stroke (32). A multicenter study suggested that high levels of serum HCY were an independent predictor of early neurological deterioration in AIS patients (33). This study concluded that HCY levels significantly influenced the risk of death within 1 year in AIS patients. The risk of death in patients with high HCY having 1.29 times (95% CI 1.16 – 1.45) compared to ones with normal HCY.

The NIHSS is a common scale used in neurology as a quantitative indicator of disease severity (34). The present study classified SS with the help of the NIHSS scale, and an NIHSS score of ≥ 9 was defined as moderate-to-severe stroke. Fischer et al. (35) suggested that patients with low NIHSS scores tended to have a better prognosis, which is consistent with the current study. The present study concluded that SS had a significant influence on death within 1 year in AIS patients. The risk of death in patients with moderate to severe stroke having 3.12 times (95% CI 1.03 – 9.83) higher than those with mild stroke.

Neurological deficits have been associated with lesions in different brain regions (36, 37), but the relationship between the number of lesions and AIS has rarely been reported. In this study, the number of lesions was innovatively included in the analysis, and the results suggested that the number of lesions was a significant factor in death within 1 year in AIS patients. The risk of death in patients with multiple lesions was 3.44 times (95% CI: 1.41 – 8.36) higher than patients with a single lesion.

There were several limitations to this study. First, the retrospective study design may have introduced selection bias, while the data imbalance that emerged from real-world studies resulted in PRC effects without the AUC number. Secondly, although our model showed good performance, its data source was limited to one medical center, which may limit its generalizability, and we will follow up with an additional multicenter study. Thirdly, further independent external validation is needed to confirm these findings. Finally, we collected AIS-related variables as comprehensively as possible, but there were still some important variables that were not available in a timely manner, which may also limit the generalizability of the study. Future research is needed to examine this issue further.

5. Conclusion

The results of this study suggest that serum inflammatory markers (CRP and HCY), SS, and NOS are independent risk factors of death within 1 year in AIS patients. The XGB algorithm showed good performance as a tool to predict death within 1 year in AIS patients. Using this web-based calculator can effectively prevent death, reduce mortality, and assist physicians in making treatment decisions.

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 study was written approved by the Ethics Committee of the Second Affiliated Hospital of Xuzhou Medical University [ethics number: [2020] 081603]. The patients/participants provided their written informed consent to participate in this study.

Author contributions

WLi, LR, and XW completed the study design. KW, WLi, and WLi performed the study and collected and analyzed the data. LG and WLi drafted the manuscript. LR, XW, KW, and HL provided the expert consultations and suggestions. CX and CY conceived of the study, participated in its design and coordination, and helped to embellish language. All authors reviewed the final version of the manuscript.

Funding

This study was supported by Scientific Research Project of Jiangsu Health Committee (No. H2019054), the Xuzhou Science and Technology Planning Project (No. KC21220), Development Fund of Affiliated Hospital of Xuzhou Medical University (No. XYFY2020013), and Shaanxi Provincial Health and Health Research Fund Project (2022E006).

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2023.1092534/full#supplementary-material>

SUPPLEMENTARY FIGURE 1
Precision recall curve (PRC) for machine learning algorithm.

References

- Ding Q, Liu S, Yao Y, Liu H, Cai T, Han L. Global, regional, and national burden of ischemic stroke, 1990–2019. *Neurology*. (2022) 98:e279–90. doi: 10.1212/WNL.00000000000013115
- Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *Lancet*. (2020) 396:1204–22. doi: 10.1016/S0140-6736(20)30925-9
- Mendelson SJ, Prabhakaran S. Diagnosis and management of transient ischemic attack and acute ischemic stroke: A review. *J Am Med Assoc*. (2021) 325:1088–98. doi: 10.1001/jama.2020.26867
- Jayaraj RL, Azimullah S, Beiram R, Jalal FY, Rosenberg GA. Neuroinflammation: friend and foe for ischemic stroke. *J Neuroinflammation*. (2019) 16:142. doi: 10.1186/s12974-019-1516-2
- Lasek-Bal A, Jedrzejowska-Szypulka H, Student S, Warsz-Wianecka A, Zareba K, Puz P, et al. The importance of selected markers of inflammation and blood-brain barrier damage for short-term ischemic stroke prognosis. *J Physiol Pharmacol*. (2019) 70:503953. doi: 10.1101/503953
- Xu S, Lu J, Shao A, Zhang JH, Zhang J. Glial cells: Role of the immune response in ischemic stroke. *Front Immunol*. (2020) 11:294. doi: 10.3389/fimmu.2020.00294
- Morotti A, Poli L, Costa P. Acute stroke. *Semin Neurol*. (2019) 39:61–72. doi: 10.1055/s-0038-1676992
- 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
- Chae SH, Kim Y, Lee KS, Park HS. Development and clinical evaluation of a web-based upper limb home rehabilitation system using a smartwatch and machine learning model for chronic stroke survivors: Prospective comparative study. *JMIR mHealth uHealth*. (2020) 8:e17216. doi: 10.2196/17216
- Castaneda-Vega S, Katiyar P, Russo F, Patzwaldt K, Schnabel L, Mathes S, et al. Machine learning identifies stroke features between species. *Theranostics*. (2021) 11:3017–34. doi: 10.7150/thno.51887
- Bonkhoff A, Hong S, Bretzner M, Schirmer M, Regenhardt R, Arsava E, et al. Association of stroke lesion pattern and white matter hyperintensity burden with stroke severity and outcome. *Neurology*. (2022) 99:e1364–79. doi: 10.1212/WNL.000000000000200926
- Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to machine learning, neural networks, and deep learning. *Transl Vis Sci Technol*. (2020) 9:14. doi: 10.1167/tvst.9.2.14
- Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. *Nat Rev Mol Cell Biol*. (2022) 23:40–55. doi: 10.1038/s41580-021-00407-0
- Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Informat Process Syst*. (2017) 2017:4765–74. doi: 10.48550/arXiv.1705.07874
- Chen H, Lundberg SM, Lee SI. Explaining a series of models by propagating Shapley values. *Nat Commun*. (2022) 13:4512. doi: 10.1038/s41467-022-31384-3
- Nohara Y, Matsumoto K, Soejima H, Nakashima N. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Comput Methods Programs Biomed*. (2022) 214:106584. doi: 10.1016/j.cmpb.2021.106584
- Tseng PY, Chen YT, Wang CH, Chiu KM, Peng YS, Hsu SP, et al. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Crit Care*. (2020) 24:478. doi: 10.1186/s13054-020-03179-9
- Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: Machine learning and the future of medicine. *J Intern Med*. (2018) 284:603–19. doi: 10.1111/joim.12822
- Orchard P, Agakova A, Pinnock H, Burton CD, Sarran C, Agakov F, et al. Improving prediction of risk of hospital admission in chronic obstructive pulmonary disease: Application of machine learning to telemonitoring data. *J Med Internet Res*. (2018) 20:e263. doi: 10.2196/jmir.9227
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. (2015) 521:436–44. doi: 10.1038/nature14539
- Shek A, Jiang Z, Teo J, Au Yeung J, Bhalla A, Richardson MP, et al. Machine learning-enabled multitrust audit of stroke comorbidities using natural language processing. *Eur J Neurol*. (2021) 28:4090–7. doi: 10.1111/ene.15071
- Ou C, Liu J, Qian Y, Chong W, Zhang X, Liu W, et al. Rupture risk assessment for cerebral aneurysm using interpretable machine learning on multidimensional data. *Front Neurol*. (2020) 11:570181. doi: 10.3389/fneur.2020.570181
- Herpich F, Rincon F. Management of acute ischemic stroke. *Crit Care Med*. (2020) 48:1654–63. doi: 10.1097/CCM.00000000000004597
- Myint PK, Clark AB, Kwok CS, Davis J, Durairaj R, Dixit AK, et al. The SOAR (Stroke subtype, Oxford Community Stroke Project classification, Age, prestroke modified Rankin) score strongly predicts early outcomes in acute stroke. *Int J Stroke*. (2014) 9:278–83. doi: 10.1111/ijss.12088
- Sha L, Xu T, Ge X, Shi L, Zhang J, Guo H. Predictors of death within 6 months of stroke onset: A model with Barthel index, platelet/lymphocyte ratio and serum albumin. *Nurs Open*. (2021) 8:1380–92. doi: 10.1002/nop.2.754
- Wafa HA, Wolfe CDA, Emmett E, Roth GA, Johnson CO, Wang Y. Burden of stroke in Europe: Thirty-year projections of incidence, prevalence, deaths, and disability-adjusted life years. *Stroke*. (2020) 51:2418–27. doi: 10.1161/STROKEAHA.120.029606
- Luo Y, Wang Z, Li J, Xu Y. Serum CRP concentrations and severity of ischemic stroke subtypes. *Can J Neurol Sci*. (2012) 39:69–73. doi: 10.1017/S0317167100012713
- Hu Z, Lai J, Chen L, Yi Y, Li R, Liao W. Can baseline C-reactive protein level predict functional outcome in acute ischaemic stroke? A meta-analysis. *Biomarkers*. (2020) 25:525–32. doi: 10.1080/1354750X.2020.1817982
- Idicula TT, Brogger J, Naess H, Waje-Andreassen U, Thomassen L. Admission C-reactive protein after acute ischemic stroke is associated with stroke severity and mortality: The “Bergen stroke study”. *BMC Neurol*. (2009) 9:18. doi: 10.1186/1471-2377-9-18
- Shantikumar S, Grant PJ, Catto AJ, Bamford JM, Carter AM. Elevated C-reactive protein and long-term mortality after ischaemic stroke: Relationship with markers of endothelial cell and platelet activation. *Stroke*. (2009) 40:977–9. doi: 10.1161/STROKEAHA.108.525105
- Luo Y, Jin H, Guo ZN, Zhang P, Zhang LY, Chen J, et al. Effect of hyperhomocysteinemia on clinical outcome and hemorrhagic transformation after thrombolysis in ischemic stroke patients. *Front Neurol*. (2019) 10:592. doi: 10.3389/fneur.2019.00592
- He Y, Li Y, Chen Y, Feng L, Nie Z. Homocysteine level and risk of different stroke types: A meta-analysis of prospective observational studies. *Nutr Metabol Cardiovasc Dis*. (2014) 24:1158–65. doi: 10.1016/j.numecd.2014.05.011
- Kwon HM, Lee YS, Bae HJ, Kang DW. Homocysteine as a predictor of early neurological deterioration in acute ischemic stroke. *Stroke*. (2014) 45:871–3. doi: 10.1161/STROKEAHA.113.004099
- Garavelli F, Ghelfi AM, Kilstein JG. Usefulness of NIHSS score as a predictor of non-neurological in-hospital complications in stroke. *Med Clin*. (2021) 157:434–7. doi: 10.1016/j.medcli.2020.07.034
- Fischer U, Baumgartner A, Arnold M, Nedeltchev K, Gralla J, De Marchis GM, et al. What is a minor stroke? *Stroke*. (2010) 41:661–6. doi: 10.1161/STROKEAHA.109.572883
- Fröhlich K, Macha K, Gerner ST, Bobinger T, Schmidt M, Dörfler A, et al. Angioedema in stroke patients with thrombolysis. *Stroke*. (2019) 50:1682–7. doi: 10.1161/STROKEAHA.119.025260
- Pillay SB, Binder JR, Humphries C, Gross WL, Book DS. Lesion localization of speech comprehension deficits in chronic aphasia. *Neurology*. (2017) 88:970–5. doi: 10.1212/WNL.0000000000003683



OPEN ACCESS

EDITED BY

Daniel Donoho,
Children's National Hospital, United States

REVIEWED BY

Ping Hu,
Second Affiliated Hospital of Nanchang
University, China
Hui Jan Tan,
National University of Malaysia, Malaysia

*CORRESPONDENCE

Yuechao Fan
✉ fyc626@163.com
Jinhong Feng
✉ feng_jh@189.com.cn

[†]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 06 January 2023

ACCEPTED 08 March 2023

PUBLISHED 03 April 2023

CITATION

Li F, Chen A, Li Z, Gu L, Pan Q, Wang P, Fan Y
and Feng J (2023) Machine learning-based
prediction of cerebral hemorrhage in patients
with hemodialysis: A multicenter, retrospective
study. *Front. Neurol.* 14:1139096.
doi: 10.3389/fneur.2023.1139096

COPYRIGHT

© 2023 Li, Chen, Li, Gu, Pan, Wang, Fan and
Feng. This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Machine learning-based prediction of cerebral hemorrhage in patients with hemodialysis: A multicenter, retrospective study

Fengda Li^{1†}, Anmin Chen^{2†}, Zeyi Li³, Longyuan Gu⁴, Qiyang Pan⁵,
Pan Wang³, Yuechao Fan^{4*} and Jinhong Feng^{6*}

¹Department of Neurosurgery, Changshu Hospital Affiliated to Soochow University, Changshu, China,

²Department of Nephrology, The First People's Hospital of Jintan, Changzhou, China, ³School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing, China, ⁴Department of Neurosurgery, Affiliated Hospital of Xuzhou Medical University, Xuzhou, China, ⁵Faculty of Informatics, Università della Svizzera italiana, Lugano, Ticino, Switzerland, ⁶Department of Nephrology, Affiliated Hospital of Xuzhou Medical University, Xuzhou, China

Background: Intracerebral hemorrhage (ICH) is one of the most serious complications in patients with chronic kidney disease undergoing long-term hemodialysis. It has high mortality and disability rates and imposes a serious economic burden on the patient's family and society. An early prediction of ICH is essential for timely intervention and improving prognosis. This study aims to build an interpretable machine learning-based model to predict the risk of ICH in patients undergoing hemodialysis.

Methods: The clinical data of 393 patients with end-stage kidney disease undergoing hemodialysis at three different centers between August 2014 and August 2022 were retrospectively analyzed. A total of 70% of the samples were randomly selected as the training set, and the remaining 30% were used as the validation set. Five machine learning (ML) algorithms, namely, support vector machine (SVM), extreme gradient boosting (XGB), complement Naïve Bayes (CNB), K-nearest neighbor (KNN), and logistic regression (LR), were used to develop a model to predict the risk of ICH in patients with uremia undergoing long-term hemodialysis. In addition, the area under the curve (AUC) values were evaluated to compare the performance of each algorithmic model. Global and individual interpretive analyses of the model were performed using importance ranking and Shapley additive explanations (SHAP) in the training set.

Results: A total of 73 patients undergoing hemodialysis developed spontaneous ICH among the 393 patients included in the study. The AUC of SVM, CNB, KNN, LR, and XGB models in the validation dataset were 0.725 (95% CI: 0.610 ~ 0.841), 0.797 (95% CI: 0.690 ~ 0.905), 0.675 (95% CI: 0.560 ~ 0.789), 0.922 (95% CI: 0.862 ~ 0.981), and 0.979 (95% CI: 0.953 ~ 1.000), respectively. Therefore, the XGBoost model had the best performance among the five algorithms. SHAP analysis revealed that the levels of LDL, HDL, CRP, and HGB and pre-hemodialysis blood pressure were the most important factors.

Conclusion: The XGB model developed in this study can efficiently predict the risk of a cerebral hemorrhage in patients with uremia undergoing long-term hemodialysis and can help clinicians to make more individualized and rational clinical decisions. ICH events in patients undergoing maintenance hemodialysis (MHD) are associated with serum LDL, HDL, CRP, HGB, and pre-hemodialysis SBP levels.

KEYWORDS

hemodialysis, uremia, intracerebral hemorrhage, machine learning, predictive models, Shapley additive explanations

1. Introduction

Maintenance hemodialysis (MHD) is the primary renal replacement therapy for patients with uremia (1). Intracerebral hemorrhage (ICH), defined as non-traumatic hemorrhage in the brain parenchyma with or without ventricles, accounts for 10–15% of all stroke cases and is an important cause of disability and death globally (2). ICH is one of the most serious complications among patients undergoing MHD. Various factors have an important impact on the occurrence and development of ICH. Recent studies have attempted to identify relevant risk factors, and lipid metabolism and inflammatory responses have been reported as important factors regulating the progression of ICH and subsequent brain injury and brain function repair. Despite the continuous development of hemodialysis technology and the gradual improvement of nursing levels, the risk of a cerebral hemorrhage in patients undergoing MHD is approximately six times higher than that in healthy individuals (3), and the mortality rate is as high as 41–47% (4). Most patients require admission to the intensive care unit (ICU) for monitoring and treatment, which imposes a serious economic burden on the family and society.

ICH often has no identifiable warning signs or symptoms. Although optimal strategies for the medical and surgical management of ICH have been investigated, survival and functional outcomes have not been significantly improved (5). Therefore, establishing risk prediction models to identify high-risk patients undergoing MHD is important for the early implementation of targeted interventions. To date, only a few studies have attempted to develop such models.

Machine learning (ML), an artificial intelligence method, uses computers to statistically learn from datasets and build corresponding models to identify relationships between various factors. In the field of medicine, ML is increasingly used through statistical learning methods to overcome possible obstacles in clinical practice (6, 7). In recent years, although ML has been used to analyze clinical data to predict the complications and adverse outcomes of critical illnesses (8–10), few efforts have been made to develop strategies for predicting the prognosis of patients with uremia undergoing dialysis, especially for predicting the risk of cerebral hemorrhage, a serious complication of dialysis. ML has shown good performance in previous studies; however, because of its “black box” nature, the effects of each feature on the final results remain unknown, and it is difficult to explain the factors that lead to a given prediction. This lack of interpretability limits the widespread application of ML methods in medical research (11, 12). Shapley additive explanation (SHAP) is a method inspired by the classical game theory that assigns a predicted value to each feature and evaluates the contribution of each feature to the results of ML models to achieve a balance between the accuracy and interpretability of the model (13).

To analyze complex variables that may be related to a cerebral hemorrhage after regular hemodialysis, we integrated the demographic data, laboratory test results, hemodialysis indicators, and other information of patients to construct a model for predicting the risk of a cerebral hemorrhage. To make the model more applicable for the diagnosis of chronic kidney disease with intracerebral hemorrhage, overcome the “black box” nature of ML,

TABLE 1 Baseline features of patients.

Variables	Non-ICH (n = 320)	ICH (n = 73)	P-value
Age (years)	57.000 (46.000, 66.000)	54.000 (48.000, 63.000)	0.384
Sex (%)			0.697
Female	108 (33.7)	27 (37.0)	
Male	212 (66.3)	46 (63.0)	
Hypertension (%)			<0.001
No	141 (44.1)	13 (17.8)	
Yes	179 (55.9)	60 (82.2)	
Diabetes mellitus (%)			0.001
No	196 (61.2)	60 (82.2)	
Yes	124 (38.8)	13 (17.8)	
Polycystic kidney (%)			0.001
No	311 (97.2)	64 (87.7)	
Yes	9 (2.8)	9 (12.3)	
Duration of dialysis (months)	41.84 (19.27, 67.99)	34.17 (18.67, 43.97)	0.014
WBCs (10 ⁹ /L)	5.60 (4.58, 6.82)	6.60 (4.91, 8.85)	<0.001
PLTs (10 ⁹ /L)	165.00 (131.75, 204.25)	145.00 (117.00, 180.00)	0.009
HGB (g/L)	106.55 (16.77)	92.43 (13.26)	<0.001
NE (10 ⁹ /L)	3.82 (3.02, 4.91)	5.16 (3.51, 7.47)	<0.001
LY (10 ⁹ /L)	1.00 (0.80, 1.30)	0.82 (0.60, 1.10)	<0.001
HCT (%)	33.27 (5.71)	31.87 (6.07)	0.061
CRP (mg/L)	5.16 (3.16, 7.20)	12.00 (4.79, 31.48)	<0.001
NLR	3.74 (2.775, 5.185)	5.40 (3.27, 12.26)	<0.001
PLR	165.79 (129.983, 214.580)	188.89 (117.00, 250.00)	0.156
ALT (U/L)	9.00 (6.00, 14.00)	10.00 (7.00, 15.00)	0.136
AST (U/L)	12.00 (9.00, 15.00)	13.00 (10.00, 17.00)	0.069
TP (g/L)	66.60 (62.80, 70.53)	70.30 (64.10, 74.60)	<0.001
ALB (g/L)	40.20 (37.58, 42.70)	41.00 (37.50, 43.80)	0.329
BUN (mmol/L)	23.54 (18.07, 29.03)	20.12 (14.87, 26.48)	0.003
Scr (umol/L)	787.15 (650.45, 1,010.25)	746.00 (587.00, 905.00)	0.071
CysC mg/L	5.29 (4.58, 5.91)	4.76 (3.78, 5.38)	<0.001

(Continued)

TABLE 1 (Continued)

Variables	Non-ICH (<i>n</i> = 320)	ICH (<i>n</i> = 73)	<i>P</i> -value
UA (umol/L)	368.00 (296.50, 431.75)	365.00 (325.70, 409.00)	0.817
TG (mmol/L)	1.29 (0.91, 2.00)	1.53 (1.14, 2.49)	<0.001
TC (mmol/L)	3.67 (3.07, 4.25)	3.55 (3.31, 4.04)	0.933
LDL (mmol/L)	1.77 (1.51, 1.89)	1.31 (1.21, 1.44)	<0.001
HDL (mmol/L)	1.78 (1.25, 2.38)	1.09 (0.86, 1.47)	<0.001
K (mmol/L)	4.80 (4.26, 5.37)	4.77 (4.24, 5.25)	0.971
Na (mmol/L)	137.60 (135.40, 140.00)	136.70 (134.90, 138.30)	0.053
Ca (mmol/L)	2.155 (2.01, 2.27)	2.180 (2.00, 2.43)	0.288
P (mmol/L)	1.775 (1.41, 2.13)	1.650 (1.32, 2.16)	0.503
Calcium-phosphorus product (mg/dL)	46.90 (36.45, 57.742)	44.02 (33.97, 63.88)	0.498
eGFR (mL/min)	5.60 (4.42, 7.36)	6.10 (4.84, 8.33)	0.095
Hemodialysis vascular access (%)			0.302
Arteriovenous fistula	256 (80.00)	62 (84.93)	
Artificial blood vessel	9 (2.81)	0 (0.00)	
Central venous catheter	55 (17.19)	11 (15.07)	
Total anticoagulant (IU)	4,500.00 (4,000.00, 5,000.00)	4,500.00 (4,000.00, 5,000.00)	0.421
Blood flow rate (mL/min)	240.00 (220.00, 250.00)	250.00 (230.00, 260.00)	<0.001
Pre-hemodialysis SBP (mmHg)	142.00 (130.00, 155.00)	160.00 (148.00, 173.00)	<0.001
Pre-hemodialysis DBP (mmHg)	80.00 (75.00, 88.00)	82.00 (77.00, 89.00)	0.237
Post-hemodialysis SBP (mmHg)	139.52 ± 18.03	151.41 ± 20.82	<0.001
Post-hemodialysis DBP (mmHg)	80.29 ± 8.72	78.44 ± 9.51	0.108

WBCs, white blood cells; PLTs, platelets; HGB, hemoglobin; Ne, neutrophil; Ly, lymphocyte; HCT, hematocrit; CRP, C-reactive protein; NLR, neutrophil-to-lymphocyte ratio; PLR, platelet-to-lymphocyte ratio; ALT, alanine aminotransferase; AST, aspartate aminotransferase; TP, serum total protein; ALB, serum albumin; BUN, blood urea nitrogen; Scr, serum creatinine; CysC, cystatin C; eGFR, estimated glomerular filtration rate; UA, uric acid; TG, triglyceride; TC, total cholesterol; LDL, low-density lipoprotein; HDL, high-density lipoprotein; K, blood potassium; Na, blood sodium; Ca, blood calcium; P, blood phosphorus.

and explore the relationship between each feature and its clinical significance, we used the extreme gradient boosting (XGBoost) algorithm to develop the model (14). SHAP was used to provide a more intuitive global and local explanation of the model to understand the prediction of the model and improve the clinical understanding of the risk of a cerebral hemorrhage in patients with hemodialysis.

2. Materials and methods

2.1. Study population and data source

Patients with end-stage kidney disease undergoing hemodialysis from August 2014 to August 2022 at the Affiliated Hospital of Xuzhou Medical University, Xuzhou Central Hospital, and the Second Affiliated Hospital of Xuzhou Medical University were recruited for the study. According to the occurrence of ICH, the patients were divided into ICH and non-ICH groups.

2.2. Data collection

The inclusion criteria were as follows: (a) patients diagnosed with uremia according to chronic kidney disease (CKD) staging and recommendations or the Kidney Disease Outcomes Quality Initiative (KDOQI) guidelines formulated by the American Kidney Foundation, that is, patients with estimated glomerular filtration rate (eGFR) of <15 mL/(min·1.73 m²) diagnosed with CKD stage 5, which is the uremia stage (15); (b) patients receiving hemodialysis regularly, those aged ≥18 years, those with dialysis age of ≥3 months, and dialysis frequency of three times per week and 4 h per dialysis; and (c) patients with ICH confirmed *via* a CT examination of the head. The exclusion criteria were as follows: (a) patients with severe failure of the heart, lung, and other organs, blood system diseases, autoimmune diseases, and malignant tumors; (b) patients with primary subarachnoid hemorrhage, secondary cerebral hemorrhage, such as trauma, intracranial tumors, ICH caused by hemorrhage after an ischemic stroke, and severe coagulation dysfunction; (c) patients on antiplatelet drugs, hormones, immunosuppressants, and antibacterial agents in the past 1 month; and (d) patients with missing clinical data. Based on the diagnosis and inclusion and exclusion criteria, 393 patients with end-stage kidney disease complicated with cerebral hemorrhage owing to long-term hemodialysis were included. Of these 393 patients, 73 patients were included in the ICH group, whereas 320 patients were included in the non-ICH group. Because this study had a retrospective design, there was no security-related risk. The present study was approved by the Ethics Committee of the Affiliated Hospital of Xuzhou Medical University.

2.3. Inclusion of observed variables

The clinical data of patients were collected with reference to clinical experience, reported literature, and medical records in the electronic medical record systems of the three centers. Data

regarding the following five aspects were collected: (1) demographic data (sex and age); (2) vascular risk factors (hypertension, diabetes, polycystic kidney disease, and duration of dialysis); (3) baseline blood pressure (systolic blood pressure [SBP] and diastolic blood pressure [DBP] before and after dialysis); (4) treatment during hemodialysis (including anticoagulant dosage, dialysis access, and blood flow velocity); and (5) laboratory tests (white blood cells [WBCs], platelets [PLTs], hemoglobin

[HGB], neutrophils [Nes], lymphocytes [Lys], hematocrit [HCT], C-reactive protein [CRP], neutrophil-to-lymphocyte ratio [NLR], platelet-to-lymphocyte ratio [PLR], alanine aminotransferase [ALT], aspartate aminotransferase [AST], serum total protein [TP], serum albumin [ALB], blood urea nitrogen [BUN], serum creatinine [Scr], cystatin C [CysC], eGFR, uric acid [UA], triglyceride [TG], total cholesterol [TC], low-density lipoprotein [LDL], high-density lipoprotein [HDL], blood potassium [K],

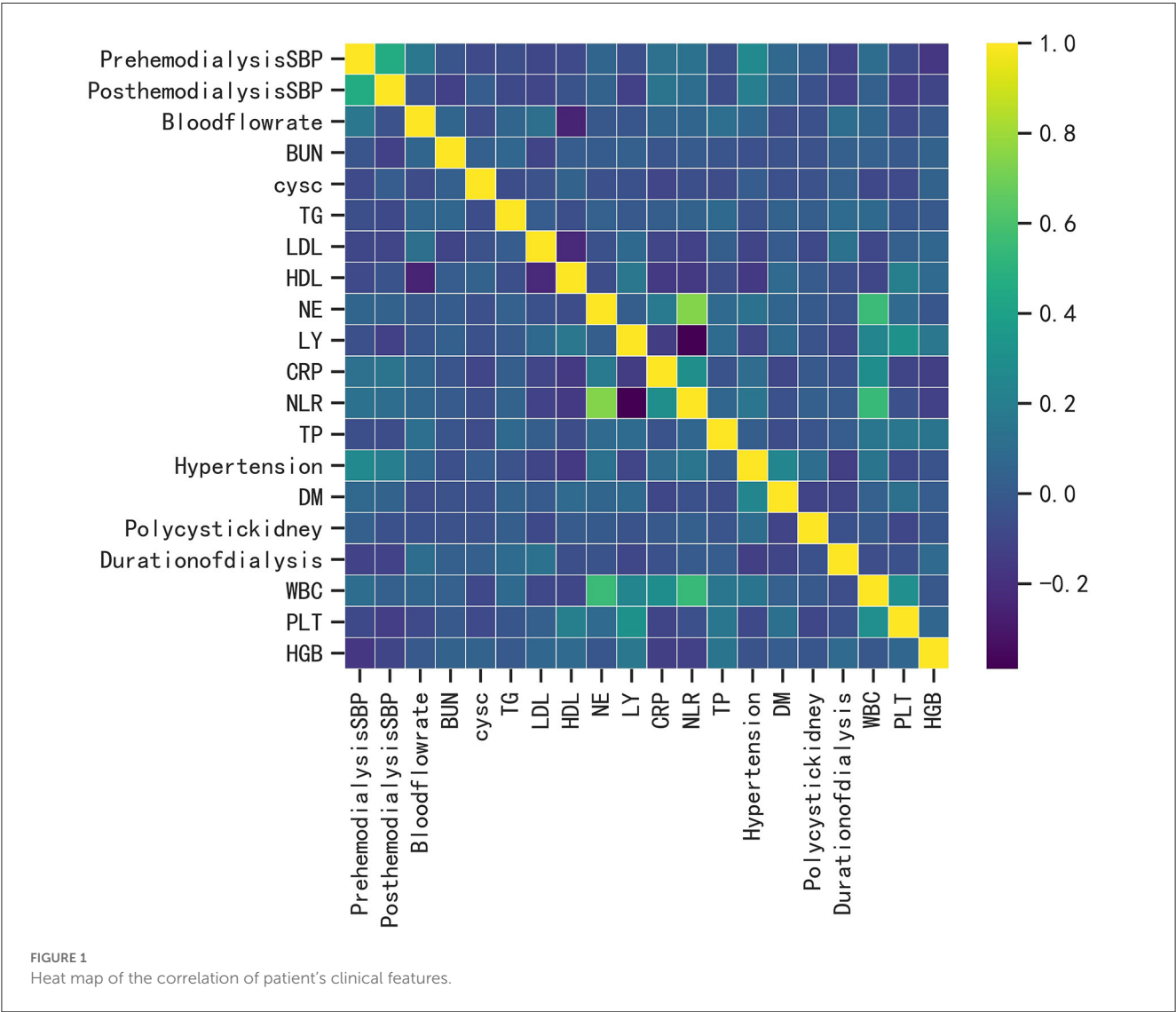


TABLE 2 Comparison of the predictive performance of five machine learning algorithms in the validation set.

Different algorithms	Accuracy (95%CI)	Precision (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)	F1-score (95%CI)
SVM	0.631 (0.546–0.715)	0.309 (0.272–0.345)	0.753 (0.603–0.902)	0.645 (0.485–0.806)	0.422 (0.377–0.466)
CNB	0.784 (0.741–0.827)	0.457 (0.399–0.514)	0.733 (0.652–0.815)	0.786 (0.716–0.857)	0.556 (0.502–0.610)
KNN	0.785 (0.766–0.803)	0.342 (0.248–0.437)	0.652 (0.599–0.706)	0.695 (0.657–0.733)	0.427 (0.336–0.518)
LR	0.835 (0.813–0.857)	0.539 (0.509–0.568)	0.861 (0.807–0.916)	0.853 (0.799–0.907)	0.659 (0.637–0.682)
XGboost	0.939 (0.926–0.952)	0.949 (0.910–0.988)	0.932 (0.913–0.951)	0.952 (0.930–0.973)	0.938 (0.921–0.956)

SVM, support vector machine; CNB, complement naive Bayes; KNN, k nearest neighbors; LR, logistic regression; XGBoost, extreme gradient boost.

blood sodium [Na], blood calcium [Ca], calcium-phosphorus product, and blood phosphorus [P]).

2.4. Selection of machine learning models

Before constructing ML models, the original clinical data were normalized. Normalization can improve the speed of gradient descent to find the optimal solution, and the algorithm for Euclidean distance can effectively improve the accuracy. In this study, the min-max normalization method was used to normalize the characteristic values of clinical data to the range of (0,1).

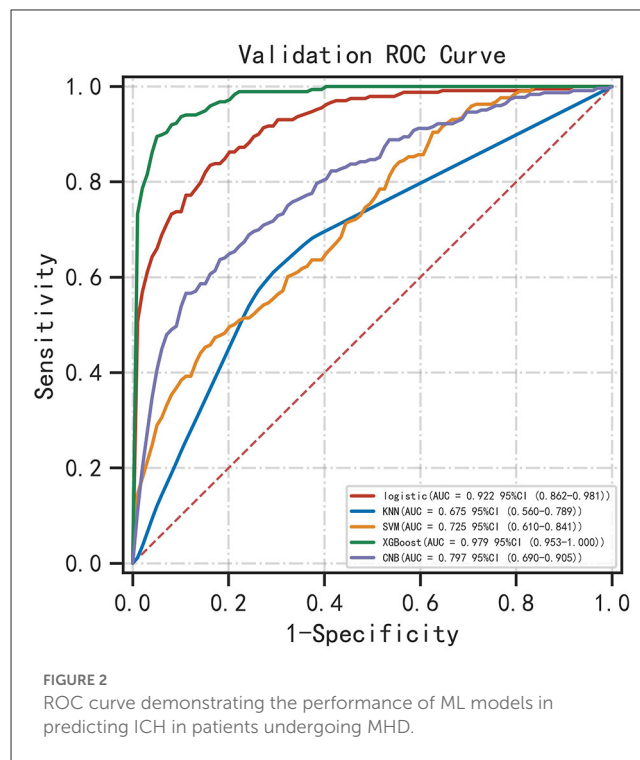
Approximately 70% of the samples in the dataset were randomly selected as the training set, whereas the remaining 30% of the samples were used as the validation set. The dataset is represented as $D = \{(x_i, y_i), i = 1, 2, \dots, N\}$, where x_i is $[x_{i1}, x_{i2}, x_{i3}, \dots, x_{ip}]$, which is a row vector with input variables (or features) of real value as its elements, and $y_i \in \{0, 1\}$ is a scalar with the output of an integer value as its element. The task in hand was a binary classification problem, that is, the generation of a model ($y = f[x]$) in the training set. The model was subsequently verified in the validation set to predict $\hat{y}_k = f(x_k)$. The predicted output \hat{y}_k should be similar to the actual output. All models were tested using Python.

We applied five ML algorithms to model the data: logistic regression (LR), support vector machine (SVM), K-nearest neighbor (KNN), complement Naive Bayes (CNB), and XGBoost. To be able to ensure that the training samples selected for multiple-model training were consistent, we generalized the performance of each model over multiple training sessions using a resampling training/validation mechanism. The XGBoost (version 1.2.1), lightGBM (version 3.2.1), and sklearn (version 0.22.1) packages were used for developing the ML models. For the RF algorithm, “ntree” was set to 100, and “mtree” was set to 3. To avoid overfitting and enhance interpretability, the maximum tree depth was set to eight nodes in the XGBoost algorithm. In addition, to evaluate the predictive accuracy of various ML models, accuracy, precision, sensitivity, specificity, F1 score, and the area under the receiver operating characteristic curve (ROC) were evaluated.

SHAP is a “model interpretation” package developed based on Python. To understand the results of the model output, the SHAP package was used to interpret and sort the features of the trained model and examine the contribution of each element in the features to the model.

2.5. Statistical analysis

The R software (version 4.02) was used for data processing and statistical analysis. Categorical variables were expressed in terms of quantity and percentage and were compared using Fisher's exact test or the chi-square test. For continuous variables, the Shapiro-Wilk test was initially used to determine whether the variables conformed to a normal distribution, and the independent sample *t*-test (conforming to a normal distribution) was subsequently used for comparing the data, which were expressed as mean \pm standard deviation. The Mann-Whitney *U*-test was used to



compare data with non-normal distribution, which were expressed as the median (first and third quartiles). A $P < 0.05$ was considered statistically significant.

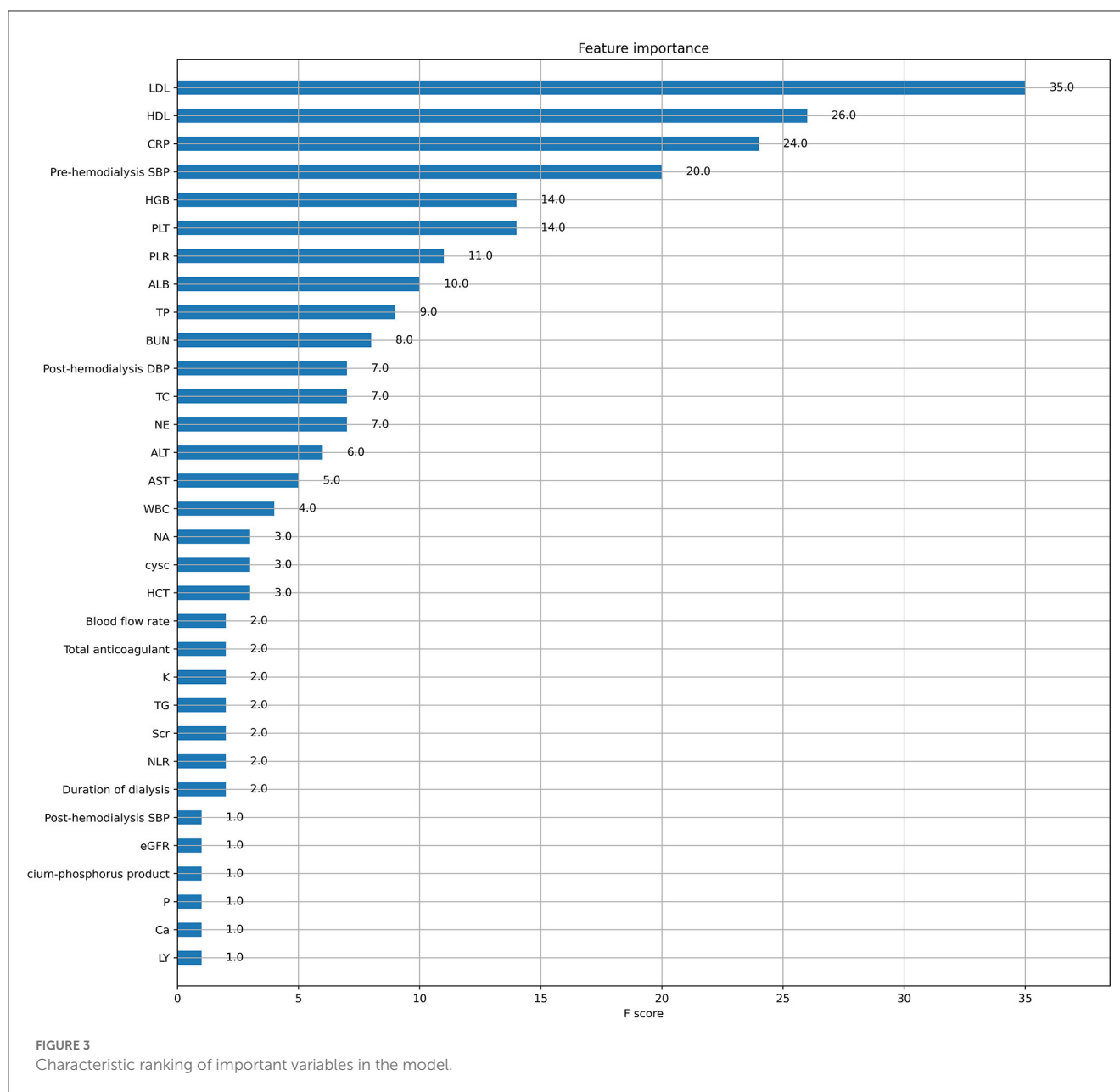
3. Results

3.1. Baseline patient characteristics

A total of 393 patients were included in this study, and the baseline characteristics of the ICH and non-ICH groups are shown in Table 1. In terms of demographic characteristics, no significant differences were observed in the sex and age of patients between the two groups. The history of diabetes and polycystic kidney disease was a significant variable in terms of underlying diseases. The blood flow rate and SBP before and after dialysis were important variables in terms of dialysis indicators. Laboratory indices, such as the levels of CRP, LDL, and HDL, were significantly different between the two groups. We further constructed a heat map demonstrating Spearman correlation coefficients to visualize the correlation between variables with differences (Figure 1).

3.2. Comparison of the predictive performance of all models

Five ML algorithms were used to construct predictive models. The training set was used to create and train the models. All ML models were tested in the test set, and their accuracy, precision, sensitivity, specificity, and F1 score were compared. The XGBoost model had the highest accuracy, precision, sensitivity, specificity, and F1 score (0.939, 0.949,



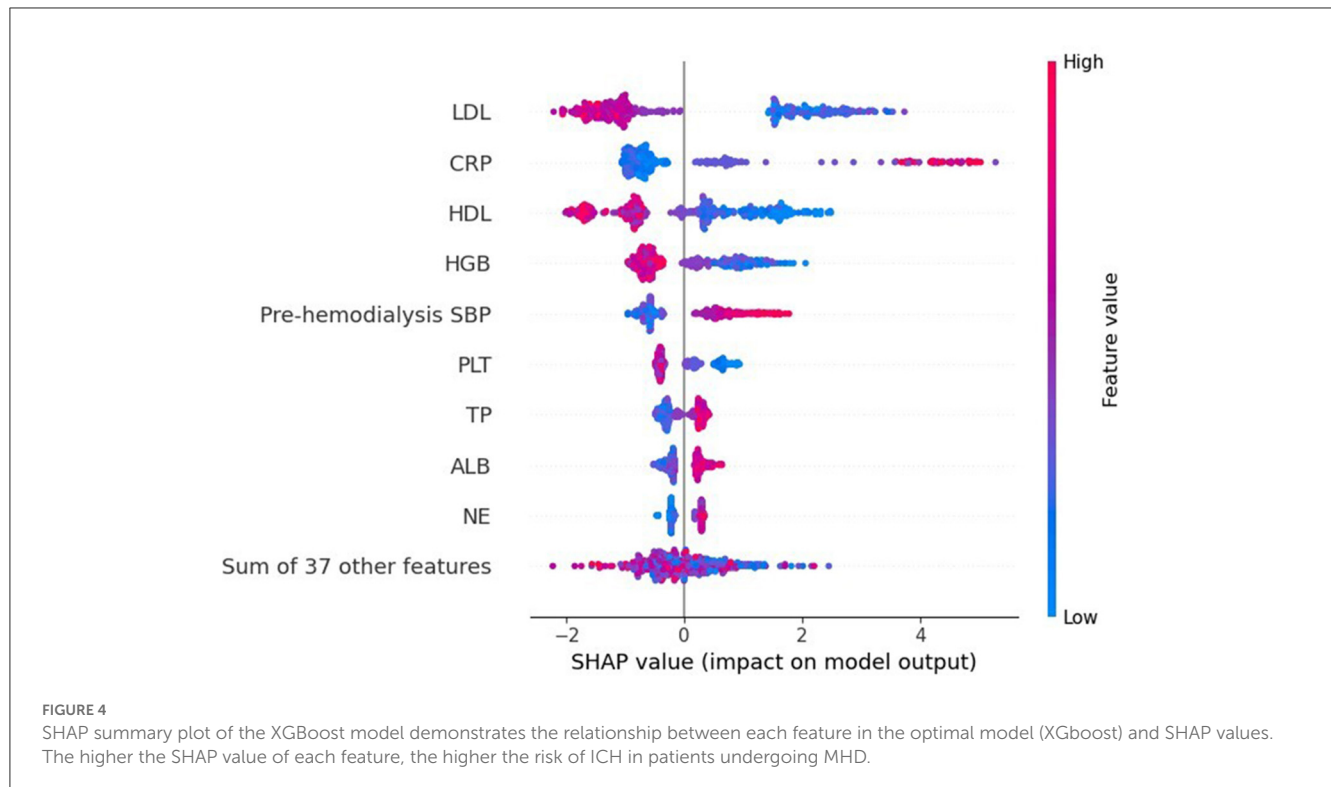
0.932, 0.952, and 0.938, respectively) (Table 2). Figure 2 shows a ROC curve demonstrating the predictive performance of all models. The XGBoost model (AUC = 0.979; 95% CI, 0.953–1.000) demonstrated optimal performance in the validation set. Therefore, the XGBoost model can be considered an ideal model for predicting the risk of ICH in patients undergoing MHD.

3.3. Explainable analysis of overall features

XGboost was used to rank the importance of features. Figure 3 shows the ranking of the most important variables in the model. The top five variables were LDL, HDL, CRP, pre-dialysis SBP, and HGB. The

interpretation of the impact of these features is roughly consistent with that reported in previous studies and clinician perception.

Figure 4 shows a characteristic density scatter plot, which demonstrates the effects of the main features in the dataset on the predictive performance of the model. The abscissa represents the SHAP value, which represents the contribution of a feature in the model to the overall output. SHAP values <0, equal to 0, and >0 represent negative, no, and positive contributions, respectively. The left ordinate represents the features sorted by importance. The color of the right ordinate, from blue to red, represents the feature values from low to high. Lower LDL levels, higher CRP levels, lower HDL levels, lower HGB levels, and higher pre-dialysis SBP have higher SHAP values, indicating a higher likelihood of developing ICH.



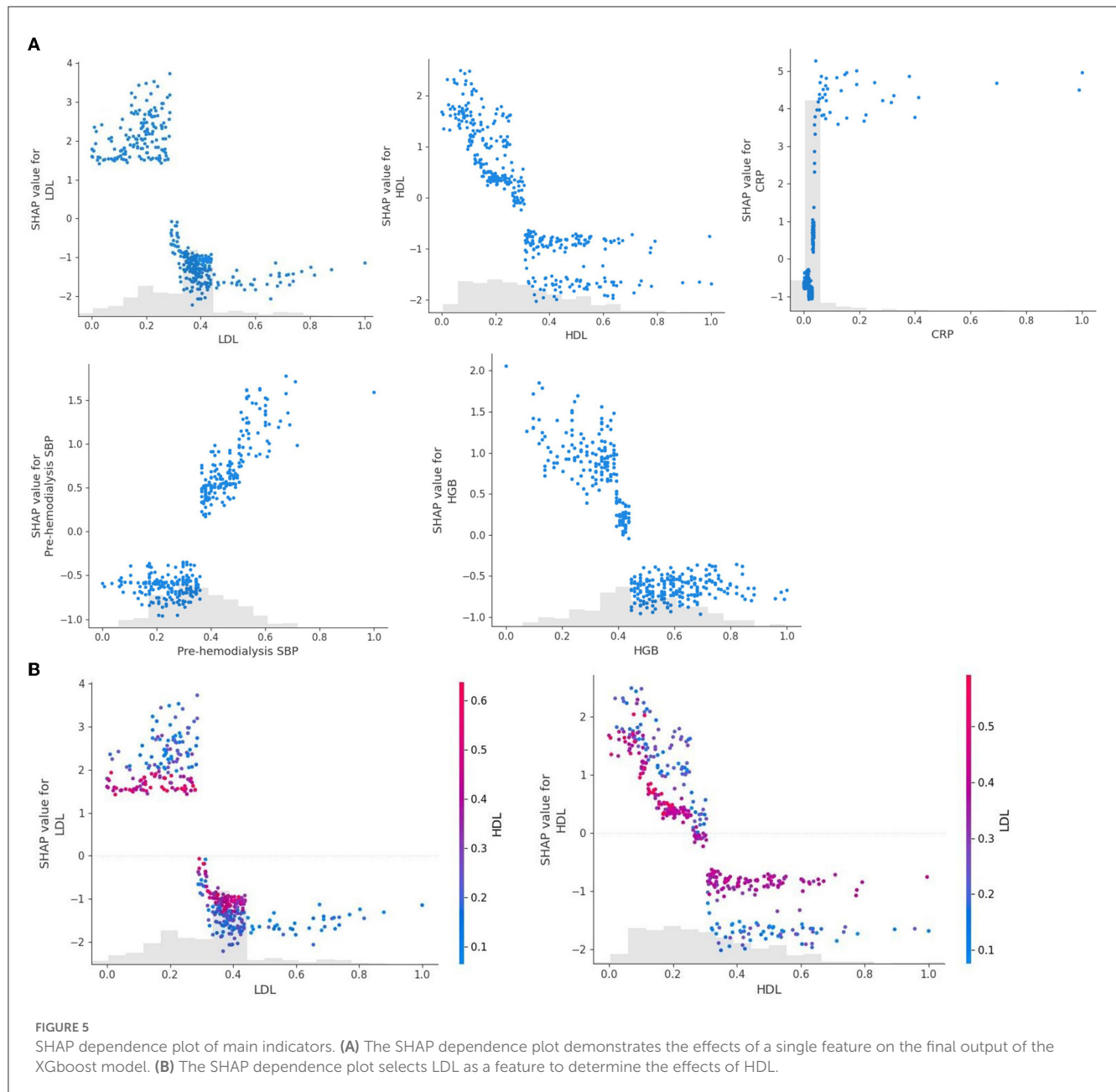
3.4. Explainable analysis of individual features

As shown in Figure 5, the SHAP dependence plot demonstrates the effects of a single feature on the final output of the XGboost model and can be used to select the most significant features of the model. CRP levels and pre-dialysis SBP were positively correlated with SHAP values, that is, the larger the values, the higher the risk of bleeding. However, the levels of LDL, HDL, and HGB were negatively correlated with SHAP values, indicating that the smaller the values, the higher the risk of bleeding (Figure 5A). We selected LDL as a feature to determine the effects of HDL. The red and blue dots represent high and low HDL levels, respectively. After the data were normalized, it was found that when LDL was less than the critical value of 0.3, regardless of HDL levels, the SHAP value of LDL was always greater than zero. In addition, when HDL was greater than the critical value, the SHAP value of HDL was always less than zero (Figure 5B). The cutoff level of LDL is 1.572 mmol/L in actual clinical practice. If this threshold is exceeded, the possibility of ICH decreases. However, if this threshold is not exceeded, the possibility of ICH increases. In addition, the values of all main features are distributed differently in different ranges and vary greatly in some regions. It remains unclear whether these conditions have some specific significance, which may have important implications for clinical outcomes. The feature dependence plot provides information within a given range, showing the trend of possible results. However, it is noteworthy that the plot suggests correlation and not causality. Therefore, it is necessary to integrate the results with clinical experience and specific conditions to determine whether they can be used to develop adjunctive intervention strategies.

In addition, SHAP can be used to analyze the influencing factors of a cerebral hemorrhage in each patient. Figure 6 shows the interpretation of the XGBoost model for the prediction of two cases. Specifically, the arrows show the effects of each factor on prediction. Features that increase the risk of developing ICH are shown in red, and those that reduce the risk are shown in blue. The stripe length of each feature indicates the importance of the feature when making predictions. The longer the stripes, the greater the contribution of the feature to the prediction. After combining the influence of all factors, the corresponding prediction score of each factor was calculated. Figure 6A demonstrates the contribution of different features to prediction in a patient correctly predicted to have ICH. CRP, LDL, and HGB had the largest contribution (red), indicating that they were the main causes of cerebral hemorrhage in the patient. The second patient was accurately predicted to have no ICH (Figure 6B), with LDL, CRP, and pre-dialysis SBP identified as protective factors. Although there were some risk factors, the patient had no cerebral hemorrhage.

4. Discussion

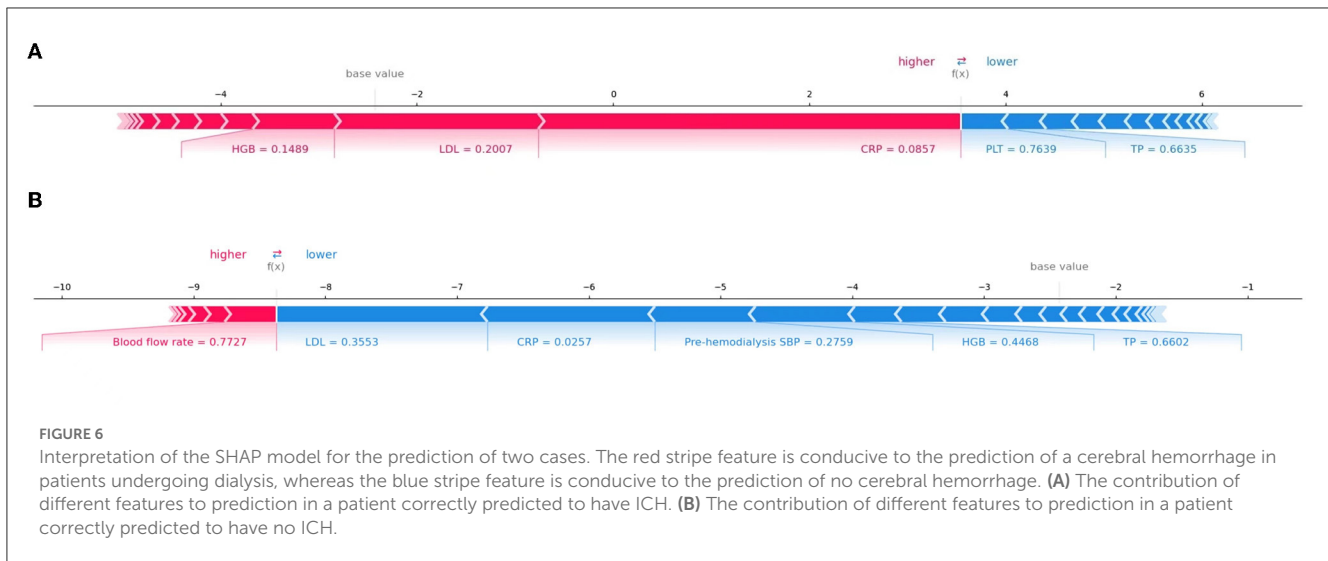
Intracerebral hemorrhage is characterized by a high rate of disability and death, which greatly increases the economic burden on families and society, so it is essential to investigate the factors influencing the complications of ICH events in MHD patients. Many scholars have identified the risk factors of ICH and hematoma expansion in patients undergoing MHD and screened variables, such as serum calcium, serum creatinine, and serum antiplatelet agents, *via* multivariate logistic regression (16, 17). Unlike many previous studies, the present study innovatively



used ML algorithms to screen for variables, and to the best of our knowledge, this is the first study to report the development of an ML-based predictive model to evaluate the probability of concurrent ICH events in patients undergoing MHD. In addition, we also applied four mainstream machine learning models, namely, LR, SVM, KNN, and CNB, to compare the predictive performance of the XGBoost algorithm with these machine learning methods.

XGBoost is a lifting algorithm based on tree models. Since its establishment in 2016, it has been used to deal with non-linear relationships and complex interactions between variables owing to its higher prediction accuracy and faster operation speed (14). The XGBoost algorithm has been widely used in the medical field, especially for the prediction of critical illnesses. Po-Yu Tseng et al. used the combination of RF and XGboost to predict the risk of acute kidney injury after cardiac

surgery, and the final AUC value was 0.843 (8). Pan et al. used XGBoost to predict the mortality of critically ill patients with COVID-19 admitted to the ICU. The AUC values of the training and validation sets were 0.86 and 0.92, respectively (18). The findings of the present study suggest that XGBoost can effectively improve the prediction of ICH in patients undergoing MHD. In this study, the predictors considered to be related to ICH in actual clinical practice and literature were included; patient information was collected as comprehensively as possible; abnormal indicators of various metabolic disorders were refined; ML algorithms were used to analyze variables; and finally, the ROC-AUC value of the optimal model (XGBoost) was as high as 0.979 (Figure 2), with the highest prediction accuracy and significantly better performance than other mainstream machine learning models.



In addition, in this study, we used SHAP to interpret the results of ML models. Emphasis is placed on features that have the greatest impact on outcome measures, thus helping clinicians to realize the rationale behind predicted outcomes early enough to initiate prompt intervention. The results showed that changes in LDL, HDL, CRP, SBP, and HGB levels were the main predictors of ICH in patients undergoing MHD, which was consistent with clinical studies.

Lipid is an indispensable neutral fat in the human body. To date, numerous studies have investigated the relationship between lipid metabolism and ICH. Lipid metabolism disorders in patients undergoing long-term hemodialysis are closely related to the occurrence of a cerebral hemorrhage (19, 20), which is consistent with the results of this study. The Genetic and Environmental Risk Factors for Hemorrhagic Stroke (GERFHS) reported a 33% reduction in the risk of a cerebral hemorrhage in patients with higher cholesterol levels, and a retrospective study (21) reported a significantly increased risk of hemorrhagic stroke in patients with lower HDL levels. The mechanism may be explored because lower LDL-C levels are closely associated with an increased number of cerebral microbleeds (CMBs) (22). Lobar CMBs are mainly associated with cerebral amyloid angiopathy (CAA) (23). The ϵ 4 allele variation of apolipoprotein E (APOE) is a known genetic risk factor for CAA. Genetic studies have shown a higher rate of reduction in LDL-C concentrations with the APOE ϵ 4 genotype vector (24). Recent studies have also shown that higher LDL-C genetic risk scores are associated with a higher prevalence of multiple lobar microbleeds (25). CMBs are independent risk factors for ICH and strong predictors of future cerebral hemorrhage (26). In addition, cholesterol is related to physiological processes such as vascular wall construction. Extremely low cholesterol levels may destroy the integrity of intracranial vascular endothelial cells, aggravate vascular endothelial damage, and increase the risk of cerebral hemorrhage (27). HDL is considered a protective factor for atherosclerosis (28), and low HDL levels can aggravate the progression of atherosclerosis, thus increasing the risk of a cerebral hemorrhage.

CRP is an important part of the immune system and one of the signs of acute inflammation (29). In this study, CRP

levels were significantly different between the ICH and non-ICH groups, and CRP was highly correlated with ICH, which is consistent with the findings of previous studies (30–32). Patients undergoing MHD often have comorbid inflammation, which may lead to endothelial damage and atherosclerosis (33, 34), thereby increasing the morbidity and mortality of cerebrovascular diseases (35). Genetic studies have shown that the significantly reduced expression of haplotype H5 in the CRP genotype is closely associated with hemorrhagic stroke (36). CRP induces endothelial dysfunction by directly destroying the blood–brain barrier (BBB) and induces monocytes to release proinflammatory cytokines, leading to increased vascular permeability and cerebral hemorrhage (37, 38).

According to the model results of this study, the SBP before daily hemodialysis in the cerebral hemorrhage group was higher than that in the control group, which is consistent with the conclusion that hypertension is a risk factor for cerebral hemorrhage in MHD patients as reported in previous studies. Hypertension is a known traditional risk factor for ICH (39). In patients with chronic kidney disease, renal function and excretion are impaired, blood volume is increased, renin–angiotensin–aldosterone system is activated in a feedback manner, and water and sodium retention is aggravated. In this study, the higher SBP before dialysis in patients with ICH may be related to inadequate dialysis. In addition, during hemodialysis, the greater hemodynamic changes and the excretion of antihypertensive drugs will aggravate hypertension, resulting in increased pressure on cerebral arteries. When the pressure on the vascular wall exceeds the pressure, the cerebral vessels rupture and bleed, causing cerebral hemorrhage.

Patients undergoing MHD are predisposed to anemia owing to factors such as reduced erythropoietin synthesis (40). HGB is the main indicator reflecting the anemic status of humans. Recent studies have reported that the HGB level of patients with MHD is negatively correlated with the risk of a cerebral hemorrhage (41–43), which is consistent with the results of this study. The underlying mechanisms may include vasoconstriction (44), platelet aggregation (45, 46), and cytotoxic reaction caused by chronic hypoxia (44), leading to brain dysfunction or damage.

This study has some limitations. First, although this study had a multicenter design, it only includes patients from three hospitals in Xuzhou, China. In future studies, we will include datasets from different regions and hospitals for external testing to improve the generalization ability of the model. Second, the number of patients with and without ICH was not well-balanced, which may have led to impaired prediction. Considering that deep learning has been widely used in the medical community in recent years, we will use deep learning models to incorporate a wider range of data in future studies. Overall, compared with traditional models, the prediction model developed in this study contains more information and has better predictive accuracy. In addition, the visualization of results based on SHAP can, to a great extent, alleviate the “black box” problem.

5. Conclusion

A predictive ML model was developed based on XGBoost, and SHAP was used to explain the clinical significance of each risk factor in predicting the occurrence of ICH in patients undergoing MHD. ICH events in patients undergoing MHD are associated with serum LDL, HDL, CRP, HGB, and pre-hemodialysis SBP levels. The combination of the XGBoost algorithm and SHAP can provide a clear explanation for risk prediction, which has great application value in future clinical research. This combination can help clinicians to implement early clinical interventions, provide more comprehensive information for the long-term management of patients undergoing MHD, and prevent and reduce the risk of ICH.

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 the Ethics Committee of the Affiliated Hospital of

Xuzhou Medical University. Written informed consent from the patients/participants or patients/participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

FL and AC conceptualized the study, outlined the study design, collected data, analyzed and interpreted results, and wrote the manuscript. ZL and QP preprocessed input data, built machine learning models, analyzed data, and wrote the manuscript. LG collected data and preprocessed input data. YF, JF, and PW helped to adjust the ideas of the manuscript, suggested changes, and revised the manuscript. All authors agreed to take responsibility for their contributions and read and approved the final manuscript.

Funding

The research was sponsored by the National Natural Science Foundation (General Program) Grant No. 61972211, China and the National Key Research and Development Project Grant No. 2020YFB1804700, China.

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Himmelfarb J, Vanholder R, Mehrotra R, Tonelli M. The current and future landscape of dialysis. *Nat Rev Nephrol.* (2020) 16:573–85. doi: 10.1038/s41581-020-0315-4
- de Oliveira Manoel AL. Surgery for spontaneous intracerebral hemorrhage. *Crit Care.* (2020) 24:45. doi: 10.1186/s13054-020-2749-2
- Wang HH, Hung SY, Sung JM, Hung KY, Wang JD. Risk of stroke in long-term dialysis patients compared with the general population. *Am J Kidney Dis.* (2014) 63:604–11. doi: 10.1053/j.ajkd.2013.10.013
- Wyld M, Webster AC. Chronic kidney disease is a risk factor for stroke. *J Stroke Cerebrovasc Dis.* (2021) 30:105730. doi: 10.1016/j.jstrokecerebrovasdis.2021.105730
- Magid-Bernstein J, Girard R, Polster S, Srinath A, Romanos S, Awad IA, et al. Cerebral hemorrhage: pathophysiology, treatment, and future directions. *Circ Res.* (2022) 130:1204–29. doi: 10.1161/CIRCRESAHA.121.319949
- Van Calster B, Wynants L. Machine learning in medicine. *N Engl J Med.* (2019) 380:2588. doi: 10.1056/NEJMc1906060
- Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of medicine. *J Intern Med.* (2018) 284:603–19. doi: 10.1111/joim.12822
- Tseng PY, Chen YT, Wang CH, Chiu KM, Peng YS, Hsu SP, et al. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Crit Care.* (2020) 24:478. doi: 10.1186/s13054-020-03179-9
- Li W, Dong Y, Liu W, Tang Z, Sun C, Lowe S, et al. A deep belief network-based clinical decision system for patients with osteosarcoma. *Front Immunol.* (2022) 13:1003347. doi: 10.3389/fimmu.2022.1003347
- Peng J, Zou K, Zhou M, Teng Y, Zhu X, Zhang F, et al. An Explainable artificial intelligence framework for the deterioration risk prediction of hepatitis patients. *J Med Syst.* (2021) 45:61. doi: 10.1007/s10916-021-01736-5

11. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA*. (2017) 318:517–8. doi: 10.1001/jama.2017.7797
12. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng*. (2018) 2:749–60. doi: 10.1038/s41551-018-0304-0
13. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst*. (2017) 30.
14. Chen T, Guestrin C. “Xgboost: A scalable tree boosting system” in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. (2016) p. 785–794. doi: 10.1145/2939672.2939785
15. National Kidney Foundation. KDOQI clinical practice guideline for hemodialysis adequacy: 2015 update. *Am J Kidney Dis*. (2015) 66:884–930. doi: 10.1053/j.ajkd.2015.07.015
16. Ozelsancak R, Micozkadioglu H, Torun D, Tekkarismaz N. Cerebrovascular events in hemodialysis patients; a retrospective observational study. *BMC Nephrol*. (2019) 20:466. doi: 10.1186/s12882-019-1629-y
17. Kitamura M, Tateishi Y, Sato S, Kitamura S, Ota Y, Muta K, et al. Association between serum calcium levels and prognosis, hematoma volume, and onset of cerebral hemorrhage in patients undergoing hemodialysis. *BMC Nephrol*. (2019) 20:210. doi: 10.1186/s12882-019-1400-4
18. Pan P, Li Y, Xiao Y, Han B, Su L, Su M, et al. Prognostic assessment of COVID-19 in the intensive care unit by machine learning methods: model development and validation. *J Med Internet Res*. (2020) 22:e23128. doi: 10.2196/23128
19. Ferro CJ, Mark PB, Kanbay M, Sarafidis P, Heine GH, Rossignol P, et al. Lipid management in patients with chronic kidney disease. *Nat Rev Nephrol*. (2018) 14:727–49. doi: 10.1038/s41581-018-0072-9
20. Amarenco P, Bogousslavsky J, Callahan A. 3rd, Goldstein LB, Hennerici M, Rudolph AE, et al. High-dose atorvastatin after stroke or transient ischemic attack. *N Engl J Med*. (2006) 355:549–59. doi: 10.1016/j.jvs.2006.10.008
21. Shen Y, Shi L, Nauman E, Katzmazyk PT, Price-Haywood EG, Bazzano AN, et al. Inverse association between HDL (High-Density Lipoprotein) cholesterol and stroke risk among patients with type 2 diabetes mellitus. *Stroke*. (2019) 50:291–7. doi: 10.1161/STROKEAHA.118.023682
22. Ma C, Gurol ME, Huang Z, Lichtenstein AH, Wang X, Wang Y, et al. Low-density lipoprotein cholesterol and risk of intracerebral hemorrhage: a prospective study. *Neurology*. (2019) 93:e445–57. doi: 10.1212/WNL.0000000000007853
23. Schrag M, Kirshner H. Management of intracerebral hemorrhage: JACC focus seminar. *J Am Coll Cardiol*. (2020) 75:1819–31. doi: 10.1016/j.jacc.2019.10.066
24. Phuath CL, Raffeld MR, Ayres AM, Gurol ME, Viswanathan A, Greenberg SM, et al. APOE polymorphisms influence longitudinal lipid trends preceding intracerebral hemorrhage. *Neurol Genet*. (2016) 2:e81. doi: 10.1212/NXG.0000000000000081
25. Akoudad S, Ikram MA, Portegies ML, Adams HH, Bos D, Hofman A, et al. Genetic loci for serum lipid fractions and intracerebral hemorrhage. *Atherosclerosis*. (2016) 246:287–92. doi: 10.1016/j.atherosclerosis.2016.01.024
26. Charidimou A, Shams S, Romero JR, Ding J, Veltkamp R, Horstmann S, et al. Clinical significance of cerebral microbleeds on MRI: a comprehensive meta-analysis of risk of intracerebral hemorrhage, ischemic stroke, mortality, and dementia in cohort studies (v1). *Int J Stroke*. (2018) 13:454–68. doi: 10.1177/1747493017751931
27. Lyu J, Yang EJ, Shim JS. Cholesterol trafficking: an emerging therapeutic target for angiogenesis and cancer. *Cells*. (2019) 8:389. doi: 10.3390/cells8050389
28. Ouimet M, Barrett TJ, Fisher EA. HDL and reverse cholesterol transport. *Circ Res*. (2019) 124:1505–18. doi: 10.1161/CIRCRESAHA.119.312617
29. Pathak A, Agrawal A. Evolution of C-reactive protein. *Front Immunol*. (2019) 10:943. doi: 10.3389/fimmu.2019.00943
30. Xue Y, Zhang L, Fan Y, Li Q, Jiang Y, Shen C. C-reactive protein gene contributes to the genetic susceptibility of hemorrhagic stroke in men: a case-control study in chinese han population. *J Mol Neurosci*. (2017) 62:395–401. doi: 10.1007/s12031-017-0945-6
31. Löppönen P, Qian C, Tetri S, Juvela S, Huhtakangas J, Bode MK, et al. Predictive value of C-reactive protein for the outcome after primary intracerebral hemorrhage. *J Neurosurg*. (2014) 121:1374–9. doi: 10.3171/2014.7.JNS132678
32. Bader ER, Pana TA, Barlas RS, Metcalf AK, Potter JF, Myint PK. Elevated inflammatory biomarkers and poor outcomes in intracerebral hemorrhage. *J Neurol*. (2022) 269:6330–41. doi: 10.1007/s00415-022-11284-8
33. Sá Martins V, Aguiar L, Dias C, Lourenço P, Pinheiro T, Velez B, et al. Predictors of nutritional and inflammation risk in hemodialysis patients. *Clin Nutr*. (2020) 39:1878–84. doi: 10.1016/j.clnu.2019.07.029
34. Maraj M, Kuśnierz-Cabala B, Dumnicka P, Gala-Błazińska A, Gawlik K, Pawlica-Gosiewska D, et al. Malnutrition, inflammation, atherosclerosis syndrome (MIA) and diet recommendations among end-stage renal disease patients treated with maintenance hemodialysis. *Nutrients*. (2018) 10:69. doi: 10.3390/nu10010069
35. Bihl JC, Zhang C, Zhao Y, Xiao X, Ma X, Chen Y, et al. Angiotensin-(1-7) counteracts the effects of Ang II on vascular smooth muscle cells, vascular remodeling and hemorrhagic stroke: role of the NF- κ B inflammatory pathway. *Vascul Pharmacol*. (2015) 73:115–23. doi: 10.1016/j.vph.2015.08.007
36. Wang Q, Ding H, Tang JR, Zhang L, Xu YJ, Yan JT, et al. C-reactive protein polymorphisms and genetic susceptibility to ischemic stroke and hemorrhagic stroke in the Chinese Han population. *Acta Pharmacol Sin*. (2009) 30:291–8. doi: 10.1038/aps.2009.14
37. Sproston NR, Ashworth JJ. Role of C-reactive protein at sites of inflammation and infection. *Front Immunol*. (2018) 9:754. doi: 10.3389/fimmu.2018.00754
38. Di Napoli M, Slevin M, Poppa-Wagner A, Singh P, Lattanzi S, Divani AA. Monomeric C-Reactive protein and cerebral hemorrhage: from bench to bedside. *Front Immunol*. (2018) 9:1921. doi: 10.3389/fimmu.2018.01921
39. Chaudhary N, Pandey AS, Wang X, Xi G. Hemorrhagic stroke-Pathomechanisms of injury and therapeutic options. *CNS Neurosci Ther*. (2019) 25:1073–4. doi: 10.1111/cns.13225
40. Babitt JL, Lin HY. Mechanisms of anemia in CKD. *J Am Soc Nephrol*. (2012) 23:1631–4. doi: 10.1681/ASN.2011111078
41. Diedler J, Sykora M, Hahn P, Heerlein K, Schölzke MN, Kellert L, et al. Low hemoglobin is associated with poor functional outcome after non-traumatic, supratentorial intracerebral hemorrhage. *Crit Care*. (2010) 14:R63. doi: 10.1186/cc8961
42. Kuragano T, Matsumura O, Matsuda A, Hara T, Kiyomoto H, Murata T, et al. Association between hemoglobin variability, serum ferritin levels, and adverse events/mortality in maintenance hemodialysis patients. *Kidney Int*. (2014) 86:845–54. doi: 10.1038/ki.2014.114
43. Milonitis H, Papavasileiou V, Eskandari A, D’Ambrogio-Remillard S, Ntaios G, Michel P. Anemia on admission predicts short- and long-term outcomes in patients with acute ischemic stroke. *Int J Stroke*. (2015) 10:224–30. doi: 10.1111/ijss.12397
44. Zhang S, Pan X, Wei C, Wang L, Cheng Y, Hu Z, et al. Associations of anemia with outcomes in patients with spontaneous intracerebral hemorrhage: a meta-analysis. *Front Neurol*. (2019) 10:406. doi: 10.3389/fneur.2019.00406
45. Daurman HL, Lessard D, Yarzebski J, Gore JM, Goldberg RJ. Bleeding complications in patients with anemia and acute myocardial infarction. *Am J Cardiol*. (2005) 96:1379–83. doi: 10.1016/j.amjcard.2005.06.088
46. Lisan T, Caldwell SH, Burroughs AK, Northup PG, Senzolo M, Stravitz RT, et al. Hemostasis and thrombosis in patients with liver disease: the ups and downs. *J Hepatol*. (2010) 53:362–71. doi: 10.1016/j.jhep.2010.01.042



OPEN ACCESS

EDITED BY

Jean-Claude Baron,
University of Cambridge, United Kingdom

REVIEWED BY

Yang Song,
Óbuda University, Hungary
Zhigang Liang,
Yantai Yuhuangding Hospital, China

*CORRESPONDENCE

Bingying Du
✉ 15800614142@163.com
Xiaoying Bi
✉ bixiaoying2013@163.com

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

RECEIVED 25 January 2023

ACCEPTED 25 May 2023

PUBLISHED 09 June 2023

CITATION

Xu Y, Sun X, Liu Y, Huang Y, Liang M, Sun R, Yin G, Song C, Ding Q, Du B and Bi X (2023) Prediction of subjective cognitive decline after corpus callosum infarction by an interpretable machine learning-derived early warning strategy. *Front. Neurol.* 14:1123607. doi: 10.3389/fneur.2023.1123607

COPYRIGHT

© 2023 Xu, Sun, Liu, Huang, Liang, Sun, Yin, Song, Ding, Du and Bi. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Prediction of subjective cognitive decline after corpus callosum infarction by an interpretable machine learning-derived early warning strategy

Yawen Xu[†], Xu Sun[†], Yanqun Liu[†], Yuxin Huang, Meng Liang, Rui Sun, Ge Yin, Chenrui Song, Qichao Ding, Bingying Du* and Xiaoying Bi*

Department of Neurology, Changhai Hospital, Second Military Medical University, Shanghai, China

Background and purpose: Corpus callosum (CC) infarction is an extremely rare subtype of cerebral ischemic stroke, however, the symptoms of cognitive impairment often fail to attract early attention of patients, which seriously affects the long-term prognosis, such as high mortality, personality changes, mood disorders, psychotic reactions, financial burden and so on. This study seeks to develop and validate models for early predicting the risk of subjective cognitive decline (SCD) after CC infarction by machine learning (ML) algorithms.

Methods: This is a prospective study that enrolled 213 (only 3.7%) CC infarction patients from a nine-year cohort comprising 8,555 patients with acute ischemic stroke. Telephone follow-up surveys were carried out for the patients with definite diagnosis of CC infarction one-year after disease onset, and SCD was identified by Behavioral Risk Factor Surveillance System (BRFSS) questionnaire. Based on the significant features selected by the least absolute shrinkage and selection operator (LASSO), seven ML models including Extreme Gradient Boosting (XGBoost), Logistic Regression (LR), Light Gradient Boosting Machine (LightGBM), Adaptive Boosting (AdaBoost), Gaussian Naïve Bayes (GNB), Complement Naïve Bayes (CNB), and Support vector machine (SVM) were established and their predictive performances were compared by different metrics. Importantly, the SHapley Additive exPlanations (SHAP) was also utilized to examine internal behavior of the highest-performance ML classifier.

Results: The Logistic Regression (LR)-model performed better than other six ML-models in SCD predictability after the CC infarction, with the area under the receiver characteristic operator curve (AUC) of 77.1% in the validation set. Using LASSO and SHAP analysis, we found that infarction subregions of CC infarction, female, 3-month modified Rankin Scale (mRS) score, age, homocysteine, location of angiostenosis, neutrophil to lymphocyte ratio, pure CC infarction, and number of angiostenosis were the top-nine significant predictors in the order of importance for the output of LR-model. Meanwhile, we identified that infarction subregion of CC, female, 3-month mRS score and pure CC infarction were the factors which independently associated with the cognitive outcome.

Conclusion: Our study firstly demonstrated that the LR-model with 9 common variables has the best-performance to predict the risk of post-stroke SCD due to CC infarction. Particularly, the combination of LR-model and SHAP-explainer could aid in achieving personalized risk prediction and be served as a decision-making tool for early intervention since its poor long-term outcome.

KEYWORDS

corpus callosum infarction, cognitive impairment, machine learning, subjective cognitive decline, Shapley additive explanations

1. Introduction

The corpus callosum (CC) is the largest commissural bridge of white-matter fibers between bilateral hemispheres (1), accompanied by a unique anterior and posterior double circulation system and abundant collateral arteries (2). Because of the sufficient blood supply, CC infarction is extremely rare and accounts for barely 2.3–8.0% of cerebral ischemic stroke (3–5). Because of its unique physiological structure and function, the manifestations of CC infarction are variable and lacking of specificity. Due to these special and complex characteristics, misdiagnosis and delayed treatment are not uncommon for CC infarction (6). Interestingly, we previously found that, compared to general basal ganglia infarction, patients with CC infarction had lower National Institutes of Health Stroke Scale scores and better recovery at the time of discharge, while the one-year mortality is higher with poorer long-term prognosis (5). Cognitive impairment is one of the main causes of poor long-term prognosis in patients with CC cerebral infarction. Unfortunately, due to its occult exacerbation process, patients often do not pay enough attention to it in the early stage, and miss the optimal intervention period, resulting in irreversible cognitive impairment.

Subjective cognitive decline (SCD) is an individual's self-report of cognitive decline and is nowadays thought to be a precursor to various common cognitive disorders in clinic, such as mild cognitive impairment (MCI) (7) and Alzheimer's disease (AD) (8). Recent researches have revealed that compared with age-matched healthy controls, patients with SCD suffer a 4.5–6 times higher risk of developing into MCI or AD (9, 10). Compared to universally-known post-stroke cognitive impairment (PSCI), SCD places more emphasis on the patient's subjective perceptions and timely feedback from caregivers, making it easier to identify and intervene early. Meanwhile, our, as well as others' previous studies have proved that, white matter lesions (WMLs) are important pathological mechanisms for cognitive dysfunctions (5, 11–14). As an extremely rare subtype of stroke with prominent WMLs, CC infarction is likely to become a potential driver of SCD and other symptomatic cognitive decline. Therefore, aiming to restore brain health and cognitive abilities as long as possible, this at-risk group is recognized as an eligible target population for early intervention strategies (15, 16).

The role of physicians has always been to synthesize the data available to them to identify prognosis patterns that guide early intervention. Machine learning (ML) is a new rising technical foundation of artificial intelligence, which enables the computer to learn the rules hidden in the data automatically (17). Several studies have revealed that ML-based models are promising in predicting the diagnosis, prognosis or recurrence of ischemic stroke, what's more, those models are also widely used in the field of psychology, biomechanics and so on (18–23). Nevertheless, it still lacks of ML-based evidence on SCD prediction after cerebral infarction. What's more, the “black-box” character of ML-technique hinders

clinicians to have a good understand of the predictive decision, namely failure in accountability (24). To this end, we proposed an interpretable strategy combining ML algorithm with SHapley Additive exPlanations (SHAP) to provide consistent and locally accurate attribution values for each feature within each prediction model. It's calculated by comparing the predicting discrepancy in all possible combinations containing and withholding each feature and provide a unique report individually (25).

Here, with the largest sample of CC infarction to date, this is an exploratory study that for the first time emphasizes the clinical feasibility to individually predict the occurrence of one-year SCD after CC infarction by using ML methods. We also attempt to apply SHAP-value for explaining the importance and influence of each predictor contributing to the optimal model's outcome. We expect this ML-derived early warning system and SHAP-based framework of interpretation could help clinicians to better counsel patients, conduct targeted follow-up and determine personalized interventional measures.

2. Methods

2.1. Participants

The design of this study is presented in Figure 1. A total of 8,555 ischemic stroke patients were collected from Shanghai Changhai Hospital between July 2012 and June 2021. Among them, 314 (3.7%) patients with acute CC infarction were enrolled. The exclusion criteria were as follows: (i) age under 30 or above 80 years old, (ii) cognitive impairment precedes CC infarction, (iii) follow-up period was less than 1 year, or loss to follow-up, (iv) serious medical complications, (v) incomplete neuroimaging materials, (vi) acceptance of thrombolytic therapy or interventional therapy, and (vii) failure to sign written informed consent. Ultimately, 213 patients with acute CC infarction were included for final analysis. This study was approved by the Changhai Hospital Ethics Committee (NO. CHEC2021-1021).

2.2. Clinical and imaging assessment

Basic clinical and imaging information of enrolled patients were obtained from Electronic Medical Record (EMR) management system. A list of these variables was shown in Supplementary Table S1, including: demographic characteristics (age, sex, body mass index[BMI]), vascular risk factors (hypertension, diabetes mellitus, prior stroke or transient ischemic attack, heart diseases, smoking, alcoholism), stroke severity on admission (time from onset to hospital, NIH stroke scale [NIHSS] scores), laboratory tests (alanine transaminase [ALT], low-density lipoprotein [LDL], high-density lipoprotein [HDL], cholestenone, triglyceride, creatine, urea, uric acid,

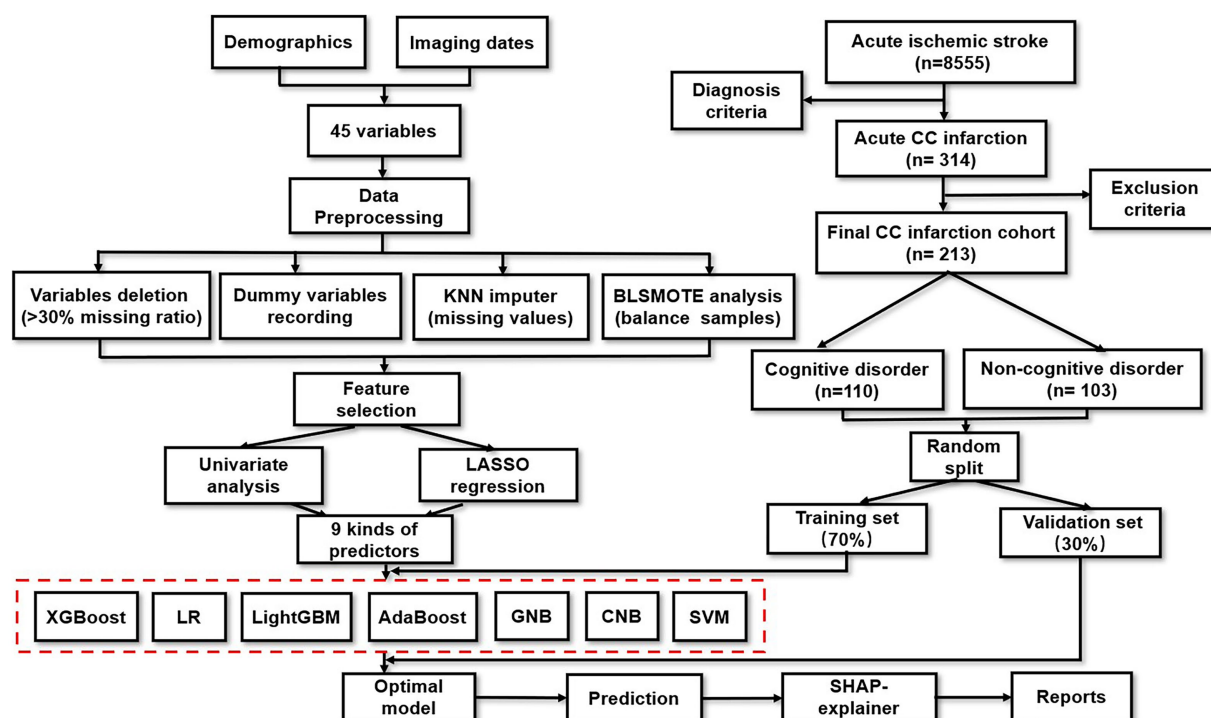


FIGURE 1
Schematic representation of the study design and modeling workflow.

Glucose [Glu]), (thyroid-stimulating hormone [TSH], triiodothyronine [T3], thyroxine [T4], erythrocyte, leukocyte, neutrophil to lymphocyte ratio [NLR], hemoglobin, thrombocyte, erythrocyte sedimentation rate [ESR], C-reactive protein [CRP], homocysteine [Hcy], glycosylated hemoglobin [HbA1c], fibrinogen [FIB], D-dimer), imaging examination assessment (pure CC infarction, infarction subregion of CC, other infarction areas, location of angiostenosis, number of angiostenosis, extracranial carotid plaque, TOAST subtype (26)), functional status (Modified Rankin scale [mRS] at 3-month), secondary prevention and recurrence (rehabilitation treatment, regular secondary prevention and recurrent stroke).

In detail, rehabilitation here referred to a series of standardized rehabilitation therapy obtained in rehabilitation hospitals, which mainly focuses on the motor and language function. Moreover, it also included lifestyle modification and taking medication exactly as prescribed at Discharge Notes, as well as additional carotid surgery or stenting, repairment for closure of patent foramen ovale, and surgery for intracranial or vertebral stenosis if necessary (27).

Additionally, the corresponding neuroimaging evidences were collected from both (i) MRI (Magnetom Impact 3.0 T, Siemens, Berlin, Germany), including T1-imaging, T2-imaging and diffusion-weighted imaging (DWI), and (ii) MR-angiography (MAGNETOM Skyra 3.0 T, Siemens) or CT-angiography (Aquilion One, Toshiba, Tokyo, Japan). As shown in Figure 2, the patients could be divided into 2 groups according to DWI patterns: pure callosal infarcts and complex callosal infarcts. The former was further subdivided into following subgroups: (i) Pure genu infarction of the corpus callosum, (ii) Pure body infarction of the corpus callosum, and (iii) Pure

splenium infarction of the corpus callosum according to the subregions of CC.

2.3. Cognitive dysfunction definition

Telephone follow-up surveys were carried out for the patients with definite diagnosis of CC infarction one-year after onset. According to the cognitive decline module of the Behavioral Risk Factor Surveillance System (BRFSS), which is the largest ongoing health survey system in the world (28). SCD was identified by the question of BRFSS, “During the past 12 months, have you experienced confusion or memory loss that is happening more often or is getting worse?” (29–31). If respondents had a clear cognitive complaint compared with the self-perception before stroke, they were classified as suffering from post-stroke SCD, otherwise they were distinguished as non-SCD. Additionally, there were five detailed questions of aggravating confusion or memory decline mentioned in the BRFSS questionnaire, including: (1) the frequency of giving up daily household activities or common chores, (2) the frequency of requirement of assistance with these daily activities, (3) the frequency of getting help, just as you wanted, (4) the frequency of work, volunteer, or social activities disturbed by the confusion or memory disorder, and (5) whether having sought medical attention for this (29, 31). These SCD-related outcomes evaluated by a five-point scale (Always, usually, sometimes, rarely, never) were dichotomized to determine if these outcomes were challenge (assigned as 1) vs. if they rarely or never happened (assigned as 0) (28). Consequently, the patients would better realize whether they had problems with

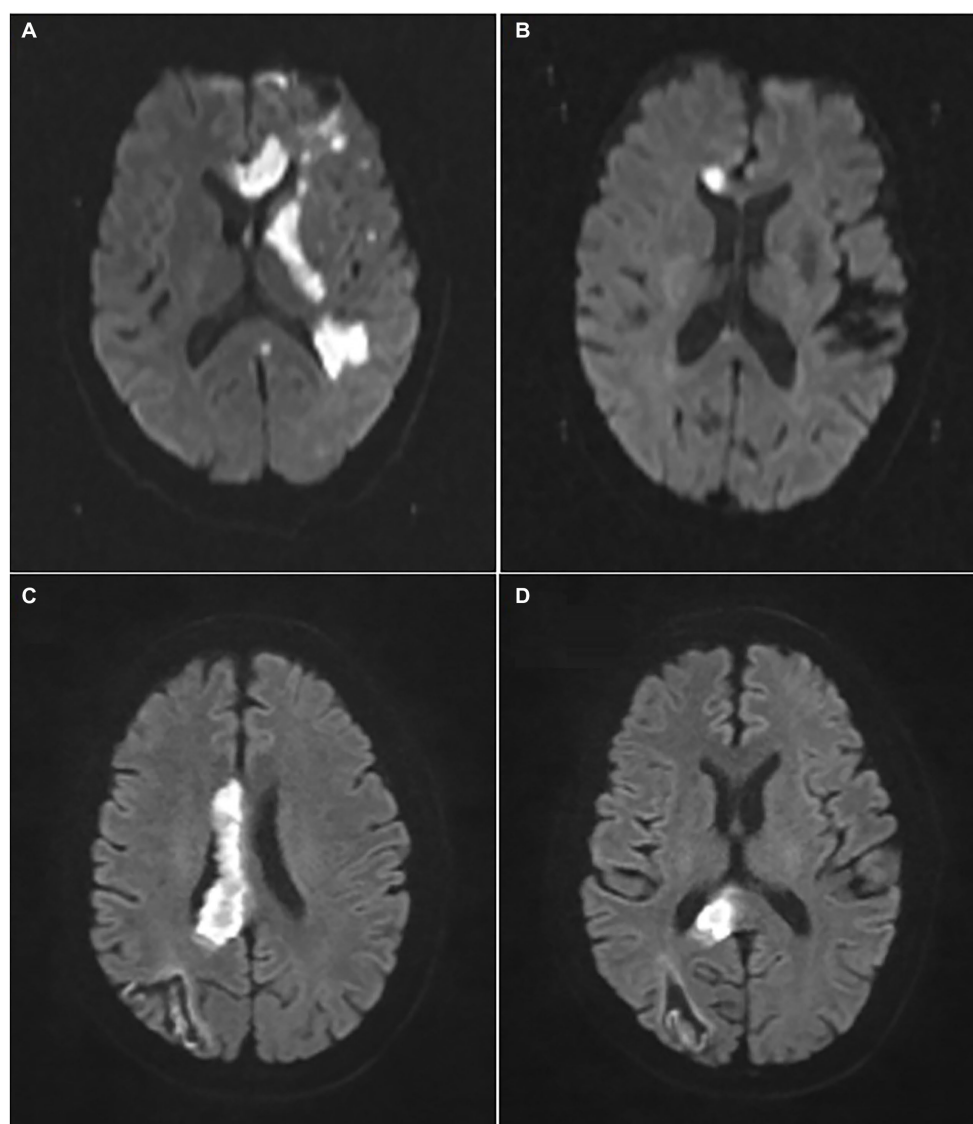


FIGURE 2
Representative images of pure and complex callosal infarction. **(A)** Complex callosal infarction. **(B)** Pure genu infarction of the corpus callosum. **(C)** Pure body infarction of the corpus callosum. **(D)** Pure splenium infarction of the corpus callosum.

post-stroke SCD and SCD related functional impairment through our telephone survey.

2.4. Machine learning

2.4.1. Features selection

Least absolute shrinkage and selection operator (LASSO) was used to select variables among high-dimensional data based on the penalty method. The originally small coefficients were compressed to 0 after compressing (32). Thereafter, regarded as non-significant variables, the corresponding variables of these coefficients were directly discarded (33). LASSO regression is also usually characterized by variable selection and complexity adjustment for construction of ML models while avoiding overfitting. However, the most ML methods could not process data with missing values, so we imputed the dataset by KNN before LASSO regression. In our study, this binary

logistic regression (LASSO) model is helpful to screen out significant predictors of SCD after acute CC infarction.

2.4.2. Machine learning models

Then, the dataset was randomly divided into training set and validation set. As in most cases, the training set accounted for 70% and the validation set accounted for 30% (34). Seven comprehensive and up-to-date ML algorithms were thereafter used to develop the predictive models, including Extreme Gradient Boosting (XGBoost), Logistic Regression (LR), Light Gradient Boosting Machine (LightGBM), Adaptive Boosting (AdaBoost), Gaussian Naïve Bayes (GNB), Complement Naïve Bayes (CNB), and Support vector machine (SVM). For each ML-based model, five-fold cross-validation was performed to evaluate the generalization ability (35), and the optimal hyperparameters were selected subsequently. Additionally, the following indicators are calculated to comprehensively evaluate the performance of different models: area under the curve (AUC)-value,

accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F1 scores.

2.4.3. Personalized interpretation

Specifically, we further utilized a novel approach to explain the output of the highest-performing ML model, namely Shapley Additive explanation (SHAP), rooted in Shapley value. Calculated the marginal contribution of a feature when it is added to the “black-box” model, then the SHAP value takes the average value considering the different marginal contribution of the feature in all permutations of individuals (36). A feature with a positive SHAP value improves the output value, and those larger numerical values make greater contributions (37). In our study, the SHAP summary plot, the importance ranking, and the SHAP dependence plot of the relevant covariates were used to improve the interpretability. SHAP explainer was suitable to visualize the black-box ML algorithms on the basis of the cooperative game theory (36). The advantage of SHAP method is be able to explain how much and in which direction each predictor influences the optimal ML-model's output. It concluded that, the core idea of SHAP-explainer is to calculate the marginal contribution of features to model output, and then to explain the “black-box” model from global and local levels (38).

2.5. Data preprocessing

Firstly, indicators including ESR, CRP, TSH, T3, T4 were excluded because of the high missing ratio (over 30% (39), respectively). Secondly, categorical variables were encoded into dummy variables, and the details were as follows: (i) TOAST subtypes were converted into range 1–5 (LAA = 1, CE = 2, SAO = 3, ODC = 4, UND = 5), (ii) Infarction region of CC were divided into range 1–5 (rostrum = 1, genu = 2, body = 3, splenium = 4, at least two of rostrum, genu, body and splenium = 5), and (iii) Other infarction areas were turned into range 0–5 (none = 0, frontal lobe = 1, parietal lobe = 2, temporal lobe = 3, occipital lobe = 4, others = 5), (iv) Location of angiostenosis were encoded into range 0–4 (none = 0, ICA = 1, VBA = 2, both of ICA and VBA = 3), etc. After that, remaining indicators were processed by K-nearest-neighbor (KNN) analysis to impute their missing values (40). In the end, the Borderline-1 SMOTE (BLSMOTE) algorithm was also adopted to balance the samples between the SCD group and non-SCD group in an absolute fairness (for 50%, respectively), which would improve the reliability or classifying performance of the ML-models (41).

2.6. Statistical analysis

Continuous data were uniformly described as mean (SD) or median (IQR), while categorical data were presented as n (%). Baseline characteristics were compared between the SCD group and non-SCD group after CC infarction by Chi-square test (categorical variables), two-sample *t*-test (continuous variables with symmetric distribution), Mann–Whitney *U* test (continuous variables with asymmetric distribution), or Welch's *t*-test (continuous variables with heterogeneity of variance), as appropriate. Then, variables with a relatively remarkable ($p < 0.1$) association with cognitive outcome in univariable analysis were further analyzed by multivariable analysis with a traditional forward stepwise selection. All statistical analyzes were

performed using programming language R package (version 3.6.3, <https://cran.r-project.org/bin/windows/base/>) and all ML-relevant workflows were performed using python (version 3.7, <https://www.python.org/getit/>); $p < 0.05$ indicates statistical significance.

3. Results

3.1. Demographics

The baseline demographical, clinical, biochemical and neuroimaging characteristics of 213 patients (75 female) with acute CC infarction were summarized in [Supplementary Table S1](#). The average age at baseline was 63 [55, 69] years. After 1 year follow-up period, 110 subjects developed into post-stroke SCD, while the remaining were no-complaint (NC) patients. Compared to NC participants, SCD patients tended to be slightly older (63 [58, 71] vs. 61 [51, 68] years, $p = 0.012$), had a higher percentage of female (45.4% vs. 24.2%, $p = 0.001$), and higher mRS scores at 3 month (1 [1–3] vs. 1 [0–2], $p < 0.001$). Pure CC infarction seems to be more likely to cause post-stroke SCD ($p = 0.030$). Meanwhile, the group with more than two subregions involvement of CC infarction were especially vulnerable to post-stroke SCD ($p = 0.001$). Furthermore, patients with post-stroke SCD were prone to have multiple angiostenosis with both of internal carotid artery (ICA) and vertebral basilar artery (VBA) involved ($p = 0.009$).

3.2. Multivariable analysis of risk factors

According to traditional forward selection, we found that female (OR: 3.344; 95% CI: 1.656–6.998; $p = 0.001$), 3-month mRS scores (OR: 1.380; 95% CI: 1.109–1.736; $p = 0.005$), pure CC infarction (OR: 4.823; 95% CI: 1.531–17.919; $p = 0.011$) were the eligible independent risk factors for SCD after acute CC infarction ([Table 1](#)). Compared with the patients with acute infarction of the genu, patients with infarction of the splenium but not rostrum or body were more likely to have cognitive deterioration during follow-up (OR: 3.058; 95% CI: 1.221–8.183; $p = 0.020$). Furthermore, the patients with at least two subregions of CC infarction were more susceptible to post-stroke SCD than those only with lesions in the genu (OR: 7.370; 95% CI: 2.649–22.124; $p < 0.001$).

3.3. Performance of machine learning models

Based on the predictors selected by LASSO in the supplementary materials ([Supplementary Figure S1](#)), different artificial intelligence (AI) -derived models were constructed ([Table 2](#)). According to the metrics, the AUC and accuracy of the LR model were obviously better than those of the other six models, respectively. Therefore, the logistic model was selected as the most prominent one for predicting SCD after acute CC infarction, which achieved an AUC of 0.771 (± 0.042) and accuracy of 0.703 (± 0.050) in the validation set. The ROC-curves and Forest map of AUC values for the LR and the other models were shown in [Figure 3](#).

Moreover, we calculated the contribution of each predictor to LR model by SHAP algorithm, which can simultaneously reveal the

power and direction of these factors. Thereafter, features were ranked on the basis of the absolute SHAP values over all samples (Figure 4A). As is depicted in Figure 4B, high values of infarction subregions of CC, female, 3-month mRS score and pure CC infarction have positive impact on the output of LR model, indicating the acceleration of cognitive deficit after acute CC infarction. Importantly also, age, HCY, NLR, location and number of angiostenosis were the other top-9 predictors for post-stroke SCD based on Shapely value.

SHAP model is a relatively all-powerful ML-model interpretation method, which can also be used for personalized interpretation. That means, individual patient predictions can be extracted to visualize which features played a role in their cognitive decline and what their feature values were. For instance, Figure 4C exhibits a subject with a predicted possibility of 74% for SCD after CC infarction by LR-model. The plot explains that location of angiostenosis = 3.0 (both of ICA and VBA), infarction subregion of CC = 5.0 (at least two of rostrum, genu, body, and splenium) and female = 1 (female) are the most remarkable values contributing to the increased chance of cognitive disorder,

while 3-month mRS score = 0 is just the opposite. Ultimately, the result indicated a high-risk of post-stroke SCD for this subject, and the follow-up result confirmed cognitive impairment outcome, which means true positive. Similarly, Figure 4D exhibits a case with a predicted possibility of 37% for post-stroke SCD, in other words, that means a possibility of 63% for non-SCD after CC infarction. The most essential positive contributors towards adverse cognitive outcome are NLR = 1.9 and HCY = 12.3. Inversely, the negative contributors involve location of angiostenosis = 0.0 (none) and age = 62.0. Therefore, the LR-based algorithm's result was low-risk of SCD after CC infarction for this subject, and the actual outcome was identified as non-cognitive impairment (true negative).

4. Discussion

The presence of SCD is known to be associated with a high risk for objective cognitive decline and even clinical progression to symptomatic disease stages (42, 43). Effective intervention to delay or prevent pathologic cognitive decline may best be targeted at the earliest symptomatic disease stage, such as SCD, in which cognitive function is still relatively preserved (44). This is an exploratory study that for the first time focuses on post-stroke SCD of rare CC infarction via an interpretable machine learning-derived early warning strategy.

After multivariate adjustment for potential confounders, we found that female, 3-month mRS scores, pure CC infarction and infarction subregion of CC independently correlated with the incidence of SCD. Interestingly, our previous study has reported that males had a higher incidence of CC infarction (5), while in the current cohort, we found females were more susceptible to SCD after this specific infarction. Reasons for this phenomenon may include: (1) females in the present study had an older onset-age of CC infarction than males (64 [58,71] vs. 62 [55,69]), (2) women are usually considered to have a larger corpus callosum volume (45–47), indicating that callosum may play a more important role in maintaining brain function of females, (3) women tend to have higher cortisol but lower estradiol levels in menopausal period (48). Indeed, scholars have well-clarified that higher serum cortisol is correlated with more severe microstructural WMLs, particularly in CC, while estrogen are thought to promote the remyelination, and the latter in turn is strongly associated with general cognitive capacity (49–51). Meanwhile, a strong interaction between serum cortisol and cerebral atrophy among females, but not males was also identified (52). Richa et al. (53) once reported that the MoCA scores (between 4–8 weeks post-infarct) were obviously correlated to

TABLE 1 Multivariate logistic regression for the risk factors of post-stroke SCD after callosal infarction.

Variables	Odds ratio (95% CI)	<i>p</i> value
Age, year	1.024 (0.995–1.055)	0.114
Female (yes vs. no)	3.344 (1.656–6.998)	0.001
Uric acid, umol/L	0.999 (0.997–1.001)	0.497
Hypertension (yes vs. no)	1.266 (0.615–2.619)	0.522
3-month mRS scores	1.380 (1.109–1.736)	0.005
Pure CC infarction (yes vs. no)	4.823 (1.531–17.919)	0.011
Number of angiostenosis		
none	1 (ref)	
Seldom	0.537 (0.169–1.662)	0.284
Multiple	1.711 (0.723–4.141)	0.225
Infarction subregion of CC		
Genu	1 (ref)	
Body	3.347 (0.747–15.754)	0.116
Splenium	3.058 (1.221–8.183)	0.020
At least two of above subregions	7.370 (2.649–22.124)	<0.001

mRS, Modified Rankin scale; CC, Corpus callosum. In this table, the term of Multiple means the number of narrowed or occluded vessels is greater than two, and the term of Seldom means the level was between none and multiple.

TABLE 2 Comparison of predictive effects of different machine learning algorithms.

Models	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV	F1 score
XGBoost	0.722 (0.035)	0.618 (0.024)	0.722 (0.100)	0.717 (0.120)	0.704 (0.068)	0.578 (0.022)	0.705 (0.045)
LR	0.771 (0.042)	0.703 (0.050)	0.763 (0.094)	0.730 (0.108)	0.661 (0.077)	0.765 (0.044)	0.703 (0.064)
LightGBM	0.655 (0.066)	0.615 (0.093)	0.757 (0.145)	0.559 (0.230)	0.644 (0.104)	0.604 (0.105)	0.676 (0.042)
AdaBoost	0.691 (0.059)	0.648 (0.053)	0.669 (0.070)	0.688 (0.095)	0.628 (0.087)	0.668 (0.090)	0.642 (0.051)
GNB	0.752 (0.047)	0.700 (0.045)	0.701 (0.085)	0.768 (0.087)	0.722 (0.062)	0.683 (0.062)	0.709 (0.062)
CNB	0.668 (0.044)	0.573 (0.051)	0.711 (0.185)	0.627 (0.187)	0.616 (0.051)	0.562 (0.072)	0.641 (0.079)
SVM	0.647 (0.042)	0.594 (0.026)	0.774 (0.175)	0.496 (0.204)	0.588 (0.048)	0.609 (0.069)	0.660 (0.094)

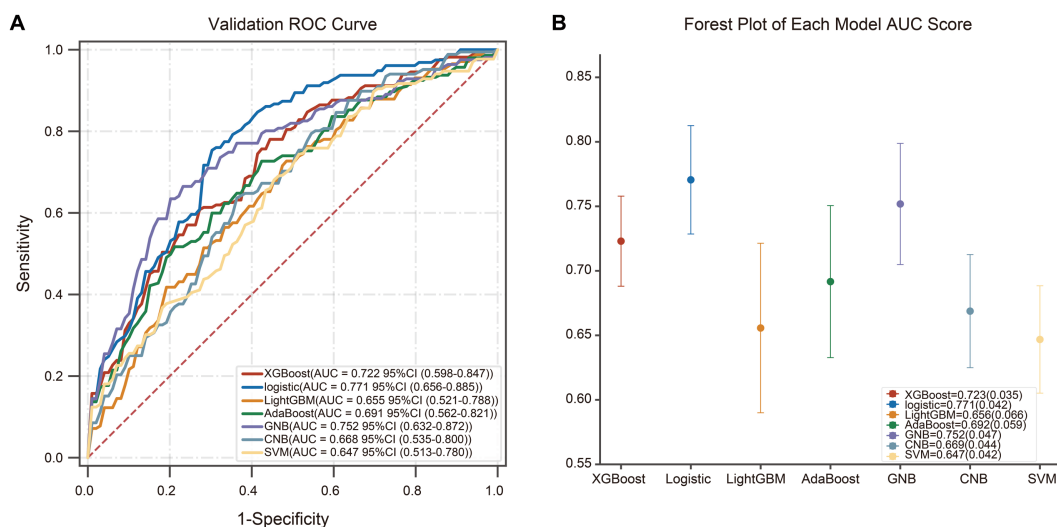


FIGURE 3

The ROC-curves and Forest map of AUC values for seven models. (A) The ROC curves for the different machine learning algorithms, and LR model yielded the greatest AUC among all the models. (B) The Forest map of AUC values of the seven models. The dots mean the AUC-value of each model, and the confidence intervals are depicted by the vertical lines.

mRS scores (at the same follow-up points) among the stroke patients. Then, our results showed that 3-month mRS scores were related to longer-time cognitive outcome due to CC infarction.

The structure of CC can be divided into four classical parts in the order from front to back: rostrum, genu, body and splenium (54). Consistent with previous reports (5), we found that the incidence of 'pure' CC infarction was rare, while the mental disturbance and cognitive dysfunction were more prominent than 'complex' CC infarction. The mechanisms of the discrepancy are still unclear, perhaps the atypical symptoms and insufficient distinguishment of MRI scan made it difficult to draw sufficient attention and appropriate prevention of 'pure' CC infarction. Meanwhile, we reported for the first time that acute infarction in the splenium had a higher tendency of cognitive decline than that in the genu. As the most vulnerable area of the CC infarction, the splenium is more vulnerable to insufficient blood supply, and the splenium lesions were known to be related with cognitive disorder, aphasia, homonymous hemianopsia, alien hand syndrome and so on (54). Therefore, we believe that the splenium plays an important role in the high incidence of SCD caused by CC infarction to some extent. What's more, patients with at least two subregions of CC infarction were more susceptible to SCD than those only had lesions in the genu. This result is well understood given that the more structural damage CC is, the more disrupted the fiber connections and information transmission between the bilateral hemispheres. Besides, evidence showing that the infarction in body or splenium of CC could lead to disturbed executive capacity, attention and calculation (55), which may provide a side note for our viewpoint.

Besides multivariable analysis, LASSO analysis was also adopted to select potential risk predictors by eliminating irrelevant features. It is universally accepted that age was a risk factor of cognition damage after various types of ischemic stroke (56). Except of age, evidence linking high HCY (HHCY) and cognitive decline is profuse (57). It is known that, HHCY is not only associated with WMLs, but also the progression of WMLs (58). In the meantime, extensive intracranial vascular stenosis is another promotor for SCD after CC infarction.

Cerebral angiostenosis/occlusion has already been proved to induce hypoperfusion and impaired executive dysfunctions, such as working memory, attention, cognitive flexibility, planning, thought organization and implementation (59). This phenomenon indicates that appropriate increase of cerebral blood flow may help prevent post-stroke SCD. Interestingly, NLR is often known as a risk factor for PSCI (60). However, we found that NLR is negatively associated with self-report cognitive decline, indicating that NLR is likely to act as a compensatory neuroprotective response in the early stage of CC infarction. Biological mechanisms between NLR and risk of post-stroke SCD have not been explored before and warrants further clarifications, especially in the condition of CC infarction.

In our study, the combination of LASSO regression and ML-based models was beneficial to identify the optimal configuration to predict whether it is vulnerable to develop SCD after CC infarction or not. Then, the seven ML algorithms were assessed by several metrics, comprising AUC-value, accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), as well as F1 scores. Apart from global explanation, the well-accepted local explanation, SHAP was also implemented to interpret how a complex black-box ML model makes a prediction (61). By incorporating the individualized patient profile, the level of contribution and directionality of specific input features were visualized (62). As shown in the Table 1, the LR model seemed to be the best-performance classifier with the highest scores of AUC-value (77.1%), accuracy (70.3%), sensitivity (76.3%) and NPV (76.5%). In addition, acceptable values of specificity, PPV and F1 score (all above 65.0%) were achieved in the validation set. Taken together, we selected the LR model as the optimal algorithm with the best generalization ability. At the same time, we suggested that we should treat this problem dialectically and choose appropriate predictive classifier according to different clinical needs.

The strength of our research is that the cohort has the largest sample of CC infarction in the world, and the datasets are non-synthetic, which is more likely to be objective and effective as a screening tool. Unlike studies focused on each risk factor individually

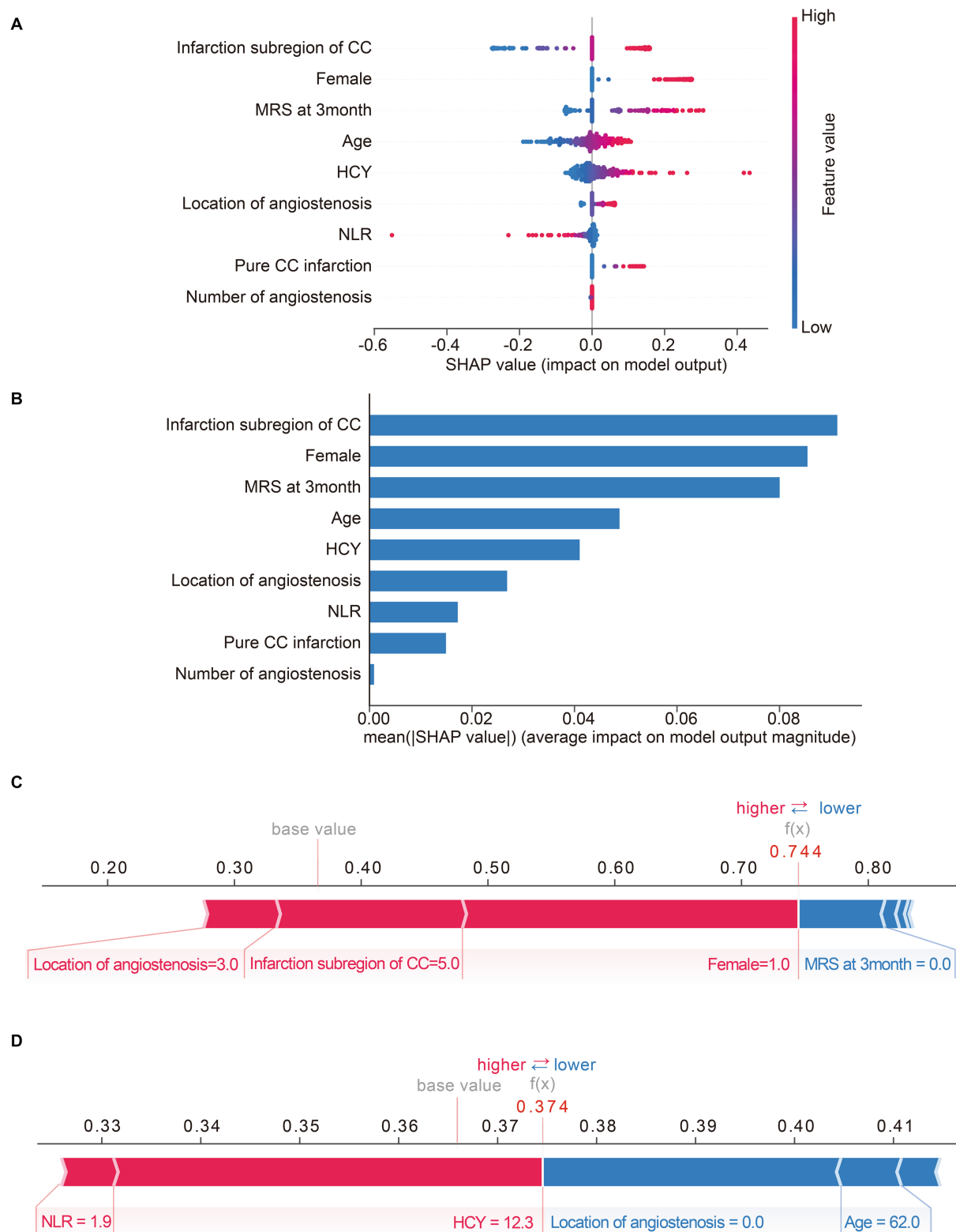


FIGURE 4

Matrix plots of the top nine important features and the SHAP plots for two selected patients. (A) The SHAP summary plot of LR model. Each dot represents a SHAP-value for a feature. The red color means high value, while the blue means low. The positive SHAP-value represents an increased risk of post-stroke SCD for the output of LR model, and vice versa. (B) The histogram of mean absolute SHAP values of top-nine important features of LR model. The longer the bar, the larger impact the feature has on the output. (C,D) SHAP force plots for two selected patients. Feature values colored red are pushing the predictive outcome towards cognitive impairment, while feature values colored blue are just the opposite. The associated Shapley value of each feature is visualized by the length of an arrow, and the longer of the arrow, the more significant the feature value is.

or its pathophysiological interpretation (63, 64), we aimed to encompass a large combination of variables from real-world clinical situations once. The variables we used, including demographics, laboratory and radiological findings were all easily accessible for clinicians, which could assist with the early prediction and prevention for suspected post-stroke SCD. Additionally, an interpretable and explainable ML model was created with the help of SHAP-explainer, promoting to make individualized clinical decisions.

There are some limitations that still needed to be ironed out in our study. Firstly, although this investigation had the maximal population of patients with acute CC infarction, the sample size was still needed to be added. Secondly, we did not exploit the different cognitive abilities separately, such as orientation, calculation, executive abilities, long-term and short-term memory and attention, etc. Thirdly, the follow-up period is not long enough to verify the proportion of patients with SCD who eventually converted to PSCI. Therefore, multicenter prospective cohorts with detailed cognitive domains impairment are needed in the future.

5. Conclusion

In conclusion, the present study screened out 9 key features associated with post-SCD and developed a LR-model which can improve the prediction accuracy of one-year SCD after CC infarction. What's more, the individual report generated by SHAP facilitate the early implementation of primary prevention measures. Based on these techniques, we are even expected to continue to individually predict the long-term effects of different clinical drugs on cognitive impairment to shape a brighter future for patients with CC infarction.

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 authors.

Ethics statement

The studies involving human participants were reviewed and approved by Changhai Hospital Ethics Committee (No.

CHEC2021-1021). The patients/participants provided their written informed consent to participate in this study.

Author contributions

YX, XS, and XB: conceived and designed the study. YX, XS, YL, YH, ML, RS, GY, CS, QD, BD, and XB: performed the study. YX, XS, YL, BD, and XB: revised the article for intellectual content. YX, XS, and BD: wrote the article. All authors contributed to the article and approved the submitted version.

Funding

The work was financially supported by the National Natural Science Foundation of China (81871040 and 82101563), the Clinical Research Plan of SHDC (no. SHDC2020CR1038B), Scientific research project of Shanghai Health Commission (20214Y0500), and the Youth Program of Naval Medical University (2021JCQN10).

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2023.1123607/full#supplementary-material>

References

- De León Reyes NS, Bragg-Gonzalo L, Nieto M. Development and plasticity of the corpus callosum. *Development*. (2020) 147:dev189738. doi: 10.1242/dev.189738
- Türe U, Yaşargil MG, Krisht AF. The arteries of the corpus callosum: a microsurgical anatomic study. *Neurosurgery*. (1996) 39:1075–84. doi: 10.1097/00006123-199612000-00001
- Chrysikopoulos H, Andreou J, Roussakis A, Pappas J. Infarction of the corpus callosum: computed tomography and magnetic resonance imaging. *Eur J Radiol*. (1997) 25:2–8. doi: 10.1016/s0720-048x(96)01155-2
- Giroud M, Dumas R. Clinical and topographical range of callosal infarction: a clinical and radiological correlation study. *J Neurol Neurosurg Psychiatry*. (1995) 59:238–42. doi: 10.1136/jnnp.59.3.238
- Sun X, Li J, Fan C, Zhang H, Si Y, Fang X. Clinical, neuroimaging and prognostic study of 127 cases with infarction of the corpus callosum. *Eur J Neurol*. (2019) 26:1075–81. doi: 10.1111/ene.13942
- Gelibter S, Genchi A, Callea M, Anzalone N, Galantucci S, Volonté MA. Corpus callosum infarction: radiological and histological findings. *J Neurol*. (2020) 267:3418–20. doi: 10.1007/s00415-020-10224-8
- Jessen F, Spottke A, Boecker H, Brosseron F, Buerger K, Catak C. Design and first baseline data of the DZNE multicenter observational study on predementia Alzheimer's disease (DELCODE). *Alzheimers Res Ther*. (2018) 10:15. doi: 10.1186/s13195-017-0314-2
- Rabin LA, Smart CM, Crane PK, Amariglio RE, Berman LM, Boada M. Subjective cognitive decline in older adults: an overview of self-report measures used across 19 international research studies. *J Alzheimers Dis*. (2015) 48:S63–86. doi: 10.3233/JAD-150154
- Abner EL, Kryscio RJ, Caban-Holt AM, Schmitt FA. Baseline subjective memory complaints associate with increased risk of incident dementia: the PREADVICE trial. *J Prev Alzheimers Dis*. (2015) 2:11–6. doi: 10.14283/jpad.2015.37

10. Reisberg B, Shulman MB, Torossian C, Leng L, Zhu W. Outcome over seven years of healthy adults with and without subjective cognitive impairment. *Alzheimers Dement.* (2010) 6:11–24. doi: 10.1016/j.jalz.2009.10.002
11. Des Portes V, Rolland A, Velazquez-Dominguez J, Peyric E, Cordier MP, Gaucherand P. Outcome of isolated agenesis of the corpus callosum: a population-based prospective study. *Eur J Paediatr Neurol.* (2018) 22:82–92. doi: 10.1016/j.ejpn.2017.08.003
12. Huynh-Le MP, Tibbs MD, Karunamuni R, Salans M, Tringale KR, Yip A. Microstructural injury to Corpus callosum and Intrahemispheric White matter tracts correlate with attention and processing speed decline after brain radiation. *Int J Radiat Oncol Biol Phys.* (2021) 110:337–47. doi: 10.1016/j.ijrobp.2020.12.046
13. Platten BM, Martola J, Fink K, Ouellette R, Piehl F, Granberg T. MRI-based manual versus automated Corpus callosum volumetric measurements in multiple sclerosis. *J Neuroimaging.* (2020) 30:198–204. doi: 10.1111/jon.12676
14. Sidtis JJ, Volpe BT, Holtzman JD, Wilson DH, Gazzaniga MS. Cognitive interaction after staged callosal section: evidence for transfer of semantic activation. *Science (New York, NY).* (1981) 212:344–6. doi: 10.1126/science.6782673
15. Jessen F, Amariglio RE, van Boxtel M, Breteler M, Ceccaldi M, Chételet G. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. *Alzheimers Dement.* (2014) 10:844–52. doi: 10.1016/j.jalz.2014.01.001
16. Smart CM, Karr JE, Areshenkoff CN, Rabin LA, Hudon C, Gates N. Non-pharmacologic interventions for older adults with subjective cognitive decline: systematic review, Meta-analysis, and preliminary recommendations. *Neuropsychol Rev.* (2017) 27:245–57. doi: 10.1007/s11065-017-9342-8
17. Deo RC. Machine learning in medicine. *Circulation.* (2015) 132:1920–30. doi: 10.1161/CIRCULATIONAHA.115.001593
18. 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
19. Lee H, Lee EJ, Ham S, Lee HB, Lee JS, Kwon SU. Machine learning approach to identify stroke within 4.5 hours. *Stroke.* (2020) 51:860–6. doi: 10.1161/STROKEAHA.119.027611
20. Vodencarevic A, Weingärtner M, Caro JJ, Ukalovic D, Zimmermann-Rittereiser M, Schwab S. Prediction of recurrent ischemic stroke using registry data and machine learning methods: the Erlangen stroke registry. *Stroke.* (2022) 53:2299–306. doi: 10.1161/STROKEAHA.121.036557
21. Xu D, Song Y, Meng Y, István B, Gu Y. Relationship between firefighter physical fitness and special ability performance: predictive research based on machine learning algorithms. *Int J Environ Res Public Health.* (2020) 17:7689. doi: 10.3390/ijerph17207689
22. Vélez JI. Machine learning based psychology: advocating for a data-driven approach. *Int J Psychol Res (Medellin).* (2021) 14:6–11. doi: 10.21500/20112084.5365
23. Phellan R, Hachem B, Clin J, Mac-Thiong JM, Duong L. Real-time biomechanics using the finite element method and machine learning: review and perspective. *Med Phys.* (2021) 48:7–18. doi: 10.1002/mp.14602
24. Watson DS, Krutzinna J, Bruce IN, Griffiths CE, McInnes IB, Barnes MR. Clinical applications of machine learning algorithms: beyond the black box. *BMJ.* (2019) 364:l886. doi: 10.1136/bmj.l886
25. Liu S, Schlesinger JJ, McCoy AB, Reese TJ, Steitz B, Russo E, et al. New onset delirium prediction using machine learning and long short-term memory (LSTM) in electronic health record. *J Am Med Inform Assoc.* (2022) 30:120–31. doi: 10.1093/jamia/ocac210
26. Adams HP Jr, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL. 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
27. Diener HC, Hankey GJ. Primary and secondary prevention of ischemic stroke and cerebral hemorrhage: JACC focus seminar. *J Am Coll Cardiol.* (2020) 75:1804–18. doi: 10.1016/j.jacc.2019.12.072
28. Xu P, Zhang F, Cheng J, Huang Y, Ren Z, Ye R, et al. The relationship between physical activity and subjective cognitive decline: evidence from the behavioral risk factor surveillance system (BRFSS). *J Affect Disord.* (2023) 328:108–15. doi: 10.1016/j.jad.2023.02.045
29. Gupta S. Racial and ethnic disparities in subjective cognitive decline: a closer look, United States, 2015–2018. *BMC public health* (2021) 21:1173. doi: 10.1186/s12889-021-11068-1
30. Burns SP, Mueller M, Magwood G, White BM, Lackland D, Ellis C. Racial and ethnic differences in post-stroke subjective cognitive decline exist. *Disabil Health J.* (2019) 12:87–92. doi: 10.1016/j.dhjo.2018.08.005
31. Taylor CA, Bouldin ED, Greenlund KJ, McGuire LC. Comorbid chronic conditions among older adults with subjective cognitive decline, United States, 2015–2017. *Innov Aging.* (2020) 4:igz045. doi: 10.1093/geroni/igz045
32. Martínez-Laperche C, Buces E, Aguilera-Morillo MC, Picornell A, González-Rivera M, Lillo R. A novel predictive approach for GVHD after allogeneic SCT based on clinical variables and cytokine gene polymorphisms. *Blood Adv.* (2018) 2:1719–37. doi: 10.1182/bloodadvances.2017011502
33. Laukhtina E, Schuettfort VM, D'Andrea D, Pradere B, Quhal F, Mori K. Selection and evaluation of preoperative systemic inflammatory response biomarkers model prior to cytoreductive nephrectomy using a machine-learning approach. *World J Urol.* (2022) 40:747–54. doi: 10.1007/s00345-021-03844-w
34. Tseng PY, Chen YT, Wang CH, Chiu KM, Peng YS, Hsu SP. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Crit Care.* (2020) 24:478. doi: 10.1186/s13054-020-03179-9
35. Ye WL, Shen C, Xiong GL, Ding JJ, Lu AP, Hou TJ. Improving docking-based virtual screening ability by integrating multiple energy auxiliary terms from molecular docking scoring. *J Chem Inf Model.* (2020) 60:4216–30. doi: 10.1021/acs.jcim.9b00977
36. Czub N, Paclawski A, Szłęk J, Mendyk A. Curated database and preliminary AutoML QSAR model for 5-HT1A receptor. *Pharmaceutics.* (2021) 13:1711. doi: 10.3390/pharmaceutics13101711
37. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng.* (2018) 2:749–60. doi: 10.1038/s41551-018-0304-0
38. Duckworth C, Chmiel FP, Burns DK, Zlatev ZD, White NM, Daniels TWV. Using explainable machine learning to characterise data drift and detect emergent health risks for emergency department admissions during COVID-19. *Sci Rep.* (2021) 11:23017. doi: 10.1038/s41598-021-02481-y
39. Zhao Y, Chen Q, Liu T, Luo P, Zhou Y, Liu M. Development and validation of predictors for the survival of patients with COVID-19 based on machine learning. *Front Med.* (2021) 8:683431. doi: 10.3389/fmed.2021.683431
40. Bania RK, Halder A. R-HEFS: rough set based heterogeneous ensemble feature selection method for medical data classification. *Artif Intell Med.* (2021) 114:102049. doi: 10.1016/j.artmed.2021.102049
41. Fotouhi S, Asadi S, Kattan MW. A comprehensive data level analysis for cancer diagnosis on imbalanced data. *J Biomed Inform.* (2019) 90:103089. doi: 10.1016/j.jbi.2018.12.003
42. Jessen F, Wolfsgruber S, Wiese B, Bickel H, Mösch E, Kaduszkiewicz H. AD dementia risk in late MCI, in early MCI, and in subjective memory impairment. *Alzheimers Dement.* (2014) 10:76–83. doi: 10.1016/j.jalz.2012.09.017
43. Snitz BE, Wang T, Cloonan YK, Jacobsen E, Chang CH, Hughes TF. Risk of progression from subjective cognitive decline to mild cognitive impairment: the role of study setting. *Alzheimers Dement.* (2018) 14:734–42. doi: 10.1016/j.jalz.2017.12.003
44. Rabin LA, Smart CM, Amariglio RE. Subjective cognitive decline in preclinical Alzheimer's disease. *Annu Rev Clin Psychol.* (2017) 13:369–96. doi: 10.1146/annurev-clinpsy-032816-045136
45. Ardekani BA, Figarsky K, Sidtis JJ. Sexual dimorphism in the human corpus callosum: an MRI study using the OASIS brain database. *Cereb Cortex.* (2013) 23:2514–20. doi: 10.1093/cercor/bhs253
46. Eliot L, Ahmed A, Khan H, Patel J. Dump the "dimorphism": comprehensive synthesis of human brain studies reveals few male-female differences beyond size. *Neurosci Biobehav Rev.* (2021) 125:667–97. doi: 10.1016/j.neubiorev.2021.02.026
47. Potvin O, Mouiha A, Dieumegarde L, Duchesne S. Corrigendum to 'Normative data for subcortical volumes over the lifetime of the adult human brain' [NeuroImage 137 (2016) 9–20]. *NeuroImage.* (2018) 183:994–5. doi: 10.1016/j.neuroimage.2018.09.020
48. Mezzullo M, Gambineri A, Di Dalmazi G, Fazzini A, Magagnoli M, Baccini M. Steroid reference intervals in women: influence of menopause, age and metabolism. *Eur J Endocrinol.* (2021) 184:395–407. doi: 10.1530/EJE-20-1147
49. He Q, Luo Y, Lv F, Xiao Q, Chao F, Qiu X. Effects of estrogen replacement therapy on the myelin sheath ultrastructure of myelinated fibers in the white matter of middle-aged ovariectomized rats. *J Comp Neurol.* (2018) 526:790–802. doi: 10.1002/cne.24366
50. Luo Y, Xiao Q, Chao F, He Q, Lv F, Zhang L. 17 β -estradiol replacement therapy protects myelin sheaths in the white matter of middle-aged female ovariectomized rats: a stereological study. *Neurobiol Aging.* (2016) 47:139–48. doi: 10.1016/j.neurobiolaging.2016.07.023
51. Penke L, Maniega SM, Bastin ME, Valdés Hernández MC, Murray C, Royle NA. Brain white matter tract integrity as a neural foundation for general intelligence. *Mol Psychiatry.* (2012) 17:1026–30. doi: 10.1038/mp.2012.66
52. Echouffo-Tcheugui JB, Conner SC, Himali JJ, Maillard P, DeCarli CS, Beiser AS. Circulating cortisol and cognitive and structural brain measures: the Framingham heart study. *Neurology.* (2018) 91:e1961–70. doi: 10.1212/WNL.0000000000006549
53. Sharma R, Mallick D, Llinas RH, Marsh EB. Early post-stroke cognition: in-hospital predictors and the association with functional outcome. *Front Neurol.* (2020) 11:613607. doi: 10.3389/fneur.2020.613607
54. Katsuki M, Kato H, Niizuma H, Nakagawa Y, Tsunoda M. Homonymous Hemianopsia due to the infarction in the splenium of the Corpus callosum. *Cureus.* (2021) 13:e19574. doi: 10.7759/cureus.19574
55. Huang X, Du X, Song H, Zhang Q, Jia J, Xiao T. Cognitive impairments associated with corpus callosum infarction: a ten cases study. *Int J Clin Exp Med.* (2015) 8:21991–8.
56. Pendlebury ST, Rothwell PM Oxford Vascular Study. Incidence and prevalence of dementia associated with transient ischaemic attack and stroke: analysis of the

population-based Oxford vascular study. *Lancet Neurol.* (2019) 18:248–58. doi: 10.1016/S1474-4422(18)30442-3

57. Hannibal L, Blom HJ. Homocysteine and disease: causal associations or epiphenomenons? *Mol Asp Med.* (2017) 53:36–42. doi: 10.1016/j.mam.2016.11.003

58. Vermeer SE, van Dijk EJ, Koudstaal PJ, Oudkerk M, Hofman A, Clarke R. Homocysteine, silent brain infarcts, and white matter lesions: the Rotterdam scan study. *Ann Neurol.* (2002) 51:285–9. doi: 10.1002/ana.10111

59. Zhao JH, Tian XJ, Liu YX, Yuan B, Zhai KH, Wang CW. Executive dysfunction in patients with cerebral hypoperfusion after cerebral angiostenosis/occlusion. *Neurol Med Chir.* (2013) 53:141–7. doi: 10.2176/nmc.53.141

60. Lee M, Lim JS, Kim CH, Lee SH, Kim Y, Hun Lee J. High neutrophil-lymphocyte ratio predicts post-stroke cognitive impairment in acute ischemic stroke patients. *Front Neurol.* (2021) 12:693318. doi: 10.3389/fneur.2021.693318

61. Zheng Y, Guo Z, Zhang Y, Shang J, Yu L, Fu P. Rapid triage for ischemic stroke: a machine learning-driven approach in the context of predictive, preventive and personalised medicine. *EPMA J.* (2022) 13:285–98. doi: 10.1007/s13167-022-00283-4

62. Kim S, Jeon E, Yu S, Oh K, Kim C, Song T. Interpretable machine learning for early neurological deterioration prediction in atrial fibrillation-related stroke. *Sci Rep.* (2021) 11:20610. doi: 10.1038/s41598-021-99920-7

63. Chander RJ, Lim L, Handa S, Hiu S, Choong A, Lin X. Atrial fibrillation is independently associated with cognitive impairment after ischemic stroke. *J Alzheimers Dis.* (2017) 60:867–75. doi: 10.3233/JAD-170313

64. Wang Q, Wang K, Ma Y, Li S, Xu Y. Serum Galectin-3 as a potential predictive biomarker is associated with Poststroke cognitive impairment. *Oxidative Med Cell Longev.* (2021) 2021:5827812–7. doi: 10.1155/2021/5827812



OPEN ACCESS

EDITED BY

Jean-Claude Baron,
University of Cambridge, United Kingdom

REVIEWED BY

Lucas Alexandre Ramos,
Academic Medical Center, Netherlands
Shubham Misra,
Yale University, United States

*CORRESPONDENCE

Xiaoning Li
✉ lixiaoning456@163.com

RECEIVED 08 September 2022

ACCEPTED 25 May 2023

PUBLISHED 13 June 2023

CITATION

Li Q, Chi L, Zhao W, Wu L, Jiao C, Zheng X,
Zhang K and Li X (2023) Machine learning
prediction of motor function in chronic stroke
patients: a systematic review and
meta-analysis. *Front. Neurol.* 14:1039794.
doi: 10.3389/fneur.2023.1039794

COPYRIGHT

© 2023 Li, Chi, Zhao, Wu, Jiao, Zheng, Zhang
and Li. This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Machine learning prediction of motor function in chronic stroke patients: a systematic review and meta-analysis

Qinglin Li¹, Lei Chi², Weiying Zhao¹, Lei Wu³, Chuanxu Jiao⁴,
Xue Zheng¹, Kaiyue Zhang¹ and Xiaoning Li^{2*}

¹Second Clinical Medical School, Heilongjiang University of Chinese Medicine, Harbin, Heilongjiang, China, ²Department of Acupuncture, The Second Affiliated Hospital of Heilongjiang University of Chinese Medicine, Harbin, Heilongjiang, China, ³Department of Acupuncture, The Third Affiliated Hospital of Zhejiang Chinese Medical University, Hangzhou, Zhejiang, China, ⁴Department of Neurorehabilitation, Taizhou Enze Medical Center Luqiao Hospital, Taizhou, Zhejiang, China

Background: Recent studies have reported that machine learning (ML), with a relatively strong capacity for processing non-linear data and adaptive ability, could improve the accuracy and efficiency of prediction. The article summarizes the published studies on ML models that predict motor function 3–6 months post-stroke.

Methods: A systematic literature search was conducted in PubMed, Embase, Cochrane and Web of Science as of April 3, 2023 for studies on ML prediction of motor function in stroke patients. The quality of the literature was assessed using the Prediction model Risk Of Bias Assessment Tool (PROBAST). A random-effects model was preferred for meta-analysis using R4.2.0 because of the different variables and parameters.

Results: A total of 44 studies were included in this meta-analysis, involving 72,368 patients and 136 models. Models were categorized into subgroups according to the predicted outcome Modified Rankin Scale cut-off value and whether they were constructed based on radiomics. C-statistics, sensitivity, and specificity were calculated. The random-effects model showed that the C-statistics of all models were 0.81 (95% CI: 0.79; 0.83) in the training set and 0.82 (95% CI: 0.80; 0.85) in the validation set. According to different Modified Rankin Scale cut-off values, C-statistics of ML models predicting Modified Rankin Scale > 2 (used most widely) in stroke patients were 0.81 (95% CI: 0.78; 0.84) in the training set, and 0.84 (95% CI: 0.81; 0.87) in the validation set. C-statistics of radiomics-based ML models in the training set and validation set were 0.81 (95% CI: 0.78; 0.84) and 0.87 (95% CI: 0.83; 0.90), respectively.

Conclusion: ML can be used as an assessment tool for predicting the motor function in patients with 3–6 months of post-stroke. Additionally, the study found that ML models with radiomics as a predictive variable were also demonstrated to have good predictive capabilities. This systematic review provides valuable guidance for the future optimization of ML prediction systems that predict poor motor outcomes in stroke patients.

Systematic review registration: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022335260, identifier: CRD42022335260.

KEYWORDS

machine learning, model prediction, stroke, motor function, systematic review

1. Introduction

Stroke is an acute cerebrovascular disease caused by sudden rupture of intracranial vessels or vascular obstruction preventing blood from flowing into the brain and thereby leading to brain tissue damage. Based on its pathological pattern, stroke can be classified into ischemic stroke (IS) and hemorrhagic stroke (HS). The Global Burden of Diseases, Injuries, and Risk Factors Study 2017 (GBD 2017) reported that stroke resulted in 6.17 million deaths and is the second leading cause of death and disability worldwide (1). According to the 2021 Guideline for the Prevention of Stroke in Patients With Stroke and Transient Ischemic Attack From the American Stroke Association (ASA), high blood pressure, diet, abdominal obesity, physical inactivity and smoking represent 82% (2) of the population-attributable risk (PAR) in patients with IS and HS. Although most IS patients have received effective treatments, many of them still suffer certain functional impairment after treatment. Motor function outcome in stroke survivors, as a primary determinant of the burden of stroke, directly determines their quality of life. Furthermore, physical disability is a key factor for the occurrence of mental disorders, such as depression, which occurs in 33% of stroke survivors (3). Therefore, the motor function-related outcome is one of the greatest concerns for stroke patients and their families. Clinically, it is extremely significant for clinicians to judge the prognosis of stroke patients and make a long-term treatment plan for those with a poor motor function outcome (4).

With rapid advances in medical and health informatization, medical data in a larger scale can be divided into more types, and the health care field has also entered a new era of big data. Due to the large scale, diversified types and high hidden value of medical data, ML algorithms have been widely used in the medical field (5–8). ML can be defined as a subfield of artificial intelligence (AI) that uses computerized algorithms to automatically improve performance through an iterative learning process or experience (i.e., data collection) (9). Different from traditional prediction models that use selected variables for calculation, ML techniques can easily incorporate a large number of variables to describe the complex and unpredictable nature of human physiology in a clearer way. Therefore, ML may be helpful for clinical prediction and identification of new prognostic markers (10). In recent years, many ML methods have been applied to the diagnosis and assessment of stroke (11, 12), including the evaluation of stroke severity (13), analysis of cerebral edema (14), prediction of hematoma expansion (15), and incidence prediction (16). Therefore, ML model predictions not only aid in disease analysis, prevention, diagnosis, and patient monitoring, but also help clinicians handle massive amounts of data in a more accurate and efficient manner (17).

In the literature, the published systematic reviews lack ML model prediction analysis of motor function 3–6 months post-stroke, especially regarding model predictive capabilities in different outcome cut-off values to accurately determine efficacy of ML predictions. Additionally, individual original studies may not be able to statistically assess the robustness of prediction results. Therefore, this systematic review and meta-analysis was conducted to assess the performance of current ML models as clinical tools for

predicting medium- and long-term recovery of motor function in stroke patients.

2. Methods

This systematic review and meta-analysis was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2020 (PRISMA 2020), and prospectively registered in PROSPERO (CRD42022335260).

2.1. Search strategy

A comprehensive and systematic search was conducted in PubMed, Embase, Cochrane, and Web of Science databases. The retrieval was as of April 3, 2023. A researcher (Xiaoning Li) designed the keywords and search strategy of this systematic review, and both subject headings and free words were searched. The complete search strategy can be found in the [Supplementary Table S1](#).

2.2. Inclusion and exclusion criteria

2.2.1. Inclusion criteria

(1) Patients were diagnosed with IS or HS on CT or MRI. IS included large vessel occlusion, anterior circulation infarction and posterior circulation infarction. (2) ML was used to predict motor function in patients 3–6 months post-stroke.

(3) The Modified Rankin Scale was used as the outcome measure. (4) Aged 18 and older. (5) Articles written in English or translated into English. (6) Randomized controlled trials, cohort studies, case-control studies.

2.2.2. Exclusion criteria

(1) Patients were clumsy in physical activities or unable to function independently before stroke (Premorbid Modified Rankin Scale ≥ 2). (2) Cerebral hemorrhage resulted from secondary causes, such as cerebral trauma and subarachnoid hemorrhage. (3) Prediction models applied clinical scoring rather than ML. (4) Case reports, protocols, editorials, and perspectives that have no original data.

2.3. Literature screening and data extraction

All of the retrieved studies were imported into Endnote for management. After automatic and manual removal of duplicates, two researchers (Weiyang Zhao and Xue Zheng) independently assessed remaining articles. Titles and abstracts were preliminarily screened before the full texts were downloaded. Then we read the full texts to select eligible studies that meet the inclusion criteria. If there was any dissent on a study, a third researcher (Lei Chi) was consulted to assist in determining whether to include it. Before data extraction, a sheet of standard data extraction was prepared,

including data source, Modified Rankin Scale cut-off value that was defined as a poor outcome, outcome prediction time, missing data processing methods, sample sizes of the training and validation sets, validation set confirmation way, internal and external validation information, predictors and their number, as well as ML model types. Data of the accuracy metrics were also collected, including sensitivity, specificity, receiver operator characteristic (ROC), area under the curve (AUC) and other.

2.4. Quality analysis

The ROB of included studies was assessed using PROBAST (18). It involves four major domains: participants, predictors, outcomes and statistic analysis, and reflects the overall ROB and applicability. The four domains include two, three, six and nine signaling questions, respectively. Signaling questions are answered as yes/probable yes (Y/PY), no/probably no (N/PN), or no information (NI). If a domain is answered with at least a N/PN, it is considered at high ROB. When all of the four domains were rated as low ROB, the overall ROB is deemed to be low. Two researchers (CJ and KZ) independently carried out ROB assessment in accordance with PROBAST. Then their assessment results were cross checked. Any dissent was consulted to a third researcher (LW) for final determination.

2.5. Data analysis

We performed a meta-analysis of the metrics (C-statistics and accuracy) for evaluating ML models. If C-statistic lacked 95% confidence interval (CI) and standard error (SE), we referred to the study by Debray TP et al. (19) to estimate its standard error. In case of inaccurate original data, we calculated based on sensitivity and specificity in combination with the sample size of each molecular subtype and model. Given the difference in variables and parameters in ML models, a random-effects model was preferred to perform the meta-analysis. This meta-analysis was conducted using R4.2.0 (R development Core Team, Vienna, <http://www.R-project.org>). A subgroup analysis by ML models and Modified Rankin Scale threshold was performed in our systematic review and meta-analysis. Heterogeneity was quantified by calculating I^2 as a percentage. A low level of heterogeneity was present when I^2 was 25%, a moderate level when I^2 was 50% and a high level when I^2 was 75% (20). Publication bias was examined by creating a funnel plot and Begg's bias test.

3. Results

3.1. Literature search

A total of 23,594 articles were initially searched from PubMed, Embase, Cochrane and Web of Science. In the screening process, not Modified Rankin Scale outcome, Modified Rankin Scale outcome time <90 days, not clearly identify prediction standard all exclusion conditions. After screening, a total of 44

papers were eligible (21–64). The selection process is shown in Figure 1.

3.2. Characteristics of included studies

A total of 44 eligible studies were included in this systematic review, involving 72,368 patients and 136 prediction models. All prediction models were internally validated, and 26 of them were externally validated. Additionally, among the 28 eligible studies, 15 studies were multi-centered (27–29, 39, 45, 47, 52, 55, 58–64); five studies extracted their original data from databases (24, 25, 32, 37, 57); and the remaining 24 studies were single-centered. Most eligible studies focused on IS and 7 studies (23, 32, 39, 40, 48, 60, 63) on HS. A study (32) used the same four models to predict the motor function in patients with IS and HS respectively. The primary outcome was Modified Rankin Scale at 3 to 6 months after the onset of stroke. Due to distinct purposes, these included studies defined poor motor function in a different manner. Specifically, it was defined as Modified Rankin Scale >1 in 4 studies (26, 31, 45, 58), Modified Rankin Scale >3 in 4 (39, 53, 60, 63) studies, Modified Rankin Scale >4 in two (25, 42) studies, and Modified Rankin Scale >2 in other 34 studies. In this study, a total of 13 primary studies using radiomics as predictive factors were identified (27, 30, 36, 37, 49, 51, 52, 54–57, 61, 62), from which 33 prediction models were extracted. Among them, 20 models primarily used MRI images as predictive factors.

In the training set, models of LR (Logistic Regression), RF (Random Forest), SVM (Support Vector Machine), XGB (Extreme Gradient Boosting), ANN (Artificial Neural Networks), DT (Decision Tree), GBM (Gradient Boosting Machine), DNN (Deep Neural Network), ADB (Adaptive Boosting), KNN (K-nearest Neighbors), CNN (Convolutional Neural Network) were applied. C-statistic was used for analysis. Meanwhile, the sensitivity and specificity of LR, RF, SVM, XGB, ANN, DT, ADB, CNN, Naive Bayes were calculated. In contrast, the validation set applied LR, ANN, RF, SVM, XGB, DT, GBM, KNN and GLM and used C-statistics to evaluate these models. The sensitivity and specificity of LR, RF, SVM, XGB, ANN, DT, and KNN were also calculated. Furthermore, these included studies were published from 2017 to 2023, which generally shows an increasing trend year by year. This reflects the rising popularity of ML prediction. Characteristics of included studies are presented in the [Supplementary Tables S2–S4](#).

3.3. Quality assessment

In terms of the evaluation of case source, 99 of 136 models came from retrospective case-control studies, which was rated as high ROB. The high bias of data in retrospective studies caused limited accuracy of models in prediction. Thus these models were at high ROB. The other 37 models extracted clinical data from prospective cohort studies or registration data platforms, thereby rated at low ROB. Regarding the assessment of predictors, 67 models were rated at high ROB. Since researchers knew both predictors and data results in retrospective studies, there was a high ROB in such studies. For the outcome, 109 models were rated at low ROB,

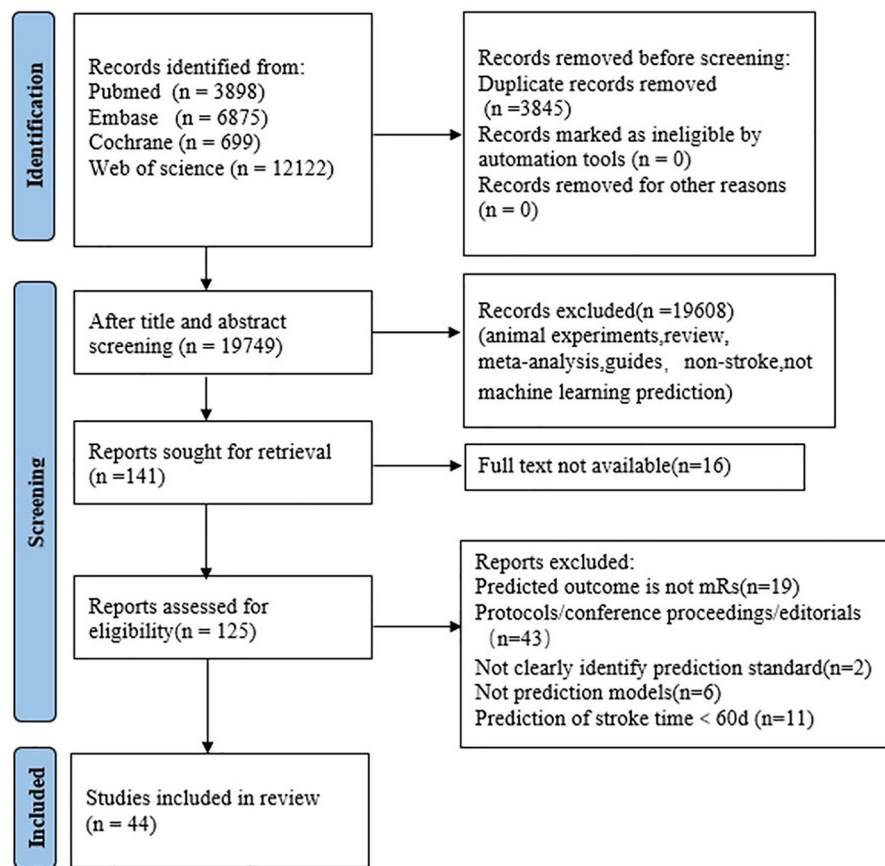


FIGURE 1
PRISMA flow diagram of study selection.

and 22 models were at unclear ROB for failing to report whether the predictor information was unclear at the time of outcome determination. Lastly, as for the analysis, 88 models were at high ROB, among them 60 models were rated at high ROB because of sample size <100 or events per variable (EPV) <10, and 43 models at unclear ROB due to a failure to elucidate the processing method of missing data, data complexity and optimal fitting method. The ROB evaluation is presented in Figure 2.

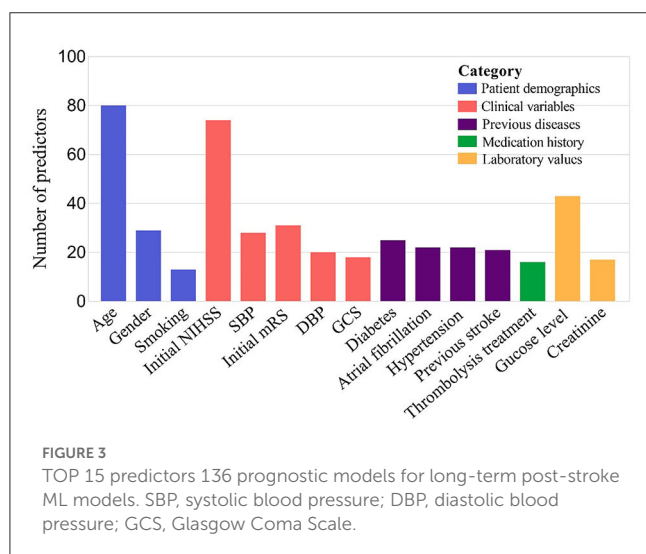
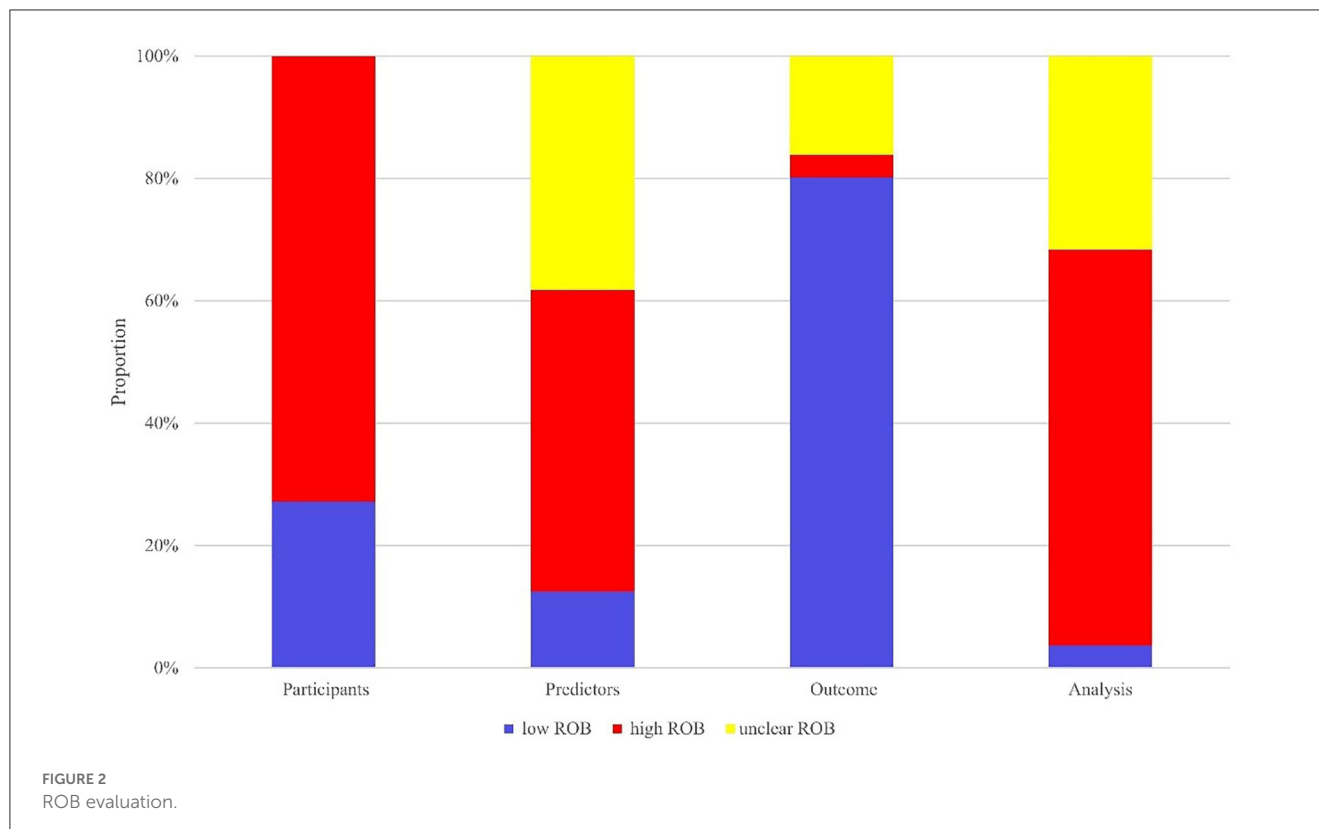
3.4. Variable ordering

The present study summarized and ordered the predictors in included models. All predictors were divided into five categories, which comprehensively covered all-round information of stroke patients. This was helpful for clinicians to provide targeted secondary prevention and health guidance for corresponding patients in the future. Among all predictors, age (patient demographics) was used most frequently in prediction models, followed by initial NIHSS (clinical variables), glucose level (laboratory values), initial Modified Rankin Scale (clinical variables). In terms of the medication history, the frequency of thrombolysis treatment was high (shown in Figure 3).

3.5. Machine learning outcomes

Firstly, a random-effects model was used to combine the C-statistics in ML models. The overall C-statistic for the 96 models in the training set was 0.81 (95% CI: 0.79; 0.83). The 96 models can be classified into 12 types, which were ordered according to their frequency of application to prediction. Among the 12 model types, the LR model was used 29 times in all eligible studies to predict post-stroke motor function, and the C-statistic was 0.81 (95% CI: 0.78; 0.85). ANN had the best performance in prediction with a C-statistic of 0.91 (95% CI: 0.86; 0.95). The overall C-statistic for 71 models in the validation set was 0.82 (95% CI: 0.80; 0.85). The 71 models were divided into 10 types, which were also ranked according to their application frequency in prediction. Among the nine model types, LR models were used most frequently, namely 20 times, to predict post-stroke motor function, with a C-statistic of 0.82 (95% CI: 0.78; 0.87). The performance of each model can be seen in Table 1 and Figure 4.

To minimize the heterogeneity of data sources, the present study also performed C-statistic pooling based on different Modified Rankin Scale cutoff thresholds. Among the training set models with Modified Rankin Scale >2, there were 85 in total with a C-statistic of 0.81 (95% CI: 0.78; 0.84), and among the validation set models, there were 48 in total with a C-statistic of 0.84 (95% CI:



0.81; 0.87). In both the prediction model of training and validation sets, the ANN prediction performed the best with C-statistics of 0.91 (95% CI: 0.86; 0.95) and 0.89 (95% CI: 0.83; 0.96), respectively, as shown in Figure 5. For models with Modified Rankin Scale > 1, there were 6 training set models, with an overall C-statistic of 0.78 (95% CI: 0.77; 0.80). There were 14 validation set models, with an overall C-statistic of 0.79 (95% CI: 0.77; 0.82). For models with Modified Rankin Scale > 3, there were 5 training set models, with an overall C-statistic of 0.83 (95% CI: 0.78; 0.88). There were 4 validation set models, all of which were LR models, with an overall

C-statistic of 0.87 (95% CI: 0.82; 0.92). There were no training set models with Modified Rankin Scale > 4, but there were 5 validation set models, with an overall C-statistic of 0.79 (95% CI: 0.78; 0.81). The performance of each model within the subgroups can be seen in Table 2.

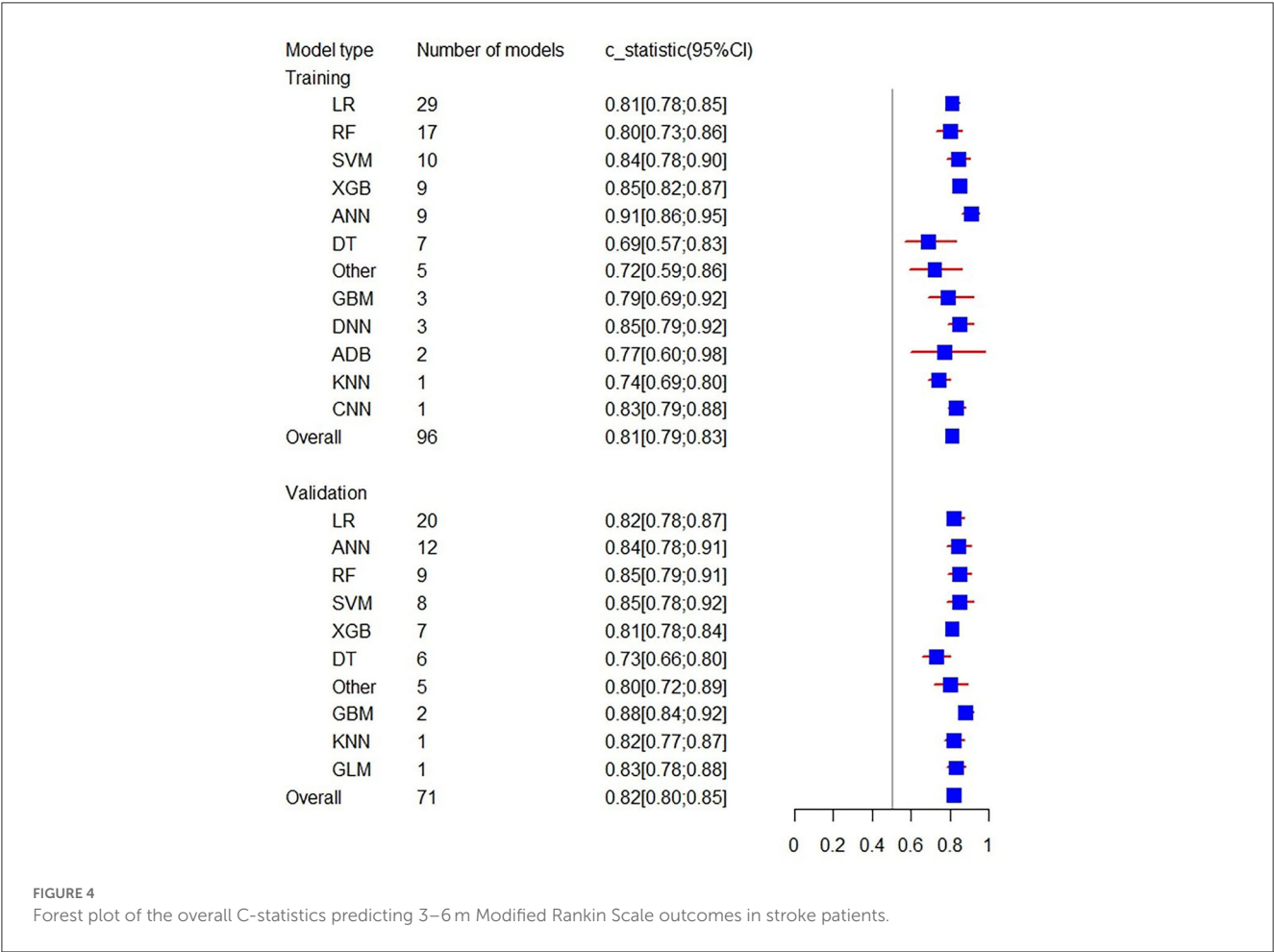
Additionally, the C-statistics were also combined for radiomics-based machine learning prediction models. The training set included a total of 20 prediction models from 8 categories. The overall C-statistic was 0.81 (95% CI: 0.78; 0.84). The most numerous type was the LR model, which also had the best predictive performance with a C-statistic of 0.86 (95% CI: 0.82; 0.91). There were 13 validation set models, with an overall C-statistic of 0.87 (95% CI: 0.83; 0.90). Similarly, the LR model had the best predictive performance with a C-statistic of 0.91 (95% CI: 0.88; 0.95). The performance of the remaining models is provided in Table 3 and Figure 6.

3.6. Sensitivity and specificity

In order to avoid data imbalance, the sensitivity and specificity of models in prediction were analyzed. The training set included 39 models in total with the overall sensitivity of 0.72 (95% CI: 0.70; 0.75). LR were used 15 times, with a sensitivity of 0.74 (95% CI: 0.67; 0.79). Additionally, the overall specificity was 0.77 (95% CI: 0.74; 0.80). The validation set included 40 models, and the overall sensitivity and specificity were 0.74 (95% CI: 0.69; 0.79) and 0.72 (95% CI: 0.66; 0.77), respectively. Sensitivity and specificity analyses were also conducted for subgroups with different Modified Rankin

TABLE 1 Overall C-statistics of machine learning models.

Model	Training			Validation		
	Number of models	Sample size	C-statistics (95%CI)	Number of models	Sample size	C-statistics (95%CI)
LR	29	17,459	0.81 [0.78; 0.85]	20	5,819	0.82 [0.78; 0.87]
RF	17	53,869	0.80 [0.73; 0.86]	9	43,629	0.85 [0.79; 0.91]
SVM	10	49,840	0.84 [0.78; 0.90]	8	43,466	0.85 [0.78; 0.92]
XGB	9	10,150	0.85 [0.82; 0.87]	7	3,138	0.81[0.78; 0.84]
ANN	9	86,701	0.91 [0.86; 0.95]	12	83,384	0.84 [0.78; 0.91]
DT	7	3,136	0.69 [0.57; 0.83]	6	3,342	0.73 [0.66; 0.80]
Other	5	2,972	0.72[0.59; 0.86]	5	1,115	0.80 [0.72; 0.89]
GBM	3	1,778	0.79 [0.69; 0.92]	2	326	0.88 [0.84; 0.92]
DNN	3	5,516	0.85 [0.79; 0.92]	NA	NA	NA
ADB	2	1,113	0.77 [0.60; 0.98]	NA	NA	NA
KNN	1	293	0.74	1	297	0.82
CNN	1	322	0.83	NA	NA	NA
GLM	NA	NA	NA	1	251	0.83
Overall	96	23,3149	0.81 [0.79; 0.83]	71	18,4767	0.82 [0.80; 0.85]



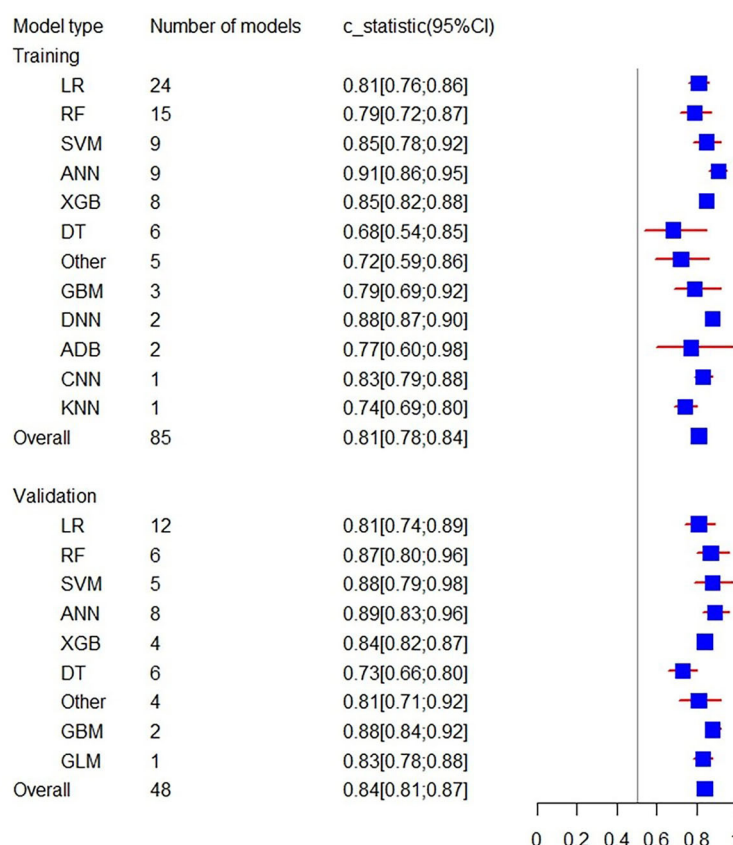


FIGURE 5

Forest plot of the C-statistics predicting 3–6m Modified Rankin Scale >2 outcomes in stroke patients.

Scale thresholds. The sensitivity and specificity of ML models and subgroups were presented in Table 4 and Figure 7.

3.7. Publication bias

In the meta-analysis of C-statistic, no publication bias was found in the funnel plots for both the training set and validation set of ML models for predicting motor function 3–6 months after stroke. The results of Begg test showed that $P = 0.473$ in the training set, $P = 0.909$ in the validation set. The funnel plots are shown in Figures 8, 9. In the meta-analysis of the diagnostic 4-fold table, there was publication bias in the training set of ML models for predicting motor function 3–6 months after stroke, while no publication bias was found in the validation set. The Begg test results showed that $P = 0.01$ in the training set, $P = 0.31$ in the validation set. The funnel plots are shown in Figures 10, 11.

4. Discussion

The study reviews the performance of ML models in the prediction of motor function recovery in patients 3–6 months post-stroke. In the case of Modified Rankin Scale >1, >2, >3, >4, the model's predictive performance was favorable. The C-statistics for

models with predictive factors based on radiomics were 0.81 (95% CI: 0.78; 0.84) in the training set and 0.87 (95% CI: 0.83; 0.90) in the validation set. The overall sensitivity and specificity of the models were both over 0.70 and relatively balanced. The study makes up the gap and deficiency in current researches on the prediction of motor function recovery in stroke patients, which has significant instruction for clinical practice. In the quality assessment part, the ROB of original studies was analyzed in accordance with the PROBAST standard. This provides detailed suggestions for researchers to design model prediction tests in the future, which is conducive to standardization and unification.

According to this analysis of predictors, age and initial NIHSS, are the most critical predictors for the prognosis of motor function in patients with stroke for 90 to 180 days, followed by glucose level, initial Modified Rankin Scale. This is consistent with the results of previous studies (2, 65, 66) on prediction models and meta-analysis of predictors. A large number of studies show that aged patients with acute IS have higher mortality and poorer quality of life than their young counterparts. For instance, both cerebral infarction complicated with pulmonary infection and hemorrhagic transformation after cerebral infarction are more likely to cause relatively poor outcomes in aged patients (67, 68). NIHSS score has been widely recognized as a key determinant of the prognosis in patients with acute IS in China and abroad (69). Furthermore, a previous study (70) on motor function outcome in stroke patients

TABLE 2 C-statistics of subgroups machine learning models (according to threshold of Modified Rankin Scale).

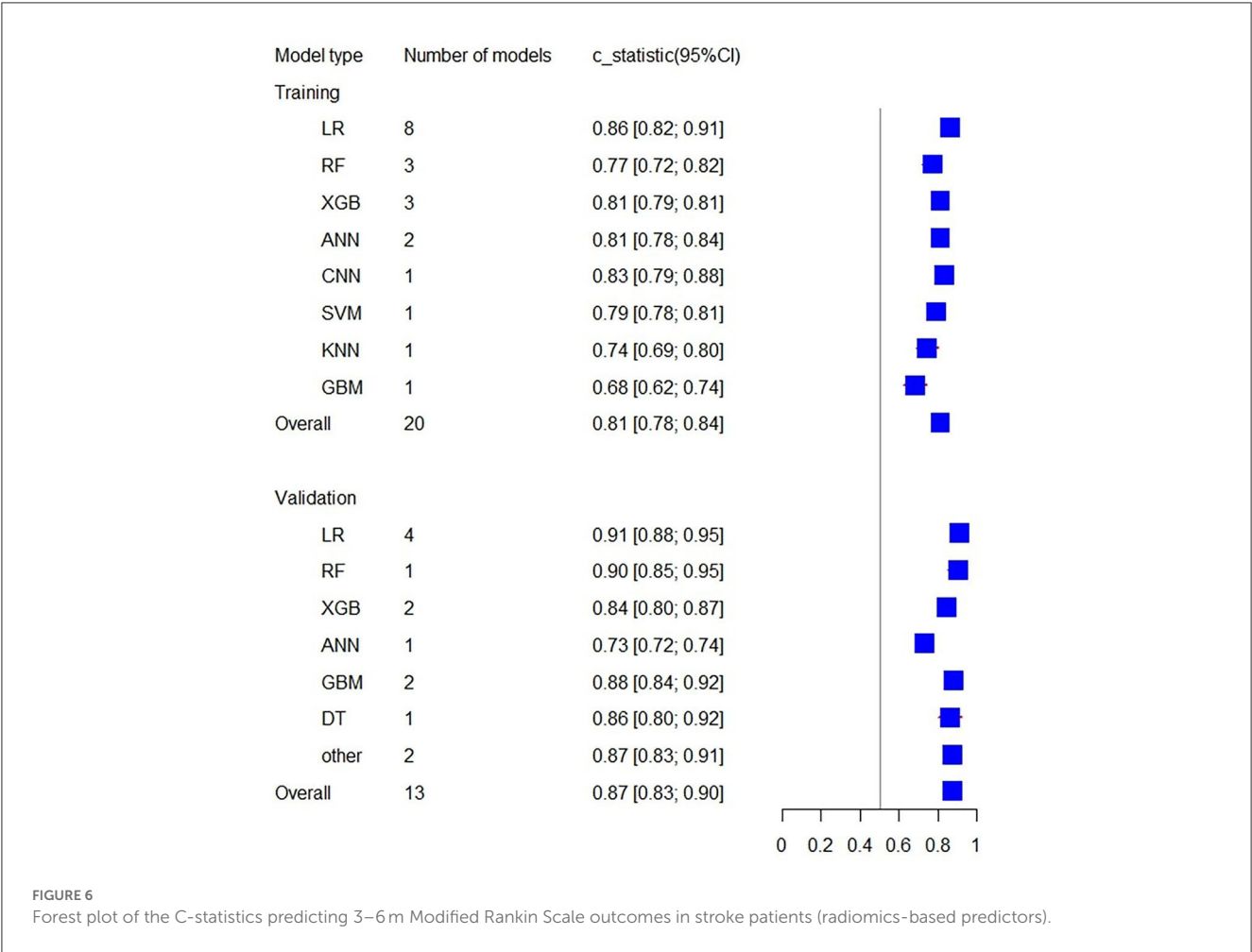
Subgroup	Training			Validation		
	Number of models	Sample size	C-statistics (95%CI)	Number of models	Sample size	C-statistics (95%CI)
Modified Rankin Scale >1						
LR	2	2,012	0.76 [0.74; 0.79]	3	1,148	0.82 [0.76; 0.88]
XGB	1	1,524	0.81	2	678	0.78 [0.72; 0.84]
RF	1	1,524	0.78	2	678	0.79 [0.73; 0.86]
DNN	1	1,524	0.78	NA	NA	NA
SVM	1	1,524	0.77	2	678	0.81 [0.73; 0.90]
KNN	NA	NA	NA	1	297	0.82
ANN	NA	NA	NA	3	443	0.66 [0.53; 0.82]
Other	NA	NA	NA	1	31	0.75
Overall	6	8,108	0.78 [0.77; 0.80]	14	3,953	0.79 [0.77; 0.82]
Modified Rankin Scale >2						
LR	24	13,742	0.81 [0.76; 0.86]	12	2,578	0.81 [0.74; 0.89]
RF	15	52,067	0.79 [0.72; 0.87]	6	41,425	0.87 [0.80; 0.96]
SVM	9	48,316	0.85 [0.78; 0.92]	5	41,262	0.88 [0.79; 0.98]
ANN	9	86,701	0.91 [0.86; 0.95]	8	81,415	0.89 [0.83; 0.96]
XGB	8	8,626	0.85 [0.82; 0.88]	4	934	0.84 [0.82; 0.87]
DT	6	2,858	0.68 [0.54; 0.85]	6	3,342	0.73 [0.66; 0.80]
Other	5	2,972	0.72 [0.59; 0.86]	4	1,084	0.81 [0.71; 0.92]
GBM	3	1,778	0.79 [0.69; 0.92]	2	326	0.88 [0.84; 0.92]
DNN	2	3,992	0.88 [0.87; 0.90]	NA	NA	NA
ADB	2	1,113	0.77 [0.60; 0.98]	NA	NA	NA
CNN	1	322	0.83	NA	NA	NA
KNN	1	293	0.74	NA	NA	NA
GLM	NA	NA	NA	1	251	0.83
Overall	85	222,780	0.81 [0.78; 0.84]	48	172,617	0.84 [0.81; 0.87]
Modified Rankin Scale >3						
LR	3	1,705	0.86 [0.84; 0.89]	4	567	0.87 [0.82; 0.92]
RF	1	278	0.82	NA	NA	NA
DT	1	278	0.75	NA	NA	NA
Overall	5	2,261	0.83 [0.78; 0.88]	4	567	0.87 [0.82; 0.92]
Modified Rankin Scale >4						
ANN	NA	NA	NA	1	1,526	0.81
LR	NA	NA	NA	1	1,526	0.80
RF	NA	NA	NA	1	1,526	0.80
XGB	NA	NA	NA	1	1,526	0.78
SVM	NA	NA	NA	1	1,526	0.77
Overall	NA	NA	NA	5	7,630	0.79 [0.78; 0.81]

show that initial measures were found to be the most significant predictors of upper limb recovery; odds ratio 14.84 (95% CI 9.08–24.25) and 38.62 (95% CI 8.40–177.53) respectively.

AI has been widely applied in the diagnosis, classification, and prediction of stroke. One of its biggest advantages is that it can process data endlessly and can perform faster than traditional

TABLE 3 C-statistics of radiomics-based predictors machine learning models.

Model	Training			Validation		
	Number of models	Sample size	C-statistics (95%CI)	Number of models	Sample size	C-statistics (95%CI)
LR	8	4,055	0.86 [0.82; 0.91]	4	350	0.91 [0.88; 0.95]
RF	3	4,595	0.77 [0.72; 0.82]	1	163	0.90
XGB	3	4,725	0.81 [0.79; 0.81]	2	456	0.84 [0.80; 0.87]
ANN	2	3,251	0.81 [0.78; 0.84]	1	74	0.73
CNN	1	322	0.83	NA	NA	NA
SVM	1	3,001	0.79	NA	NA	NA
KNN	1	293	0.74	NA	NA	NA
GBM	1	293	0.68	2	326	0.88 [0.84; 0.92]
DT	NA	NA	NA	1	163	0.86
Other	NA	NA	NA	2	326	0.87 [0.83; 0.91]
Overall	20	20,535	0.81 [0.78; 0.84]	13	1858	0.87 [0.83; 0.90]



computer-aided detection and diagnosis (CAD) (71). Although there are more and more studies on post-stroke motor function prognosis, there is still a lack of prediction and guidance for neurological motor function, especially for model evaluation after stratifying prediction outcomes. A previous systematic review (72) collected and summarized clinical prognosis trials for patients with large vessel occlusion undergoing thrombectomy, and predicted 90-day Modified Rankin Scale for 802 patients.

TABLE 4 Sensitivity and specificity of overall and subgroup machine learning models.

Subgroup	Training				Validation			
	Number of models	Sample size	Sen (95%CI)	Spe (95%CI)	Number of models	Sample size	Sen (95%CI)	Spe (95%CI)
Model type								
LR	15	10,637	0.74 [0.67;0.79]	0.77 [0.73;0.82]	11	4,204	0.75 [0.69;0.80]	0.72 [0.63;0.80]
RF	6	2,598	0.71 [0.68; 0.74]	0.78 [0.64; 0.88]	5	3,058	0.73 [0.61; 0.83]	0.73 [0.59; 0.83]
SVM	4	4,592	0.70 [0.63;0.76]	0.77 [0.65;0.86]	6	3,189	0.72 [0.59;0.81]	0.76 [0.65;0.85]
Other	3	387	0.75 [0.67;0.81]	0.78 [0.71;0.83]	2	538	0.67 [0.64;0.71]	0.71 [0.68;0.74]
XGB	3	5,022	0.73 [0.70;0.75]	0.77 [0.70;0.83]	5	2,682	0.69 [0.54; 0.80]	0.78 [0.62; 0.88]
ANN	3	3,513	0.76 [0.73; 0.79]	0.81 [0.78; 0.84]	8	2,720	0.72 [0.59; 0.82]	0.72 [0.59; 0.83]
DT	2	724	0.80 [0.65; 0.90]	0.82 [0.67; 0.91]	2	1,014	0.77 [0.74; 0.79]	0.59 [0.57; 0.61]
ADB	1	614	0.73	0.60	NA	NA	NA	NA
CNN	1	322	0.67	0.87	NA	NA	NA	NA
NB	1	150	0.75	0.68	NA	NA	NA	NA
KNN	0	NA	NA	NA	1	297	1.00	0.10
Modified Rankin Scale								
>1	0	NA	NA	NA	13	3,483	0.85 [0.76;0.91]	0.57 [0.46;0.68]
>2	38	27,724	0.72 [0.69;0.75]	0.77 [0.74;0.80]	20	6,302	0.68 [0.62;0.72]	0.79 [0.72;0.84]
>3	1	835	0.75	0.70	2	287	0.84 [0.75;0.90]	0.68 [0.63;0.72]
>4	0	NA	NA	NA	5	7,630	0.65 [0.56;0.73]	0.80 [0.72;0.86]
Overall	39	28,559	0.72 [0.70;0.75]	0.77 [0.74;0.80]	40	17,702	0.74 [0.69;0.79]	0.72 [0.66;0.77]

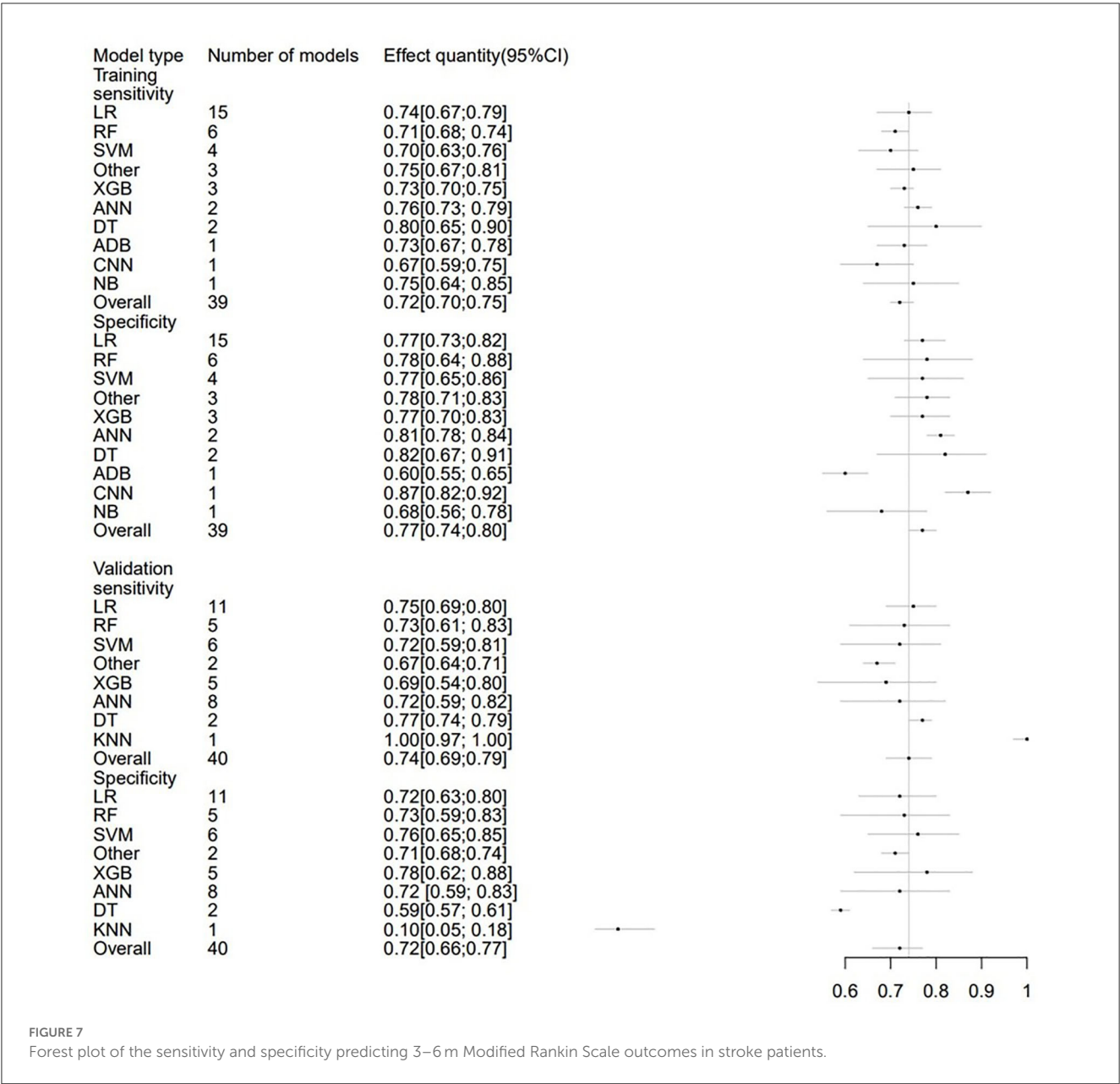


FIGURE 7 Forest plot of the sensitivity and specificity predicting 3–6 m Modified Rankin Scale outcomes in stroke patients.

The random-effects model showed an AUC of 0.846 (95% CI 0.686–0.902), indicating good predictive performance. However, as most of the limited number of included studies used SVM, a comprehensive comparison of the predictive performance of various models to determine the best model was not possible. Similarly, earlier research (73) in rehabilitation medicine on the prediction of post-stroke function recovery confirmed the application of ML prediction ability in clinical settings but did not provide specific C-statistic values, making it difficult to accurately assess predictive performance. Therefore, conducting a meta-analysis of ML motor function prediction models classified by prognosis outcome and predictive factors is both necessary and valuable, as it deepens the understanding of earlier research and further clarifies the ideal application value of ML in predicting post-stroke outcomes.

For stroke prediction, most existing ML algorithms use binary classification to evaluate the outcome indicator. Conventionally, when Modified Rankin Scale are 0–2, functional outcomes are usually defined as “good”; when Modified Rankin Scale 3–6, functional outcomes are typically defined as “poor.” Studies usually measure Modified Rankin Scale at 90 days post-stroke (38, 74, 75). However, with different patient situations in clinical settings and various research objectives, studies have started using Modified Rankin Scale threshold values such as 0–1 VS 2–6, 0–3 VS 4–6, and conducting later follow-ups. New ML algorithms that incorporate these results will provide greater assistance to clinicians.

The advancement of machine learning has made it possible to transform subjective visual interpretation into objective evaluation driven by image data. Radiomics has emerged in this context.

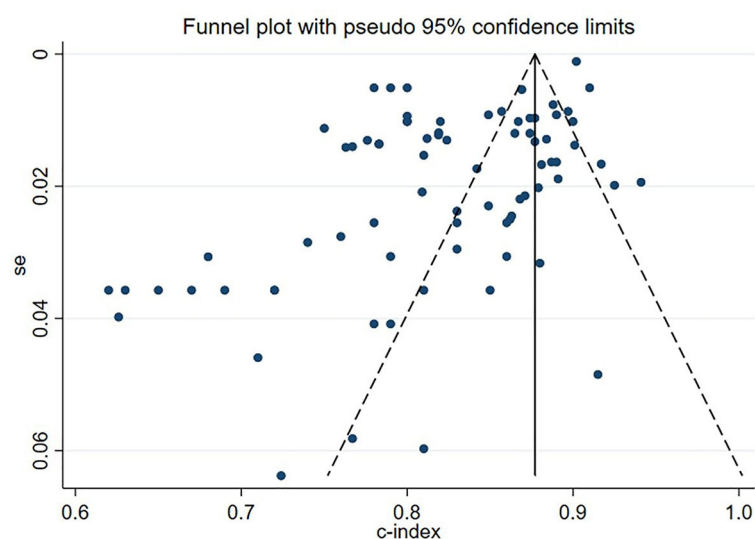


FIGURE 8

Funnel plot for publication bias of C-statistics in the training set.

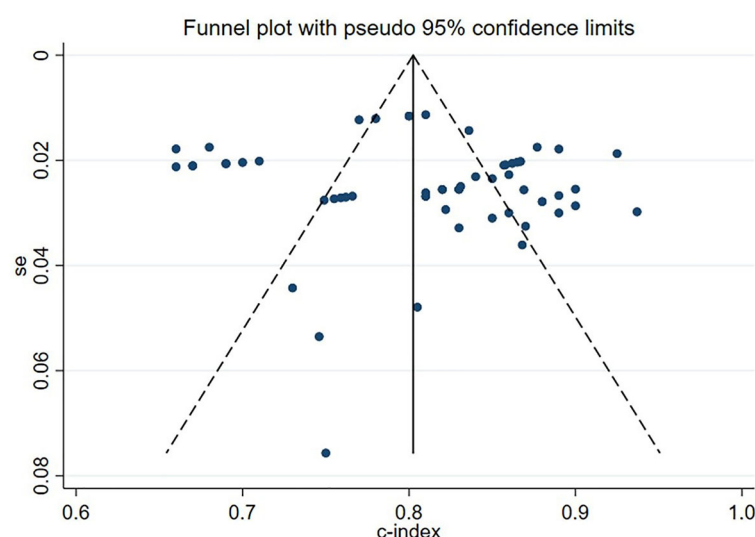


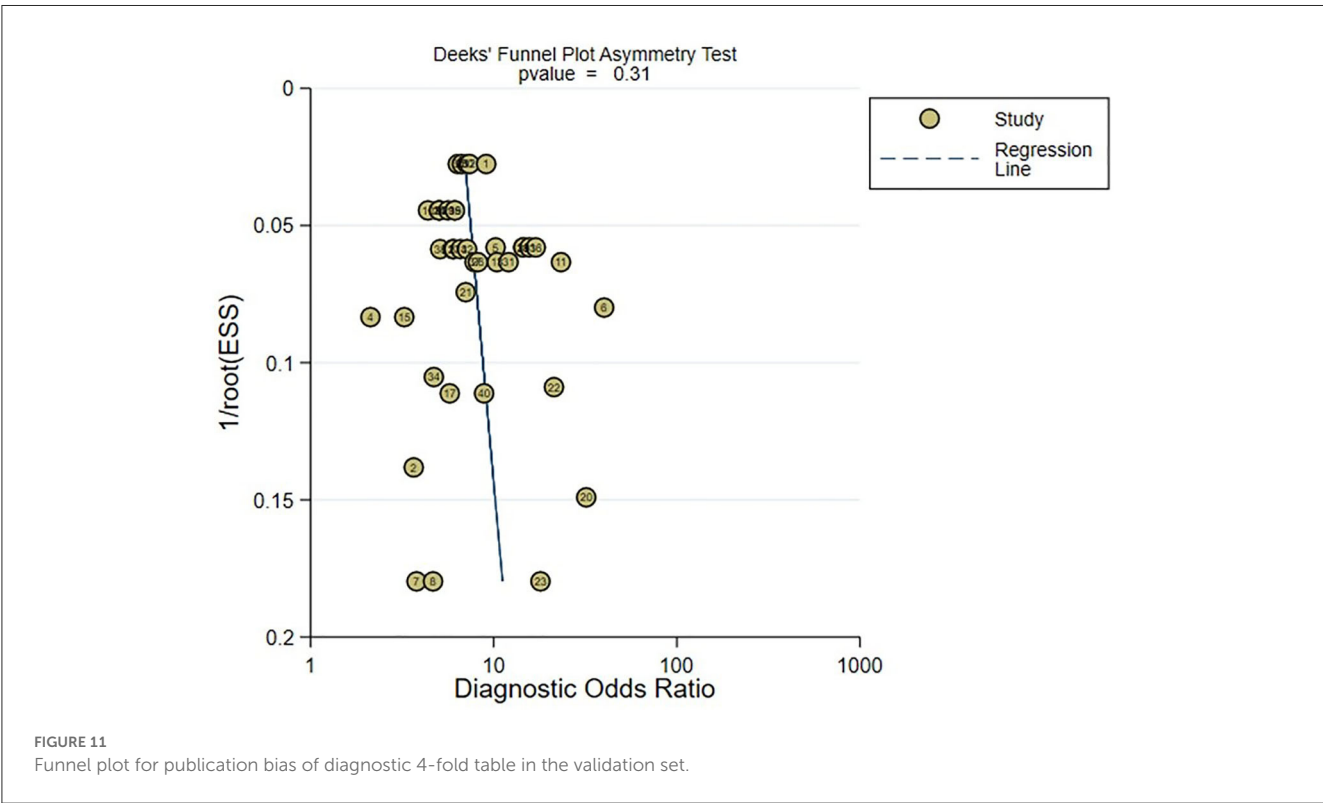
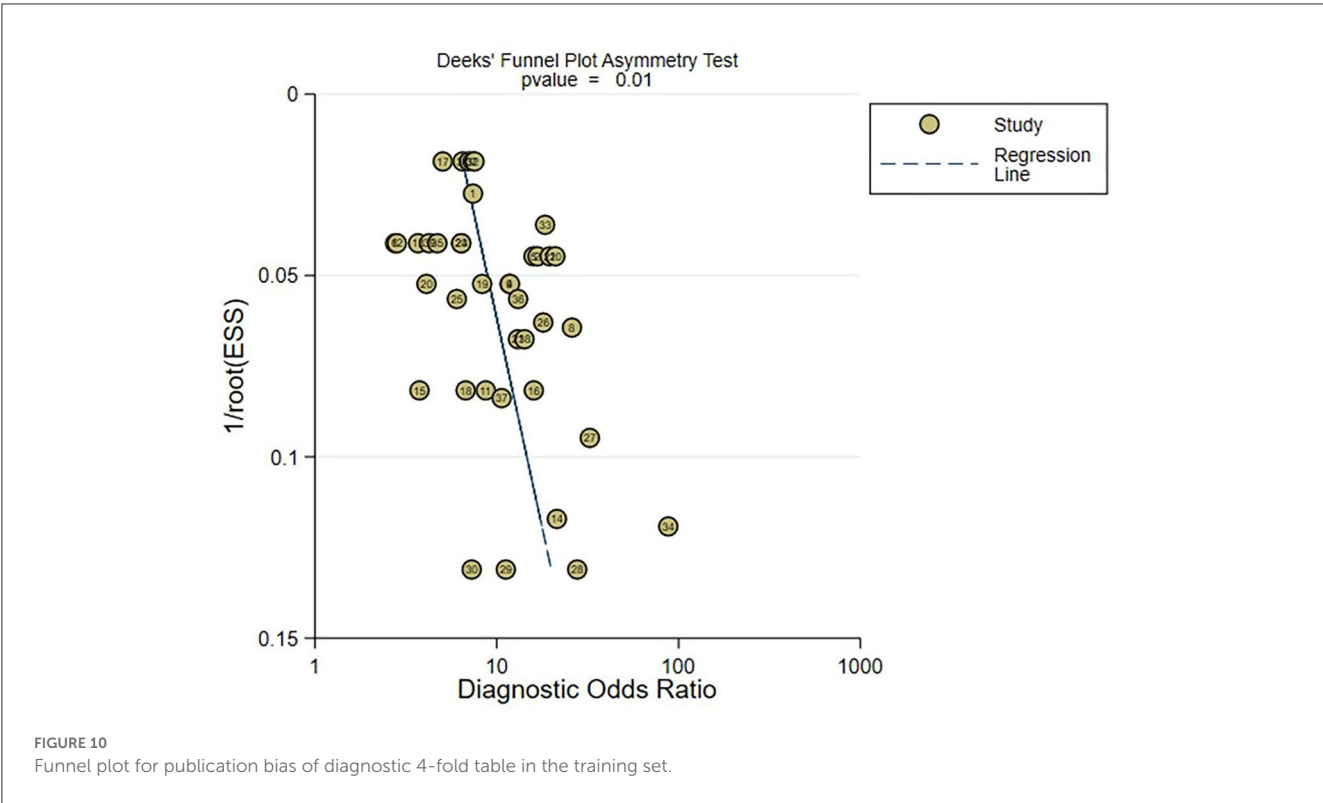
FIGURE 9

Funnel plot for publication bias of C-statistics in the validation set.

Radiomics is a computer-aided process that can extract a large number of quantitative features from biomedical images in an objective, repeatable, and high-throughput manner (76, 77). These features can be combined with other medical information such as demographics, clinical, histological, or genomic data to improve clinical treatment decision-making and accelerate the progress of precision medicine. A systematic review (78) reveals that the artificial intelligence coupling CNN with image feature has greater sensitivity, up to 83%. ML not only offers promising applications in medical imaging by learning information features and patterns from structured input data, but also promotes the emergence of deep learning (DL) and demonstrates its excellent performance in medical image processing (79, 80). XinruiWang's latest study

(81) analyzed ML models to predict the volume of core infarct tissue in AIS patients based on basic CT or MRI imaging at admission. DL models outperformed traditional ML classifiers, with the best performance observed in DL algorithms combined with CT data. Currently, the pooled dice similarity coefficient score of the included ML models for final infarct prediction based on ML was 0.50 (95% CI 0.39–0.61).

In theory, different imaging modalities and parameters provide different diagnostic and prognostic information that can complement each other. Therefore, adjusting and optimizing parameters for multimodal imaging data in radiomics can improve the overall predictive performance. In addition, the study combined clinical data such as clinical symptom assessment, medical history,



and laboratory examinations (82, 83). Multidimensional input information consisting of both imaging and clinical data has the potential to establish better prediction models, which is a direction for future research.

4.1. Limitation

However, there are some limitations to the present study that need to be considered. First, due to the different types of algorithms

and parameter adjustments, there is inevitably a high degree of heterogeneity between studies. To minimize the heterogeneity, we conducted subgroup analysis according to different Modified Rankin Scale cut-off values and analyzed the performance of predictive factors based on different categories. Moreover, from the summary plot of variables, it can be observed that the predictive factors in each study are similar to some degree and selected from five dimensions. Second, although ML has enormous potential in the computing function of huge data, its “black box” characteristic restricts clinicians from trusting the ML prediction. Meanwhile, due to the instability of association between impact factors, ML model requires plenty of samples to improve its accuracy (84, 85). Third, while using the Modified Rankin Scale as a functional outcome measure can directly elucidate the functional levels, it fails to express the details of various post-stroke neurological symptoms, such as dysarthria and pure sensory stroke. At last, from the literature quality assessment summary table, retrospective case studies are in the majority, leading to a high ROB and poor performance of prediction models. Therefore, in order to avoid high ROB, future clinical studies on ML prediction should collect data from clinical registration platforms or prospective clinical studies. Additionally, the design of clinical protocol should meet the requirement of EPV ≥ 20 to ensure the reliability of the results of the prediction model.

5. Conclusion

In this study, we conducted a systematic review and meta-analysis of the current research using ML algorithms to predict post-stroke motor function 3–6 months. Due to its good predictive performance, sensitivity, and specificity, ML can be used as an evaluation tool for predicting motor function after stroke. Additionally, the study found that ML models with radiomics as predictive variables also demonstrated good predictive capabilities. The multidimensional input information consisting of both imaging and clinical data has the potential to establish better prediction models that can guide clinical work.

References

- Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*. (2017) 390:1211–59. doi: 10.1016/S0140-6736(17)32154-2
- Kleindorfer DO, Towfighi A, Chaturvedi S, Cockroft KM, Gutierrez J, Lombardi-Hill D, et al. Guideline for the prevention of stroke in patients with stroke and transient ischemic attack: a guideline from the American heart association/American stroke association. *Stroke*. (2021) 52:e364–467. doi: 10.1161/STR.0000000000000375
- Paolucci S, Iosa M, Coiro B, Venturiero V, Savo A, De Angelis D, et al. Post-stroke depression increases disability more than 15% in ischemic stroke survivors: a case-control study. *Front Neurol*. (2019) 10:926. doi: 10.3389/fneur.2019.00926
- Hung MC, Hsieh CL, Hwang JS, Jeng JS, Wang JD. Estimation of the long-term care needs of stroke patients by integrating functional disability and survival. *PLoS ONE*. (2013) 8:e75605. doi: 10.1371/journal.pone.0075605
- Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. *Psychol Med*. (2019) 49:1426–48. doi: 10.1017/S0033291719000151
- Li B, Feridooni T, Cuen-Ojeda C, Kishibe T, de Mestral C, Mamdani M, et al. Machine learning in vascular surgery: a systematic review and critical appraisal. *NPJ Dig Med*. (2022) 5:7. doi: 10.1038/s41746-021-00552-y
- Windisch P, Koehli C, Rogers S, Schröder C, Förster R, Zwahlen DR, et al. Machine learning for the detection and segmentation of benign tumors of the central nervous system: a systematic review. *Cancers*. (2022) 14:2676. doi: 10.3390/cancers14112676
- Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vas Neurol*. (2017) 2:230–43. doi: 10.1136/svn-2017-000101
- Mainali S, Darsie ME, Smetana KS. Machine learning in action: stroke diagnosis and outcome prediction. *Front Neurol*. (2021) 12:734345. doi: 10.3389/fneur.2021.734345
- Rubbert C, Patil KR, Beseoglu K, Mathys C, May R, Kaschner MG, et al. Prediction of outcome after aneurysmal subarachnoid haemorrhage using data from patient admission. *Eur Radiol*. (2018) 28:4949–58. doi: 10.1007/s00330-018-5505-0

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.

Author contributions

QL wrote the main manuscript and fully participated in all analyses. XL, LC, and LW contributed to the study concept and design. WZ, CJ, XZ, and KZ participated in literature search, data extraction, and quality assessment. All authors have read and approved the final manuscript.

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2023.1039794/full#supplementary-material>

11. 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
12. 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 Dig Med*. (2018) 1:9. doi: 10.1038/s41746-017-0015-z
13. 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
14. 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
15. 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
16. Arslan AK, Colak C, Sarihan ME. Different medical data mining approaches based prediction of ischemic stroke. *Comput Methods Programs Biomed*. (2016) 130:87–92. doi: 10.1016/j.cmpb.2016.03.022
17. Patel UK, Anwar A, Saleem S, Malik P, Rasul B, Patel K, et al. Artificial intelligence as an emerging technology in the current care of neurological disorders. *J Neurol*. (2021) 268:1623–42. doi: 10.1007/s00415-019-09518-3
18. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med*. (2019) 170:W1–w33. doi: 10.7326/M18-1377
19. Debray TP, Damen JA, Riley RD, Snell K, Reitsma JB, Hooft L, et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. *Stat Methods Med Res*. (2019) 28:2768–86. doi: 10.1177/0962280218785504
20. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. (2003) 327:557–60. doi: 10.1136/bmj.327.7414.557
21. Zihni E, Madai VI, Livne M, Galinovic I, Khalil AA, Fiebach JB, et al. Opening the black box of artificial intelligence for clinical decision support: a study predicting stroke outcome. *PLoS ONE*. (2020) 15:e0231166. doi: 10.1371/journal.pone.0231166
22. Xie Y, Jiang B, Gong E, Li Y, Zhu G, Michel P, et al. JOURNAL CLUB: use of gradient boosting machine learning to predict patient outcome in acute ischemic stroke on the basis of imaging, demographic, and clinical information. *AJR Am J Roentgenol*. (2019) 212:44–51. doi: 10.2214/AJR.18.20260
23. 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
24. van Os HJA, Ramos LA, Hilbert A, van Leeuwen M, van Walderveen MAA, Kruijff 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
25. Ramos LA, Kappelhof M, van Os HJA, Chalos V, Van Kranendonk K, Kruijff ND, et al. Predicting poor outcome before endovascular treatment in patients with acute ischemic stroke. *Front Neurol*. (2020) 11:580957. doi: 10.3389/fneur.2020.580957
26. Park D, Jeong E, Kim H, Pyun HW, Kim H, Choi YJ, et al. Machine learning-based three-month outcome prediction in acute ischemic stroke: a single cerebrovascular-specialty hospital study in South Korea. *Diagnostics*. (2021) 11:1909. doi: 10.3390/diagnostics11101909
27. 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
28. 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
29. Nezu T, Hosomi N, Yoshimura K, Kuzume D, Naito H, Aoki S, et al. Predictors of stroke outcome extracted from multivariate linear discriminant analysis or neural network analysis. *J Atheroscler Thromb*. (2022) 29:99–110. doi: 10.5551/jat.59642
30. Moulton E, Valabregue R, Lehericy S, Samson Y, Rosso C. Multivariate prediction of functional outcome using lesion topography characterized by acute diffusion tensor imaging. *NeuroImage Clin*. (2019) 23:101821. doi: 10.1016/j.nicl.2019.101821
31. Lin X, Lin S, Cui X, Zou D, Jiang F, Zhou J, et al. Prediction-driven decision support for patients with mild stroke: a model based on machine learning algorithms. *Front Neurol*. (2021) 12:761092. doi: 10.3389/fneur.2021.761092
32. Lin CH, Hsu KC, Johnson KR, Fann YC, Tsai CH, Sun Y, et al. Evaluation of machine learning methods to stroke outcome prediction using a nationwide disease registry. *Comput Methods Programs Biomed*. (2020) 190:105381. doi: 10.1016/j.cmpb.2020.105381
33. Liang Y, Li Q, Chen P, Xu L, Li J. comparative study of back propagation artificial neural networks and logistic regression model in predicting poor prognosis after acute ischemic stroke. *Open Med*. (2019) 14:324–30. doi: 10.1515/med-2019-0030
34. Li X, Pan X, Jiang C, Wu M, Liu Y, Wang F, et al. Predicting 6-month unfavorable outcome of acute ischemic stroke using machine learning. *Front Neurol*. (2020) 11:539509. doi: 10.3389/fneur.2020.539509
35. Kim C, Lee SH, Lim JS, Kim Y, Jang MU, Oh MS, et al. Impact of 25-hydroxyvitamin D on the prognosis of acute ischemic stroke: machine learning approach. *Front Neurol*. (2020) 11:37. doi: 10.3389/fneur.2020.00037
36. Jiang B, Zhu G, Xie Y, Heit JJ, Chen H, Li Y, et al. Prediction of clinical outcome in patients with large-vessel acute ischemic stroke: performance of machine learning vs. SPAN-100. *AJNR Am J Neuroradiol*. (2021) 42:240–6. doi: 10.3174/ajnr.A6918
37. 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.compbimed.2019.103516
38. 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
39. Hall AN, Weaver B, Liotta E, Maas MB, Faigle R, Mroczek DK, et al. Identifying modifiable predictors of patient outcomes after intracerebral hemorrhage with machine learning. *Neurocrit Care*. (2021) 34:73–84. doi: 10.1007/s12028-020-00982-8
40. Guo R, Zhang R, Liu R, Liu Y, Li H, Ma L, et al. machine learning-based approaches for prediction of patients' functional outcome and mortality after spontaneous intracerebral hemorrhage. *J Personal Med*. (2022) 12:112. doi: 10.3390/jpm12010112
41. Feng X, Hua Y, Zou J, Jia S, Ji J, Xing Y, et al. Intelligible models for healthcare: predicting the probability of 6-month unfavorable outcome in patients with ischemic stroke. *Neuroinformatics*. (2021) 3:6. doi: 10.1007/s12021-021-09535-6
42. Chiu IM, Zeng WH, Cheng CY, Chen SH, Lin CR. Using a multiclass machine learning model to predict the outcome of acute ischemic stroke requiring reperfusion therapy. *Diagnostics*. (2021) 11:80. doi: 10.3390/diagnostics11010080
43. Chiu HW, Huang YT, Cheng CA, editors. Using artificial neural network to predict functional recovery of patients treated by intravenous thrombolysis in acute ischemic stroke. *World Congress on Medical Physics and Biomedical Engineering 2018* Springer. (2019). doi: 10.1007/978-981-10-9035-6_60
44. Chi NF, Chang TH, Lee CY, Wu YW, Shen TA, Chan L, et al. Untargeted metabolomics predicts the functional outcome of ischemic stroke. *J Formos Med Assoc*. (2021) 120:234–41. doi: 10.1016/j.jfma.2020.04.026
45. Bacchi S, Zerner T, Oakden-Rayner L, Kleinig T, Patel S, Jannes J. Deep Learning in the Prediction of Ischaemic Stroke Thrombolysis Functional Outcomes: A Pilot Study. *Acad Radiol*. (2020) 27:e19–23. doi: 10.1016/j.acra.2019.03.015
46. Alawieh A, Zarakat F, Alawieh MB, Chatterjee AR, Spiotta A. Using machine learning to optimize selection of elderly patients for endovascular thrombectomy. *J Neurointerv Surg*. (2019) 11:847–51. doi: 10.1136/neurintsurg-2018-014381
47. Alaka SA, Menon BK, Brobbey A, Williamson T, Goyal M, Demchuk AM, et al. Functional outcome prediction in ischemic stroke: a comparison of machine learning algorithms and regression models. *Front Neurol*. (2020) 11:889. doi: 10.3389/fneur.2020.00889
48. Phan TG, Chen J, Beare R, Ma H, Clissold B, Van Ly J, et al. Classification of different degrees of disability following intracerebral hemorrhage: a decision tree analysis from VISTA-ICH collaboration. *Front Neurol*. (2017) 8:64. doi: 10.3389/fneur.2017.00064
49. Zhang XG, Wang JH, Yang WH, Zhu XQ, Xue J, Li ZZ, et al. Nomogram to predict 3-month unfavorable outcome after thrombectomy for stroke. *BMC Neurol*. (2022) 22:111. doi: 10.1186/s12883-022-02633-1
50. Zhang C, Zhang W, Huang Y, Qiu J, Huang ZX, A. Dynamic nomogram to predict the 3-month unfavorable outcome of patients with acute ischemic stroke. *Risk Manag Healthc Policy*. (2022) 15:923–34. doi: 10.2147/RMHP.S361073
51. Yao Z, Mao C, Ke Z, Xu Y. An explainable machine learning model for predicting the outcome of ischemic stroke after mechanical thrombectomy. *J Neurointerv Surg*. (2022) 4:9598. doi: 10.1136/jnis-2022-019598
52. Moulton E, Valabregue R, Piotin M, Marnat G, Saleme S, Lapergue B, et al. Interpretable deep learning for the prognosis of long-term functional outcome post-stroke using acute diffusion weighted imaging. *J Cereb Blood Flow Metabol*. (2023) 43:198–209. doi: 10.1177/0271678X221129230
53. Ding GY, Xu JH, He JH, Nie ZY. Clinical scoring model based on age, NIHSS, and stroke-history predicts outcome 3 months after acute ischemic stroke. *Front Neurol*. (2022) 13:935150. doi: 10.3389/fneur.2022.935150
54. Zhou Y, Wu D, Yan S, Xie Y, Zhang S, Lv W, et al. Feasibility of a clinical-radiomics model to predict the outcomes of acute ischemic stroke. *Korean J Radiol*. (2022) 23:811–20. doi: 10.3348/kjr.2022.0160
55. Xu Q, Zhu Y, Zhang X, Kong D, Duan S, Guo L, et al. Clinical features and FLAIR radiomics nomogram for predicting functional outcomes after thrombolysis in ischaemic stroke. *Front Neurosci*. (2023) 17:1063391. doi: 10.3389/fnins.2023.1063391
56. Tao Z, Zhou F, Zhang H, Qian M. Value of MRI T2 FLAIR Vascular Hyperintensities combined with DWI ASPECTS in predicting the prognosis of

- acute cerebral infarction with endovascular treatment. *Curr Med Imaging*. (2023) 3:3813. doi: 10.2174/1573405619666230201103813
57. Ramos LA, van Os H, Hilbert A, Olabarriaga SD, van der Lugt A, Roos Y, et al. Combination of radiological and clinical baseline data for outcome prediction of patients with an acute ischemic stroke. *Front Neurol*. (2022) 13:809343. doi: 10.3389/fneur.2022.809343
 58. Ping Z, Min L, Qiuyun L, Xu C, Qingke B. Prognostic nomogram for the outcomes in acute stroke patients with intravenous thrombolysis. *Front Neurosci*. (2022) 16:1017883. doi: 10.3389/fnins.2022.1017883
 59. Li J, Zhu W, Zhou J, Yun W, Li X, Guan Q, et al. A Presurgical unfavorable prediction scale of endovascular treatment for acute ischemic stroke. *Front Aging Neurosci*. (2022) 14:942285. doi: 10.3389/fnagi.2022.942285
 60. Li J, Luo D, Peng F, Kong Q, Liu H, Chen M, et al. ANAID-ICH nomogram for predicting unfavorable outcome after intracerebral hemorrhage. *CNS Neurosci Ther*. (2022) 28:2066–75. doi: 10.1111/cns.13941
 61. Kniep HC, Elsayed S, Nawabi J, Brooks G, Meyer L, Bechstein M, et al. Imaging-based outcome prediction in posterior circulation stroke. *J Neurol*. (2022) 269:3800–9. doi: 10.1007/s00415-022-11010-4
 62. Jabal MS, Joly O, Kallmes D, Harston G, Rabinstein A, Huynh T, et al. Interpretable machine learning modeling for ischemic stroke outcome prediction. *Front Neurol*. (2022) 13:884693. doi: 10.3389/fneur.2022.884693
 63. Huang X, Wang D, Zhang Q, Ma Y, Li S, Zhao H, et al. Development and validation of a clinical-based signature to predict the 90-day functional outcome for spontaneous intracerebral hemorrhage. *Front Aging Neurosci*. (2022) 14:904085. doi: 10.3389/fnagi.2022.904085
 64. Hu J, Fang Z, Lu X, Wang F, Zhang N, Pan W, et al. Influence factors and predictive models for the outcome of patients with ischemic stroke after intravenous thrombolysis: a multicenter retrospective cohort study. *Oxid Med Cell Longev*. (2022) 2022:3363735. doi: 10.1155/2022/3363735
 65. Kwon Y, Norby FL, Jensen PN, Agarwal SK, Soliman EZ, Lip GY, et al. Association of smoking, alcohol, and obesity with cardiovascular death and ischemic stroke in atrial fibrillation: the atherosclerosis risk in communities (ARIC) study and cardiovascular health study (CHS). *PLoS ONE*. (2016) 11:e0147065. doi: 10.1371/journal.pone.0147065
 66. Palaodimou L, Lioutas VA, Lambadiari V, Paraskevas GP, Voumvourakis K, Tsigoulis G. Glycemia management in acute ischemic stroke: current concepts and novel therapeutic targets. *Postgrad Med*. (2019) 131:423–37. doi: 10.1080/00325481.2019.1651206
 67. Roy-O'Reilly M, McCullough LD. Age and sex are critical factors in ischemic stroke pathology. *Endocrinology*. (2018) 159:3120–31. doi: 10.1210/en.2018-00465
 68. Al Khathaami AM, Al Bdah B, Alnosair A, Alturki A, Alrebdi R, Alwayili S, et al. Predictors of poor outcome in embolic stroke of undetermined source. *Neurosciences*. (2019) 24:164–7. doi: 10.17712/nsj.2019.3.20190005
 69. Ortiz GA, Sacco RL, JWEoCT. National institutes of health stroke scale (nihss). *J Physiotherapy*. (2007) 3:1–9. doi: 10.1002/9780471462422.eoct400
 70. Coupar F, Pollock A, Rowe P, Weir C, Langhorne P. Predictors of upper limb recovery after stroke: a systematic review and meta-analysis. *Clin Rehabil*. (2012) 26:291–313. doi: 10.1177/0269215511420305
 71. 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
 72. Teo YH, Lim I, Tseng FS, Teo YN, Kow CS, Ng ZHC, et al. Predicting clinical outcomes in acute ischemic stroke patients undergoing endovascular thrombectomy with machine learning: a systematic review and meta-analysis. *Clin Neuroradiol*. (2021) 31:1121–30. doi: 10.1007/s00062-020-00990-3
 73. Campagnini S, Arienti C, Patrini M, Liuzzi P, Mannini A, Carrozza MC. Machine learning methods for functional recovery prediction and prognosis in post-stroke rehabilitation: a systematic review. *J Neuroeng Rehabil*. (2022) 19:54. doi: 10.1186/s12984-022-01032-4
 74. 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
 75. 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
 76. Avanzo M, Stancanella J, El Naqa I. Beyond imaging: The promise of radiomics. *Physica medica: PM: an international journal devoted to the applications of physics to medicine and biology. J Italian Assoc Biomed Phys (AIFB)*. (2017) 38:122–39. doi: 10.1016/j.ejmp.2017.05.071
 77. Yip SS, Aerts HJ. Applications and limitations of radiomics. *Phys Med Biol*. (2016) 61:R150–66. doi: 10.1088/0031-9155/61/13/R150
 78. 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
 79. Choy G, Khalilzadeh O, Michalski M, Do S, Samir AE, Panykh OS, et al. Current applications and future impact of machine learning in radiology. *Radiology*. (2018) 288:318–28. doi: 10.1148/radiol.2018171820
 80. Chavva IR, Crawford AL, Mazurek MH, Yuen MM, Prabhat AM, Payabvash S, et al. Deep learning applications for acute stroke management. *Ann Neurol*. (2022) 92:574–87. doi: 10.1002/ana.26435
 81. Wang X, Fan Y, Zhang N, Li J, Duan Y, Yang B. Performance of machine learning for tissue outcome prediction in acute ischemic stroke: a systematic review and meta-analysis. *Front Neurol*. (2022) 13:910259. doi: 10.3389/fneur.2022.910259
 82. Robben D, Boers AMM, Marquering HA, Langezaal L, Roos Y, van Oostenbrugge RJ, et al. Prediction of final infarct volume from native CT perfusion and treatment parameters using deep learning. *Med Image Anal*. (2020) 59:101589. doi: 10.1016/j.media.2019.101589
 83. Winder AJ, Siemonsen S, Flottmann F, Thomalla G, Fiehler J, Forkert ND. Technical considerations of multi-parametric tissue outcome prediction methods in acute ischemic stroke patients. *Sci Rep*. (2019) 9:13208. doi: 10.1038/s41598-019-49460-y
 84. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. *PLoS Med*. (2018) 15:e1002689. doi: 10.1371/journal.pmed.1002689
 85. 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



OPEN ACCESS

EDITED BY

Jean-Claude Baron,
University of Cambridge, United Kingdom

REVIEWED BY

Gang Fang,
Guangzhou University, China
Mario Versaci,
Mediterranea University of Reggio Calabria, Italy

*CORRESPONDENCE

Yazhou Wu
✉ asiawu@tmmu.edu.cn

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

RECEIVED 04 February 2023

ACCEPTED 22 May 2023

PUBLISHED 21 June 2023

CITATION

Ye W, Chen X, Li P, Tao Y, Wang Z, Gao C, Cheng J, Li F, Yi D, Wei Z, Yi D and Wu Y (2023) OEDL: an optimized ensemble deep learning method for the prediction of acute ischemic stroke prognoses using union features. *Front. Neurol.* 14:1158555. doi: 10.3389/fneur.2023.1158555

COPYRIGHT

© 2023 Ye, Chen, Li, Tao, Wang, Gao, Cheng, Li, Yi, Wei, Yi and Wu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

OEDL: an optimized ensemble deep learning method for the prediction of acute ischemic stroke prognoses using union features

Wei Ye^{1†}, Xicheng Chen^{1†}, Pengpeng Li^{1†}, Yongjun Tao^{2†}, Zhenyan Wang¹, Chengcheng Gao¹, Jian Cheng³, Fang Li¹, Dali Yi^{1,4}, Zeliang Wei¹, Dong Yi¹ and Yazhou Wu^{1*}

¹Department of Health Statistics, College of Preventive Medicine, Army Medical University, Chongqing, China, ²Department of Neurology, Taizhou Municipal Hospital, Taizhou, Zhejiang, China, ³Department of Radiology, Taizhou Municipal Hospital, Taizhou, Zhejiang, China, ⁴Department of Health Education, College of Preventive Medicine, Army Medical University, Chongqing, China

Background: Early stroke prognosis assessments are critical for decision-making regarding therapeutic intervention. We introduced the concepts of data combination, method integration, and algorithm parallelization, aiming to build an integrated deep learning model based on a combination of clinical and radiomics features and analyze its application value in prognosis prediction.

Methods: The research steps in this study include data source and feature extraction, data processing and feature fusion, model building and optimization, model training, and so on. Using data from 441 stroke patients, clinical and radiomics features were extracted, and feature selection was performed. Clinical, radiomics, and combined features were included to construct predictive models. We applied the concept of deep integration to the joint analysis of multiple deep learning methods, used a metaheuristic algorithm to improve the parameter search efficiency, and finally, developed an acute ischemic stroke (AIS) prognosis prediction method, namely, the optimized ensemble of deep learning (OEDL) method.

Results: Among the clinical features, 17 features passed the correlation check. Among the radiomics features, 19 features were selected. In the comparison of the prediction performance of each method, the OEDL method based on the concept of ensemble optimization had the best classification performance. In the comparison to the predictive performance of each feature, the inclusion of the combined features resulted in better classification performance than that of the clinical and radiomics features. In the comparison to the prediction performance of each balanced method, SMOTEENN, which is based on a hybrid sampling method, achieved the best classification performance than that of the unbalanced, oversampled, and undersampled methods. The OEDL method with combined features and mixed sampling achieved the best classification performance, with 97.89, 95.74, 94.75, 94.03, and 94.35% for Macro-AUC, ACC, Macro-R, Macro-P, and Macro-F1, respectively, and achieved advanced performance in comparison with that of methods in previous studies.

Conclusion: The OEDL approach proposed herein could effectively achieve improved stroke prognosis prediction performance, the effect of using combined data modeling was significantly better than that of single clinical

or radiomics feature models, and the proposed method had a better intervention guidance value. Our approach is beneficial for optimizing the early clinical intervention process and providing the necessary clinical decision support for personalized treatment.

KEYWORDS

MRI, radiomics, deep learning, ensemble learning, metaheuristic algorithms, ischemic stroke

Introduction

In recent years, with the increasingly serious aging phenomenon globally, the incidence of major chronic diseases represented by ischemic strokes has also increased (1). Stroke is still the second leading cause of death in the world and the number one cause of acquired long-term disability, especially in China, which is the greatest challenge of stroke in the world, ranking third among the leading causes of death in China, second only to malignant tumors and heart disease (2). Acute ischemic stroke (AIS) is associated with high morbidity, high mortality, and poor prognoses. They have become a major public health problem that cannot be ignored and have brought a great burden to the economy and society. In the context of limited medical resources, it is necessary to prioritize the implementation of nursing care for patients with poor prognoses, thereby reducing the incidence of disability (3). In the era of precise diagnosis and individualized treatment, prognostic classification has become an important strategy for stroke management (4, 5). The early prediction of prognoses is of great significance for improving the efficiency of stroke disease diagnosis and treatment and improving the levels of disease prevention and control (6).

In the past, prognosis evaluations in clinical practice mostly relied on the manual judgments of physicians, which required high-end medical technology and much physician experience, and the prediction effect of this approach was unstable, which limited its clinical promotion (7, 8). As a new non-invasive technique, radiomics can extract high-throughput quantitative information from traditional medical images, enabling the assessment of internal tumor textures that cannot be captured by visual assessments (9, 10). Radiomics aims to extract quantitative and high-dimensional data from digital biomedical images to facilitate the comprehensive exploration of disease information and progression, and it has been widely used in a variety of clinical fields (11, 12). However, previous studies of this kind were mostly limited to radiomics alone and failed to comprehensively predict disease prognoses with clinical and radiomics features (13–17). At present, there is still a lack of relevant research focusing on the predictive value of combined features for stroke prognosis, which has broad research prospects.

Compared with traditional prediction models, deep learning-based prediction models, represented by deep neural networks (DNNs), long short-term memory recurrent neural networks (LSTM-RNNs), and deep belief networks (DBNs), can automate and accurately analyze a large number of features and are suitable for various medical fields (18–21). Ensemble learning has the

advantages of fast operation and high accuracy and has been widely used in numerous fields, such as medical treatment, healthcare, and information technology (22). Single machine learning and deep learning have the problems of limited convergence effect and difficulty in optimizing hyperparameters, which affects the improvement of prediction efficiency. Deep ensemble learning is expected to solve these problems and improve the accuracy of the model. Compared with shallow learning models and individual learning models, ensemble deep learning models can perform better on multiple learning tasks. They can also extract deeper essential features during the learning process, which can effectively improve the accuracy of the model prediction results (23).

In previous radiomics studies, the applications of deep ensemble models were relatively lacking (24–26). If the selected network structure and parameter settings are not appropriate, this may increase the complexity of the model and reduce its overall operating efficiency (27). Hence, the parameter optimization and layer number setting steps of deep ensemble models are still key issues that need to be solved (28). To improve the optimization accuracy of these models and reduce the time required for the optimization process, such research usually requires the use of metaheuristic algorithms as optimization strategies (29, 30). However, traditional algorithms often have problems such as slow convergence speeds and ease of falling into local optima (31, 32). Research on optimization algorithms with novel optimization mechanisms, accurate solution methods, and robust computing power is still an important direction for feature and parameter selection.

This study aims to build a stroke prognosis prediction model in a deeply integrated way to provide a reference for the diagnosis and prevention of stroke. We adopt an ensemble concept involving data, methods, and algorithms and achieve excellent classification performance. Our contributions and the innovations of this study can be summarized as follows.

(1) In terms of data fusion, we innovatively extract, select, and fuse clinical features and imaging features. The combined data are beneficial to fully extract information and provide early warning for the prognosis of stroke more comprehensively and effectively.

(2) In terms of the categorical outcome, multicategorical outcome variables (normal group, mild group, and moderate-severe group) are used in this study. Compared with that of two-classification approaches, the multi-classification method is conducive to improving the pertinence of the classification, which is conducive to accurate prognosis judgment and intervention guidance.

(3) In terms of model construction, we innovatively construct the optimized ensemble of deep learning (OEDL) method. We comprehensively selected and integrated multiple deep learning methods to maximize the advantages of each method and verified the performance of the model for classification prediction. Our proposed model increases the diversity of prognosis prediction methods, enriches the methodological content of deep ensemble learning, provides new methods and ideas in its research field and clinical decision support for personalized intervention.

(4) For model optimization, we design a new Big Bang optimization algorithm (BBOA), which aims to implement the optimization process efficiently and accurately and then improve the efficiency of the feature selection and parameter search processes.

Materials and methods

The data that support the findings of this study are available from the corresponding author upon reasonable request. This study includes the following steps. (1) Data source and feature extraction: The clinical features and radiomics features are extracted in turn. (2) Data processing and feature fusion: The data filling, data noise reduction, data standardization, data screening, data splicing, data balancing, and related steps are performed. (3) Model construction: Clinical features, imaging features, and combined features from the data are included in turn. In this method, the concept of deep integration is used for modeling, and the base learner and the integration mode are selected in turn. (4) Model optimization: The proposed improved metaheuristic algorithm is used to improve the efficiency of the parameter search. Our technical route is shown in Figure 1.

Data source and feature extraction

This retrospective study was approved by the ethics committee of the Taizhou Municipal Hospital, and the requirement to obtain informed consent was waived. A total of 477 acute ischemic stroke (AIS) patients admitted to the Department of Neurology, Taizhou Municipal Hospital, Zhejiang Province, from January 2020 to April 2021 were recruited. The inclusion criteria were as follows: those who met the AIS diagnostic criteria, had complete clinical data, age more than 18 years. The patient had the first episode, and the MRI images were clear and without artifacts. Severe liver and kidney dysfunction, blood system diseases, malignant tumors, immune system diseases and other diseases, other serious nervous system diseases, and failure to cooperate with clinical treatment or follow-up as required by law were exclusion criteria. A total of 441 cases were eventually included. To ensure that the sample size met the needs of deep learning, we implemented balancing processing for the study data.

The prognosis groupings were based on the National Institute of Health Stroke Scale (NIHSS) at the time of discharge and could be divided into three groups (33, 34): a normal group (<1 point) with 106 cases; a mild group (1–4 points) with 289 cases; and a moderate-severe group (≥ 5 points) with 46 cases. In the following text, we refer to the normal, mild, and model severe groups as

groups A, B, and C, respectively. The NIHSS scores could reflect the degrees of neurological deficit in patients and were used as prognostic indicators in this study.

The clinical data included NIHSS score at admission, disease type, OSCP classification, sex, age, body mass index (BMI), systolic blood pressure (SBP), left ventricular hypertrophy (LVH), homocysteinemia, history of hypertension, history of diabetes, history of coronary heart disease (CHD), history of atrial fibrillation, history of drinking, history of smoking, serum total cholesterol (TC), and low-density lipoprotein (LDL). The distribution of baseline data of each group is shown in Section Results of clinical feature selection of the results. Because of the first onset, relevant characteristics such as “stroke history” were not included in this article.

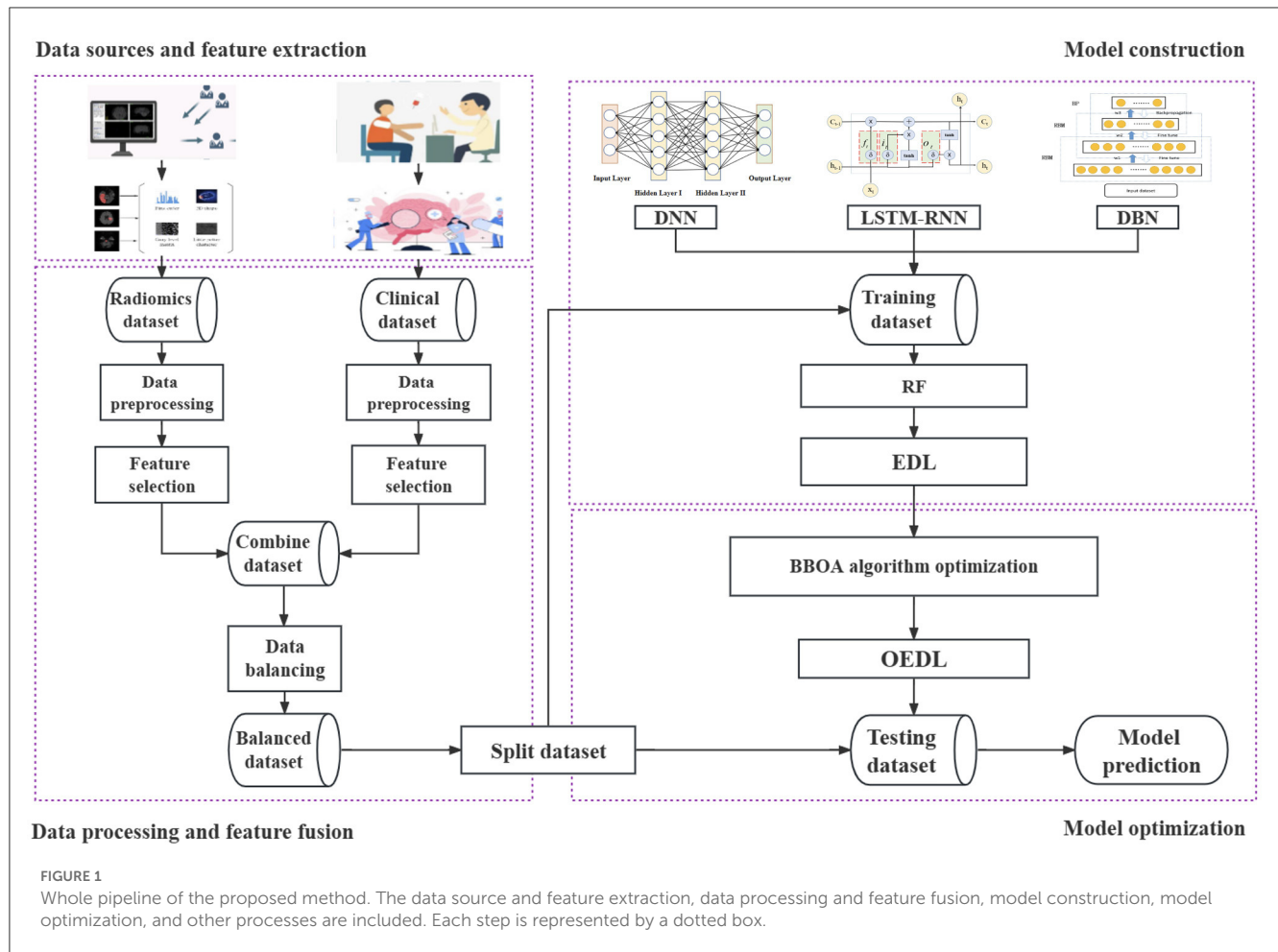
The image data were obtained from cranial MR images, and a Philips Achieve 1.5T scanner was used to obtain these data. The axial DWI sequence was acquired from all patients. To obtain DW images, the following parameters were used: the echo time was 101 ms, the repetition time was 3,211 ms, the number of excitations was 1, the slice thickness was 5 mm, the slice spacing was 1 mm, the acquisition matrix was 230×230 , and the field of vision was $23 \text{ cm} \times 23 \text{ cm}$.

Each patient's first MR image was collected after admission. Two attending physicians independently segmented the regions of interest (ROIs) from the lesions, and ITK-SNAP 3.6.0 software was used for segmentation to obtain the 3D structural data of the lesions. The radiomics features of each annotated lesion were then obtained by using a radiomics analysis tool (the Pyradiomics package). The 2D mask labeling process for each patient is shown in Figure 2. The radiomics features included shape features (14 features), first-order statistics (162 features), gray-level dependence matrix features (GLDM features, 126 features), gray-level cooccurrence matrix features (GLCM features, 216 features), gray-level run length matrix features (GLRLM features, 144 features), gray-level size zone matrix features (GLSZM features, 144 features), and neighboring gray-tone difference matrix features (NGTDM features, 45 features). Finally, 17 clinical features and 851 radiomics features were initially included in this study.

Data processing and feature fusion

After feature extraction, data preprocessing was performed, including data filling, data noise reduction, data standardization, data screening, data splicing, data balancing, and other steps, as shown in Figure 3.

First, in data imputation, we used multiple imputation methods (35). Multiple imputation is a commonly used method to deal with missing values in data. Its basic principle is to generate multiple complete data sets through simulation, and each data set uses different methods to impute missing values. In this study, we use a multiple imputation method based on the multiple Monte Carlo Method to imputation the samples to reduce the impact of missing data on model construction. The multiple Monte Carlo method is a statistical method that uses multiple independent random samples to estimate the expected value. Suppose the expected value $E[f(x)]$



of some function $f(x)$, where $x \in R^d$ is a vector of dimension d . The formula for the multiple Monte Carlo method is as follows:

$$\bar{f}_n = \frac{1}{N} \sum_{i=1}^N f(x_n, i) \quad (1)$$

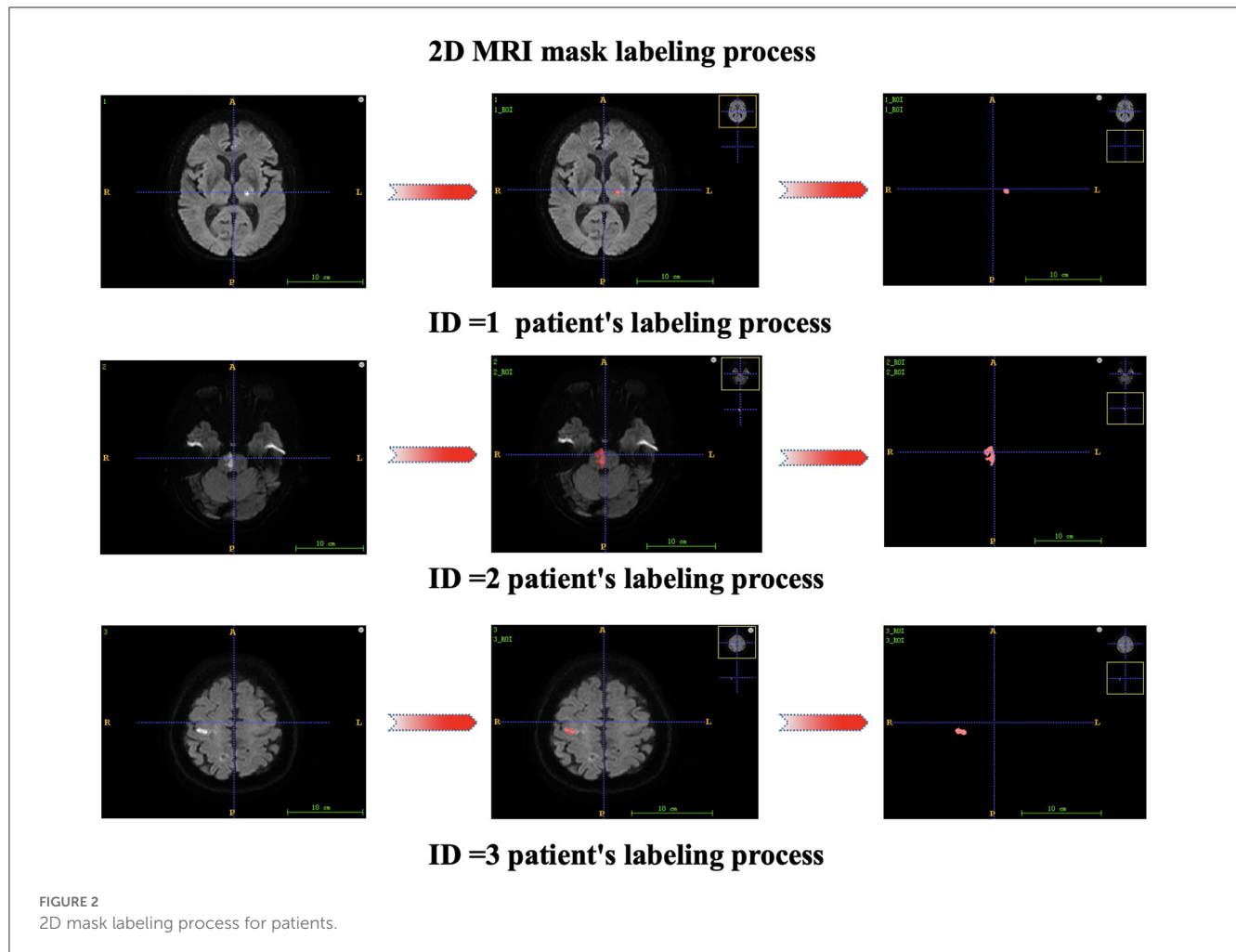
$$\bar{f}_m = \frac{1}{M} \sum_{j=1}^M f_{n(j)} \quad (2)$$

where \bar{f}_n is the sample mean value of $f(x)$ obtained from the n -th sampling, x_n, i is the i -th sample point obtained from the n -th sampling, and N is the number of samples for each sampling; \bar{f}_m is the multiple Monte Carlo estimator obtained by averaging \bar{f}_n obtained from n samples for m times, where $n(j)$ is the sample set used for the j -th sample.

Second, for high-dimensional imaging features, a large number of useless noise features will affect the screening of meaningful features and increase the difficulty of model construction (36). In data denoising, we chose the variance selection filtering method to perform variance-based feature screening (37) and then filtered out features with small differences. The variance of each feature was

calculated, and features with variances greater than the threshold were selected. If the variance is small, it means that there is a small difference between these samples with respect to the feature, and this feature is not conducive to sample discrimination. We filtered features with zero or less variance to preferentially exclude features with lower contributions.

Third, the data were standardized by distinguishing clinical features from radiologic features. (A) Clinical characteristics were assigned to a range between 0 and 1 by one-hot encoding because one-hot encoding can extend the values of discrete features to a Euclidean geometry space and thus fuse standardized imaging features (38, 39). The mathematical formula for one-hot encoding is as follows. Let the value of a discrete feature x with n different values be $\{x_1, x_2, \dots, x_n\}$; then, the one-hot encoding of this feature is an n -dimensional vector v , where if and only if the value of x is x_i , $v[i] = 1$; if x is not x_i , $v[i] = 0$. (B) For radiomics features, we selected a normalization method for feature processing, aiming to eliminate dimensional differences between different features, avoid possible deviation in the model training, improve the convenience of data processing, and speed up the model convergence (40, 41). Normalization helps to ensure dimensional unity between different features, thus improving the robustness and generalizability of the model (42, 43). Normalization refers to scaling the data so that it falls within a specific interval. Standardization (Sz) is the



transformation of data into a normal distribution with a mean of 0 and a standard deviation of 1 (44). Suppose that there are N samples, each sample has n features, and the value of the i -th feature of all N samples is $x_{i1}, x_{i2}, \dots, x_{in}$. Then, the standardized mathematical formula of the feature is as follows:

$$Sz = \frac{x_{ij} - \mu_i}{\sigma_i} \quad (3)$$

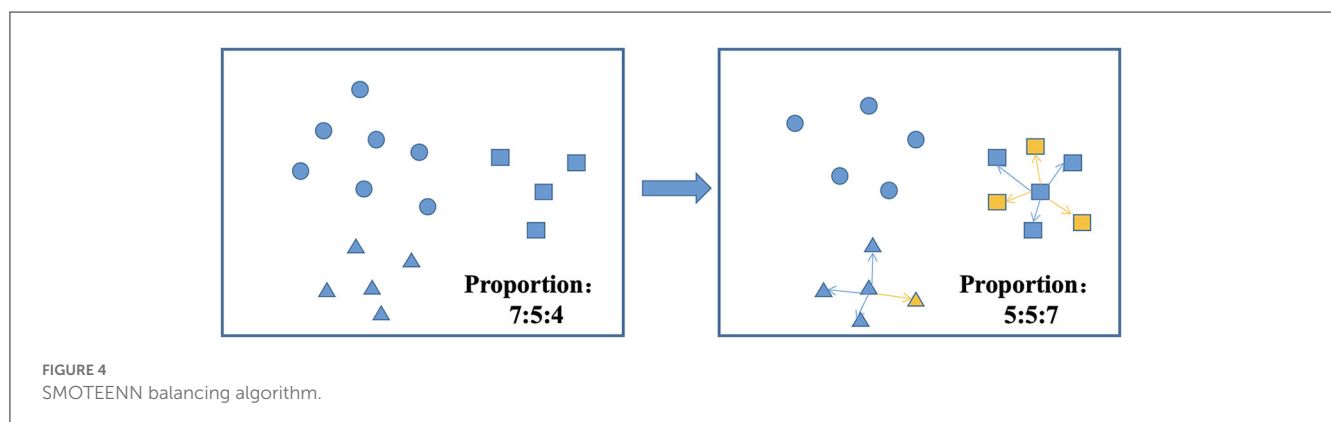
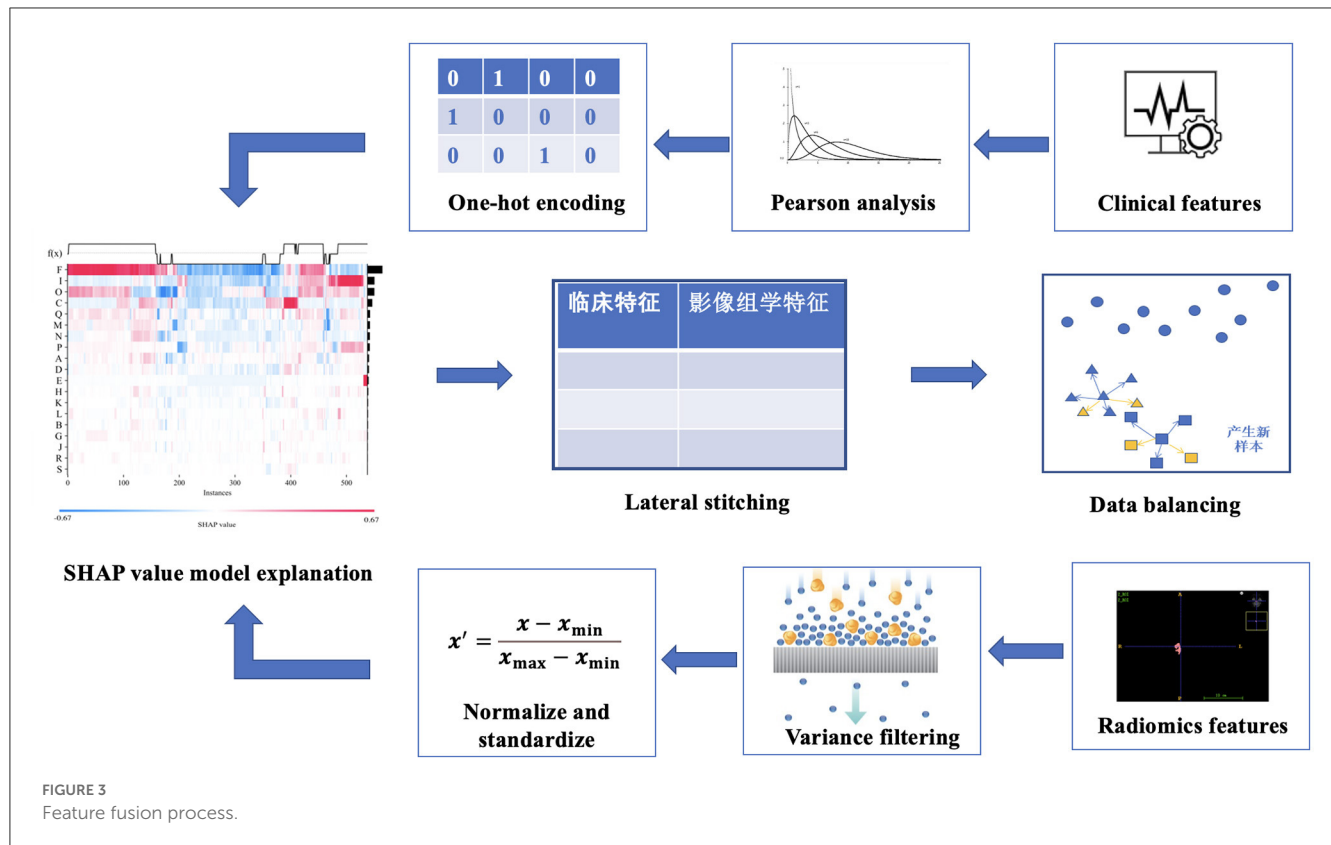
where μ_i denotes the mean of the i -th feature over all N samples and σ_i denotes the standard deviation of the i -th feature over all N samples. For the i -th feature in each sample, a new value can be obtained from this formula, representing the relative size and distribution of the feature across the entire data set.

Fourth, we use the embedded method to filter and reduce the dimensions of the data. The LightGBM and XGBoost algorithms are selected to perform feature importance scoring and selection, the top 50 most important features in terms of weight are screened out, and the features appearing in both methods and the top 10 features in terms of weight in each method are sorted out. (A) LightGBM is a gradient-boosting framework based on a decision tree (DT). It uses a node segmentation strategy based on leaves, seeks the leaf with the largest gain among all the current leaves, and

finally generates a boosted tree (45, 46). The LightGBM algorithm is based on the selection of partition points based on the histogram algorithm and reduces the number of samples and features required in the training and learning processes through two methods, namely, gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB), to maintain high learning performance and reduce the resource occupation in terms of time and space in the training process (47, 48). Let X^s be the input space, s be the feature dimension, and Y be the output space. The given training dataset is $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, where $\vec{x}_i = (\vec{x}_i^{(1)}, \vec{x}_i^{(2)}, \dots, \vec{x}_i^{(s)})$, $i = 1, 2, \dots, n$ represents the input instance and $\{g_1, g_2, \dots, g_n\}$ represents the negative gradient direction of the loss function relative to the model output at each enhancement iteration. Let n represent the number of samples, and let O be the training set of the DT on a node. Then, the information gain $V_{j|o}(d)$ (49) of feature j at node d can be defined as

$$V_{j|o}(d) = \frac{1}{n_0} \left(\frac{(\sum_{x_i \in o: x_{ij} \leq dg_i})^2}{n_{l|o}^j(d)} + \frac{(\sum_{x_i \in o: x_{ij} > dg_i})^2}{n_{r|o}^j(d)} \right) \quad (4)$$

where $n_0 = \sum I(x_i \in o)$, $n_{l|o}^j(d) = \sum I[x_i \in o] : x_{ij} \leq dg_i$, and $n_{r|o}^j(d) = \sum I[x_i \in o : x_{ij} > dg_i]$. (B) The XGBoost algorithm can generate a second-order Taylor expansion of the utilized



loss function and obtain the optimal solution for the regular term outside the loss function (50, 51). The larger the weight of a feature and the more times it is selected by the boosted tree, the more important the feature is considered to be (52, 53). Suppose that the model has t DTs, n represents the total number of samples, f_t represents the t -th regression tree, F represents the collective space of all DTs, and \hat{y}_i^t represents the total predicted value for the i -th sample after adding the outputs of the t DTs. Then, the predicted value of XGBoost (54) can be expressed as

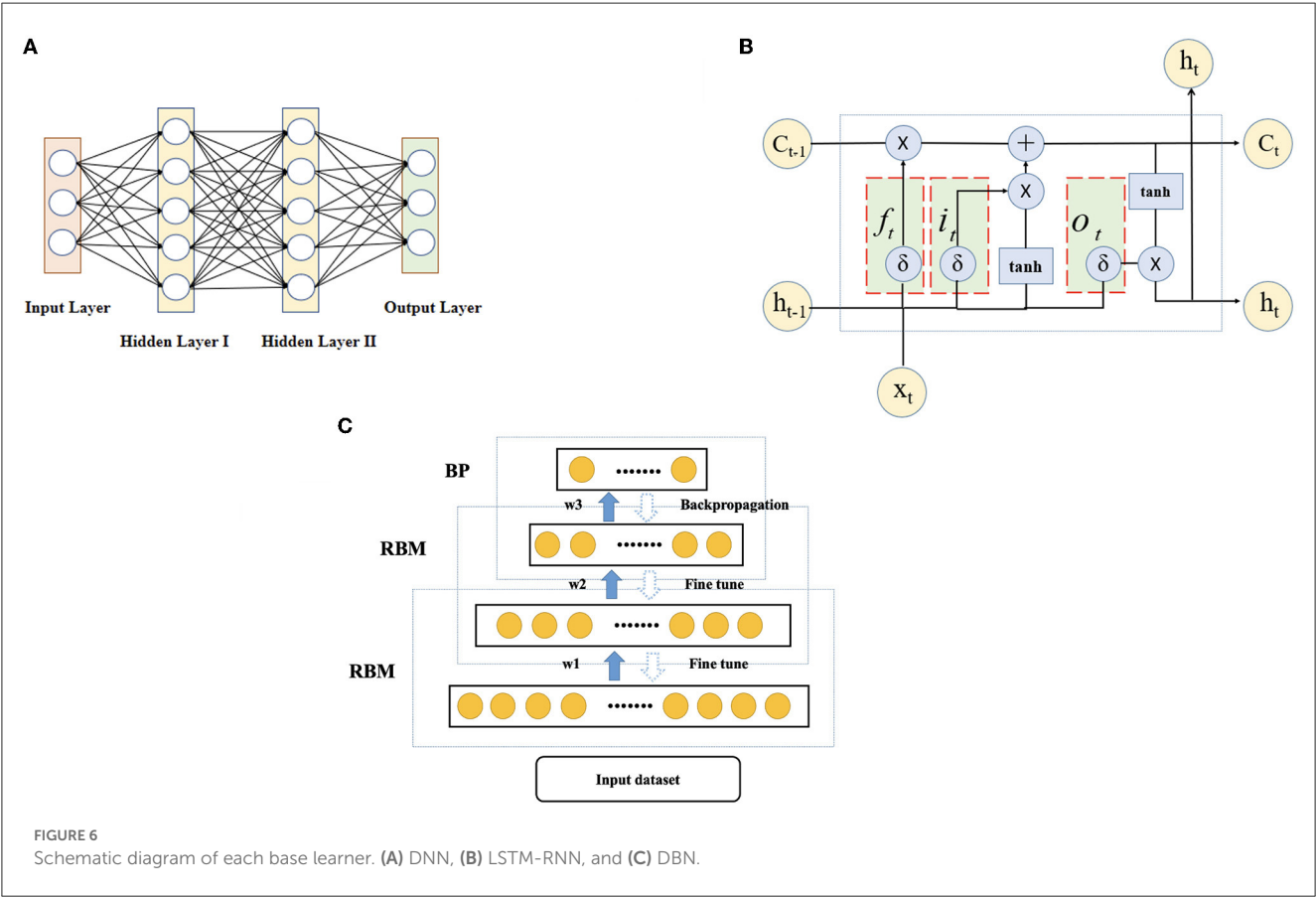
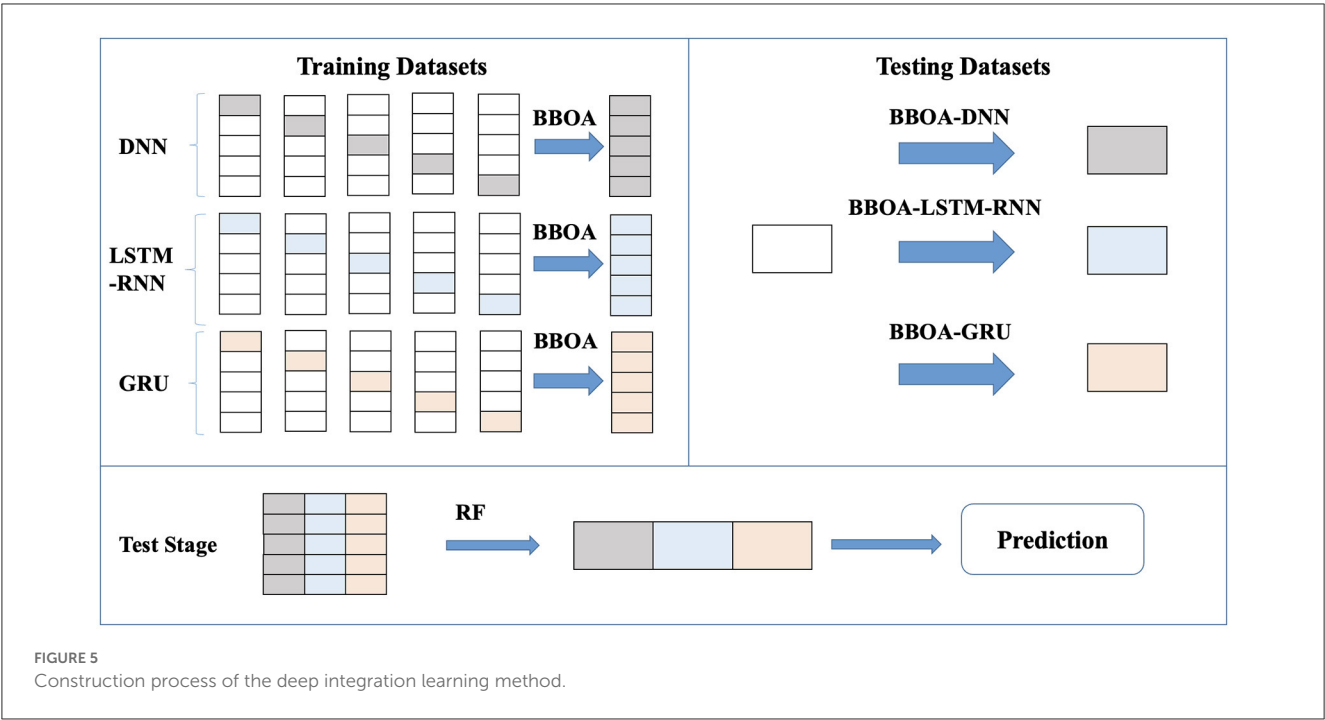
$$\hat{y}_i^{(t)} = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i), f_k \in F, i \in n \quad (5)$$

Its loss function is

$$L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^t \Omega(f_k) \quad (6)$$

$$\Omega(f_k) = \lambda T + \frac{1}{2} \lambda \|w\|^2 \quad (7)$$

where l represents the error between the predicted value and the actual value, T and w represent the number and weight of the leaf nodes, respectively, and γ and λ represent regularization coefficients. The k -th tree is represented by k , and the complexity of k trees is represented by $\sum_{k=1}^t \Omega(f_k)$. (C) The Pearson



correlation coefficient can measure the strength and direction of the linear relationship between two variables, and different correlation coefficients can be selected according to different data characteristics (55). If two features have a high correlation, this indicates that the information contained in the two features is highly similar, and too much similar information can reduce the performance of the chosen algorithm (56). Hence, only one feature must be reserved for features whose correlations are higher than

a certain threshold. To avoid the negative impact of collinearity features on outcome variables, we randomly retained only one of many features with Pearson correlation coefficients greater than the threshold (0.9 in our study). (D) In this study, the SHAP model interpreter tool is used to explain the operation mechanism of the model. SHAP can construct a weighted explanatory model to calculate the contribution of each feature to the results (57, 58). In the interpretation of radiomics and clinical features using LightGBM and XGBoost, respectively, each sample can generate a predictive value, and the SHAP value is expressed as $f(x)$, which can represent the numerical value assigned to each feature in a sample. Red represents features that act positively, and blue represents features that act negatively (5). After the screening of clinical and radiomics features, the combined features were constructed by stitching.

Fifth, there are three common approaches to dealing with class imbalance: undersampling, oversampling, and hybrid sampling techniques. Undersampling techniques include the random undersampling technique, and oversampling techniques include the random oversampling, SMOTE, adaptive synthetic (ADASYN), and borderline-SMOTE techniques (59). SMOTEENN is a method that combines oversampling and undersampling to handle both sample imbalance and noisy data. The SMOTE method increases the number of minority class samples by random oversampling, while the ENN method reduces the number of majority class samples by removing majority class samples. The combination of these two methods can better balance the class distribution in the dataset, thus improving the performance of the classifier (60, 61). The balancing algorithm can balance the number of samples for each classification, thus effectively improving the prediction performance of the model with unbalanced datasets (62, 63). Figure 4 shows the process of the SMOTEENN balancing algorithm, which not only synthesizes new samples for minority classes but also prunes duplicate samples to improve the difference between groups.

Model construction and optimization

The content in this section can be divided into the selection of the base learner, model construction, model optimization, and other steps. The construction process combines the ideas of ensemble learning and deep learning to construct an ensemble of deep learning (EDL) model with a multilayer cascade structure. The optimized ensemble of deep learning (OEDL) model is established by adding an optimization algorithm. The model is built as shown in Figure 5.

First, the selection of base learners is needed. DNN, LSTM-RNN and DBN are used as base learners, and the schematic diagram of each base learner is shown in Figure 6. (A) A deep neural network (DNN) refers to a neural network with more than one hidden layer [64]. The input layer and hidden layer, hidden layer and hidden layer, and hidden layer and output layer all have linear relationships, which can be expressed as

$$y_i = \sigma(\sum w_i^n x_i^n + b_i) \quad (8)$$

where y_i is the next neuron, x_i is a feature or neuron connected to y_i , σ is an activation function in a layer, n is the number of neurons or features connected to the neuron, w_i is a weight coefficient between a feature and a neuron or between neurons, and b is a constant. (B) Long short-term memory (LSTM) is proposed to solve the problem of vanishing or exploding gradients in recurrent neural networks (64, 65). The unit structure records the patient characteristic information of the current state by introducing a new internal state and carries out internal information transmission. First, an input gate i_t , a forget gate f_t , and an output gate o_t are calculated by using the patient characteristic information x_t of the current state and the hidden state h_{t-1} of the last time. Then, the input gate i_t and the forget gate f_t are used to control the retained historical characteristic information and the current state characteristic information of the patient, respectively, to obtain a new C_t . Finally, the input gate o_t is used to transfer the patient characteristic information of the internal state to the hidden state h_t . To achieve the classification effect, an RNN fully connected layer is added behind the LSTM unit to construct an LSTM-RNN to obtain a multi-classification result. (C) Deep belief networks (DBNs) are probabilistic generative models consisting of multiple layers of restricted Boltzmann machines. The main structure combines several layers of RBM and one layer of a BP network and outputs the results by the BP network. The specific steps are as follows. First, the features are trained in each layer of the RBM network separately in an unsupervised manner to ensure that the feature information is reused and retained. Then, the trained features enter the BP network to train the classifier through supervision. Finally, a backpropagation network fine-tunes the training error information direction of each RBM layer so that optimization can be achieved throughout the whole network. The parameters of the DBN are given by w (connection weights), b (visible unit bias), and c (hidden unit bias). The probability of input vector v and output vector h is given by

$$p(v, h) = \frac{e^{-E(v, h)}}{Z} \quad (9)$$

where $-E(v, h)$ is the energy function

$$E(v, h) = -b^T v - c^T h - h^T W v \quad (10)$$

Z is the normalizing factor obtained by summing the numerator over all possible statuses of h and v :

$$Z = \sum_{v, h} -E(v, h) \quad (11)$$

Second, after considering bagging, boosting, stacking, and other methods (see the section Results), we chose the stacking algorithm as the ensemble method in this study, and the model constructed by it is named as the ensemble of deep learning (EDL). Stacking is a method that combines the outputs of multiple base learners according to a certain combination strategy (66, 67). We chose classical and representative deep learning models such as DNN, LSTM-RNN, and DBN as the base learner (68–70) and

```

1: -----Step 1: Input
   Dataset D = {(x1(1), x2(1), ..., y1(1)), (x1(2), x2(2), ..., y2(2)), (x1(n), x2(n), ..., yn(n))};
   Primary learning algorithm :
   PLA = {DNN, LSTM-RNN, DBN};
   Secondary learning algorithm: RandomForest(RF)
3: -----Step 2: Process
1: split D: Train_data, Testing_data
2: for t = 1, 2, ..., T do
3:     h(t) = Stratified Fold (Train_data);
4: end for
5: New_Train_data = ∅;
6: for i in PLA do
7:     for t = 1, 2, ..., T do
8:         Zit = h(t) (PLA(i));
9:     end for
10:    New_Train_data = ∅ ∪ ((Z1i, Z1i, ..., Z1i), yi);
11: end for
12: New_Test_data = ∅;
13: for i in PLA do
14:     for t = 1, 2, ..., T do
15:         Zit = h(t) (PLA(i));
16:     end for
17:    New_Test_data = ∅ ∪ ((Z1i, Z1i, ..., Z1i), yi);
18: end for
19: Training_RF = RF(New_Train_data)
3: -----Step 3: Output
Testing_RF = Training_RF
   (h(1)(New_Test_data(1)), h(2)(New_Test_data(1)), ...,
   h(T)(New_Test_data(1)))

```

Algorithm 1. Stacking pseudocode.

random forest (RF) as the meta-learner. After the training of each base learner was completed, we used the stacking algorithm for analysis; that is, the outputs of multiple base learners were taken as a new dataset that was incorporated into the meta-learner (random forest was selected in this study) for learning and prediction. We integrated the results of the three neural networks and formed probabilities for the three classifications to obtain the final prediction for each sample. The deep learning system was iterated 100 times, and finally, the optimal model was selected by using a greedy strategy. The pseudocode for the stacking algorithm is shown in Algorithm 1.

The innovations of the OEDL method proposed in this study can be reflected in the following aspects. (A) When splitting the training set and the test set, the random stratification method is improved to the label percentage stratification approach to achieve the effect of label balancing. (B) During data selection, we selected clinical features and radiomics features in turn, analyzed clinical features and high-dimensional, abstract radiomics information as combined features, and finally built a combined feature model. (C) We chose the method of deep integration and comprehensively utilized the advantages of each deep learning model to improve its effectiveness and generalization. (D) We innovatively used an improved metaheuristic algorithm (see Section Model training for details) for optimization purposes to ensure the excellence of the classification results.

Third, we performed model optimization based on a newly proposed optimization algorithm. The above EDL proposal combines deep learning and ensemble learning ideas, but there is still the problem of the slow hyperparameter search. To solve this problem, we considered introducing a metaheuristic algorithm. Based on the stacking idea and the framework of particle swarm optimization, we proposed the big bang optimization algorithm (BBOA), which aims to solve the parameter optimization problem in deep networks and applied it to the OEDL method. During the analysis, the algorithm draws on the particle swarm optimization algorithm and the black hole theory of the Big Bang (71), as shown in Figure 7. In the process of constructing the algorithm, we used a sinusoidal chaotic map, an adaptive inertia weight, a greedy strategy, and other optimization methods. The symbol descriptions of the algorithm are shown in Table 1.

(A) The galaxy's initial position can be expressed as

$$x_j^H = U_x + m_{ij}^H * (D_x - U_x) \quad (12)$$

$$x_j^L = U_x + m_{ij}^L * (D_x - U_x) \quad (13)$$

$$m_{ij}^H = a(m_{i-1j}^H)^2 \sin(\pi m_{i-1j}^H) \quad (14)$$

$$m_{ij}^L = a(m_{i-1j}^L)^2 \sin(\pi m_{i-1j}^L) \quad (15)$$

In the study, Formula (14) and Formula (15) add the sinusoidal chaotic map, which acts as an initial randomization to make the distribution range of the star group more dispersed (72). A sinusoidal chaotic map is a non-linear map that can produce chaotic phenomena (73), where a is any constant, and the initial values m_{ij}^H and m_{ij}^L can be any number, but for the depth rule suitable for the deep learning neural network, the initial value is a random integer between 0 and 100, and the chaotic map is rounded to indicate the number of neurons in the deep learning.

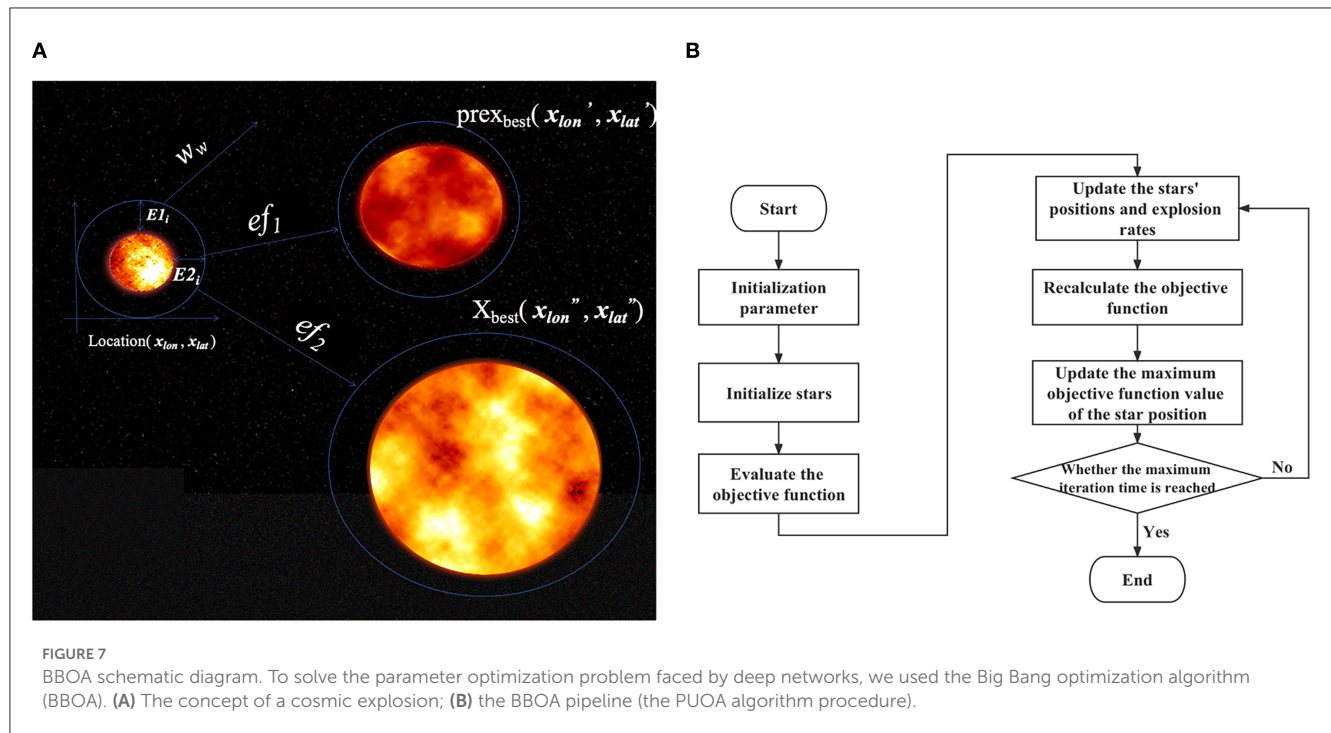
(B) The rates of expansion of the galaxies can be expressed as

$$E_i^H = w_w E_{1i}^{t-1} + ef_1 \cdot r(y_{oi}^t - H_i^t) + ef_2 \cdot r(y_{oi}^t - H_i^t) \quad (16)$$

$$E_i^L = w_w E_{2i}^{t-1} + ef_1 \cdot r(y_{oi}^t - L_i^t) + ef_2 \cdot r(y_{oi}^t - L_i^t) \quad (17)$$

$$w_w = w_s - (w_s - w_e) \frac{t}{T} \quad (18)$$

In the study, Formula (18) is the added adaptive inertia weight, and its function is to regulate the initial expansion speed. Adaptive inertia weights are a variant of inertia weights. Each galaxy should constantly consider its historical and global best position when updating its expansion speed. The adaptive inertia weight can dynamically adjust the value of the inertia weight according to the historical state of the galaxy so that the algorithm converges to the



optimal solution faster. The function of the adaptive inertia weight is to regulate the initial expansion speed.

(C) The transformation of the production expansion center of galaxies affected by higher expansion velocities is expressed as

$$x(H)_i = x(H)_i + E_i^H \quad (19)$$

$$x(L)_i = x(L)_i + E_i^L \quad (20)$$

(D) The optimal solution is the velocity of the largest star in the universe, which can be expressed as

$$Y_o = \begin{cases} y_o, & \text{if } y_o > Y_o; \\ Y_o, & \text{else.} \end{cases} \quad (21)$$

In this study, to update the optimal solution, the greedy strategy is used, which can be expressed as

$$y_o = \max(y_o^t) \quad (22)$$

BBOA was used in OEDL to optimize the number of hidden layers and the number of neurons in each layer of the DNN, LSTM-RNN, and DBN. The stacking integration algorithm was used to integrate the three models after each model was optimized.

Model training

This research was carried out on a Linux workstation equipped with a GPU. The software platform was based on Python

3.7. The proposed algorithms were implemented based on the TensorFlow 2.8 framework. The GPU was used to accelerate the training process. Among the study population, 70% of the data were randomly selected for the training set, and the remaining 30% were used as the test set. Statistical analysis was performed using Python 3.7.0, SPSS 26.0 (SPSS Inc., Chicago, IL, USA).

For various algorithm models, the algorithm was implemented using data split based on a ratio of 7:3 for training and testing. The model was fitted with the training set, the hyperparameters except for the number of hidden layers and the number of neurons were determined by the grid search method, and the best parameter model was selected after 50% cross-validation. Finally, the model was tested with the test set to evaluate the generalizability of each model. The hyperparameters determined by the grid search method in this study were based on a learning rate of 10⁻⁴, a batch size of 20, a momentum term of 0.9, and 1,000 epochs. In addition, the imbalanced distribution of the sample size in each category will lead to the prediction bias of the model. To eliminate this effect, we used the SMOTEENN algorithm to enhance the fused features.

In this study, the implementation of the deep learning network was based on the Keras package in TensorFlow 2.8. The Adam optimizer was used to optimize the gradient of the deep learning model, and the cross-entropy loss function was combined with the softmax activation function to obtain better classification results. The neural network of the three base learners was initially set as a double layer, and the number of neurons in each layer was 10. The BBOA optimizes the numbers of layers and neurons of each base learner to fix the model.

The evaluation indicators in the classification model were obtained based on a confusion matrix, and they include the

TABLE 1 Algorithm process symbol description.

Notation	Meaning
x_j^H	Initial position of the j -th galaxy (horizontal axis)
x_j^L	Initial position of the j -th galaxy (vertical axis)
m_{ij}^H	Initial position of the i -th star in the j -th galaxy (horizontal axis)
m_{ij}^L	Initial position (vertical axis) of the i -th star in the j -th galaxy
U_x	Upper boundary of the galaxy (horizontal axis)
D_x	Lower boundary of the galaxy (vertical axis)
ef_1	The speed of the fastest star in the galaxy
ef_2	The speed of the fastest star in the universe
r	Random number between 0 and 1
E_j^H	Galactic expansion speed (horizontal axis)
E_j^L	Expansion speed of the galaxies (vertical axis)
H_i^t	Time position of the largest star t in the galaxy (horizontal axis)
L_i^t	Time position of the largest star t in the galaxy (vertical axis)
y_{oi}^t	The speed of the fastest star in the galaxy
Y_{oi}^t	The speed of the fastest star in the universe
w_s	The initial inertia of stars in the galaxies
w_e	The final inertia of stars in the galaxies
t	Evolution time (number of iterations)
T	Total evolution time (iterations)
w_w	Adaptive inertia weight
Y_o	Velocity of the largest star in the universe (optimal solution)
y_o	The velocity of the largest star in the galaxy

following four basic indicators: “true positive” (TP) means that the prediction is true and the actual value is also true; “true negative” (TN) means that the prediction is false and the actual value is also false; “false positive” (FP) means that the prediction is true but the actual value is false; “false negative” (FN) means that the prediction is false but the actual value is true. Among multiple classes, each class i has values TP_i , TN_i , FP_i , and FN_i . T_iP_i represents that the true class i is correctly predicted as class i , and F_jP_i represents that the true class j is incorrectly predicted to be class i .

In this study, the evaluation indices include Macro-AUC, accuracy (ACC), macrosensitivity (Macro-R), macrospecificity (Macro-P), and Macro-F1 score (Macro-F1) (74). The ROC curve for each classification was plotted with the true positive rate of each classification as the vertical axis and the false positive rate of each classification as the horizontal axis. The area under the ROC curve of each category is the AUC value of each category, and Macro-AUC is the sum of all types of areas and the average. The value range is $[0 \sim 1]$. The greater the value is, the more accurate the classification. The indicators can be expressed as follows:

$$ACC = \frac{\sum_{i=1}^{n=3} TP_i}{\sum_{i=1}^{n=3} (TP_i + FP_i)} \quad (23)$$

$$macro_PRE = \frac{1}{n} \sum_{i=1}^{n=3} \left(\frac{TP_i}{TP_i + FP_i} \right) \quad (24)$$

$$macro_SEN = \frac{1}{n} \sum_{i=1}^{n=3} \left(\frac{TP_i}{TP_i + FN_i} \right) \quad (25)$$

$$macro_F1score = \frac{2 \cdot macro_PRE \cdot macro_SEN}{macro_PRE + macro_SEN} \quad (26)$$

Results

The results include the results of clinical feature selection, imaging feature selection, comparison of the prediction performance of each method, comparison of the prediction performance of each feature, comparison of the prediction performance of each balanced method, and comparison with previous studies.

Results of clinical feature selection

In the clinical feature selection stage, 17 features were included in the model, and all of these features passed the correlation test (as shown in Figure 8). The clinical characteristics according to the discharge NIHSS classification are presented in Table 2.

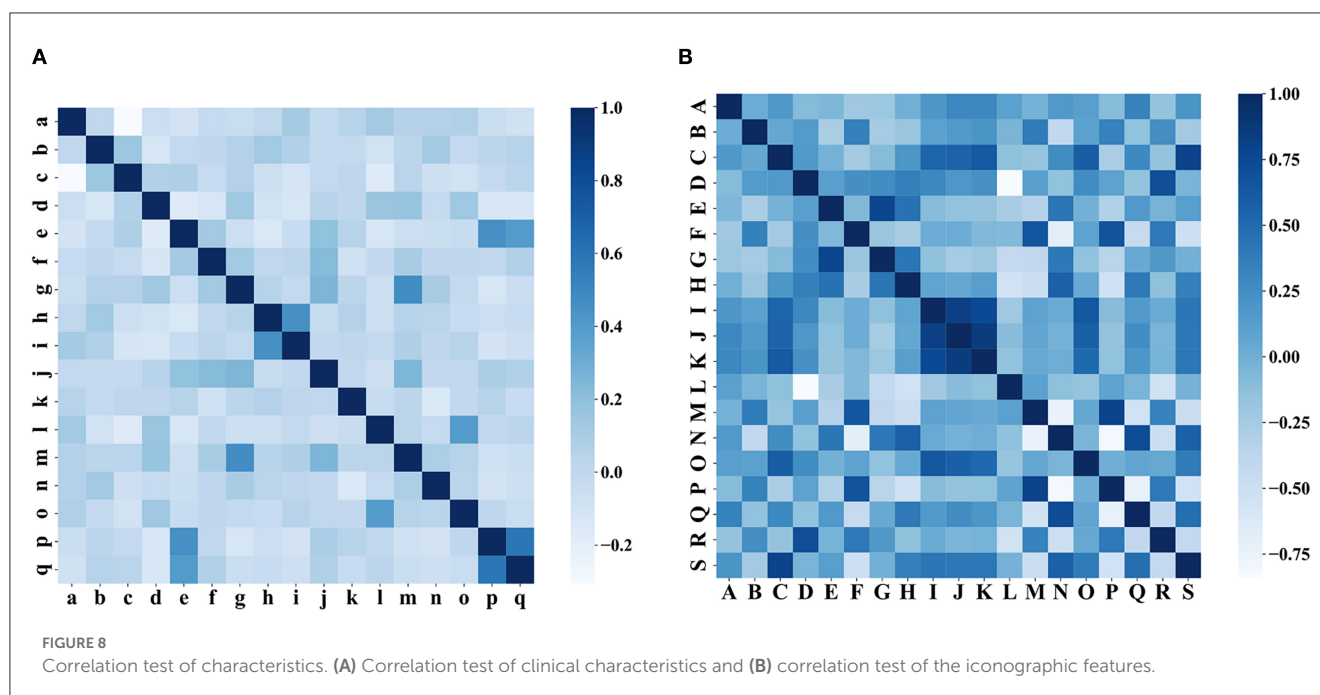
Results of radiomics feature selection

In radiomics feature selection, we first selected 328 features from 851 features using variance selection (threshold 0.3), used the LightGBM and XGBoost algorithms to screen out 81 more important features with the top 50 weights, and sorted out 19 features that appeared in both methods and the top 10 features in their respective methods. After the correlation test, 19 image features were selected. A rose plot was drawn based on the 19 features and their importance weights to the model, as shown in Figure 9.

To visualize the importance of the selected features, we used SHAP to illustrate the degrees to which these features influenced the prediction results, as shown in Figure 10. The SHAP value represents the contribution of each feature to the final prediction and can effectively explain the model prediction for each sample. The feature ranking (y-axis) represents the importance of the prediction model, and the corresponding SHAP value (x-axis) represents the degree of feature influence.

Comparison of the prediction performance of each method

Using the joint dataset as an example, the classification results of these methods are compared. The results show that the OEDL constructed based on the concept of ensemble optimization



obtained the best classification performance, and its Macro-AUC, ACC, Macro-R, Macro-P, and Macro-F1 reached 97.89, 95.74, 94.75, 94.03, and 94.35%, respectively, as shown in Table 3.

Comparison of the prediction performance of each feature

The classification results of EDL and OEDL were compared. The results show that compared with that using the clinical and radiomics features, the method using the combined data had better classification performance, and the EDL method achieved a Macro-AUC of 96.68% and an ACC of 92.55%. The OEDL method achieved a Macro-AUC of 97.89% and an ACC of 95.74%, as shown in Table 4. We also visualized the classification results of the OEDL method with the three features in the form of ROC curves, as shown in Figure 11.

Comparison of the prediction performance of each balanced method

The classification results of the combined features and the OEDL method were compared. Compared with the unbalanced, oversampled, and undersampled techniques, SMOTEENN based on a mixed sampling method achieved the best classification performance, and its Macro-AUC, ACC, Macro-R, Macro-P, and Macro-F1 reached 97.89, 95.74, 94.75, 94.03, and 94.35%, respectively, as shown in Table 5.

Comparison with previous studies

To demonstrate the advanced performance of the method proposed in this study, we reviewed relevant research in the field

of AIS classification and prediction and compared the AUC and ACC of each study. Although the datasets, classification numbers, and other aspects of these studies differed, the differences in the results have some implications for the excellence of the methods. The comparison results show that the proposed method has significantly better classification performance in terms of AUC, ACC, and other aspects than that of previous studies, and it has better classification advantages. For more information, see Table 6.

Discussion

AIS is one of the many diseases that endangers the health of Chinese residents. It is difficult and expensive to check, and it is difficult to evaluate the early prognosis (75). We used joint features to train the OEDL model to predict the prognosis of AIS, which is of great significance to improve the diagnosis and prevention system of AIS and promote the optimal allocation of medical resources.

In terms of data collection and processing, we used clinical features and radiomics features creatively to build joint features, and we built a complete and feasible data processing operation process. Compared with the clinical and imaging feature models, the method using the combined data had better classification performance, and the EDL method achieved a Macro-AUC of 96.68% and an ACC of 92.55%. The OEDL method obtained a Macro-AUC of 97.89% and an ACC of 95.74%. Joint-feature modeling produces better results than single-feature modeling. The reason for the analysis is that simple clinical feature model information is easy to collect, but prediction efficiency is limited due to clinical feature instability; radiomics features can be used to achieve high prediction efficiency, but the inclusion of purely influencing omics features is limited; joint feature modeling can incorporate more comprehensive and objective information. Feature selection is the process of finding the feature subset that

TABLE 2 Clinical characteristics according to the discharge NIHSS category.

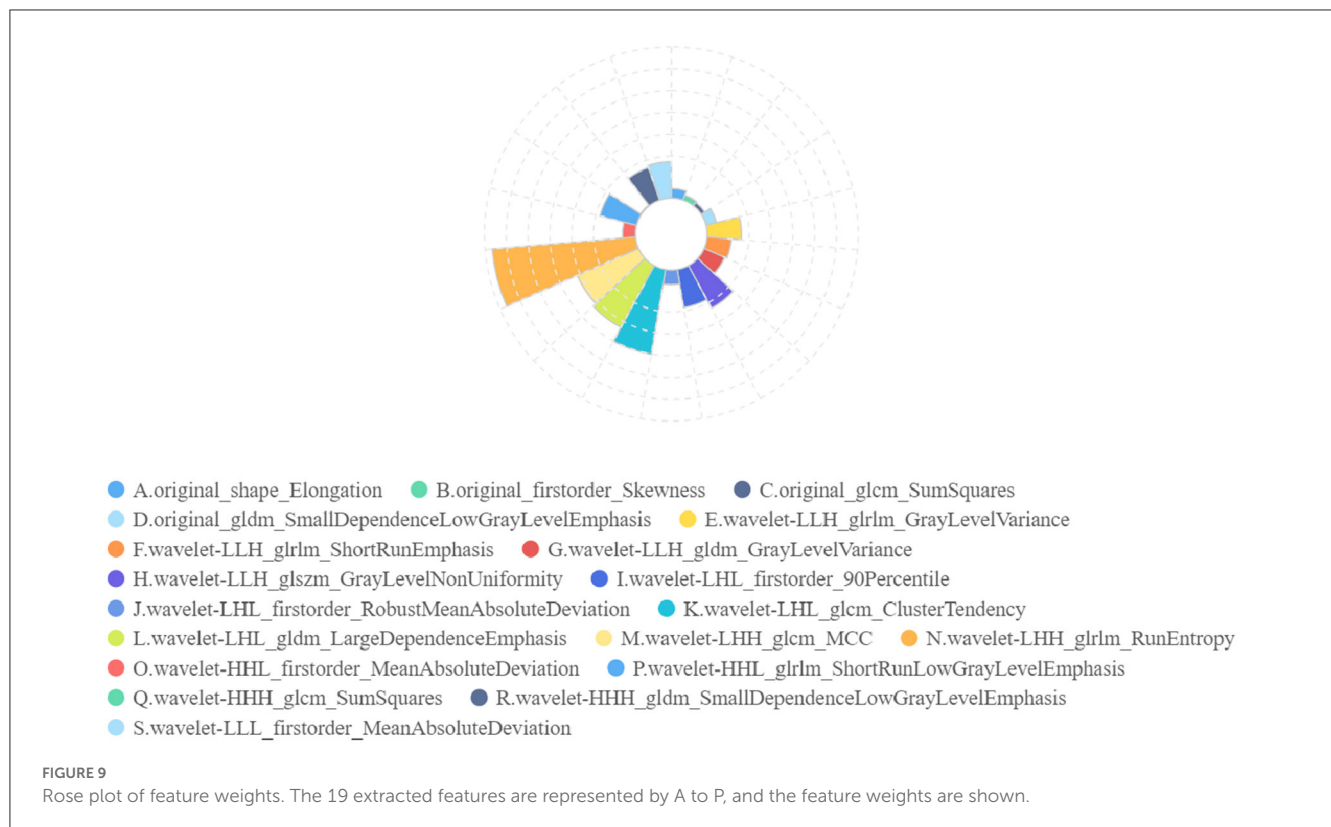
Number	Clinical factors		Group A (n = 106)	Group B (n = 289)	Group C (n = 46)
a	NIHSS on admission		0 (0, 1)	2 (1, 3)	7 (5, 9)
b	Position	Telencephalon	53 (50.00%)	122 (42.21%)	22 (47.83%)
		Diencephalon	15 (14.15%)	91 (31.49%)	15 (32.61%)
		Cerebellum	23 (21.70%)	5 (1.73%)	0 (0.00%)
		Brain stem	15 (14.15%)	71 (24.57%)	9 (19.57%)
c	OCSP typing	TACI	0 (0.00%)	2 (0.69%)	9 (19.57%)
		PACI	22 (20.75%)	116 (40.14%)	26 (56.52%)
		LACI	44 (41.51%)	79 (27.34%)	10 (21.74%)
		POCI	40 (37.74%)	92 (31.83%)	1 (2.17%)
d	Age		67.17 ± 11.63	66.92 ± 12.09	65.63 ± 14.82
e	Sex	Male	66 (62.26%)	167 (57.79%)	24 (52.17%)
		Female	40 (37.74%)	122 (42.21%)	22 (47.83%)
f	BMI		24.29 ± 3.10	24.13 ± 3.94	25.29 ± 4.15
g	SBP		160.79 ± 14.96	162.90 ± 13.97	161.85 ± 18.70
h	TC		4.48 ± 1.05	4.60 ± 1.11	4.38 ± 1.14
i	LDL		2.61 ± 0.92	2.70 ± 1.01	2.64 ± 1.14
j	LVH	Yes	61 (57.55%)	161 (55.71%)	31 (67.39%)
		No	45 (42.45%)	128 (44.29%)	15 (32.61%)
k	HCY	Yes	10 (9.43%)	30 (10.38%)	6 (13.04%)
		No	96 (90.57%)	259 (89.62%)	40 (86.96%)
l	AF	Yes	9 (8.49%)	21 (7.27%)	10 (21.74%)
		No	97 (91.51%)	268 (92.73%)	36 (78.26%)
m	Hypertension	Yes	93 (87.74%)	265 (91.70%)	43 (93.48%)
		No	13 (12.26%)	24 (8.30%)	3 (6.52%)
n	Diabetes	Yes	35 (33.02%)	124 (42.91%)	23 (50.00%)
		No	71 (66.98%)	165 (57.09%)	23 (50.00%)
o	CHD	Yes	6 (5.66%)	17 (5.88%)	6 (13.04%)
		No	100 (94.34%)	272 (94.12%)	40 (86.96%)
p	Smoking	Yes	32 (30.19%)	62 (21.45%)	13 (28.26%)
		No	74 (69.81%)	227 (78.55%)	33 (71.74%)
q	Drinking	Yes	25 (23.58%)	52 (17.99%)	8 (17.39%)
		No	81 (76.42%)	237 (82.01%)	38 (82.61%)

Groups A, B, and C refer to the normal group, mild group, and moderate-severe group, respectively. SBP, systolic blood pressure; TC, total cholesterol; LDL, low-density lipoprotein; LVH, left ventricular hypertrophy; HCY, homocysteinemia; AF, atrial fibrillation; CHD, coronary heart disease; TACI, OCSP classification, total anterior circulation infarcts; PACI, partial anterior circulation infarcts; LACI, lacunar circulation infarcts; and POCI, posterior circulation infarcts.

yields the best model performance, which is conducive to removing redundant features and avoiding the risk of overfitting. Our data collection and processing techniques can be actively promoted in future radiomics research.

Traditional methods for predicting AIS prognosis are shallow and deep machine learning methods. Their ability to represent complex problems is limited, as is their learning ability. To design a new OEDL and apply it to the prediction of AIS prognosis, we creatively combined the ideas of deep learning, integrated

learning, and metaheuristic optimization. A comparison of the prediction performance of the various methods shows that the best classification performance was obtained by OEDL based on ensemble optimization, with Macro-AUC, ACC, Macro-R, Macro-P, and Macro-F1 reaching 97.89, 95.74, 94.75, 94.03, and 94.35%, respectively. The main reasons can be analyzed as follows. (1) In complex problems, the deep learning model can outperform the traditional shallow learning model in terms of feature learning ability. Deep learning multilayer networks can effectively represent



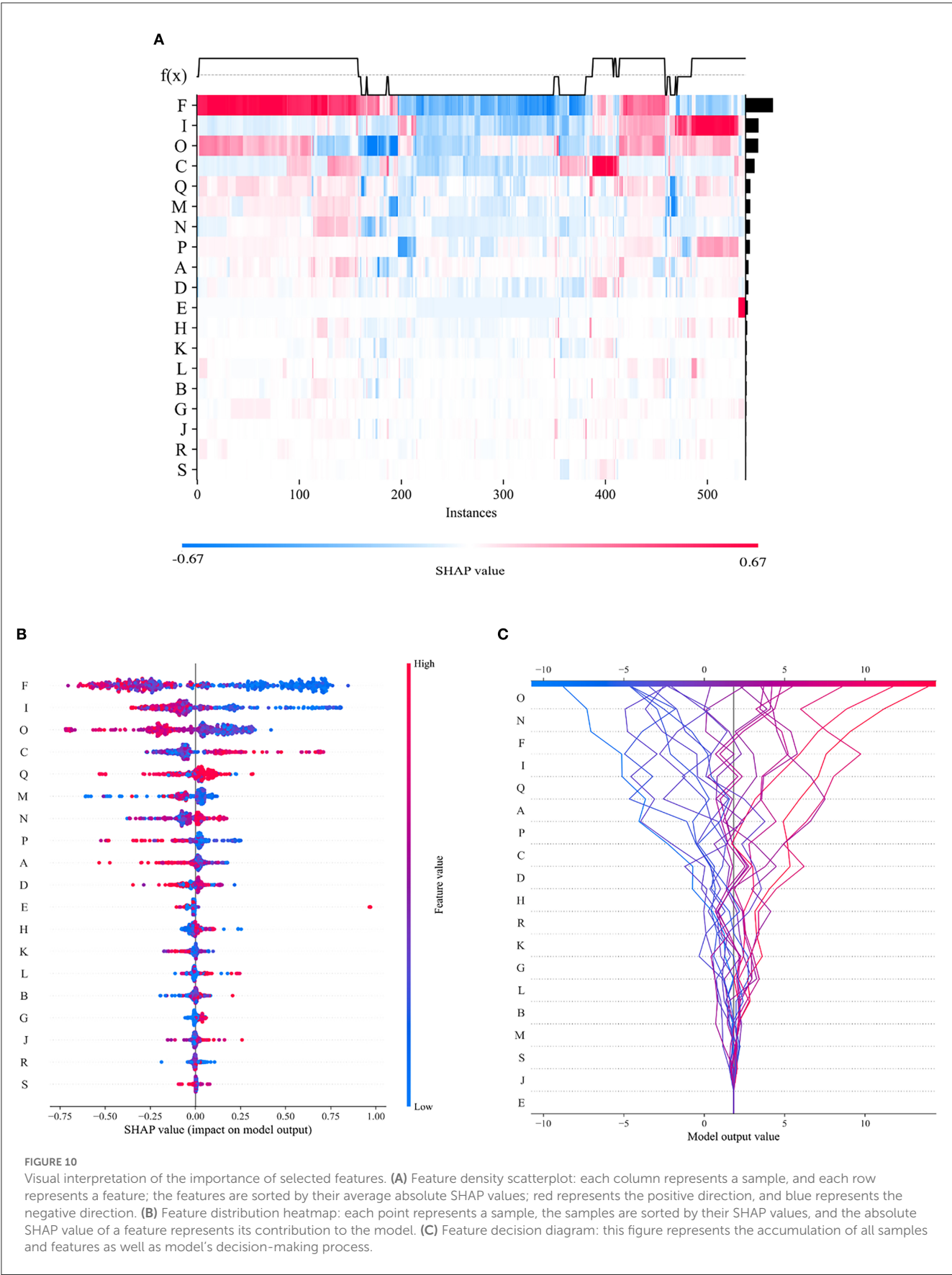
the complexity of prediction results and are adept at discovering complex relationships between a large number of input features, resulting in high prediction performance (76). (2) When compared to a single learner, the advantage of integrated learning is that it ensures classifier diversity and richness, as well as better prediction effect and stability through stacking combination (77). (3) We developed a new parameter optimization strategy based on the traditional metaheuristic algorithm to address the problem of superparameter optimization in machine learning algorithms. Our optimization algorithm can effectively avoid the problem that traditional optimization methods have of falling into a local optimal solution, and it can also effectively improve the model efficiency (78).

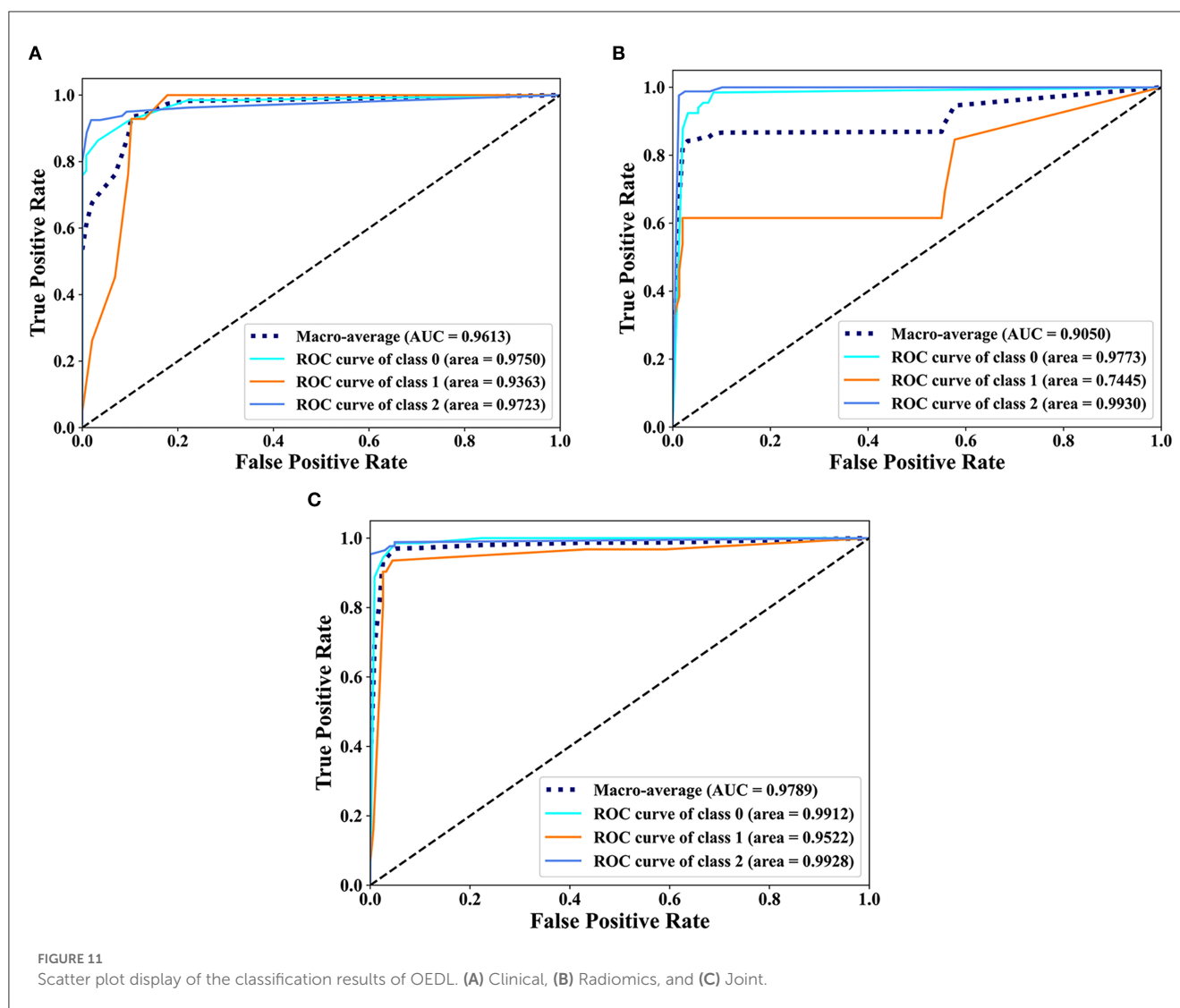
In deep learning, the quality and quantity of data have a crucial impact on the training effect of the model. If the training data are imbalanced, i.e., the number of samples in some classes is too small, then the model will be biased toward those classes with a high proportion during the training process and will perform poorly for those classes with a low proportion. This results in poor model performance with test data may lead to overfitting. Therefore, we introduced a data balancing method to ensure the balance of the training data. The classification results of various balance methods were compared. Compared with the unbalanced, oversampled, and undersampled techniques, SMOTEENN based on mixed sampling can achieve the best classification performance. The results suggest that SMOTEENN, which combines undersampling and oversampling, is the most suitable balancing technique for this study.

A comparison with actual scenarios can be described as follows. Qiu (16) used a linear SVM method to analyze the optimal imaging

group thrombus characteristics of IV protease recanalization with AIS patients on noncontrast CT (NCCT) and CT angiography and obtained (0.85 ± 0.03) ACC in the comparison of actual data. Multiple regression and machine learning models were used by Alaka (17) to predict the related dysfunction of AIS patients after intravascular therapy. Using an internal dataset, the model had an AUC of 0.65–0.72, and using an external dataset, the model had an AUC of 0.66–0.71. Hofmeister (26) investigated the predictive value of radiomics features extracted from clots on the first thrombosis recanalization using SVM, with an ACC of 0.88. Wang (24) obtained an ACC of 0.73 by using the modified Rankin scale (mRS) to predict the prognosis of AIS. Traditional methods for complex problems have limited expression and learning ability, so it is necessary to design a deep integration model with a multilevel cascade structure to improve the model's learning ability in complex problems (35). When compared to that of other single and integrated methods, OEDL can achieve the best classification performance when compared to the control method. The proposed method outperformed previous methods in terms of classification performance (AUC, ACC, etc.) and classification superiority. Furthermore, to address the problem of poor interpretability that frequently exists in deep learning (79), we used interpretable machine learning technology to understand the model's applicability to clinical prediction, with the goal of revealing the reasons behind the prediction results.

In a clinical sense, the combined feature model we developed can serve as a reliable clinical diagnostic tool for predicting stroke prognoses. Our modeling method is more suitable for the clinical model application scenario; it is convenient for radiologists to understand the differences between clinical features,





morphological features, and high-dimensional omics features, as well as diagnostic performance differences. In addition, when building the training set, we also built a data validation set and performed in-model validation at a single center. This study confirmed the validity and scientific nature of the combined data, provided an important reference for similar subsequent studies, and facilitated further verification through the use of more external multi-center data. Compared with traditional radiomics analysis, our combined feature model could extract more statistical features, thereby providing a comprehensive stroke description. In addition, computerized tools overcome the instability of human empirical judgments, allowing clinicians to quickly and accurately predict long-term outcomes.

This study still has room for improvement. First, this study is a single-center, retrospective study with a limited sample size, and it is expected that a multi-center study with larger samples will be implemented in future to further verify the generalizability of the model. Second, lesion labeling comes from manual delineation and may be affected by the subjective judgment of investigators. Subsequent semiautomatic or fully automatic labeling algorithms need to be further explored to

improve the stability and consistency of feature extraction. Third, the radiomics features constructed in this study are based on noncontrast-enhanced MR only, requiring further advanced MR acquisitions such as contrast-enhanced DWI to obtain a high level of evidence for clinical application. Fourth, more efficient image preprocessing tools (80) need to be incorporated to improve the robustness and versatility of the method.

Conclusion

In conclusion, using a combination of clinical features and radiomics, we developed and validated a set of methods for the early prediction of stroke prognoses. We combined DNN ideas with ensemble learning to use OEDL as an effective tool for the early and non-invasive prediction of prognosis levels, thereby optimizing the clinical decision-making process and improving the efficiency of clinical intervention. The ideas in this study can provide new research directions for the effective establishment of stroke prevention and control mechanisms.

TABLE 3 Comparison of the classification effects of different methods (%).

Type	Name	Macro-AUC	ACC	Macro-R	Macro-P	Macro-F1
Machine learning	DT	90.53	75.53	59.62	50.37	54.41
	SVM	97.69	82.45	72.56	82.71	74.11
	RF	95.90	87.77	79.67	85.28	81.14
Deep learning	DNN	93.13	82.96	83.26	83.02	83.12
	LSTM-RNN	94.56	84.81	84.81	84.66	84.63
	DBN	94.39	83.70	83.70	83.61	83.61
Deep learning + Ensemble learning	Hard-voting	95.36	87.23	86.86	87.01	86.90
	Soft-voting	93.45	87.23	86.86	87.10	86.92
	EDL	96.68	92.55	92.10	91.42	91.72
OEDL	OEDL	97.89	95.74	94.75	94.03	94.35

EDL represents a deep ensemble learning model based on DNN, LSTM-RNN, DBN, and stacking ensemble; OEDL is an optimization algorithm based on EDL and BBOA. These bold characters represent the predictive performance of the optimal method.

TABLE 4 Comparison of classification performance of various feature combinations (%).

Feature	Model	Macro-AUC	ACC	Macro-R	Macro-P	Macro-F1
Clinical	EDL	97.15	88.30	87.60	86.78	86.82
Radiomics	EDL	90.79	90.74	74.10	80.28	75.82
Joint	EDL	96.68	92.55	92.10	91.42	91.72
Clinical	OEDL	96.13	90.43	90.57	89.29	89.35
Radiomics	OEDL	90.50	93.21	82.19	86.27	83.87
Joint	OEDL	97.89	95.74	94.75	94.03	94.35

EDL represents a deep ensemble learning model based on DNN, LSTM-RNN, DBN, and stacking ensemble; OEDL is an optimization algorithm based on EDL and BBOA. These bold characters represent the predictive performance of the optimal method.

TABLE 5 Comparison of the classification performance of various balancing methods (%).

Method	Type	Macro-AUC	ACC	Macro-R	Macro-P	Macro-F1
Original	None	92.60	81.16	85.01	74.62	78.75
Random Oversample	Oversampling	91.84	80.37	80.37	81.14	80.49
Random Undersample	Oversampling	82.78	73.81	71.82	74.52	72.47
SMOTE	Undersampling	90.98	85.93	86.10	85.85	85.85
ADASYN	Undersampling	95.74	84.84	85.05	85.35	84.28
Borderline-SMOTE	Undersampling	96.10	87.04	87.12	87.68	86.75
SMOTEENN	Mixed-sampling	97.89	95.74	94.75	94.03	94.35

TABLE 6 Comparison of classification performance with previous studies.

Authors	Data	Methods	Number of categories	AUC (%)	ACC (%)
Qiu (16)	Radiomics	SVM	2	85	-
Alaka (17)	Clinical	LR + Machine Learning	2	66–71	-
Hofmeister (26)	Clinical + Radiomics	SVM	2	88.1	-
Wang (24)	Clinical + Radiomics	LR	2	73	-
Ours	Clinical + Radiomics	OEDL	3	97.89	95.74

These bold characters represent the predictive performance of the optimal method.

Data availability statement

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

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the patients/participants or patients/participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

WY: conceptualization, data curation, resources, methodology, software, formal analysis, validation, investigation, and writing—original draft. XC: methodology, software, formal analysis, validation, investigation, writing—original draft, and editing and polishing. PL: data curation, resources, software, and formal analysis. YT: software, validation, investigation, and data curation. ZW: validation, investigation, data curation, and editing and polishing. CG: investigation, resources, and writing—polishing. JC: validation and resources. FL: formal analysis and validation. DaY and ZW: writing—polishing. DoY: conceptualization, investigation, validation,

and resources. YW: funding acquisition, conceptualization, investigation, resources, methodology, writing—review and editing, project administration, and supervision. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the National Natural Science Foundation of China (Nos. 81872716 and 82173621) and Key Project of Chongqing Graduate Education Reform (yjg192040).

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- 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
- Katan M, Luft A. Global burden of stroke. *Semin Neurol.* (2018) 38:208–11. doi: 10.1055/s-0038-1649503
- Purroy F, Montalà N. Epidemiología del ictus en la última década: revisión sistemática. *Rev Neurol.* (2021) 73:321. doi: 10.33588/rn.7309.2021138
- Pohl M, Hesszenberger D, Kapus K, Meszaros J, Feher A, Varadi I, et al. Ischemic stroke mimics: a comprehensive review. *J Clin Neurosci.* (2021) 93:174–82. doi: 10.1016/j.jocn.2021.09.025
- Karatzetou S, Tsiptsios D, Terzoudi A, Aggeloussis N, Vadikolias K. Transcranial magnetic stimulation implementation on stroke prognosis. *Neurol Sci.* (2022) 43:873–88. doi: 10.1007/s10072-021-05791-1
- Lau L-H, Lew J, Borschmann K, Thijs V, Ekinici EI. Prevalence of diabetes and its effects on stroke outcomes: a meta-analysis and literature review. *J Diabetes Investig.* (2019) 10:780–92. doi: 10.1111/jdi.12932
- Vasudeva K, Munshi A. miRNA dysregulation in ischaemic stroke: focus on diagnosis, prognosis, therapeutic and protective biomarkers. *Eur J Neurosci.* (2020) 52:3610–27. doi: 10.1111/ejn.14695
- Tang G, Cao Z, Luo Y, Wu S, Sun X. Prognosis associated with asymptomatic intracranial hemorrhage after acute ischemic stroke: a systematic review and meta-analysis. *J Neurol.* (2022) 269:3470–81. doi: 10.1007/s00415-022-11046-6
- Harding-Theobald E, Louissaint J, Maraj B, Cuaresma E, Townsend W, Mendiratta-Lala M, et al. Systematic review: radiomics for the diagnosis and prognosis of hepatocellular carcinoma. *Aliment Pharmacol Ther.* (2021) 54:890–901. doi: 10.1111/apt.16563
- Sotoudeh H, Sarraimi AH, Roberson GH, Shafaat O, Sadaatpour Z, Rezaei A, et al. Emerging applications of radiomics in neurological disorders: a review. *Cureus.* (2021) 13:e20080. doi: 10.7759/cureus.20080
- Wang X, Xie T, Luo J, Zhou Z, Yu X, Guo X. Radiomics predicts the prognosis of patients with locally advanced breast cancer by reflecting the heterogeneity of tumor cells and the tumor microenvironment. *Breast Cancer Res.* (2022) 24:20. doi: 10.1186/s13058-022-01516-0
- Chen Q, Xia T, Zhang M, Xia N, Liu J, Yang Y. Radiomics in stroke neuroimaging: techniques, applications, and challenges. *Aging Dis.* (2021) 12:143. doi: 10.14336/AD.2020.0421
- Hatami N, Cho T-H, Mechtouff L, Eker OF, Rousseau D, Frindel C. CNN-LSTM based multimodal MRI and clinical data fusion for predicting functional outcome in stroke patients. 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Glasgow, Scotland, United Kingdom: IEEE (2022). p. 3430–4. doi: 10.1109/EMBC48229.2022.9871735
- Sarioglu O, Sarioglu FC, Capar AE, Sokmez DF, Mete BD, Belet U. Clot-based radiomics features predict first pass effect in acute ischemic stroke. *Interv Neuroradiol.* (2022) 28:160–8. doi: 10.1177/15910199211019176
- Wen X, Li Y, He X, Xu Y, Shu Z, Hu X, et al. Prediction of malignant acute middle cerebral artery infarction via computed tomography radiomics. *Front Neurosci.* (2020) 14:708. doi: 10.3389/fnins.2020.00708
- Qiu W, Kuang H, Nair J, Assis Z, Najm M, McDougall C, et al. Radiomics-based intracranial thrombus features on CT and CTA predict recanalization with intravenous alteplase in patients with acute ischemic stroke. *AJNR Am J Neuroradiol.* (2019) 40:39–44. doi: 10.3174/ajnr.A5918
- Alaka SA, Menon BK, Brobbey A, Williamson T, Goyal M, Demchuk AM, et al. Functional outcome prediction in ischemic stroke: a comparison of machine learning algorithms and regression models. *Front Neurol.* (2020) 11:889. doi: 10.3389/fneur.2020.00889
- Dashtbani Moghari M, Zhou L, Yu B, Young N, Moore K, Evans A, et al. Efficient radiation dose reduction in whole-brain CT perfusion imaging using a 3D GAN: performance and clinical feasibility. *Phys Med Biol.* (2021) 66:075008. doi: 10.1088/1361-6560/abe917

19. Kriegeskorte N, Golan T. Neural network models and deep learning. *Curr Biol*. (2019) 29:R231–6. doi: 10.1016/j.cub.2019.02.034
20. Liu X, Liu C, Huang R, Zhu H, Liu Q, Mitra S, et al. Long short-term memory recurrent neural network for pharmacokinetic-pharmacodynamic modeling. *CP*. (2021) 59:138–46. doi: 10.5414/CP203800
21. Chen X, Li T-H, Zhao Y, Wang C-C, Zhu C-C. Deep-belief network for predicting potential miRNA-disease associations. *Brief Bioinformatics*. (2021) 22:bbaa186. doi: 10.1093/bib/bbaa186
22. Wu O, Winzeck S, Giese A-K, Hancock BL, Etherton MR, Bouts MJRJ, et al. Big data approaches to phenotyping acute ischemic stroke using automated lesion segmentation of multi-center magnetic resonance imaging data. *Stroke*. (2019) 50:1734–41. doi: 10.1161/STROKEAHA.119.025373
23. Poirion OB, Jing Z, Chaudhary K, Huang S, Garmire LX. DeepProg: an ensemble of deep-learning and machine-learning models for prognosis prediction using multi-omics data. *Genome Med*. (2021) 13:112. doi: 10.1186/s13073-021-00930-x
24. Wang H, Sun Y, Ge Y, Wu P-Y, Lin J, Zhao J, et al. Clinical-radiomics nomogram for functional outcome predictions in ischemic stroke. *Neurol Ther*. (2021) 10:819–32. doi: 10.1007/s40120-021-00263-2
25. Tang N, Zhang R, Wei Z, Chen X, Li G, Song Q, et al. Improving the performance of lung nodule classification by fusing structured and unstructured data. *Information Fusion*. (2022) 88:161–74. doi: 10.1016/j.inffus.2022.07.019
26. Hofmeister J, Bernava G, Rosi A, Vargas MI, Carrera E, Montet X, et al. Clot-based radiomics predict a mechanical thrombectomy strategy for successful recanalization in acute ischemic stroke. *Stroke*. (2020) 51:2488–94. doi: 10.1161/STROKEAHA.120.030334
27. Bahaddad AA, Ragab M, Ashary EB, Khalil EM. Metaheuristics with deep learning-enabled Parkinson's disease diagnosis and classification model. *J Healthc Eng*. (2022) 2022:9276579. doi: 10.1155/2022/9276579
28. Gao C, Zhang R, Chen X, Yao T, Song Q, Ye W, et al. Integrating Internet multisource big data to predict the occurrence and development of COVID-19 cryptic transmission. *NPJ Digit Med* (2022) 5:161. doi: 10.1038/s41746-022-00704-8
29. Dehghani M, Trojovská E, Zušák T. A new human-inspired metaheuristic algorithm for solving optimization problems based on mimicking sewing training. *Sci Rep*. (2022) 12:17387. doi: 10.1038/s41598-022-22458-9
30. Kumar Pandey R, Aggarwal S, Nath G, Kumar A, Vaferi B. Metaheuristic algorithm integrated neural networks for well-test analyses of petroleum reservoirs. *Sci Rep*. (2022) 12:16551. doi: 10.1038/s41598-022-21075-w
31. Alshareef AM, Alsini R, Alsieni M, Alrowais F, Marzouk R, Abunadi I, et al. Optimal deep learning enabled prostate cancer detection using microarray gene expression. *J Healthc Eng*. (2022) 2022:7364704. doi: 10.1155/2022/7364704
32. Nematzadeh S, Kiani F, Torkamanian-Afshar M, Aydin N. Tuning hyperparameters of machine learning algorithms and deep neural networks using metaheuristics: a bioinformatics study on biomedical and biological cases. *Comput Biol Chem*. (2022) 97:107619. doi: 10.1016/j.compbiolchem.2021.107619
33. Larsen K, Jæger HS, Hov MR, Thorsen K, Solyga V, Lund CG, et al. Streamlining acute stroke care by introducing national institutes of health stroke scale in the emergency medical services: a prospective cohort study. *Stroke*. (2022) 53:2050–7. doi: 10.1161/STROKEAHA.121.036084
34. Elsaid N, Bigliardi G, Dell'Acqua ML, Vandelli L, Ciolli L, Picchetto L, et al. Evaluation of stroke prognostication using age and NIH Stroke Scale index (SPAN-100 index) in delayed intravenous thrombolysis patients (beyond 45 hours). *J Stroke Cerebrovasc Dis*. (2022) 31:106384. doi: 10.1016/j.jstrokecerebrovasdis.2022.106384
35. Ye W, Zhang L, Zhang W, Wu X, Yi D, Wu Y, et al. Comparison of single imputation and multiple imputation methods for missing data in different oncogene expression profiles. *Biostat Epidemiol*. (2022) 6:113–27. doi: 10.1080/24709360.2021.2023805
36. Staartjes VE, Kernbach JM, Stumpo V, van Niftrik CHB, Serra C, Regli L. Foundations of feature selection in clinical prediction modeling. *Acta Neurochir Suppl*. (2022) 134:51–7. doi: 10.1007/978-3-030-85292-4_7
37. Bommert A, Welchowski T, Schmid M, Rahnenführer J. Benchmark of filter methods for feature selection in high-dimensional gene expression survival data. *Brief Bioinform*. (2022) 23:bbab354. doi: 10.1093/bib/bbab354
38. Riaz S, Latif S, Usman SM, Ullah SS, Algarni AD, Yasin A, et al. Malware detection in internet of things (IoT) devices using deep learning. *Sensors*. (2022) 22:9305. doi: 10.3390/s22239305
39. Diao X, Huo Y, Zhao S, Yuan J, Cui M, Wang Y, et al. Automated ICD coding for primary diagnosis via clinically interpretable machine learning. *Int J Med Inform*. (2021) 153:104543. doi: 10.1016/j.ijmedinf.2021.104543
40. Salvi M, Michielli N, Molinari F. Stain color adaptive normalization (SCAN) algorithm: separation and standardization of histological stains in digital pathology. *Comput Methods Programs Biomed*. (2020) 193:105506. doi: 10.1016/j.cmpb.2020.105506
41. Dorgalaleh A, Favaloro EJ, Bahraimi M, Rad F. Standardization of prothrombin time/international normalized ratio (PT/INR). *Int J Lab Hematol*. (2021) 43:21–8. doi: 10.1111/ijlh.13349
42. Miyoshi T, Tanaka H. Standardization of normal values for cardiac chamber size in echocardiography. *J Med Ultrason*. (2001) 49:21–33. doi: 10.1007/s10396-021-01147-6
43. Risso D. Normalization of single-cell RNA-seq data. *Methods Mol Biol*. (2021) 2284:303–29. doi: 10.1007/978-1-0716-1307-8_17
44. Wang S-F, Ku F-L, Zhang H-X. Improved PCA facial recognition with bootstrap and data standardization in small sample case. *Proceedings of 2011 International Conference on Computer Science and Network Technology*. Harbin, China: IEEE (2011). p. 2618–22. doi: 10.1109/ICCNSNT.2011.6182504
45. Zeng X. Length of stay prediction model of indoor patients based on light gradient boosting machine. *Comput Intell Neurosci*. (2022) 2022:9517029. doi: 10.1155/2022/9517029
46. Zhang C, Lei X, Liu L. Predicting metabolite-disease associations based on lightGBM model. *Front Genet*. (2021) 12:660275. doi: 10.3389/fgene.2021.660275
47. Du Z, Zhong X, Wang F, Uversky VN. Inference of gene regulatory networks based on the Light Gradient Boosting Machine. *Comput Biol Chem*. (2022) 101:107769. doi: 10.1016/j.compbiolchem.2022.107769
48. Zeng H, Yang C, Zhang H, Wu Z, Zhang J, Dai G, et al. LightGBM-based EEG analysis method for driver mental states classification. *Comput Intell Neurosci*. (2019) 2019:3761203. doi: 10.1155/2019/3761203
49. Hu X, Yin S, Zhang X, Menon C, Fang C, Chen Z, et al. Blood pressure stratification using photoplethysmography and light gradient boosting machine. *Front Physiol*. (2023) 14:1072273. doi: 10.3389/fphys.2023.1072273
50. Gan L. XGBoost-based e-commerce customer loss prediction. *Comput Intell Neurosci*. (2022) 2022:1858300. doi: 10.1155/2022/1858300
51. Shin H. XGBoost regression of the most significant photoplethysmogram features for assessing vascular aging. *IEEE J Biomed Health Inform*. (2022) 26:3354–61. doi: 10.1109/JBHI.2022.3151091
52. Song X, Zhu J, Tan X, Yu W, Wang Q, Shen D, et al. XGBoost-based feature learning method for mining COVID-19 novel diagnostic markers. *Front Public Health*. (2022) 10:926069. doi: 10.3389/fpubh.2022.926069
53. Zhang J, Ma X, Zhang J, Sun D, Zhou X, Mi C, et al. Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. *J Environ Manage*. (2023) 332:117357. doi: 10.1016/j.jenvman.2023.117357
54. Khan MS, Salsabil N, Alam MGR, Dewan MAA, Uddin MZ. CNN-XGBoost fusion-based affective state recognition using EEG spectrogram image analysis. *Sci Rep*. (2022) 12:14122. doi: 10.1038/s41598-022-18257-x
55. Eslami T, Saeed F. Fast-GPU-PCC: a GPU-based technique to compute pairwise pearson's correlation coefficients for time series data-fMRI study. *High Throughput*. (2018) 7:11. doi: 10.3390/ht7020011
56. Lu L-L, Hu X-J, Yang Y, Xu S, Yang S-Y, Zhang C-Y, et al. Correlation of myopia onset and progression with corneal biomechanical parameters in children. *World J Clin Cases*. (2022) 10:1548–56. doi: 10.12998/wjcc.v10.i5.1548
57. Dickinson Q, Meyer JG. Positional SHAP (PoSHAP) for interpretation of machine learning models trained from biological sequences. *PLoS Comput Biol*. (2022) 18:e1009736. doi: 10.1371/journal.pcbi.1009736
58. Nohara Y, Matsumoto K, Soejima H, Nakashima N. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Comput Methods Programs Biomed*. (2022) 214:106584. doi: 10.1016/j.cmpb.2021.106584
59. Wu Y, Fang Y. Stroke prediction with machine learning methods among older Chinese. *Int J Environ Res Public Health*. (2020) 17:1828. doi: 10.3390/ijerph17061828
60. Wang B, Zhang C, Wong YD, Hou L, Zhang M, Xiang Y. Comparing resampling algorithms and classifiers for modeling traffic risk prediction. *Int J Environ Res Public Health*. (2022) 19:13693. doi: 10.3390/ijerph192013693
61. Xing M, Zhang Y, Yu H, Yang Z, Li X, Li Q, et al. Predict DLBCL patients' recurrence within two years with Gaussian mixture model cluster oversampling and multi-kernel learning. *Comput Methods Programs Biomed*. (2022) 226:107103. doi: 10.1016/j.cmpb.2022.107103
62. Sajjadian M, Lam RW, Milev R, Rotzinger S, Frey BN, Soares CN, et al. Machine learning in the prediction of depression treatment outcomes: a systematic review and meta-analysis. *Psychol Med*. (2021) 51:2742–51. doi: 10.1017/S0033291721003871
63. Xu Y, Ye W, Song Q, Shen L, Liu Y, Guo Y, et al. Using machine learning models to predict the duration of the recovery of COVID-19 patients hospitalized in Fangcang shelter hospital during the Omicron BA. 22 pandemic. *Front Med (Lausanne)*. (2022) 9:1001801. doi: 10.3389/fmed.2022.1001801
64. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. (1997) 9:1735–80. doi: 10.1162/neco.1997.9.8.1735
65. Gers FA, Schmidhuber J, Cummins F. Learning to forget: continual prediction with LSTM. *Neural Comput*. (2000) 12:2451–71. doi: 10.1162/089976600300015015
66. Attallah O. A deep learning-based diagnostic tool for identifying various diseases via facial images. *Digital Health*. (2022) 8:205520762211244. doi: 10.1177/20552076221124432

67. Yan J, Cai X, Chen S, Guo R, Yan H, Wang Y. Ensemble learning-based pulse signal recognition: classification model development study. *JMIR Med Inform.* (2021) 9:e28039. doi: 10.2196/28039
68. Kim J, Kang U, Lee Y. Statistics and deep belief network-based cardiovascular risk prediction. *Health Inform Res.* (2017) 23:169. doi: 10.4258/hir.2017.23.3.169
69. Mahbobi M, Kimiagari S, Vasudevan M. Credit risk classification: an integrated predictive accuracy algorithm using artificial and deep neural networks. *Ann Oper Res.* (2021). doi: 10.1007/s10479-021-04114-z
70. Abdel-Nasser M, Mahmoud K. Accurate photovoltaic power forecasting models using deep LSTM-RNN. *Neural Comput & Applic.* (2019) 31:2727–40. doi: 10.1007/s00521-017-3225-z
71. Zhang J, Sheng J, Lu J, Shen L, UCPSO. A uniform initialized particle swarm optimization algorithm with cosine inertia weight. *Comput Intell Neurosci.* (2021) 2021:1–18. doi: 10.1155/2021/8819333
72. Mirjalili S, Gandomi AH. Chaotic gravitational constants for the gravitational search algorithm. *Appl Soft Comput.* (2017) 53:407–19. doi: 10.1016/j.asoc.2017.01.008
73. Zhu H, Ge J, Qi W, Zhang X, Lu X. Dynamic analysis and image encryption application of a sinusoidal-polynomial composite chaotic system. *Math Comput Simul.* (2022) 198:188–210. doi: 10.1016/j.matcom.2022.02.029
74. Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. *Phys Med Biol.* (2018) 63:07TR01. doi: 10.1088/1361-6560/aab4b1
75. Ma Q, Li R, Wang L, Yin P, Wang Y, Yan C, et al. Temporal trend and attributable risk factors of stroke burden in China, 1990–2019: an analysis for the Global Burden of Disease Study 2019. *Lancet Public Health.* (2021) 6:e897–906. doi: 10.1016/S2468-2667(21)00228-0
76. Muhammad Usman S, Khalid S, Bashir S. A deep learning based ensemble learning method for epileptic seizure prediction. *Comput Biol Med.* (2021) 136:104710. doi: 10.1016/j.combiomed.2021.104710
77. Cheng L-H, Hsu T-C, Lin C. Integrating ensemble systems biology feature selection and bimodal deep neural network for breast cancer prognosis prediction. *Sci Rep.* (2021) 11:14914. doi: 10.1038/s41598-021-92864-y
78. Zhang R, Gao C, Chen X, Li F, Yi D, Wu Y. Genetic algorithm optimised Hadamard product method for inconsistency judgement matrix adjustment in AHP and automatic analysis system development. *Expert Syst Appl.* (2023) 211:118689. doi: 10.1016/j.eswa.2022.118689
79. Ogami C, Tsuji Y, Seki H, Kawano H, To H, Matsumoto Y, et al. An artificial neural network–pharmacokinetic model and its interpretation using Shapley additive explanations. *CPT Pharmacometrics Syst Pharmacol.* (2021) 10:760–8. doi: 10.1002/psp4.12643
80. Versaci M, Angiulli G, di Barba P, Morabito FC. Joint use of eddy current imaging and fuzzy similarities to assess the integrity of steel plates. *Open Physics.* (2020) 18:230–40. doi: 10.1515/phys-2020-0159

Frontiers in Neurology

Explores neurological illness to improve patient care

The third most-cited clinical neurology journal explores the diagnosis, causes, treatment, and public health aspects of neurological illnesses. Its ultimate aim is to inform improvements in patient care.

Discover the latest Research Topics

[See more →](#)

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact

