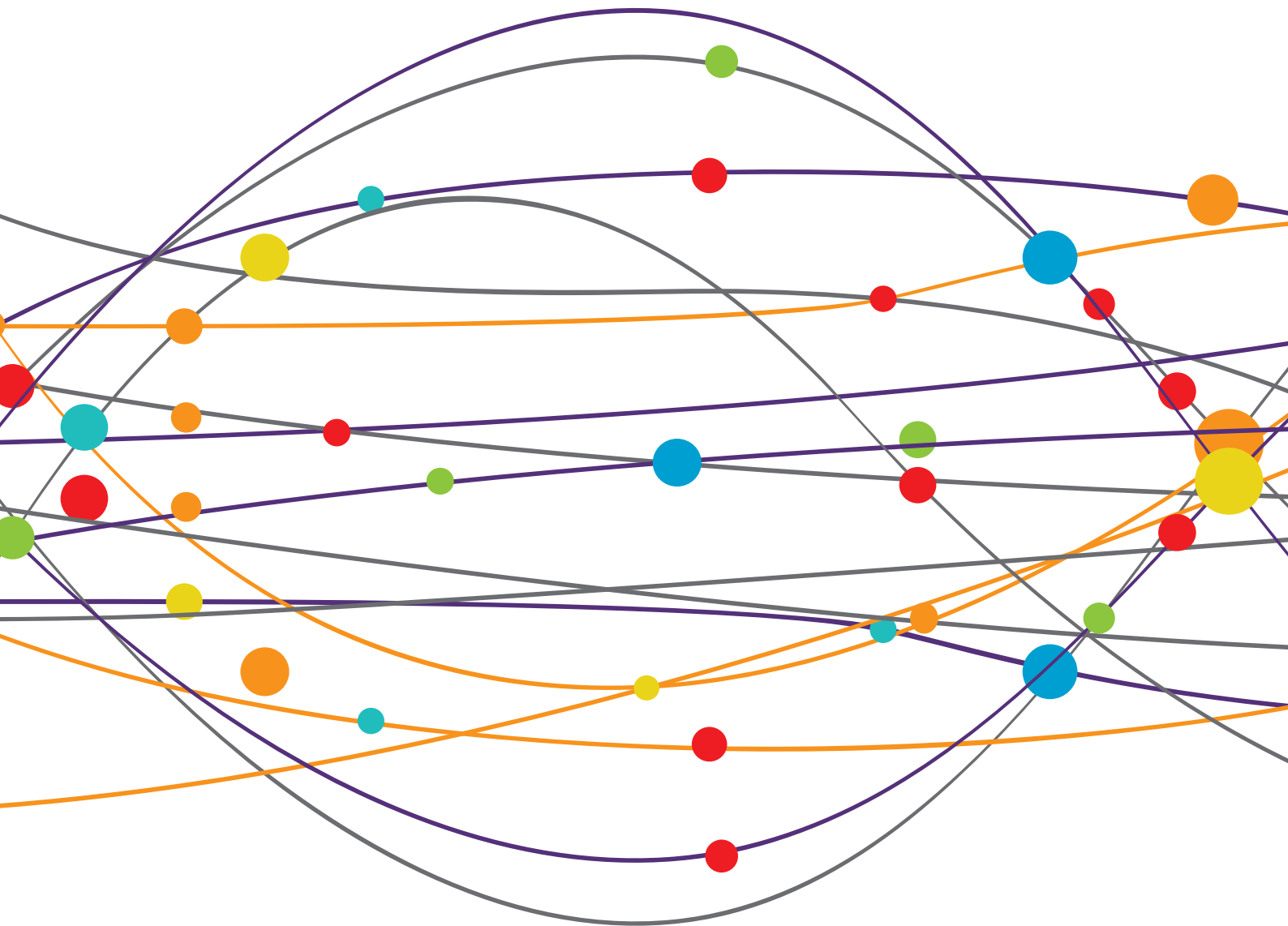


MACHINE LEARNING AND DECISION SUPPORT IN STROKE

EDITED BY: Fabien Scalzo and David S. Liebeskind
PUBLISHED IN: Frontiers in Neurology





frontiers

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 978-2-88963-846-8

DOI 10.3389/978-2-88963-846-8

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: researchtopics@frontiersin.org

MACHINE LEARNING AND DECISION SUPPORT IN STROKE

Topic Editors:

Fabien Scalzo, University of California, Los Angeles, United States

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

Citation: Scalzo, F., Liebeskind, D. S., eds. (2020). Machine Learning and Decision Support in Stroke. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88963-846-8

Table of Contents

- 05 Editorial: Machine Learning and Decision Support in Stroke**
David S. Liebeskind and Fabien Scalzo
- 07 Multimodal MRI-Based Triage for Acute Stroke Therapy: Challenges and Progress**
Oh Young Bang, Jong-Won Chung, Jeong Pyo Son, Wi-Sun Ryu, Dong-Eog Kim, Woo-Keun Seo, Gyeong-Moon Kim and Yoon-Chul Kim
- 16 Optimal Delay Time of CT Perfusion for Predicting Cerebral Parenchymal Hematoma After Intra-Arterial tPA Treatment**
Bing Wu, Nan Liu, Max Wintermark, Mark W. Parsons, Hui Chen, Longting Lin, Shuai Zhou, Gang Hu, Yongwei Zhang, Jun Hu, Ying Li, Zihua Su, Xinhuai Wu and Guangming Zhu
- 24 Application of Machine Learning to Automated Analysis of Cerebral Edema in Large Cohorts of Ischemic Stroke Patients**
Rajat Dhar, Yasheng Chen, Hongyu An and Jin-Moo Lee
- 33 A Bayesian Network Model for Predicting Post-stroke Outcomes With Available Risk Factors**
Eunjeong Park, Hyuk-jae Chang and Hyo Suk Nam
- 44 Relating Acute Lesion Loads to Chronic Outcome in Ischemic Stroke—An Exploratory Comparison of Mismatch Patterns and Predictive Modeling**
Simon Habegger, Roland Wiest, Bruno J. Weder, Pasquale Mordasini, Jan Gralla, Levin Häni, Simon Jung, Mauricio Reyes and Richard McKinley
- 55 ISLES 2016 and 2017-Benchmarking Ischemic Stroke Lesion Outcome Prediction Based on Multispectral MRI**
Stefan Winzeck, Arsany Hakim, Richard McKinley, José A. A. D. S. R. Pinto, Victor Alves, Carlos Silva, Maxim PISOV, Egor Krivov, Mikhail Belyaev, Miguel Monteiro, Arlindo Oliveira, Youngwon Choi, Myunghee Cho Paik, Yongchan Kwon, Hanbyul Lee, Beom Joon Kim, Joong-Ho Won, Mobarakol Islam, Hongliang Ren, David Robben, Paul Suetens, Enhao Gong, Yilin Niu, Junshen Xu, John M. Pauly, Christian Lucas, Mattias P. Heinrich, Luis C. Rivera, Laura S. Castillo, Laura A. Daza, Andrew L. Beers, Pablo Arbelaezs, Oskar Maier, Ken Chang, James M. Brown, Jayashree Kalpathy-Cramer, Greg Zaharchuk, Roland Wiest and Mauricio Reyes
- 75 Predicting Outcome of Endovascular Treatment for Acute Ischemic Stroke: Potential Value of Machine Learning Algorithms**
Hendrikus J. A. van Os, Lucas A. Ramos, Adam Hilbert, Matthijs van Leeuwen, Marianne A. A. van Walderveen, Nyika D. Kruyt, Diederik W. J. Dippel, Ewout W. Steyerberg, Irene C. van der Schaaf, Hester F. Lingsma, Wouter J. Schonewille, Charles B. L. M. Majoie, Silvia D. Olabarriaga, Koos H. Zwinderman, Esmee Venema, Henk A. Marquering, Marieke J. H. Wermer and the MR CLEAN Registry Investigators
- 83 Decision Criteria for Large Vessel Occlusion Using Transcranial Doppler Waveform Morphology**
Samuel G. Thorpe, Corey M. Thibeault, Nicolas Canac, Seth J. Wilk, Thomas Devlin and Robert B. Hamilton

- 93 *Machine Learning in Acute Ischemic Stroke Neuroimaging***
Haris Kamal, Victor Lopez and Sunil A. Sheth
- 99 *Learning to Predict Ischemic Stroke Growth on Acute CT Perfusion Data by Interpolating Low-Dimensional Shape Representations***
Christian Lucas, André Kemmling, Nassim Bouteldja, Linda F. Aulmann, Amir Madany Mamlouk and Mattias P. Heinrich
- 114 *Stroke Lesion Outcome Prediction Based on MRI Imaging Combined With Clinical Information***
Adriano Pinto, Richard Mckinley, Victor Alves, Roland Wiest, Carlos A. Silva and Mauricio Reyes
- 124 *Convolutional Neural Networks for Direct Inference of Pharmacokinetic Parameters: Application to Stroke Dynamic Contrast-Enhanced MRI***
Cagdas Ulas, Dhritiman Das, Michael J. Thrippleton, Maria del C. Valdés Hernández, Paul A. Armitage, Stephen D. Makin, Joanna M. Wardlaw and Bjoern H. Menze
- 138 *Early Identification of High-Risk TIA or Minor Stroke Using Artificial Neural Network***
Ka Lung Chan, Xinyi Leng, Wei Zhang, Weinan Dong, Quanli Qiu, Jie Yang, Yannie Soo, Ka Sing Wong, Thomas W. Leung and Jia Liu
- 145 *STIR-Net: Deep Spatial-Temporal Image Restoration Net for Radiation Reduction in CT Perfusion***
Yao Xiao, Peng Liu, Yun Liang, Skylar Stolte, Pina Sanelli, Ajay Gupta, Jana Ivanidze and Ruogu Fang



Editorial: Machine Learning and Decision Support in Stroke

David S. Liebeskind* and Fabien Scalzo

Neurovascular Imaging Research Core and UCLA Stroke Center, Department of Neurology, University of California, Los Angeles, Los Angeles, CA, United States

Keywords: stroke, imaging, machine learning, ischemia, perfusion

Editorial on the Research Topic

Machine Learning and Decision Support in Stroke

Real-world, intelligent application of artificial intelligence (AI) for augmented stroke care has dramatically expanded in the last several years. In this volume, we use the term AI to refer to its popularized definition, more precisely referred to as machine learning (ML) and related to the automatic learning of a computational model from a previously acquired labeled dataset. The use of ML for decision support in stroke has been adopted as a key research priority in most academic institutions while industry partners have flourished, regulatory agencies have adapted to keep pace and clinicians across the globe look to implement these tools to improve stroke care. The application of AI and its combination with imaging has the potential to transform how stroke care is delivered worldwide. Over the last decade, AI has attained several technology milestones by capitalizing on big data, providing mobile solutions, relying on cloud computing, implementing full automation, and advanced graphical interfaces for better visualization. AI has been a priority investment area for healthcare providers as it is expected to increase productivity, efficiency, and geographical reach of healthcare delivery. AI may also improve the experience of stroke care providers, accelerating decision support with imaging and enabling them to spend more time in direct patient care. Exuberance and seemingly miraculous applications of AI abound in stroke, from detecting the earliest signs of ischemia on CT to delivering key metrics on perfusion or blood flow in specific areas of the brain. This clinical perspective may be quite different from the application of AI in radiology or pathology subspecialties, where potential existential threats abound. At present, most advanced imaging software modules in stroke have revolutionized the use of imaging data, yet AI is often only used in extremely limited aspects. In fact, the use of AI in imaging has been balanced with applications to rapidly glean essential information from electronic health records. The numerous AI methods in this volume reflect predominantly academic collaborations that mirror the continually expanding list of novel software products available on the market. Rapid FDA clearance has facilitated numerous commercial products in recent years, but expansive marketing claims may have subsequently overemphasized their impact in saving lives due to stroke. Unfortunately, there have been gaps in rigorous validation and post-marketing surveillance, with prime emphasis on the ability to modernize and simplify clinical decision making.

The broad array of topics in this volume speak to innovation by content experts who understand far beyond the simple or isolated imaging data. Bang et al. provide their perspective on the development of multimodal MRI triage strategies, including obstacles and achievements along the way. Kamal et al. provide a survey of various ML applications in imaging of acute ischemic stroke. Novel methods to predict CT perfusion lesion growth are offered by Lucas et al.. Almost a dozen more original research articles describe unique ways to apply AI, using innovative methods on CT, MRI, and TCD in a wide variety of settings from around the world to potentially streamline stroke care by advancing decision support (Chan et al.; Dhar et al.; Habegger et al.; Park et al.; Pinto et al.;

OPEN ACCESS

Edited and reviewed by:

Jean-Claude Baron,
University of Cambridge,
United Kingdom

*Correspondence:

David S. Liebeskind
dliebeskind@mednet.ucla.edu

Specialty section:

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

Received: 20 April 2020

Accepted: 04 May 2020

Published: 29 May 2020

Citation:

Liebeskind DS and Scalzo F (2020)
Editorial: Machine Learning and
Decision Support in Stroke.
Front. Neurol. 11:486.
doi: 10.3389/fneur.2020.00486

Thorpe et al.; Ulas et al.; van Os et al.; Winzeck et al.; Wu et al. Xiao et al.). These reports chronicle the remarkable progress achieved in AI for stroke imaging, reflecting an early sea change preceding the expected tsunami of AI stroke imaging tools to flood bedside encounters in coming years.

Gaps undoubtedly remain as AI methods represent only a fraction of these modern tools. Automation of image processing is extremely valuable, but the ML component remains limited. Most AI tools for stroke imaging rapidly generate an existing variable commonly used, such as an ASPECTS score or ischemic core lesion volume. However, the methods, definitions, and ultimate reliability often remain remiss or ill-defined. In such scenarios, there is not much “machine” learning or deep learning, while clinicians will not learn unless they contrast automated results with their own review of the images. Radiology worklists may be quickly sorted to facilitate their priority reads, but sensitivity and specificity of these methods are both critical. From the bedside, some clinicians have argued that an “eyeball” method to rapidly glance at the imaging to recognize simple patterns may be enough. The clinical context is therefore essential. For example, standard definitions of core and penumbra on CT perfusion generated by automated software have been developed and validated predominantly in acute, complete occlusion of a proximal artery such as the MCA. Such defined metrics for the delayed perfusion of penumbra or decreased cerebral blood flow are not applicable in more subacute strokes, in the presence of stenoses or in other territories where the collateral flow patterns are different. The exact volume of such lesions is also extremely simplistic, as the topology or pattern is essential information that is poorly captured by machine, yet readily seen by expert eyes. The utility of imaging in stroke for decision support has always been driven by the most subtle findings and focused around the clinical context. For instance, recognizing FLAIR vascular hyperintensities in the distal MCA territory may be critical information for the patient presenting with transient or mild

hemi-neglect due to right hemispheric ischemia. Unfortunately, machine learning must be trained by such rich data that incorporates these critical clinical contextual data. Similarly, the development of AI tools should be prompted to address the most pertinent clinical questions in decision support. Clinician input and continual development in light of this perspective is therefore key. The papers in this volume largely reflect this perspective from the bedside application in real-world scenarios. Validation of nascent tools also requires human, clinical expertise to link results with the underlying pathophysiology. As a result, AI stroke imaging methods will inevitably retain dependence on over-reading by experts and they cannot replicate the role of core lab, detailed adjudications.

The future of AI or ML in decision support with imaging is undoubtedly a key facet of modernization in the delivery of stroke care. Modernization, however, does not equate with a complete replacement of current practice or the role of expertise. Electricity was discovered and harnessed to modernize many aspects of daily life after billions of years of electromagnetic energy on our planet. Similarly, AI of stroke imaging will require much clinical expertise to continually modernize stroke care.

AUTHOR CONTRIBUTIONS

DL and FS contributed to the creation of this article, from initial drafts to critical revisions, and final production.

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.

Copyright © 2020 Liebeskind and Scalzo. 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.



Multimodal MRI-Based Triage for Acute Stroke Therapy: Challenges and Progress

Oh Young Bang^{1,2*}, Jong-Won Chung¹, Jeong Pyo Son², Wi-Sun Ryu³, Dong-Eog Kim³, Woo-Keun Seo¹, Gyeong-Moon Kim¹ and Yoon-Chul Kim⁴

¹ Department of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea,

² Department of Health Sciences and Technology, Samsung Advanced Institute for Health Sciences and Technology,

Sungkyunkwan University, Seoul, South Korea, ³ Stroke Center and Korean Brain MRI Data Center, Dongguk University Ilsan Hospital, Goyang, South Korea, ⁴ Samsung Medical Center, Clinical Research Institute, Seoul, South Korea

OPEN ACCESS

Edited by:

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

Reviewed by:

Claus Ziegler Simonsen,
Aarhus University Hospital, Denmark
Maurizio Acampa,
Azienda Ospedaliera Universitaria
Senese, Italy

*Correspondence:

Oh Young Bang
ohyoung.bang@samsung.com

Specialty section:

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

Received: 15 May 2018

Accepted: 29 June 2018

Published: 24 July 2018

Citation:

Bang OY, Chung J-W, Son JP,
Ryu W-S, Kim D-E, Seo W-K,
Kim G-M and Kim Y-C (2018)
Multimodal MRI-Based Triage for
Acute Stroke Therapy: Challenges and
Progress. *Front. Neurol.* 9:586.
doi: 10.3389/fneur.2018.00586

Revascularization therapies have been established as the treatment mainstay for acute ischemic stroke. However, a substantial number of patients are either ineligible for revascularization therapy, or the treatment fails or is futile. At present, non-contrast computed tomography is the first-line neuroimaging modality for patients with acute stroke. The use of magnetic resonance imaging (MRI) to predict the response to early revascularization therapy and to identify patients for delayed treatment is desirable. MRI could provide information on stroke pathophysiologies, including the ischemic core, perfusion, collaterals, clot, and blood-brain barrier status. During the past 20 years, there have been significant advances in neuroimaging as well as in revascularization strategies for treating patients with acute ischemic stroke. In this review, we discuss the role of MRI and post-processing, including machine-learning techniques, and recent advances in MRI-based triage for revascularization therapies in acute ischemic stroke.

Keywords: stroke, MRI, endovascular treatment, machine learning, triage

INTRODUCTION

Revascularization therapies, including rt-PA and EVT, have been established as the mainstay of treatment for acute ischemic stroke. It has become clear that consideration of heterogeneity among stroke patients is of importance in these therapies. Neuroimaging has been used as a triage tool for revascularization therapy in patients with acute stroke. The use of magnetic resonance imaging (MRI) for predicting the response to early revascularization therapy and for identifying patients in whom delayed treatment is appropriate is desirable.

During the past 20 years, there have been significant advances in both neuroimaging as well as in revascularization strategies for treating patients with acute ischemic stroke. In this review, the role of MRI discussed and the recent advances in MRI-based triage for revascularization therapies and in post-processing, including machine learning techniques, in these patients.

NEUROIMAGING STUDIES IN THE ACUTE STROKE INTERVENTION FIELD: RESULTS FROM RANDOMIZED CONTROLLED TRIALS

Previous intravenous rt-PA trials [the NINDS rt-PA (1) and ECASS-III (2) trials] have used non-contrast computed tomography (NCCT) images. Randomized controlled trials (RCTs) of EVT have implemented MRI or computed tomography perfusion/angiography (CTP/CTA)

techniques, in addition to the NCCT Alberta Stroke Program Early CT Score (ASPECTS). However, the results of RCTs and the rapid evolution of neuroimaging techniques have led to significant changes in the international guidelines for neuroimaging in acute ischemic stroke over time (**Figure 1**).

Multicenter prospective MRI studies, including the DEFUSE and EPITHET trials for intravenous rt-PA given more than 3 h post-stroke, reported a significant association between recanalization and reduced infarct growth in patients with an MR perfusion–diffusion mismatch (3, 4). Expert consensus on the Acute Stroke Imaging Research (ASIR) Roadmap reported on the methodological issues in perfusion and penumbral imaging (5). The 2009 American Heart Association/American Stroke Association (AHA/ASA) guidelines recommended that MRI or CTA should be performed in conjunction with vascular and perfusion studies for EVT (6).

The IMS III (7), MR RESCUE (8), and SYNTHESIS expansion (9) trials were multicenter, prospective RCTs that failed to show a benefit from EVT for acute ischemic stroke. The potential reasons for this failure include time delays to angiographic reperfusion and the inclusion of patients with large core or no large vessel occlusion. In addition, the MR RESCUE trial used an algorithm that implements multiple diffusion-weighted imaging (DWI) and MR perfusion (MRP) parameters, but failed to show that patients with a favorable penumbral pattern on neuroimaging benefitted from EVT (8). Given the lessons from these aforementioned three RCTs, published 2013, the ASIR Roadmap II provided guidelines for the use of imaging in stroke clinical trials (10). In the AHA/ASA 2013 guidelines, the recommendations were changed to “CTP and MRP may be considered” (11).

Further phase III RCTs were conducted in 2015; these included the MR CLEAN (12), ESCAPE (13), EXTEND-IA (14), SWIFT PRIME (15) and REVASCAT (16) trials. The findings of these RCTs demonstrated overwhelming evidence of the benefit of EVT for treatment of acute ischemic stroke with a small core (as measured by the ASPECTS) and large vessel occlusion. The ASIR Roadmap III proposed the optimal imaging profile for EVT,

based on the results of these recent positive RCTs: the presence of large vessel occlusion, a smaller core, good collaterals, and a large penumbra (17). After the success of the ASPECTS-based RCTs of EVT in 2015, the recommendations were again changed to “the benefits of CTP and MRP are unknown and need further studies” (18). For patients eligible for EVT, the 2016 AHA/ASA guidelines require the absence of bleeding and an ASPECTS of 6 points or more in NCCT, as well as the presence of causative occlusion of the internal carotid artery or proximal middle cerebral artery (18).

Very recently, the results of the phase III RCTs of EVT in an extended time-window showed a significant and remarkable functional recovery with EVT vs. that with medical treatment in carefully selected patients (19, 20). EVT was initiated between 6 and 16 h after onset in patients with a target mismatch in the DEFUSE 3 trial (20), and 6–24 h after onset in patients with mismatch between clinical presentation and DWI/CTP in the DAWN trial (19). In these trials, the benefits of EVT persisted [or even increased, i.e., “late-window paradox” (21)] across the period when patients had a small core and large salvageable tissues. Based on these trials, the new 2018 guidelines recommended that CTP or DWI/MRP scans be obtained if the patient presents more than 6 h after his/her last known normal status and has large vessel occlusion (LVO), and to perform EVT when eligibility criteria from these trials were met (22).

IMPLICATIONS OF MRI-BASED TRIAGE ON THE NUMBER OF PATIENTS RECEIVING EVT

The beneficial effect of EVT has been confirmed in selected patients with acute ischemic stroke. However, a substantial proportion of patients are EVT ineligible (only 7–13% of acute ischemic stroke patients are eligible for EVT) (23), have failure of reperfusion (TICI 0–2a in 14–41% in five recent phase III RCTs) (12–16), or have futile reperfusion (26–49% showed a poor outcome, despite successful recanalization) (24). These findings indicate that imaging-guided tailored treatment may be beneficial in acute ischemic stroke. Although there have been significant advances in CT techniques, the advantages of MRI techniques make MRI more desirable for use (25, 26). The advantages and limitations of CT-based triage for EVT as well as the recent advances in MRI techniques are summarized in **Table 1**.

In EVT-eligible patients, MRI-based triage may increase the efficacy of EVT, at the expense of decreasing the number of patients receiving EVT by excluding patients with large lesions on DWI (27) (DWI is superior to any CT techniques in imaging the infarct core) (25). In contrast, MRI-based triage can also increase EVT use in patients considered ineligible under the current guidelines, as follows.

First, wake-up stroke occurs in one-fifth of patients with stroke; it was estimated that 58,000 patients with wake-up strokes presented to an emergency department in the U.S. in 2005 (3 million wake-up stroke cases worldwide) (28). In these cases, less time might have elapsed from the onset of stroke because circadian variation for stroke is well-known, with most

Abbreviations: ASIR, acute stroke imaging research; ASPECTS, Alberta Stroke Program Early CT score; DAWN, Diffusion-Weighted Imaging or Computerized Tomography Perfusion Assessment with Clinical Mismatch in the Triage of Wake-Up and Late Presenting Strokes Undergoing Neurointervention with Trevo; DEFUSE, Diffusion and Perfusion Imaging Evaluation for Understanding Stroke Evolution study; DWI, diffusion-weighted image; ECASS, European Cooperative Acute Stroke Study; EPITHET, Echoplanar Imaging Thrombolytic Evaluation Trial; ESCAPE, Endovascular treatment for Small Core and Anterior circulation Proximal occlusion with Emphasis on minimizing CT to recanalization times; EXTEND-IA, Extending the Time for thrombolysis in Emergency Neurological Deficits–Intra-Arterial; FLAIR, fluid attenuation inversion recovery; HERMES, Highly Effective Reperfusion evaluated in Multiple Endovascular Stroke; IMS-III, Interventional Management of Stroke III; MR CLEAN, Multicenter Randomized Clinical trial of Endothelial treatment for Acute ischemic stroke; MRP, magnetic resonance perfusion; MR RESCUE, Mechanical Retrieval and Recanalization of Stroke Clots Using Embolectomy; REVASCAT, Randomized Trial of Revascularization With Solitaire FR Device vs. Best Medical Therapy in the Treatment of Acute Stroke Due to Anterior Circulation Large-Vessel Occlusion Presenting Within 8 h of Symptom Onset; SWIFT PRIME, Solitaire With the Intention For Thrombectomy as Primary Endovascular treatment for acute ischemic stroke; SYNTHESIS, intra-arterial vs. systemic thrombolysis for acute ischemic stroke.

	Multicenter RCTs	Major development or findings	Guideline or Expert consensus
1995	NINDS rt-PA		
1997		CT angiography	
1999	Intra-arterial fibrinolysis (PROACT-II)		
2001		non-contrast CT ASPECTS	
2007	Multicenter prospective MRI studies (DEFUSE and EPITHET study)	Defining MRP-DWI mismatch MRP-derived blood-brain barrier image	
2008	rt-PA window extended to 4.5 h (ECASS-III)		ASIR roadmap (penumbra image)
2009	Failure of intravenous desmoteplase trial (DIAS-2)	DWI-FLAIR mismatch for wake-up stroke	AHA/ASA 2009 guidelines (MRI or CTA should be performed along with vascular and perfusion studies for EVT)
2010		Automated post-process (e.g., RAPID) ASL techniques	
2013	Failure of three EVT trials (IMS-III, MR RESCUE, SYNTHESIS EXP)	Futile recanalization in large core and poor collaterals or at a later time point	ASIR roadmap II (guideline for image in RCTs)
2014		Short (6 min) MR acquisition time MRP-based collateral imaging techniques Clot imaging techniques	AHA/ASA 2013 guidelines (CTP and MRP may be considered beyond the time window for IVT)
2015	Benefit of EVT in five RCTs (MR CLEAN, ESCAPE, SWIFT PRIME, EXTEND-IA, REVASCAT)	Optimal imaging profile is LVO, small core, good collaterals, and large penumbra	ASIR roadmap III (optimal image profile for EVT) AHA/ASA 2016 guidelines (The benefits of CTP and MRP are unknown and need further studies)
2017		Machine learning based imaging techniques	
2018	Benefit of EVT in unclear onset/ <24 h (DAWN), and <16 h (DEFUSE 3) Benefit of IVT in unclear onset stroke (WAKE-UP)	'Late window paradox'	AHA/ASA 2018 guidelines (CTP or MRP are need beyond 6 h)
Future			

FIGURE 1 | Major randomized controlled trials (RCTs) and corresponding guidelines in the field of acute stroke therapy. rt-PA, recombinant tissue plasminogen activator; ASPECTS, Alberta Stroke Program Early CT Score; MRP, magnetic resonance perfusion; DWI, diffusion-weighted image; FLAIR, fluid attenuation inversion recovery; ASIR, Acute Stroke Imaging Research; AHA/ASA, American Heart Association/American Stroke Association; DIAS-2, Desmoteplase in acute ischemic stroke-2; RAPID, RAPid postprocess for Perfusion and Diffusion; ASL, arterial spin labeling; LVO, large vessel occlusion; EVT, endovascular therapy. Please see Glossary for other abbreviations.

cerebrovascular events known to occur during the morning (29). When also considering cases with an unknown time of symptom onset, for example, non-witnessed stroke with aphasia or disturbance of consciousness, it appears that the time of symptom onset is unknown in 14–35% of patients with acute stroke (28, 30–32). It is possible that some of these patients could benefit from revascularization therapies. A population-based study has shown that more than one-third of wake-up strokes would have been eligible for thrombolysis if arrival time were not a factor (28). MRI features combining fluid attenuation inversion recovery (FLAIR) sequences with DWI have been investigated as a surrogate marker for lesion age and a DWI-positive/FLAIR-negative mismatch pattern was identified in patients within 4.5 h of stroke onset in the middle cerebral artery territory, with high predictive values (33, 34). A recent RCT (WAKE-UP) showed that in patients with acute ischemic stroke with an unknown time of onset, intravenous rt-PA guided by a mismatch between DWI and FLAIR in the region of ischemia resulted in a significantly better functional outcome than the control group (35).

Second, patients could receive EVT if they have ASPECTS of ≥ 6 on NCCT and LVO on CTA, and if treatment can be initiated within 6 h of symptom onset (18). However, a significant proportion of patients arrived late at a comprehensive stroke center, where EVT can be performed, and some patients may require a longer procedural time. Data from US academic medical centers have shown that one-third of patients arrived

more than 6 h after symptom onset (36). The ESCAPE trial showed that EVT improved functional outcomes and reduced mortality in patients with a small infarct core and moderate-to-good collateral circulation, up to 12 h after symptom onset (13). The HERMES investigators performed a meta-analysis of individual patient data from five recent RCTs of EVT, to test whether EVT is efficacious across a diverse population (i.e., a lower ASPECTS or longer onset-to-groin puncture time, etc.) (37). Patients with small cores (high ASPECTS) had a slower decline in benefit with longer symptom onset-to-reperfusion, than patients with larger infarct cores (38). In addition, reperfusion is related to a positive clinical outcome only if adequate collateralization can prevent infarction until the vessel can be recanalized. A good collateral status could thus feasibly extend the time-window for EVT (39–41). Patients with good collaterals as assessed by MRI showed a favorable outcome in terms of infarct growth at day 7 and modified Rankin score at day 90 (42, 43). Therefore, inclusion of patients with good collaterals, but not in those with larger cores, the time-window for EVT may be extended. Indeed, the results of the DAWN and DEFUSE 3 trials have extended the time-window in these patients to 16–24 h (19, 20).

Lastly, EVT is not recommended in patients with ASPECTS of <6 points on NCCT, according to the current guidelines (18). Data from the ECASS II study, in which 800 patients were randomized to rt-PA or placebo within 6 h of symptom onset,

TABLE 1 | Requirements for neuroimaging for EVT and recent advances in MRI techniques.

Requirements	Imaging triages		
	Non-contrast CT	Advanced CT (CTA, CTP)	Advances in MRI
IS IT POSSIBLE?			
Fast	Short scan time No need for post-processing	10–15 min for complete protocol	Short (6 min) acquisition time Automatic fast post-processing
Available	Widely available 24-7-365	Generally greater availability	Available in all comprehensive stroke centers
Reliable			
Standardization	Good	Parameters are being defined	Surrogate markers and their threshold and sequence are being defined
Inter-rate reliability	Modest (especially in the hyperacute phase)	Good	Machine-learning is being applied
IS IT NECESSARY?			
Stroke pathophysiology	ASPECTS for determining the extent of the core Dense MCA sign	CTP for core and penumbra CTA for collateral CTP for BBB damage	Perfusion pattern (Expected salvageable tissue) Collateral status (Predict response to treatment) BBB damage (Avoid hemorrhagic complications) Clot to treat (Need for adjuvant therapy)
Save more (proportion of extra cases) Unclear onset (1 in 3 patients) Delayed EVT (1 in 4 patients) ASPECTS <6 (1 in 8 patients)	Not applicable	Guided by not applicable CTA collaterals CTP, CTA ASPECTS	Guided by rFLAIR, Collateral, MRP Collateral, MRP DWI, Collateral, MRP
IS IT TESTED?			
Confirmed in RCTs	Role of ASPECTS confirmed	CTP (mismatch)-based EVT confirmed	MRI (mismatch)-based EVT confirmed

CTA, computed tomography angiography; CTP, computed tomography perfusion; MCA, middle cerebral artery; BBB, blood–brain barrier; RCT, randomized controlled trial. Please see Glossary for other abbreviations.

showed that the median ASPECTS value was 9, and that about one in seven patients showed ASPECTS of <6 on NCCT (44). Interestingly, the effect of rt-PA on functional outcome was not influenced by baseline ASPECTS, although patients with low ASPECTS have a substantially increased risk of thrombolysis-related parenchymal hemorrhage (44). In the DEFUSE 3 trial, there was no difference in the effect of EVT according to the ASPECTS score (<8 vs. ≥8) (20), which suggests that advanced image-guided selection could be considered in patients who have a low ASPECTS.

ADVANCES IN MRI TECHNIQUES FOR IMAGE PROCESSING AND INDIVIDUAL STROKE PATHOPHYSIOLOGY

There have been significant advances in MRI techniques in terms of availability, acquisition (scanning and post-processing) time, direct visualization of cardinal features (the 4 Cs; i.e., [tissue]-clock, clot, collaterals, and core), and machine learning-based algorithm implementation.

Availability

An NCCT scan is usually one of the first tests done in the evaluation of acute stroke. MRI takes longer and is

often not available under emergency conditions, while NCCT has advantages in terms of fast acquisition time, widespread availability, and ease of interpretation in an emergency setting. However, MRI is available in all comprehensive stroke centers, where EVT can be performed. One single-center study showed that MRI-based triage for EVT is feasible in terms of the scan-to-groin puncture time, with acceptable rates of poor outcome and symptomatic hemorrhage (45). A recent randomized trial (General or Local Anesthesia in Intra Arterial Therapy, GOLIATH) showed that MRI selection for endovascular therapy can be accomplished rapidly and within a similar time frame as computed tomography-based selection (46). The door-to-MRI time can be reduced by a quality improvement process (47). Moreover, although a comprehensive MRI protocol can be implemented in ~20 min, a fast MRI protocol can be implemented in about 6 min, rivaling the time of any comprehensive acute stroke CT protocol (48). This fast MRI protocol includes DWI, FLAIR, gradient echo, and MR angiography, and MRP. The CTP image requires additional imaging time (2–3 min) and post-processing time (5–15 min) (49).

For clinical use, automated software that allows fast post-processing is mandatory, and is increasingly being used in clinical trials. For example, the RAPid postprocess for Perfusion

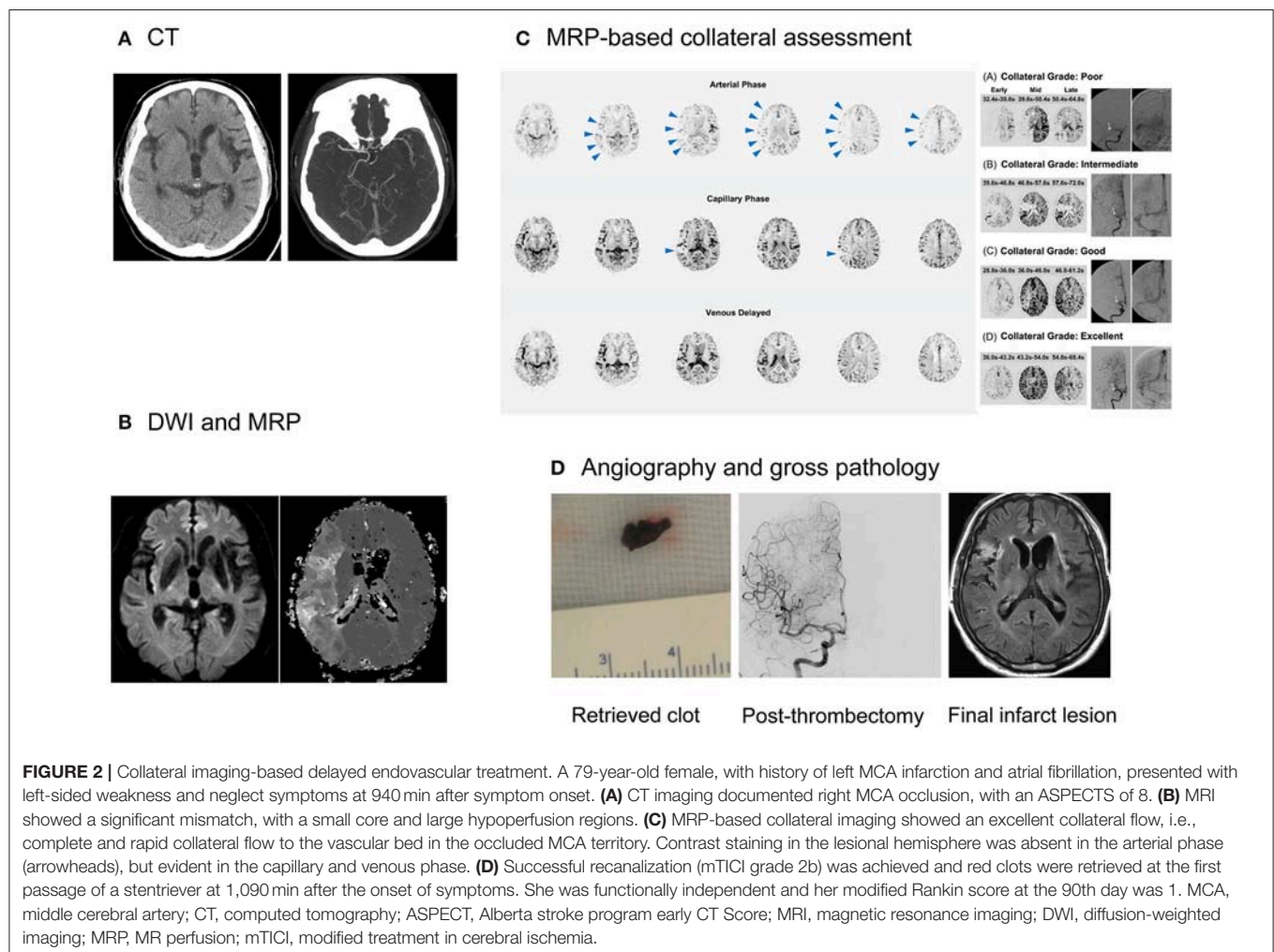
and Diffusion (RAPID; Rapid Software Corporation, Grapevine, TX, USA), an automated software package for performing quantitative evaluation of the apparent diffusion coefficient to estimate the ischemic core and an MRP threshold of $T_{\max} > 6$ s for defining critical hypoperfusion (50), has been used in several RCTs. Similarly, automated software for collateral assessment, the Fast Analysis SysTem for COLLaterals (FAST-COLL) (42), has been developed; it requires <5 min, and allows a clinical decision to be made at the workstation or bedside, based on the collateral grade.

Visualization of Individual Stroke Pathophysiology

Aside from demonstrating a perfusion-diffusion mismatch and delineating penumbral and irreversibly infarcted regions, MRI could provide additional information on the blood–brain barrier (BBB), collaterals, and clot. MRI techniques for determining the age of the infarct was mentioned above. Representative cases are presented in **Figures 2, 3**.

Collateral Flow

Better collaterals are associated with improved clinical and radiological outcomes, while poor collaterals are linked to hemorrhagic complications and poor recanalization rates after revascularization therapy for acute ischemic stroke (51). A collateral flow map derived from MRP source data can be generated by automatic post-processing (42). This study showed good correlation between MRI-based collateral grade and conventional angiography-based collateral grade, indicating that pretreatment MRI-based collateral evaluation could replace conventional angiographic evaluation in the angio-suite, which may require >20 min before EVT can be initiated. The role of MRI-based collateral imaging (FAST-COLL) is currently being tested in a prospective observational study (Clinicaltrials.identifier NCT02668627), to evaluate whether MRI-based collateral imaging is feasible and can predict the response to EVT in a wide range of patients with acute ischemic stroke. This study is evaluating the early infarct grow rate and eligibility according to the DAWN and DEFUSE 3 criteria, depending on the pretreatment MRI-based collateral grades. It is interesting that a significant proportion of patients who are



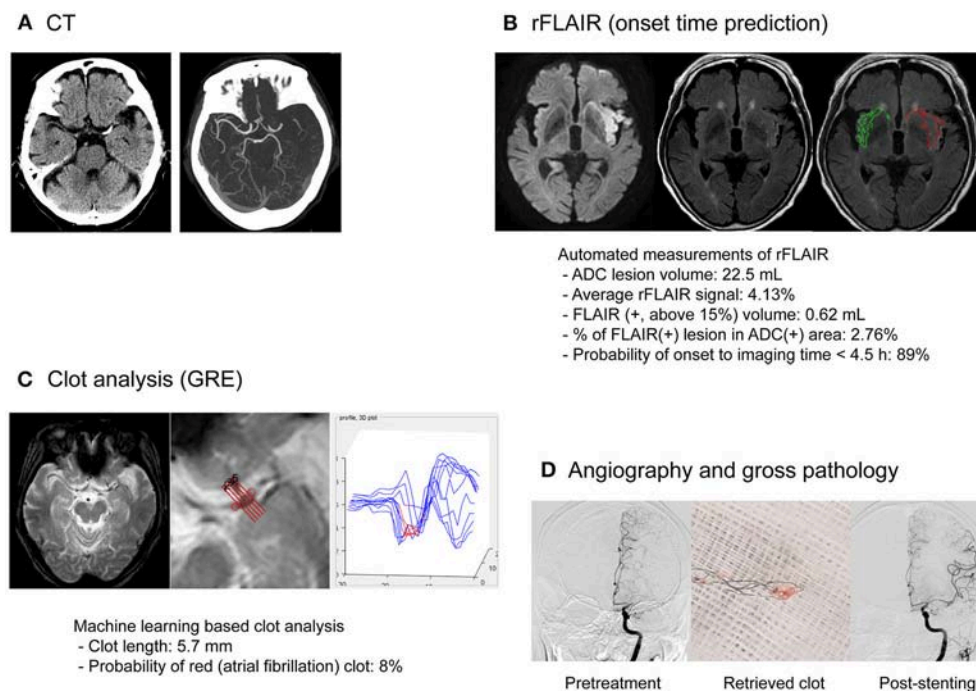


FIGURE 3 | MRI-based endovascular treatment of a stroke patient with unclear symptom onset time and atherosclerotic occlusion of the middle cerebral artery. An 81-year-old male presented with right-sided weakness and global aphasia, with an unclear onset time. **(A)** On the CT image, left MCA occlusion was documented with an ASPECTS of 8. The patient underwent multimodal MRI using a fast (9 min) MRI protocol. **(B)** Minimal signal change was observed on FLAIR imaging and the probability of stroke onset within 4.5 h, based on relative FLAIR signal analysis, was high. **(C)** Clot analysis using GRE imaging and machine-learning techniques showed a low possibility of atrial fibrillation as cause of the clot. **(D)** Whitish clots were retrieved during endovascular treatment. Successful recanalization (mTICI grade 2b) was achieved with rescue stenting after repetitive reocclusion with a stentriever. His modified Rankin score at the 90th day was 1. CT, computed tomography; MCA, middle cerebral artery, ASPECTS, Alberta Stroke Program Early CT Score; ADC, apparent diffusion coefficient; FLAIR, fluid attenuation inversion recovery; GRE, gradient echo; mTICI, modified treatment in cerebral ischemia.

eligible according to the DEFUSE 3 criteria did not meet the DAWN criteria for eligibility (20, 21).

Clot Treatment

Successful reperfusion may be associated with the histopathology of occlusive thrombi, including the existence of atheromatous gruel and the proportion of erythrocyte components (52). Intracranial atherosclerosis is particularly prevalent in Asians, and is associated with frequent EVT failure. In this condition, adjuvant therapy, such as the use of a GP IIb/IIIa inhibitor or permanent stent placement, may be needed (53). Although previous studies have attempted to predict the response to revascularization therapy using CT imaging of the clot, a recent study showed a lack of association between CT-based clot images and the histopathology of thrombi, and stroke etiology (54). MRI can identify clots with high specificity and can measure the clot burden more accurately than CT images. Blooming artifacts caused by paramagnetic materials in GRE or susceptible weighted images have been associated with cardioembolic stroke (55, 56).

BBB Derangement

GRE images are as accurate as CT at detecting acute hemorrhage in patients with acute stroke. BBB permeability dysfunction often precedes hemorrhagic transformation (57). Gadolinium

contrast agents are routinely used to detect BBB disruptions in patients with strokes or tumors. Using a simple post-processing algorithm that employs pretreatment MRP source data, MRI permeability images can visualize BBB dysfunction and identify patients at risk of hemorrhagic transformation, with high specificity (57). A multicenter study tested various MRP-derived permeability measures in acute stroke patients and showed that MRI permeability images may be used in clinical practice (58). The multicenter DEFUSE 2 and MR RESCUE trials showed that the amount of BBB disruption seen on pretreatment MRI is associated with the severity of intracranial bleeding after EVT (59, 60).

Machine Learning

Machine learning is an approach used to achieve artificial intelligence goals. The field of artificial intelligence has evolved significantly with the introduction of a number of sophisticated algorithms, some of which are capable of self-learning. Application of artificial intelligence in the stroke field is increasing, and is used in the prediction of stroke [e.g., risk factors (61) and fine particulate matter (PM_{2.5}) (62)] and pervasive health monitoring, by using smart monitoring devices embedded in the living environment (e.g., real-time monitoring via smartphone for adherence to oral anticoagulant treatment)

(63). It can be particularly helpful in decision-making in every step of EVT for acute ischemic stroke: in clinical and imaging recognition of acute ischemic stroke in the ambulance or emergency room (64), and in predicting the outcome after EVT (65).

Machine learning-based assessment has advantages over simple visual estimation as follows. First, the application of deep learning to create an algorithm for automated detection of abnormal neuroimaging findings can improve inter-rater correlation. For example, the NCCT ASPECTS is widely used worldwide and there have been efforts to increase the inter-rater reliability, including implementation of a teaching program. However, inter-rater reliability has been reported to be low, particularly in a hyperacute setting. One recent systematic review showed that, in patients considered for EVT, there may be insufficient agreement between clinicians for the ASPECTS to be used reliably as a criterion for treatment decisions (66). Although the agreement could be higher among stroke experts than stroke trainees, it may still be lower than that achieved with a machine learning-based decision algorithm developed using various neuroimaging data. A computer algorithm that would automatically read a brain CT scan and automatically generate an ASPECTS (e-ASPECTS) may improve reliability in terms of calculating the score. A multicenter trial has shown that e-ASPECTS was non-inferior to neuroradiologists in determining the ASPECTS score using NCCT images obtained from acute stroke patients (67).

Second, machine-learning techniques facilitate the merging of information from various MR sequences. Integration of information from several MRI sequences could improve the role of MRI-based triage for EVT. For example, although combining DWI and FLAIR data showed high predictive values for identifying patients within 4.5 h of symptom onset, adding information on collaterals or perfusion improved the accuracy of predicting the time from symptom onset within 4.5 h (68, 69). Combining data from quantitative image analyses with other types of data (e.g., clinical or laboratory) can provide models that are helpful in the choice of work up, prediction of outcome or response to treatment.

Finally, because deep learning uses numerous imaging features that are most predictive for certain types of stroke pathophysiology, rather than explicitly detecting clinical features with which stroke physicians are familiar (e.g., conventional DWI), both physicians and patients have to trust a “black box” to determine a disease state. With increased numbers of MR sequences, many variables may influence how a machine defines referable stroke pathophysiologies. These include heterogeneous populations comprising different races (with different stroke subtypes, along with normal

variations), heterogeneous prestroke conditions (e.g., preexisting atherosclerotic changes, white matter changes and other age-related changes) and novel features of specific sequences. Because the software does not detect the implication of the algorithm generated by the machine learning process, the stroke neurologist and radiologist need to provide the background and interpretation.

CONCLUSIONS AND PERSPECTIVES

CT has been the standard neuroimaging modality for determining whether EVT should be applied due to the limited availability and standardization and time required for acquisition or post-processing of MRI data. However, more information may be available within the permitted time-windows if rapid acquisition of MRI data, automated fast post-processing and machine learning-based decision algorithms can be implemented. The role of imaging in EVT may shift from “go/no go” to “how to go.” With the advances in transformative technologies (such as machine learning and artificial intelligence) along with a better understanding of stroke pathophysiology and MRI physics, it is highly likely that multimodal MRI information could guide treatment strategies for patients with acute ischemic stroke.

As technology becomes more complex, selection of the technology used becomes more important. Stroke physicians will need to understand the evolution of such technology. Appropriate RCTs are required to verify the usefulness of imaging-based algorithms for the selection of EVT before these techniques can be incorporated into routine clinical practice.

AUTHOR CONTRIBUTIONS

OB, the corresponding author, established the study idea, wrote the manuscript, and made critical revisions to the manuscript with substantive intellectual content. J-WC, JS, and Y-CK established the study idea, analyzed the data, and wrote the manuscript. W-SR, D-EK, W-KS, and G-MK established the database and made critical revisions to the manuscript with substantive intellectual content.

ACKNOWLEDGMENTS

This research was supported by grants from the Global Research Lab (GRL) program (NRF-2015K1A1A2028228) and the Bio-SPC program (NRF-2018M3A9G1024808) of the National Research Foundation, funded by the Korean government, Republic of Korea.

REFERENCES

1. National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. *N Engl J Med.* (1995) 333:1581–7. doi: 10.1056/NEJM199512143332401
2. Hacke W, Kaste M, Bluhmki E, Brozman M, Davalos A, Guidetti D, et al. Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke. *N Engl J Med.* (2008) 359:1317–29. doi: 10.1056/NEJMoa0804656
3. Albers GW, Thijs VN, Wechsler L, Kemp S, Schlaug G, Skalabrinn E, et al. Magnetic resonance imaging profiles predict clinical response to

- early reperfusion: the diffusion and perfusion imaging evaluation for understanding stroke evolution (DEFUSE) study. *Ann Neurol.* (2006) 60:508–17. doi: 10.1002/ana.20976
4. Davis SM, Donnan GA, Parsons MW, Levi C, Butcher KS, Peeters A, et al. Effects of alteplase beyond 3 h after stroke in the echoplanar imaging thrombolytic evaluation trial (EPITHET): a placebo-controlled randomised trial. *Lancet Neurol.* (2008) 7:299–309. doi: 10.1016/S1474-4422(08)70044-9
 5. Wintermark M, Albers GW, Alexandrov AV, Alger JR, Bammer R, Baron JC, et al. Acute stroke imaging research roadmap. *Stroke* (2008) 39:1621–8. doi: 10.1161/STROKEAHA.107.512319
 6. Latchaw RE, Alberts MJ, Lev MH, Connors JJ, Harbaugh RE, Higashida RT, et al. Recommendations for imaging of acute ischemic stroke: a scientific statement from the American Heart Association. *Stroke* (2009) 40:3646–78. doi: 10.1161/STROKEAHA.108.192616
 7. Broderick JP, Palesch YY, Demchuk AM, Yeatts SD, Khatri P, Hill MD, et al. Endovascular therapy after intravenous t-PA versus t-PA alone for stroke. *N Engl J Med.* (2013) 368:893–903. doi: 10.1056/NEJMoa1214300
 8. Kidwell CS, Jahan R, Gornbein J, Alger JR, Nenov V, Ajani Z, et al. A trial of imaging selection and endovascular treatment for ischemic stroke. *N Engl J Med.* (2013) 368:914–23. doi: 10.1056/NEJMoa1212793
 9. Ciccone A, Valassori L, Nichelatti M, Sgoifo A, Ponzio M, Sterzi R, et al. Endovascular treatment for acute ischemic stroke. *N Engl J Med.* (2013) 368:904–13. doi: 10.1056/NEJMoa1213701
 10. Wintermark M, Albers GW, Broderick JP, Demchuk AM, Fiebach JB, Fiehler J, et al. Acute stroke imaging research roadmap II. *Stroke* (2013) 44:2628–39. doi: 10.1161/STROKEAHA.113.002015
 11. Jauch EC, Saver JL, Adams HP Jr., Bruno A, Connors JJ, Demaerschalk BM, et al. Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* (2013) 44:870–947. doi: 10.1161/STR.0b013e318284056a
 12. Berkhemer OA, Fransen PS, Beumer D, van den Berg LA, Lingsma HF, Yoo AJ, et al. A randomized trial of intraarterial treatment for acute ischemic stroke. *N Engl J Med.* (2015) 372:11–20. doi: 10.1056/NEJMoa1411587
 13. Goyal M, Demchuk AM, Menon BK, Eesa M, Rempel JL, Thornton J, et al. Randomized assessment of rapid endovascular treatment of ischemic stroke. *N Engl J Med.* (2015) 372:1019–30. doi: 10.1056/NEJMoa1414905
 14. Campbell BC, 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
 15. Saver JL, Goyal M, Bonafe A, Diener HC, Levy EI, Pereira VM, et al. Stent-retriever thrombectomy after intravenous t-PA vs. t-PA alone in stroke. *N Engl J Med.* (2015) 372:2285–95. doi: 10.1056/NEJMoa1415061
 16. 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
 17. Warach SJ, Luby M, Albers GW, Bammer R, Bivard A, Campbell BC, et al. Acute stroke imaging research roadmap III Imaging selection and outcomes in acute stroke reperfusion clinical trials: consensus recommendations and further research priorities. *Stroke* (2016) 47:1389–98. doi: 10.1161/STROKEAHA.115.012364
 18. Powers WJ, Derdeyn CP, Biller J, Coffey CS, Hoh BL, Jauch EC, et al. American Heart Association/American Stroke Association focused update of the 2013 guidelines for the early management of patients with acute ischemic stroke regarding endovascular treatment: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* (2015) 46:3020–35. doi: 10.1161/STR.0000000000000074
 19. 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
 20. 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
 21. Albers GW. Late window paradox. *Stroke* (2018) 49:768–71. doi: 10.1161/STROKEAHA.117.020200
 22. Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K, et al. 2018 guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* (2018) 49:e46–110. doi: 10.1161/STR.0000000000000158
 23. Chia NH, Leyden JM, Newbury J, Jannes J, Kleinig TJ. Determining the number of ischemic strokes potentially eligible for endovascular thrombectomy: a population-based study. *Stroke* (2016) 47:1377–80. doi: 10.1161/STROKEAHA.116.013165
 24. Molina CA. Futile recanalization in mechanical embolectomy trials: a call to improve selection of patients for revascularization. *Stroke* (2010) 41:842–3. doi: 10.1161/STROKEAHA.110.580266
 25. Kohrmann M, Schellinger PD. Acute stroke triage to intravenous thrombolysis and other therapies with advanced CT or MR imaging: pro MR imaging. *Radiology* (2009) 251:627–33. doi: 10.1148/radiol.2513081074
 26. Leslie-Mazwi TM, Hirsch JA, Falcone GJ, Schaefer PW, Lev MH, Rabinov JD, et al. Endovascular stroke treatment outcomes after patient selection based on magnetic resonance imaging and clinical criteria. *JAMA Neurol.* (2016) 73:43–9. doi: 10.1001/jamaneurol.2015.3000
 27. Wisco D, Uchino K, Saqqur M, Gebel JM, Aoki J, Alam S, et al. Addition of hyperacute MRI AIDS in patient selection, decreasing the use of endovascular stroke therapy. *Stroke* (2014) 45:467–72. doi: 10.1161/STROKEAHA.113.003880
 28. Mackey J, Kleindorfer D, Sucharew H, Moomaw CJ, Kissela BM, Alwell K, et al. Population-based study of wake-up strokes. *Neurology* (2011) 76:1662–7. doi: 10.1212/WNL.0b013e318219fb30
 29. Elliott WJ. Circadian variation in the timing of stroke onset: a meta-analysis. *Stroke* (1998) 29:992–6.
 30. Serena J, Davalos A, Segura T, Mostacero E, Castillo J. Stroke on awakening: looking for a more rational management. *Cerebrovasc Dis.* (2003) 16:128–33. doi: 10.1159/000070592
 31. Moradiya Y, Janjua N. Presentation and outcomes of “wake-up strokes” in a large randomized stroke trial: analysis of data from the International Stroke Trial. *J Stroke Cerebrovasc Dis.* (2013) 22:e286–92. doi: 10.1016/j.jstrokecerebrovasdis.2012.07.016
 32. Reid JM, Dai D, Cheripelli B, Christian C, Reidy Y, Gubitz GJ, et al. Differences in wake-up and unknown onset stroke examined in a stroke registry. *Int J Stroke.* (2015) 10:331–5. doi: 10.1111/ijis.12388
 33. Thomalla G, Rossbach P, Rosenkranz M, Siemonsen S, Krutzmann A, Fiehler J, et al. Negative fluid-attenuated inversion recovery imaging identifies acute ischemic stroke at 3 hours or less. *Ann Neurol.* (2009) 65:724–32. doi: 10.1002/ana.21651
 34. Thomalla G, Cheng B, Ebinger M, Hao Q, Tournias T, Wu O, et al. DWI-FLAIR mismatch for the identification of patients with acute ischaemic stroke within 4.5 h of symptom onset (PRE-FLAIR): a multicentre observational study. *Lancet Neurol.* (2011) 10:978–86. doi: 10.1016/S1474-4422(11)70192-2
 35. Thomalla G, Simonsen CZ, Boutitie F, Andersen G, Berthezene Y, Cheng B, et al. MRI-guided thrombolysis for stroke with unknown time of onset. *N Engl J Med.* (2018). doi: 10.1056/NEJMoa1804355. [Epub ahead of print].
 36. Lichtman JH, Watanabe E, Allen NB, Jones SB, Dostal J, Goldstein LB. Hospital arrival time and intravenous t-PA use in US Academic Medical Centers, 2001–2004. *Stroke* (2009) 40:3845–50. doi: 10.1161/STROKEAHA.109.562660
 37. Goyal M, Menon BK, van Zwam WH, Dippel DW, Mitchell PJ, Demchuk AM, et al. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *Lancet* (2016) 387:1723–31. doi: 10.1016/S0140-6736(16)00163-X
 38. Saver JL, Goyal M, van der Lugt A, Menon BK, Majoie CB, Dippel DW, et al. Time to treatment with endovascular thrombectomy and outcomes from ischemic stroke: a meta-analysis. *JAMA* (2016) 316:1279–88. doi: 10.1001/jama.2016.13647
 39. Galimanis A, Jung S, Mono ML, Fischer U, Findling O, Weck A, et al. Endovascular therapy of 623 patients with anterior circulation stroke. *Stroke* (2012) 43:1052–7. doi: 10.1161/STROKEAHA.111.639112
 40. Ribo M, Flores A, Rubiera M, Pagola J, Sargento-Freitas J, Rodriguez-Luna D, et al. Extending the time window for endovascular procedures according to collateral pial circulation. *Stroke* (2011) 42:3465–9. doi: 10.1161/STROKEAHA.111.623827
 41. Hwang YH, Kang DH, Kim YW, Kim YS, Park SP, Liebeskind DS. Impact of time-to-reperfusion on outcome in patients with poor collaterals. *Am J Neuroradiol.* (2015) 36:495–500. doi: 10.3174/ajnr.A4151

42. Kim SJ, Son JP, Ryoo S, Lee MJ, Cha J, Kim KH, et al. A novel magnetic resonance imaging approach to collateral flow imaging in ischemic stroke. *Ann Neurol.* (2014) 76:356–69. doi: 10.1002/ana.24211
43. Son JP, Lee MJ, Kim SJ, Chung JW, Cha J, Kim GM, et al. Impact of slow blood filling via collaterals on infarct growth: comparison of mismatch and collateral status. *J Stroke.* (2017) 19:88–96. doi: 10.5853/jos.2016.00955
44. Dzialowski I, Hill MD, Coutts SB, Demchuk AM, Kent DM, Wunderlich O, et al. Extent of early ischemic changes on computed tomography (CT) before thrombolysis: prognostic value of the Alberta Stroke Program Early CT Score in ECASS II. *Stroke* (2006) 37:973–8. doi: 10.1161/01.STR.0000206215.62441.56
45. Simonsen CZ, Sorensen LH, Karabegovic S, Mikkelsen IK, Schmitz ML, Juul N, et al. MRI before intraarterial therapy in ischemic stroke: feasibility, impact, and safety. *J Cereb Blood Flow Metab.* (2014) 34:1076–81. doi: 10.1038/jcbfm.2014.57
46. Simonsen CZ, Yoo AJ, Rasmussen M, Sorensen KE, Leslie-Mazwi T, Andersen G, et al. Magnetic resonance imaging selection for endovascular stroke therapy: workflow in the GOLIATH trial. *Stroke* (2018) 49:1402–6. doi: 10.1161/STROKEAHA.118.021038
47. Shah S, Luby M, Poole K, Morella T, Keller E, Benson RT, et al. Screening with MRI for accurate and rapid stroke treatment: SMART. *Neurology* (2015) 84:2438–44. doi: 10.1212/WNL.0000000000001678
48. Nael K, Khan R, Choudhary G, Meshksar A, Villablanca P, Tay J, et al. Six-minute magnetic resonance imaging protocol for evaluation of acute ischemic stroke: pushing the boundaries. *Stroke* (2014) 45:1985–91. doi: 10.1161/STROKEAHA.114.005305
49. Lee JS, Demchuk AM. Choosing a hyperacute stroke imaging protocol for proper patient selection and time efficient endovascular treatment: lessons from recent trials. *J Stroke.* (2015) 17:221–8. doi: 10.5853/jos.2015.17.3.221
50. Straka M, Albers GW, Bammer R. Real-time diffusion-perfusion mismatch analysis in acute stroke. *J Magn Reson Imaging.* (2010) 32:1024–37. doi: 10.1002/jmri.22338
51. Bang OY, Goyal M, Liebeskind DS. Collateral circulation in ischemic stroke: assessment tools and therapeutic strategies. *Stroke* (2015) 46:3302–9. doi: 10.1161/STROKEAHA.115.010508
52. Hashimoto T, Hayakawa M, Funatsu N, Yamagami H, Satow T, Takahashi JC, et al. Histopathologic analysis of retrieved thrombi associated with successful reperfusion after acute stroke thrombectomy. *Stroke* (2016) 47:3035–7. doi: 10.1161/STROKEAHA.116.015228
53. Baek JH, Kim BM, Kim DJ, Heo JH, Nam HS, Song D, et al. Importance of truncal-type occlusion in stentriever-based thrombectomy for acute stroke. *Neurology* (2016) 87:1542–50. doi: 10.1212/WNL.0000000000003202
54. Brinjikji W, Duffy S, Burrows A, Hacke W, Liebeskind D, Majoie CB, et al. Correlation of imaging and histopathology of thrombi in acute ischemic stroke with etiology and outcome: a systematic review. *J Neurointerv Surg.* (2016) 9:529–34. doi: 10.1136/neurintsurg-2016-012391
55. Kang DW, Jeong HG, Kim DY, Yang W, Lee SH. Prediction of stroke subtype and recanalization using susceptibility vessel sign on susceptibility-weighted magnetic resonance imaging. *Stroke* (2017) 48:1554–9. doi: 10.1161/STROKEAHA.116.016217
56. Zhang R, Zhou Y, Liu C, Zhang M, Yan S, Liebeskind DS, et al. Overestimation of susceptibility vessel sign: a predictive marker of stroke cause. *Stroke* (2017) 48:1993–6. doi: 10.1161/STROKEAHA.117.016727
57. Bang OY, Buck BH, Saver JL, Alger JR, Yoon SR, Starkman S, et al. Prediction of hemorrhagic transformation after recanalization therapy using T2*-permeability magnetic resonance imaging. *Ann Neurol.* (2007) 62:170–6. doi: 10.1002/ana.21174
58. Scalzo F, Alger JR, Hu X, Saver JL, Dani KA, Muir KW, et al. Multi-center prediction of hemorrhagic transformation in acute ischemic stroke using permeability imaging features. *Magn Reson Imaging* (2013) 31:961–9. doi: 10.1016/j.mri.2013.03.013
59. Leigh R, Christensen S, Campbell BC, Marks MP, Albers GW, Lansberg MG, et al. Pretreatment blood-brain barrier disruption and post-endovascular intracranial hemorrhage. *Neurology* (2016) 87:263–9. doi: 10.1212/WNL.0000000000002862
60. Nael K, Knitter JR, Jahan R, Gornbein J, Ajani Z, Feng L, et al. Multiparametric magnetic resonance imaging for prediction of parenchymal hemorrhage in acute ischemic stroke after reperfusion therapy. *Stroke* (2017) 48:664–70. doi: 10.1161/STROKEAHA.116.014343
61. Weng SF, Reys J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS ONE* (2017) 12:e0174944. doi: 10.1371/journal.pone.0174944
62. Ong BT, Sugiura K, Zettsu K. Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for predicting PM2.5. *Neural Comput Appl.* (2016) 27:1553–66. doi: 10.1007/s00521-015-1955-3
63. Labovitz DL, Shafner L, Reyes Gil M, Virmani D, Hanina A. Using artificial intelligence to reduce the risk of nonadherence in patients on anticoagulation therapy. *Stroke* (2017) 48:1416–9. doi: 10.1161/STROKEAHA.116.016281
64. Abedi V, Goyal N, Tsivgoulis G, Hosseinichimeh N, Hontecillas R, Bassaganya-Riera J, et al. Novel screening tool for stroke using artificial neural network. *Stroke* (2017) 48:1678–81. doi: 10.1161/STROKEAHA.117.017033
65. Asadi H, Dowling R, Yan B, Mitchell P. Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy. *PLoS ONE* (2014) 9:e88225. doi: 10.1371/journal.pone.0088225
66. Farzin B, Fahed R, Guilbert F, Poppe AY, Daneault N, Durocher AP, et al. Early CT changes in patients admitted for thrombectomy: intrarater and interrater agreement. *Neurology* (2016) 87:249–56. doi: 10.1212/WNL.0000000000002860
67. Nagel S, Sinha D, Day D, Reith W, Chapot R, Papanagiotou P, et al. e-ASPECTS software is non-inferior to neuroradiologists in applying the ASPECT score to computed tomography scans of acute ischemic stroke patients. *Int J Stroke* (2017) 12:615–22. doi: 10.1177/1747493016681020
68. Wouters A, Dupont P, Christensen S, Norrving B, Laage R, Thomalla G, et al. Association between time from stroke onset and fluid-attenuated inversion recovery lesion intensity is modified by status of collateral circulation. *Stroke* (2016) 47:1018–22. doi: 10.1161/STROKEAHA.115.012010
69. Wouters A, Dupont P, Norrving B, Laage R, Thomalla G, Albers GW, et al. Prediction of stroke onset is improved by relative fluid-attenuated inversion recovery and perfusion imaging compared to the visual diffusion-weighted imaging/fluid-attenuated inversion recovery mismatch. *Stroke* (2016) 47:2559–64. doi: 10.1161/STROKEAHA.116.013903

Conflict of Interest Statement: 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.

Copyright © 2018 Bang, Chung, Son, Ryu, Kim, Seo, Kim and Kim. 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.



Optimal Delay Time of CT Perfusion for Predicting Cerebral Parenchymal Hematoma After Intra-Arterial tPA Treatment

Bing Wu^{1†}, Nan Liu^{2†}, Max Wintermark³, Mark W. Parsons^{4,5}, Hui Chen², Longting Lin⁴, Shuai Zhou^{1,6}, Gang Hu¹, Yongwei Zhang⁷, Jun Hu⁸, Ying Li², Zihua Su⁹, Xinhui Wu^{1,6*} and Guangming Zhu^{3*}

OPEN ACCESS

Edited by:

Fabien Scalzo,
University of California, Los Angeles,
United States

Reviewed by:

Asaf Honig,
University of British Columbia, Canada
Li Xiong,
Harvard Medical School,
United States

*Correspondence:

Guangming Zhu
zhugmdc@aliyun.com
Xinhui Wu
wuxinhui_beijing@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: 04 May 2018

Accepted: 27 July 2018

Published: 21 August 2018

Citation:

Wu B, Liu N, Wintermark M, Parsons MW, Chen H, Lin L, Zhou S, Hu G, Zhang Y, Hu J, Li Y, Su Z, Wu X and Zhu G (2018) Optimal Delay Time of CT Perfusion for Predicting Cerebral Parenchymal Hematoma After Intra-Arterial tPA Treatment. *Front. Neurol.* 9:680. doi: 10.3389/fneur.2018.00680

¹ Department of Radiology, PLA Army General Hospital, Beijing, China, ² Department of Neurology, PLA Army General Hospital, Beijing, China, ³ Neuroradiology Section, Department of Radiology, Stanford University, Stanford, CA, United States, ⁴ School of Medicine and Public Health, University of Newcastle, Newcastle, NSW, Australia, ⁵ Department of Neurology, John Hunter Hospital, University of Newcastle, Newcastle, NSW, Australia, ⁶ Inner Mongolia Medical University, Hohhot, China, ⁷ Department of Neurology, Changhai Hospital, Second Military Medical University, Shanghai, China, ⁸ Department of Neurology, Southwest Hospital, Third Military Medical University, Chongqing, China, ⁹ GE Healthcare, Beijing, China

Background and Purpose: Cerebral hemorrhage is a serious potential complication of stroke revascularization, especially in patients receiving intra-arterial tissue-type plasminogen activator (tPA) therapy. We investigated the optimal pre-intervention delay time (DT) of computed tomography perfusion (CTP) measurement to predict cerebral parenchymal hematoma (PH) in acute ischemic stroke (AIS) patients after intra-arterial tissue plasminogen activator (tPA) treatment.

Methods: The study population consisted of a series of patients with AIS who received intra-arterial tPA treatment and had CTP and follow-up computed tomography/magnetic resonance imaging (CT/MRI) to identify hemorrhagic transformation. The association of increasing DT thresholds (>2, >4, >6, >8, and >10 s) with PH was examined using receiver operating characteristic (ROC) analysis and logistic regression.

Results: Of 94 patients, 23 developed PH on follow-up imaging. Receiver operating characteristic analysis revealed that the greatest area under the curve for predicting PH occurred at DT > 4 s (area under the curve, 0.66). At this threshold of > 4 s, DT lesion volume ≥ 30.85 mL optimally predicted PH with 70% sensitivity and 59% specificity. DT > 4 s volume was independently predictive of PH in a multivariate logistic regression model ($P < 0.05$).

Conclusions: DT > 4 s was the parameter most strongly associated with PH. The volume of moderate, not severe, hypo-perfusion on DT is more strongly associated and may allow better prediction of PH after intra-arterial tPA thrombolysis.

Keywords: stroke, hemorrhage transformation, CT scan, perfusion imaging, delay time

INTRODUCTION

While significant advances have been made regarding emergent treatment of acute ischemic stroke (AIS), the different therapeutic revascularization options remain associated with an increased risk of hemorrhage. While intravenous tissue-type plasminogen activator (tPA) is effective at reanalyzing more distal thrombi (1, 2), endovascular reperfusion therapy in a 6 or 12 h time window, can be effective for those patients with more proximal intracranial artery occlusion (3–6). The revascularization increases the risk of Hemorrhagic transformation (HT) (7–9), and HT is nearly 5 times more common for patients receiving intravenous thrombolysis compared to controls, and even more for patients receiving intra-arterial thrombolysis (10). Symptomatic intracranial hemorrhage (sICH) transformation or cerebral parenchymal hematoma (PH) is the most serious complication after revascularization therapy. The identification and possible exclusion of patients at high risk for sICH or PH will significantly reduce the complication rate (11).

Several clinical risk factors, such as age, diabetes mellitus, infarct volume, and anticoagulant or antiplatelet therapy are associated with HT. Computed tomography (CT) and magnetic resonance imaging (MRI), including perfusion imaging, provide detailed assessment of acute stroke pathophysiology, and has led to the establishment of imaging predictive parameters for HT after thrombolysis. Relative cerebral blood flow (rCBF), relative mean transit time, cerebral blood volume (CBV), Tmax, delay time (DT), and permeability parameters have been found to be associated with hemorrhagic transformation (12–16). A prior study by Yassi et al (15) showed that extremely long Tmax was independently predictive of PH for the patient with or without thrombolysis. However, these studies focused on patients who received intravenous tPA, rather than on patients who received intra-arterial tPA treatment.

In this study, we sought to identify the optimal pre-intervention DT parameter for prediction of PH after AIS intra-arterial tPA therapy.

METHODS

Patients

In this study, the clinical and imaging data were obtained from 3 participating institutions: the PLA Army General Hospital, Beijing; Changhai Hospital, Shanghai; and Southwest Hospital, Chongqing. All data contributed to the study were completely anonymized. The institutional review boards of the three institutions approved the study. Consecutive patients with signs and symptoms suggesting hemispheric stroke from January 2011 to January 2014 were retrospectively identified. Inclusion criteria were as follows: (1) AIS with occlusion of the M1 segment of the middle cerebral artery, the internal carotid, or both; (2) an admission National Institutes of Health Stroke Scale (NIHSS) score between 4 and 22; (3) with CT imaging indicating stroke, including non-contrast-CT, CT angiography (CTA), and CT perfusion (CTP), upon admission; (4) intra-arterial tPA thrombolysis with <12 h from onset; (5) availability of MRIs or

CT scans taken within 7 days after therapy to assess HT. The patients who received IV-tPA were not included in the study, considering data consistency. A flow chart delineating patient selection is shown in **Figure 1**. The demographic and clinical variables were recorded as follows: age, sex, medical history, vascular risk factors, routine blood tests, time from onset to imaging, time from symptom onset to treatment, NIHSS score upon admission, and modified Rankin Score (mRS) at 90 days. The mRS was assessed in the outpatient clinic or by the telephone and the death was coded as 6. Stroke mechanisms were subtyped using the TOAST (Trial of Org 10172 in Acute Stroke Treatment) classification and were diagnosed by 2 stroke neurologists (N.L. and H.C.) in consensus.

Imaging Protocol

CTP studies were all obtained on 64-slice CT scanners. Each CTP study involved successive gantry rotations performed in cine mode with 45 time-points acquired each 1.33 s (total acquisition, 60 s), with intravenous administration of 40–50 mL of iodinated contrast material (Ultravist 370; Bayer HealthCare; Berlin, Germany) at an injection rate of 4–5 mL/s followed by a 40-mL saline push. Total CTP coverage was 40 mm. Acquisition parameters were 80 kVp and 100 mAs.

Digital subtraction angiography (DSA) was performed using a biplane cerebral angiographic system. Images were acquired during injection of the internal and external carotid arteries and ≥ 1 vertebral artery. Imaging was performed through the entire arterial and venous phases to evaluate the collateral circulation. All patients underwent intra-arterial tPA thrombolysis without mechanical embolectomy at the discretion of the attending neurologist (Y.Z., J.H., and G.Z.).

Imaging Processing

All perfusion CT data were analyzed with the commercial software (MiStar, Apollo Medical Imaging Technology) (17, 18). Perfusion data were processed using a single value deconvolution algorithm with delay and dispersion correction. The actual delay time (DT) was calculated by a modified singular value deconvolution approach by looping through a series of DT values (19). The cerebral blood flow and cerebral blood volume were determined by the peak height and area under the curve of the input residue function. Arterial input function and venous outflow function were automatically selected by the software from the non-stroke middle cerebral artery/anterior cerebral artery and superior sagittal sinus, respectively. The volume of increasing DT thresholds (2, >4, >6, >8, and >10 s), the relative cerebral blood flow <40% within the DT >3 s, and cerebral blood volume < 2 mL/100 g within the DT >3 s, were automatically calculated (17).

Receiver operating characteristic (ROC) analysis was performed using PH as the outcome variable, and the lesion volume was defined using a particular DT, CBV, or rCBF threshold in each individual patient. These thresholds were iterated across the range of values present in the data to determine the threshold for each parameter that generated the highest area under the curve (AUC). This optimal threshold was taken forward in the analysis to compare the sensitivity

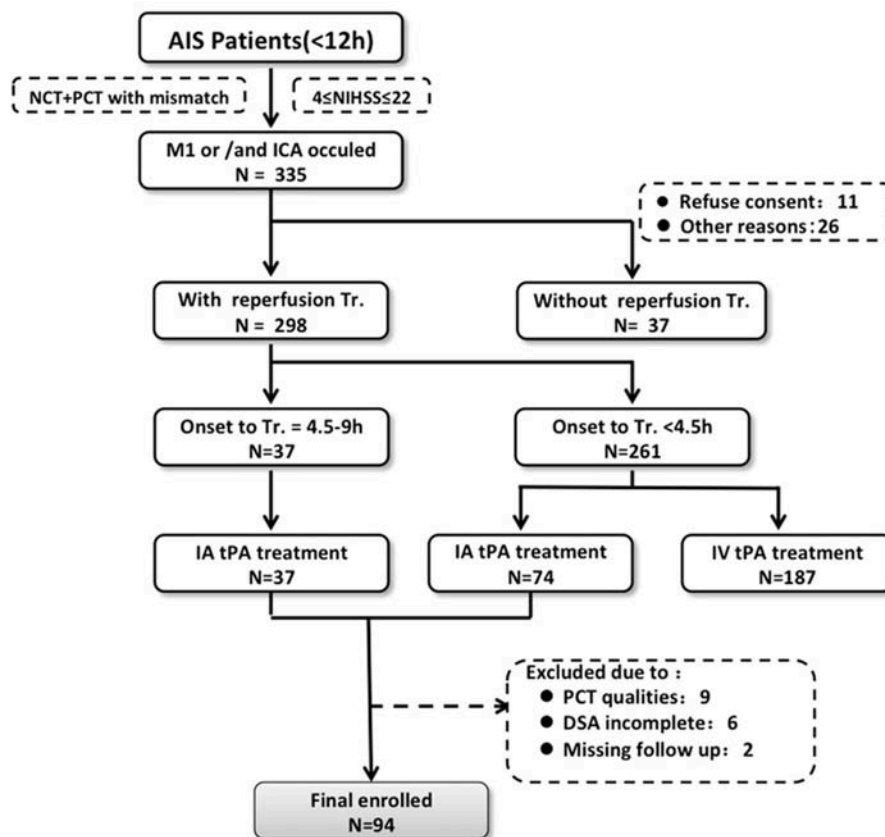


FIGURE 1 | Flow chart outlining patient selection and exclusion criteria. AIS indicates acute ischemic stroke; DSA, digital subtraction angiography; IA, intra-arterial; IV tPA, intravenous tissue-type plasminogen activator; NCT, noncontrast-CT; NIHSS, National Institutes of Health Stroke Scale; CTP, computed tomography perfusion; and Tr, treatment.

and specificity of each perfusion parameter. The Youden Index (sensitivity+specificity–1) was then calculated for this optimized threshold to determine the optimal volume of rCBF, CBV, and DT to predict the development of PH.

It has previously been demonstrated that poor baseline collaterals and successful therapeutic recanalization may result in clinically significant hemorrhagic complications (20). The angiographic collateral was scaled as follows: grade 0 (no collaterals visible to the ischemic site), 1 (slow collaterals to the periphery of the ischemic site with persistence of some of the defects), 2 (rapid collaterals to the periphery of ischemic site with persistence of some of the defects and to only a portion of the ischemic territory), 3 (collaterals with slow but complete angiographic blood flow of the ischemic bed by the late venous phase), and 4 (complete and rapid collateral blood flow to the vascular bed in the entire ischemic territory by retrograde perfusion). Vascular reperfusion was graded based on the Thrombolysis in Cerebral Infarction (TICI) classification: 0 (no perfusion), 1 (penetration with minimal perfusion), 2a (less than 67% perfusion); 2b (more than 67% perfusion), and 3 (complete perfusion of the affected vascular territory). Reperfusion status was classified as ER+ (positive for early reperfusion; TICI score 2b to 3 within 12 h of symptom onset)

or ER–. The angiographic collateral and TICI scores were rated by consensus of a neurologist (G.Z.) and a neuroradiologist (B.W.). The influence of collaterals and recanalization on PH was analyzed in distinct case scenarios relative to baseline collateral grade at angiography (0–1 vs. 2–4) and recanalization (TICI scale, ER+ vs. ER–): (1) good collaterals and no recanalization, (2) poor collaterals and no recanalization, (3) good collaterals and successful recanalization, and (4) poor collaterals with successful recanalization.

Outcome Measurement

All patients received follow-up CT or MRI as part of the routine. However the time interval depended on the patient's condition. Follow-up imaging (MRI or CT within 7 days) was independently assessed for hemorrhagic transformation by 2 stroke neurologists (N.L. and H.C.), who then reached consensus using the European Cooperative Acute Stroke Study (ECASS) scoring system (21). This classifies hemorrhagic infarction 1 (HI1) as small petechiae along the periphery of the infarct region, hemorrhagic infarction 2 (HI2) as confluent petechiae within the infarct, without space-occupying effect, parenchymal hemorrhage 1 (PH1) as bleeding ≤30% of the infarcted area, with mild space-occupying effect, and parenchymal hemorrhage 2 (PH2) as bleeding >30% of the

infarcted area, with space-occupying effect. PH included both PH1 and PH2. The readers of neuroradiological imaging were blinded to all clinical data.

Statistical Analyses

Statistical analysis was performed using commercially available IBM SPSS Statistics (version 22.0.0.0; IBM Corp, Armonk, NY). The Mann–Whitney U–test was used to determine the association between baseline clinical characteristics or imaging parameters with parenchymal hemorrhage outcome. ROC analysis was performed to determine the optimal CTP parameter for prediction of PH. Sensitivity, specificity, positive predictive value, negative predictive value, and likelihood ratios were determined for each parameter at different volumetric thresholds. The best performing CTP parameters in ROC analysis were subsequently tested in a multivariate logistic regression model including age, baseline NIHSS score, and poor collaterals and recanalization. The other variables were tested in a univariate model logistic regression. The variable would be added to the multivariate model when it was statistically significant in univariate model. For perfusion lesions, we divided the lesion volumes into 4 groups according to interquartile cutoff points of the distribution of volumes of perfusion delay. Correlation between DT >2 s, >4 s, >6 s, >8 s, and >10 s were also analyzed with the correlation coefficient matrix. $P < 0.05$ was considered statistically significant.

RESULTS

Of 111 consecutive ischemic stroke patients imaged with multi-modal CT and intra-arterial thrombolysis treatment, 94 were included in the analysis. The reasons for exclusion were severely motion-degraded perfusion data ($n = 9$), incomplete DSA imaging ($n = 6$), and 2 patients could not be contacted during follow-up. Follow-up MRI and CT were performed in 75 and 19 patients, respectively. PH occurred in 23 patients overall (24.5%). **Table 1** shows baseline clinical characteristics, stroke risk factors, and time to imaging for the study patients.

Male percent, atrial fibrillation, hyperlipidemia, ASPECTS on NCT, collateral flow score, and volume of CBV <2 mL/100 g were significantly different among types of HT, but not associated with PH, as shown in **Table 1**. The volume of DT >4 s was significantly associated with PH.

The ROC analysis for association of PH across the range of values in each CTP parameter identified DT >4 s as the optimal threshold for further analysis (area under the curve = 0.657; $P = 0.024$), followed by rCBF <40% (area under the curve = 0.587; $P = 0.212$), and CBV <2 mL/100 g (area under the curve = 0.523; $P = 0.742$). Mean DT >4 s volume was 27.7 mL (interquartile range [IQR], 17.0–40.0) in the no-PH group and 36.6 mL (IQR, 25.3–51.7) in the PH group ($P = 0.024$; Mann–Whitney U test) (**Table 2**).

Based on the ROC analysis and Youden Index, the optimal volume of DT >4 s for the association with PH was ≥ 30.4 mL. There were 49 of 94 (52%) patients with DT >4 s volumes of <30.4 mL, of whom 7 (14.3%) patients developed PH compared with the overall rate of 24.5%. This indicated a low risk of

TABLE 1 | Demographic parameters and other relevant information ($n = 94$).

Clinical and imaging characteristics	No PH	PH	P-value
	N = 71	N = 23	
Male, %	40 (56.3)	14 (60.9)	0.8101
Age, years	68.0 (53.0–77.0)	68.0 (55.0–73.0)	0.8156
NIHSS score at admission	16.0 (11.0–19.0)	17.0 (13.0–19.0)	0.2254
Hypertension, %	44 (62.0)	14 (60.9)	0.9247
Diabetes mellitus, %	10 (14.1)	4 (17.4)	0.7398
Hyperlipidemia, %	24 (33.8)	3 (13.0)	0.0666
Atrial fibrillation, %	19 (27.1)	3 (13.0)	0.2581
CAD, %	15 (22.1)	3 (13.0)	0.5456
Current statin administration, %	19 (26.8)	3 (13.0)	0.2587
Stroke mechanism			
Cardioembolic stroke, %	23 (32.4)	5 (21.7)	0.1644
Large artery disease, %	20 (28.2)	12 (52.2)	
Other type of stroke, %	10 (14.1)	1 (4.4)	
Undetermined	18 (25.4)	5 (21.7)	
Categories, %			
Time from onset to CT imaging, hours	5.5 (3.5–7.8)	6.0 (4.8–11.5)	0.3011
ASPECTS score on NCT	8.0 (7.0–9.0)	7.0 (6.0–9.0)	0.1003
HMCAS on NCT, %	35 (55.6)	17 (73.9)	0.1425
Collateral Flow Scores	2.0 (1.0–3.0)	2.0 (2.0–3.0)	0.3106
DT >2 s, mL	55.7 (47.0–70.7)	59.8 (46.2–70.2)	0.4442
DT >4 s, mL	27.7 (17.0–40.0)	36.6 (25.3–51.7)	0.0243*
DT >6 s, mL	11.1 (2.2–21.3)	17.6 (8.1–34.2)	0.054
DT >8 s, mL	5.1 (0.3–10.1)	6.7 (2.5–19.9)	0.0854
DT >10 s, mL	1.6 (0.0–4.6)	2.2 (0.3–10.6)	0.1157
rCBF <40%, mL	16.2 (4.3–23.4)	16.9 (6.1–30.5)	0.2132
CBV <2, mL	17.2 (3.5–32.3)	21.9 (5.1–32.5)	0.7447
Site of occlusion			
ICA, %	5 (7.0)	3 (13.0)	0.2687
M1, %	53 (74.7)	13 (56.5)	
ICA & M1, %	13 (18.3)	7 (30.4)	
Successful recanalization (TIMI:2–3)	57 (80.3)	19 (82.6)	0.8053

ASPECTS indicates Alberta Stroke Program Early CT Score; CAD, coronary artery disease; CT, computed tomography; HMCAS, hyperdense middle cerebral artery sign; NCT, noncontrast-CT; NIHSS, National Institutes of Health Stroke Scale; DT, delay time; PH, parenchymal hematoma; rCBF, relative cerebral blood flow; CBV, cerebral blood volume; ICA, internal carotid artery; M1, middle cerebral artery; TIMI, thrombolysis in myocardial infarct. * $P < 0.05$.

PH in this group, with a negative predictive value of 0.86 (95% confidence interval, 0.73–0.94) and a negative likelihood ratio of 0.51 (0.27–0.98). For DT >4 s volumes ≥ 30.4 mL, the sensitivity for PH was 0.70 (0.49–0.84), the specificity was 0.59 (0.48–0.70), and the positive likelihood ratio was 1.70 (1.15–2.51).

A large area of moderate, not severe, perfusion delay (DT >4 s) on the pretreatment CTP were independently associated with PH; compared with patients with the lowest DT >4 s volume quartile, those with second, third, and fourth

TABLE 2 | Receiver operating characteristic analysis.

	PH		
	AUC	95% CI	P-value
DT > 2 s	0.554	0.415–0.692	0.442
DT > 4 s	0.657	0.525–0.789	0.024*
DT > 6 s	0.634	0.498–0.77	0.054
DT > 8 s	0.620	0.484–0.755	0.086
DT > 10 s	0.608	0.47–0.747	0.120
rCBF < 40%	0.587	0.446–0.728	0.212
CBV < 2	0.523	0.392–0.654	0.742

DT indicates delay time; AUC, area under the curve; CI, confidence interval; rCBF, relative cerebral blood flow; CBV, cerebral blood volume; * $P < 0.05$.

quartile were approximately 1.3, 3.3, and 3.6 times more likely to develop PH, respectively.

We also found significant correlations between DT > 2 s, > 4 s, > 6 s, > 8 s, and > 10 s, with correlation coefficients ranging from 0.906 to 0.992.

In backward stepwise elimination logistic regression, including age, baseline NIHSS score, poor collaterals and recanalization, and DT > 4 s, only the volume of DT > 4 s was independently associated with PH (odds ratio (OR) 1.04 per 1 mL increase in DT > 4 s volume [95% confidence interval (CI), 1.01–1.06]; $P = 0.011$). **Table 3** shows the results of the logistic regression models and the ORs for PH.

Two illustrative cases with areas of DTs, with subsequent PH after intra-arterial thrombolysis corresponding to the site of abnormal DT are shown in **Figure 2**.

Based on ECASS hemorrhagic transformation classification, the **Supplementary Table 1** shows baseline clinical characteristics, stroke risk factors, and time to imaging for the study patients. The died cases (rate) post-procedure of No HT, HI1, HI2, PH1, and PH2 were 12(26%), 3(33%), 3(19%), 2(14%), and 4(44%), respectively. The ROC analysis for association of any HT across the range of values in each CTP parameter identified rCBF < 40% (area under the curve = 0.632; $P = 0.027$), and CBV < 2 mL /100 g (area under the curve = 0.650; $P = 0.012$). It has previously been demonstrated that rCBF < 40% or CBV < 2 mL /100 g on CTP corresponds closely to ischemic core (as shown in **Supplementary Table 2**). Although rCBF < 40% or CBV < 2 mL /100 g could not be significantly associated with the PH in ROC analysis, they were significantly associated with any HT.

DISCUSSION

Our study has three main findings. Firstly, moderate hypo-perfusion (DT > 4 s) predicted PH in patients with endovascular thrombolysis better than other DT values. Secondly, DT > 4 s was better than rCBF < 40% or CBV < 2 mL /100 g for PH prediction after IA tPA thrombolysis. Thirdly, the volume of DT > 4 s was an independent factor to predict PH, which was more significant than collateral grading and recanalization status.

Although clinical factors are useful in the decision making process before IA tPA thrombolysis administration, in practice,

TABLE 3 | Logistic regression analysis for PH.

Variable	Multivariate Model	
	Odds ratio (CI)	P-value
NIHSS	1.09 (0.98–1.21)	0.110
Age	0.99 (0.95–1.03)	0.597
DT > 4 s, mL ^a	1.04 (1.01–1.06)	0.011*
Poor collaterals and recanalization	1.04 (0.59–1.82)	0.901
CTP**		
DT > 4 s		
Q2	1.33 (0.26–6.74)	0.728
Q3	3.33 (0.76–14.66)	0.111
Q4	3.56 (0.80–15.72)	0.094
DT > 6 s		
Q2	3.50 (0.63–19.54)	0.153
Q3	4.32 (0.79–23.59)	0.091
Q4	5.60 (1.04–30.20)	0.045*
DT > 8 s		
Q2	2.75 (0.61–12.29)	0.187
Q3	1.75 (0.37–8.37)	0.481
Q4	3.56 (0.80–15.72)	0.094
DT > 10 s		
Q2	3.50 (0.63–19.54)	0.153
Q3	4.32 (0.79–23.59)	0.091
Q4	5.60 (1.04–30.20)	0.045*

CI indicates 95% confidence interval; CTP, CT perfusion imaging; DT, delay time; and NIHSS, National Institutes of Health Stroke Scale.

^a Odds ratio (OR) given for each 1 mL increase in DT > 4 s volume.

* $P < 0.05$.

**Compared with Q1.

most of these are insufficient in isolation to predict PH. The non-contrast CT before treatment is useful to exclude patients with more than 1/3 of the MCA territory infarct, but of limited value to detect patients at risk of PH in the remaining patients who go on to receive treatment. Although MRI is not routinely available in the acute setting, MRI-based parameters for prediction of post-thrombolysis HT were found in many previous studies, including diffusion-weighted imaging-based lesion volume, severe hypoperfusion measured by high Tmax (22), regional very low CBV (23), and increased permeability (24–26).

CTP is widely available and rapidly accessible in most stroke centers, and thus lends itself to the clinical decision-making process. In this study, the reason we chose DT, instead of Tmax, was that the Tmax value could be dependent on various factors including arterial delay and dispersion and tissue transit time and dispersion. To compensate for arterial delay and dispersion effects, a vascular transport model involving an arterial transport function with a delay time and a relative dispersion has been proposed (19). Thus, there is marked variability in lesion volume prediction among various deconvolution techniques (19, 27).

This study demonstrated that moderate hypo-perfusion (DT > 4 s) predicted PH in patients with endovascular thrombolysis better than other DT values. A prior study by Yassi et al. (15), correlating Tmax and PH, showed that extreme hypo-perfusion

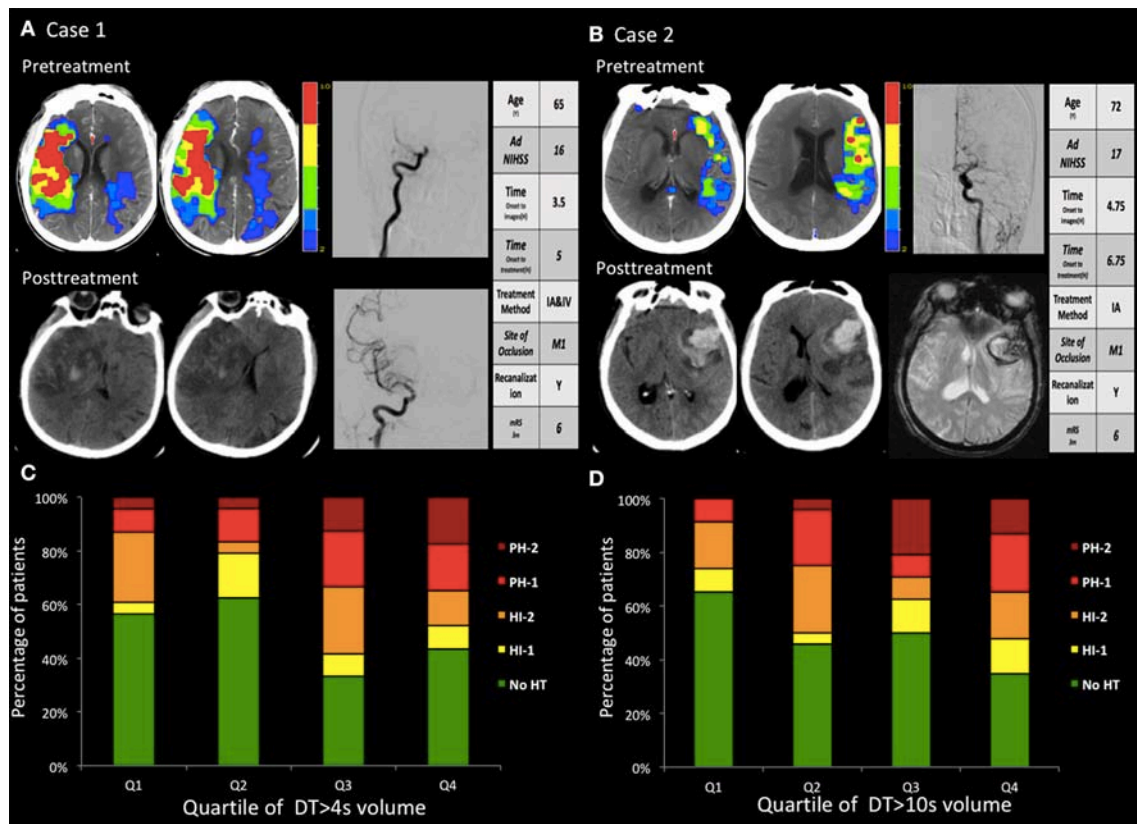


FIGURE 2 | (A,B) two cases with hemorrhagic transformation. **(A)** Pretreatment CTP images show a marked large area with perfusion delay. The day after recanalization with intravenous and endovascular treatment, CT findings revealed a PH. The PH was located within the mildly and severely hypoperfused regions. **(B)** Pretreatment CTP images show small areas with large mild perfusion delay and small severe perfusion delay. After recanalization with endovascular treatment, CT and GRE findings revealed a large area of PH. The PH was located within the mildly hypoperfused regions. **(C,D)** The volume in the regions with baseline mild and severe perfusion delay ($DT > 4$ s and $DT > 10$ s) correlated with the PH. Q1 represents the patients with the lowest quartile volume, whereas Q4 indicates patients with the highest quartile volume.

lesion ($T_{max} > 14$ s volumes of > 5 mL) and thrombolysis were both independently predictive of PH for the patient with or without thrombolysis. OR values of the $T_{max} > 14$ s volumes and the thrombolysis were 4.3 and 10.1, respectively. Thus, thrombolysis had much more influence on the PH than T_{max} . IA treatment was more likely to be performed in patients with severe neurological deficits and perfusion delay, and PH might reflect increasing stroke severity and more aggressive treatment (22). When all cases were treated with endovascular thrombolysis, the influence of aggressive treatment disappeared, and this may explain the discrepancy in findings between Yassi's study and our study. Secondly, the correlations between $DT > 2$ s, > 4 s, > 6 s, > 8 s, and > 10 s were statistically high from 0.906 to 0.992; therefore, the results might also be much dependent on different study populations. Moreover, although the thresholds at more extreme reductions in CBF, CBV, T_{max} , and regional very low CBV have been previously shown to be useful in MRI, they were to be of limited utility using CTP, mainly because of a lack of sensitivity to very low levels of contrast in the severely hypo-perfused region, which produces a relatively large region of undetectable CBF and CBV.

Reperfusion of the ischemic core is an important cofactor in the pathogenesis of hemorrhage (10, 28). The CBV < 2 mL/100 g, and CBF $< 40\%$ could be equal to the ischemic core in previous studies (17, 29). However, in this study, CBF $< 40\%$ and CBV < 2 mL/100 g could not be able to predict PH statistically in ROC analysis. Poor collaterals and recanalization were also found to correlate with HT (20), but in this study, we were unable to conclude that the poor collateral circulation and therapeutic recanalization had more PH. This study found that $DT > 4$ s may be a better predictor of PH after IA tPA thrombolysis therapy. Therefore $DT > 4$ s reflect perhaps both collaterals and reperfusion, and indicate that there is still some blood reaching the tissue, independent of the source of these collaterals or reperfusion. Additionally, instead of having to consider collaterals, reperfusion, and severe ischemia, now a single parameter can be assessed: $DT > 4$ s.

The ECASS III trial used the following definition for symptomatic HT: any blood in the brain or intracranially associated with a clinical deterioration 4 NIHSS points that was identified as the predominant cause of neurological deterioration (30). However, it is often difficult to differentiate the predominant

cause of neurological deterioration in clinical practice (22). The different clinical outcomes after different subtypes of HT illustrates the difficulty in defining symptomatic HT precisely and clearly (11, 31). Also, patients with HI often showed no symptoms. Thus, in this study, we used the PH as the outcome of symptomatic HT.

Limitations include the modest numbers of patients in the cohort. Due to the retrospective nature, PH was measured in different time intervals. There was a significantly increased mortality rate of our cohort, especially in those patients with PH2. It might be related to the prolonged treatment window time and more severe state of illness as the mean NIHSS was greater than 15. Its small sample size hampers proper multivariate analysis. Second, the patient cohort was collected from three different hospitals retrospectively, so inconsistency on imaging parameters may exist, although the CT scanners were all 64-slice scanners and used the same cine mode, and the iodinated contrast material was the same brand with similar injection methods. Also, the whole brain was not covered in the CTP studies, as 64-slice CT scanners were used in this study (32). And dual-energy CT was not used in this study. Third, quantitative analysis of other values of perfusion maps (CBF, MTT, CBV) with a range of thresholds was not included in our study. We chose to focus on DT on the basis of previously published data indicating its value as a predictor of PH (15).

In conclusion, the results of this study indicate that the moderate perfusion delay rather than severe delay was independently associated with PH after endo-vascular thrombolysis. Although the ischemic core on CTP is useful in the pretreatment prediction of HT, the moderate hypo-perfusion on DT is more strongly associated and may allow better prediction of PH after endovascular thrombolysis. Perfusion imaging may be significant not only for the fate of cerebral tissues, but also for the prevention of PH. Further studies are needed for a better understanding of the pathogenesis of PH.

ETHICS STATEMENT

All procedures performed in the studies involving human participants were in accordance with the ethical

standards of the institutional review boards of PLA Army General Hospital and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

AUTHOR'S NOTE

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

AUTHOR CONTRIBUTIONS

BW: study concept and design, analysis and interpretation of data, statistical analysis. NL: study concept and design, acquisition of data. MW, MP, HC, LL, YL, and ZS: study concept and design, critical revision of manuscript. SZ, GH, YZ, and JH: acquisition of data, analysis and interpretation of data. XW and GZ: study concept and design, critical revision of manuscript for intellectual content.

FUNDING

The National Natural Science Foundations of China (Grant No. 81371286, No. 81501024, No.81601035, and No.81771273) and Science Cooperative Foundation of China and Hungary (HCSCF-2016-2) supported this study.

ACKNOWLEDGMENTS

We would like to thank Statistical Elite Studios [www.tjstat.com/] for helping with statistical analyses.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2018.00680/full#supplementary-material>

REFERENCES

- Bhatia R, Hill MD, Shobha N, Menon B, Bal S, Kochar P, et al. Low rates of acute recanalization with intravenous recombinant tissue plasminogen activator in ischemic stroke: real-world experience and a call for action. *Stroke* (2010) 41:2254–8. doi: 10.1161/STROKEAHA.110.592535
- Menon BK, Campbell BC V, Levi C, Goyal M. Role of imaging in current acute ischemic stroke workflow for endovascular therapy. *Stroke* (2015) 46:1453–61. doi: 10.1161/STROKEAHA.115.009160
- Berkhemer OA, Fransen PSS, Beumer D, van den Berg LA, Lingsma HF, Yoo AJ, et al. A randomized trial of intraarterial treatment for acute ischemic stroke. *N Engl J Med.* (2015) 372:11–20. doi: 10.1056/NEJMoa1411587
- Campbell BC V, 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
- Goyal M, Demchuk AM, Menon BK, Eesa M, Rempel JL, Thornton J, et al. Randomized assessment of rapid endovascular treatment of ischemic stroke. *N Engl J Med.* (2015) 372:1019–30. doi: 10.1056/NEJMoa1414905
- Saver JL, Jahan R, Levy EI, Jovin TG, Baxter B, Nogueira RG, et al. Solitaire flow restoration device versus the Merci Retriever in patients with acute ischaemic stroke (SWIFT): a randomised, parallel-group, non-inferiority trial. *Lancet* (2012) 380:1241–9. doi: 10.1016/S0140-6736(12)61384-1
- Kase CS, Furlan AJ, Wechsler LR, Higashida RT, Rowley HA, Hart RG, et al. Cerebral hemorrhage after intra-arterial thrombolysis for ischemic stroke: the PROACT II trial. *Neurology* (2001) 57:1603–10. doi: 10.1212/WNL.57.9.1603
- von Kummer R, Broderick JP, Campbell BC V, Demchuk A, Goyal M, Hill MD, et al. The heidelberg bleeding classification: classification of bleeding events after ischemic stroke and reperfusion therapy. *Stroke* (2015) 46:2981–6. doi: 10.1161/STROKEAHA.115.010049

9. Khatri R, McKinney AM, Swenson B, Janardhan V. Blood-brain barrier, reperfusion injury, and hemorrhagic transformation in acute ischemic stroke. *Neurology* (2012) 79(13 Suppl):S52–7. doi: 10.1212/WNL.0b013e3182697e70
10. Sussman ES, Connolly ESJ. Hemorrhagic transformation: a review of the rate of hemorrhage in the major clinical trials of acute ischemic stroke. *Front. Neurol.* (2013) 4:69. doi: 10.3389/fneur.2013.00069
11. Singer OC, Humpich MC, Fiehler J, Albers GW, Lansberg MG, Kastrup A, et al. Risk for symptomatic intracerebral hemorrhage after thrombolysis assessed by diffusion-weighted magnetic resonance imaging. *Ann Neurol.* (2008) 63:52–60. doi: 10.1002/ana.21222
12. Jain AR, Jain M, Kanthala AR, Damania D, Stead LG, Wang HZ, et al. Association of CT perfusion parameters with hemorrhagic transformation in acute ischemic stroke. *AJNR. Am. J Neuroradiol.* (2013) 34:1895–900. doi: 10.3174/ajnr.A3502
13. Souza LCS, Payabvash S, Wang Y, Kamalian S, Schaefer P, Gonzalez RG, et al. Admission CT perfusion is an independent predictor of hemorrhagic transformation in acute stroke with similar accuracy to DWI. *Cerebrovasc Dis.* (2012) 33:8–15. doi: 10.1159/000331914
14. Shinoyama M, Nakagawara J, Yoneda H, Suzuki M, Ono H, Kunitzugu I, et al. Initial “TTP Map-Defect” of computed tomography perfusion as a predictor of hemorrhagic transformation of acute ischemic stroke. *Cerebrovasc Dis Extra.* (2013) 3:14–25. doi: 10.1159/000346113
15. Yassi N, Parsons MW, Christensen S, Sharma G, Bivard A, Donnan GA, et al. Prediction of poststroke hemorrhagic transformation using computed tomography perfusion. *Stroke* (2013) 44:3039–43. doi: 10.1161/STROKEAHA.113.002396
16. Aviv RI, d’Este CD, Murphy BD, Hoppyan JJ, Buck B, Mallia G, et al. Hemorrhagic transformation of ischemic stroke: prediction with CT perfusion. *Radiology* (2009) 250:867–77. doi: 10.1148/radiol.2503080257
17. 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
18. Parsons MW, Pepper EM, Chan V, Siddique S, Rajaratnam S, Bateman GA, et al. Perfusion computed tomography: prediction of final infarct extent and stroke outcome. *Ann Neurol.* (2005) 58:672–9. doi: 10.1002/ana.20638
19. 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
20. Bang OY, Saver JL, Kim SJ, Kim G-M, Chung C-S, Ovbiagele B, et al. Collateral flow averts hemorrhagic transformation after endovascular therapy for acute ischemic stroke. *Stroke* (2011) 42:2235–9. doi: 10.1161/STROKEAHA.110.604603
21. Renou P, Sibon I, Tourdias T, Rouanet F, Rosso C, Galanaud D, et al. Reliability of the ECASS radiological classification of postthrombolysis brain haemorrhage: a comparison of CT and three MRI sequences. *Cerebrovasc Dis.* (2010) 29:597–604. doi: 10.1159/000312867
22. Kim JH, Bang OY, Liebeskind DS, Ovbiagele B, Kim GM, Chung CS, Lee KH, Saver JL. Impact of baseline tissue status (diffusion-weighted imaging lesion) versus perfusion status (severity of hypoperfusion) on hemorrhagic transformation. *Stroke* (2010) 41:e135–42. doi: 10.1161/STROKEAHA.109.563122
23. Campbell BC V, Christensen S, Butcher KS, Gordon I, Parsons MW, Desmond PM, et al. Regional very low cerebral blood volume predicts hemorrhagic transformation better than diffusion-weighted imaging volume and thresholded apparent diffusion coefficient in acute ischemic stroke. *Stroke* (2010) 41:82–8. doi: 10.1161/STROKEAHA.109.562116
24. Leigh R, Christensen S, Campbell BC V, Marks MP, Albers GW, Lansberg MG. Pretreatment blood-brain barrier disruption and post-endovascular intracranial hemorrhage. *Neurology* (2016) 87:263–9. doi: 10.1212/WNL.0000000000002862
25. Leigh R, Jen SS, Hillis AE, Krakauer JW, Barker PB. Pretreatment blood-brain barrier damage and post-treatment intracranial hemorrhage in patients receiving intravenous tissue-type plasminogen activator. *Stroke* (2014) 45:2030–5. doi: 10.1161/STROKEAHA.114.005249
26. Aksoy D, Bammer R, Mlynash M, Venkatasubramanian C, Eyngorn I, Snider RW, et al. Magnetic resonance imaging profile of blood-brain barrier injury in patients with acute intracerebral hemorrhage. *J Am Heart Assoc.* (2013) 2:e000161. doi: 10.1161/JAHA.113.000161
27. Calamante F, Christensen S, Desmond PM, Ostergaard L, Davis SM, Connelly A. The physiological significance of the time-to-maximum (Tmax) parameter in perfusion MRI. *Stroke* (2010) 41:1169–74. doi: 10.1161/STROKEAHA.110.580670
28. Campbell BC V, Christensen S, Parsons MW, Churilov L, Desmond PM, Barber PA, et al. Advanced imaging improves prediction of hemorrhage after stroke thrombolysis. *Ann Neurol.* (2013) 73:510–9. doi: 10.1002/ana.23837
29. Wintermark M, Flanders AE, Velthuis B, Meuli R, van Leeuwen M, Goldsher D, et al. Perfusion-CT assessment of infarct core and penumbra: receiver operating characteristic curve analysis in 130 patients suspected of acute hemispheric stroke. *Stroke* (2006) 37:979–85. doi: 10.1161/01.STR.0000209238.61459.39
30. Hacke W, Kaste M, Bluhmki E, Brozman M, Davalos A, Guidetti D, et al. Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke. *N Engl J Med.* (2008) 359:1317–29. doi: 10.1056/NEJMoa0804656
31. Berger C, Fiorelli M, Steiner T, Schabitz WR, Bozzao L, Bluhmki E, et al. Hemorrhagic transformation of ischemic brain tissue: asymptomatic or symptomatic? *Stroke* (2001) 32:1330–5. doi: 10.1161/01.STR.32.6.1330
32. 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

Conflict of Interest Statement: ZS was employed by company GE Healthcare, Beijing.

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.

Copyright © 2018 Wu, Liu, Wintermark, Parsons, Chen, Lin, Zhou, Hu, Zhang, Hu, Li, Su, Wu and Zhu. 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.



Application of Machine Learning to Automated Analysis of Cerebral Edema in Large Cohorts of Ischemic Stroke Patients

Rajat Dhar^{1*}, Yasheng Chen^{2†}, Hongyu An³ and Jin-Moo Lee²

¹ Division of Neurocritical Care, Department of Neurology, Washington University in St. Louis, St. Louis, MO, United States,

² Division of Cerebrovascular Diseases, Department of Neurology, Washington University in St. Louis, St. Louis, MO, United States, ³ Department of Radiology, Washington University in St. Louis, St. Louis, MO, United States

OPEN ACCESS

Edited by:

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

Reviewed by:

W. Taylor Kimberly,
Harvard Medical School,
United States
Maurizio Acampa,
Azienda Ospedaliera Universitaria
Senese, Italy

*Correspondence:

Rajat Dhar
dharr@wustl.edu

[†]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: 03 May 2018

Accepted: 30 July 2018

Published: 21 August 2018

Citation:

Dhar R, Chen Y, An H and Lee J-M
(2018) Application of Machine
Learning to Automated Analysis of
Cerebral Edema in Large Cohorts of
Ischemic Stroke Patients.
Front. Neurol. 9:687.
doi: 10.3389/fneur.2018.00687

Cerebral edema contributes to neurological deterioration and death after hemispheric stroke but there remains no effective means of preventing or accurately predicting its occurrence. Big data approaches may provide insights into the biologic variability and genetic contributions to severity and time course of cerebral edema. These methods require quantitative analyses of edema severity across large cohorts of stroke patients. We have proposed that changes in cerebrospinal fluid (CSF) volume over time may represent a sensitive and dynamic marker of edema progression that can be measured from routinely available CT scans. To facilitate and scale up such approaches we have created a machine learning algorithm capable of segmenting and measuring CSF volume from serial CT scans of stroke patients. We now present results of our preliminary processing pipeline that was able to efficiently extract CSF volumetrics from an initial cohort of 155 subjects enrolled in a prospective longitudinal stroke study. We demonstrate a high degree of reproducibility in total cranial volume registration between scans ($R = 0.982$) as well as a strong correlation of baseline CSF volume and patient age (as a surrogate of brain atrophy, $R = 0.725$). Reduction in CSF volume from baseline to final CT was correlated with infarct volume ($R = 0.715$) and degree of midline shift (quadratic model, $p < 2.2 \times 10^{-16}$). We utilized generalized estimating equations (GEE) to model CSF volumes over time (using linear and quadratic terms), adjusting for age. This model demonstrated that CSF volume decreases over time ($p < 2.2 \times 10^{-13}$) and is lower in those with cerebral edema ($p = 0.0004$). We are now fully automating this pipeline to allow rapid analysis of even larger cohorts of stroke patients from multiple sites using an XNAT (eXtensible Neuroimaging Archive Toolkit) platform. Data on kinetics of edema across thousands of patients will facilitate precision approaches to prediction of malignant edema as well as modeling of variability and further understanding of genetic variants that influence edema severity.

Keywords: ischemic stroke, machine learning, cerebral edema, image analysis and processing, CT scan, CSF volume, GEE

INTRODUCTION

Over 10 million persons suffer a stroke each year worldwide (1). Most of these patients have at least one brain imaging study performed during their acute hospitalization, primarily for diagnostic purposes on presentation (2). Follow-up scans are often obtained to evaluate the size of infarction, degree of cerebral edema, as well as exclude the development of hemorrhagic transformation (3). Computed tomography (CT) is the most frequently employed modality for acute stroke imaging due to its widespread availability, lower cost, and greater speed of scanning, especially important in acutely unstable patients where “time is brain” (4). Although conventional CT does not have the ability of magnetic resonance imaging (MRI) to detect hyper-acute stroke, its ability to track progression of infarction and edema after stroke are comparable while affording greater temporal resolution with serial imaging (5). This practice means that there is a massive global imaging dataset of stroke patients with information on stroke location, infarct size, development of edema, and hemorrhagic transformation. While these parameters can be assessed by human raters, such evaluation is not scalable when leveraging imaging data from thousands of patients.

Cerebral edema develops around regions of brain infarction within the first week after stroke. This pathologic increase in brain water and hemispheric volume can lead to mass effect and is the major cause of death and neurological worsening after stroke (6). Development of edema is usually heralded by abrupt mental status worsening 2 days or more after admission, when herniation and midline shift have already developed (7). However, this process actually begins in the first hours after stroke and evolves continually and progressively over the first few days. At first decreases in blood and cerebrospinal fluid (CSF) compartments within the cranial compartment compensate for this increase in brain volume. However, once this has been exhausted, decompensation with worsening rapidly follows. Current measures of edema such as midline shift (MLS) or neurological deterioration capture only this decompensated state and not the critical early stages of edema before worsening. Further, assessing edema utilizing only MLS neglects the full spectrum of edema, including those with increased brain volume who never develop MLS. Measures of lesion volume either requires MRI (not feasible in all stroke patients) or can be estimated using CT; however, hypodensity on CT may be subtle early on and represents a variable combination of infarct plus edema. It is only the latter component that contributes to swelling and risk of herniation, and so lesion volume (even on MRI) only partially predicts risk of herniation (8).

We have proposed a sensitive quantitative metric of edema severity that can be extracted from CT imaging at variable time points after stroke (9). This leverages the reciprocal biologic relationship between increase in brain volume due to swelling and proportional decrease in CSF volume as compensation. CSF is pushed out of hemispheric sulci, cerebral ventricles, and the basal cisterns as edema develops in the hours and days after stroke. The reduction in CSF volume precedes the development of midline shift and clinical worsening due to edema. We

demonstrated that the volume of CSF displaced up to the time of maximal edema closely correlated with extent of midline shift.

We have also developed an automated algorithm to segment CSF from CT scans of stroke patients (10). This critical step employed random forest-based machine learning (ML) trained on manually delineated scans. Features integrated into the ML platform include Haar-like patterns of pixels. This supervised learning approach was able to rapidly and reliably measure CSF volume on serial CT scans from two sites in our preliminary testing, performing significantly better than simple threshold-based models for CSF segmentation which were confounded by density of infarction mimicking CSF. Correlations of automated CSF volumes to ground-truth values exceeded 0.95, with volumes that closely approximated actual CSF values after active contour refinement. This automated approach facilitates the translation of this metric to studies evaluating edema in large numbers of stroke patients. Exploring the variability in quantifiable edema severity between patients will not only unlock opportunities for precise prediction of malignant edema at earlier time points but also provide the basis for understanding the genetic basis of cerebral edema. Such studies require thousands of stroke patients with serial imaging to undergo CSF-based edema measurement. We now present a proof-of-principle application of a processing algorithm capable of handling large datasets of CT scans and extracting CSF volumes for such analyses.

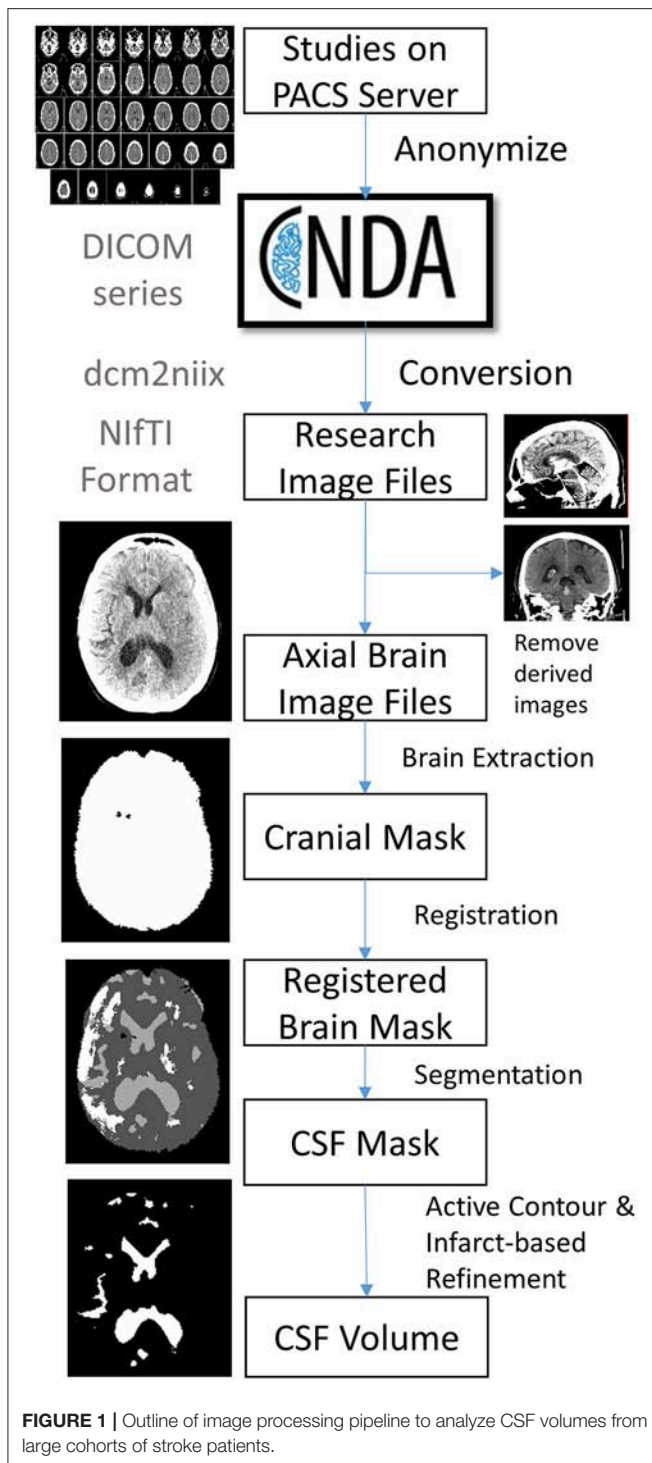
MATERIALS AND METHODS

Subjects and Data Collection

Patients with a diagnosis of ischemic stroke who were admitted to Barnes-Jewish Hospital were screened for enrollment into the Genetics of Neurological Instability after Ischemic Stroke (GENISIS) study if they presented within 6 h of symptom onset. Subjects provided informed consent for data collection, including acute stroke imaging. Clinical data collected included age and NIHSS at baseline. All head CT imaging performed on subjects enrolled between 2009 and 2014 was then extracted from the clinical radiology server. We included only those with at least one follow-up scan performed during their hospitalization. **Figure 1** shows the steps involved in a processing pipeline capable of uploading, evaluating, processing, and extracting CSF volumes from these scans. All scans (including baseline CT on presentation and each follow-up scan available) were uploaded from the hospital's Picture Archiving and Communication System (PACS) server to Central Neuroimaging Data Archive (CNDA), where they were stored in Digital Imaging and Communications in Medicine (DICOM) format (11). All studies were de-identified during the upload process using a standard algorithm integrated into the upload pipeline. FU scans were reviewed for presence of visible infarct as well as graded for degree of cerebral edema (CED grade 0, no infarct visible; 1, focal swelling up to 1/3 of cerebral hemisphere; 2, focal swelling of >1/3 of cerebral hemisphere; 3, swelling with midline shift) (12).

DICOM Conversion

DICOM images were converted to NIfTI (Neuroimaging Informatics Technology Initiative) format in bulk using the



dcm2niix software. Multiple DICOM-encoded brain slices from a single scanner sequence were compiled into a single 3-dimensional NIfTI file. The header of the newly created NIfTI file also stores the image dimensions (e.g., $\sim 512 \times 512 \times 32$) and pixel dimensions (e.g., $\sim 0.42 \times 0.42 \times 5$ -mm). The conversion also labels the resulting file using the subject identifier (assigned during upload) plus date and time of each scan (extracted from

the DICOM metadata). Due to inconsistency in storing slice thickness in CT metadata, conversion extracts pixel height not from slice thickness and/or spacing between slices (reported inconsistently in metadata) but by calculating the actual distance between two consecutive slices. Conversion of CT images poses additional unique complexities: images are often acquired with slice axis oblique to the scanning table (13). This *gantry tilt* would result in a skewed 3D stack of images if this is not resolved using trigonometry and resampling (as is performed during conversion). This resampling to a consistent plane is also important for accurate co-registration of scans within a given patient. Some CT series may also be acquired with varying slice thicknesses, typically with thinner slices in the posterior fossa. Such inconsistency cannot be handled by the NIfTI format, which requires uniform slice thickness when storing imaging data. The conversion algorithm recognizes such variable inter-slice distances and interpolates to a uniform thickness in the resulting NIfTI file. We also store additional metadata not captured in the NIfTI header (such as scanner, protocol, method of conversion) in a brain imaging data structure (BIDS) accessory file (14).

Image Selection

Each patient often has multiple series performed as part of a single session. Derived images were excluded automatically from conversion using the “-i y” switch in *dcm2niix*. However, selection of axial brain images required some manual review of converted NIfTI files to exclude bone windows and additional series that were not analyzed (e.g., angiographic images).

Infarct Review

Each follow-up scan was also manually reviewed for presence and location of visible infarcts as well as presence and degree of midline shift (at level of the septum pellucidum). Visible stroke-related hypodensities were outlined in MRICro and saved as image masks. Infarct location was categorized as cortical, subcortical, both cortical and subcortical, lacunar (subcortical with diameter <15 mm), or other.

Brain Extraction and Perimeter Registration

Further anonymization of images was ensured by removal of all structures external to the cranial cavity (i.e., skull stripping). This was accomplished by k-means clustering of pixel intensities to segregate brain, skull, and all external pixels. Skull and external regions were then excluded to yield a mask of just the intracranial contents. This image was then registered to a brain template that consisted of 15 brain images of stroke patients with manually outlined cranial perimeter to include all supratentorial structures as well as basal cisterns, but specifically excluding portions of the posterior fossa (e.g., cerebellum) on the same slices. Each subjects baseline brain scan was registered to each of these atlas brains using the Advanced Normalized Toolkit (ANTs) and pixels were included if they matched to the atlas masks in over half of the template scans. This registered baseline scan was then registered to each follow-up scan and non-matching brain regions were excluded.

CSF Segmentation

The brain mask was then segmented using the CSF classifier that we previously trained using random forest machine learning (10). This segmentation was then refined using an active contour method and cleaned using a manually drawn mask of the infarct hypodensity (if present on follow-up scans). Results are summarized in JPEG snapshots of the resulting CSF mask overlaid onto the CT images for manual review of segmentation accuracy on serial scans (see **Figure 2** for results of CSF segmentation in one representative subject).

Volumetric Analyses

The number of pixels in each compartment (intracranial compartment, CSF, infarct) is extracted from each image mask. This is then converted into volume using the pixel dimensions in the image header. Results from each scan are compiled into an exportable data file. This was analyzed in R (R: a language and environment for statistical computing). CSF and infarct volumes were analyzed in milliliters (ml) as well as a proportion of total cranial volume (%). The maximum change in CSF volume was calculated using the lowest measured volume as a percentage of the baseline volume.

Dynamic CSF Volumetric Modeling

Generalized estimating equation (GEE) was used to model the temporal CSF volume changes using the multiple CT scans from this patient cohort. In this study, due to an irregular time interval between the scans from different subjects, we employed a Markov working correlation structure, $\text{corr}(y_{i,j}, y_{i,k}) = a^{|t_{i,j} - t_{i,k}|}$, where $y_{i,j}$ and $y_{i,k}$ are CSF volumes of patient i at $t_{i,j}$ and $t_{i,k}$. Besides its capability to model irregular time interval between the scans, this working correlation structure also takes the assumption that

the correlation between the measurements from the same subject weakens with an increased time interval ($0 < a < 1$) (15). In this study, the model we employed for statistical inference include age, time from stroke onset (T), and a dichotomized cerebral edema grade (CED grade 3 vs. grade 0, 1, 2), which is given as $E(y_{i,j}) = b_0 + b_1 * t_{i,j} + b_2 * t_{i,j} * t_{i,j} + b_3 * \text{age}_i + b_4 * \text{ced}_i$. These coefficient ($b_0 \sim b_4$ and a) are calculated through a two stage solutions. P -values were computed with a robust covariance structure.

RESULTS

The cohort included 155 subjects, whose demographics are shown in **Table 1**. Registration failed in two subjects, who were excluded from segmentation and analysis. This left a total of 397 scans analyzed for cranial cavity and CSF volumes. Median time from stroke onset to first scan was just over 1 h (IQR 0.8–2.4 h) while time from baseline to first follow-up scan was a median of 21 h (IQR 6–42 h); 55 subjects had three or more scans performed serially after stroke. One hundred (66%) had one or more scans performed at least 24 h after stroke onset. In one case the only FU scan was over 1 week after stroke; this subject was excluded. The majority of infarcts were cortical or both cortical and subcortical. In those with at least 24-h follow-up, median volume of visible infarct-related hypodensity was 73 ml (IQR 5–203). Median volume was 22 ml for subcortical infarcts, 49.6 ml for cortical infarcts, and 219 ml for infarcts affecting both cortical and subcortical regions.

Swelling with midline shift (i.e., CED grade 3) was demonstrated in 32 (32%) of those with scans beyond 24 h. Median MLS was 6.5 mm in this subgroup (IQR 4.0–9.5) compared to 0 in other CED grades. Registration was able

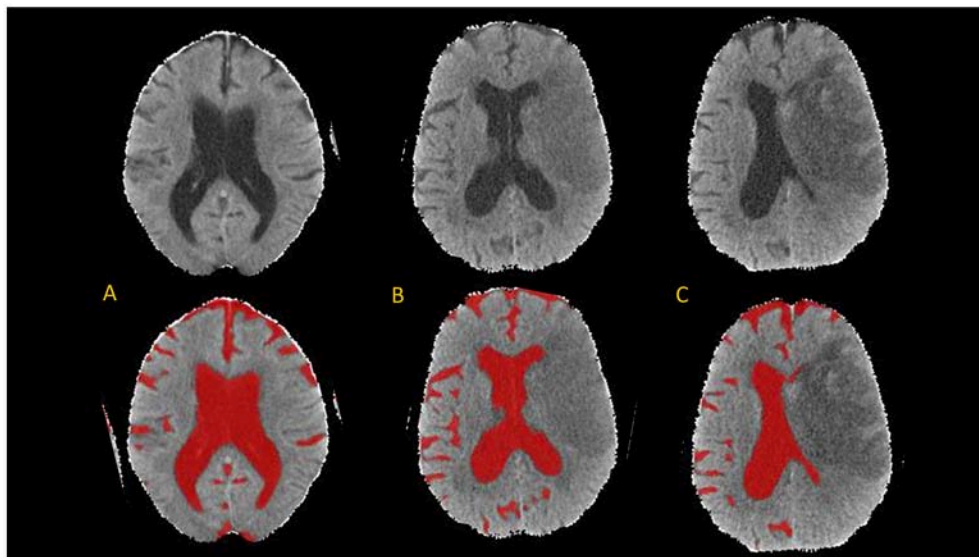


FIGURE 2 | Axial brain slices from head CT (top) and results of CSF segmentation from a 82-year old woman with initial NIHSS of 18. Baseline CT (**A**) was performed within 1 h of stroke onset (CSF volume 224 ml). First follow-up CT (**B**) was performed at 20-h (CSF volume 150 ml) and second follow-up CT (**C**) at 110-h (CSF volume 105 ml).

to extract a consistent cranial mask across serial scans; we demonstrated a strong correlation between baseline and FU cranial volumes ($r = 0.98$, $p < 2 \times 10^{-16}$; **Figure 3**). There was also a good correlation between baseline CSF volume (as percent of cranial volume) and age of patient ($r = 0.74$, $p < 2 \times 10^{-16}$; **Figure 4**).

The maximal reduction in CSF volume (as percentage of baseline) was associated with degree of midline shift developing (**Figure 5**). In fact, there appeared to be a non-linear (quadratic) relationship, whereby minimal midline shift developed despite a mild-moderate CSF volume loss. Beyond the point at which 30–40% of the total baseline CSF had been lost, it appears that

midline shift rapidly develops. Peak CSF volume loss was also correlated with infarct volume in a linear fashion (**Figure 6**) and was significantly greater in those stroke patients with infarcts affecting both cortical and subcortical structures and minimal in those with lacunar infarcts (**Figure 7**).

In the GEE model, we found that CSF volume was independently affected by all three variables: age, time from stroke onset and CED grade. CSF volume increases with age ($b_1 = 3.01$ cc/year, $p < 10^{-16}$) and is lower in those with CED grade 3 ($b_3 = -32.57$ cc, $p = 4 \times 10^{-4}$). CSF volume also decreased over time (-22 cc/day, $p = 2 \times 10^{-13}$) but there was also a second-order quadratic time factor significant in CSF evolution ($p = 6 \times 10^{-10}$). The evolution of CSF volume over time in CED grades is shown in **Figure 8**.

TABLE 1 | Cohort of 155 stroke subjects with baseline and follow-up CT scans.

Variable	
Age	67 ± 14 years
Gender, male	82 (53%)
NIHSS	11 (IQR 6–16)
Time from stroke onset to first CT	1.25 h (IQR 0.8–2.4)
Number with 2/3/4/5 serial scans	153/55/24/14
Infarct location:	
Cortical only	24
Subcortical only	8
Both cortical and subcortical	42
Lacune	7
No infarct seen	12
Unable to assess (no scan beyond 24 h)	56
CED grade* 0/1/2/3	17/34/16/32

*In those with imaging at least at 24 h or beyond.

DISCUSSION

Here we present the initial results of a machine learning-based pipeline to analyze large numbers of serial CT brain images in order to quantify the progression of cerebral edema after ischemic stroke. We applied our random forest-based segmentation algorithm within a broader image processing pipeline to measure CSF volumes in almost 400 CT scans, with failure of scan registration in only two of over 150 subjects with a variety of stroke locations and volumes. We are now working to refine our registration parameters to deal with these rare failures, including addition of shrink factors, smoothing parameters, and affine registration (16). In the remainder cranial registration was robust, with tight correlation of volumes between baseline and repeat images. We also demonstrated a clear relationship between proportion of the cranium comprising CSF (as a surrogate for brain atrophy) and patient age (17).

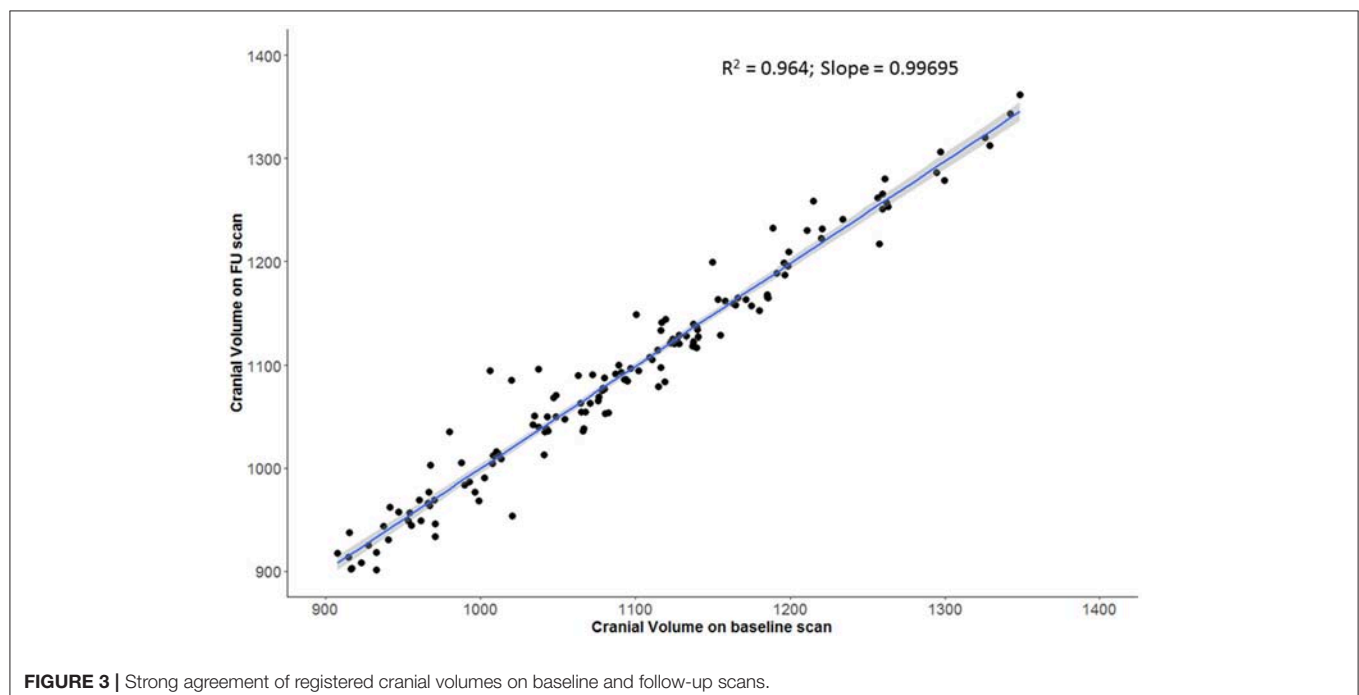


FIGURE 3 | Strong agreement of registered cranial volumes on baseline and follow-up scans.

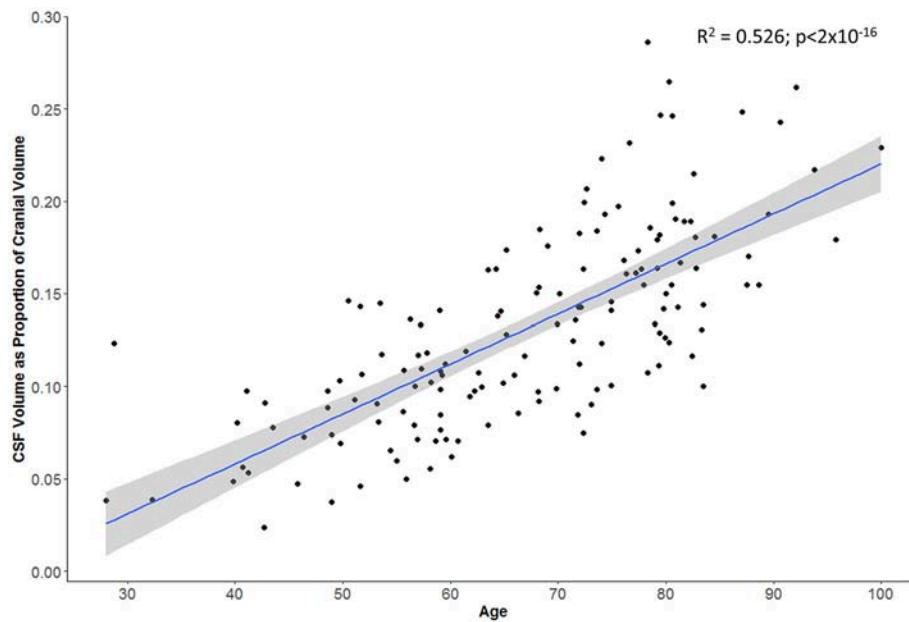


FIGURE 4 | Subject's age correlates with proportion of cranial cavity comprised by CSF on baseline head CT (gray zone represents 95% confidence interval for predictions from the linear regression model).

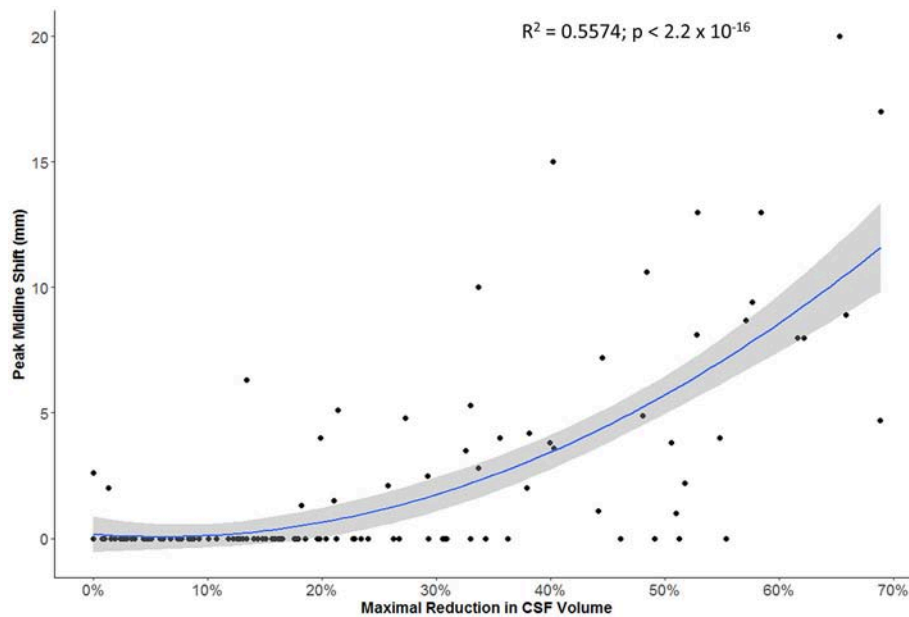
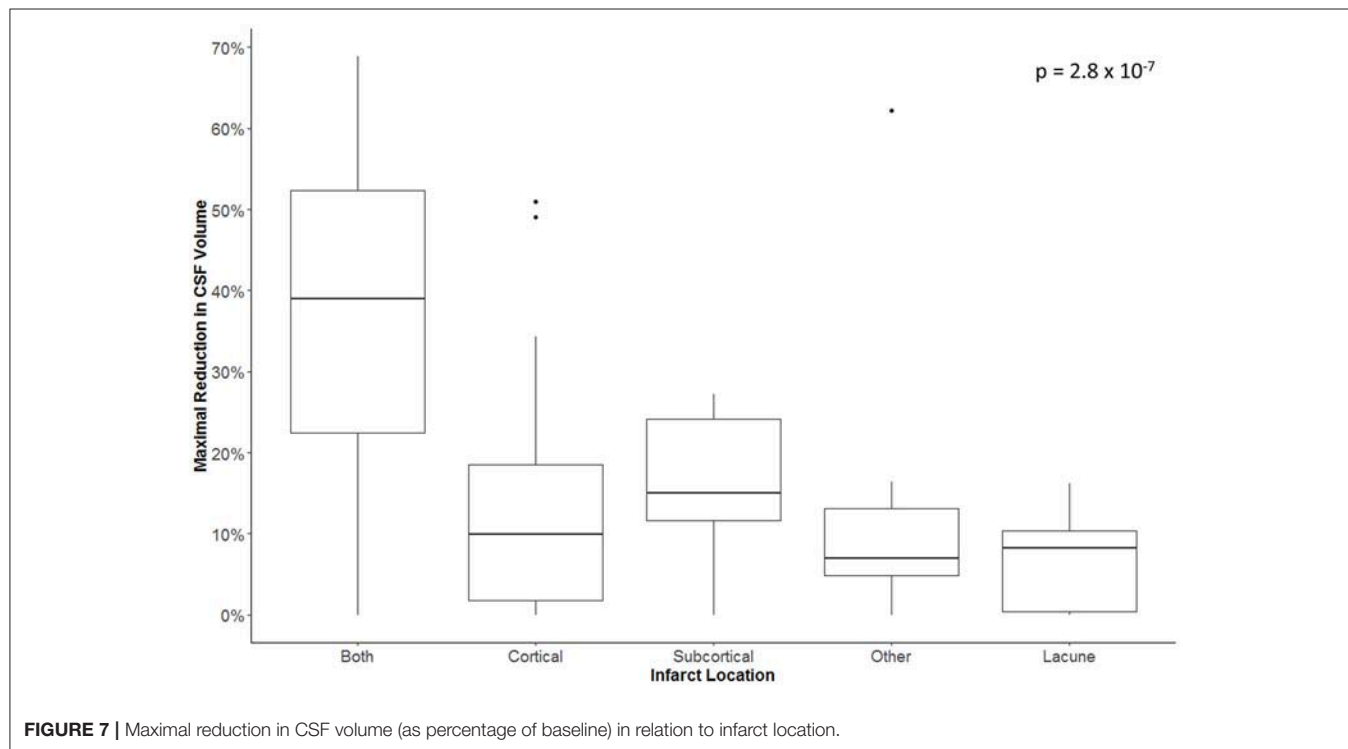
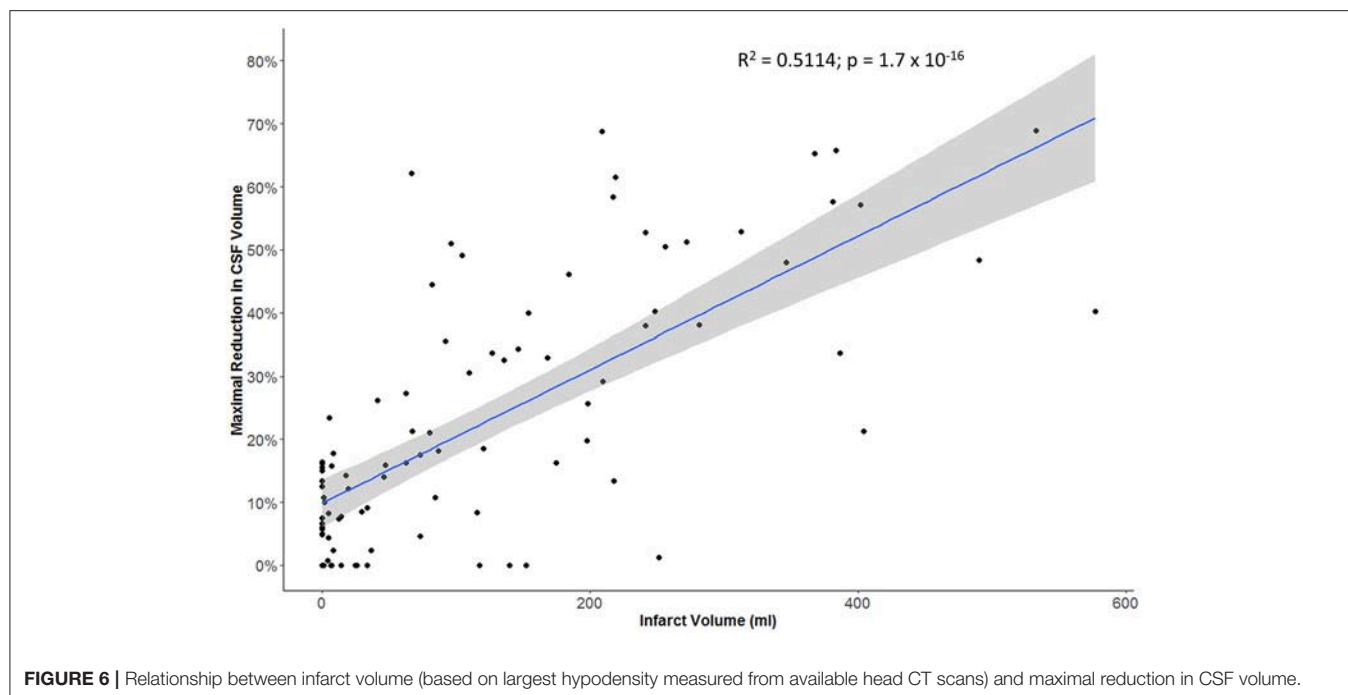


FIGURE 5 | Relationship between maximal reduction in CSF volume (relative to baseline) and peak degree of midline shift (gray zone represents 95% confidence interval for predictions from the quadratic regression model).

More importantly, we further demonstrated that our metric of CSF volume reduction is a strong marker not only of stroke volume but of the eventual development of midline shift. There was more CSF volume loss in those with larger infarcts affecting both cortical and subcortical structures. However, it appears that midline shift only develops once some degree of compensation

afforded by CSF loss has been exhausted. Beyond this threshold, midline shift rapidly develops, as illustrated by our quadratic modeling.

Furthermore, we used longitudinal GEE modeling to demonstrate that CSF volume generally decreased over time after stroke. Even adjusting for age and time from baseline



CT, we confirmed that those with significant CED had greater reductions in CSF volume than those without CED. CSF volumes do not change appreciably over time in those with small infarcts (**Figure 8**) with while those with larger infarcts (CED grades 2 and 3) appear to exhibit gradual but progressive reductions in CSF volumes of between 25 and 50% relative

to baseline. Furthermore, those with CED grade 3 (who, by definition, ultimately develop MLS) seem to manifest a continued downward trajectory between 24 and 48 h after stroke. This group appears to reach an asymptote of maximal CSF reduction of about half baseline volume by 48 h. This volume reduction would represent approximately one hemisphere of

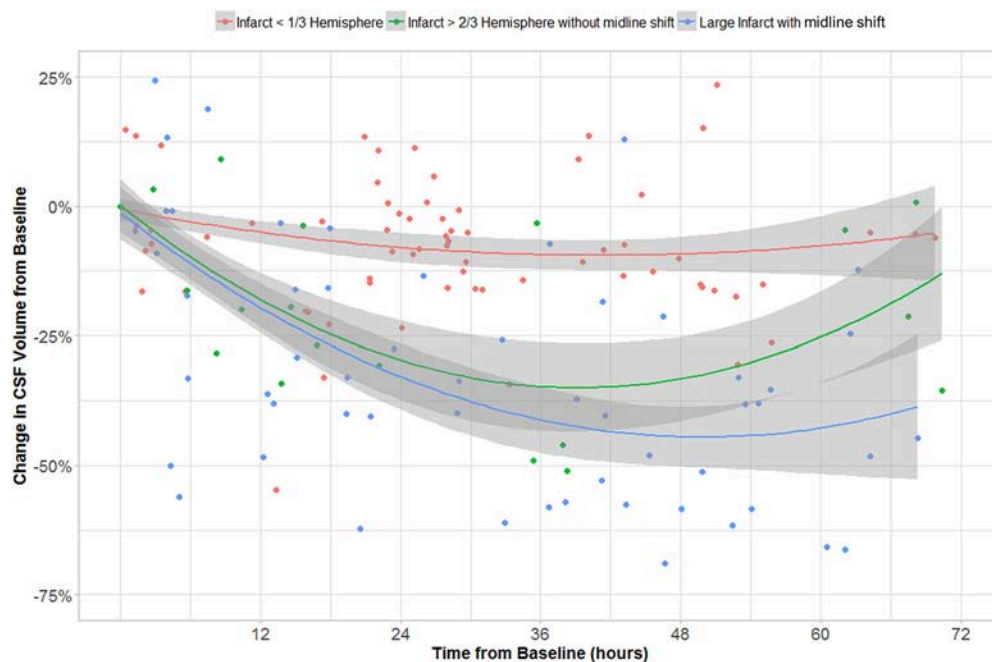


FIGURE 8 | CSF volume (as percentage of volume on baseline CT) over time in groups with CED grades 0–1, 2, and 3.

CSF, appropriate to a process that is likely to produce edema predominantly involving the ipsilateral hemisphere. As our analysis relating MLS with CSF volume loss suggests, there is potential for greater decompensation (with development of MLS) once this proportion of CSF volume has been exhausted. As we accumulate more volumetric data across more stroke patients, we plan to perform more sophisticated analyses that evaluate the interaction of edema severity with rate of CSF volume reduction, incorporating and modeling the effect of further covariates such as NIHSS.

This study provides proof-of-principle that we can automate brain imaging data analysis and obtain meaningful volumetric data on large cohorts of stroke patients. Such an approach, leveraging routinely obtained clinical imaging data or imaging obtained in clinical trials to advance the science of stroke is the pathway to realizing the potential of big data in brain imaging (18). One notable challenge of sharing brain imaging data is ensuring anonymization. In our pipeline this is accomplished by both robust de-identification of DICOM metadata prior to scan transmission to our centralized repository as well as skull stripping during brain extraction and registration. This latter process has also been accomplished previously using similar methods (19).

While this study demonstrates the feasibility of an imaging pipeline to deal with large volumes of CT data, there are a number of refinements required before it can manage big imaging data from large multi-site repositories. In this preliminary test application, we only utilized data from a single site with existing upload capabilities from PACS to our analysis server.

In future we will leverage the existing resources of CNDA to import and archive scans from multiple sites. This imposes other challenges to imaging harmonization as scans are obtained with various protocols under varying sequence names and even in disparate languages. We are currently developing a convolutional neural network (CNN) approach to intelligently but automatically select the appropriate scan from a number of CT series performed concurrently. Subsequent steps in processing such as brain registration and segmentation also need to be automated and we are working on a Docker container-based approach to integrating processing modules (20). We are also working to provide internal quality control checks and means of project-level data visualization to further refine the processing. A further challenge to full automation is the need for manual delineation of infarct hypodensity. We are now developing a CNN-based method of segmenting stroke lesions from serial CT scans (21). Such refinements will be key to successfully scaling up these processes to thousands of CT scans and realizing the potential of big data in stroke. With respects to cerebral edema, this will allow us to precisely predict the course of individual patients from early CSF changes while simultaneously utilizing this imaging-based endophenotype (rate of edema formation) as the basis for powerful genetic studies to develop new targeted therapies to prevent edema.

ETHICS STATEMENT

The study protocol was approved the Human Research Protection Office at Washington University

in St. Louis. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

RD drafted the initial manuscript and collected the clinical data. YC performed the image processing, statistical analyses, and

revised the manuscript. HA and J-ML supervised this project, reviewed the manuscript, and made critical revisions.

FUNDING

This work was funded by the National Institutes of Neurological Disorders and Stroke through grants to RD (K23 NS099440) and J-ML (R01 NS085419).

REFERENCES

- Feigin VL, Krishnamurthi RV, Parmar P, Norrving B, Mensah GA, Bennett DA, et al. Update on the global burden of ischemic and hemorrhagic stroke in 1990–2013: the GBD 2013 study. *Neuroepidemiology* (2015) 45:161–76. doi: 10.1159/000441085
- McAlister FA, Fisher BW, Houston SC. The timing of computed tomography in acute stroke: a practice audit. *Can Assoc Radiol J.* (1997) 48:123–9.
- Larrue V, von Kummer R, del Zoppo G, Bluhmki E. Hemorrhagic transformation in acute ischemic stroke: potential contributing factors in the european cooperative acute stroke study. *Stroke* (1997) 28:957–60. doi: 10.1161/01.STR.28.5.957
- Tong E, Hou Q, Fiebach JB, Wintermark M. The role of imaging in acute ischemic stroke. *Neurosurg Focus* (2014) 36:E3. doi: 10.3171/2013.10.FOCUS13396
- Barber PA, Hill MD, Eliasziw M, Demchuk AM, Pexman JH, Hudon ME, et al. Imaging of the brain in acute ischaemic stroke: comparison of computed tomography and magnetic resonance diffusion-weighted imaging. *J Neurol Neurosurg Psychiatry* (2005) 76:1528–33. doi: 10.1136/jnnp.2004.059261
- Silver FL, Norris JW, Lewis AJ, Hachinski VC. Early mortality following stroke: a prospective review. *Stroke* (1984) 15:492–96. doi: 10.1161/01.STR.15.3.492
- Pullicino PM, Alexandrov AV, Shelton JA, Alexandrova NA, Smurawska LT, Norris JW. Mass effect and death from severe acute stroke. *Neurology* (1997) 49:1090–5. doi: 10.1212/WNL.49.4.1090
- Batley TW, Karki M, Singhal AB, Wu O, Sadaghiani S, Campbell BC, et al. Brain edema predicts outcome after nonlacunar ischemic stroke. *Stroke* (2014) 45:3643–8. doi: 10.1161/STROKEAHA.114.006884
- Dhar R, Yuan K, Kulik T, Chen Y, Heitsch L, An H, et al. CSF Volumetric analysis for quantification of cerebral edema after hemispheric infarction. *Neurocrit Care* (2016) 24:420–7. doi: 10.1007/s12028-015-0204-z
- Chen Y, Dhar R, Heitsch L, Ford A, Fernandez-Cadenas I, Carrera C, et al. Automated quantification of cerebral edema following hemispheric infarction: application of a machine-learning algorithm to evaluate CSF shifts on serial head CTs. *Neuroimage Clin.* (2016) 12:673–80. doi: 10.1016/j.nicl.2016.09.018
- Gurney J, Olsen T, Flavin J, Ramaratnam M, Archie K, Ransford J, et al. The Washington University central neuroimaging data archive. *Neuroimage* (2017) 144:287–93. doi: 10.1016/j.neuroimage.2015.09.060
- Strbian D, Meretoja A, Putaala J, Kaste M, Tatlisumak T. Cerebral edema in acute ischemic stroke patients treated with intravenous thrombolysis. *Int J Stroke* (2013) 8:529–34. doi: 10.1111/j.1747-4949.2012.00781.x
- Li X, Morgan PS, Ashburner J, Smith J, Rorden C. The first step for neuroimaging data analysis: DICOM to NIFTI conversion. *J Neurosci Methods* (2016) 264:47–56. doi: 10.1016/j.jneumeth.2016.03.001
- Gorgolewski KJ, Auer T, Calhoun VD, Craddock RC, Das S, Duff EP, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci Data* (2016) 3:160044. doi: 10.1038/sdata.2016.44
- Ratcliffe SJ, Shults J. GEEQBOX: a MATLAB toolbox for generalized estimating equations and quasi-least squares. *J Stat Softw.* (2008) 25:1–14. doi: 10.18637/jss.v025.i14
- Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* (2011) 54:2033–44. doi: 10.1016/j.neuroimage.2010.09.025
- Orellana C, Ferreira D, Muehlboeck JS, Mecocci P, Vellas B, Tsolaki M, et al. Consortium AddNeuronMed, Initiative for the Alzheimer's Disease Neuroimaging. Measuring global brain atrophy with the brain volume/cerebrospinal fluid index: normative values, cut-offs and clinical associations. *Neurodegener Dis.* (2016) 16:77–86. doi: 10.1159/000442443
- Liebeskind DS, Albers GW, Crawford K, Derdeyn CP, George MS, Palesch YY, et al. Imaging in StrokeNet: realizing the potential of big data. *Stroke* (2015) 46:2000–6. doi: 10.1161/STROKEAHA.115.009479
- Muschelli J, Ullman NL, Mould WA, Vespa P, Hanley DF, Crainiceanu CM. Validated automatic brain extraction of head CT images. *Neuroimage* (2015) 114:379–85. doi: 10.1016/j.neuroimage.2015.03.074
- Damon SM, Boyd BD, Plassard AJ, Taylor W, Landman BA. DAX—the next generation: towards one million processes on commodity hardware. *Proc SPIE Int Soc Opt Eng.* (2017) 2017:101380C. doi: 10.1117/12.2254371
- Ronneberger O, Fisher P, Brox T. U-Net: convolutional neural networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, editors. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. Lecture Notes in Computer Science*, Vol. 9351. Cham: Springer International Publishing (2015).

Conflict of Interest Statement: 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.

Copyright © 2018 Dhar, Chen, An and Lee. 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.



A Bayesian Network Model for Predicting Post-stroke Outcomes With Available Risk Factors

Eunjeong Park¹, Hyuk-jae Chang² and Hyo Suk Nam^{3*}

¹ Cardiovascular Research Institute, College of Medicine, Yonsei University, Seoul, South Korea, ² Department of Cardiology, College of Medicine, Yonsei University, Seoul, South Korea, ³ Department of Neurology, College of Medicine, Yonsei University, Seoul, South Korea

OPEN ACCESS

Edited by:

Fabien Scalzo,
University of California, Los Angeles,
United States

Reviewed by:

Jens Fiehler,
Universitätsklinikum
Hamburg-Eppendorf, Germany
Katharina Stibrant Sunnerhagen,
University of Gothenburg, Sweden

*Correspondence:

Hyo Suk Nam
hsnam@yuhs.ac

Specialty section:

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

Received: 30 May 2018

Accepted: 02 August 2018

Published: 07 September 2018

Citation:

Park E, Chang H-j and Nam HS (2018)
A Bayesian Network Model for
Predicting Post-stroke Outcomes With
Available Risk Factors.
Front. Neurol. 9:699.
doi: 10.3389/fneur.2018.00699

Bayesian network is an increasingly popular method in modeling uncertain and complex problems, because its interpretability is often more useful than plain prediction. To satisfy the core requirement in medical research to obtain interpretable prediction with high accuracy, we constructed an inference engine for post-stroke outcomes based on Bayesian network classifiers. The prediction system that was trained on data of 3,605 patients with acute stroke forecasts the functional independence at 3 months and the mortality 1 year after stroke. Feature selection methods were applied to eliminate less relevant and redundant features from 76 risk variables. The Bayesian network classifiers were trained with a hill-climbing searching for the qualified network structure and parameters measured by maximum description length. We evaluated and optimized the proposed system to increase the area under the receiver operating characteristic curve (AUC) while ensuring acceptable sensitivity for the class-imbalanced data. The performance evaluation demonstrated that the Bayesian network with selected features by wrapper-type feature selection can predict 3-month functional independence with an AUC of 0.889 using only 19 risk variables and 1-year mortality with an AUC of 0.893 using 24 variables. The Bayesian network with 50 features filtered by information gain can predict 3-month functional independence with an AUC of 0.875 and 1-year mortality with an AUC of 0.895. We also built an online prediction service, Yonsei Stroke Outcome Inference System, to substantialize the proposed solution for patients with stroke.

Keywords: stroke, bayesian network, prognostic model, machine learning classification, decision support techniques, imbalanced data

INTRODUCTION

A stroke is the second most common cause of death in the world and a leading cause of long-term disability. Patients with stroke have higher mortality than age- and sex-matched subjects who have not experienced a stroke. It is also reported that strokes recur in 6–20% of patients, and approximately two-thirds of stroke survivors continue to have functional deficits that are associated with diminished quality of life (1). Such disability after stroke can be measured by the modified Rankin scale that categorizes functional ability from 0 to 6 (2–4). To discriminate the effect of clinical treatment for patients with ischemic stroke, a score on the modified Rankin scale 0–2 is widely applied for the indication of functional independence after stroke (2).

There are many prognostic models for the functional outcomes and risk of death after stroke. However, an agreed set of guidelines or reporting for the development of prognostic score models are currently unavailable. In a recent systematic review of clinical prediction models, the discriminative performances of models were still unsatisfactory, with the AUC values ranging from 0.60 to 0.72, which are similar to the predictability of experienced clinicians (5).

The prediction of prognosis needs to employ a variety of statistical, probabilistic, and optimization techniques to learn patterns from large, complex, and unbalanced medical data. This complexity challenges researchers to apply machine learning techniques to diagnose and predict the progress of the disease (6, 7). Machine learning has been expected to dramatically improve prognosis, and certain applications have achieved remarkable results (7). These applications have employed various machine learning techniques including a deep neural network (8), support vector machine (8, 9), decision trees (10), and ensemble methods (11, 12) to classify diseases, level of deficits, and mortality. Selecting the optimal solution for a decision problem should consider the unique pattern of a data set and the specific characteristics of the problem (13).

The Bayesian network, a machine learning method, predicts and describes classification based on the Bayes theorem (14). Bayesian networks are widely used in medical decision support for their ability to intuitively encapsulate cause and effect relationships between factors that are stored in medical data (15, 16). With these characteristics of conditional probabilities, the Bayesian network can provide interpretable classifiers by logic inherent in a decision support (17, 18). The parameters and their dependences with conditional probabilities of the Bayesian network can be provided either by experts' knowledge (16, 19) or by automatic learning from data (20, 21). In addition, Bayesian networks can be used to query any given node in the network and are therefore substantially more useful in clinics compared with classifiers built based on specific outcome variables (22).

In this study, our aim was to investigate the usefulness of a machine learning method to forecast functional recovery for independent activities and 1-year mortality in patients with acute ischemic stroke. We also introduced an online inference system for predicting functional independence at 3 months and mortality in 1 year of patients with stroke based on the proposed Bayesian network.

MATERIALS AND METHODS

Data Set

Subjects for this study were selected from consecutive patients with acute ischemic stroke who had been registered in the Yonsei Stroke Registry over a 6.5-year period (January 2007 to June 2013). The Yonsei Stroke Registry is a prospective hospital-based registry for patients with acute ischemic stroke or transient ischemic attack within 7 days after symptom onset (23).

During admission, all patients were thoroughly investigated for medical history, clinical manifestations, and the presence of vascular risk factors. Every patient was evaluated with 12-lead electrocardiography, chest x-ray, lipid profiles, and standard

blood tests. All registered patients underwent brain imaging studies including brain computed tomography (CT) and/or MRI. Angiographic studies using CT angiography, magnetic resonance angiography, or digital subtraction angiography were included in the standard evaluation. Additional blood tests for coagulopathy or prothrombotic conditions were performed in patients younger than 45 years. Transesophageal echocardiography was included in the standard evaluation, except in patients with decreased consciousness, impending brain herniation, poor systemic condition, inability to accept an esophageal transducer because of swallowing difficulty or tracheal intubation, or lack of informed consent (24). Transthoracic echocardiography, heart CT, and Holter monitoring were also performed in selected patients (25). When a patient was admitted more than twice because of recurrent strokes, only data for the first admission were used for this study. Initial stroke severity was determined by National Institute of Health Stroke Scale (NIHSS) scores and score tertiles were used for the analysis.

Hypertension was defined as resting systolic blood pressure ≥ 140 mm Hg or diastolic blood pressure ≥ 90 mm Hg after repeated measurements during hospitalization or currently taking antihypertensive medication. Diabetes mellitus was defined as fasting plasma glucose values ≥ 7 mmol/L or taking an oral hypoglycemic agent or insulin. Hyperlipidemia was diagnosed as a fasting serum total cholesterol level ≥ 6.2 mmol/L, low-density lipoprotein cholesterol ≥ 4.1 mmol/L, or currently taking a lipid-lowering drug after a hyperlipidemia diagnosis. A current smoker was defined as an individual who smoked at the time of stroke or had quit smoking 1 year before treatment (26). The collection of variables during admission including clinical, imaging, and laboratory data were used in statistical analysis and Bayesian network modeling.

Stroke classification was determined during weekly conferences based on the consensus of stroke neurologists. Data including clinical information, risk factors, imaging study findings, laboratory analyses, and other special evaluations were collected. Along with these data, prognosis during hospitalization and long-term outcomes were also determined. Data were entered into a web-based registry. Stroke subtypes were identified according to the Trial of ORG 10172 in Acute Stroke Treatment (TOAST) classification (27).

For target variables in classification, we collected the outcome variables for patients who were followed in the outpatient clinic or by a structured telephone interview at 3 months and every year after discharge. Short-term functional outcomes at 3 months were determined based on the modified Rankin scale. Major disability was defined as a score on the modified Rankin scale of 3–6, as a poor outcome at 3 months after stroke. Deaths among subjects from January 2001 to December 31, 2013, were confirmed by matching the information in the death records and identification numbers assigned to the subjects at birth (5). We obtained data for the date and causes of death from the Korean National Statistical Office, which were identified based on death certificates (28, 29). The institutional review board of Severance Hospital, Yonsei University Health System, approved this study and waived the patients' informed consent

because of a retrospective design and observational nature of this study.

Bayesian Networks

The collected data set was used to construct Bayesian networks for predicting post-stroke outcomes. We extracted a total of 76 random variables of each instance for patient data. A Bayesian network consists of a directed acyclic graph whose nodes represent random variables and links express dependences between nodes. Suppose random variables $V_i \in V$ ($1 \leq i \leq n$). A Bayesian network is described as a directed acyclic graph $G = (V, A, P)$ with links $A \subseteq V \times V$ and P a joint probability distribution. P , a joint probability over V , is described as

$$P(V) = \prod_{V_i \in V} P(V_i | \pi(V_i)),$$

where $\pi(V_i)$ is the set of parent nodes of V_i .

Training Bayesian network classifiers is the process of parameter learning to find optimal Bayesian structures estimating parameter set of P that best represents given data set with labeled instances (13). Given a data set D with variable V_i , the observed distribution P_D is described as a joint probability distribution over D . The learning process now measures and compares the quality of Bayesian networks to evaluate how well the represented distribution explains the given data set. The log-likelihood is the basic common value used for measuring the quality of a Bayesian network as follows:

$$LL(\mathcal{B}|D) = \sum_{V_i} \log(P(V_i | \pi_{\mathcal{B}}(V_i))),$$

where \mathcal{B} is the Bayesian network over D and $\pi_{\mathcal{B}}(V_i)$ is parent nodes of V_i in \mathcal{B} (13, 30).

Diverse quality measurement methods have been investigated (31). The algorithm searched the best Bayesian network based on the Bayesian information criterion (32), Bayesian Dirichlet equivalence score (19), Akaike information criterion (AIC) (33), and the maximum description length (MDL) scores (30, 34). In this study, we used the MDL score to evaluate the quality of a Bayesian network. The MDL score is described as

$$MDL = -LL(\mathcal{B}|D) + \frac{\log N}{2} \cdot |\mathcal{B}|,$$

where N is the number of instances in D , and $|\mathcal{B}|$ is the number of parameters in \mathcal{B} . The smaller the MDL score, the better the network. The search algorithm, greedy hill-climbing algorithm (35) in our study, selects the best Bayesian network by calculating MDL scores of candidate networks. For the type of Bayesian network structure, we constructed tree-augmented network (TAN) structures that restrict the number of parents to two nodes (36).

Prediction Process

The entire process of a Bayesian network-based prediction system is shown in **Figure 1**. A total of 76 features were extracted from the Yonsei Stroke Registry and data preparation process

filtered records with missing outcome variables and exclusion criteria. For feasible prediction service in clinical environment, we performed two different feature selection methods.

Feature selection or dimension reduction is the process of reducing the number of random variables under consideration by obtaining a set of principal variables (37, 38). Feature selection improves the overfitting problem caused by irrelevant or redundant variables that may strongly bias the performance of the classifier. The definition of feature selection in formal expression is described in Drugan and Wiering (30) and Hruschka et al. (39). In many studies, feature selection methods are categorized into filters, wrappers, or embedded methods that are applied to the data set in advance of the training learning algorithm, or to embed feature selection in the learning process (37, 40). Filter methods select features based on a performance measure regardless of the employed data modeling algorithm. The filter approach selects random variables based on information gain score, ReliefF, or correlation-based method by ranking variables or searching subset of variables. Information gain measures the amount of entropy as a measure of uncertainty reduced by knowing a feature (41–43); ReliefF evaluates the worth of an attribute by repeatedly sampling an instance and considering the value of the given attribute for the nearest instance of the same and the different class (44, 45); and correlation evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them (46, 47). Unlike the filter approach, wrapper methods measure the usefulness of a subset of features by actually training a model on it. We evaluated the performance of Bayesian networks with a reduced variable set selected by information gain and Bayesian network algorithms that are popular in filter and wrapper methods (42, 48, 49).

First, we tested the Bayesian network classifier with features chosen by information gain based on entropy of each feature. The other feature selection method, considering the characteristics of Bayesian network classifiers, reduces the variable set by evaluating the performance of the Bayesian network classifier in cross-validation in which a search algorithm extracts a subset of attributes to maximize AUC in prediction (**Figure 1**). The optimization for AUC is to solve the imbalance between the number of survival and mortal subjects.

Using the reduced variables by feature selection, the system constructed a Bayesian network prediction model to search optimal Bayesian network structures and parameters. We evaluated the performance of prediction algorithms using (1) a basic tree-augmented Bayesian network, (2) a tree-augmented Bayesian network with features filtered by information gain, and (3) a tree-augmented Bayesian network with features filtered by the wrapper of a Bayesian network. The performances of all Bayesian networks and predictive models were evaluated based on the AUC, specificity, and sensitivity of 10-fold cross-validations (50). We also implemented an online prediction system for post-stroke outcomes embedding the trained classifiers. In the validation process, we bound the minimum sensitivity as 0.50 to utilize the trained classifiers in real-world applications with imbalanced data.

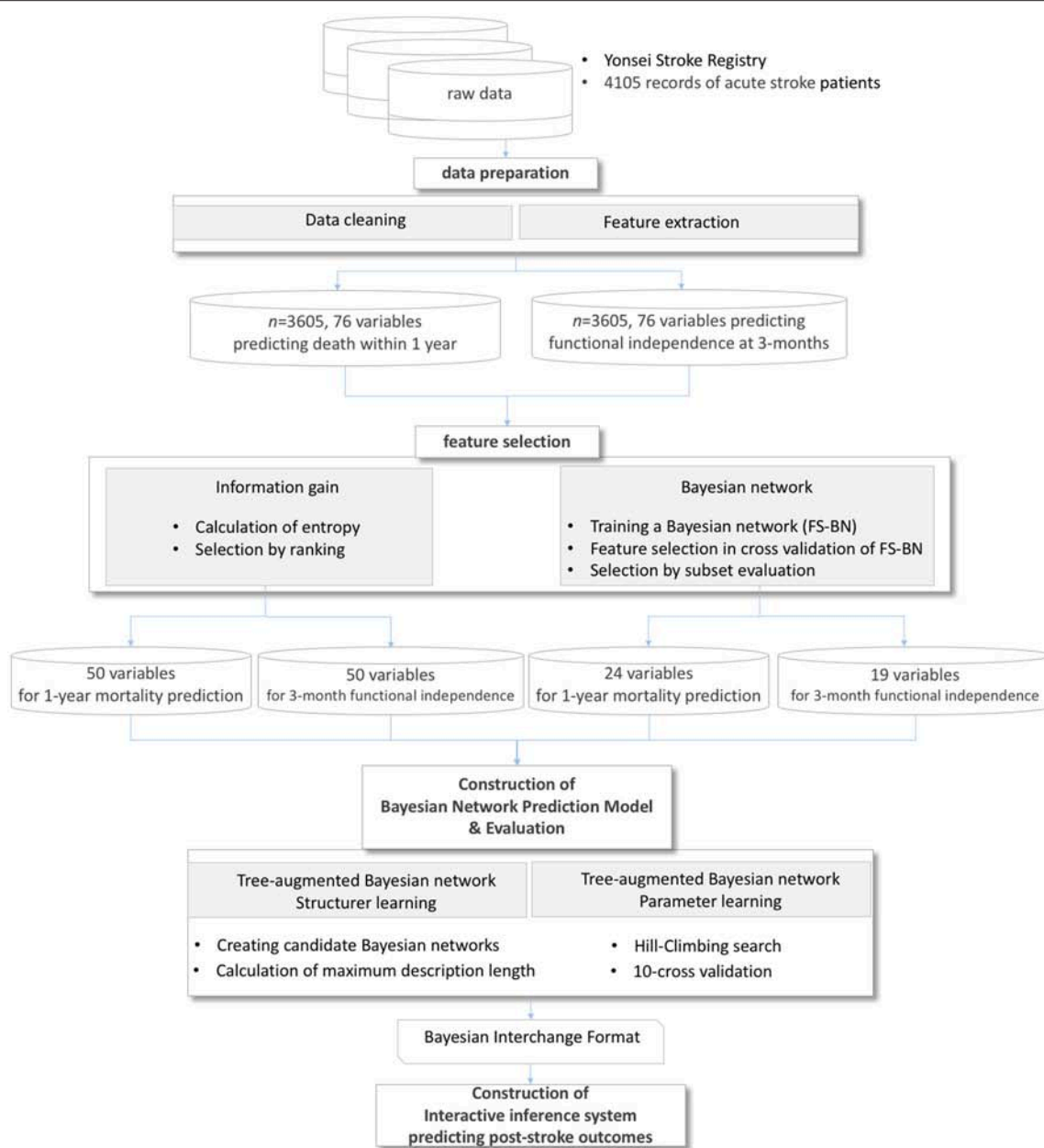


FIGURE 1 | Process of a prediction system for post-stroke outcomes.

RESULTS

Statistical Characteristics

During the study period, 4,105 consecutive patients with acute ischemic stroke or transient ischemic attack were registered to the Yonsei Stroke Registry. Exclusion criteria of this study were patients with the stroke subtypes other than cryptogenic stroke including transient ischemic attack ($n = 326$), foreigner ($n = 48$), missing data ($n = 29$), follow-up loss ($n = 97$). After exclusion, a total of 3,605 patients were finally enrolled for this study. The mean age was 65.9 ± 12.6 years, and 60.7% were men. A comparison of demographic characteristics between the outcome

at 3 months and death within 1 year is shown at **Table 1**. Patients with poor outcome were older, more likely to be women, not a current smoker, frequently had old stroke, hypertension, atrial fibrillation, congestive heart failure, peripheral artery obstructive disease, or anemia. Thrombolysis or endovascular mechanical thrombectomy, symptomatic intracranial hemorrhage, and herniation are frequent in patients with poor outcome. Laboratory data showed that patients with poor outcome showed lower hemoglobin, hematocrit, albumin, prealbumin, body weight and higher ESR, fibrinogen, hsCRP, and D-dimer level. The differences of demographics of patients between survival and

TABLE 1 | Demographic characteristics and comparison of outcome at 3 months and death within 1 year.

	Total	Outcome at 3 months			Death within 1 year		
	(N = 3,605)	Good outcome (N = 2,653)	Poor outcome (N = 952)	p	No (N = 3,171)	Yes (N = 434)	p
Age	65.9 ± 12.6	64.0 ± 12.3	71.2 ± 11.9	< 0.001	64.8 ± 12.4	73.9 ± 11.2	< 0.001
Sex				< 0.001			0.016
F	1,416 (39.3%)	969 (36.5%)	447 (47.0%)		1,222 (38.5%)	194 (44.7%)	
M	2,189 (60.7%)	1,684 (63.5%)	505 (53.0%)		1,949 (61.5%)	240 (55.3%)	
Hypertension	2,675 (74.2%)	1,940 (73.1%)	735 (77.2%)	0.015	2,675 (74.2%)	1,940 (73.1%)	0.023
Diabetes	1,144 (31.7%)	827 (31.2%)	317 (33.3%)	0.243	1,144 (31.7%)	827 (31.2%)	0.282
Hypercholesterolemia	747 (20.7%)	554 (20.9%)	193 (20.3%)	0.726	685 (21.6%)	62 (14.3%)	0.001
Current smoking	856 (23.7%)	704 (26.5%)	152 (16.0%)	< 0.001	856 (23.7%)	704 (26.5%)	< 0.001
Old stroke	472 (13.1%)	301 (11.3%)	171 (18.0%)	< 0.001	401 (12.6%)	71 (16.4%)	0.038
Atrial fibrillation	813 (22.6%)	482 (18.2%)	331 (34.8%)	< 0.001	623 (19.6%)	190 (43.8%)	< 0.001
Coronary artery disease	811 (22.5%)	603 (22.7%)	208 (21.8%)	0.608	717 (22.6%)	94 (21.7%)	0.701
Congestive heart failure	184 (5.1%)	110 (4.1%)	74 (7.8%)	< 0.001	134 (4.2%)	50 (11.5%)	< 0.001
Peripheral artery obstructive disease	110 (3.1%)	60 (2.3%)	50 (5.3%)	< 0.001	85 (2.7%)	25 (5.8%)	0.001
Initial NIHSS score	5.6 ± 6.3	3.4 ± 4.0	11.5 ± 7.5	< 0.001	4.8 ± 5.4	11.5 ± 8.4	< 0.001
TOAST				< 0.001			< 0.001
LAC	321 (8.9%)	285 (10.7%)	36 (3.8%)		312 (9.8%)	9 (2.1%)	
LAA	741 (20.6%)	504 (19.0%)	237 (24.9%)		661 (20.8%)	80 (18.4%)	
CE	991 (27.5%)	688 (25.9%)	303 (31.8%)		823 (26.0%)	168 (38.7%)	
SOD	89 (2.5%)	68 (2.6%)	21 (2.2%)		80 (2.5%)	9 (2.1%)	
UT	668 (18.5%)	498 (18.8%)	170 (17.9%)		587 (18.5%)	81 (18.7%)	
UN	785 (21.8%)	607 (22.9%)	178 (18.7%)		703 (22.2%)	82 (18.9%)	
UI	10 (0.3%)	3 (0.1%)	7 (0.7%)		5 (0.2%)	5 (1.2%)	
Anemia	617 (17.1%)	361 (13.6%)	256 (26.9%)	< 0.001	450 (14.2%)	167 (38.5%)	< 0.001
Thrombolysis	485 (13.5%)	272 (10.3%)	213 (22.4%)	< 0.001	377 (11.9%)	108 (24.9%)	< 0.001
Symtomatic ICH	92 (2.6%)	10 (0.4%)	82 (8.6%)	< 0.001	43 (1.4%)	49 (11.3%)	< 0.001
Herniation	105 (2.9%)	3 (0.1%)	102 (10.7%)	< 0.001	38 (1.2%)	67 (15.4%)	< 0.001
Body weight	62.9 ± 11.1	64.0 ± 10.9	60.0 ± 11.2	< 0.001	63.6 ± 11.0	57.8 ± 10.8	< 0.001
hgb	13.8 ± 2.0	14.0 ± 1.9	13.3 ± 2.2	< 0.001	14.0 ± 1.9	12.7 ± 2.3	< 0.001
hct	40.6 ± 5.6	41.1 ± 5.3	39.3 ± 6.1	< 0.001	41.0 ± 5.3	37.9 ± 6.5	< 0.001
esr	23.9 ± 22.2	21.2 ± 20.1	31.3 ± 25.8	< 0.001	22.1 ± 20.6	36.5 ± 28.8	< 0.001
pt	1.0 ± 0.5	1.0 ± 0.3	1.0 ± 0.7	0.123	1.0 ± 0.5	1.0 ± 0.2	0.002
Albumin	4.2 ± 0.5	4.3 ± 0.4	4.0 ± 0.5	< 0.001	4.3 ± 0.4	3.9 ± 0.6	< 0.001
Prealbumin	223.7 ± 72.6	239.0 ± 69.9	205.6 ± 71.6	< 0.001	233.3 ± 69.8	186.8 ± 71.4	< 0.001
Fibrinogen	322.8 ± 94.3	316.1 ± 83.9	341.5 ± 116.8	< 0.001	320.1 ± 88.5	342.5 ± 128.5	0.001
hsCRP	11.3 ± 48.4	7.5 ± 49.7	22.2 ± 42.7	< 0.001	9.2 ± 48.4	27.3 ± 45.5	< 0.001
D-dimer	779.0 ± 3846.1	418.4 ± 1704.4	1788.2 ± 6834.6	< 0.001	464.5 ± 1759.3	3079.8 ± 9723.3	< 0.001

death within 1 year were similar with functional outcome at 3 months. D-dimer levels were significantly higher in patients who died within 1 year compared with survivors (3079.8 ± 9723.3 vs. 464.5 ± 1759.3 , $p < 0.001$).

Structure and Parameters of Bayesian Networks

As we described in **Figure 1**, two different feature selection techniques were performed in our experiment: variables selected by information gain with ranking or variables selected by a

wrapper embedding Bayesian network with greedy stepwise subset selection in cross-validation. The top-ranked variables in the filter by information gain and the wrapper of the Bayesian network in forecasting functional independence at 3 months are shown in **Figures 2A,B**, and variables for predicting 1-year mortality are shown in **Figures 2C,D**. The most affective factor for functional recovery prediction was Initial NIHSS, while D-dimer ranked top in 1-year mortality prediction. The common variables for predicting post-stroke outcomes were Initial NIHSS, D-dimer, hsCPR, and Age. However, the subset-searching algorithm selects a method differently from the

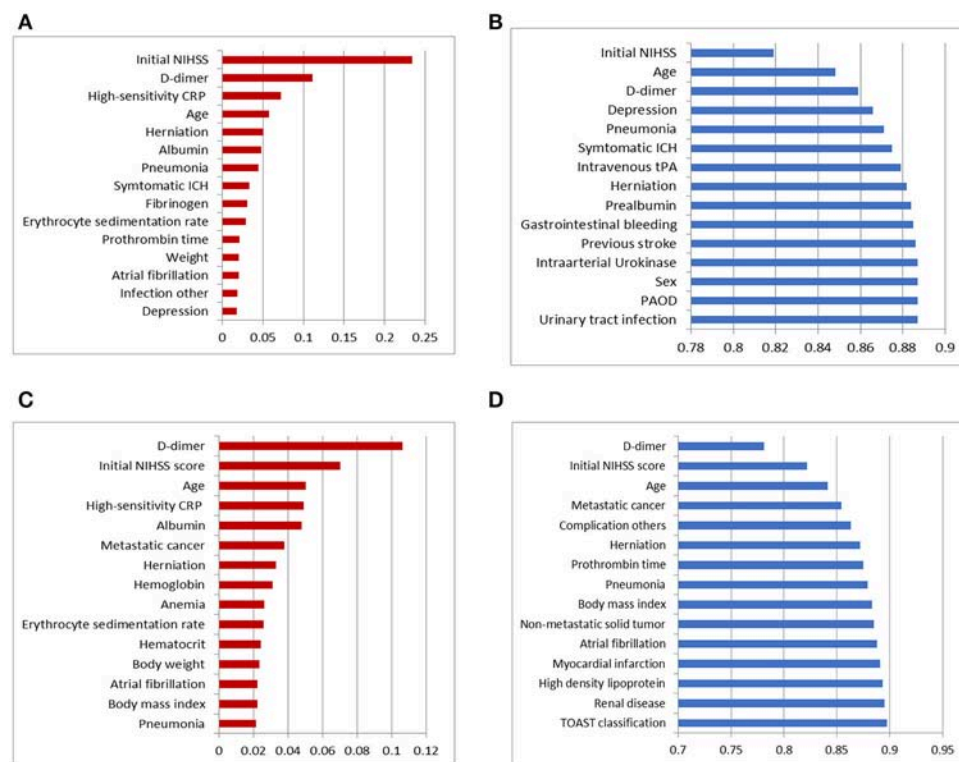


FIGURE 2 | Top 15 variables in dimension reduction for post-stroke outcome prediction: **(A)** variables filtered by ranks of information gain for predicting functional independence at 3 months, **(B)** variables selected by the wrapper of the Bayesian network classifier with greedy subset selection for predicting functional independence at 3 months, **(C)** variables filtered by ranks of information gain for predicting 1-year mortality, and **(D)** variables selected by the wrapper of the Bayesian network classifier with greedy subset selection for predicting 1-year mortality.

ranking method that evaluates the individual variables separately; thus, certain variables were excluded from the selected subset even though their ranks are high in individual evaluation.

Using the result of feature selection, we trained three tree-augmented Bayesian network classifiers; (1) Tree-augmented Bayesian network with the entire dataset, (2) tree-augmented Bayesian network with features filtered by ranking of information gain, and (3) tree-augmented Bayesian network with features filtered by the wrapper of the Bayesian network classifier (see **Figure 3**). The predictive performance for 3-month outcomes is shown in **Figure 3A**. The classifier trained with features chosen by the Bayesian network's subset evaluation performs in prediction of 3-month functional recovery with the specificity of 0.931, accuracy of 0.643, and AUC of 0.889 (95% CI, 0.879–0.899) although the sensitivity (0.643) is slightly lower than other algorithms. The tree-augmented Bayesian network without feature selection achieved the AUC of 0.875 (95% CI, 0.864–0.886), but the highest sensitivity of 0.684; and the Bayesian network with features by ranking of information gain obtained the AUC of 0.875 (95% CI, 0.864–0.886) and mid-level performance between two other algorithms. The Bayesian network classifier with feature selection achieved best performance in most metrics except sensitivity, although it reduced the variable set from 76 variables to 19 variables, resulting in a great reduction in model construction time.

In prediction of 1-year mortality, AUCs of three algorithms were not significantly different (0.892 with 95% CI, 0.872–0.912; 0.895 with CI, 0.875–0.915; and 0.893 with CI, 0.873–0.913). All algorithms achieved higher specificities in predicting 1-year mortality than those for the prediction of functional independence (0.915 vs. 0.897 with a basic Bayesian network, 0.915 vs. 0.898 with a Bayesian network with features filtered by information gain, and 0.943 vs. 0.931 with a Bayesian network with features chosen by the wrapper of the Bayesian network classifier). The Bayesian network algorithm with feature selection for 1-year mortality cuts out the entire variable set to 24 variables that curtail network construction time. The final Bayesian networks predicting functional recovery and 1-year mortality are shown in **Figures 4, 5**, respectively.

Online Interactive System for Predicting Post-stroke Outcomes

To realize decision support using Bayesian network classifiers, we embedded our final Bayesian networks into an online inference system, Y-SOIS (Yonsei-Stroke Outcome Inference System, https://www.hed.cc/?a=Yonsei_SOIS), that enables answering post-stroke outcomes when users provide available risk variables. **Figure 6** shows the screenshots of Y-SOIS.

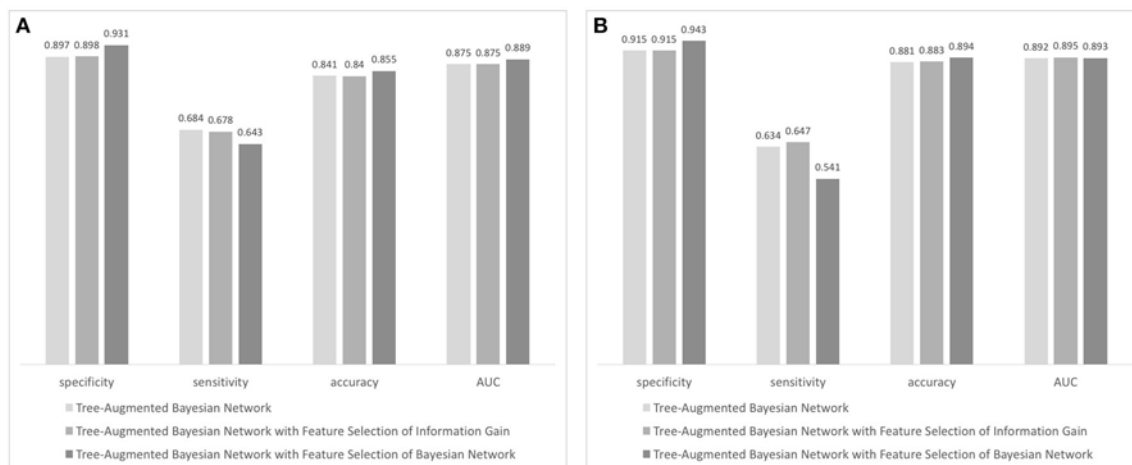


FIGURE 3 | Performance evaluation of Bayesian network-based classifiers: **(A)** performance of classifiers forecasting 90-day functional independence and **(B)** performance of classifiers for 1-year mortality prediction.

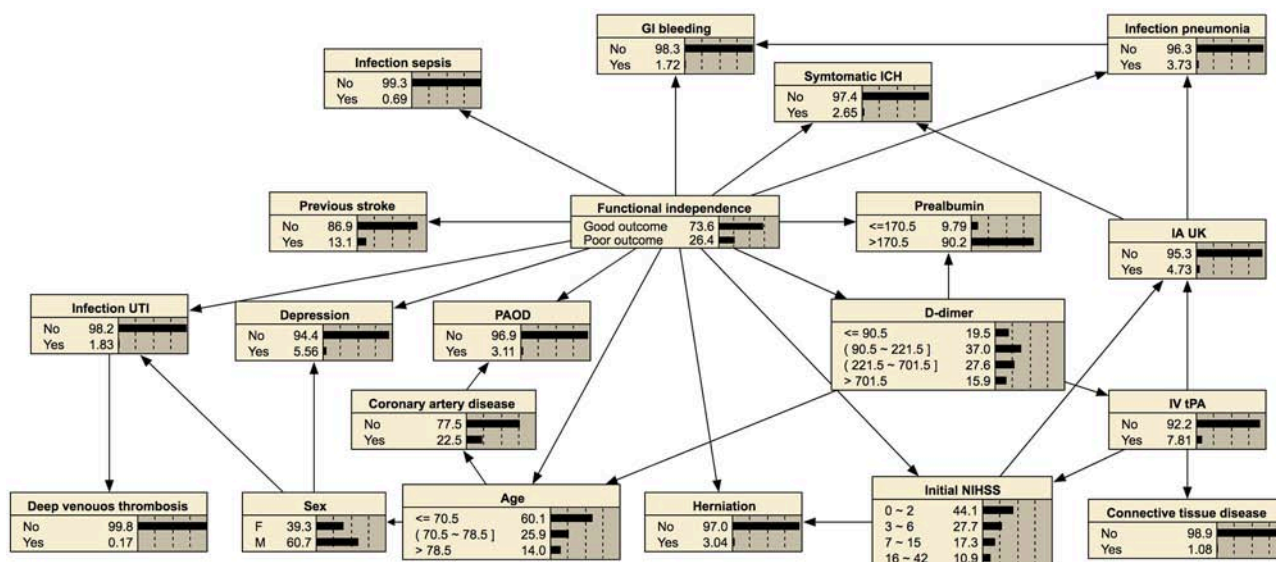


FIGURE 4 | Bayesian network for predicting functional independence at 3 months. The tree-augmented Bayesian network used 19 variables selected by the wrapper of the Bayesian network for prediction.

DISCUSSION

Interpretability is a core requirement for machine learning models in medicine, because both patients and physicians need to understand the reason behind a prediction (51). This study presents an evaluation of Bayesian networks in providing post-stroke outcomes estimates based on the collected demographic data, lab result, and initial neurological assessment. The stroke-specific variables were selected from a large stroke registry, and our experiment filtered those variables into the Bayesian network-suitable reduced set. The trained Bayesian networks were embedded in our online prediction system.

Strength of a Bayesian Network on Stroke Outcome Measurements

Research on stroke outcomes is essential for both clinical care and policy development, because approximately two-thirds of stroke survivors continue to experience functional deficits and approximately 1 of 10 patients died within 1 year (5). The prediction of post-stroke outcomes thus requires high accuracy in classification along with the understandable result that can be explained to patients. A Bayesian network can intuitively make connections between variables in medical data and provide interpretable determination in medical decision (17, 18). Therefore, Bayesian networks are well suited for

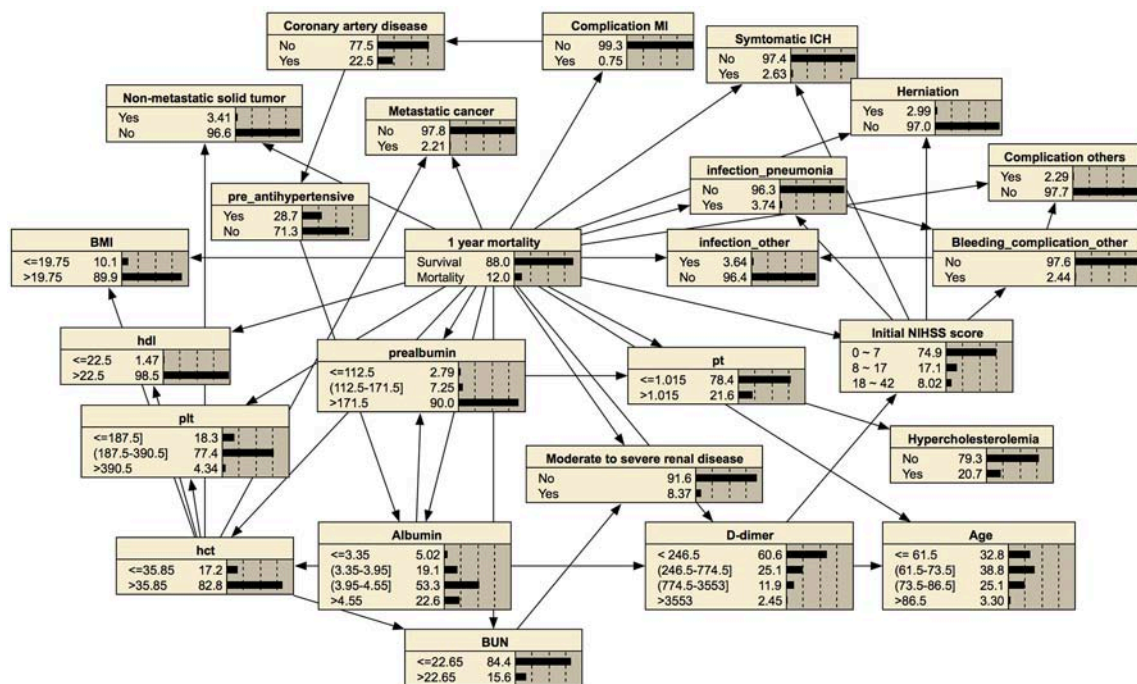


FIGURE 5 | Bayesian network for predicting 1-year mortality. The tree-augmented Bayesian network used 24 variables selected by the wrapper of the Bayesian network for prediction.

representing uncertainty and causality in prediction for patients with stroke. In recent machine learning studies, a Bayesian neural network is focused on a state of the art method which estimates predictive uncertainty (52). In Kendall and Gal (53), a Bayesian deep learning framework combines input-dependent aleatoric uncertainty together with epistemic uncertainty, to solve the black-box problem in deep learning. Constructing Bayesian networks enables medical diagnosis or prediction with incomplete and partially correct statistics, because it determines causes and effects based on the conditional probability between variables (54).

Prediction With Imbalanced Data

Often real-world data sets are predominately composed of normal instances with only a small percentage of interesting instances; therefore, class imbalance is one of the most important challenges (55). Our study also has heavily unbalanced classes in mortality prediction (3,171:434). Suppose entire positive instances were classified into negative class; then the accuracy is 0.880 in 1-year mortality prediction, although mortality is not predicted at all. Most machine learning algorithms train classifiers mainly searching for higher accuracy; therefore, the minority class is less considered in the training process. To challenge this imbalanced classification, a number of techniques have been proposed (56): oversampling approaches create minority instances by simple duplication or synthetic-minority oversampling technique (SMOTE) (57–59); certain classifiers with undersampling beat oversampling (60); cost-sensitive methods weigh higher penalty on misclassification

of the minority class (61); and bagging, boosting, and hybrid approaches utilize feedback from misclassification in previous stages of learning (62).

In addition to the capability of interpretable prediction and reduced uncertainty, a Bayesian network is strong machine learning in classifying an imbalanced data set as investigated in Drummond and Holte (60) and Monsalve-Torra et al. (63). In Monsalve-Torra et al. (63), the Bayesian network outperformed radial basis function and multilayer perceptron in sensitivity. In our experiment, the learning process searched the best Bayesian network structure and parameters for the highest AUC while it guarantees at least 0.5 in sensitivity. A more computation-expensive searching algorithm such as repeated hill climbing might be helpful to increase sensitivity in classification.

Visualized Probability of Outcomes After Stroke

Bayesian networks can also provide a visual graph structure. We constructed a tree-augmented Bayesian network structure that shows an association between nodes. This visualization of conditional probability might be helpful for clinical reasoning. For example, a Bayesian network can provide the association among symptomatic intracranial hemorrhage, higher initial NIHSS score, or higher 1-year mortality with conditional probability, as shown in Figure 5. Therefore, our prediction model of post-stroke outcomes differs from the black-box concept of other machine learning methods (54).

A

Yonsei Stroke Outcome Inference System
Powered by AutoNetica™

Site: Yonsei_SOIS

Topic: SOIS-3-month mRS Reload

Restart Save Case Restore Case

Functional Independence

Questions

Prealbumin: ☐ ≤170.5 ☐ >170.5 ☐ (Skip)
 IA UK: ☐ No ☐ Yes ☐ (Skip)
 Infection pneumonia: ☐ No ☐ Yes ☐ (Skip)
 Depression: ☐ No ☐ Yes ☐ (Skip)
 Infection sepsis: ☐ No ☐ Yes ☐ (Skip)
 Infection UTI: ☐ No ☐ Yes ☐ (Skip)
 Previous stroke: ☐ No ☐ Yes ☐ (Skip)
 PAOD: ☐ No ☐ Yes ☐ (Skip)

Results

Functional Independence: Good outcome = 18.4%

Answered

IV tPA: ☐ No ☒ Yes ☐ Unknown
 Age: ☐ ≤70.5 ☒ (70.5 ~ 78.5] ☐ >78.5 ☐ Unknown
 Herniation: ☒ No ☐ Yes ☐ Unknown
 GI bleeding: ☒ No ☐ Yes ☐ Unknown
 Symptomatic ICH: ☐ No ☒ Yes ☐ Unknown
 D-dimer: ☐ ≤90.5 ☐ (90.5 ~ 221.5] ☐ (221.5 ~ 701.5] ☒ >701.5 ☐ Unknown
 Initial NIHSS: ☐ 0 ~ 2 ☒ 3 ~ 6 ☐ 7 ~ 15 ☐ 16 ~ 42 ☐ Unknown

B

Yonsei Stroke Outcome Inference System
Powered by AutoNetica™

Site: Yonsei_SOIS

Topic: SOIS-1-year mortality Reload

Restart Save Case Restore Case

Questions

Metastatic cancer: ☐ No ☐ Yes ☐ (Skip)
 prealbumin:
 pt: ☐ ≤1,015 ☐ >1,015 ☐ (Skip)
 plt:
 infection_other: ☐ Yes ☐ No ☐ (Skip)
 Complication others: ☐ Yes ☐ No ☐ (Skip)
 Moderate to severe renal disease: ☐ No ☐ Yes ☐ (Skip)
 BUN: ☐ ≤22.65 ☐ >22.65 ☐ (Skip)

Results

1 year mortality: Survival = 38.1%

Answered

D-dimer: (774.5-3553]
 Age: (61.5-73.5]
 Initial NIHSS score: 8 ~ 17
 Herniation: ☐ Yes ☒ No ☐ Unknown
 Symptomatic ICH: ☒ No ☐ Yes ☐ Unknown
 infection_pneumonia: ☐ No ☒ Yes ☐ Unknown
 BMI: ☒ ≤19.75 ☐ >19.75 ☐ Unknown

FIGURE 6 | Screenshots of an online prediction system, Y-SOIS (Yonsei Stroke Outcome Inference System). **(A)** Y-SOIS forecasts the functional independence at 3 months and **(B)** Y-SOIS forecasting the 1-year mortality.

The reduction of dimension is also helpful to visualize inference of prediction. The results demonstrated that the Bayesian network classifier with a reduced variable set can adapt the size of a network for better interpretability with a minimal or better impact on other performance.

Predictors of Post-stroke Outcomes

In this study, the information gain analysis showed that “D-dimer” was the highest feature in predicting 1-year mortality. We previously reported that a high D-dimer level by itself appeared to be associated with an increased risk of mortality

(64). D-dimer can be found to be elevated in various thrombotic and inflammatory conditions, including ischemic heart disease, infection, or malignancy. These conditions are frequently found in patients with stroke and can increase the risk of mortality (65). However, patients with comorbid diseases were frequently excluded from the clinical trials, so there are no guidelines and evidence whether to treat or not patients with serious comorbid diseases in real clinical practice. In this respect, providing information of the impact of the comorbid condition with a Bayesian network might be helpful to predict the outcomes.

LIMITATIONS AND FUTURE DIRECTION

This study was conducted in a single university hospital and focused on those of East Asian descent. To provide generalizability on our prediction system, we will include various cohorts including different ethnics or patients who received thrombolysis or endovascular thrombectomy. We have plan to apply the interpretable prediction for the SECRET (SElection CRiteria in Endovascular thrombectomy and Thrombolytic therapy) study, which is a nationwide registry for hyperacute stroke. Consecutive patients who received intravenous thrombolysis and/or endovascular thrombectomy were registered (Clinical Trial Registration: NCT02964052). Bayesian network analysis of this specific condition can be used to predict outcome in patients with hyperacute stroke. We will also enlarge our training data including data of various populations by applying the proposed solution to global data archives. Additive risk predictors might be selected as determinant features in a

Bayesian network, and it makes the prediction system more applicable in a global clinical environment.

AUTHOR CONTRIBUTIONS

HN designed the study; EP analyzed the data and wrote the manuscript; and H-jC and HN contributed to data interpretation and revising the manuscript.

FUNDING

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2017R1D1A1B03029014) and grant funded by the Korea government (MSIP) (2016R1C1B2016028) and the National Fire Agency, Republic of Korea (MPSS-2015-70).

REFERENCES

- Ko Y, Park JH, Kim WJ, Yang MH, Kwon OK, Oh CW, et al. The long-term incidence of recurrent stroke: single hospital-based cohort study. *J Korean Neurol Assoc.* (2009) 27:110–5.
- 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. *New Engl J Med.* (2018) 378:708–18. doi: 10.1056/NEJMoa1713973
- Anderson CS, Woodward M, Chalmers J. More on low-dose versus standard-dose intravenous alteplase in acute ischemic stroke. *New Engl J Med.* (2018) 378:1465–6. doi: 10.1056/NEJMc1801548
- Banks JL, Marotta CA. Outcomes validity and reliability of the modified Rankin scale: implications for stroke clinical trials: a literature review and synthesis. *Stroke* (2007) 38:1091–6. doi: 10.1161/01.STR.0000258355.23810.c6
- Nam HS, Kim HC, Kim YD, Lee HS, Kim J, Lee DH, et al. Long-term mortality in patients with stroke of undetermined etiology. *Stroke* (2012) 43:2948–56. doi: 10.1161/STROKEAHA.112.661074
- Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer Inf.* (2006) 2:117693510600200030. doi: 10.1177/117693510600200030
- Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. *New Engl J Med.* (2016) 375:1216. doi: 10.1056/NEJMp1606181
- Asadi H, Dowling R, Yan B, Mitchell P. Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy. *PLoS ONE* (2014) 9:e88225. doi: 10.1371/journal.pone.0088225
- Magnin B, Mesrob L, Kinkingnéhun S, Péligrini-Issac M, Colliot O, Sarazin M, et al. Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. *Neuroradiology* (2009) 51:73–83. doi: 10.1007/s00234-008-0463-x
- Tjortjjs C, Saraae M, Theodoulidis B, Keane J. Using T3, an improved decision tree classifier, for mining stroke-related medical data. *Methods Inf Med.* (2007) 46:523–9. doi: 10.1160/ME0317
- Ward MM, Pajevic S, Dreyfuss J, Malley JD. Short-term prediction of mortality in patients with systemic lupus erythematosus: classification of outcomes using random forests. *Arthritis Care Res.* (2006) 55:74–80. doi: 10.1002/art.21695
- Statnikov A, Wang L, Aliferis CF. A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. *BMC Bioinform.* (2008) 9:319. doi: 10.1186/1471-2105-9-319
- Witten IH, Frank E, Hall MA, Pal CJ. *Data Mining: Practical Machine Learning Tools and Techniques*. Burlington, MA: Morgan Kaufmann (2016).
- Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. *Mach Learn.* (1997) 29:131–63. doi: 10.1023/A:1007465528199
- Lucas PJ, Van der Gaag LC, Abu-Hanna A. Bayesian networks in biomedicine and health-care. *Artif Intell Med.* (2004) 30:201–14. doi: 10.1016/j.artmed.2003.11.001
- Nikovski D. Constructing Bayesian networks for medical diagnosis from incomplete and partially correct statistics. *IEEE Trans Knowl Data Eng.* (2000) 12:509–16. doi: 10.1109/69.868904
- Letham B, Rudin C, McCormick TH, Madigan D. Interpretable classifiers using rules and bayesian analysis: building a better stroke prediction model. *Ann Appl Stat.* (2015) 9:1350–71. doi: 10.1214/15-AOAS848
- Jansen R, Yu H, Greenbaum D, Kluger Y, Krogan NJ, Chung S, et al. A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science* (2003) 302:449–53. doi: 10.1126/science.1087361
- Heckerman D, Geiger D, Chickering DM. Learning bayesian networks: the combination of knowledge and statistical data. *Mach Learn.* (1995) 20:197–243. doi: 10.1007/BF00994016
- Uusitalo L. Advantages and challenges of Bayesian networks in environmental modelling. *Ecol Model.* (2007) 203:312–8. doi: 10.1016/j.ecolmodel.2006.11.033
- Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. *Artif Intell Med.* (2001) 23:89–109. doi: 10.1016/S0933-3657(01)00077-X
- Sesen MB, Nicholson AE, Banares-Alcantara R, Kadir T, Brady M. Bayesian networks for clinical decision support in lung cancer care. *PLoS ONE* (2013) 8:e82349. doi: 10.1371/journal.pone.0082349
- Lee BI, Nam HS, Heo JH, Kim DI. Yonsei stroke registry. *Cerebrovasc Dis.* (2001) 12:145–51. doi: 10.1159/000047697
- Cho HJ, Choi HY, Kim YD, Nam HS, Han SW, Ha JW, et al. Transoesophageal echocardiography in patients with acute stroke with sinus rhythm and no cardiac disease history. *J Neurol Neurosurg Psychiatry* (2010) 81:412–5. doi: 10.1136/jnnp.2009.190322
- Yoo J, Yang JH, Choi BW, Kim YD, Nam HS, Choi HY, et al. The frequency and risk of preclinical coronary artery disease detected using multichannel cardiac computed tomography in patients with ischemic stroke. *Cerebrovasc Dis.* (2012) 33:286–94. doi: 10.1159/000334980
- Song TJ, Kim J, Lee HS, Nam CM, Nam HS, Kim YD, et al. Distribution of cerebral microbleeds determines their association with impaired kidney function. *J Clin Neurol.* (2014) 10:222–8. doi: 10.3988/jcn.2014.10.3.222
- Adams HP Jr, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke* (1993) 24:35–41. doi: 10.1161/01.STR.24.1.35

28. Khang Y-H, Lynch JW, Kaplan GA. Health inequalities in Korea: age- and sex-specific educational differences in the 10 leading causes of death. *Int J Epidemiol.* (2004) 33:299–308. doi: 10.1093/ije/dyg244
29. Kim HC, Choi DP, Ahn SV, Nam CM, Suh I. Six-year survival and causes of death among stroke patients in Korea. *Neuroepidemiology* (2009) 32:94–100. doi: 10.1159/000177034
30. Drugan MM, Wiering MA. Feature selection for Bayesian network classifiers using the MDL-FS score. *Int J Approx Reason.* (2010) 51:695–717. doi: 10.1016/j.ijar.2010.02.001
31. Liu Z, Malone B, Yuan C. Empirical evaluation of scoring functions for Bayesian network model selection. *BMC Bioinformatics* (2012) 13(Suppl. 15):S14. doi: 10.1186/1471-2105-13-S15-S14
32. Schwarz G. Estimating the dimension of a model. *Ann Stat.* (1978) 6:461–4. doi: 10.1214/aos/1176344136
33. Vrieze SI. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychol Methods* (2012) 17:228. doi: 10.1037/a0027127
34. Lam W, Bacchus F. Learning Bayesian belief networks: an approach based on the MDL principle. *Comput Intell.* (1994) 10:269–93. doi: 10.1111/j.1467-8640.1994.tb00166.x
35. Tsamardinos I, Brown LE, Aliferis CF. The max-min hill-climbing Bayesian network structure learning algorithm. *Mach Learn.* (2006) 65:31–78. doi: 10.1007/s10994-006-6889-7
36. Chinnasamy A, Sung W-K, Mittal A. Protein structure and fold prediction using tree-augmented naive Bayesian classifier. *J Bioinform Comput Biol.* (2005) 3:803–19. doi: 10.1142/S0219720005001302
37. Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res.* (2003) 3:1157–82.
38. Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. *science.* (2000) 290:2323–6. doi: 10.1126/science.290.5500.2323
39. Hruschka ER, Hruschka ER, Ebecken NF. Feature selection by Bayesian networks. In: *Conference of the Canadian Society for Computational Studies of Intelligence*. London, ON: Springer (2004).
40. Jović A, Brkić K, Bogunović N. A review of feature selection methods with applications. In: *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) Opatija* (2015).
41. Lee C, Lee GG. Information gain and divergence-based feature selection for machine learning-based text categorization. *Informat Proc Manag.* (2006) 42:155–65. doi: 10.1016/j.ipm.2004.08.006
42. Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah MH, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. *Eur Heart J.* (2016) 38:500–7. doi: 10.1093/eurheartj/ehw188
43. Lei S. A feature selection method based on information gain and genetic algorithm. In: *Computer Science and Electronics Engineering (ICCSEE), 2012 International Conference on Hangzhou*: IEEE (2012).
44. Kononenko I. Estimating attributes: analysis and extensions of RELIEF. In: *European Conference on Machine Learning*. Catania: Springer (1994).
45. Robnik-Šikonja M, Kononenko I. An adaptation of Relief for attribute estimation in regression. In: *Proceedings of the 14th International Conference on Machine Learning (ICML)*. San Francisco, CA (1997).
46. Hall MA. Correlation-based feature selection of discrete and numeric class machine learning. In: *ICML '00 Proceedings of the Seventeenth International Conference on Machine Learning*. San Francisco, CA (2000).
47. Yu L, Liu H. Feature selection for high-dimensional data: a fast correlation-based filter solution. In: *Proceedings of the 20th International Conference on Machine Learning (ICML)*. Washington, DC (2003).
48. Bermejo P, Gamez JA, Puerta JM. Improving incremental wrapper-based subset selection via replacement and early stopping. *Int J Pattern Recogn Artif Intell.* (2011) 25:605–25. doi: 10.1142/S0218001411008804
49. Kohavi R, John GH. Wrappers for feature subset selection. *Artif Intell.* (1997) 97:273–324. doi: 10.1016/S0004-3702(97)00043-X
50. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. *Stat Surv.* (2010) 4:40–79. doi: 10.1214/09-SS054
51. Valdes G, Luna JM, Eaton E, Simone II CB, Ungar LH, Solberg TD. MediBoost: a patient stratification tool for interpretable decision making in the era of precision medicine. *Sci Rep.* (2016) 6:37854. doi: 10.1038/srep37854
52. Lakshminarayanan B, Pritzel A, Blundell C. Simple and scalable predictive uncertainty estimation using deep ensembles. In: *Advances in Neural Information Processing Systems*. Long Beach, CA (2017).
53. Kendall A, Gal Y. What uncertainties do we need in bayesian deep learning for computer vision? In: *Advances in Neural Information Processing Systems*. Long Beach, CA (2017).
54. Korb KB, Nicholson AE. *Bayesian Artificial Intelligence*. Boca Raton, FL: CRC Press (2010).
55. Maldonado S, Weber R, Famili F. Feature selection for high-dimensional class-imbalanced data sets using Support Vector Machines. *Inf Sci.* (2014) 286:228–46. doi: 10.1016/j.ins.2014.07.015
56. Fernández A, García S, Herrera F. Addressing the classification with imbalanced data: open problems and new challenges on class distribution. In: *International Conference on Hybrid Artificial Intelligence Systems*. Wroclaw: Springer (2011).
57. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res.* (2002) 16:321–57. doi: 10.1613/jair.953
58. Han H, Wang W-Y, Mao B-H. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In: *International Conference on Intelligent Computing*. Hefei: Springer (2005).
59. Ho KC, Speier W, El-Saden S, Liebeskind DS, Saver JL, Bui AA, et al. Predicting discharge mortality after acute ischemic stroke using balanced data. In: *AMIA Annual Symposium Proceedings, American Medical Informatics Association*. Washington DC (2014).
60. Drummond C, Holte RC. C4. 5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling. In: *Workshop on Learning from Imbalanced Datasets II*. Washington DC: Citeseer (2003).
61. Weiss GM, McCarthy K, Zabar B. Cost-sensitive learning vs. sampling: which is best for handling unbalanced classes with unequal error costs? In: *IEEE International Conference on Data Mining* (2007) p. 35–41.
62. Manual N. *Netica V5. 18*. Vancouver, BC: Norsys Software Corp. (2015).
63. Monsalve-Torra A, Ruiz-Fernandez D, Marin-Alonso O, Soriano-Payá A, Camacho-Mackenzie J, Carreño-Jaimes M. Using machine learning methods for predicting in-hospital mortality in patients undergoing open repair of abdominal aortic aneurysm. *J Biomed Inf.* (2016) 62:195–201. doi: 10.1016/j.jbi.2016.07.007
64. Kim YD, Song D, Nam HS, Lee K, Yoo J, Hong G-R, et al. D-dimer for prediction of long-term outcome in cryptogenic stroke patients with patent foramen ovale. *Thromb Haemost.* (2015) 114:614–22. doi: 10.1160/TH14-12-1040
65. Adam SS, Key NS, Greenberg CS. D-dimer antigen: current concepts and future prospects. *Blood* (2009) 113:2878–87. doi: 10.1182/blood-2008-06-165845

Conflict of Interest Statement: 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.

Copyright © 2018 Park, Chang and Nam. 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.



Relating Acute Lesion Loads to Chronic Outcome in Ischemic Stroke—An Exploratory Comparison of Mismatch Patterns and Predictive Modeling

Simon Habegger^{1*}, Roland Wiest¹, Bruno J. Weder¹, Pasquale Mordasini¹, Jan Gralla¹, Levin Häni², Simon Jung^{3,4}, Mauricio Reyes⁵ and Richard McKinley¹

¹ Support Center for Advanced Neuroimaging, Institute for Diagnostic and Interventional Neuroradiology, Inselspital, University of Bern, Bern, Switzerland, ² Department of Neurosurgery, Inselspital, University of Bern, Bern, Switzerland, ³ Department of Neurology, Inselspital, University of Bern, Bern, Switzerland, ⁴ Neurovascular Imaging Research Core, Department of Neurology, University of California, Los Angeles, Los Angeles, CA, United States, ⁵ Institute for Surgical Technology and Biomechanics, University of Bern, Bern, Switzerland

OPEN ACCESS

Edited by:

Jean-Claude Baron,
University of Cambridge,
United Kingdom

Reviewed by:

Nishant K. Mishra,
Icahn School of Medicine at Mount
Sinai, United States
Muhib Khan,
Michigan State University,
United States
Vincent Thijs,
Florey Institute of Neuroscience and
Mental Health, Australia

*Correspondence:

Simon Habegger
Simon.Habegger@insel.ch

Specialty section:

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

Received: 02 March 2018

Accepted: 13 August 2018

Published: 11 September 2018

Citation:

Habegger S, Wiest R, Weder BJ, Mordasini P, Gralla J, Häni L, Jung S, Reyes M and McKinley R (2018) Relating Acute Lesion Loads to Chronic Outcome in Ischemic Stroke—An Exploratory Comparison of Mismatch Patterns and Predictive Modeling. *Front. Neurol.* 9:737. doi: 10.3389/fneur.2018.00737

Objectives: To investigate the relationship between imaging features derived from lesion loads and 3 month clinical assessments in ischemic stroke patients. To support clinically implementable predictive modeling with information from lesion-load features.

Methods: A retrospective cohort of ischemic stroke patients was studied. The dataset was dichotomized based on revascularization treatment outcome (TICI score). Three lesion delineations were derived from magnetic resonance imaging in each group: two clinically implementable (threshold based and fully automatic prediction) and 90-day follow-up as final groundtruth. Lesion load imaging features were created through overlay of the lesion delineations on a histological brain atlas, and were correlated with the clinical assessment (NIHSS). Significance of the correlations was assessed by constructing confidence intervals using bootstrap sampling.

Results: Overall, high correlations between lesion loads and clinical score were observed (up to 0.859). Delineations derived from acute imaging yielded on average somewhat lower correlations than delineations derived from 90-day follow-up imaging. Correlations suggest that both total lesion volume and corticospinal tract lesion load are associated with functional outcome, and in addition highlight other potential areas associated with poor clinical outcome, including the primary somatosensory cortex BA3a. Fully automatic prediction was comparable to ADC threshold-based delineation on the successfully treated cohort and superior to the Tmax threshold-based delineation in the unsuccessfully treated cohort.

Conclusions: The confirmation of established predictors for stroke outcome (e.g., corticospinal tract integrity and total lesion volume) gives support to the proposed methodology—relating acute lesion loads to 3 month outcome assessments by way of correlation. Furthermore, the preliminary results indicate an association of further brain regions and structures with three month NIHSS outcome assessments. Hence,

prediction models might observe an increased accuracy when incorporating regional (instead of global) lesion loads. Also, the results lend support to the clinical utilization of the automatically predicted volumes from FASTER, rather than the simpler DWI and PWI lesion delineations.

Keywords: stroke recovery, lesion load, correlation, FASTER, atlas-based regional image analysis

BACKGROUND

In 2013, 18.3 million ischemic stroke survivors were reported world-wide. The incidence of ischemic stroke in the same year was stated to be 6.9 million and the disease claimed 3.3 million lives worldwide (1). The global burden of ischemic stroke has increased with respect to incidence (37%), number of death (21%), and DALYs lost (18%) over the last two decades (2). Ischemic stroke has an enormous individual, socioenvironmental and economic impact; improvements in stroke treatment and rehabilitation may therefore be of great societal interest.

An accurate assessment of likely neurological deficits after an acute stroke is important for various reasons, including setting attainable treatment goals, correctly and accurately informing patients and relatives, planning facility discharge, and assessing impact on daily living (3). Additionally, if this assessment is available at the acute stage, it may be possible to better stratify patients who are eligible for mechanical thrombectomy. Total lesion volume has been found to be an independent 90 days predictor of neurological outcome (4), and lesion topography is related to recovery and outcome prognosis (5–8). The Alberta Stroke Program Early Computed Tomography Score (ASPECTS) was created to quantify ischemic changes in ten regions along the middle cerebral artery (9). It is linked to 3 month functional outcome as measured by mRS (10) and values > 6 were found to be predictive for functional independence at 3 months and 1 year post-stroke (11). Diffusion-weighted imaging (DWI) provides an early depiction of size and location of an ischemic lesion. DWI lesion volume is an independent predictor of Barthel Index (BI) quantified outcome (12) and the power of prediction models may be increased by incorporating it as a feature (13).

Existing models predicting clinical outcome from acute imaging have taken into account load on the corticospinal tract (14–16) and lesion location (17, 18). In order for such a model to be useful for treatment selection, the model must operate within the acute time-frame. Images must be processed with little or no human interaction, meaning that manual lesion delineation is impossible. Systems providing fast automated definitions of tissue-at-risk would, on the other hand, be feasible in the acute setting. This paper investigates lesion-load features based on three different lesion delineations to demonstrate the plausibility of automatically linking lesion loads to clinical outcome. First, we analyze the correlation between observed lesion loads and outcome, as given by a manual segmentation of 90 day follow-up imaging. Second, we analyze the correlation between predicted lesion load at the acute stage and outcome: in

addition to the standard threshold-based concepts of core and penumbra, we derive predicted lesion loads from the prediction maps of a previously proposed in-house developed software (19). We perform an exploratory analysis of the plausibility of automatically linking lesion loads to clinical outcome, using a small retrospective cohort to relate a large number of imaging features to outcome. We conjecture that (i) a significant relationship exists between lesion load features and clinical assessments, where both are assessed in the chronic phase of the disease, (ii) that (i) is still valid if the features are extracted from an automatic lesion delineation at the acute stage, and, hence, that the method is clinically applicable, and (iii) that the lesion prediction maps from a previously proposed in-house developed software (19) are superior to a simple threshold-derived lesion delineation (i.e., $T_{max} > 6$ s) for finding a relationship between lesion loads at the acute phase and clinical outcome assessments.

MATERIALS AND METHODS

Study Ethics

The study is based on data from the Bernese stroke registry, a prospectively collected database approved by the Kantonale Ethikkommission Bern, some aspects of which have been reported previously (20–23). All patients were treated for an acute ischemic stroke at the University Hospital of Bern between 2005 and 2013. The study was performed according to the ethical guidelines of the Canton of Bern with approval of our institutional review board (Kantonale Ethikkommission Bern).

Inclusion Criteria

Patients were included in this analysis if: (i) a diagnosis of ischemic stroke was established by MR imaging with an identifiable lesion on DWI and perfusion imaging, (ii) a proximal occlusion of the middle cerebral artery (M1 or M2 segment) was documented on digital subtraction angiography, (iii) endovascular therapy was attempted, either by intra-arterial thrombolysis (before 2010) or by mechanical thrombectomy (since 2010), (iv) pre-treatment MRI was performed with sufficient quality (i.e., no motion artifacts), (v) the imaging data were recorded completely into the picture archiving and communication system, (vi) the patients had a minimum age of 18 years at the time of stroke. Patients were excluded if they received only purely diagnostic angiography. Patients with a stenosis or occlusion of the carotid artery were excluded as well. Revascularization success was stratified retrospectively according to the TICI score by two examiners blinded for clinical data (24). Stroke severity for these patients was assessed at admission according to the National Institutes of Health Stroke

Abbreviations: NIHSS, National Institute of Health Stroke Scale; TICI, Thrombolysis in Cerebral Infarction Scale; ADC, Apparent Diffusion Coefficient.

Scale (NIHSS) scale. We aimed to identify all patients with a 3-month axial T2-weighted follow-up image in order to define the final extent of infarction. The inclusion/exclusion criteria did not depend on lesion location, nor was the data selected according to predetermined impairments.

Clinical Assessment

The degree of recovery was determined with standard scores (NIHSS and mRS) that are routinely available in clinical stroke registries.

Dataset Splitting

After endovascular therapy, success of the intervention can be determined via the TICI score (24). In the study at hand, patients were dichotomized according to endovascular therapy outcome into successful and unsuccessful revascularization. Successful revascularization was ascribed to patients with a TICI score 2b-3 whereas the unsuccessful revascularization was assigned to TICI scores 0-2a.

Pipeline

The processing pipeline used in this paper is depicted in **Figure 1**. The individual steps are briefly discussed in the following sections.

Image Acquisition

Imaging data were acquired on either a 1.5T (Siemens Magnetom Avanto) or 3T MRI system (Siemens Magnetom Verio). Patients received whole brain DWI, (24 slices, thickness 5 mm, repetition time 3,200 ms, echo time 87 ms, number of averages 3, matrix 256×256 , flip angle 90) yielding images for b values of 0 s/mm² and 1,000 s/mm² as well as ADC maps that were calculated automatically. Standard dynamic susceptibility contrast-enhanced perfusion MRI (gradient-echo echoplanar imaging sequence, repetition time 1,410 ms, echo time 30 ms, field of view 230×230 mm, voxel size: $1.8 \times 1.8 \times 5.0$ mm, slice thickness 5.0 mm, 19 slices, 80 acquisitions, flip angle 90) was acquired. Images were acquired during the first pass of a standard bolus of 0.1 mmol/kg gadobutrol (Gadovist, Bayer Healthcare). Contrast medium was injected at a rate of 5 ml/s followed by a 20 ml bolus of saline at a rate of 5 ml/s. In addition, an axial T2-weighted turbo-spin echo sequence (TR 3760–4100 ms, TE 85–100 ms, flip angle 150) and contrast-enhanced T1-weighted sequence [1.5T system: spin-echo sequence (TR 663 ms, TE 17 ms, flip angle 90), 3T system: gradient-echo sequence (TR 250 ms, TE 2.67 ms, flip angle 70)], a time-of-flight angiography and a first pass Gd-MRA were acquired, with T2-weighted imaging and TOF angiography performed before contrast injection.

Pre-processing

The main pre-processing step in this work was image normalization to warp the utilized images into the MNI152 space with a $2 \times 2 \times 2$ mm resolution. This was done in Matlab (MATLAB R2014a, The MathWorks, Inc., Natick, Massachusetts, United States) with SPM12 (Wellcome Trust Centre for Neuroimaging, University College London).

Lesion Delineation

We compared four lesion maps: expert manual segmentation of 90 day T2 MRI, threshold-based manual segmentation of acute ADC, automated threshold-based segmentation of Tmax imaging, and machine-learning-based prediction of chronic lesion load from acute imaging (using our in-house software tool, FASTER).

FASTER

FASTER (19) is a recently proposed stroke lesion estimation method. Given acute stroke imaging data, FASTER produces two predicted lesion maps, representing successful and unsuccessful revascularization. The required inputs to FASTER consist of diffusion-weighted, T2 and T1w (contrast enhanced) sequences and dynamic susceptibility contrast perfusion. FASTER provides a threshold-independent estimation by using two machine-learning models trained on cases with a TICI of 3 and 0, respectively. The fully automated software makes stroke outcome prediction feasible in any clinical setting, provided the necessary imaging data is accessible. Using FASTER, a prediction of tissue damage in the case of successful revascularization was calculated for the patients with TICI 0-2a, and a prediction of tissue damage in the case of unsuccessful revascularization was calculated for the patients with TICI 2b-3.

Follow-Up Segmentation

Manual segmentation was performed on the T2-weighted 3 month follow-up images. Manual regions of interest were drawn to the maximal extent of the final infarction, including areas with hemorrhagic transformation, but excluding regions already hyperintense on acute T2 imaging. The boundaries of the infarctions were manually delineated for every single transversal slice. The 3 month follow-up lesion was chosen as the definition of final infarction, rather than the lesion in the early acute phase of lesion evolution, since apparent lesion size in the early acute phase is known to overestimate final lesion volume (25). T2 was chosen as the modality for identifying the final lesion, since it was more widely available than a FLAIR follow-up image in the retrospective data used.

Core and Tmax > 6 s Delineations

The infarct core was manually segmented based on a threshold of $ADC < 600 \times 10^{-6} \text{ mm}^2/\text{s}$ (26). The perfusion deficit was computed as a pre-processing step of FASTER, using a threshold of $T_{max} > 6 \text{ s}$ (27).

Lesion Load Features

Previous studies that investigated the relationship between imaging biomarkers and clinical outcome assessments have revealed the importance of white-matter structures (15, 16, 28, 29). Therefore, it was important to choose an atlas for feature definition that identifies white-matter structures, especially, left and right corticospinal tract. To this end the Juelich histological atlas (30–32) from FSL 5.0 was selected, which encompasses 29 white-matter and 92 gray-matter regions (**Figure 2**).

Following normalization, the lesion segmentations and predictions were overlaid onto the Juelich structural atlas.

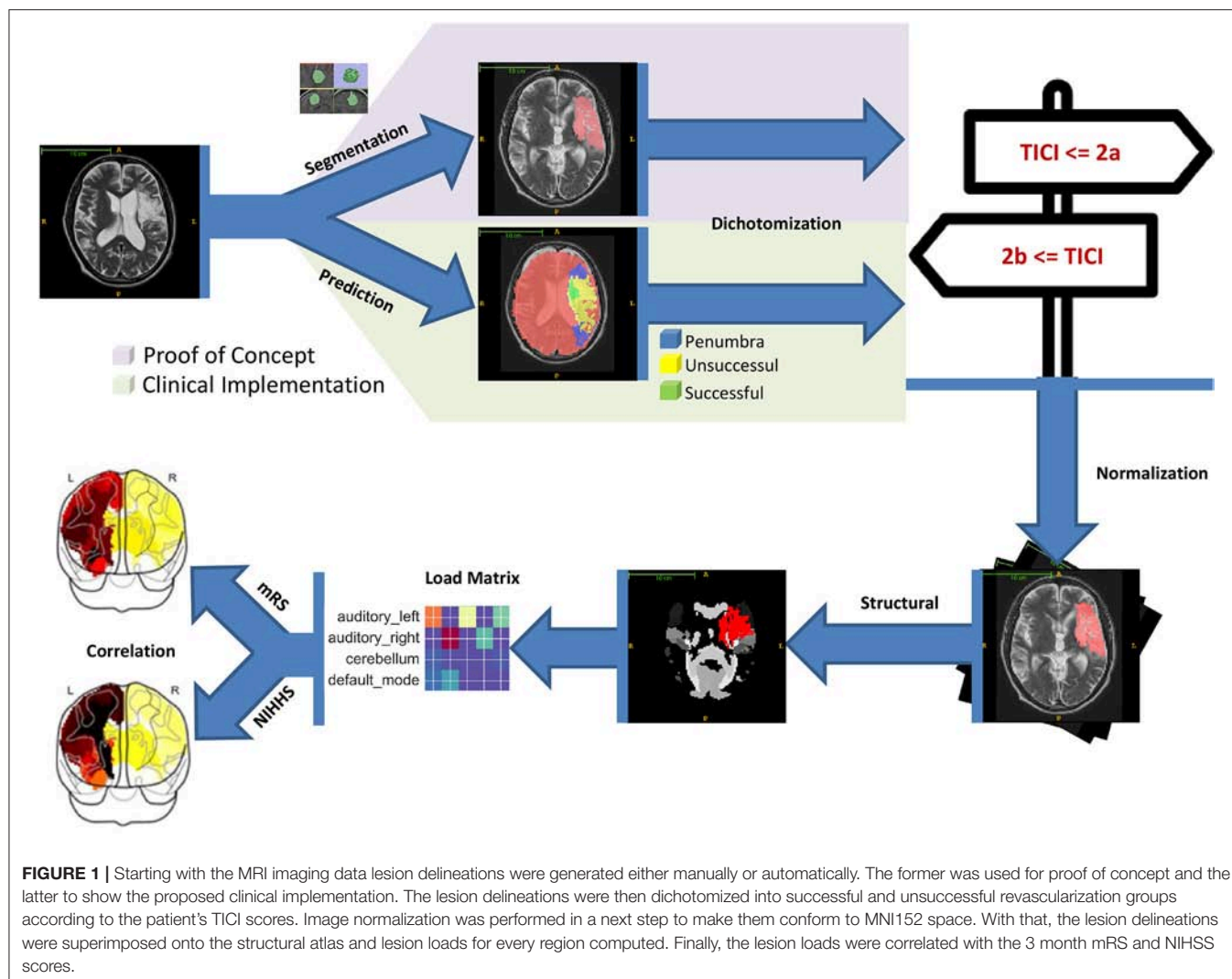


FIGURE 2 | The images depict axial, coronal and sagittal slices of a normalized brain with the overlapped Juelich histological atlas. Both gray and white matter structures can be seen.

After that, a load percentage was computed for every structural region of interest. The lesion load denotes the percentage of the region that was affected by the lesion.

Correlation Analysis

Lesion loads were created on a per-patient level for every atlas defined region based on the lesion segmentation—i.e., 121 lesion

loads per patient. The loads designate the percentage of the respective region affected by the lesion.

The correlation between the 3 month clinical assessments and the lesion load of every atlas-defined region was assessed.

The presented work rests on the premise that a linear relationship between continuous lesion loads and continuous 3 month clinical assessment scores exists. As a result, the Pearson correlation coefficient was selected to assess the relationship.

Correlations were also calculated between total volume and outcome.

In a first step, we correlated lesion loads with both 3 month NIHSS and mRS scores. The whole dataset was considered (i.e., no grouping into successful and unsuccessful) and the lesion loads were based on the 3 month follow-up segmentation. The subsequent analysis was then carried out with only the superior clinical assessment that emerged from this analysis. In a second step, we evaluated the various lesion delineations by correlating the respective lesion loads with the superior outcome measure from the first step (i.e., 3 month NIHSS or mRS). For this part of the analysis the dataset was split according to dichotomized revascularization success. In this step, we calculated correlations with the lesion loads as generated by manually segmented 90-day follow-up imaging and as generated both by thresholds and by FASTER on acute imaging in order to validate our hypotheses.

A result of having as many as 121 brain regions delineated by the atlas is that the lesion load data is relatively sparse. Our data thus fails to satisfy the normality assumptions of parametric statistical tests, leading to most of the correlations we observed being significant. To better observe the true significance of our findings, we used the non-parametric statistical method of bootstrap sampling (33) to construct confidence intervals (CI) for the obtained correlation values. One thousand samples were drawn from the original data (for every ROI, respectively) and individually correlated with the clinical outcome scores. The obtained correlations were then used to create a 95% confidence interval. If the confidence interval included zero correlation the original correlation was termed statistically insignificant and significant otherwise. Since we consider our study an exploratory one, examining the feasibility of linking lesion loads to NIHSS (rather than a study to identify which regions are linked to NIHSS), we do not perform any correction for multiple comparisons. As a result, this study should not be seen as identifying any individual stroke damage locations as “significantly” related to NIHSS, but rather as providing a group of regions which may, in a subsequent study of a larger group of patients, be viewed as good candidate regions on which to focus.

RESULTS

We analyzed 55 patients in total: the successful revascularization cohort contained 35 patients, whereas the unsuccessful revascularization cohort consisted of 18 patients. The successfully revascularized cohort entails 13 female and 23 male patients. The mean age of the group is 61 ± 12 years; minimum and maximum ages are 35 and 81, respectively. In the unsuccessfully revascularized cohort the mean age is 59 ± 14 years; minimum

and maximum ages are 18 and 76, respectively. The group is composed of 7 female and 12 male patients.

Lesion Distribution

Figure 3 depicts normalized lesion distributions grouped by underlying lesion delineation (rows) and revascularization outcome (columns).

Correlations Analysis

This section presents the results of the correlation between lesion loads and 3-month clinical assessments.

Three Month NIHSS vs. mRS on Complete Dataset

The results are depicted in **Figure 4** both visually and in tabular form (only top ten regions).

A Wilcoxon signed-rank test comparing NIHSS and mRS correlations revealed a statistically significant difference between the two samples (p -value < 0.05).

Comparison of Lesion Delineations

Since NIHSS was found to yield significantly higher correlations than mRS, we used NIHSS as a measure of clinical outcome in the remainder of our experiments.

A correlation analysis was performed with respect to the 3 month NIHSS on the split dataset. The correlations are shown as overlay to a glass brain and in tabular form in **Figure 5**. The first row depicts the normalized lesion distributions for successful and unsuccessful revascularization based on the follow-up segmentation, whereas the subsequent rows with the respective tables present correlations. Lesion distributions (already presented in **Figure 3**) are shown only as a guide for interpretation. Correlations that were found significant according to the bootstrap CI are marked with an asterisk in the designated column of the table.

DISCUSSION

We proposed a methodology suitable for investigating a hypothesized relationship between affected functional brain regions and 3 month clinical outcome in ischemic stroke patients. Other studies have examined such a relationship on a voxel based level—i.e., voxel-based lesion symptom mapping (VLSM) for multiple sclerosis lesions (34). Another study found, through VLSM, a specific motor pathway influence on mRS outcome and a reflection of lateralized functions such as neglect and aphasia (35). Although the study included 101 patients it was nevertheless designated to be exploratory; this is a result of two limitations associated with VLSM: generally a large number of voxels must be considered and there is functional cross-dependence between the voxels. These two limitations can be alleviated with a regional approach, where voxels are grouped into functionally meaningful regions.

We analyzed the lesion distributions of 55 stroke patients, of which 35 were successfully revascularized (TICI 2b-3) and 18 were unsuccessfully revascularized (TICI 0-2a). As expected, the outcome prediction maps from FASTER show strong dissimilarities between successfully revascularized patients

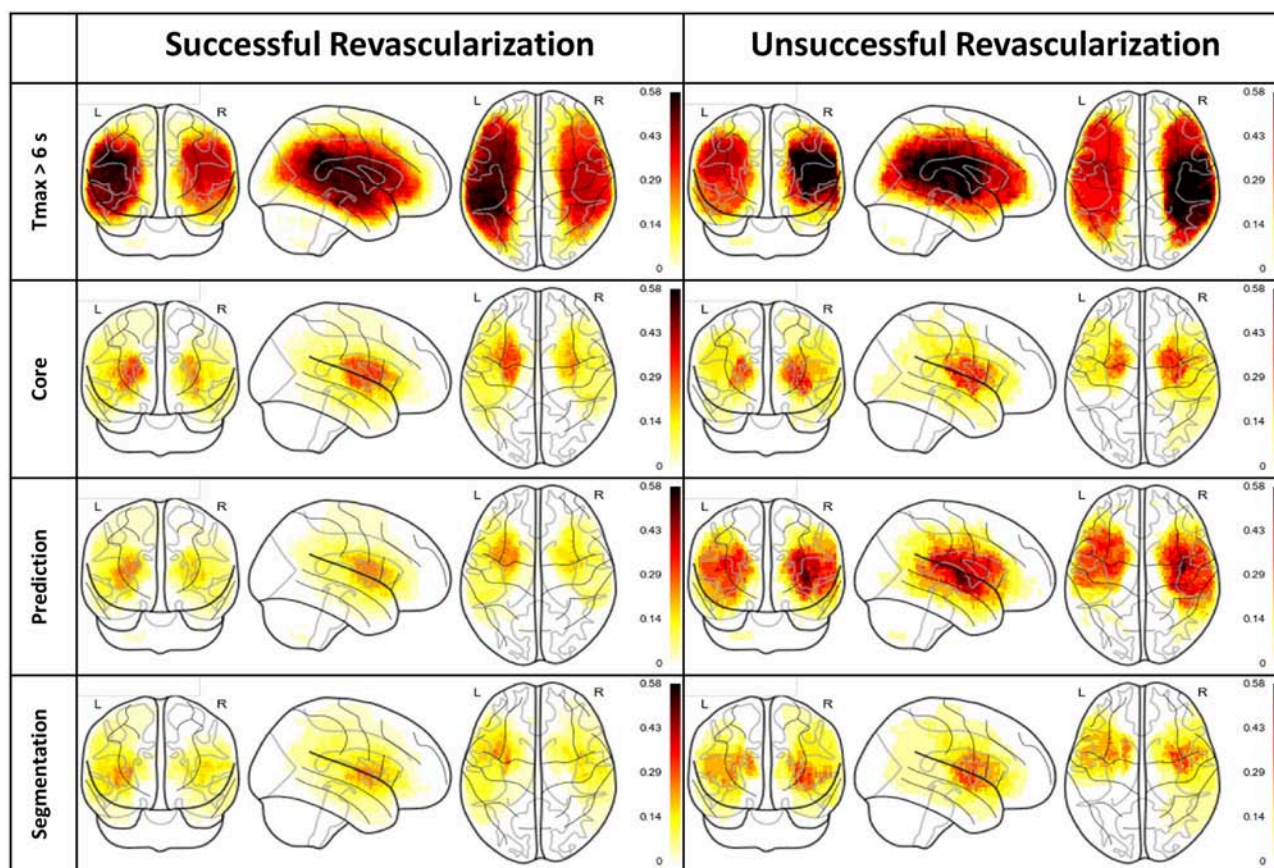


FIGURE 3 | Lesion distributions horizontally and vertically grouped by revascularization outcome and type of lesion delineation, respectively. The distributions are normalized so that the values are confined to the range 0 (i.e., not affected by any lesion in the cohort) and 1 (i.e., affected by all lesions in the cohort).

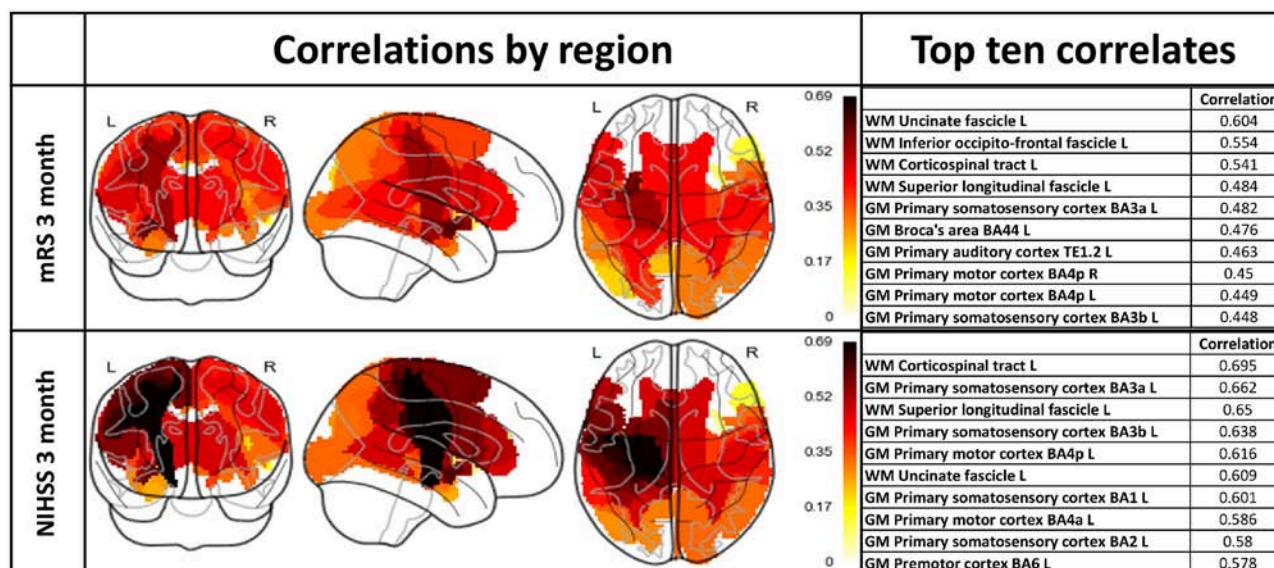
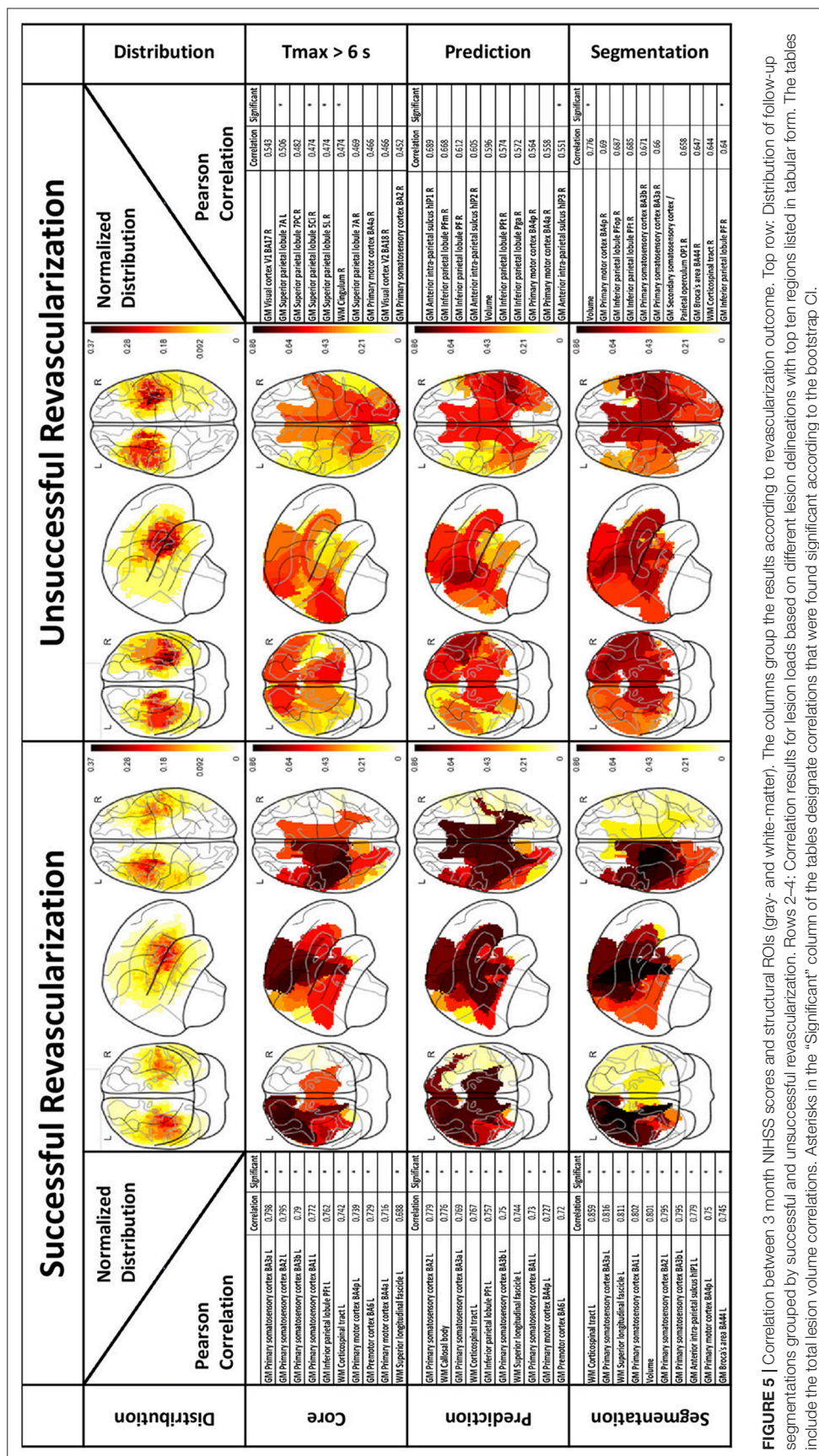


FIGURE 4 | Lesion load correlations with 3 month mRS and NIHSS scores including the whole patient cohort. Left column: Visualization of regional lesion load correlations with clinical assessments. Right column: Top ten correlating regions with respect to outcome scores.



(where the prediction is based on a model trained on patients with TICI 3) and unsuccessfully revascularized patients (where the prediction is based on a model trained on patients with TICI 0). The segmentation distribution displays the same disparity between treatment outcome cohorts: this effect is less pronounced, which can perhaps be explained by the effect of collateral circulation preserving tissue which might otherwise have been lost, or the effects of partial revascularization in patients with a TICI greater than zero.

Our results show substantial differences in the revascularized and unrevascularized cohorts, with respect to correlations between lesion load and functional outcome, which is to be expected. In the case of successful revascularization, the tissue damage is mostly limited to tissue lying in the ischemic core, which is well-identified by diffusion-weighted imaging. This can be observed in the similarities between the loads observed in the ischemic core, FASTER prediction and final outcome segmentation. In this cohort, regions in the sensorimotor system were strongly correlated with outcome, as were the three white-matter tracts (superior longitudinal fascicle, corticospinal tract, and callosal body) as well as the total lesion volume. For the unrevascularized cohort, the situation is less well-defined. First, when considering the relationship between follow-up segmentation and NIHSS, only two of the top-ten correlated regions were significantly correlated according to the bootstrap confidence interval. The results arising from the Tmax and FASTER segmentations should be viewed in this light, since we should not expect assessments at the acute phase to be more accurate than the final lesion volume. Nevertheless, we can make a qualitative assessment of the most correlated regions: these differ not only in values but also in appearing regions. Correlations between prediction-based lesion loads and NIHSS were in general higher than between Tmax-based lesion-loads and NIHSS, but more correlations were found to be significant using Tmax than using prediction. We hypothesize that the former effect would persist in larger cohorts, and that the second would not, but this will require analysis of larger cohorts of unsuccessfully revascularized patients.

A major finding from our study is the difference in relationships between atlas based lesion load mapping and the different outcome measures (mRS and NIHSS). These scores have been developed to ensure simplicity, reliability and validity in its clinical application (36), and were tested for consistency (37) and reliability (38). The modified Rankin scale has been found to relate more closely to quality of life compared to the NIHSS (39). If, however, the focus is on specific neurological, rather than global disabilities, the NIHSS should be used (40). Our analysis revealed a statistically significant difference between the obtained correlation values, with correlations between lesion loads and NIHSS being higher than with mRS. A similarity in recognizable contours can be perceived, e.g., corticospinal tract and uncinate fascicle. However, different regions dominate and the heatmap as well as the tabular listing reveals that, in general, higher correlations were achieved with the NIHSS score. This might be an important finding for future investigations as recent studies focused on the mRS (35, 41). Moreover, white-matter structures dominate the top correlates with mRS. The NIHSS correlations

also exhibit the importance of white-matter structures (50% white-matter regions in top six); the highest correlation was achieved by the left corticospinal tract. Besides the white-matter structures, both correlation lists exhibit the importance of sensorimotor structures. The importance of motor regions as well as white-matter structures with respect to outcome has been shown (35, 42). The observed correlation values together with the prominence of established brain areas are a strong support for the methodology of correlating lesion loads with NIHSS outcome assessment. The predominant white-matter tracts that appeared in our analysis were the corticospinal tract, cingulum, callosal body and superior longitudinal fascicle. The role of the corticospinal tract has been investigated and its importance with respect to stroke recovery outcome is established (14, 16, 43–47). This is consistent with the role of the corticospinal tract as the main pathway for afferent and efferent signal conduction between brain and limbs. The effect of an impact on the other white-matter structures is less clearly understood and can range from mental disorder (anterior cingulum) to working memory impairment (posterior cingulum) and loss of language integration ability (superior longitudinal fasciculus). Also, the dorsal Anterior Cingulate Cortex (dACC) was found to send task specific modulatory signals to the Supplementary Motor Area (48), which is an indication of its involvement in motor control.

Total acute lesion volume as measured through perfusion or DWI imaging has been found to be positively correlated with clinical assessments (12, 49, 50). Our results support these findings, but suggest that total lesion volume may not be the most important feature in predicting functional outcome: total lesion volume was surpassed by at least four gray-matter regions and/or white-matter structures in five out of the six analyzed situations (only in case of the unsuccessful segmentation result was the volume the top correlate). Total lesion volume for the various delineations ranked 1st (0.776, significant), 5th (0.801, significant), and 5th (0.596, insignificant) for segmentation unsuccessful, segmentation successful and prediction unsuccessful, respectively. This suggests that while lesion volume is informative and carries predictive weight, it accounts only partially for the observed recovery. The importance of lesion location in predicting stroke outcome is supported in the literature, specifically in studies where the focus is on particular structures and brain regions (e.g., corticospinal tract) (14–17, 28, 43, 44, 46, 47). A study including loads on 132 cortical and subcortical regions constructed a decision tree associating regional lesion loads with outcome; the total lesion volume played a role in the left hemisphere, according to the study (41). Like our study, the number of patients precludes identifying loads in any one region as being significantly associated with outcome.

Our results confirm the significant and intertwined role of stroke lesion volume and location, and suggest that a number of other gray-matter regions and white-matter structures carry predictive information and may have the potential to increase the predictive accuracy.

A lateralization of the correlation values was evident in the results, with good correlations between left-hemispheric regions

for successful thrombectomy and right-hemispheric regions for unsuccessful thrombectomy.

Studies have shown that the importance of regions with respect to outcome may depend on hemispheric idiosyncrasies: limbic, default mode and language areas in the left hemisphere, and visuospatial and motor regions in the right hemisphere (41). Areas of lateralized brain functions together with motor areas were found to influence functional outcome (35). However, we do not consider that our study has the statistical power to conclude that the observed effect is due to any biological reason, and assume that the lateralization is a result of the small patient cohort and, therefore, due to randomness in the lesion distribution with respect to laterality. This conclusion should be investigated and confirmed on a larger dataset. Similarly, the difference in statistical significance between unsuccessful and successful revascularized patient cohorts is likely a reflection of the sample sizes of the two cohorts, which was 18 and 35 for unsuccessful and successful groups, respectively. It is important to keep in mind the meaning of a 95% confidence interval derived from bootstrapping: by randomly sampling repeatedly from our available data, we simulate a large number of experiments. The confidence interval for the correlation coefficient is defined as the 2.5th and 97.5th percentiles of the derived correlations. Since the lesion loads were sparse, for many of these trials no patients with non-zero lesion load were randomly selected, and this accounts, in the main, for the large number of non-significant correlations. The results that the FASTER lesion predictions are superior to simple threshold-based delineations (in particular in case of unsuccessful revascularization) and that the predictive value of total lesion volume is surpassed by lesion loads must be confirmed in future investigations, ideally with substantially larger patient cohorts. We note that there may be brain regions not identified by this study which are nonetheless important for predicting stroke outcome. Regions which due to cohort size were not loaded in our dataset cannot be considered in our analysis: this again necessitates an increased cohort size. Additionally, lesion load on additional regions not represented in the Juelich atlas (e.g., the basal ganglia) may also influence clinical outcome. As a next step, atlases that encompass the deep brain nuclei, the inferior temporal lobe and frontal lobe areas should be included into the analysis.

The study investigated the relationship between lesion loads and clinical outcome assessment with the Pearson correlation coefficient. While we think the employed measure is applicable to the data, various alternative measures might be worth considering in future investigations.

Finally, correlations for each region were considered individually, while the effects of loads on networks of regions may not be adequately explained by loads on the constituent parts of those networks. Furthermore, adjacent regions will tend to be loaded together, and it is possible that loads on a given region may be predictive of clinical outcome, simply by their proximity to adjacent eloquent areas. Detailed NIHSS scores of the 15 functional elements and supplementary quality of life scales, providing functionality of subjects in daily life, were not available in the cohort under investigation.

Outlook

Our analysis lays a solid foundation for exploring the relationship between neurological assessments (mRS and NIHSS) and lesion site and extent. This knowledge can be used to focus on and extract from imaging a selection of important features to make reliable predictions on neurological scores. The keywords here, we believe, are “focus” and “selection” as only information that is focused and manageable allows for a model that can be built from a feasible amount of patient data. Further, it brings the advantage of permitting a model with enough simplicity that the prediction process can be understood and further insights gained.

CONCLUSION

This study investigated the relationship between imaging-based lesion loads and 3 month clinical assessments. We analyzed various lesion delineations used for the computation of the lesion loads. Regions known to be associated with stroke outcome were confirmed and new potentially informative areas suggested. The results support the clinical utilization of the automatically predicted volumes from FASTER over the simpler DWI and PWI lesion delineations.

AUTHOR CONTRIBUTIONS

SH: Study design, data preparation and analysis, statistical analysis and writing of the manuscript; RW: Study design, concept, data interpretation, and writing of the manuscript. LH: Data preparation; JG: Manuscript discussion and review; PM: Manuscript discussion and review, BW: Data interpretation, manuscript discussion and review, SJ: Data preparation, manuscript discussion and review, MR: Study design, concept, data interpretation and writing of the manuscript, RM: Study design, concept, data interpretation, statistical analysis, and writing of the manuscript.

FUNDING

This work was supported by following grants: Swiss National Foundation (SNF) grant 320030L_170060 Stroke treatment goes personalized: Gaining added diagnostic yield by computer-assisted treatment selection (the STRAY-CATS project, JG), SNF grant 32003B_160107 Effects of serotonergic neuromodulation on behavioral recovery and motor network plasticity after cortical ischemic stroke: a longitudinal, placebo-controlled study (RW) and Swiss heart foundation grants 2016 and 2017 related to the projects above.

ACKNOWLEDGMENTS

We are grateful to all the stroke patients who gave consent to utilize their data. We thank all the staff of the departments of neuroradiology and neurology at the Inselspital.

REFERENCES

- Feigin VL, Krishnamurthi RV, Parmar P, Norrving B, Mensah GA, Bennett DA. Update on the global burden of ischemic and hemorrhagic stroke in 1990–2013: the GBD 2013 study. *Neuroepidemiology* (2015) 45:161–176. doi: 10.1159/000441085
- Krishnamurthi RV, Feigin VL, Forouzanfar MH, Mensah GA, Connor M, Bennett DA, et al. Articles global and regional burden of first-ever ischaemic and haemorrhagic stroke during 1990–2010: findings from the global burden of disease study 2010 *Lancet Glob Health* (2013) 1:e259–81. doi: 10.1016/S2214-109X(13)70089-5
- Veerbeek JM, Kwakkel G, van Wegen EE, Ket JC, Heymans MW. Early prediction of outcome of activities of daily living after stroke. *Stroke* (2011) 42:1482–8. doi: 10.1161/STROKEAHA.110.604090
- Vogt G, Laage R, Shuaib A, Schneider A, VISTA Collaboration. Initial lesion volume is an independent predictor of clinical stroke outcome at day 90: an analysis of the virtual international stroke trials archive (VISTA) database. *Stroke* (2012) 43:1266–72. doi: 10.1161/STROKEAHA.111.646570
- Borsody M, Warner Gargano J, Reeves M, Jacobs B, MASCOTS Insula-Stroke Substudy Group. Infarction involving the insula and risk of mortality after stroke. *Cerebrovasc Dis.* (2009) 27:564–71. doi: 10.1159/000214220
- Ois A, Cuadrado-Godia E, Solano A, Perich-Alsina X, Roquer J. Acute ischemic stroke in anterior choroidal artery territory. *J Neurol Sci.* (2009) 281:80–4. doi: 10.1016/j.jns.2009.02.323
- Schiemanck S, Kwakkel G, Post M, Kappelle L, Prevo A. Impact of internal capsule lesions on outcome of motor hand function at one year post-stroke. *J Rehabil Med.* (2008) 40:96–101. doi: 10.2340/16501977-0130
- Bathey TW, Karki M, Singhal AB, Wu O, Sadaghiani S, Campbell BC, et al. Brain edema predicts outcome after nonlacunar ischemic stroke. *Stroke* (2014) 45:3643–8. doi: 10.1161/STROKEAHA.114.006884
- Barber PA, Demchuk AM, Zhang J, Buchan AM. Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy. *The Lancet* (2000). 355:1670–74. doi: 10.1016/S0140-6736(00)02237-6
- Heiss WD. Contribution of neuro-imaging for prediction of functional recovery after ischemic stroke. *Cerebrovasc Dis.* (2017) 44:266–76. doi: 10.1159/000479594
- Alexander LD, Pettersen JA, Hoppyan JJ, Sahlas DJ, Black SE. Long-term prediction of functional outcome after stroke using the alberta stroke program early computed tomography score in the subacute stage. *J Stroke Cerebrovasc Dis.* (2012) 21:737–44. doi: 10.1016/j.jstrokecerebrovasdis.2011.03.010
- Thijs VN, Lansberg MG, Beaulieu C, Marks MP, Moseley ME, Albers GW. Is early ischemic lesion volume on diffusion-weighted imaging an independent predictor of stroke outcome? a multivariable analysis. *Stroke* (2000) 31:2597–602. doi: 10.1161/01.STR.31.11.2597
- Hand PJ, Wardlaw JM, Rivers CS, Armitage PA, Bastin ME, Lindley RI, et al. MR diffusion-weighted imaging and outcome prediction after ischemic stroke. *Neurology* (2006) 66:1159–63. doi: 10.1212/01.wnl.0000202524.43850.81
- Feng W, Wang J, Chhatbar PY, Doughty C, Landsittel D, Lioutas VA, et al. Corticospinal tract lesion load: an imaging biomarker for stroke motor outcomes. *Ann Neurol.* (2015). 78:860–70. doi: 10.1002/ana.24510
- Zhu LL, Lindenberg R, Alexander MP, Schlaug G. Lesion load of the corticospinal tract predicts motor impairment in chronic stroke. *Stroke* (2010) 41:910–15. doi: 10.1161/STROKEAHA.109.577023
- Lindenberg R, Renga V, Zhu LL, Betzler F, Alsop D, Schlaug G. Structural integrity of corticospinal motor fibers predicts motor impairment in chronic stroke. *Neurology* (2010) 74:280–87. doi: 10.1212/WNL.0b013e3181ccc6d9
- Munsch F, Sagnier S, Asselineau J, Bigourdan A, Guttmann CR, Debruxelles S, et al. Stroke location is an independent predictor of cognitive outcome. *Stroke* (2016) 47:66–73. doi: 10.1161/STROKEAHA.115.011242
- Jones PS, Pomeroy VM, Wang J, Schlaug G, Tulasi Marrapu S, Geva S, et al. Does stroke location predict walk speed response to gait rehabilitation? *Hum Brain Mapp.* (2015) 37:689–703. doi: 10.1002/hbm.23059
- McKinley R, Häni L, Gralla J, El-Koussy M, Bauer S, Arnold M, et al. Fully automated stroke tissue estimation using random forest classifiers (FASTER). *J Cereb Blood Flow Metab.* (2016) 37:2728–41. doi: 10.1177/0271678X16674221
- Arnold M, Kappeler L, Nedeltchev K, Brekenfeld C, Fischer U, Kaserue B, et al. Recanalization and outcome after intra-arterial thrombolysis in middle cerebral artery and internal carotid artery occlusion. *Stroke* (2007) 38:1281–5. doi: 10.1161/01.STR.0000259711.13490.23
- Jung S, Mono ML, Fischer U, Galimanis A, Findling O, De Marchis GM, et al. Three-month and long-term outcomes and their predictors in acute basilar artery occlusion treated with intra-arterial thrombolysis. *Stroke* (2011) 42:1946–51. doi: 10.1161/STROKEAHA.110.606038
- Jung S, Schindler K, Findling O, Mono ML, Fischer U, Gralla J, et al. Adverse effect of early epileptic seizures in patients receiving endovascular therapy for acute stroke. *Stroke* (2012) 43:1584–1590. doi: 10.1161/STROKEAHA.111.645358
- Galimanis A, Jung S, Mono ML, Fischer U, Findling O, Weck A, et al. Endovascular therapy of 623 patients with anterior circulation stroke. *Stroke* (2012) 43:1052–7. doi: 10.1161/STROKEAHA.111.639112
- Higashida RT, Furlan AJ, Roberts H, Tomsick T, Connors B, Barr J, et al. Trial design and reporting standards for intra-arterial cerebral thrombolysis for acute ischemic stroke. *Stroke* (2003) 34:e109–37. doi: 10.1161/01.STR.0000082721.62796.09
- Gaudinski MR, Henning EC, Miracle A, Luby M, Warach S, Latour LL. Establishing final infarct volume: stroke lesion evolution past 30 days is insignificant. *Stroke* (2008) 39:2765–8. doi: 10.1161/STROKEAHA.107.512269
- Straka M, Albers GW, Bammer R. Real-time diffusion-perfusion mismatch analysis in acute stroke. *J Magn Reson Imaging* (2010) 32:1024–37. doi: 10.1002/jmri.22338
- Olivet JM, Mlynash M, Thijs VN, Purushotham A, Kemp S, Lansberg MG, et al. Geography, structure, and evolution of diffusion and perfusion lesions in Diffusion and perfusion imaging Evaluation For Understanding Stroke Evolution (DEFUSE). *Stroke* (2009) 40:3245–51. doi: 10.1161/STROKEAHA.109.558635
- DeVetten G, Coutts SB, Hill MD, Goyal M, Eesa M, O'Brien B, et al. Acute corticospinal tract wallerian degeneration is associated with stroke outcome. *Stroke* (2010) 41:751–6. doi: 10.1161/STROKEAHA.109.573287
- Jang SH. Prediction of motor outcome for hemiparetic stroke patients using diffusion tensor imaging: a review. *Neurorehabilitation* (2010) 27:367–72. doi: 10.3233/NRE-2010-0621
- Eickhoff SB, Paus T, Caspers S, Grosbras MH, Evans AC, Zilles K, et al. Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. *Neuroimage* (2007) 36:511–21. doi: 10.1016/j.neuroimage.2007.03.060
- Eickhoff SB, Heim S, Zilles K, Amunts K. Testing anatomically specified hypotheses in functional imaging using cytoarchitectonic maps. *Neuroimage* (2006) 32:570–82. doi: 10.1016/j.neuroimage.2006.04.204
- Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, et al. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* (2005) 25:1325–35. doi: 10.1016/j.neuroimage.2004.12.034
- Carpenter J, Bithell J. Bootstrap confidence intervals : when, which, what ? a practical guide for medical statisticians. *Stat Med.* (2000) 19:1141–64. doi: 10.1002/(SICI)1097-0258(20000515)19:9<1141::AID-SIM479>3.0.CO;2-F
- Altermatt A, Gaetano L, Magon S, Häring DA, Tomic D, Wuerfel J, et al. Clinical correlations of brain lesion location in multiple sclerosis: voxel-based analysis of a large clinical trial dataset. *Brain Topogr.* (2018) 31:886–94. doi: 10.1007/s10548-018-0652-9
- Cheng B, Forkert ND, Zavaglia M, Hilgetag CC, Golsari A, Siemonsen S, et al. Influence of stroke infarct location on functional outcome measured by the modified rankin scale. *Stroke* (2014) 45:1695–702. doi: 10.1161/STROKEAHA.114.005152
- Brott T, Adams HP, Olinger CP, Marler JR, Barsan WG, Biller J, et al. Measurements of acute cerebral infarction: a clinical examination scale. *Stroke* (1989) 20:864–70.
- Goldstein LB, Bertels C, Davis JN. Interrater reliability of the NIH stroke scale. *Arch Neurol.* (1989) 46:660–2. doi: 10.1001/archneur.1989.00520420080026
- Josephson SA, Hills NK, Johnston SC. NIH Stroke scale reliability in ratings from a large sample of clinicians. *Cerebrovasc Dis.* (2006) 22:389–95. doi: 10.1159/000094857

39. Ali M, Fulton R, Quinn T, Brady M, VISTA Collaboration. How well do standard stroke outcome measures reflect quality of life?: a retrospective analysis of clinical trial data. *Stroke* (2013) 44:3161–5. doi: 10.1161/STROKEAHA.113.001126
40. Harrison JK, McArthur KS, Quinn TJ. Assessment scales in stroke: clinimetric and clinical considerations. *Clin Interv Aging* (2013) 8:201–11. doi: 10.2147/CIA.S32405
41. Yassi N, Churilov L, Campbell BC, Sharma G, Bammer R, Desmond PM, et al. The association between lesion location and functional outcome after ischemic stroke. *Int J Stroke* (2015) 10:1270–6. doi: 10.1111/ijs.12537
42. Corbetta M, Ramsey L, Callejas A, Baldassarre A, Hacker CD, Siegel JS, et al. Common behavioral clusters and subcortical anatomy in stroke. *Neuron* (2015) 85:927–41. doi: 10.1016/j.neuron.2015.02.027
43. Schaechter JD, Fricker ZP, Perdue KL, Helmer KG, Vangel MG, Greve DN, et al. Microstructural status of ipsilesional and contralesional corticospinal tract correlates with motor skill in chronic stroke patients. *Hum Brain Mapp*. (2009) 30:3461–74. doi: 10.1002/hbm.20770
44. Stinear CM, Barber PA, Smale PR, Coxon JP, Fleming MK, Byblow WD. Functional potential in chronic stroke patients depends on corticospinal tract integrity. *Brain* (2006) 130:170–80. doi: 10.1093/brain/awl333
45. Stinear C. Prediction of recovery of motor function after stroke. *Lancet Neurol*. (2010) 9:1228–32. doi: 10.1016/S1474-4422(10)70247-7
46. Pineiro R, Pendlebury ST, Smith S, Flitney D, Blamire AM, Styles P, et al. Relating MRI changes to motor deficit after ischemic stroke by segmentation of functional motor pathways. *Stroke* (2000) 31:672–9. doi: 10.1161/01.STR.31.3.672
47. Binkofski F, Buccino G, Posse S, Seitz RJ, Rizzolatti G, Freund H. A fronto-parietal circuit for object manipulation in man: evidence from an fMRI-study. *Eur J Neurosci*. (1999) 11:3276–86. doi: 10.1046/j.1460-9568.1999.00753.x
48. Asemi A, Ramaseshan K, Burgess A, Diwadkar VA, Bressler SL. Dorsal anterior cingulate cortex modulates supplementary motor area in coordinated unimanual motor behavior. *Front Hum Neurosci*. (2015) 9:309. doi: 10.3389/fnhum.2015.00309
49. Baird AE, Lövblad KO, Dashe JF, Connor A, Burzynski C, Schlaug G, et al. Clinical correlations of diffusion and perfusion lesion volumes in acute ischemic stroke. *Cerebrovasc Dis*. (2000) 10:441–8. doi: 10.1159/000016105
50. Tong DC, Yenari MA, Albers GW, O'Brien M, Marks MP, Moseley ME. Correlation of perfusion- and diffusion-weighted MRI with NIHSS score in acute. *Neurology* (1998) 50:864–70.

Conflict of Interest Statement: 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.

Copyright © 2018 Habegger, Wiest, Weder, Mordasini, Gralla, Häni, Jung, Reyes and McKinley. 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.



ISLES 2016 and 2017-Benchmarking Ischemic Stroke Lesion Outcome Prediction Based on Multispectral MRI

OPEN ACCESS

Edited by:

Fabien Scalzo,
University of California, Los Angeles,
United States

Reviewed by:

Jens Fiehler,
Universitätsklinikum
Hamburg-Eppendorf, Germany
Ivana Galinovic,
Centrum für Schlaganfallforschung
Berlin (CSB), Germany

*Correspondence:

Stefan Winzeck
sw742@cam.ac.uk
Mauricio Reyes
mauricio.reyes@istb.unibe.ch

Specialty section:

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

Received: 04 May 2018

Accepted: 27 July 2018

Published: 13 September 2018

Citation:

Winzeck S, Hakim A, McKinley R, Pinto JAADS, Alves V, Silva C, Pisov M, Krivov E, Belyaev M, Monteiro M, Oliveira A, Choi Y, Paik MC, Kwon Y, Lee H, Kim B, Won JH, Islam M, Ren H, Robben D, Suetens P, Gong E, Niu Y, Xu J, Pauly JM, Lucas C, Heinrich MP, Rivera LC, Castillo LS, Daza LA, Beers AL, Arbelaez P, Maier O, Chang K, Brown JM, Kalpathy-Cramer J, Zaharchuk G, Wiest R and Reyes M (2018) ISLES 2016 and 2017-Benchmarking Ischemic Stroke Lesion Outcome Prediction Based on Multispectral MRI. *Front. Neurol.* 9:679. doi: 10.3389/fneur.2018.00679

Stefan Winzeck^{1*}, Arsany Hakim², Richard McKinley², José A. A. D. S. R. Pinto³, Victor Alves³, Carlos Silva³, Maxim Pisov^{4,5}, Egor Krivov⁵, Mikhail Belyaev⁵, Miguel Monteiro⁶, Arlindo Oliveira⁶, Youngwon Choi⁷, Myunghee Cho Paik⁷, Yongchan Kwon⁷, Hanbyul Lee⁷, Beom Joon Kim⁸, Joong-Ho Won⁷, Mobarakol Islam⁹, Hongliang Ren⁹, David Robben¹⁰, Paul Suetens¹⁰, Enhao Gong¹¹, Yilin Niu¹², Junshen Xu¹¹, John M. Pauly¹¹, Christian Lucas¹³, Mattias P. Heinrich¹³, Luis C. Rivera¹⁴, Laura S. Castillo¹⁴, Laura A. Daza¹⁴, Andrew L. Beers¹⁵, Pablo Arbelaez¹⁴, Oskar Maier¹³, Ken Chang¹⁵, James M. Brown¹⁵, Jayashree Kalpathy-Cramer¹⁵, Greg Zaharchuk¹⁶, Roland Wiest² and Mauricio Reyes^{17*}

¹ University Division of Anaesthesia, Department of Medicine, University of Cambridge, Cambridge, United Kingdom,

² Support Center of Advanced Neuroimaging (SCAN), Institute of Diagnostic and Interventional Neuroradiology, University of Bern, Inselspital, Bern University Hospital, Bern, Switzerland, ³ CEMES-UMinho Research Unit, University of Minho, Braga, Portugal, ⁴ Moscow Institute of Physics and Technology, Dolgoprudny, Russia, ⁵ Institute for Information Transmission Problems (RAS), Moscow, Russia, ⁶ Instituto de Engenharia de Sistemas e Computadores Investigação e Desenvolvimento, Lisbon, Portugal, ⁷ Department of Statistics, Seoul National University, Seoul, South Korea, ⁸ Department of Neurology and Cerebrovascular Center, Seoul National University Bundang Hospital, Seongnam, South Korea, ⁹ Department of Biomedical Engineering, National University of Singapore, Singapore, Singapore, ¹⁰ ESAT-PSI, KU Leuven, Leuven, Belgium, ¹¹ Electrical Engineering and Radiology, Stanford University, Stanford, CA, United States, ¹² Computer Science, Tsinghua University, Beijing, China, ¹³ Institute of Medical Informatics, Universität zu Lübeck, Lübeck, Germany, ¹⁴ Biomedical Engineering, University of Los Andes, Bogotá, Colombia, ¹⁵ Athinoula A. Martinos Center for Biomedical Imaging, Harvard, MA, United States, ¹⁶ Department of Radiology, Stanford University, Stanford, CA, United States, ¹⁷ Medical Image Analysis, Institute for Surgical Technology and Biomechanics, University of Bern, Bern, Switzerland

Performance of models highly depend not only on the used algorithm but also the data set it was applied to. This makes the comparison of newly developed tools to previously published approaches difficult. Either researchers need to implement others' algorithms first, to establish an adequate benchmark on their data, or a direct comparison of new and old techniques is infeasible. The Ischemic Stroke Lesion Segmentation (ISLES) challenge, which has ran now consecutively for 3 years, aims to address this problem of comparability. ISLES 2016 and 2017 focused on lesion outcome prediction after ischemic stroke: By providing a uniformly pre-processed data set, researchers from all over the world could apply their algorithm directly. A total of nine teams participated in ISLES 2015, and 15 teams participated in ISLES 2016. Their performance was evaluated in a fair and transparent way to identify the state-of-the-art among all submissions. Top ranked teams almost always employed deep learning tools, which were predominately convolutional neural networks (CNNs). Despite the great efforts, lesion outcome prediction persists challenging. The annotated data set remains publicly available and new approaches can be compared directly via the online evaluation system, serving as a continuing benchmark (www.isles-challenge.org).

Keywords: stroke, stroke outcome, machine learning, deep learning, benchmarking, datasets, MRI, prediction models

1. INTRODUCTION

Defining the location and extent of a stroke lesion is an essential step toward acute stroke assessment. Of special interest is the development of a lesion over time, as this could provide valuable information about tissue outcome after stroke onset. Modern magnetic resonance imaging (MRI) techniques, including diffusion and perfusion imaging, have shown great value to distinguish between acutely infarcted tissue (known as “core”) and hypo-perfused tissue (known as “penumbra”). However, available automated methods used to characterize core and penumbra regions from MRI information lack accuracy and cannot correctly capture the complexity of the image information. Hence, there is a great need for advanced data analysis techniques that identify these regions and predict tissue outcome in a more reproducible and accurate way. Eventually, such tools will be available to support clinicians in their decision-making process (e.g., deciding for or against thrombolytic therapy). In recent years machine learning methods for medical image computing have shown unprecedented levels of progress. The area of supervised machine learning (i.e., where computer models are trained based on existing pre-annotated datasets) and particular deep learning, has gained much popularity and has shown great potential for medical applications where image quantification and interpretation is important for the decision making process (1). Along with this, the benchmarking of machine learning techniques for medical image computing has become a central area of interest at the annual Medical Image Computing and Computer Assisted Intervention (MICCAI) conference, where algorithms are tested and evaluated using curated datasets and common evaluation metrics. The ISLES challenge was created as an effort to raise the interest of the medical image computing community to make progress on the challenging aspects of stroke lesion characterisation from MRI data. The work of Maier and colleagues summarizes the lessons learned from the ISLES 2015 edition (2), where image analysis at the subacute and acute stages provided insights as to how accurate machine learning approaches could characterize core and penumbra regions. In the following years the discussions happening among interdisciplinary teams at the ISLES challenge, allowed the community to advance toward the challenge of stroke lesion prediction from MRI data. This is of great interest in a clinical routine, as the responsible physician needs to decide quickly, whether the particular stroke patient could benefit from an interventional treatment (i.e., thrombectomy or thrombolysis). This decision is often draw on basis of lesion appearance, the time passed since stroke onset and the clinicians personal experience. Objective methods that reliably predict lesions and clinical outcome only from the acute scans would be a powerful tool to support and accelerate decision making during the critical phase.

1.1. Current Methods

From the literature review presented by Maier et al. (2), summarizing the state of the art until 2016, the recent machine learning methods for stroke lesion segmentation and outcome prediction clearly show the transition from classic machine

learning tools [e.g., (3, 4)] to approaches based on deep learning (5–10). Generally, the accuracy of those methods is tightly connected to the data set they have been applied to and prevent a direct comparison. For this reason, the development of a publicly available benchmarking, such as ISLES is crucial to facilitate the analysis of current machine learning technologies and leverage research lines to improve them. The ISLES challenge held in 2016 and 2017 have hosted a total of 24 teams participating in the lesion segmentation and outcome prediction sub-tasks. In this article, we present the main results and findings in benchmarking machine learning approaches presented at ISLES 2016 and 2017. The ISLES challenges feature 75 cases from two different centers, including perfusion and diffusion imaging (Raw Perfusion, CBF, CBV, TTP, Tmax, ADC, MTT) as well as clinical information (time-since-stroke, time-to-treatment, TIC1 and mRS scores). Through reference annotations produced by two clinical experts, and a set of quantitative metrics and qualitative expert evaluations, we analyse and describe common strategies and approaches leading to best algorithmic performance. We present the progress of these algorithms, and current challenges that these techniques need to overcome in order to integrate them into the time-critical clinical workflow of stroke patients.

1.2. Motivation for ISLES and Challenge Setup

Automated methods for lesion segmentation and prediction are part of an active research field. Since results are highly dependent on the size and quality of the used data, comparison of independent validation methods is challenging. In order to compare different automated methods, researchers typically have to reimplement algorithms presented in previous publications, which is known to be a difficult task due to the complexity of the algorithms, and lack of detailed description of their implementation. Although the community is changing and provides more frequently open source code is more frequently provided, benchmarking remains time consuming. For these reasons, computational challenges aim to provide a platform allowing a fair and on going validation of various methods tackling a predefined problem. The ISLES challenge follows this direction by providing a stroke imaging database and benchmarking platform that facilitates the comparisons of new algorithms for lesion segmentation and prediction. ISLES was launched for the first time in 2015 and was successfully continued in the subsequent 2 years. Researchers interested in this challenge could register online and download the imaging data via the SICAS Medical Image Repository (SMIR) platform (11). The training data was provided in a preprocessed format that allowed teams to apply their algorithms directly without need of pre-processing. Furthermore, this ensured that performance differences are mainly driven by the prediction models, rather than different preprocessing steps. Eventually, methods could be directly compared and ranked on a leaderboard to discover the most successful approach.

1.2.1. ISLES 2016

While the focus for ISLES 2015 lied on ischemic stroke lesion segmentation (2), ISLES 2016 aimed for the outcome prediction

TABLE 1 | Participants of ISLES 2016 (more details and main features of each method see **Appendix** ISLES16-A1 to ISLES16-A7).

CH-UBE	University of Bern, Switzerland Incorporating time to reperfusion into the FASTER (3) model of stroke tissue-at-risk
DE-UZL	Institute of Medical Informatics, Universität zu Lübeck, Germany Random forests for stroke lesion and clinical outcome prediction
HK-CUH	Department of Computer Science and Engineering, The Chinese University of Hong Kong Residual Volumetric Network for Ischemic Stroke Lesion Segmentation
KR-SUC*	Department of Statistics, Seoul National University, Korea
KR-SUK*	Deep Convolutional Neural Network Approach for Brain Lesion Segmentation
KR-SUL*	
PK-PNS	Pakistan Institute of Nuclear Science and Technology, Islamabad, Pakistan Segmentation of Ischemic Stroke Lesion using Random Forests in Multi-modal MRI Images
UK-CVI	CVIP, Comp. at School of Science and Eng., University of Dundee, UK Combination of CNN and Hand-crafted feature for Ischemic Stroke Lesion Segmentation
US-SFT	University of Southern California, Fractal Analytics, TopicIQ A Deep-Learning Based Approach for Ischemic Stroke Lesion Outcome Prediction

*These methods are variants of a single method.

of lesions. Therefore, multispectral MRI data from acute phase of 35 stroke patients were provided together with lesion maps annotated on 3–9 month follow-up scans. After a period of several weeks, participating teams (See **Table 1**) were asked to apply their algorithm to 19 unseen test data. The lesion labels for the test data were generated by two raters independently, and merged via the STAPLE algorithms (12) to generate a fused ground-truth dataset. On basis of the performance on this test data set, methods were ranked to define a winner of the challenge. As a second task, teams were asked to predict the clinical mRS score, which denotes the degree of disability. Upon analysis of the results, we acknowledge that the latter task required more data for a reliable statistical analysis, which is why they are not presented in this paper. However, the reader is referred to the official website for ISLES 2016¹ for more details.

1.2.2. ISLES 2017

Similarly to ISLES 2016, in 2017 participants were asked to predict lesion outcome on MRI data. The data set of ISLES 2016 was expanded to a total of 43 patients for the training phase, and 32 cases for methods evaluation (see **Table 3**). For the additional 13 test cases, added in 2017, only one groundtruth was generated (in contrast to the other 19 cases from ISLES 2016, for which two annotations per cases exist). For ISLES 2017, participants were asked to submit an abstract, describing their approach, until August 2017. Mid August the test data was distributed and teams had the chance to apply their models and submit their final prediction 2 weeks later. Participating teams and their submitted

TABLE 2 | Participants of ISLES 2017 (more details and main characteristic of each method see **Appendix** ISLES17-A1 to ISLES17-A14).

AAMC	Athinoula A. Martinos Center, USA Ensembling 3D U-Nets For Ischemic Stroke Lesion Segmentation
HKU-1	Hong Kong University of Science and Technology, China Deep Adversarial Networks for Stroke Lesion Segmentation
HKU-2	Hong Kong University of Science and Technology, China Stochastic Dense Network for Brain Lesion Segmentation
INESC	INESC-ID, Portugal Fully Convolutional Neural Network for 3D Stroke Lesion Segmentation
KU	Korea University, Korea Gated Two-Stage Convolutional Neural Networks for Ischemic Stroke Lesion Segmentation
KUL	KU Leuven, Belgium Dual-scale Fully Convolutional Neural Network for Final Infarct Prediction
MIPT	Moscow Institute of Physics and Technology, Russia Neural Networks Stroke Lesion Segmentation
NEU	NEUROPHET Inc. Seoul, South Korea Combination of U-Net and Densely Connected Convolutional Networks
NUS	National University of Singapore, Singapore Fully Convolutional Network with Hypercolumn Features for Brain Lesion Segmentation
SNU-1*	Seoul National University, Korea
SNU-2*	Schemic Stroke Lesion Segmentation with Convolutional Neural Networks for Small Data
SU	Stanford University, USA Multi-scale Patch-wise 3D CNN for Ischemic Stroke Lesion Segmentation
UA	Universidad de los Andes, Colombia Volumetric Multimodality Neural Network For Ischemic Stroke Segmentation
UL	University of Luebeck, Germany 2D Multi-Scale Res-Net for Stroke Segmentation
UM	Universito of Minho, Portugal Combining Clinical Information for Stroke Lesion Outcome Prediction using Deep Learning

*These methods are variants of a single method.

abstract titles can be found in **Table 2**, along with main features of each method (detailed description of methodology in **Appendix**).

The access to the ISLES data remains open so that future research efforts can easily be compared against the existing benchmark.

1.3. Data and Methods

1.3.1. Data Acquisition and Pre-processing

Subjects used for the database, were patients treated for acute ischemic stroke at the University Hospital of Bern or at the UMC Freiburg between 2005 and 2015. Diagnosis of ischemic stroke was performed by identification of lesions on DWI and PWI MR imaging. Digital subtraction angiography was employed to

¹<http://www.isles-challenge.org/ISLES2016/>

document proximal occlusion of the middle cerebral artery (M1 or M2 segment).

Patient inclusion criteria considered: (I) Identification of ischemic stroke lesions on DWI and PWI imaging, (II) proximal occlusion of the middle cerebral artery (M1 or M2 segment) documented on digital subtraction angiography, (III) attempt for endovascular therapy was undertaken, either by intra-arterial thrombolysis (before 2010) or by mechanical thrombectomy (since 2010), (IV) no motion artifacts during pretreatment MR imaging, and (V) patients had a minimum age of 18 years at the time of stroke. Patients were excluded if they had undergone a purely diagnostic angiography and if stenosis or occlusion of the carotid artery were found.

MR imaging was performed on a 1.5T (Siemens Magnetom Avanto), and on a 3T MRI system (Siemens Magnetom Trio). The stroke protocol encompassed whole brain DWI, (24 slices, thickness 5 mm, repetition time 3,200 ms, echo time 87 ms, number of averages 2, matrix 256*256) yielding isotropic b0 and b1000 as well as apparent diffusion coefficient maps (ADC) that were calculated automatically. Additionally, a T2 image was acquired for each case, which was not released to participants but used later for the generation of the groundtruth lesion outcome delineations (section 1.3.2) For PWI, the standard dynamic-susceptibility contrast enhanced perfusion MRI (gradient-echo echo-planar imaging sequence, repetition time 1,410 ms, echo time 30 ms, field of view 230*230 mm, voxel size: 1.8*1.8*5.0 mm, slice thickness 5.0 mm, 19 slices, 80 acquisitions) was acquired. PWI images were acquired during first pass of a standard bolus of 0.1 mmol/kg gadobutrol (Gadovist, Bayer Schering Pharma, Berlin, Germany). Contrast medium was injected at a rate of 5 ml/s followed by a 20 ml bolus of saline at a rate of 5 ml/s. Perfusion maps were obtained by block-circular singular value decomposition using the Olea-Sphere software v2.3 (Olea Medical, La Ciotat). Raw PWI images were also released to participants in the form of a single 4D Nifti image, to allow teams interested in using a different parametric map reconstruction method. All PWI maps (rBF, rBV, TTP, Tmax, MTT) were rigidly registered to the ADC image and automatically skull-stripped (2) to extract the brain area only. We remark, this alignment step was performed to standardize the pre-processing step, hence, to factor out this pre-processing step from the evaluation of results. The cohort curated in 2016 was then extended into a larger dataset for the challenge in 2017. **Table 3** summarizes the ISLES 2016 and 2017 dataset.

1.3.2. Groundtruth Lesion Outcome Segmentation

The lesion outcome status was manually segmented by a board-certified neuroradiologist using 3D Slicer v4.5.0-1, and based

on the 90-day follow-up T2 image. Regions of maximal extent of the final infarction, including haemorrhagic transformation but excluding hyper-intense areas on the acute T2 image (i.e., infarctions due to previous CVI), were delineated on every transversal slice. The 90-day follow-up lesion was chosen to be delineated, since it yields a more reliable final lesion volume than the apparent lesion volume that is observable on subacute images. Groundtruth images were converted into the Nifti format for distribution to participants. For the 19 test cases of ISLES 2016, two lesion annotations were generated by individual raters, and subsequently merged via STAPLE algorithm (12).

1.3.3. Lesion Characteristics

We performed a correlation analysis to assess a possible connection between clinical variables and the performance of the automated methods. The evaluation was conducted for ISLES 2017 submitted methods. **Table 4** summarizes the collected information.

1.3.4. Evaluation Metrics

As quantitative evaluation metrics of the presented methods, we calculated the Dice score as a measure of overlap between manually outlined and automatically predicted lesions. To further shed light on the algorithm's effect we computed precision and sensitivity scores. With TP, true positives; FP, false positive and FN, false negative; the metrics were defined as followed:

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

TABLE 4 | Summary of lesion characteristics for ISLES 2017 Data.

Lesion count	mean [min, max] = 2.46 [1, 14]
Lesion volume	mean [min, max] = 37.83 ml [1.6 ml, 160.4 ml]
Lesion localisation in Lobes	for all 32 cases lesions were located in more than one lobe
Lesion localisation	$n_{subcortical}=3, n_{cortical}=29$
Involved territory	$n_{MCA}=29, n_{MCA+PCA} = 1, n_{multiple}=1$
Midline shift	not present for any of the 32 cases
Laterality	$n_{left}=16, n_{right}=16$
White matter lesions*	$n_0=9, n_1=10, n_2=8, n_3=5$

n, number of cases with given feature; MCA, middle cerebral artery; PCA, posterior cerebral artery.

*Fazekas Classification: 0, absent; 1, punctuate; 2, beginning confluent areas; 3, large confluence.

TABLE 3 | Details of the ISLES 2016 & 2017 Data.

	2016	2017
Number of cases	35 training and 19 testing	43 training and 32 testing
Number of expert segmentations for training and testing sets	1 (training), 2 (testing)	1 (training), 1 (testing)
MRI sequences	ADC, rBF, rBV, MTT, TMAX, TTP, Raw PWI	ADC, rBF, rBV, MTT, TMAX, TTP, Raw PWI

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (3)$$

Alongside these, we measured the maximum surface distance between automatically defined volume and the manually delineated groundtruth volumes by means of the Hausdorff distance (HD). Denoting A_S and B_S as the surface voxels of groundtruth and segmentation volume, respectively, we calculated:

$$HD(A_S, B_S) = \max \left\{ \max_{a \in A_S} \min_{b \in B_S} d(a, b), \max_{b \in B_S} \min_{a \in A_S} d(b, a) \right\} \quad (4)$$

As distance measure $d(\cdot, \cdot)$ we used the Euclidean distance.

Additionally, the average symmetric surface distance (ASSD) was computed for ISLES 2016:

$$ASSD(A_S, B_S) = \frac{ASD(A_S, B_S) + ASD(B_S, A_S)}{2} \quad (5)$$

with the average surface distance (ASD) defined as:

$$ASD(A_S, B_S) = \frac{\sum_{a \in A_S} \min_{b \in B_S} d(a, b)}{|A_S|} \quad (6)$$

1.3.5. Ranking Approach

In order to rank participant's submission for ISLES 2017, we focused on Dice score, as it combines both precision and sensitivity into one metric, and the HD metric. First, both measurements were computed for each patient data individually. Then, all participants were ranked for each metric separately on a case-wise basis such that a high Dice score and a low HD resulted in a high rank. The mean of both ranks yielded a case specific rank. A participant's total rank is obtained by averaging the ranks over all cases (see **Figure 1**). Ranks for ISLES 2016 were computed in the same way for both available groundtruths. Furthermore, ASSD was included alongside Dice and HD for ISLES 2016. In case where teams were not submitting all testing results, the Dice scores were completed with 0 and a large (i.e., $1e+5$) value was set for HD. All unsuccessful segmentation (Dice = 0), were always ranked last. Segmentations with the exact same metrics received the same rank.

1.3.6. Fusion and Thresholding of Softmax Maps

Fusing the output of several classifiers has been shown to yield better results than the single classifiers. This concept is the foundation for ensemble learners, such as random forest (13), and has also been shown to be beneficial for tumor lesion segmentation (14, 15). In theory, each different model could provide valuable, complementary information to enhance the overall segmentation performance. All submitted methods for ISLES 2017 were deep neuronal networks. These include by design a final classification layer, which is commonly a softmax function that provides voxel-wise output values between [0, 1] (further referred to as softmax maps). This output can be interpreted as a probability of voxel belonging to its given class (in this case healthy or lesion tissue). To leverage potential benefit of several submitted models, we averaged the softmax maps of the top five and top three ranked methods for each individual case, followed by its thresholding at the 0.5 mark. Moreover, the softmax maps were thresholded at various levels and subsequently binarised in order to analyse the robustness of methods. Finally, the Dice score was computed between these binary images and the groundtruth.

1.3.7. Statistical Analysis

To assess statistical differences between the submitted methods we applied a Friedman test, a non-parametric, one-way analysis of variances for repeated measurements, and *post-hoc* Dunn test for multiple comparison between teams. For all tests we used GraphPad PRISM Version 5.0.1. The levels of significant differences are marked in plots with asterisks (* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$).

2. RESULTS

2.1. ISLES 2016

2.1.1. Inter-observer Variance

The annotated volumes by rater 1 range from median [Q1, Q3] = 16.7 [6.1, 41.6] ml, and for rater 2 from median [Q1, Q3] = 9.0 [2.9, 36.8] ml, revealed the tendency of rater 1 having

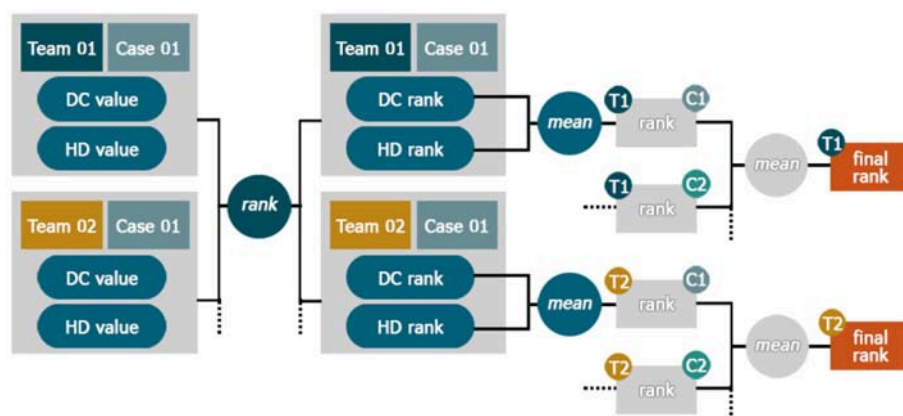


FIGURE 1 | Ranking Scheme. The teams were sorted by their different performance metrics e.g., Dice score (DC) and assigned a rank value per case. Ranks for each team were then separately averaged on a case-wise basis. The final team's rank was then calculated as the mean of all its case-ranks.

segmented more tissue as lesion than rater 2. In 18 out of 19 cases, rater 1 outlined larger lesion volumes, which holds true especially for rather small lesions. Comparing the overlap between manually outlined lesions of both raters yielded an average Dice score of 0.58 ± 0.20 , with median $[Q1, Q3] = 0.62 [0.39, 0.77]$. The relative low coherence between the experts' annotations shows the difficulty of outlining the follow-up image.

2.1.2. Leaderboard and Statistical Analysis

Table 5 shows the ranking of the submitted methods. Only four (KR-SUC, CH-UBE, HK-CHU, PK-PNS) out of nine teams managed to get a successful lesion prediction (Dice > 0) for all 19 cases. The ranking reflects mostly the teams' Dice ranks, except for CH-UBE which was ranked on fourth place despite the second lowest average Dice score (not shown in table). This can be explained by the relative good HD (not shown in table) in comparison to the last ranked teams (see **Table 5** places 7–9).

Analysing the Dice scores across all methods showed that almost all methods are superior to that of US-SFT, which was ranked last. Only PK-PNS, which came second to last, was not found statistically different from US-SFT. The winning approaches (KR-SUC, KR-SUK, KR-SUL) achieved also significantly higher Dice scores than PK-PNS. All methods ranked in second cluster of groups (CH-UBE, DE-UZL, HK-CUH, UK-CVI) did not show statistically significant differences to one and another (see **Figure 2**).

Comparing the Dice scores directly for both manual annotations individually, revealed a positive bias toward the groundtruth generated by the second rater. For all teams the average Dice for both groundtruths varied around five percentage points (see **Figure 3**).

2.2. ISLES 2017

2.2.1. Leaderboard

Only one (SU) of the 15 teams was able to predict stroke lesions (Dice > 0) consistently for all 32 cases. Examining the average Dice and HD rank for each time separately, revealed that the

second ranking team (UL) yielded a lower Dice rank than the following two teams (i.e., HKU-1 and INESC). However, UL achieved the best HD rank, which secured its second place (see **Table 6**).

2.2.2. Dice, Precision and Sensitivity

Table 7 summarizes the participating teams' performance, measured by Dice score, precision and sensitivity, highlighting the strengths of different models. Team KUL's model was the most precise while showing lower sensitivity. AAMC's model showed the highest sensitivity while lacking in precision. Although HKU-1 achieved the highest mean Dice score, it was ranked third seemingly due to a lower HD rank (compare **Table 6**). Even top ranking models reached a low average Dice score of around 0.3, underlining the substantial difficulty of lesion forecasting.

Analysing the Dice score per case disclosed a wide range of quality of lesion outcome prediction. While there are a few cases (28–32) where the average Dice score was above 0.5, the majority of cases turned out to be hard to predict. For 14 cases at least one team achieved a prediction that was overlapping with the groundtruth by 50% (**Figure 5**). For six cases (1–5, 9) none of the teams reached the overall mean Dice score (0.23).

TABLE 5 | Leaderboard ISLES 2016: The rank specifies the final value to order methods relative to each other by performance.

Place	Team	Rank	Dice rank	HD rank	ASSD rank	Cases
1	KR-SUL	3.03	3.37	2.79	2.92	18/19
2	KR-SUC	3.57	3.58	3.71	3.42	18/19
3	KR-SUK	3.82	3.74	4.13	3.61	19/19
4	CH-UBE	3.95	4.26	3.76	3.82	19/19
5	DE-UZL	4.21	4.21	3.82	4.61	19/19
6	UK-CVI	4.08	5.11	4.68	5.45	16/19
7	HK-CHU	5.59	5.08	5.53	6.16	19/19
8	PK-PNS	6.48	6.34	7.58	5.55	12/19
9	US-SFT	8.07	8.03	8.03	8.16	11/19

Dice, HD, and ASSD rank are the average achieved ranks for each participating team per case. The last column gives the number of successfully (Dice > 0) predicted lesions. Best mean values printed in bold.

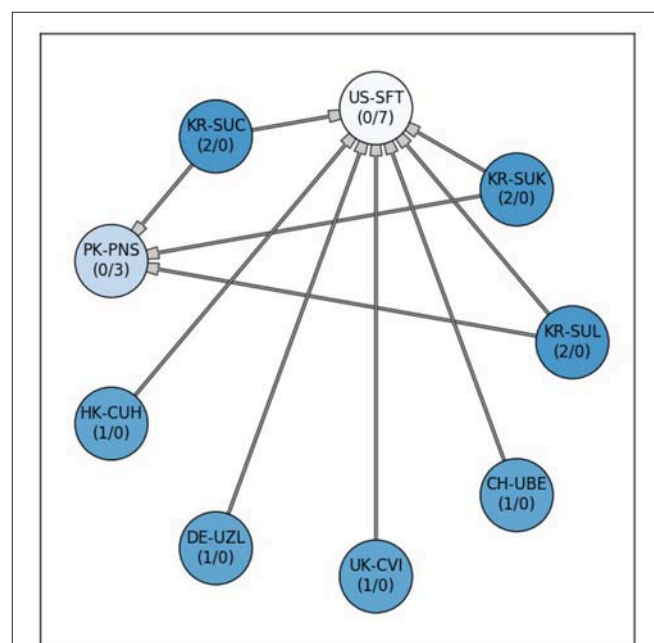


FIGURE 2 | Significant differences between the 9 submitted methods for ISLES 2016. Each node stands for one participating team. A connection between the nodes represents a significant difference between both lesion prediction models. Methods at the tail side of the arrow indicate superiority to the corresponding connected one. The stronger or weaker a model is the more outgoing or incoming connections (#outgoing/#incoming, respectively), are associated with a team's node. Additionally, the node's color saturation indicates the strength of a method (differences in Friedman test rank sum), with better methods appearing more saturated (i.e., darker blue). All methods, except for PK-PNS, are significantly better than US-SFT (post-hoc Dunn test $p < 0.05$).

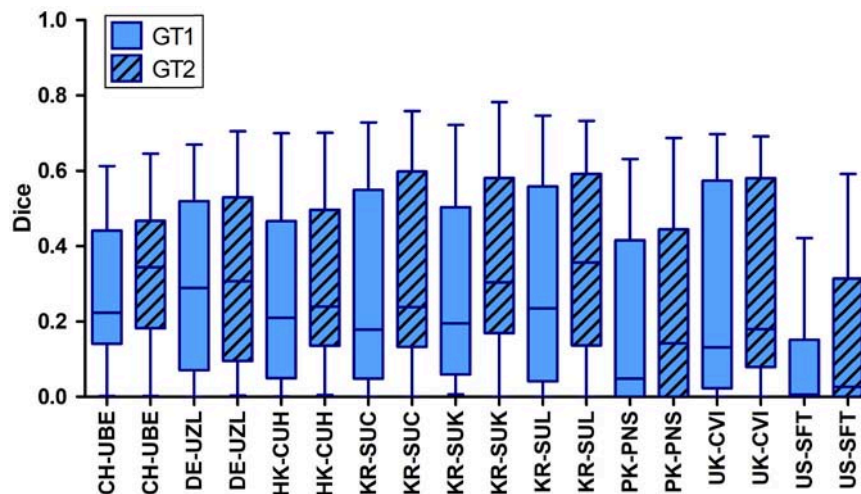


FIGURE 3 | Distribution of Dice scores computed between the automatic lesion predictions and both groundtruths (GT1 and GT2) individually for ISLES 2016. For all teams the Dice scores computed with respect to rater 1 were significantly lower than those calculated with respect to the 2nd groundtruth (GT2).

TABLE 6 | Leaderboard ISLES 2017: While the rank denotes the final value used to sort the teams performances relative to each other.

Place	Team	Rank	Dice rank	HD rank	Cases
1	SNU-2	5.25	4.53	5.97	30/32
2	UL	5.42	6.16	4.69	29/32
3	HKU-1	5.55	5.09	6.00	29/32
4	INESC	5.92	5.00	6.84	31/32
5	KUL	6.03	6.19	5.88	30/32
6	SNU-1	6.47	6.25	6.69	29/32
7	UM	6.58	6.31	6.84	31/32
8	MIPT	6.72	6.34	7.09	30/32
9	SU	7.20	7.09	7.31	32/32
10	KU	8.75	10.09	7.41	28/32
11	AAMC	9.05	8.63	9.47	27/32
12	UA	9.78	9.31	10.25	29/32
13	NUS	9.95	9.50	10.41	29/32
14	NEU	10.44	11.88	9.00	16/32
15	HKU-2	11.80	12.50	11.09	14/32

Dice and HD rank are the average achieved ranks for each participating team. The cases column denotes the number of successfully ($DC > 0$) predicted lesions. Best mean values printed in bold.

TABLE 7 | Average Dice score, precision and sensitivity for individual teams across all 32 cases for ISLES 2017.

Place	Team	Dice	Precision	Sensitivity
1	SNU-2	0.31 ± 0.23	0.36 ± 0.27	0.45 ± 0.31
2	UL	0.29 ± 0.21	0.34 ± 0.26	0.51 ± 0.33
3	HKU-1	0.32 ± 0.23	0.34 ± 0.27	0.39 ± 0.28
4	INESC	0.30 ± 0.22	0.34 ± 0.27	0.51 ± 0.31
5	KUL	0.27 ± 0.22	0.44 ± 0.33	0.39 ± 0.31
6	SNU-1	0.28 ± 0.23	0.36 ± 0.31	0.41 ± 0.31
7	UM	0.29 ± 0.22	0.26 ± 0.24	0.61 ± 0.28
8	MIPT	0.27 ± 0.20	0.31 ± 0.28	0.39 ± 0.29
9	SU	0.26 ± 0.21	0.28 ± 0.25	0.56 ± 0.26
10	KU	0.17 ± 0.16	0.23 ± 0.28	0.36 ± 0.33
11	AAMC	0.23 ± 0.22	0.19 ± 0.20	0.62 ± 0.37
12	UA	0.19 ± 0.16	0.27 ± 0.25	0.21 ± 0.18
13	NUS	0.19 ± 0.16	0.29 ± 0.26	0.23 ± 0.22
14	NEU	0.11 ± 0.16	0.17 ± 0.25	0.12 ± 0.17
15	HKU-2	0.05 ± 0.10	0.17 ± 0.28	0.05 ± 0.11

All evaluation measures are given in mean \pm standard deviation. Best mean values printed in bold.

2.2.3. Statistical Comparison of Team Performances

Figure 6 shows the comparison of the team's Dice scores on the test data set. Each method, represented as node, connects to other methods when a statistical differences in the Dice scores was found. Methods associated to nodes with more outgoing and less incoming connections can be considered stronger than other with less outgoing or more incoming connections. The nodes for stronger models were further grouped and indicated by a more saturated color. This visually highlights the winning team SNU-2 that showed overall higher Dice scores for the prediction

lesions than the other six teams, while none of the other methods were significantly better. This is closely followed by HKU-1 and INESC having each five outgoing edges. The two worst methods (NEU, HKU-2) failed to predict the lesions for several subjects completely, resulting in poor performance inferior to most teams (9 and 10 respectively).

2.2.4. Performance of Single Models Vs. Ensembles

As mentioned in section 1.3.6 we fused the softmax maps to create an ensemble of the top five ($E5 = \text{SNU-1, SNU-2, UL, INESC, KUL}$) and top three ($E3 = \text{SNU-2, UL, INESC}$) ranking

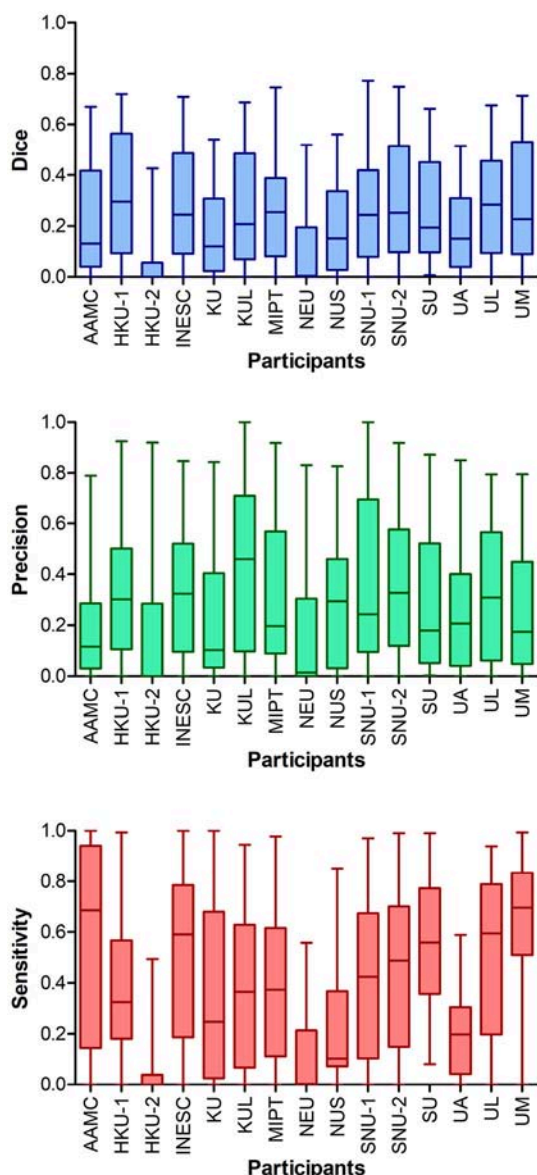


FIGURE 4 | Performance metrics for all teams of ISLES 2017. Higher ranking teams (e.g., 1st place SNU-2) achieved Dice scores > 0.7 for some cases, however, overall Dice scores clustered around 0.2–0.3. The two teams ranked last (NEU and HKU-2) showed much lower Dice scores than all other teams, which was a consequence of the low number of successful submissions. The model of UM seemed to be most sensitive to detect lesions, but lacks in precision.

teams² and compared both ensembles to their individual models. All included models had no significantly different Dice score distributions in comparison to each other (see **Figure 6**).

Figure 7 shows that Dice scores of both ensembles were very similarly distributed as the single models' Dice scores.

²No softmax maps from team HKU-1 were made available, which is why we include the next ranked team on the list i.e., SNU-1 for E5 and INESC for E3.

Ensemble E3 did not result in an improved performance, although the median Dice score (0.28) was higher in comparison to ensemble E5 (0.25) and to the winning team SNU-2 (0.26). Likewise, its mean precision was higher (0.34), although not statistically significant, than most single models (SNU-1, SNU-2, UL, INESC). However, the mean sensitivity of E3 (0.51) could be raised over the one from SNU-1 (0.44).

In contrast, ensemble E5 yielded a significantly better mean Dice score (0.31) than UL (0.28, $p < 0.05$) and SNU-1 (0.26, $p < 0.01$). Among the five teams, whose models were used to build the ensemble, SNU-1 was ranked the lowest, explaining why E5 performed significantly better than SNU-1 by itself. While the ensemble's sensitivity was not improved, combining all softmax maps together significantly increased the precision over four single models ($p < 0.01$, INES, SNU-1, SNU-2, UL).

Figure 8 displays an example of the different participants' softmax maps as well as the fused softmax maps of both ensembles (E3 & E5). While softmax maps from INESC and SNU-2 showed similar certainty values through out the predicted lesion, the other three teams' softmax maps appeared to be more heterogeneous. In contrast to the smooth and blob-like structures predicted by SNU-1, SNU-2, INESC and KUL, UL's model provided a greater detail for boundaries. This is also cohesive with the findings, that UL has the highest HD rank (see **Table 6**) as this metric is considering closeness of boundaries. Dice scores of the lesion predictions for this particular patient could not be improved by ensembles ($\text{Dice}_{E5} = 0.76$, $\text{Dice}_{E3} = 0.73$) in comparison to the single teams ($\text{Dice}_{SNU-1} = 0.76$, $\text{Dice}_{SNU-2} = 0.74$, $\text{Dice}_{UL} = 0.60$, $\text{Dice}_{INESC} = 0.70$, $\text{Dice}_{KUL} = 0.69$).

2.2.5. Analysis of Robustness of Lesion Outcome Prediction

We computed Dice scores between the manually outlined lesion groundtruth and differently thresholded and binarised softmax maps for the top five ranking teams. For four teams (SNU-2, UL, INESC & SNU-1) the Dice scores seemed to be fairly robust and centered around the initial threshold of 0.5. SNU-2's and INESC's prediction vary only in about 1 percentage point for different threshold values (see Appendix: **Table A1**). As an exception, KUL's softmax layer thresholded at a lower level of 0.3 resulted in a higher Dice score (0.28) compared to the the lower Dice (0.26) at a threshold level of 0.5. This effect is coherent with previous findings (see **Table 7** and **Figure 4**) that KUL produces highly precise predictions with relative low sensitivity. Thresholding at a lower level could assign more voxels to the lesion class, hence increased the model's sensitivity and effectively improve Dice scores.

2.2.6. Correlation of Lesion Volumes

When comparing predicted lesion volumes with the manually outlined lesion volumes for the top five ranked teams as mentioned in section 2.2.4, we found a significant correlation only for SNU-1 (Spearman coefficient $r = 0.39$) and for SNU-2 (Spearman coefficient $r = 0.37$). All other teams submission and the ensembles did not correlate with the human rater's annotations, with Spearman coefficients ranging from 0.28 (UL)

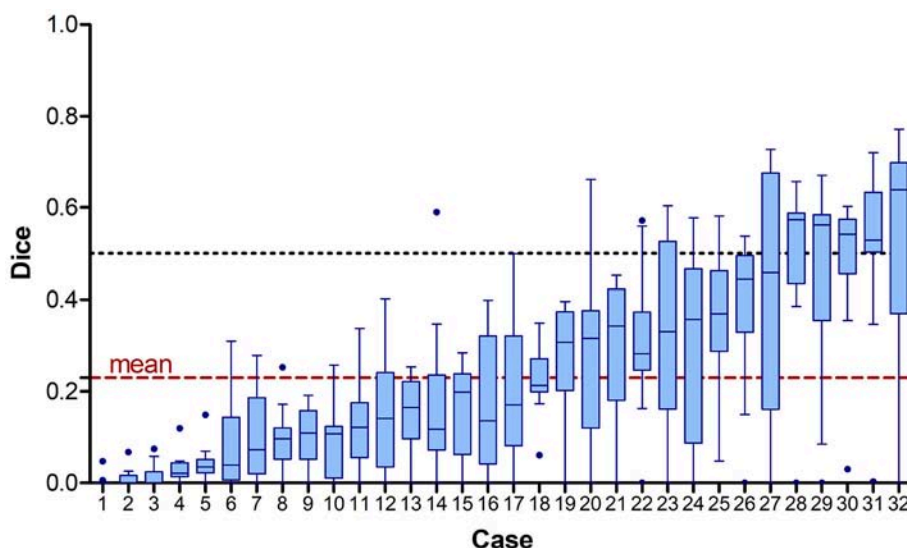


FIGURE 5 | Achieved Dice scores for each case across all 15 participating teams sorted by mean value. The dashed line shows the overall mean Dice score of 0.23 (red) and the 0.5 mark (black). Note that the case numbers were assigned according to ascending mean Dice score.

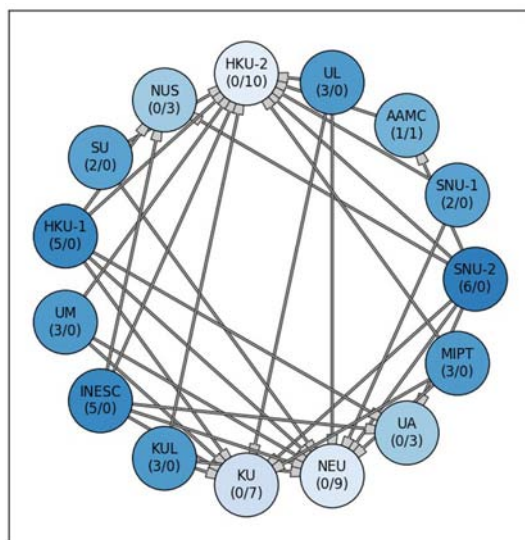


FIGURE 6 | Significant differences between the 15 submitted methods at ISLES 2017. Each node stands for one participating team. A connection between two nodes represents a significant difference between both lesion prediction models, whereas the methods at the tail side was superior. The stronger or weaker a models the more outgoing or incoming connections (#outgoing/#incoming), are associated with a team's node. Additionally, the node color saturation indicates the strength of a method, with better methods appearing more saturated. Differences between methods were assessed via non-parametric ANOVA with repeated measurements (Friedman test) and subsequent, pair-wise comparison with Dunn test ($p < 0.05$).

highest for UL (0.72), INESC (0.71) and E3 (0.70), and lowest for KUL (0.41) and SNU-1 (0.55). Mid-range Spearman coefficients were found for SNU-2 (0.59) and E5 (0.68).

3. DISCUSSION

3.1. Current Performance of Stroke Lesion Outcome Prediction Methods

In ISLES 2016, results showed that deep learning models outperformed *Random Classification Forests* (RF). However, no conclusive superiority of deep learning was found against other machine learning approaches, as demonstrated by CNN-based approaches also ranking in the low tier ranks. Analysing precision and sensitivity revealed the tendency of models to yield over-estimated lesion segmentations. The large variability within the assessed metrics could be explained by the strong correlation between performance and lesion sizes.

Discussions during the ISLES 2016 session led to the decision to enrich the existing ISLES dataset to further encourage participation of the computer science community. Especially, data driven approaches such as deep learning algorithms could truly benefit from larger data sets. Consequently, in ISLES 2017 the training and testing dataset were extended versions of the training and testing sets used in ISLES 2016. For both years, data were provided in minimally pre-processed format. This should allow a more direct comparison of different stroke prediction models, without the influence of any specific pre-processing steps. Of course advanced processing could foster the tissue outcome prediction, however we argue that our focus for the challenges lies on the model development. Furthermore, the applied pre-processing steps were kept to a minimum and are commonly accepted techniques, such as co-registration. This did not prevent participants to further process the provided

to 0.35 (E5). As expected, the Dice scores of all models correlated significantly with the lesion volumes, such that the higher the volume the higher the Dice scores. Spearman coefficients were

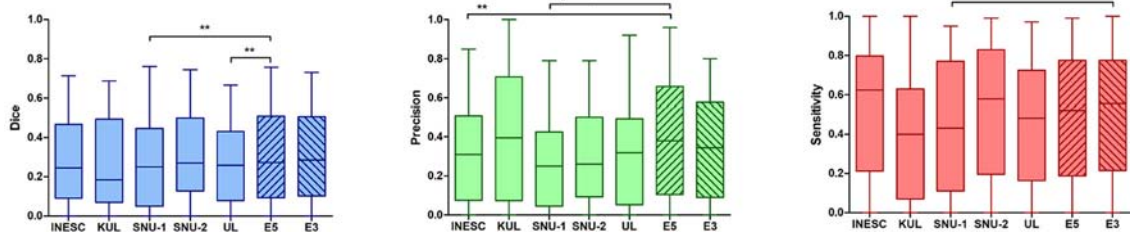


FIGURE 7 | Statistical comparison of lesion prediction performance of single models vs. ensembles. **Left:** An ensemble of five models (E5) could improve the Dice score in comparison to the two weaker models (SNU-1 $p < 0.01$, UL $p < 0.05$). This effect was, however, not observed when building an ensemble with three models (E3). **Middle:** The ensemble E5 significantly gained precision in contrast to most of the single models (SNU-1 $p < 0.01$, SNU-2 $p < 0.05$, UL $p < 0.001$, INESC $p < 0.01$). KUL's precision was higher or similar to that of the ensembles, showing no significant difference. **Right:** The ensemble E3 was found to be more sensitive to predict lesion than SNU-1's model. Overall the models show a fair ability to detect lesions. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

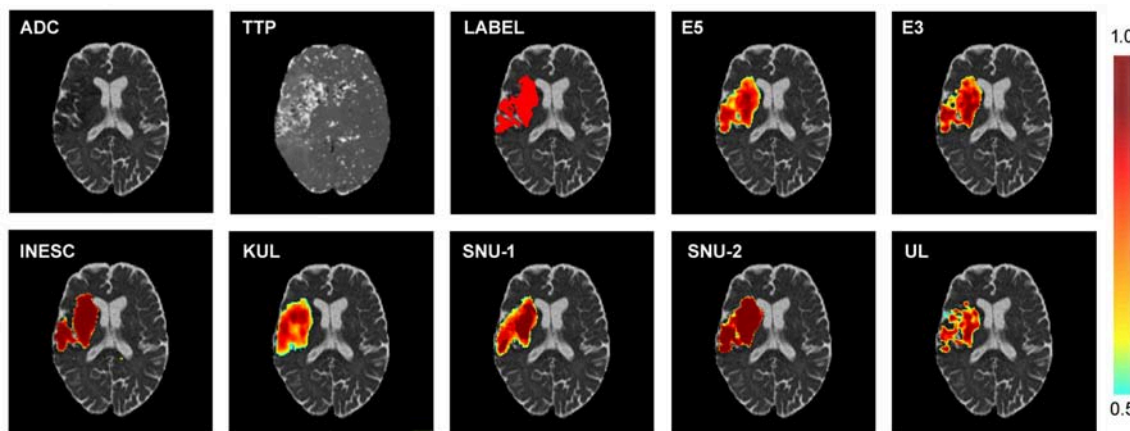


FIGURE 8 | Example of different softmax maps of one patient. Top row: Diffusion (ADC) and perfusion (TTP) scan and the corresponding manual lesion annotation (LABEL) and the softmax maps of the ensembles of the top five (E5) and top three (E3) ranked teams. Bottom row: Softmax maps of the top five ranking teams. Both shape and certainty (see color bar) of the predicted lesion vary between the different participants.

data. Although teams also had partly access to raw data (i.e., raw perfusion data), all of them preferred to work with the pre-processed data.

All participating teams of ISLES 2017 suggested a deep learning approach, with top ranked methods featuring CNN architectures. Despite the increased size of the training data, the overall performance was surprisingly not much different than for ISLES 2016. Top ranked models were found to operate on a similar level, sharing similar architectures and system characteristics. Even ensembles of different CNNs were not strong enough to boost the performance further. These results suggest that CNNs' performance may have reached a plateau on this dataset. Future investigation need to strongly focus on improved training strategies for CNNs or on development of new methodologies to advance stroke lesion outcome prediction. Enhancing the performance especially for small sized lesions and incorporating non-imaging information could bear a strong potential for improvement.

It has been shown that ensemble approaches or fusion of results can improve segmentation predictions (14, 15). Our

findings suggest that the ensemble approaches had a tendency to perform better than single models. Despite the unimproved sensitivity of the ensembles, combining all softmax maps together significantly increased the precision over four single models. This suggests a reduction of false positive predictions. However the effect was not strong enough to result in statistical improvement over the highest ranked single method. It was also not entirely clear which model contributed to enhance or worsen the performance. In fact, the submissions for ISLES 2017 included single as well as ensembles of neural networks, but the ranking did not reflect an overall superiority of ensemble methods. Although the combination of several weak classifiers can cancel out individual model's limitations, it is nonetheless important to build an ensemble of strong methods to leverage benefits and justify increased computational costs of an ensemble based approaches.

Examining each participating team's softmax maps was motivated to analyse their potential to describe their correctness and certainty to perform the task. As these models are intended to provide a prediction of stroke lesion outcome, we postulate

that model calibration is an important aspect for future analysis of deep learning models used for stroke lesion prediction. Particularly, it will have to be investigated how model capacity, regularization and normalization can affect model's calibration, despite apparent increases in model's accuracy (16).

Our findings support the use of different ranking metrics and align with the findings reported in Maier et al. (2). For example the team UL was ranked second in ISLES 2017 thanks to its top HD rank, despite being assigned a relative low Dice rank of 6.16, which would equate the fourth place on the leaderboard.

Overall, the difficulty of the task is reflected by the low Dice scores, with top methods averaging a Dice rank of 0.3. The low Dice scores of the models can be explained by the inherent challenges of the prediction task. Contrary to stroke lesion segmentation, stroke lesion outcome models are trained to predict the lesion status at a 90-days follow-up image based on the acute imaging information. Inherently, many factors contribute to tissue recovery or infarction, which are not explicitly nor implicitly characterized in the imaging information acquired at time of the stroke infarct.

3.2. Limitations and Remaining Challenges

Looking at the evolution of ISLES over the past 3 years, a clear convergence of methodology is observable. While for ISLES 2015 and 2016, still classic machine learning models, such as RF were explored, all submissions of ISLES 2017 offered a variation of CNNs. With their undeniable benefits and success, deep learning methods have set new state-of-the-art benchmarks in many disciplines. Although at present time, this would be the sensible direction to develop further techniques for stroke lesion segmentation and outcome prediction, future challenges will need to encourage exploration of more diverse models. Particularly, we remark the importance of designing methodologies capable of incorporating clinical and physiological prior information on stroke infarction and recovery.

The comparison of the automatic lesion outcome prediction with both expert annotations separately (ISLES 2016) showed a systematic bias toward a higher accordance with rater 2. While this emphasizes the importance of a common database to compare algorithms, it also unveils the general underlying dilemma of supervised learning methods and the intrinsic inter-rater variability observed in medical imaging applications. In best case, algorithms that learn solely from human annotation will only ever be as good as the best human rater and inevitable learn humans' fallacy. Overcoming this limitation calls for semi- and unsupervised learning techniques to teach the computer to detect abnormal brain tissue more accurately, as well as to consider inter-rater variability as source of information during the learning process (17). Nonetheless, a fair and consistent evaluation of such methods has yet to be established. Furthermore, our evaluation is challenged by the different levels of expertise in each team. Although there is a clear tendency that CNNs provide overall better results than RF, some CNNs were ranked lowest. This rather suggest potential deficiencies in the training scheme than a deficiency of this model class in general.

Another challenge is the interpretability of the output of the applied models. Although models are desired to predict lesions

with high precision and confidence level, there may lay valuable information in a models uncertainty for clinical decision making. Regarding lesion outcome prediction, uncertainty could give for example a better indicator of tissue at risk of infarction (e.g., naively thought: high certainty means high risk of becoming lesion tissue, while low certainty may reflect tissue likely to be healthy in future). For future challenges we recommend to ask teams to submit non-binary output maps (e.g., softmax maps) that support such analysis. Most methods work indeed best when incorporating multi-parametric information, however, the database will need to be explored, as in Pereira et al. (18) to gain knowledge on which MR sequences are important and to what extent.

4. CONCLUSION

Over the past years, the ISLES team was able to build an increasingly larger MRI database for ischemic stroke lesion MRI. With this publicly available dataset and a continuously open evaluation system, ISLES has the potential to serve as a standard benchmark framework, where researchers can test their algorithms against an existing pool of described and compared methods (14 ISLES 2015 methods for lesion segmentation, and 28 ISLES 2015 & 2016 and 2017 methods for lesion outcome prediction). Despite the great efforts and accomplishments present at ISLES, automatic segmentation of stroke lesions, and more so lesion outcome prediction remain challenging tasks. Deep learning approaches have great potential to leverage clinical routine for stroke lesion patients, but last years of progress at ISLES indicate that further developments are needed to support clinical decision making by incorporating imaging and readily-available non-imaging clinical information, collateral flow modeling, and further improve the interpretability of deep learning systems used for the clinical decision making process of stroke patients.

ETHICS STATEMENT

All datasets were fully anonymised through skull-stripping and removal of all patient informations by means of conversion of dice files to nifty volumetric files following the regulations of the Swiss Law for Human Research. Further information added below for sake of completeness (In German). Anonymisierung: Unter anonymisiertem biologischem Material und anonymisierten gesundheitsbezogenen Daten ist die irreversible Aufhebung des Personenbezuges zu verstehen. Eine solche liegt dann vor, wenn Material bzw. Daten überhaupt nicht oder nur mit einem un-verhältnismässig grossen Aufwand an Zeit, Kosten und Arbeitskraft der betreffenden Person zugeordnet werden können (vgl. Art. 3 Bst. i HFG und Art. 25 Abs. 1 HFV). Wann den Anforderungen an eine korrekte Anonymisierung Genüge getan ist, ist je nach Einzelfall zu entscheiden: Die Streichung nur des Namens kann bei einer sehr grossen Datenmenge (grosse Personenpopulation) genügen, auch wenn andere Parameter (z.B. Geburtsjahr) verbleiben. Ist die betroffene Population jedoch sehr klein, so ist

das Entfernen nur des Namens nicht ausreichend (vgl. Botschaft zum HFG, S. 8096). Insbesondere unkenntlich gemacht oder gelöscht werden müssen Namen, Adresse, Geburtsdatum und eindeutig kennzeichnende Identifikationsnummern (Art. 25 Abs. 2 HFG). Das im ursprünglichen Art. 14 HFG (vgl. Botschaft zum HFG, S. 8105) vorgesehene Verbot der Anonymisierung von biologischem Material bzw. Personendaten bei Forschungsprojekten mit Bezug zu schweren Krankheiten wurde auf Antrag der vorberatenden Kommission vom Nationalrat gestrichen (vgl. Amtliches Bulletin des Nationalrats, 09.079, Verhandlung vom 10.03.2011). Hintergrund war vermutlich das in Art. 32 Abs. 3 HFG festgelegte Informations- und Widerspruchsrecht der Patienten bei Forschung mit anonymisiertem biologischen Material und genetischen Daten. Dadurch sind die Patienten nämlich ausreichend geschützt, ein zusätzliches Verbot schien vor diesem Hintergrund wohl obsolet. Mit Streichung des ursprünglichen Artikels 14 HFG ist die Forschung mit anonymisiertem biologischem Material also auch bei Forschungsprojekten mit Bezug zu schweren Krankheiten zulässig, sofern die betroffenen Personen vorgängig korrekt informiert und auf ihr Widerspruchsrecht hingewiesen wurden.

AUTHOR CONTRIBUTIONS

SW, AH, and MR collected data, performed analysis, wrote manuscript. RM, RW, JP, VA, CS, MP, MB, EK, MM, AO, YC,

YK, MP, BK, J-HW, MI, HR, DR, PS, YN, EG, JX, JMP, GZ, EK, CL, MH, LC, PA, AB, KC, JB, JK-C, LR, LD and OM performed analysis, wrote manuscript.

FUNDING

Fundacao para a Ciencia e Tecnologia (FCT), Portugal (scholarship number PD/BD/113968/2015). FCT with the UID/EEA/04436/2013, by FEDER funds through COMPETE 2020, POCI-01-0145-FEDER-006941. NIH Blueprint for Neuroscience Research (T90DA022759/R90DA023427) and the National Institute of Biomedical Imaging and Bioengineering (NIBIB) of the National Institutes of Health under award number 5T32EB1680. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. PAC-PRECISE-LISBOA-01-0145-FEDER-016394. FEDER-POR Lisboa 2020-Programa Operacional Regional de Lisboa PORTUGAL 2020 and Fundação para a Ciência e a Tecnologia. GPU computing resources provided by the MGH and BWH Center for Clinical Data Science Graduate School for Computing in Medicine and Life Sciences funded by Germany's Excellence Initiative [DFG GSC 235/2]. National Research National Research Foundation of Korea (NRF) MSIT, NRF-2016R1C1B1012002, MSIT, No. 2014R1A4A1007895, NRF-2017R1A2B4008956 Swiss National Science Foundation-DACH 320030L_163363.

REFERENCES

- Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* (2017) 42:60–88. doi: 10.1016/j.media.2017.07.005
- Maier O, Menze BH, von der Gablentz J, Hani L, Heinrich MP, Liebrand M, et al. ISLES 2015-A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. *Med Image Anal.* (2017) 35:250–69. doi: 10.1016/j.media.2016.07.009
- McKinley R, Häni L, Gralla J, El-Koussy M, Bauer S, Arnold M, et al. Fully automated stroke tissue estimation using random forest classifiers (FASTER). *J Cereb Blood Flow Metab.* (2017) 37:2728–41. doi: 10.1177/0271678X16674221
- Pustina D, Coslett H, Turkeltaub PE, Tustison N, Schwartz MF, Avants B. Automated segmentation of chronic stroke lesions using LINDA: lesion identification with neighborhood data analysis. *Human Brain Mapp.* (2016) 37:1405–21. doi: 10.1002/hbm.23110
- Guerrero R, Qin C, Oktay O, Bowles C, Chen L, Joules R, et al. White matter hyperintensity and stroke lesion segmentation and differentiation using convolutional neural networks. *NeuroImage Clin.* (2018) 17:918–34. doi: 10.1016/j.nicl.2017.12.022
- Lucas C, Maier O, Heinrich MP. Shallow fully-connected neural networks for ischemic stroke-lesion segmentation in MRI. In: Maier-Hein, geb. Fritzsche K, Deserno, geb. Lehmann T, Handels H, Tolxdorff T, editors. *Bildverarbeitung für die Medizin 2017*. Berlin; Heidelberg: Springer (2017). p. 261–6.
- Kamnitsas K, Ledig C, Newcombe VF, Simpson JP, Kane AD, Menon DK, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal.* (2017) 36:61–78. doi: 10.1016/j.media.2016.10.004
- Chen L, Bentley P, Rueckert D. Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks. *NeuroImage Clin.* (2017) 15:633–43. doi: 10.1016/j.nicl.2017.06.016
- Menze BH, Van Leemput K, Lashkari D, Riklin-Raviv T, Geremia E, Alberts E, et al. A generative probabilistic model and discriminative extensions for brain lesion segmentation—with application to tumor and stroke. *IEEE Trans Med Imaging* (2016) 35:933–46. doi: 10.1109/TMI.2015.2502596
- Havaei M, Guizard N, Larochelle H, Jodoin PM. Deep learning trends for focal brain pathology segmentation in MRI. In: *Machine Learning for Health Informatics*. Springer (2016). pp. 125–48.
- Kistler M, Bonaretti S, Pfahrer M, Niklaus R, Büchler P. The virtual skeleton database: an open access repository for biomedical research and collaboration. *J Med Int Res.* (2013) 15:e245. doi: 10.2196/jmir.2930
- Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging* (2004) 23:903–21. doi: 10.1109/TMI.2004.828354
- Breiman L. Random forests. *Mach Learn.* (2001) 45:5–32. doi: 10.1023/A:1010933404324
- Kamnitsas K, Bai W, Ferrante E, McDonagh S, Sinclair M, Pawlowski N, et al. Ensembles of multiple models and architectures for robust brain tumour segmentation. *arXiv [Preprint]*. arXiv:171101468. (2017).
- Crimi A, Bakas S, Kuijff H, Menze B, Reyes M. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. In: *Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017, September 14, 2017, Revised Selected Papers*. Vol. 10670. Quebec City, QC: Springer (2018).
- Chuan G, Geoff P, Yu S, Kilian Q W. On Calibration of Modern Neural Networks. CoRR. 2017;abs/1706.04599. Available online at: <http://arxiv.org/abs/1706.04599>
- Jungo A, Meier R, Ermis E, Blatti-Moreno M, Herrmann E, Wiest R, et al. On the Effect of Inter-observer Variability for a Reliable Estimation of Uncertainty of Medical Image Segmentation. *arXiv [Preprint]*. arXiv:180602562 (2018).

18. Pereira, Raphael M, Richard M, Roland W, Victor A, Carlos S, et al. Enhancing interpretability of automatically extracted machine learning features: application to a RBM-Random Forest system on brain lesion segmentation. *Med Image Anal.* (2018) 44:228–44. doi: 10.1016/j.media.2017.12.009
19. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu (2017) p. 2881–90.
20. Paszke A, Chaurasia A, Kim S, Culurciello E. Enet: a deep neural network architecture for real-time semantic segmentation. *arXiv [Preprint]*. arXiv:160602147 (2016).
21. Sudre CH, Li W, Vercauteren T, Ourselin S, Cardoso MJ. Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Québec: Springer (2017). p. 240–8.
22. Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, et al. Brain tumor segmentation with deep neural networks. *Med Image Anal.* (2017) 35:18–31. doi: 10.1016/j.media.2016.05.004
23. Li W, Manivannan S, Akbar S, Zhang J, Trucco E, McKenna SJ. Gland segmentation in colon histology images using hand-crafted features and convolutional neural networks. In: *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on. IEEE*. Prague (2016) p. 1405–8.
24. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Athens: Springer (2016). p. 424–32.
25. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: *Advances in Neural Information Processing Systems*. Montreal, QC (2014). p. 2672–80.
26. Avants BB, Tustison NJ, Stauffer M, Song G, Wu B, Gee JC. The insight toolKit image registration framework. *Front Neuroinformatics* (2014) 8:44. doi: 10.3389/fninf.2014.00044
27. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv [Preprint]*. arXiv:1409.1556 (2014).
28. Chen H, Yu L, Dou Q, Shi L, Mok VC, Heng PA. Automatic detection of cerebral microbleeds via deep learning based 3d feature representation. In: *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on IEEE*. New York, NY (2015). p. 764–7.
29. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV (2016). p. 770–8. doi: 10.1109/CVPR.2016.90
30. Choi Y, Kwon Y, Lee H, Kim BJ, Paik MC, Won JH. Ensemble of Deep Convolutional Neural Networks for Prognosis of Ischemic Stroke. In: *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Athens: Springer (2016). p. 231–43.
31. Luc P, Couprie C, Chintala S, Verbeek J. Semantic segmentation using adversarial networks. *arXiv [Preprint]*. rXiv:161108408. (2016).
32. Maier O, Wilms M, von der Gablentz J, Krämer UM, Münte TF, Handels H. Extra tree forests for sub-acute ischemic stroke lesion segmentation in MR sequences. *J Neurosci Methods* (2015) 240:89–100. doi: 10.1016/j.jneumeth.2014.11.011
33. Milletari F, Navab N, Ahmadi S. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. CoRR. Available online at: <http://arxiv.org/abs/1606.04797> (2016).
34. Varma M, Zisserman A. A statistical approach to texture classification from single images. *Int J Comput Vis.* (2005) 62:61–81. doi: 10.1007/s11263-005-4635-4
35. Huang G, Liu Z, Weinberger KQ, van der Maaten L. Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Vol. 1. Honolulu (2017). p. 3.
36. Ronneberger O, Fischer P, Brox T. U-net: Convolutional Networks for Biomedical Image Segmentation. Munich: CoRR; Springer (2015). p. 234–41.
37. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston, MA (2015). p. 3431–40.
38. Chen H, Qi X, Cheng JZ, Heng PA. Deep Contextual Networks for Neuronal Structure Segmentation. In: Phoenix A, Dale S, Michael W, editors. *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence* (2016). p. 1167–1173.
39. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. CoRR. Available online at: <http://arxiv.org/abs/1512.03385> (2015).
40. Frakes WB, Baeza-Yates R. *Information Retrieval: Data Structures & Algorithms*. Vol. 331. Upper Saddle River, NJ: Prentice Hall Englewood Cliffs (1992).
41. Cho K, Van Merriënboer B, Bahdanau D, Bengio Y. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv [Preprint]*. arXiv:1409.1259 (2014).

Conflict of Interest Statement: 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.

Copyright © 2018 Winzeck, Hakim, McKinley, Pinto, Alves, Silva, Pisov, Krivov, Belyaev, Monteiro, Oliveira, Choi, Paik, Kwon, Lee, Kim, Won, Islam, Ren, Robben, Suetens, Gong, Niu, Xu, Pauly, Lucas, Heinrich, Rivera, Castillo, Daza, Beers, Arbelaes, Maier, Chang, Brown, Kalpathy-Cramer, Zaharchuk, Wiest and Reyes. 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.

A. APPENDIX

This following sections briefly summarizes the participants' algorithms.

A.1. ISLES 2016

A.1.1. ISLES16-A1. CH-UBE - Incorporating Time to Reperfusion Into the FASTER Model of Stroke Tissue-at-Risk

Authors: *Richard McKinley, Roland Wiest, and Mauricio Reyes*

In a recent paper, we introduced the tool FASTER (Fully Automated Stroke Tissue Estimation using Random Forests) (3), which aims to give an assessment of the tissue at risk in acute stroke beyond the usual paradigm of predefined thresholds on single maps. The FASTER system assesses the likelihood of tissue damage using decision forest classifiers, mapping local statistical features of perfusion and diffusion imaging onto maps of the tissue predicted to be lost even if reperfusion is established, and the tissue predicted to be lost only if there is no reperfusion. These models are trained only on extreme cases, in which reperfusion was total and rapid (TICI 3), or completely absent (TICI 0). In this work we attempt to go further, predicting the likely tissue loss in the case of TICI grades 1-2b, by interpolating between the two predictions yielded by FASTER, and incorporating the time to revascularization.

A.1.1.1. Acknowledgments

The authors acknowledge the support of the Schweizerische Herzstiftung.

TABLE A1 | Dice score dependency of threshold for softmax maps.

Team	Thresholds								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
INESC	0.28	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29
KUL	0.22	0.26	0.28	0.27	0.26	0.23	0.20	0.15	0.02
SNU-1	0.20	0.23	0.25	0.26	0.26	0.27	0.23	0.20	0.16
SNU-2	0.29	0.29	0.30	0.30	0.30	0.30	0.30	0.30	0.30
UL	0.19	0.24	0.26	0.27	0.28	0.28	0.27	0.25	0.21

TABLE A2 | Overview of methods of participants of ISLES 2016.

CH-UBE	Random Forest classifier integrating time to reperfusion
DE-UZL	Random Forests classifier
HK-CUH	U-Net architecture; summation instead of concatenation of different pathways
KR-SUC	Ensemble of U-Net architecture and fully convolutional neural network
KR-SUK	
KR-SUL	
PK-PNS	Random Forest classifier
UK-CVI	Combination of CNN and hand-crafted features
US-SFT	U-Net architecture

A.1.2. ISLES16-A2. DE-UZL - Random Forests for Stroke Lesion and Clinical Outcome Prediction

Authors: *Oskar Maier and Heinz Handels*

Ischemic stroke is caused by an obstruction in the cerebral blood supply and, if diagnosed early, part of the under-perfused tissue can potentially be salvaged. Since the available treatment options are not risk-free, the decision has to be made individually, depending on the potential gain and under great time restriction. The prediction of the final lesion outcome in form of A binary mask (Task I) and the prediction of the clinical outcome in form of the modified Rankin Scale (mRS) (Task II) are therefore of great clinical interest. The ISLES 2016 challenge offers a public dataset and associated expert groundtruth to allow researchers to compare their methods in these two fields directly and fairly. Our contribution works with carefully selected features extracted from the MR sequences and used to train a RF. The data consists of multi-spectral (ADC, PWI maps and raw PWI 4D volumes) scans and associated clinical measures. The final lesion outcome as delineated in a 90 days follow-up scan (Task I) and the 90 days mRS score (Task II) serve as groundtruths. More details on the data can be found on www.isles-challenge.org. Task I: Lesion outcome prediction From each MR sequence we extract the features previously presented in (32), but furthermore employ a hemispheric difference measure to make use of the pseudo-quantitative values provided by the PWI maps. For voxel-wise classification we employ RFs. Task II: Clinical outcome prediction Based on the segmentation results from Task I, we extract lesion characteristics as well as local image features from the supplied cases to train a regression forest. Applied, this yields a prediction of the mRS score for a formerly unseen case. Our method has been shown to provide competitive lesion segmentation results in glimo segmentation as well as acute and semi-acute stroke in the previous year's edition of the ISLES challenge. The results from this year's challenge will show if the advantages of our flexible design also extend to outcome prediction.

A.1.3. ISLES16-A3. HK-CUH - Residual Volumetric Network for Ischemic Stroke Lesion Segmentation

Authors: *Lequan Yu and Pheng-Ann Heng*

We propose a 3D CNNs based method for lesion outcome prediction. The proposed 3D network takes advantage of fully convolutional architecture to perform efficient, end-to-end, volume-to-volume training. More importantly, we introduce the recent proposed residual learning technique into our network, which can alleviate vanishing gradients problem and improve the performance of our network. It employs 3D fully convolutional architecture and is organized in a residual learning scheme. The layers of our network are all implemented with a 3D manner (under caffe library), thus the network can highly preserve and deeply exploit the 3D spatial information of the input volumetric data. We adopt small convolution kernels with size of $3 \times 3 \times 3$ in convolutional layers. Each convolutional layer is followed by a rectified linear unit (ReLU). Note that we also employ batch normalization layer (BN) before each ReLU layer. The BN layer can accelerate the training process of our network. At the end of the network, we add a $1 \times 1 \times 1$ convolutional layer as a classifier to generate the segmentation results and further get

the segmentation probability volumes after passing the softmax layer. Note that our network might appear similar to U-Net, but there are differences: We use summation units instead of concatenation units when combining different paths, and thus we can reformulate our network as residual learning scheme; additionally, we adopt recently developed batch normalization technique to improve our performance.

A.1.4. ISLES16-A4. KR-SUC/KR-SUK/KR-SUL - Deep Convolutional Neural Network Approach for Brain Lesion Segmentation

Authors: *Youngwon Choi, Yongchan Kwon, Hanbyul Lee, Myunghee Cho Paik, and Joong-Ho Won*

Brain lesion segmentation is a challenging problem because the amount of lesion area is extremely small and the size of available training magnetic resonance images are limited. To handle this, we exploit millions of 3D patches and 3D convolutional kernels for our proposed model. By treating each 3D patch as training data we capitalize on spatial information and overcome the problem of limited medical data. Our final segmentation model is an ensemble of two deep convolutional neural networks inspired by fully convolutional networks and the U-Net (36). We implement the proposed model in Python with Lasagne and Keras.

A.1.5. ISLES16-A5. PK-PNS - Segmentation of Ischemic Stroke Lesion Using Random Forests in Multi-Modal MRI Images

Authors: *Qaiser Mahmood and A. Basit*

Multi-modal MRI can be used for detecting the ischemic stroke lesion and can provide quantitative assessment of lesion area. It can be established as an essential paraclinical tool for diagnosing stroke. For a quantitative analysis of stroke lesion in MRI images, clinical expert manual segmentation is still a common approach and has been employed to compute the size, shape, and volume of the stroke lesions. However, it is time-consuming, tedious, and labor-intensive task. Moreover, manual segmentation is prone to intra- and inter-observer variabilities. Herein, we present an automated segmentation method for ischemic stroke lesion segmentation in multi-modal MRI images. The method is based on an RF ensemble learning technique called random forest, which generates several classifiers and combines their results in order to make decisions. In RF, we employ several meaningful features such as intensities, entropy, gradient etc. to classify the voxels in multi-modal MRI images. The segmentation method is validated on training data, obtained from MICCAI ISLES-2016 challenge dataset. The performance of the method is evaluated relative to the manual segmentation, done by the clinical experts. The experimental results show the robustness of the segmentation method, and that it achieves reasonable segmentation accuracy for segmenting the ischemic stroke lesion in multi-modal MRI images.

A.1.6. ISLES16-A6. UK-CVI - Combination of CNN and Hand-Crafted Feature for Ischemic Stroke Lesion Segmentation

Authors: *Haocheng Shen, Siyamalan Manivannan, Roberto Annunziata, Ruixuan Wang and Jianguo Zhang*

Convolutional neural networks can automatically learn discriminative local features and give superior performance than hand-crafted features in various applications such as image classification, semantic segmentation and object detection. CNN has also been applied to MRI brain image analysis and achieved state-of-the-art results for brain tumor region segmentation (7, 22), stroke lesion segmentation (7), and microbleeds detection (28). Recently, some studies [e.g., (23)] show that hand-crafted features may provide complementary information with CNN, hence combining them with the features extracted from CNN may give improved performance than only using the features from CNN. Motivated by this, we formulate the segmentation of ischemic stroke lesion in acute MRI scans as a pixel-level classification using a combination of CNN and hand-crafted features. We used a CNN architecture which is similar to (38). It is a fully convolutional neural network containing a downsampling path and three upsampling paths. In the task of stroke lesion segmentation, there is a large variation on the size, location, and shape of lesions. Therefore, encoding information at multiple scales is necessary and preferable than considering information at only one level. The downsampling path is able to extract the abstract information with high-level semantic meaning, while the three upsampling paths are designed to capture the fine details. These three upsampled feature maps are then combined at the later stages of the CNN architecture so that the classification layer fully make use of the information appears at multiple scales (38). We use the following hand-crafted features: intensity, the hemispheric intensity difference between two symmetric pixels in the axial view, first order statistics in a $w \times w$ volume patch, maximum response filter (MR8) (34). At each 2D pixel location, these local features are extracted independently from each image modality and combined together to get a feature representation for that pixel. As there is a large variation of lesions in the dataset, it will be beneficial to train a pool of binary classifiers instead of one. Each binary classifier in this pool is designed to separate the positive (lesion) features extracted from a patient from all the negative (normal) features extracted from the same patient. In this way we believe that some rarely appeared lesions can be easily discriminated from the normal tissue compared to a binary lesion classifier which is trained using all the training data (without using patient information). In the testing time a voting strategy (averaging the top 3 probabilities obtained by the binary classifiers in the pool) is used to get the prediction of an input.

A.1.7. ISLES16-A7. US-SFT - a Deep-Learning Based Approach for Ischemic Stroke Lesion Outcome Prediction

Authors: *Ramandeep Randhawa, Ankit Modi, Parag Jain, and Prashant Warier*

The ISLES 2016 challenge aims to address two important aspects of Ischemic stroke lesion treatment prediction. The first aspect relates to segmenting the brain MRI to identify the areas with lesions and the second aspect relates to predicting the actual clinical outcome in terms of the patient's degree of disability. The input data consists of acute MRI scans and additional clinical such as TICI scores, Time Since Stroke, and Time to Treatment. To address this challenge we take a deep-learning

based approach. In particular, we first focus on the segmentation task and use an automatic segmentation model that consists of a Deep Neural Network (DNN). The DNN takes as input the MRI images and outputs the segmented image, automatically learning the latent underlying features during the training process. The DNN architectures we consider utilize many convolutional layers with small kernels, e.g., 3×3 . This approach requires fewer parameters to estimate, and allows one to learn and generalize from the somewhat limited amount of data that is provided. One of the architectures we are currently utilizing is based on the U-Net (36), which is an all-convolutional network. It acts as an auto-encoder, that first “encodes” the input image by applying combinations of convolutional and pooling operations. This is followed by the “decoding” step that up-scales the encoded images, while performing convolutions. The all-convolutional architecture of the U-Net allows it to handle input images of different dimensions as in the challenge dataset. In our experiments, we found that this architecture yielded excellent performance on the previous ISLES 2015 dataset. Although the modalities in the 2016 challenge are different, our initial training experiments have yielded promising segmentation results. Our next steps involve addressing the regression challenge. There is limited amount of labeled data for this task. Our approach will be to include these outcomes as part of the segmentation training directly. This will allow the DNN to learn latent features that can directly help with the classification task.

A.2. ISLES 2017

A.2.1. ISLES17-A1. AAMC - Ensembling 3D U-Nets For Ischemic Stroke Lesion Segmentation

Authors: *Andrew Beers, Ken Chang, James Brown, Emmett Sartor, Elizabeth Gerstner, Bruce Rosen, and Jayashree Kalpathy-Cramer*
We propose a novel deep learning architecture based on the 3D Convolutional U-Net, an architecture that has found success both in ISLES 2016 and a wide array of other tissue segmentation

applications. A typical U-Net segmentation architecture operates by convolving and downsampling input data stepwise into a low-resolution representation, and then upsampling and deconvolving that representation into to a categorical labelmap. The downsampling arm of the U-Net is also concatenated at points to the upsampling arm, resulting in a densely-connected architecture. We improve upon previous implementations of the 3D U-Net both by increasing the number of layers and convolutional filters, and by adding multiple independent downsampling arms to the network. The motivation for this chimeric structure is to increase accuracy by concatenating several unique and not necessarily correlated downsampled representations, thereby increasing the potential amount of relevant imaging biomarkers. We apply this architecture on stacked, $16 \times 16 \times 4$ voxel patches of six of the seven given image maps (ADC, CBV, CBF, MTT, TTP, Tmax) for ISLES 2017. For training, 80% of patches are drawn from the groundtruth regions, while 20% of patches are extracted from normal brain. For inference, we predict 16 overlapping output patches per voxel, average overlapping softmax outputs, and threshold those outputs into binary labels. We finally post-process the binary labels by removing small islands and applying repeated segmentation erosions and dilations.

A.2.1.1. Acknowledgments

We would like to acknowledge the GPU computing resources provided by the MGH and BWH Center for Clinical Data Science.

A.2.2. ISLES17-A2. HKU-1 - Deep Adversarial Networks for Stroke Lesion Segmentation

Tony C. W. Mok and Albert C. S. Chung

Training models that provide accurate stroke lesion segmentation for stroke assessment is challenging. Methods based on deep convolutional neural networks usually rely on

TABLE A3 | Overview of methods of participants of ISLES 2017.

AAMC	3D CNN U-Net architecture; increased number of layers and convolutional filter, multiple down-sampling path ways; anisotropic patch size of $16 \times 16 \times 4$; prediction of 16 overlapping patches per voxels, that are averaged. Morphological operations to reduce small clusters of erroneous predictions
HKU-1	U-Net architecture, including data augmentation and batch normalization, adversarial training of two deep neural networks to avoid over-fitting
HKU-2	3D CNN U-Net architecture; long short-term memory (LSTM) to capture information in 3rd dimension of MRI scans; data augmentation
INESC	V-Net architecture; new loss-function: sum of standard cross-entropy loss and dice-loss
KU	Hierarchy of 2 CNNs. 1st CNN discriminates lesion and healthy tissue, 2nd CNN only acts up on voxels where the 1st CNN was uncertain; auto-context (use of probability maps from 1st CNN)
KUL	U-net architecture; data augmentation via x-axis flip, Gaussian noise and small linear intensity transformations; ensemble of 4 networks; suppression of prediction in non-dominant hemisphere
MIPT	Ensemble of E-Net, DeepMedic, and two U-Nets; 2D and 3D architectures; weighted sum of models' predictions; data augmentation: rotation, flips, registration, and elastic co-registration to template
NEU	Combination 3D U-Net and densely connected CNN; refinement with CRF
NUS	PixelNet applied to lesion outcome prediction
SNU-1/SNU-2	Ensemble of three CNNs: U-Net, DeepMedic, pyramid scene parsing network; negative Dice score loss
SU	3D CNN with 2 scale pathways; data augmentation through rigid transformations, weighted ratios on positive and negative labels
UA	CNN with 4 scale pathways
UL	2D U-Net with skip connections; Dice loss is added up to total loss; inversely weighted loss to tackle class imbalance
UM	2D U-Net in combination with clinical information

large amounts of annotated data. The small lesion area and limited size of available acute MRI data would degrade the quality of result using such approaches due to over-fitting the training data. To deal with this problem, we adopt two deep neural networks with adversarial training (25). (31) shows that this technique could generate a regularization effect and result in less over-fitting to the training data. Our model ensemble two deep convolutional neural networks inspired by the U-Net (36). Other technique such as data augmentation and batch normalization are adopted to further improve the final results.

A.2.3. ISLES17-A3. HKU-2 - Stochastic Dense Network for Brain Lesion Segmentation

Authors: *Pei Wang and Albert C. S. Chung*

The segmentation of ischemic stroke lesion in brain MRI is quite challenging for its varying size and unknown shape. To tackle this problem, we proposed a convolutional neural network for an end-to-end, volume-to-volume lesion segmentation. Based on the 3D U-Net structure, we apply dense connection to link every two layers to well combine the low level information with the high level one. In each layer, instead of 3D convolution, we adopt long short-term memory (LSTM) to capture the information of third dimension in MRI. To further reduce the over-fitting during training process, all the dense connections between layers are stochastically established. Due to the limited dataset, data augmentation is applied to the training dataset.

A.2.4. ISLES17-A4. INESC - Fully Convolutional Neural Network for 3D Stroke Lesion Segmentation

Authors: *Miguel Monteiro¹ and Arlindo L. Oliveira*

Our approach consists of a Fully-Convolutional Neural Network (FCNN) with a V-Net (33) architecture. The V-Net architecture is a variation of the U-Net architecture (36) which is commonly used for medical imaging segmentation. This architecture consists of a contracting path and an expanding path each made up of convolution blocks. At each level of the contracting path, the image's spatial dimensions are halved and the number of channels is doubled. In the expanding path, the opposite happens. There are skip-connections between the contracting and expanding path which feed high-resolution features to the expanding path. In addition, the convolution blocks in both paths have skip connections similar to those of the ResNet (39) which make training faster and more robust. To address class imbalanced (most of the voxels are labeled as 0 in the segmentation) we proposed a novel loss function to train the network. This loss function consisted of the sum of the standard cross-entropy loss with the dice-loss. The dice-loss is calculated by taking the negative dice coefficient calculated with label probabilities instead of discrete labels which results in a number between -1 and 0 . Since the cross-entropy loss can take any positive value up to infinity, during training, it begins by dominating the overall loss function. As training progresses, it tends toward 0 , at this point the dice-loss component becomes more dominant which helps fine tune the prediction.

A.2.4.1. Acknowledgments

This work was supported by PAC - PRECISE - LISBOA-01-0145-FEDER-016394, co-funded by FEDER through POR Lisboa 2020 - Programa Operacional Regional de Lisboa PORTUGAL 2020 and Fundação para a Ciência e a Tecnologia.

A.2.5. ISLES17-A5. KU - Gated Two-Stage Convolutional Neural Networks for Ischemic Stroke Lesion Segmentation

Authors: *Jee-Seok Yoon, Eun-Song Kang, and Heung-Il Suk*

We propose a novel framework with a gated two-stage CNN for ischemic stroke lesion segmentation. Specifically, there are two CNNs in our framework. The first CNN produces a probability of being normal tissue, i.e., normal, or being ischemic stroke lesion, i.e., lesion. Based on our observation that as for the misclassified voxels in images, the ratio between probabilities of normal and lesion was low. That is, when the probabilities of normal and lesion are close to each other, it can be a good indicator of low confidence to make a decision. In this regard, we devise a gate function that computes the probability ratio between normal and lesion. When the ratio is lower than a threshold, the gate function turns on the second CNN to operate. It is noteworthy that in our second CNN, we also utilize the probabilities obtained from the first CNN as context information. In our experiments, we could validate the effectiveness of the proposed two-stage CNN architecture.

A.2.5.1. Acknowledgments

This work was supported by Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korea government (No. 2017-0-00451).

A.2.6. ISLES17-A6. KUL - Dual-Scale Fully Convolutional Neural Network for Final Infarct Prediction

Authors: *David Robben and Paul Suetens*

We perform a voxelwise classification to predict the final infarct using relative time-to-peak, ADC and the available metadata. Relative time-to-peak is calculated per voxel as the time-to-peak (TTP) minus the first quartile of the TTP within the brain mask. The given modalities have physical units that can be interpreted absolutely, hence we use per modality the same linear transformation for all subjects: subtraction by the median mean value and scaling with the median standard deviation. The metadata are normalized similarly, after converting the TICI score into a numerical value. Inspired by (7) we implement using Keras a fully convolutional neural network with two pathways, one on the original resolution and one on a lower resolution (in plane subsampled with a factor 3). Both pathways have five $3 \times 3 \times 1$ kernels and five $3 \times 3 \times 3$ kernels to account for the anisotropy of the voxel size. Both pathways and the metadata are subsequently fed into two fully connected layers before the final classification is made. The network is regularized with drop-out and l2-regularization. We augment the training data with flips along the x-axis, Gaussian noise and small linear intensity transformations. Hyperparameters are chosen by evaluating the network's performance during

cross-validation on the training set. Training is stochastic and at testing time we use an ensemble of four networks whose predictions are averaged. The predictions are thresholded at 0.5 and all voxels on the non-dominant side of the brain are suppressed.

A.2.7. ISLES17-A7. MIPT - Neural Networks Ensembles for Ischemic Stroke Lesion Segmentation

Authors: Maxim Pisov, Mikhail Belyaev, and Egor Krivov

We use four different architectures of CNNs for image segmentation: a modification of ENet (20), DeepMedic (7), and two versions of U-Net (36). ISLES-2017 problem is a challenging task because of a strong anisotropy of the data: a typical voxel size is about $1 \times 1 \times 6 \text{ mm}^3$. That's why we used E-Net and U-Net as 2D-segmentation networks: 2D slices along the axial plane were fed into them at both training and inference steps, while DeepMedic was used as a 3D-segmentation network. Based on these network architectures we built several models with different hyper-parameters. The masks predicted by these models had significantly variable geometrical properties, e.g., smooth/rough edges, smaller/bigger regions. To reduce this variability, we used a weighted sum of final models' predictions. As a preprocessing step, we cropped all the brain images to their bounding boxes and rescaled them to the shape 192×192 in the plane xOy . To overcome the dataset size limitations, we use two different data augmentation techniques: classical spatial transformations (e.g., random rotations, random flips along the coronal, and sagittal planes) and a new co-registration-based method. The main idea of the method is to map lesions from a brain with stroke to a healthy brain using elastic co-registration. To augment data in that way we used the approximately age-matched brains of healthy subjects from the Alzheimer's Disease National Initiative dataset (adni.loni.usc.edu) as templates and applied the co-registration algorithm from ANTs toolkit (26).

A.2.8. ISLES17-A8. NEU - Combination of U-Net and Densely Connected Convolutional Networks

Authors: Donghyeon Kim, Joon Ho Lee, Dongjun Jung, Jong-min Yu, and Junkil Been

Brain lesion segmentation is an advanced challenging problem which has been handled by only experienced clinician and could not be localized using a single brain imaging method. Thus, it is essential to analyze it as multi modality sense. To address this challenge, we take convolutional neural network, specially U-Net (36), 3D U-Net (24), and Densely Connected Convolutional Network (35). In feature selection, first of all, we searched the best combination of multi data sets and the best number of convolutional neural layers considering computation cost, accuracy, and overfitting problem. With different numbers of image dataset combination, each different image of training data is ensembled to learn at the front of the bridge part between encoding (convolution layer) and decoding (deconvolution layer) in the proposed network. Furthermore, we consider the type of data extraction of the images (2D and 3D patch) and refining the result such as conditional random field (CRF).

A.2.9. ISLES17-A9. NUS - Fully Convolutional Network With Hypercolumn Features for Brain Lesion Segmentation

Authors: Mobarakol Islam and Hongliang Ren

The segmentation of stroke lesion is very necessary for diagnosis, planning treatment strategies and monitoring disease progression. We propose a fully convolutional network (FCN) with hypercolumns features and sparse pixel predictions (e.g., PixelNet) for automatic brain lesion segmentation. PixelNet extracts feature from multiple layers that correspond to the same pixel and samples a modest number of pixels across a small number of images for each SGD (Stochastic gradient descent) batch update. Deep Learning (DL) models like CNN requires large training data to generalize the model where most of the biomedical problems have small available dataset. Moreover, the problem of label imbalance leads the CNN often converge to the certain labels. PixelNet deals these problems by utilizing sparse pixel prediction on a modest number of pixels. We utilize PixelNet in ISLES (Ischemic Stroke Lesion Segmentation) challenge 2017 and achieve 68% Dice accuracy as preliminary result.

A.2.10. ISLES17-A10. SNU-1 & SNU-2 - Schemic Stroke Lesion Segmentation With Convolutional Neural Networks for Small Data

Authors: Youngwon Choi, Yongchan Kwon, Myunghee Cho Paik, Beom Joon Kim, and Joong-Ho Won

Our approach to the ISLES 2017 challenge was to build an ensemble of three-dimensional CNN models predicting ultimate ischemic stroke lesions from early imaging. We employed three types of CNNs: (I) multiscale U-net (24), (II) multiscale fully-convolutional network (7, 37), and (III) pyramid scene parsing network (19). Negative Dice score, binary crossentropy and weighted binary cross-entropy (21) were used as the loss for training. The multiscale U-net architecture trained with the negative Dice score achieved the best performance among the nine combinations considered. The implementation details such as pre-processing, data augmentation, and regularization are similar to (30), which ranked the 1st place in ISLES 2016. There are two major improvements from our approach to the 2016 challenge. First, the model complexity is reduced by 60% without sacrificing the prediction performance: multiscale U-net with 40,000 parameters showed comparable performance to the 2016 model with 100,000 parameters. Second, the training process is simplified by adopting probability calibration instead of the fine-tuning step in the multiphase training (22).

A.2.10.1. Acknowledgments

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT, NRF-2016R1C1B1012002). Joong-Ho Won's research was partially supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT, No. 2014R1A4A1007895). Myunghee Cho Paik's research was supported by the National Research Foundation of Korea under grant NRF-2017R1A2B4008956.

A.2.11. ISLES17-A11. SU - Multi-scale Patch-Wise 3D CNN for Ischemic Stroke Lesion Segmentation

Authors: Yilin Niu, Enhao Gong, Junshen Xu¹, John Pauly, and Greg Zaharchuk

A deep network model was trained with 3D CNN patch-wise approaches and multi-scale structures. A three-dimensional CNN was implemented to utilize available spatial information efficiently and exploit the relationship between slices. Our patch-wise approach extracts concentric small 3D patches from multi-contrast input volumes to emphasize local voxel information, minimize unrelated distant features and handle various volume dimensions. Overlapping 3D patches were sampled from brain regions (using brain masks) at multiple scales (with 2 scale pathways using $36 \times 36 \times 5$ and $16 \times 16 \times 3$ patch size in the final implementation) to capture both local and global contextual information simultaneously (7). Rigid transformations were used for data augmentation and weighting ratios on positive and negative labels were added to ensure better data balance. The model we implemented has 7 layers, including 1 resample layer right after the inputs, 5 convolutional layers without pooling, 1 resample layer to ensure consistent resolution of the outputs from two scale pathways and 2 fully-connected layers to generate final 6×6 patch outputs. From the 43 cases in the training dataset, we split labeled data into 77% for training and 23% for validation. The Dice Score Coefficient was used as training loss and quality metrics in validation. The model is trained using tensor-flow framework on a Linux server with 2 NVIDIA GTX-1080TI GPUs.

A.2.12. ISLES17-A12. UA - Volumetric Multimodality Neural Network For Ischemic Stroke Segmentation

Authors: Laura Silvana Castillo, Laura Alexandra Daza, Luis Carlos Rivera, and Pablo Arbeláez

High level research architectures for semantic segmentation, such as VGG (27) and FCN (37), take advantage of multiple image resolutions to simultaneously extract fine details and coarse structures from the input data by using groups of convolutional layers and non-linearities, usually Rectified Linear Units (ReLU), followed by pooling operations. However, as the resolution of the image is reduced, so is the accuracy in the segmentation location. To overcome this drawback, we propose a neural network that extracts features from different input resolutions in a parallel and independent manner. Additionally, the use of a patch-wise approach helps to deal with the imbalance of the data and reduces the memory consumption. This allows us to retrieve detailed appearance data along with accurate semantic information simultaneously. Our method is based on DeepMedic (7) and V-Net (33), methods that have shown state of the art on medical image segmentation. We developed a new architecture with four parallel pathways, each one with six convolutional layers and two residual connections, to extract features on specific resolution levels. All the paths receive patches centered at the same voxel, but extracted from different versions of the image (original and downsampled by factors of six and eight). The patches have input sizes of 363, 203, 183, and 153 for the normal, medium and low resolution pathways. An upsample layer is used to make the outputs of the

same size. Finally, the results are concatenated and introduced in fully connected layers to be combined and then classified. The classification layer is a convolution with kernel size of 13.

A.2.13. ISLES17-A13. UL - 2D Multi-Scale Res-Net for Stroke Segmentation

Authors: Christian Lucas and Mattias P. Heinrich

U-Nets (36) have shown competitive performance in different biomedical tasks while being capable of segmenting objects of different scales. Ischemic strokes vary widely in location, shape, and extend of the affected tissue. We thus propose a fully-convolutional architecture based on U-Nets for segmenting transversal image slices. The challenge data has been resampled to a common resolution of $1 \times 1 \times 5$ mm and slices are zero-padded, if required. The network is provided 42 image features as input (7 MR sequences, 3 slices including both direct neighboring slices, 2 hemispheric flips). In the contracting path, fine-grained information is improved across the five scale levels of the U-Net (from 240×240 down to 15×15) by additional skip connections: the input of each level is concatenated channel-wise with the activation [similar to ResNets (29) but with concatenation] before it is downsampled and passed to the deeper level. In the upscaling path, the Dice loss at each level is computed on softmax activation and summed up to a total loss for training. The loss of foreground and background is weighted with its inverse prior probability (estimated from training data) to account for class imbalance. To speed up training, the network parameters are optimized using the ADAM algorithm. Moreover, each convolution (in both paths) is followed by a batch normalization as done before in Lucas et al. (6).

A.2.13.1. Acknowledgments

This work was supported by the Graduate School for Computing in Medicine and Life Sciences funded by Germany's Excellence Initiative [DFG GSC 235/2]. We would also like to thank Nvidia Corporation for their support by providing us with a Titan Xp graphics card.

A.2.14. ISLES17-A14. UM - Combining Clinical Information for Stroke Lesion Outcome Prediction Using Deep Learning

Authors: Adriano Pinto, Richard Mckinley, Victor Alves, Roland Wiest, Carlos A. Silva, and Mauricio Reyes

For stroke lesion outcome prediction, we propose an end-to-end deep learning method capable of merging MRI sequences with non-imaging clinical information, namely the thrombolysis in cerebral infarction (TICI) scale. Since MRI images come from different centers, as preprocessing steps we resized all MRI sequences to $256 \times 256 \times 32$. In addition, the Tmax sequence was clipped to [0, 20s] and the ADC sequence was clipped within the range of $[0, 2600] \times 10^{-6} \text{ mm}^2/\text{s}$, as values beyond these ranges are known to be biologically meaningless (3). Afterwards, all sequences were linearly scale to [0, 255]. Our architecture has two main blocks, the first is based on the 2D-Unet (36), whose output feature maps are injected in a second block composed by two layers of Gated Recurrent Units (41). The clinical domain

knowledge is incorporated at two levels: population and patient levels. The population level is coded in a custom loss function based on the F_β - score (40), having the beta parameter modeled by the TICI scale. To encompass this clinical knowledge into the testing phase we added an extra input channel that contains the TICI score. Therefore, we aim to drive the learning process of the architecture accordingly to the success of revascularization, in order to produce optimistic predictions when the predicted lesion shrinks, and pessimistic predictions when the predicted lesion increases.

A.2.14.1. Acknowledgments

Adriano Pinto was supported by a scholarship from the Fundação para a Ciência e Tecnologia (FCT), Portugal (scholarship number PD/BD/113968/2015). This work is supported by FCT with the reference project UID/EEA/04436/2013, by FEDER funds through the COMPETE 2020 - Programa Operacional Competitividade e Internacionalização (POCI) with the reference project POCI-01-0145-FEDER-006941. We acknowledge support from the Swiss National Science Foundation - DACH 320030L 163363.



Predicting Outcome of Endovascular Treatment for Acute Ischemic Stroke: Potential Value of Machine Learning Algorithms

Hendrikus J. A. van Os^{1*}, Lucas A. Ramos^{2,3}, Adam Hilbert², Matthijs van Leeuwen⁴, Marianne A. A. van Walderveen⁵, Nyika D. Kruij¹, Diederik W. J. Dippel⁶, Ewout W. Steyerberg^{7,8}, Irene C. van der Schaaf⁹, Hester F. Lingsma⁸, Wouter J. Schonewille¹⁰, Charles B. L. M. Majoie¹¹, Silvia D. Olabariaga³, Koos H. Zwinderman³, Esmee Venema^{6,8}, Henk A. Marquering² Marieke J. H. Wermer¹ and the MR CLEAN Registry Investigators

OPEN ACCESS

Edited by:

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

Reviewed by:

Mirjam R. Heldner,
Universität Bern, Switzerland
Muhib Khan,
Michigan State University,
United States

*Correspondence:

Hendrikus J. A. van Os
h.j.a.van_os@lumc.nl

Specialty section:

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

Received: 15 May 2018

Accepted: 30 August 2018

Published: 25 September 2018

Citation:

van Os HJA, Ramos LA, Hilbert A, van Leeuwen M, van Walderveen MAA, Kruij ND, Dippel DWJ, Steyerberg EW, van der Schaaf IC, Lingsma HF, Schonewille WJ, Majoie CBLM, Olabariaga SD, Zwinderman KH, Venema E, Marquering HA, Wermer MJH and the MR CLEAN Registry Investigators (2018) Predicting Outcome of Endovascular Treatment for Acute Ischemic Stroke: Potential Value of Machine Learning Algorithms. *Front. Neurol.* 9:784. doi: 10.3389/fneur.2018.00784

¹ Department of Neurology, Leiden University Medical Center, Leiden, Netherlands, ² Department of Biomedical Engineering and Physics, University of Amsterdam, Amsterdam, Netherlands, ³ Department of Clinical Epidemiology and Biostatistics, University of Amsterdam, Amsterdam, Netherlands, ⁴ Leiden Institute for Advanced Computer Sciences, Leiden University, Leiden, Netherlands, ⁵ Department of Radiology, Leiden University Medical Center, Leiden, Netherlands, ⁶ Department of Neurology, Erasmus Medical Center, Rotterdam, Netherlands, ⁷ Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, Netherlands, ⁸ Department of Public Health, Erasmus Medical Center, Rotterdam, Netherlands, ⁹ Department of Radiology, University Medical Center Utrecht, Utrecht, Netherlands, ¹⁰ Department of Neurology, Antonius Ziekenhuis, Nieuwegein, Netherlands, ¹¹ Department of Radiology and Nuclear Medicine, University of Amsterdam, Amsterdam, Netherlands

Background: Endovascular treatment (EVT) is effective for stroke patients with a large vessel occlusion (LVO) of the anterior circulation. To further improve personalized stroke care, it is essential to accurately predict outcome after EVT. Machine learning might outperform classical prediction methods as it is capable of addressing complex interactions and non-linear relations between variables.

Methods: We included patients from the Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands (MR CLEAN) Registry, an observational cohort of LVO patients treated with EVT. We applied the following machine learning algorithms: Random Forests, Support Vector Machine, Neural Network, and Super Learner and compared their predictive value with classic logistic regression models using various variable selection methodologies. Outcome variables were good reperfusion (post-mTICI ≥ 2 b) and functional independence (modified Rankin Scale ≤ 2) at 3 months using (1) only baseline variables and (2) baseline and treatment variables. Area under the ROC-curves (AUC) and difference of mean AUC between the models were assessed.

Results: We included 1,383 EVT patients, with good reperfusion in 531 (38%) and functional independence in 525 (38%) patients. Machine learning and logistic regression models all performed poorly in predicting good reperfusion (range mean AUC: 0.53–0.57), and moderately in predicting 3-months functional independence (range mean AUC: 0.77–0.79) using only baseline variables. All models performed well in predicting 3-months functional independence using both baseline and treatment

variables (range mean AUC: 0.88–0.91) with a negligible difference of mean AUC (0.01; 95%CI: 0.00–0.01) between best performing machine learning algorithm (Random Forests) and best performing logistic regression model (based on prior knowledge).

Conclusion: In patients with LVO machine learning algorithms did not outperform logistic regression models in predicting reperfusion and 3-months functional independence after endovascular treatment. For all models at time of admission radiological outcome was more difficult to predict than clinical outcome.

Keywords: ischemic stroke, prediction, machine learning, endovascular treatment, functional outcome, reperfusion

INTRODUCTION

Endovascular treatment (EVT) is effective for ischemic stroke patients with a large vessel occlusion (LVO) of the anterior circulation. EVT results in a number needed to treat of 2.6 to reduce disability by at least one level on the modified Rankin Scale (mRS) (1). A recent meta-analysis showed a positive treatment effect of EVT across patient subgroups including different age groups, varying stroke severity, sex, and stroke localization (1). However, many clinical and imaging predictors or their combinations were not considered in the subgroup analysis. Moreover, the RCTs that provided the data differed in their patient selection criteria. To further improve personalized stroke care, it is essential to accurately predict outcome and eventually differentiate between patients who will and will not benefit from EVT.

Machine learning belongs to the domain of artificial intelligence and provides a promising tool in pursuing personalized outcome prediction, which is increasingly used in medicine (2–7). The machine learning methodology allows discovering empirical patterns in data through automated algorithms. In some clinical settings machine learning algorithms outperform classical regression models, such as logistic regression, possibly through more efficient processing of non-linear relationships and complex interactions between variables (6, 8), although poorer performance has also been observed (9).

In this study, we used multiple machine learning algorithms and logistic regression with multiple variable selection methods to predict radiological and clinical outcome after EVT in a cohort of well-characterized stroke patients. We hypothesized that machine learning algorithms outperform classic multivariable logistic regression models in terms of discrimination between good and poor radiological and clinical outcome.

METHODS

Patients

We included patients registered between March 2014 and June 2016 in the Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands (MR CLEAN) Registry. The MR CLEAN Registry is an ongoing, national, prospective, open, multicenter, observational monitoring study covering all 18 stroke

intervention centers that perform EVT in the Netherlands, of which 16 participated in the MR CLEAN trial (10). The registry is a continuation of the MR CLEAN trial collaboration and includes all patients undergoing EVT (defined as entry into the angiography suite and receiving arterial puncture) for acute ischemic stroke in the anterior and posterior circulation. In the current analysis we included those patients who adhered to the following criteria: age 18 years and older, treatment in a center that participated in the MR CLEAN trial, and LVO in the anterior circulation (internal carotid artery (ICA), internal carotid artery terminus (ICA-T), middle (M1/M2) cerebral artery, or anterior (A1/A2) cerebral artery), shown by CT angiography (CTA) or digital subtraction angiography (DSA) (11).

Clinical Baseline Characteristics

We assessed the following clinical characteristics at admission: National Institutes of Health Stroke Scale (NIHSS), Glasgow Coma Scale, medical history (TIA, ischemic stroke, intracranial hemorrhage, subarachnoid hemorrhage, myocardial infarction, peripheral artery disease, diabetes mellitus, hypertension, hypercholesterolemia), smoking, laboratory tests (blood glucose, INR, creatinine, thrombocyte count, CRP), blood pressure, medication (thrombocyte aggregation inhibitors, oral anticoagulant drugs, anti-hypertensive drugs, statins), modified Rankin Score (mRS) before stroke onset, administration of intravenous tPA (yes/no), stroke onset to groin time, transfer from another hospital, and whether the patient was admitted during weekend or off hours.

Radiological Baseline Parameters

All imaging in the MR CLEAN Registry was assessed by an imaging core laboratory (11). On non-contrast CT, the size of initial lesion in the anterior circulation was assessed by the Alberta Stroke Program Early CT Score (ASPECTS). ASPECTS is a 10 point quantitative topographic score representing early ischemic change in the middle cerebral artery territory, with a scan without ischemic changes receiving an ASPECTS of 10 points (12). In addition, presence of leukoaraiosis and old infarctions, hyperdense vessel sign, and hemorrhagic transformation of the ischemic lesion were assessed on non-contrast CT.

On CTA, the core lab determined clot burden score, clot location, collaterals, and presence of intracranial atherosclerosis. The clot burden score evaluates the extent of thrombus in

the anterior circulation by location scored on a 0–10 scale. A score of 10 is normal, implying clot absence; a score of 0 implies complete multi-segment vessel occlusion (12). Presence of intracranial carotid artery stenosis, atherosclerotic occlusion, floating thrombus, pseudo-occlusion, and carotid dissection were scored on CTA of the carotid arteries. Collaterals were assessed using a 4 point scale, with 0 for absent collaterals (0% filling of the vascular territory downstream of the occlusion), 1 for poor collaterals (>0% and ≤50% filling of the vascular territory downstream of the occlusion), 2 for moderate collaterals (>50% and <100% filling of the vascular territory downstream of the occlusion), and 3 for excellent collaterals (100% filling of the vascular territory downstream of the occlusion) (13).

Treatment Specific Variables

Variables collected during EVT were type of sedation during the procedure (general or conscious), use of a balloon guiding catheter, carotid stent placement, performed procedure (DSA only or thrombectomy), and type of EVT-device (stent retriever, aspiration device, or a combination of both). In addition, data were collected on adverse events during the procedure (perforation, dissection, distal thrombosis on DSA).

Interventional DSA parameters in our dataset were occluded vessel segment (ICA: origin, cervical, petrous, cavernous, supraclinoid, M1-M4, A1, A2), arterial occlusive lesion (AOL) recanalization score before and after EVT (14), evidence of vascular injury (i.e., perforation, or dissection, vasospasm, new clot in different vascular territory or distal thrombus confirmed with imaging), and modified Thrombolysis in Cerebral Infarction (mTICI)-score before and after EVT. The mTICI-score grades the following categories of cerebral reperfusion: no reperfusion of the distal vascular territory (0), minimal flow past the occlusion but no reperfusion (1), minor partial reperfusion (2a), major partial reperfusion (2b), and complete reperfusion (3) (15). Further variables analyzed were time from stroke onset to recanalization, post-EVT stay on intensive care, high care or stroke care, NIHSS after EVT (<48 h), delta NIHSS (pre-treatment NIHSS subtracted from NIHSS <48 h after EVT) and hemicraniectomy or symptomatic intracranial hemorrhage <48 h after EVT.

Outcome

The primary radiological outcome was good reperfusion defined as modified TICI-score directly post-procedure (post-mTICI) ≥ 2b (15). The primary clinical outcome was functional independence at 3 months after stroke (mRS ≤ 2). We excluded patients in whom any of the main outcomes (3-months mRS and post-mTICI) were missing.

To investigate the full potential of Machine learning compared with conventional methods in different settings after stroke we defined two prediction settings:

First, we assessed the probability of good reperfusion and good 3-months functional independence in our cohort of patients that underwent EVT based only on variables that were available on admission before entry into the angiography suite. With this baseline prediction setting we are able to investigate the added value of machine learning for models that could

potentially support future clinical decision making regarding the performance of EVT yes or no.

Second, we tested the models for predicting 3-months functional independence in patients after EVT was performed. For this analysis we used all variables collected up to 48 h after the end of the endovascular procedure (baseline and treatment variables).

Machine Learning Algorithms

The machine learning algorithms used in our study were Random Forests, Artificial Neural Network and Support Vector Machine, because they are among the algorithms that are currently most widely and successfully used for clinical data (2–7). Each one of them represents a different algorithm “family,” each with radically different internal algorithm structures (16). Since it was not known beforehand which kind of algorithm would perform best, we chose algorithms with different internal structures to increase the probability of good discriminative performance. We also used Super Learner, which is an ensemble method that finds the optimal weighted combination of predictions of the Random Forests, Artificial Neural Network and Support Vector Machine algorithms used in this study. Ensemble methods, such as Super Learner have been shown to increase predictive performance by increasing model flexibility (17). For the implementation of all machine learning algorithms we used off-the-shelf methods in the Python module Scikit-Learn (18).

Super Learner

Super Learner is a stacking algorithm using cross-validated predictions of other models (i.e., a machine learning algorithm and logistic regression) and assigning weights to these predictions to optimize the final prediction. Super Learner's predictive performance has been found to surpass individual machine learning models in various clinical studies (17, 19, 20).

Random Forests

Random Forests consists of a collection of decision tree classifiers that are fit on random subsamples of patients and variables in the dataset. The variation of the subsampled variables creates a robust classifier. In the decision trees, each node represents a variable and splits the input data into branches based on an objective function that determines the optimal threshold for separating the outcome classes. The predictions from each tree are used as “votes,” and the outcome with the most votes is considered the predicted outcome for that specific patient (6, 21). From the Random Forests algorithm variable importances can be derived, which are the sum of weights of nodes of the trees containing a certain variable, averaged over all trees in the forest (22).

Support Vector Machine

Support Vector Machine (SVM) is a kernel-based supervised machine learning classifier which can also be used to output probabilities. The SVM works by first mapping the input data into a high dimensional variable space. For binary classification, a hyperplane is subsequently determined to separate two classes

such that the distance between the hyperplane and the closest data points is maximized (23).

Artificial Neural Network

In this study we use the multilayer perceptron, a popular class of artificial neural network architecture composed of one or more interconnected layers of neurons that process data from the input layer into predictions for the output layer. The algorithm computes a weight for each neuron based on input activation. These weights are updated by backpropagation and stochastic gradient descent (24, 25).

Logistic Regression

For logistic regression, generally a set of variables has to be selected to be included in the model. Since model performance can rely heavily on selecting the right variables, we tested five different variable selection methods prior to logistic regression. We first selected variables based on prior knowledge, a still widely used method that could be considered “classical” (26). We selected 13 variables available at baseline that were included in a previous study for a similar purpose (27) (**Supplementary Table Ia**). In addition, from baseline and treatment variables we selected 15 variables based on expert opinions of vascular neurologists and radiologists (**Supplementary Table Ib**).

We further considered four automated variable selection methods: (i) backward elimination, which is also considered to be a more classical approach (26), and three state-of-the-art variable selection methods: (ii) least absolute shrinkage and selection operator (LASSO) (28), (iii) Elastic Net, which is a modification of the LASSO found to outperform the former while still having the advantage of a similar sparsity of representation (29), and (iv) selection based on Random Forests variable importance.

Analysis Pipeline

We imputed missing values using multiple imputations by chained equations (MICE) (30). Variables with 25% missing values or more were discarded from further analysis. All remaining variables used in this study are listed in **Supplementary Tables II, III**. In total, 53 baseline variables and 30 treatment variables were used as input for machine learning algorithms and automated variable selection methods for logistic regression.

The ordinal clinical (NIHSS) and radiological (clot burden and ASPECTS) scores were presented as continuous scores in all models to increase model efficiency, and we assumed linear trends underlying the ordinal scores.

We used nested cross-validation (CV), consisting of an outer and an inner CV loop. In the outer CV loop we used stratified CV with 100 repeated random splits resulting in a training set including 80% and a test set including 20% of all patients. Each training set was used as input for the inner CV loop, consisting of 10-fold CV (31, 32). In the inner CV loop we selected variables for the logistic regression models using the different variable selection methods, and optimized hyperparameters of all machine learning models. Hyperparameters are tuning parameters specific to each machine learning algorithm whose values have to be preset and cannot be directly learned from the data. We optimized hyperparameters with the random grid search module from Scikit-Learn (18). We selected those with highest area under the receiver operating characteristic (AUC) across all internal CV folds to find the best set of selected variables and hyperparameters. **Figure 1** shows a schematic representation of our nested CV methodology.

For all Random Forests models of both prediction settings we ranked variables by decreasing variable importance. For each variable we assessed the frequency of being among the 15 most

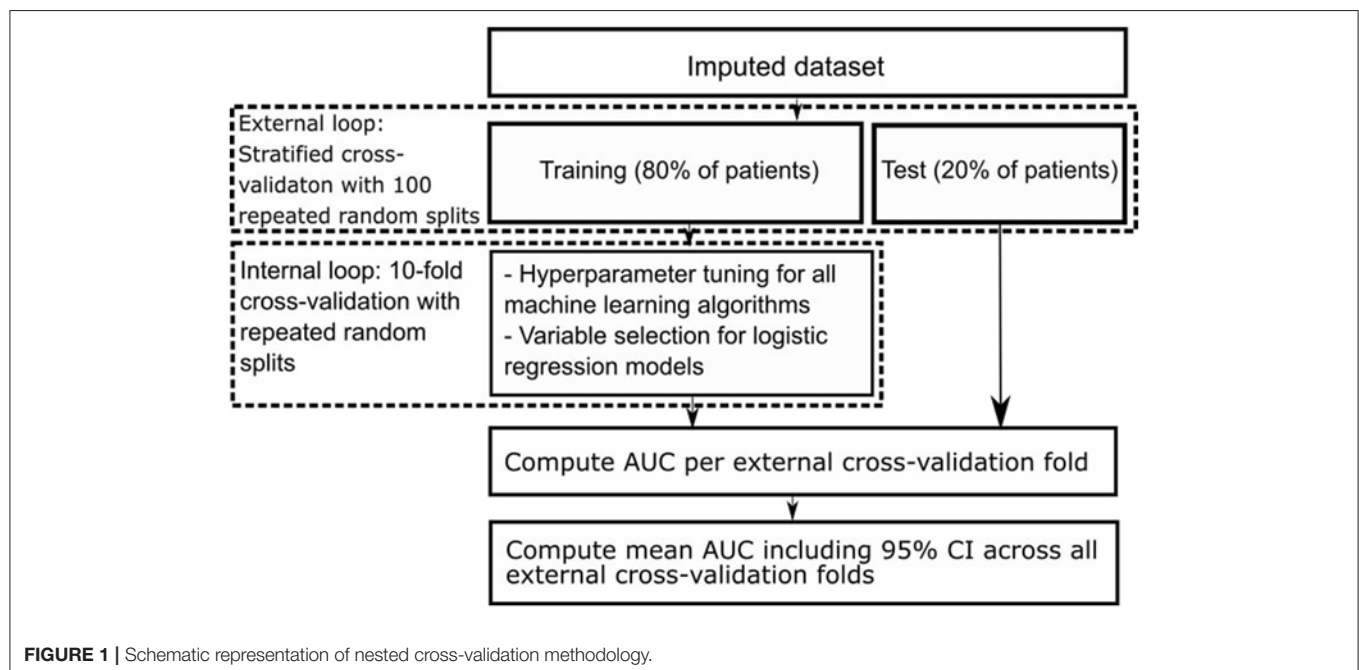


TABLE 1 | Baseline characteristics of participants.

Characteristics	All patients (<i>n</i> = 1,383)
Mean age \pm SD (years)	69.8 \pm 14.4
Men, <i>n</i> (%)	738 (53.5)
NIHSS score, median (IQR)*	16 (11–20)
Mean systolic blood pressure \pm SD (mm Hg)	150 \pm 25
MEDICAL HISTORY, <i>N</i> (%)	
Atrial fibrillation	411 (30.7)
Hypertension	697 (51.1)
Diabetes mellitus	235 (17.1)
Myocardial infarction	216 (15.9)
Peripheral artery disease	127 (9.4)
Ischaemic stroke	227 (16.5)
Hypercholesterolemia	411 (29.7)
Pre-stroke mRS > 2, <i>n</i> (%)	158 (11.6)
Smoking, <i>n</i> (%)	314 (22.9)
MEDICATION USE, <i>N</i> (%)	
DOAC**	35 (2.6)
Coumarine	179 (13.0)
Antiplatelet	461 (33.7)
Heparin	52 (3.8)
Blood pressure medication	707 (52.1)
Statin	490 (36.2)
Intravenous alteplase treatment, <i>n</i> (%)	1,054 (76.2)
ASPECTS, median (IQR)	9 (7–10)
Time from stroke onset to groin in minutes, median (IQR)	210 (160–270)
Collateral score \geq 2	764 (55)

*National Institutes of Health Stroke Scale score.

**Direct oral anticoagulant drugs.

important variables in a Random Forests model for each of the 100 external CV folds (Table 3).

Model Performance

We assessed model discrimination (the ability to differentiate between patients with good and poor outcome) with receiver operating characteristic (ROC) analyses. Because of our outer CV loop with 100 repeated random splits, we obtained 100 different AUCs from every model. We computed the average ROC-curve and mean AUC with 95% confidence intervals (CI) for all models. We evaluated differences between mean AUCs of the best performing machine learning model and best performing logistic regression model by computing the difference of means including the associated 95% CI.

RESULTS

Of the 1,627 patients registered between March 2014 and June 2016, we excluded 244 patients for this analysis because of age <18 (*n* = 2), posterior circulation stroke (*n* = 79), missing MR CLEAN trial center (*n* = 20), and missing mRS or post-mTICI (*n* = 143). Mean age was 69.8 years (SD \pm 14.4) and 738 (54%) of the 1,383 included patients were men. In total, 531 (38%) patients had good reperfusion after EVT and 525 (38%)

were functionally independent (mRS \leq 2) 3 months after stroke. Baseline characteristics are shown in Table 1.

Prediction of Good Reperfusion After EVT in Patients at Time of Admission

Discrimination between good and poor reperfusion of the best machine learning algorithm (Support Vector Machine, mean AUC: 0.55) and the best logistic regression model (using backward elimination, mean AUC: 0.57) was similar (difference of mean AUCs: 0.02; 95% CI: 0.01–0.03). Discrimination was poor for all models, with a mean AUCs ranging from 0.53 to 0.57 (Table 2). Variable selection using LASSO or Elastic Net was not possible likely because the signal-to-noise ratio was insufficient (18).

Prediction of 3-Months Functional Independence in Patients at Time of Admission

Discrimination of good functional outcome of the best machine learning algorithm (Super Learner, mean AUC: 0.79) and the best logistic regression model (using LASSO, mean AUC: 0.78) was similar (difference of mean AUCs: 0.01; 95% CI: 0.01–0.01).

Discrimination was moderate for all models, with a mean AUCs ranging from 0.77 to 0.79.

Prediction of 3-Months Functional Independence in Patients After Performance of EVT

Discrimination of good functional outcome of the best machine learning algorithm (Random Forests, mean AUC: 0.91) and the best logistic regression model (using prior knowledge, mean AUC: 0.90) was similar (difference of mean AUCs: 0.01; 95% CI: 0.00–0.01).

Discrimination was good for all models, with mean AUCs ranging from 0.88 to 0.91.

We performed a *post-hoc* analysis in patients with good reperfusion as defined by post-mTICI \geq 2b, predicting 3-months functional outcome both at time of admission and after performance of EVT. We did not find significant differences in performance between machine learning algorithms and logistic regression models in this patient subset (data not shown).

In Table 3 we show the top 15 variables based on the frequency of being among the 15 most important variables in a Random Forests model for each of the 100 external CV folds.

DISCUSSION

We found no difference in performance between best performing machine learning algorithms and best performing logistic regression models in predicting radiological or clinical outcome in stroke patients treated with EVT. For prediction of good reperfusion using variables available at baseline, all models showed a poor discriminative performance. This could indicate that reperfusion after EVT depends on characteristics not present in our variables available at time of admission, such as vascular anatomy or interventionalist related factors. Prediction

TABLE 2 | Discrimination of machine learning algorithms and logistic regression models across the various prediction settings.

Models, AUC (95% CI)*	Prediction setting (used variables: predicted outcome)		
	Baseline: post-mTICI	Baseline: mRS	All variables: mRS
Super learner	0.55 (0.54–0.56)	0.79 (0.79–0.80)	0.90 (0.90–0.91)
Random forests	0.55 (0.55–0.56)	0.79 (0.79–0.79)	0.91 (0.90–0.91)
Support vector machine	0.53 (0.53–0.54)	0.78 (0.77–0.78)	0.88 (0.88–0.89)
Neural network	0.53 (0.53–0.54)	0.77 (0.76–0.77)	0.88 (0.88–0.89)
LR: AUTOMATED SELECTION**			
Random forests	0.55 (0.55–0.56)	0.78 (0.78–0.78)	0.90 (0.90–0.90)
LASSO	NA‡	0.78 (0.78–0.79)	0.90 (0.89–0.90)
Elastic net	NA‡	0.77 (0.77–0.78)	0.89 (0.88–0.89)
Backward elimination	0.57 (0.57–0.58)	0.78 (0.77–0.78)	0.90 (0.89–0.90)
LR: prior knowledge‡	0.55 (0.55–0.58)	0.78 (0.78–0.79)	0.90 (0.90–0.90)

*Model discrimination is assessed by calculating mean Area Under the Curve (AUC) of the receiver operating characteristic across all outer cross-validation folds.

**Logistic regression using automated variable selection methods.

‡Variable selection not possible, likely due to insufficient signal-to-noise ratio.

‡Logistic regression using variables based on prior knowledge.

TABLE 3 | Variable importance of Random Forests for various prediction settings (used variables: predicted outcome).

Baseline: post-mTICI	Freq*	Baseline: mRS	Freq	All variables: mRS	Freq
RR systolic at admission	100	Age	100	NIHSS after 24–48 h	100
Duration stroke onset to groin	100	NIHSS at baseline	100	Delta NIHSS: follow-up minus baseline	100
RR diastolic at admission	100	Duration stroke onset to groin	100	Age	100
Thrombocyte count	100	Glasgow Coma Scale	100	NIHSS at baseline	100
Age	100	RR systolic at admission	100	Duration from onset to recanalization	100
Creatinine	100	CRP	100	Duration of procedure	100
CRP	100	Creatinine	100	Delta NIHSS \geq 4 points higher after EVT	100
NIHSS at baseline	100	Thrombocyte count	100	Duration stroke onset to groin	100
Clot burden score	100	RR diastolic at admission	100	Glasgow Coma Scale	100
Glasgow ComaScale	100	mRS prior to stroke	100	Creatinine	100
ASPECTS score at baseline	100	ASPECTS score at baseline	100	CRP	100
Glucose	100	Glucose	100	Thrombocyte count	100
Location: proximal M1**	74	Clot burden score	99	RR systolic at admission	100
Hyperdense artery sign on NCCT	50	Presence of leukoaraiosis	96	mRS prior to stroke	91
History of atrial fibrillation	32	Collateral score	77	RR diastolic at admission	93

NCCT, non-contrast CT; CRP, C-Reactive Protein; RR, blood pressure; NIHSS, National Institutes of Health Stroke Scale score.

*Frequency of being among the 15 most important variables in a Random Forests model for each of the 100 external CV folds.

**Location of intracranial occlusion on CTA.

of 3-months functional independence using variables known at baseline was moderate, predicting 3-months functional independence using baseline and treatment variables resulted in good performance.

We hypothesized that machine learning would outperform logistic regression models due to simultaneous assessment of a large number of variables, and more efficient processing of non-linear relations and interactions between them. Although a large number of variables (83 in total, see **Supplementary Tables II, III**) were available for analysis in the MR CLEAN Registry database, performance of best machine learning algorithms and best logistic regression models were similar. This could indicate that interactions and non-linear relationships in our dataset were of limited importance.

To interpret our results, several methodological limitations have to be considered. First, due to their great flexibility machine learning algorithms are prone to overfitting, which results in optimistic prediction performance. To account for overfitting we

used nested CV, which is considered to be an effective method for this aim (33). Second, our outer CV loop resulted in 100 AUCs per model leading to relatively small confidence intervals of mean AUCs. Although this increases the probability of statistically significant differences between mean AUCs of various models, the clinical relevance of these mean AUC differences is difficult to interpret. Because in our study mean AUC differences between models are minimal, clinical relevance of these differences is also negligible. Third, we used data from a registry. Registries might be prone to selection bias. However, we expect that selection bias in our study was minimal because the MR CLEAN Registry in principle covers all patients treated with EVT in the Netherlands. In addition, in all centers patients were treated according to national guidelines, and registration of treatment was a prerequisite for reimbursement (11).

Strong points of this study include the large sample size and standardized collection of patient data. Moreover, because of extensive hyperparameter tuning and state-of-the-art variable

selection methods, machine learning and logistic regression models were compared at their best performance. In several other studies that compared machine learning algorithms with only logistic regression methods using variables based on prior knowledge, machine learning outperformed logistic regression (6, 7, 34). Variable selection based on prior knowledge has the major drawback that predictive patterns in the data may be missed, as variable selection is strictly based on the literature and expert opinion (26). In our study however, logistic regression using variables based on prior knowledge performed similarly to logistic regression using automated variable selection methods.

The distinction between machine learning and “classical” regression methods is largely artificial. However, a clear distinction between various machine learning algorithms and logistic regression exists in terms of model transparency, which could be seen as the understanding of the mechanism by which the model works (35). Logistic regression has the advantage of transparency at the level of individual variable coefficients, since from these coefficients odds ratios can be derived. However, variable importances derived from the Random Forests algorithm also offer insight in the importance of individual variables for prediction performance (22). These variable importances take interaction between variables into account and have a similar interpretation for continuous and discrete variables, unlike odds ratios which constitute an effect per unit change of a predictor. Hence, Random Forests could be used as an efficient screening tool to pick up predictive patterns in the data that could potentially lead to further hypothesis-driven research. In **Table 3** we show the top 15 variables from either the baseline or baseline and treatment variable set, based on Random Forests variable importance. The majority of variables in **Table 3** do not overlap with the selection of variables based on prior knowledge, potentially providing researcher with additional information.

In this dataset we found no clinically relevant differences in prediction of reperfusion and 3-months functional independence across all models. However, since it is generally not known on beforehand which type of model will result in the best predictive performance in a new dataset, our methodology could be of importance in future studies. We present an analysis pipeline with both machine learning algorithms and logistic regression models including state-of-the-art variable selection methods. Assessing predictive performance of all models simultaneously enables the researcher to make the proper trade-off between predictive performance and model transparency. As our analysis pipeline is fully automated and input variables and outcome label can be altered at will, it is relatively easy to reuse in future studies. The Python code of our pipeline has been made publicly available

in an online repository (https://github.com/L-Ramos/MrClean_Machine_Learning).

ETHICS STATEMENT

The central medical ethics committee of the Erasmus Medical Centre Rotterdam, the Netherlands, evaluated the MR CLEAN Registry protocol and granted permission to carry out the study as a registry. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

HvO lead author, study design, analysis and interpretation, critical revision manuscript for important intellectual content. LR study design, analysis and interpretation, critical revision of manuscript for important intellectual content. AH, MvL, ES, HL, SO, KZ, EV, and HM study design, critical revision of manuscript for important intellectual content. NK data acquisition, critical revision of manuscript for important intellectual content. DD, MvW, IvdS, WS, and CM data acquisition, critical revision of manuscript for important intellectual content. MW supervisor of lead author, data acquisition, study design, critical revision of manuscript for important intellectual content.

FUNDING

MW was supported by a personal ZonMw VIDI grant (91717337), a Dekker Clinical Established Investigator Grant from the Netherlands Heart Foundation (2016T068) and a Fellowship grant from the Netherlands Brain Foundation (F2014(1)-22). LR was supported by an ITEA 3 grant (14003 Medolution). The MR CLEAN Registry is partially funded by unrestricted grants from Toegepast Wetenschappelijk Instituut voor Neuromodulatie, Twente University (Twin), Erasmus MC, AMC, and MUMC.

ACKNOWLEDGMENTS

See supplementary file for full details of the acknowledgements (MR CLEAN Registry acknowledgements).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2018.00784/full#supplementary-material>

REFERENCES

- Goyal M, Menon BK, van Zwam WH, Dippel DW, Mitchell PJ, Demchuk AM, et al. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *Lancet* (2016) 387:1723–31. doi: 10.1016/S0140-6736(16)00163-X
- Ferroni P, Zanzotto FM, Scarpato N, Riondino S, Nanni U, Roselli M, et al. Risk assessment for venous thromboembolism in chemotherapy-treated ambulatory cancer patients: a machine learning approach. *Med Decis Making* (2016) 37:234–42. doi: 10.1177/0272989X16662654
- Konerman MA, Zhang Y, Zhu J, Higgins PD, Lok AS, Waljee AK. Improvement of predictive models of risk of disease progression in chronic

- hepatitis c by incorporating longitudinal data. *Hepatology* (2015) 61:1832–41. doi: 10.1002/hep.27750
4. Lambin P, van Stiphout RG, Starmans MH, Rios-Velazquez E, Nalbantov G, Aerts HJ, et al. Predicting outcomes in radiation oncology–multifactorial decision support systems. *Nat Rev Clin Oncol.* (2013) 10:27–40. doi: 10.1038/nrclinonc.2012.196
 5. Mani S, Chen Y, Li X, Arlinghaus L, Chakravarthy AB, Abramson V, et al. Machine learning for predicting the response of breast cancer to neoadjuvant chemotherapy. *J Am Med Inform Assoc.* (2013) 20:688–95. doi: 10.1136/amiajnl-2012-001332
 6. Singal AG, Mukherjee A, Elmunzer BJ, Higgins PD, Lok AS, Zhu J, et al. Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma. *Am J Gastroenterol.* (2013) 108:1723–30. doi: 10.1038/ajg.2013.332
 7. Kop R, Hoogendoorn M, Teije AT, Buchner FL, Slottje P, Moons LM, et al. Predictive modeling of colorectal cancer using a dedicated pre-processing pipeline on routine electronic medical records. *Comput Biol. Med.* (2016) 76:30–8. doi: 10.1016/j.combiomed.2016.06.019
 8. 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
 9. van der Ploeg T, Nieboer D, Steyerberg EW. Modern modeling techniques had limited external validity in predicting mortality from traumatic brain injury. *J Clin Epidemiol.* (2016) 78:83–9. doi: 10.1016/j.jclinepi.2016.03.002
 10. Berkhemer OA, Fransen PS, Beumer D, van den Berg LA, Lingsma HF, Yoo AJ, et al. A randomized trial of intraarterial treatment for acute ischemic stroke. *N Engl J Med.* (2015) 372:11–20. doi: 10.1056/NEJMoa1411587
 11. Jansen IGH, Mulder M, Goldhoorn RB, investigators MCR. Endovascular treatment for acute ischaemic stroke in routine clinical practice: Prospective, observational cohort study (mr clean registry). *BMJ* (2018) 360:k949. doi: 10.1136/bmj.k949
 12. Pexman JH, Barber PA, Hill MD, Sevik RJ, Demchuk AM, Hudon ME, et al. Use of the alberta stroke program early ct score (aspects) for assessing ct scans in patients with acute stroke. *AJNR Am J Neuroradiol.* (2001) 22:1534–42.
 13. Tan IY, Demchuk AM, 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
 14. Khatri P, Neff J, Broderick JP, Khoury JC, Carrozzella J, Tomsick T, et al. Revascularization end points in stroke interventional trials: Recanalization versus reperfusion in ims-i. *Stroke* (2005) 36:2400–3. doi: 10.1161/01.STR.0000185698.45720.58
 15. Zaidat OO, Yoo AJ, Khatri P, Tomsick TA, von Kummer R, Saver JL, et al. Recommendations on angiographic revascularization grading standards for acute ischemic stroke: a consensus statement. *Stroke* (2013) 44:2650–63. doi: 10.1161/STROKEAHA.113.001972
 16. Fernandez-Delgado M, Barro CE, Baro S, Amorim D. Do we need hundreds of classifiers to solve real world classification problems? *J Machine Learn Res.* (2014) 15:3133–81. Available online at: <http://jmlr.org/papers/v15/delgado14a.html>
 17. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol.* (2007) 6:25. doi: 10.2202/1544-6115.1309
 18. *Scikit-Learn: Machine Learning in Python* (2017). Available online at: <http://scikit-learn.org/stable/> (Accessed October 17, 2017).
 19. Kreif N, Grieve R, Diaz I, Harrison D. Evaluation of the effect of a continuous treatment: a machine learning approach with an application to treatment for traumatic brain injury. *Health Econ.* (2015) 24:1213–28. doi: 10.1002/hec.3189
 20. Petersen ML, LeDell E, Schwab J, Sarovar V, Gross R, Reynolds N, et al. Super learner analysis of electronic adherence data improves viral prediction and may provide strategies for selective hiv rna monitoring. *J Acquir Immune Defic Syndr.* (2015) 69:109–18. doi: 10.1097/QAI.0000000000000548
 21. Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. *Artif Intell Med.* (2005) 34:113–27. doi: 10.1016/j.artmed.2004.07.002
 22. Breiman L. Random forests. *Machine Learn.* (2001) 45:5–32. doi: 10.1023/A:1010933404324
 23. Smola AJ, Schölkopf B. A tutorial on support vector regression. *Stat Comput.* (2004) 14:199–222. doi: 10.1023/B:STCO.0000035301.49549.88
 24. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* (2015) 521:436–44. doi: 10.1038/nature14539
 25. Bishop CM. Neural networks for pattern recognition. *J Am Stat Assoc.* (1995) 92:482.
 26. Walter S, Tiemeier H. Variable selection: current practice in epidemiological studies. *Eur J Epidemiol.* (2009) 24:733–6. doi: 10.1007/s10654-009-9411-2
 27. Venema E, Mulder M, Roozenbeek B, Broderick JP, Yeatts SD, Khatri P, et al. Selection of patients for intra-arterial treatment for acute ischaemic stroke: development and validation of a clinical decision tool in two randomised trials. *BMJ* (2017) 357:j1710. doi: 10.1136/bmj.j1710
 28. Tibshirani R. Regression shrinkage and selection via the lasso. *J Roy Stat Soc B Met.* (1996) 58:267–88.
 29. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Statist Soc B* (2005) 67:301–20. doi: 10.1111/j.1467-9868.2005.00503.x
 30. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: What is it and how does it work? *Int J Methods Psychiatr Res.* (2011) 20:40–9. doi: 10.1002/2Fmpr.329
 31. Ambrose C, McLachlan GJ. Selection bias in gene extraction on the basis of microarray gene-expression data. *Proc Natl Acad Sci USA.* (2002) 99:6562–6. doi: 10.1073/pnas.102102699
 32. Borra S, Di Ciaccio A. Measuring the prediction error: a comparison of cross-validation, bootstrap and covariance penalty methods. *Comput Stat Data Anal.* (2010) 54:2976–89. doi: 10.1016/j.csda.2010.03.004
 33. Krstajic DBL, Leahy DE, Thomas S. Cross-validation pitfalls when selecting and assessing regression and classification models. *J Cheminform.* (2014) 6:10. doi: 10.1186/1758-2946-6-10
 34. Decruyenaere A, Decruyenaere P, Peeters P, Vermassen F, Dhaene T, Couckuyt I. Prediction of delayed graft function after kidney transplantation: comparison between logistic regression and machine learning methods. *BMC Med Inform Decis Making* (2015) 15:83. doi: 10.1186/s12911-015-0206-y
 35. Lipton ZC. The mythos of model interpretability. In: *ICML Workshop Human Interpretability in Machine Learning (WHI 2016)*. New York, NY (2017).

Conflict of Interest Statement: DD reports grants from the Dutch Heart Foundation, AngioCare, Medtronic/Covidien/EV3, MEDAC/LAMEPRO, Penumbra, Top Medical/Concentric, and Stryker during conduct of the study; grants from Stryker European Operations BV, Medtronic, Dutch Heart Foundation, Brain Foundation Netherlands, The Netherlands Organisation for Health Research and Development, Health Holland Top Sector Life Sciences and Health, and consultation fees from Stryker, Bracco Imaging, and Servier, received by the Erasmus University Medical Centre, outside the submitted work. CM reports grants from TWIN, during the conduct of the study and grants from CVON/Dutch Heart Foundation, European Commission and from Stryker outside the submitted work (paid to institution), and is shareholder of Nico.lab. HM is founder and shareholder of Nico-lab.

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.

Copyright © 2018 van Os, Ramos, Hilbert, van Leeuwen, van Walderveen, Kruijt, Dippel, Steyerberg, van der Schaaf, Lingsma, Schonewille, Majoie, Olabarriaga, Zwinderman, Venema, Marquering, Wermer and the MR CLEAN Registry Investigators. 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.



Decision Criteria for Large Vessel Occlusion Using Transcranial Doppler Waveform Morphology

Samuel G. Thorpe^{1*}, Corey M. Thibeault¹, Nicolas Canac¹, Seth J. Wilk¹, Thomas Devlin² and Robert B. Hamilton¹

¹ Neural Analytics, Inc., Los Angeles, CA, United States, ² Erlanger Medical Center, Chattanooga, TN, United States

OPEN ACCESS

Edited by:

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

Reviewed by:

Qing Hao,
Icahn School of Medicine at Mount
Sinai, United States
Marek Czosnyka,
University of Cambridge,
United Kingdom

*Correspondence:

Samuel G. Thorpe
samuel.thorpe@neuralanalytics.com

Specialty section:

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

Received: 04 May 2018

Accepted: 21 September 2018

Published: 17 October 2018

Citation:

Thorpe SG, Thibeault CM, Canac N,
Wilk SJ, Devlin T and Hamilton RB
(2018) Decision Criteria for Large
Vessel Occlusion Using Transcranial
Doppler Waveform Morphology.
Front. Neurol. 9:847.
doi: 10.3389/fneur.2018.00847

Background: The current lack of effective tools for prehospital identification of Large Vessel Occlusion (LVO) represents a significant barrier to efficient triage of stroke patients and detriment to treatment efficacy. The validation of objective Transcranial Doppler (TCD) metrics for LVO detection could provide first responders with requisite tools for informing stroke transfer decisions, dramatically improving patient care.

Objective: To compare the diagnostic efficacy of two such candidate metrics: Velocity Asymmetry Index (VAI), which quantifies disparity of blood flow velocity across the cerebral hemispheres, and Velocity Curvature Index (VCI), a recently proposed TCD morphological biomarker. Additionally, we investigate a simple decision tree combining both metrics.

Methods: We retrospectively compare accuracy/sensitivity/specificity (ACC/SEN/SPE) of each method (relative to standard CT-Angiography) in detecting LVO in a population of 66 subjects presenting with stroke symptoms (33 with CTA-confirmed LVO), enrolled consecutively at Erlanger Southeast Regional Stroke Center in Chattanooga, TN.

Results: Individual VCI and VAI metrics demonstrated robust performance, with area under receiver operating characteristic curve (ROC-AUC) of 94% and 88%, respectively. Additionally, leave-one-out cross-validation at optimal identified thresholds resulted in 88% ACC (88% SEN) for VCI, vs. 79% ACC (76% SEN) for VAI. When combined, the resultant decision tree achieved 91% ACC (94% SEN).

Discussion: We conclude VCI to be superior to VAI for LVO detection, and provide evidence that simple decision criteria incorporating both metrics may further optimize.

Performance: Our results suggest that machine-learning approaches to TCD morphological analysis may soon enable robust prehospital LVO identification.

Registration: Was not required for this feasibility study.

Keywords: transcranial doppler, ultrasound, ischemic stroke, large vessel occlusion, decision tree, diagnostic biomarker

INTRODUCTION

Acute Ischemic Stroke (AIS) is the leading cause of long-term disability in the United States, and fifth leading cause of death (1). Current treatment for AIS includes the use of intravenous tissue Plasminogen Activator, and endovascular mechanical thrombectomy with a clot extraction or aspiration device. Although these therapies provide effective treatment options for Large Vessel Occlusion (LVO), their use is still limited by short time windows from symptom onset during which they are optimally effective (2–4). Indeed, only a small fraction of candidate patients who could ultimately benefit from endovascular treatment currently receive it (5). Early LVO identification is key to enabling rapid triage and transfer to comprehensive stroke centers, thus facilitating access to appropriate care. Computed Tomography Angiography (CTA) is the current gold standard for stroke diagnosis, but is limited to in-hospital use, or a low number of prohibitively expensive mobile stroke ambulances. Unfortunately, current prehospital stroke assessment scales lack reliability due to training requirements and low inherent accuracies (6, 7); causing delays in triage, transfer, and treatment.

Transcranial Doppler (TCD) ultrasound is a reliable diagnostic tool for assessing the presence and severity of LVO (8–11), which has the additional advantages of being non-invasive, inexpensive, and portable. Because it directly measures Cerebral Blood Flow Velocity (CBFV), TCD is a strong candidate technology for prehospital diagnosis and assessment of LVO. Indeed, bedside TCD examinations to detect stenosed and/or occluded intracranial vessels are routinely conducted as standard of care at many comprehensive stroke centers (12). Numerous studies have been published comparing TCD diagnosis of arterial LVO with CTA imaging; reporting sensitivity (SEN) and specificity (SPE) ranging between 79 and 98% depending on occlusion location (13–17). A limiting factor of these studies is the TCD operator's ability to locate and interpret the CBFV waveform. Such challenges have contributed to TCD being critically underutilized for stroke assessment.

For stroke diagnosis, specialized training is required to inspect flow velocity and morphology across multiple vessels and depths in both cerebral hemispheres. One of the most cited papers for stroke diagnosis using TCD was published by Demchuk et al. (10), which instructs the operator to categorize waveforms according to evidence of stroke-related pathology; namely dampened, blunted, minimal, or absent signal. A number of additional TCD exam methodologies with different criterion for LVO assessment have been published (15, 17). Typically, CBFV and power M-mode (PMD) waveforms are obtained for flow through the Middle, Anterior, and Posterior Cerebral Arteries (MCA, ACA, and PCA) in each cerebral hemisphere, as well as the Internal Carotid Arteries (ICA). Heuristic assessments are then made based on numerous features, including relative velocities, collateral flow, PMD resistance signatures, and the presence of pathological waveform morphologies.

Assessment of these categories relies heavily on qualitative interpretation by specialists which cannot be replicated by less formally trained personnel. The challenge of moving LVO detection to the prehospital setting thereby obviates the need

for objective metrics by which first responders might reliably evaluate TCD signals. An intuitive first candidate for such a metric is CBFV asymmetry, as it is already well established that velocity disparity, both between homologous vessels in opposite hemispheres (10, 18) as well as adjacent vessels in an occluded hemisphere (10, 15), can be indicative of vascular occlusion. One promising metric for LVO detection based on velocity disparity was published by (15); showing area under the Receiver Operating Characteristic curve (ROC-AUC) of 92.6%. However, their metric also relied on PMD resistance signatures as a predictive feature, which were not objectively computed, and was limited in application to occlusions of the MCA.

However intuitive, assessment of velocity asymmetry also comes with the inherent concern that velocity estimates in adjacent vessels and opposite hemispheres can be greatly impacted by anatomical variability (incident angle of the vessel and ultrasound beam), as well as by intrarater measurement inconsistency (19). Moreover, reliance on mean velocity disparity inherently discards the morphological information currently utilized in routine stroke assessment protocols. Such assessments incorporate morphological information explicitly, but in a subjective manner which requires expert interpretation. However, a number of recent studies have observed morphological changes associated with various medical conditions which are both objectively quantifiable (20–22), and independent of significant changes in mean velocity (23). Pulsatility Index is an example of a well known and widely clinically utilized morphological TCD variable (24); one which evidence suggests is not useful for detecting LVO (15).

Toward the aim of quantifying TCD waveform morphology for the purpose of LVO identification, we have recently proposed a diagnostic biomarker termed Velocity Curvature Index (VCI) (25, 26). Mathematically, it is a straightforward extension of the concept of graph curvature; one which is sensitive to the morphological structure of the pathological waveforms first described by Demchuk et al. (10). In this work we retrospectively compare the diagnostic utility of VCI to that of a standard Velocity Asymmetry Index (VAI) for the detection of LVO in a clinical subject population collected in-hospital. Additionally, we evaluate a simple decision tree classifier designed to incorporate complimentary information from both metrics. Decision trees are a commonly used diagnostic methodology in several areas of medicine (27, 28), which have previously been used with TCD variables to screen for cervical vascular injury (29). To these ends, we employ leave-one-out cross validation and subsequent sensitivity analysis to assess performance as diagnostic thresholds are weighted toward detection of true positives. Our goal is the validation of TCD-based decision criteria which are objective, intuitive, and easily communicated; allowing physicians and first responders alike a common language for LVO assessment.

MATERIALS AND METHODS

Subject Examination and Imaging

We acquired TCD waveforms from two clinical populations enrolled consecutively at Erlanger Health System's Southeast

Regional Stroke Center in Chattanooga, TN, from October 2016 through September 2017. The LVO group was comprised of patients with CTA-confirmed occlusion of the M1/M2 branches of the MCA and/or ICA (proximal extracranial or terminal intracranial segments); these occlusion locations being selected since they are the large cerebral arteries most amenable to neurovascular intervention. The In-Hospital Control group (IHC) consisted of patients who arrived at the hospital presenting with stroke symptoms, but were later confirmed negative for LVO by imaging. Patients in both groups received TCD examinations in addition to standard care (patient history, monitoring, pharmaceuticals, and CT/A/perfusion imaging). Patients were eligible for enrollment if a complete TCD exam was acquired within 4 h of CTA, and none of the following exclusions applied: (1) Head CT findings consistent with acute primary intracranial hemorrhage, (2) Hemodynamically unstable patients requiring pharmacological support for hypotension; (3) Subjects who underwent partial or full craniotomy; (4) Additional intracranial pathologies present (tumor, hydrocephalus, etc.); (5) Anticipated insufficient time to acquire a complete set of scan as described by the protocol; (6) Significant hemodynamic pharmacological agent (cocaine, amphetamine, etc.); (7) Subjects who are under arrest for a felony.

CTA examinations were performed using a GE Lightspeed VCT 64-section multidetector scanner (GE Healthcare, Milwaukee, WI) with a slice thickness of 0.625 mm, and bolus injection of 70–150 mL of Omnipaque 350 (GE Healthcare, Milwaukee, WI) contrast material (4.0 mL/s). CTA images were reformatted in the coronal and sagittal plane, and 10-mm maximum intensity projection reconstructions were rendered and sent to PACS for review. Occlusion location was determined by the radiologist on call, who was blinded to any results of the TCD examination.

Complete TCD examinations included (at minimum) one pair of left/right MCA scans at depths between 45 and 60 mm, each containing 15 or more individual beat waveforms (see **Figure 1**). Subjects for whom complete examinations were not obtained were counted as missing/indeterminate data and excluded from analysis. Data was acquired during available intervals between patient testing/treatment, and in no way impacted patient care. The TCD technician was often present during initial evaluation of the subject, and so was not entirely blinded to all clinical information or imaging results. Sample size was not predetermined for this feasibility study, being established pragmatically as the maximum number of subjects attainable in the allotted time frame. Experiment protocols were approved by University of Tennessee College of Medicine Institutional Review Board (ID: 16-097). Reporting in this manuscript is structured in accordance with the Standards for Reporting of Diagnostic Accuracy Studies [(30); see **Appendix** for detailed criteria checklist].

Waveform Processing and Feature Extraction

Recording

TCD scans were acquired by a trained technician using 2 MHz hand-held ultrasound probes. CBFV signals associated with

the left/right MCA were identified via insonation through the transtemporal window. CBFV envelopes were digitally sampled at 125 Hz and recorded throughout the entire exam. Once the CBFV signal was identified and optimized at a specific depth, waveform recordings were then made in 30-s intervals. The technician was instructed to obtain recordings for as many depths as possible between 45 and 60 mm in both the left/right cerebral hemispheres. Start times for each interval were marked by the technician using a custom event remote, which prompted a 30 s countdown to a corresponding stop event. TCD envelopes and event information were aligned using custom software (Python 2.7; Kivy 1.9) running on Windows 10.

Processing

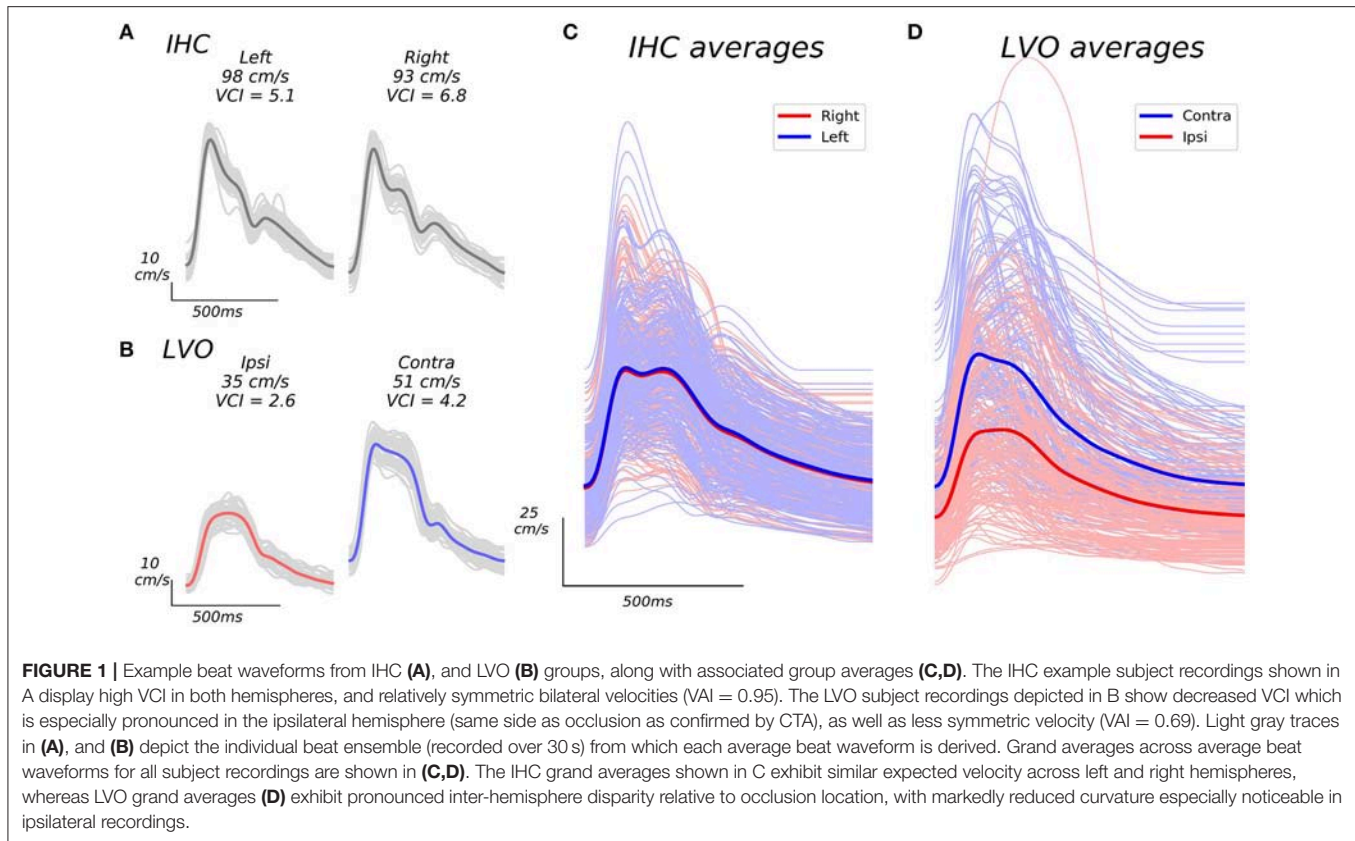
Average beat waveforms from each recorded depth interval were extracted using a combination automated beat identification algorithm with manual checking/editing. In this procedure, individual beats within each interval were first identified automatically using an internally developed beat extraction tool, and displayed to the user for manual confirmation/editing. Detected beats which lacked clear pulsatile structure and/or deviated anomalously from the group average (usually due to probe displacement during recording), were excluded. The remaining beats were then aligned and averaged, resulting in a single representative beat waveform for each recorded depth interval (see examples in **Figure 1**).

Since Doppler velocities scale with the cosine of the incident angle between the ultrasound beam and underlying blood flow (31), TCD waveforms for a given vessel with the highest measured velocities are assumed to most accurately reflect reality. In line with this reasoning, for each subject we selected a single bilateral (left/right) pair of average beat waveforms for analysis consisting of those with maximal mean velocity across all recorded depths for each hemisphere.

VCI

Curvature is a well-defined mathematical property of space curves which quantifies the degree to which a curve deviates from being “straight” at a given point. VCI is an application of the curvature metric specific to TCD which quantifies the degree to which a beat is blunted and/or dampened. Since curvature is a nonlinear function sensitive to small inflections associated with high frequency noise, we first smooth the average beat waveform via convolution with the Hanning window (9 ms). Moreover, we elect to consider only curvature associated with the beat systolic complex, where the signal-to-noise ratio is presumably greatest. The systolic complex, or “beat canopy,” comprises the proportion of the beat with the highest velocities and richest morphological structure.

To compute VCI for a given TCD beat waveform, curvature is first computed for each time point (t_i) of the smoothed beat (denoted $x(t_i)$ below) via the discretized equation for graph curvature (equation 1) expressed in terms of finite differences. Δ and δ^2 in equation 1 denote the first order backward (equation 2) and second order central (equation 3) finite difference equations. VCI, defined by equation 5, is computed as the sum of curvature taken over all individual time points comprising the beat canopy (C). The beat canopy is defined in equation 4



as the set of time points wherein velocity exceeds one quarter of its total diastolic-systolic range (t_d , and t_s denoting the time points corresponding to diastolic minimum and systolic max, respectively). Since the hypothesized effect of occlusion on the TCD waveform is to lower VCI in the occluded vessel, when assessing a bilateral pair of waveforms we take VCI as the minimum computed for each member of the pair. VCI is a positive metric which, in principle, has no upper bound, but is clearly bounded in practice (see **Figure 2A**).

$$k(t_i) = \frac{|\delta^2[x](t_i)|}{(1 + (\Delta[x](t_i))^2)^{\frac{3}{2}}} \quad (1)$$

$$\Delta[x](t_i) = x(t_i) - x(t_{i-1}) \quad (2)$$

$$\delta^2[x](t_i) = x(t_{i+1}) - 2x(t_i) + x(t_{i-1}) \quad (3)$$

$$C = \left\{ i : x(t_i) \geq x(t_d) + \frac{x(t_s) - x(t_d)}{4} \right\} \quad (4)$$

$$VCI = \sum_{i \in C} k(t_i) \quad (5)$$

VAI

Velocity Asymmetry Index is a metric which quantifies the degree to which average CBFV observed for a vessel in a given cerebral hemisphere differ from that observed in the corresponding vessel in the opposite hemisphere (see LVO example in **Figure 1**). The hypothesis that CBFV in an occluded vessel may be lower than that of the corresponding unaffected hemisphere is intuitive, but

also supported by previous work (18). For a bilateral pair of left/right average beat waveforms, denoted $x_L(t)$ (with N_L total time points), and $x_R(t)$ (with N_R time points) in equations 6 and 7, respectively, VAI (defined in equation 8) is computed as the minimum average velocity across hemispheres divided by the corresponding maximum. By definition, VAI is a positive definite metric bounded on the closed interval [0, 1].

$$\mu_L = \frac{1}{N_L} \sum_{i=1}^{N_L} x_L(t_i) \quad (6)$$

$$\mu_R = \frac{1}{N_R} \sum_{i=1}^{N_R} x_R(t_i) \quad (7)$$

$$VAI = \frac{\min(\{\mu_L, \mu_R\})}{\max(\{\mu_L, \mu_R\})} \quad (8)$$

Feature Statistical Analysis

For both VCI and VAI, resultant group samples were not normally distributed. Accordingly, we tested significance of observed differences in group distributions for each feature using the Mann-Whitney U test statistic. The U-statistic is directly proportional to the common language effect size (by a factor of the product of the group sample sizes under comparison), which is equivalent to the area under the Receiver Operating Characteristic curve (ROC-AUC). Additionally, we computed ROC curves detailing separability of subject group distributions (LVO vs. IHC) for each feature. Specifically, the ROC curves give

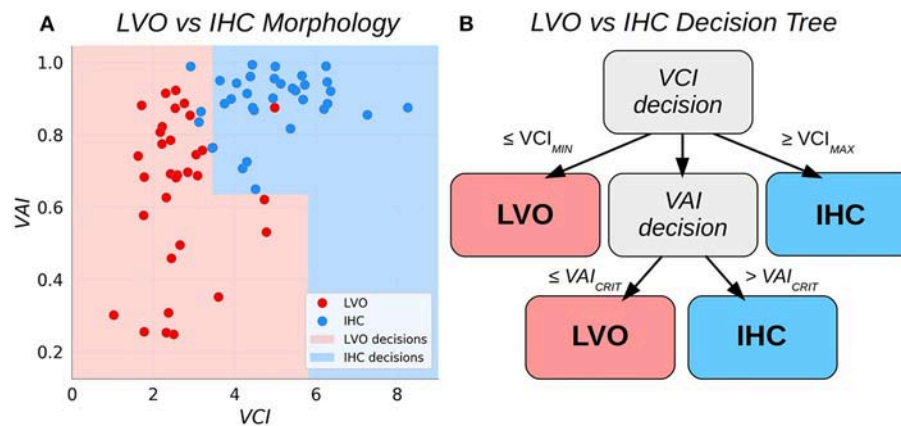


FIGURE 2 | Scatter plots relating the Velocity Asymmetry Index and Velocity Curvature Index (A), along with a decision tree partitioning the two-dimensional feature space (B).

True Positive Rate (TPR) as a function of False Positive Rate (FPR) for each possible feature threshold. To quantify binary classification performance, we computed SEN, SPE, and accuracy (ACC) at the thresholds which maximized Youden's J-statistic (32):

$$J = \text{TPR} - \text{FPR} \quad (9)$$

Additionally, we bootstrapped 95% confidence intervals (CI) on group means for each feature; accomplished by iteratively resampling each group distribution with replacement (10,000 iterations), and each time taking the mean (CI given by the 2.5th and 97.5th percentile of the resultant empirical distribution) (33). Statistical tests and ROC curves were computed using standard python libraries; SciPy version 1.0, and Scikit-learn version 0.19.1 (34), respectively.

LVO Classification

Decision Tree Classifier

We sought to combine VCI and VAI into a single binary classifier with simple and intuitive decision criteria. The approach adopted here is to augment the bilateral VCI assessment such that subject pairs with VCI less than some low critical threshold are classified as LVO, whereas pairs exceeding some high critical threshold are classified as IHC. Pairs observed to fall between these thresholds are deemed uncertain and decided then based on VAI. This procedure effectively partitions the 2D decision space into two subspaces with piece-wise linear boundaries (see **Figure 2**). The procedure for fitting the thresholds based on a given set of training data were as follows. First, the low VCI threshold (VCI_{MIN}) was fit using all the training data. Subjects with paired VCI below the threshold were predicted as LVO, and set aside. Next, the high VCI threshold (VCI_{MAX}) was fit from the remaining data. Subjects with supra-threshold VCI were predicted as IHC, and set aside. Finally, the remaining data was used to fit the VAI threshold (VAI_{CRIT}), with sub/supra-threshold subjects predicted as LVO/IHC, respectively. Specifically, each of

the three thresholds (VCI_{MIN} , VCI_{MAX} , and VAI_{CRIT}) were fit as the threshold which maximized Youden's J statistic for the data applicable to each decision (see also the sensitivity weighted J-statistic used to determine thresholds for sensitivity analysis in section Sensitivity Analysis).

Model Cross-Validation

For both individual diagnostic metrics (VAI and VCI), as well as the decision tree model, leave-one-out cross-validation was performed to assess generalization of classification performance near decision boundaries. In this iterative procedure, a single subject is removed from the pooled data, and the predictive model is derived via training (i.e., threshold optimization) with the remaining subjects. The excluded subject's data is then predicted using the trained model, and this procedure is repeated for each subject to obtain a complete set of cross-validated predictions from which to assess binary classification performance metrics (SEN/SPE/ACC).

Sensitivity Analysis

For many clinical problems, including LVO detection, diagnostic net benefit is optimized by increased detection of true positives at the cost of missing true negatives (i.e., SEN is prioritized over SPE). However, poor diagnostics for which SEN is maximized often have no clinical value, as SPE may plummet and overall ACC approaches chance. In order to assess how performance characteristics of our classifiers changed when priority is weighted toward increased sensitivity, we performed a sensitivity analysis wherein we iterated cross-validation of each model, each time incrementing classification thresholds away from the starting point of Youden's maximal J, toward increasing sensitivity. The procedure can be conceptualized as simply adjusting the thresholds up along the associated ROC curves toward increased true positive rate. Mathematically, this was accomplished by introducing a parameter (α) to modify the formula for Youden's J statistic as given in formula 9, and choosing thresholds to maximize the resultant index (J_α). Classifier performance was assessed by cross-validating each

model with α ranging from 0.5 (threshold equivalent to Youden's maximal J) to 1 (maximal sensitivity) in steps of 0.01.

$$J_{\alpha} = \alpha \text{TPR} - (1 - \alpha) \text{FPR} \quad (10)$$

RESULTS

Subject Demographics

The current analyses included 33 LVO subjects (16 female), and 33 IHC subjects (13 female), with average ages of 66.9 years (SD = 15.7), and 56.4 years (SD = 16.3), respectively. A total of 156 subject screenings were attempted at Erlanger Medical Center, of which 68 were excluded due to screening failures (time required to complete exam, subject compliance, etc.). Of subjects with sufficient initial screenings, 50 and 38 were initially enrolled in the LVO and IHC groups, respectively. Of the LVO subjects, 3 were discontinued (subject either expressed desire to discontinue, or was transferred or died before enrollment could be completed). An additional 14 LVO and 5 IHC subjects were subsequently excluded due to the presence of disqualifying criteria unknown at the time of enrollment. In the LVO group, there were 20 subjects with M1 occlusions, 3 with M2 occlusions, and 8 with ICA occlusions. An additional subject had dual occlusions of both the M1 and ICA (same hemisphere), and another additional subject had bilateral occlusions of both ICA, in addition to an M2 occlusion. TCD exams were performed an average of 33 min (SD = 20) post-CTA for IHC subjects compared to 43 min (SD = 44) post-CTA for LVO subjects (difference not significant between groups; $t = -1.15$, $p = 0.26$). At time of admittance, LVO subjects were more physiologically/cognitively impaired as assessed by National Institute of Health Stroke Scale (NIHSS), with average scores of 16.8 (SD = 6.6), compared to 1.9 (SD = 2) for IHC (differences strongly significant between groups; $t = -12.2$; $p < 0.001$). No adverse events were reported for any subjects as a result of TCD examination.

Individual Metric Statistical Comparisons

Figure 3 shows VAI and VCI metric distributions for LVO and IHC groups (A, C), and associated ROC curves depicting separability of group metrics (B, D). VAI means were greater for IHC subjects (0.89, CI = 0.86–0.92) relative to LVO (0.65, CI = 0.58–0.72). Associated ROC-AUC was observed at 88.4%, with significant group distribution differences confirmed by Mann-Whitney testing ($p < 0.001$). Similarly, VCI means were greater for IHC subjects (4.95, CI = 4.55–5.36) relative to LVO (2.66, CI = 2.38–2.97); with associated ROC-AUC observed at 94.2%. Significant group distribution differences were likewise confirmed by Mann-Whitney testing ($p < 0.001$). SEN/SPE/ACC at thresholds corresponding to Youden's maximal J are detailed in **Table 1** for both metrics.

Sensitivity Analysis

Figure 4 shows SEN, SPE, and ACC dependence on the alpha weighting parameter using leave-one-out cross-validation for each classifier. By definition, the sensitivity of each classifier increases with increased alpha. Interestingly, performance

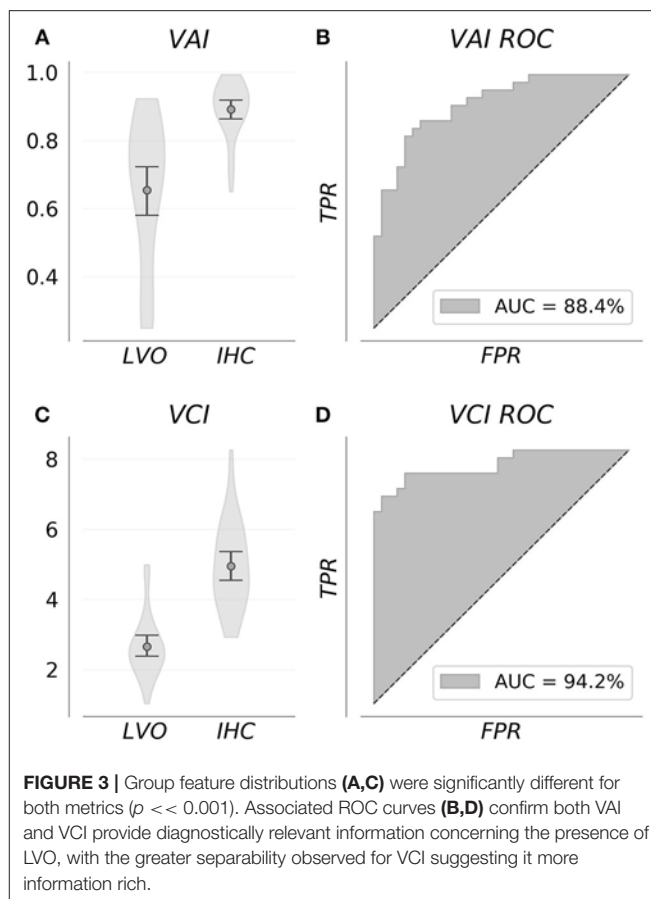


TABLE 1 | Descriptive information and performance indicators comparing LVO and IHC groups for VAI and VCI metrics at thresholds corresponding to Youden's maximal J statistic.

Metric	Mean (LVO, IHC)	CI (LVO, IHC)	AUC	SEN	SPE	ACC
LVO VS. IHC GROUP METRICS						
VAI	0.65, 0.89	(0.58–0.72), (0.86–0.92)	0.88	0.82	0.82	0.82
VCI	2.66, 4.95	(2.38–2.97), (4.55–5.36)	0.94	0.91	0.88	0.89

In column 3, CI refers to 95 percent Confidence Intervals around the respective means given in column 2.

indicator trajectories vary substantially between the VAI classifier and the other two (VCI and decision tree). For VAI, observed SPE and ACC are optimal near the maximal J (alpha = 0.5), and rapidly degrade with increased prioritization of SEN corresponding to alpha greater than 0.6. In contrast, for VCI and the decision tree, a stable range of alpha exists away from the maximal J for which SPE, and ACC are optimized and all performance indicators are stable. This range, roughly 0.6–0.8, is indicated in gray in **Figure 4**. Above this range, SEN increases for VCI and decision tree are accompanied by precipitous decreases in ACC and SPE. Together these results suggest a natural optimal

alpha range for prioritization of sensitivity for the VCI and decision tree classifiers.

Figure 5 shows cross-validated confusion matrices for each classifier with alpha specified at 0.6, which represents the start of the optimal range for the VCI and decision tree classifiers, and the tail end of the optimal range for the VAI classifier. For VAI, we observed an overall accuracy of 79%, with SEN/SPE of 76%, and 82%, respectively. For VCI, we observed an overall accuracy of 88%, with SEN/SPE of 88%, and 88%, respectively. Finally, for the decision tree we observed an overall accuracy of 91%, with SEN/SPE of 94%, and 88%, respectively. Together, these results demonstrate the superiority of the VCI classifier relative to VAI. However, within the framework of the decision tree, VAI helped to increase SEN of LVO identification relative to VCI alone. **Figure 5** results are summarized in **Table 2**.

DISCUSSION

To our knowledge, this work represents the first published LVO decision criteria based on TCD variables which can be

computed algorithmically and interpreted objectively. Results from all classifiers fall into the range observed in previous TCD studies using complex multi-vessel recording protocols (13, 17). Moreover, previous studies using predictive variables amenable to ROC analysis have not been subject to cross-validation in the manner we have performed here. Most importantly, these metrics substantially outperform stroke severity scales currently in prehospital use, which recent reviews suggest are unlikely to predict LVO with both high sensitivity and specificity (6, 7). Specifically, (6) published performance indicators for 5 such stroke assessment scales (RACE, 3ISS, LAMS, CPSSS, and PASS); reporting ACC and SEN capped at 74 and 64% across all scales. A sense of the potential for improvement upon these numbers can be gleaned from our simple decision tree which achieved cross-validated ACC and SEN exceeding 90%.

Previous work assessing TCD efficacy in detecting LVO align well with our current results. Tsivgoulis et al. (17) detected occlusions and stenoses based on the presence of the pathological waveforms described by Demchuk et al. (10) with SEN/SPE of 79 and 94%, respectively. A similar exam protocol was

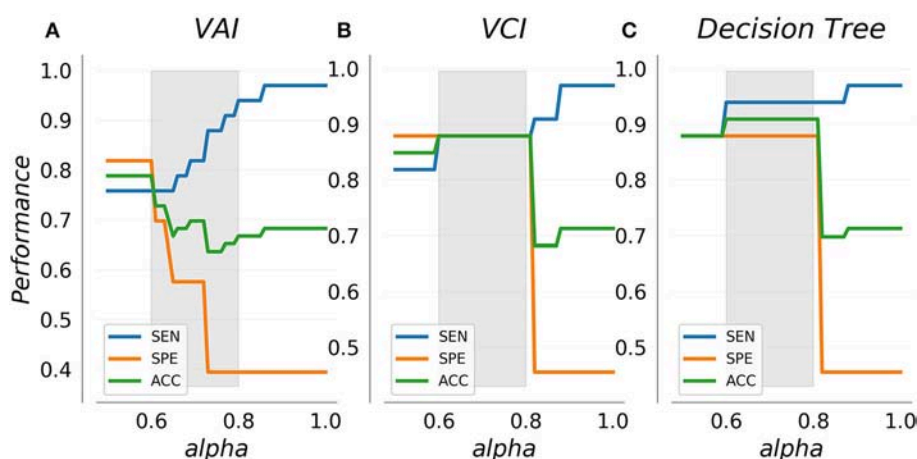


FIGURE 4 | Cross-validated performance indicators for the VAI (A) and VCI (B) metrics as well as combined decision tree classifier (C). Sensitivity increases with the alpha weighting parameter as specificity decreases. VAI specificity decreases rapidly with increased sensitivity, whereas VCI and the decision tree display a stable range (indicated in gray) wherein specificity and accuracy are optimal.

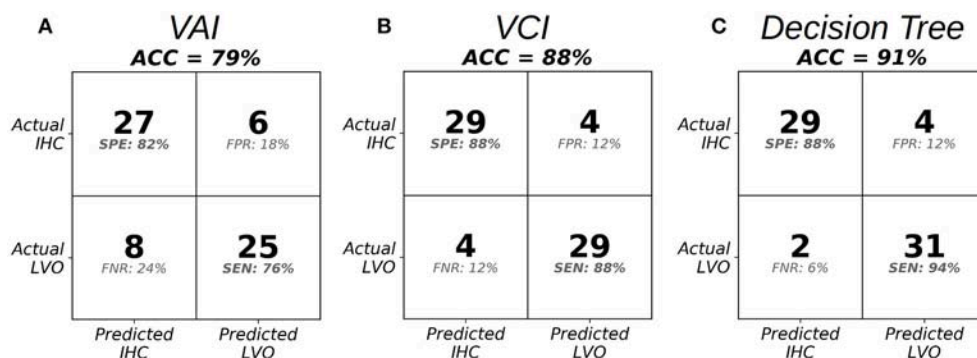


FIGURE 5 | Confusion matrices at the specified threshold corresponding to alpha = 0.6 in the optimal range for VAI (A), VCI (B), and the Decision Tree Classifier (C).

TABLE 2 | Performance indicators for leave-one-out cross-validated classifiers comparing LVO and IHC groups, with classification thresholds specified at alpha equal to 0.6.

Classifier	SEN	SPE	ACC
CROSS-VALIDATED LVO CLASSIFIER PERFORMANCE METRICS			
VAI	0.76	0.82	0.79
VCI	0.88	0.88	0.88
Decision tree	0.94	0.88	0.91

used by Brunser et al. (13), but with additional power M-mode criteria also considered, which achieved SEN/SPE of 90 and 94% detecting occlusion of any non-specific artery. The sensitivity and specificity observed for our metrics compare reasonably well with these previous results, which is especially encouraging considering our features were extracted from bilateral examination of a single vessel. Reliance on the MCA signals is pragmatic, but also represents a notable opportunity for improvement upon our current experimental paradigm. The MCA possesses the longest expected segment of probeable depths, and is thus most easily insonated and reliably located, but there is clearly more diagnostic information available in other vessels in the form of relative morphology and collateral flow which may improve performance in future experiments.

It is notable that the performance indicator curves we observed for VCI and the decision tree were extremely similar. This is a natural consequence of the decision criteria, which dictates that VAI is used only to decide “uncertain” subjects, thus serving mainly to improve upon VCI sensitivity to the degree allowable by the training data. Further work is needed to determine if there are specific occlusion types or patient demographics for which each metric works particularly better or worse, which should help to optimize decision criterion. In this configuration, VCI is doing the “heavy lifting” in our decision tree. It is an effective diagnostic metric because it is sensitive to the morphological structure of Demchuk’s minimal, blunted, and dampened flows (10). The *blunted* waveform, for example, possesses an inherently smooth (i.e., low curvature) systolic complex, and is thus readily quantified by VCI. Forthcoming work will analyze in specific detail the manner in which VCI captures the subtle morphological variations associated with pathological LVO waveforms.

Some limitations of our study and directions for future work should be noted. The primary factor which could potentially inhibit generalisability of our results is the small sample size of our study. Much further data is required to refine estimates of morphological variability inherent in LVO and clinical control patient populations. Additionally, numerous important subgroup analyses are required to determine if/how TCD morphology depends on demographic and clinical factors (age, gender, occlusion type, etc). It should also be noted that the TCD technician’s exposure to patient clinical information represents a potential source of bias which should be mitigated in future work by more thorough blinding. Finally, the relationships

between curvature, heart rate, and stroke pathology require further investigation. In our sample, LVO subjects had an average heart rate of 87.9 (SD = 22.2) beats per minute (BPM), vs. 71.2 (SD = 11) BPM for IHC; which was significantly different between groups ($t = 3.8$, $p < 0.001$). However, there was no significant correlation between heart rate and curvature within either subject group ($r = -0.006$, $p = 0.9$ for IHC; $r = -0.29$, $p = 0.1$ for LVO). Moreover, when we use heart rate itself as a predictor to distinguish between groups, we observe an AUC of 73%, considerably underperforming both the VAI (88% AUC), and VCI (94% AUC) metrics. So, while it is possible that heart rate accounts for some degree of variance between groups, it remains unclear whether the effect is causal or correlative. It is certainly plausible that the lack of blood supply characteristic of LVO causes heart rate to increase; meaning heart rate is effectively part of the diagnostic signal. Further work is needed to establish how elevated heart rate might affect VCI when occlusion is not present.

One strength of the current approach is simplicity of data acquisition and communicability of decision criteria. However, as the amount of data we acquire increases, the subtle variations we will ultimately wish to detect will undoubtedly require more complicated and abstracted models. Given the efficacy of initial results, the road map to such models is encouraging. Incorporation of depth dependent and inter-hemispheric morphological dynamics across multiple vessels might ultimately allow precise prehospital localization of occlusion, and distinction between occlusive and hemorrhagic strokes. Moreover, digital rendering of individual subject vasculature, currently possible with emergent technologies such as 3D MRA time of flight imaging, could facilitate ultra rapid mapping and scanning across multiple vessels, as well as development of anatomically realistic mathematical models of cerebral hemodynamics. Such models could dramatically increase our understanding of the fluid mechanics involved in vascular occlusion, and the associated impact on morphological biomarkers like VCI.

CONCLUSIONS

Our results suggest both VAI and VCI contain robust information concerning the presence of intracranial occlusion. Both are objective and real-time computable, and thus represent promising candidate metrics for the development of TCD-based prehospital LVO diagnostic systems. The feature distributions and classification performance indicators we observed suggest VCI may be superior to VAI for LVO detection, but even simple approaches to feature combination such as the decision tree analyzed herein may serve to further increase diagnostic accuracy. More data is needed to determine how well these decision criteria scale and generalize to a wider range of subject demographics and pathologies. Nonetheless, these results demonstrate the foundational potential for machine-learning approaches to TCD morphological analysis to enable faster and more widespread access to life saving medical intervention in the future.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of Declaration of Helsinki and The University of Tennessee College of Medicine Institutional Review Board with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the University of Tennessee College of Medicine Institutional Review Board (ID: 16-097).

AUTHOR CONTRIBUTIONS

ST had access to all data used in the study and takes responsibility for data integrity and accuracy of data analysis. SW, RH, and TD study concept and design. ST and CT analysis. ST, CT, and NC interpretation of data. ST drafting of the manuscript. CT, NC, SW, RH, and TD critical revision of the manuscript for

important intellectual content. ST statistical analysis. CT, NC, RH, SW, and TD technical or material support. RH and TD study supervision.

ACKNOWLEDGMENTS

We acknowledge and thank Leonardo Martinez-Torres, James LeVangie, and Professor Yaser Abu-Mostafa for their profound contributions to this work. In addition, the supporting efforts of Ben Delay, Michael O'Brien, Mina Ranjbaran, Shankar Radhakrishnan, Leo Petrossian, and Kian Jaleleddini were greatly appreciated.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2018.00847/full#supplementary-material>

REFERENCES

- Benjamin EJ, Blaha MJ, Chiuve SE, Cushman M, Das SR, Deo R, et al. Heart disease and stroke statistics-2017 update: a report from the American Heart Association. *Circulation* (2017) 135:e146–603. doi: 10.1161/CIR.0000000000000485
- Goyal M, Jadhav AP, Bonafe A, Diener H, Mendes Pereira V, Levy E, et al. Analysis of workflow and time to treatment and the effects on outcome in endovascular treatment of acute ischemic stroke: results from the swift prime randomized controlled trial. *Radiology* (2016) 279:888–97. doi: 10.1148/radiol.2016160204
- Saver JL, Goyal M, van der Lugt A, Menon BK, Majoie CB, Dippel DW, et al. Time to treatment with endovascular thrombectomy and outcomes from ischemic stroke: a meta-analysis. *JAMA* (2016) 316:1279. doi: 10.1001/jama.2016.13647
- Sheth SA, Jahan R, Gralla J, Pereira VM, Nogueira RG, Levy EI, et al. Time to endovascular reperfusion and degree of disability in acute stroke. *Ann. Neurol.* (2015) 78:584–93. doi: 10.1002/ana.24474
- Rai AT, Seldon AE, Boo S, Link PS, Domico JR, Tarabishy AR, et al. A population-based incidence of acute large vessel occlusions and thrombectomy eligible patients indicates significant potential for growth of endovascular stroke therapy in the USA. *J Neurointerv Surg.* (2017) 9:722–6. doi: 10.1136/neurintsurg-2016-012515
- Hastrup S, Damgaard D, Andersen G. Prehospital acute stroke severity scale to predict large artery occlusion: design and comparison with other scales. *Stroke* (2016) 47:1772–6. doi: 10.1161/strokeaha.115.012482
- Smith EE, Kent DM, Bulsara KR, Leung LY, Lichtman JH, Reeves MJ, et al. Accuracy of prediction instruments for diagnosing large vessel occlusion in individuals with suspected stroke: a systematic review for the 2018 guidelines for the early management of patients with acute ischemic stroke. *Stroke* (2018) 49:e111–22. doi: 10.1161/STR.0000000000000160
- Alexandrov AV, Demchuk AM, Burgin WS. (Insonation method and diagnostic flow signatures for transcranial power motion (M-mode) Doppler. *J Neuroimaging* (2002) 12:236–44. doi: 10.1111/j.1552-6569.2002.tb00127.x
- Alexandrov AV, Sloan MA, Tegeler CH, Newell DN, Lumsden A, Garami Z, et al. Views and reviews practice standards for transcranial doppler (TCD) ultrasound. Part II. Clinical indications and expected outcomes. *J Neuroimag.* (2010) 22:215–24. doi: 10.1111/j.1552-6569.2010.00523.x
- Demchuk AM, Christou I, Wein TH, Felberg RA, Malkoff M, Grotta JC, et al. Specific transcranial doppler flow findings related to the presence and site of arterial occlusion. *Stroke* (2000) 31:140–6. doi: 10.1161/01.STR.31.1.140
- Saqqur M, Uchino K, Demchuk AM, Molina CA, Garami Z, Calleja S, et al. Site of arterial occlusion identified by transcranial Doppler predicts the response to intravenous thrombolysis for stroke. *Stroke* (2007) 38:948–54. doi: 10.1161/01.STR.0000257304.21967.ba
- Alberts MJ, Latchaw RE, Selman WR, Shephard T, Hadley MN, Brass LM, et al. Recommendations for comprehensive stroke centers: a consensus statement from the brain attack coalition. *Stroke* (2005) 36:1597–616. doi: 10.1161/01.STR.0000170622.07210.b4
- Brunser AM, Lavados PM, Hoppe A, Lopez J, Valenzuela M, Rivas R. Accuracy of transcranial doppler compared with CT angiography in diagnosing arterial obstructions in acute ischemic strokes. *Stroke* (2009) 40:2037–41. doi: 10.1161/STROKEAHA.108.542704
- Demchuk AM, Christou I, Wein TH, Felberg RA, Malkoff M, Grotta JC, et al. Accuracy and criteria for localizing arterial occlusion with transcranial doppler. *J Neuroimaging* (2000) 10:1–12. doi: 10.1111/jon20001011
- Saqqur M, Hill MD, Alexandrov AV, Roy J, Schebel M, Krol A, et al. Derivation of power M-mode transcranial doppler criteria for angiographic proven MCA occlusion. *J Neuroimaging* (2006) 16:323–8. doi: 10.1111/j.1552-6569.2006.00055.x
- Sloan MA, Alexandrov AV, Tegeler CH, Spencer MP, Caplan LR, Feldmann E, et al. Assessment: transcranial doppler ultrasonography: report of the therapeutics and technology assessment subcommittee of the American Academy of Neurology. *Neurology* (2004) 62:1468–81. doi: 10.1212/WNL.62.9.1468
- Tsigvoulis G, Sharma VK, Lao AY, Malkoff MD, Alexandrov AV. Validation of transcranial Doppler with computed tomography angiography in acute cerebral ischemia. *Stroke* (2007) 38:1245–9. doi: 10.1161/01.STR.0000259712.64772.85
- Zanette EM, Fieschi C, Bozzao L, Roberti C, Toni D, Argentino C, et al. Comparison of cerebral angiography and transcranial Doppler sonography in acute stroke. *Stroke* (1989) 20:899–903. doi: 10.1161/01.STR.20.7.899
- Venturelli PM, Brunser AM, Mun P, Gaete J, Illanes S, Lo J, et al. Sciencedirect reliability of hand-held transcranial doppler with m-mode ultrasound in middle cerebral artery measurement. *J Med Ultrasound* (2017) 25:76–81. doi: 10.1016/j.jmu.2016.12.001
- Aggarwal S, Brooks DM, Kang Y, Linden PK, Patzer JF. Noninvasive monitoring of cerebral perfusion pressure in patients with acute liver failure using transcranial doppler ultrasonography. *Liver Trans.* (2008) 14:1048–57. doi: 10.1002/lt.21499
- Crutchfield KE, Razumovsky AY, Tegeler CH, Mozayeni BR. Differentiating vascular pathophysiological states by objective analysis of flow dynamics. *J Neuroimaging* (2004) 14:97–107. doi: 10.1111/j.1552-6569.2004.tb00224.x
- Lockhart CJ, Gamble AJ, Rea D, Hughes S, McGivern RC, Wolsley C, et al. Nitric oxide modulation of ophthalmic artery blood flow velocity

- waveform morphology in healthy volunteers. *Clin Sci.* (2006) 111:47–52. doi: 10.1042/CS20050365
23. Kurji A, Debert CT, Whitelaw WA, Rawling JM, Frayne R, Poulin MJ. Differences between middle cerebral artery blood velocity waveforms of young and postmenopausal women. *Menopause* (2006) 13:303–13. doi: 10.1097/01.gme.0000177905.94515.24
 24. De Riva N, Budohoski KP, Smielewski P, Kasprowicz M, Zweifel C, Steiner LA, et al. Transcranial doppler pulsatility index: what it is and what it isn't. *Neurocrit Care* (2012) 171:58–66. doi: 10.1007/s12028-012-9672-6
 25. Thorpe S, Thibeault CM, O'Brien, Ranjbaran M, Nicolas C, Leonardo M-T, et al. Comparison of human and robotically acquired transcranial doppler waveforms towards automated assessment of acute ischemic stroke. In: *23rd Annual Meeting of The European Society of Neurosonology and Cerebral Hemodynamics*. Prague (2018).
 26. Thorpe S, Corey M, Nicolas C, O'Brien M, Ranjbaran M, Martinez-Torres L, et al. A novel metric for transcranial doppler assessment of acute ischemic stroke. In: *European Society for Neurosonology and Cerebral Hemodynamics*. Prague (2018).
 27. Johnson CA. The decision tree in glaucoma diagnosis: our knowledge of glaucoma is rapidly expanding. new knowledge about glaucoma can lead to earlier clinical intervention. *Rev Optometry* (2009) 146:100–8.
 28. Rau C-S, Wu S-C, Chien P-C, Kuo P-J, Chen Y-C, Hsieh H-Y, et al. Prediction of mortality in patients with isolated traumatic subarachnoid hemorrhage using a decision tree classifier: a retrospective analysis based on a trauma registry system. *Int J Environ Res Public Health* (2017) 14:1420. doi: 10.3390/ijerph14111420
 29. Purvis D, Aldaghlis T, Trickey AW, Rizzo A, Sikdar S. A novel decision tree approach based on transcranial doppler sonography to screen for blunt cervical vascular injuries. *J Ultrasound Med.* (2013) 32:1023–31. doi: 10.7863/ultra.32.6.1023
 30. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *Clin Chem.* (2015) 61:1446–52. doi: 10.1373/clinchem.2015.246280
 31. Newell DW, Aaslid R. *Transcranial Doppler*, Vol. 2. New York, NY: Raven Press (1992).
 32. Youden W. Index for rating diagnostic tests. *Cancer* (1950) 3:32–5.
 33. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability.* 57th Edn. New York, NY: Chapman and Hall. (1993).
 34. Pedregosa F, Weiss R, Brucher M. Scikit-learn : machine learning in python. *J Mach Learn Res.* (2011) 12:2825–2830.

Conflict of Interest Statement: At the time that this research was conducted, ST, CT, NC, SW, and RH were employees of Neural Analytics, Inc., and TD was a paid consultant. All authors either hold stock or stock options in Neural Analytics, Inc.

Copyright © 2018 Thorpe, Thibeault, Canac, Wilk, Devlin and Hamilton. 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 in Acute Ischemic Stroke Neuroimaging

Haris Kamal*, Victor Lopez and Sunil A. Sheth

Department of Neurology, University of Texas at Houston Health Science Center, Houston, TX, United States

Machine Learning (ML) through pattern recognition algorithms is currently becoming an essential aid for the diagnosis, treatment, and prediction of complications and patient outcomes in a number of neurological diseases. The evaluation and treatment of Acute Ischemic Stroke (AIS) have experienced a significant advancement over the past few years, increasingly requiring the use of neuroimaging for decision-making. In this review, we offer an insight into the recent developments and applications of ML in neuroimaging focusing on acute ischemic stroke.

Keywords: stroke, neuroimaging, machine learning (artificial intelligence), neurosciences, support vector machina (SVM), stroke management, stroke diagnosis

OPEN ACCESS

Edited by:

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

Reviewed by:

Jens Fiehler,
Universitätsklinikum
Hamburg-Eppendorf, Germany
Maurizio Acampa,
Azienda Ospedaliera Universitaria
Senese, Italy

*Correspondence:

Haris Kamal
haris.kamal@uth.tmc.edu

Specialty section:

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

Received: 26 June 2018

Accepted: 22 October 2018

Published: 08 November 2018

Citation:

Kamal H, Lopez V and Sheth SA
(2018) Machine Learning in Acute
Ischemic Stroke Neuroimaging.
Front. Neurol. 9:945.
doi: 10.3389/fneur.2018.00945

Machine Learning (ML), considered a branch of artificial intelligence, is a field of computer science and engineering that facilitates extraction of data based on pattern recognition. A computer learns from previous mistakes after repeated analysis of data and masters tasks that were previously considered too complex for a machine to process (1). The development of these systems to interpret data in neuroimaging has provided valuable information for research in matters of the interaction, structure, and mechanisms of the brain and behavior in certain neurological disorders (2, 3).

Machine learning systems are now being implemented in the clinical neurosciences to devise imaging-based diagnostic and classification systems of neoplasms of the brain (4–6), certain psychiatric disorders (7–11), epilepsy (12, 13), neurodegenerative disorders (14–20), and demyelinating disorders (21–23). In this review, we discuss the present-day role of ML focusing on acute ischemic stroke (AIS), discussing its potential and limitations.

MACHINE LEARNING IN THE CLINICAL NEUROSCIENCES

The use of neuroimaging in the evaluation of many neurological diseases such as dementia, epilepsy, demyelinating diseases, depression, and schizophrenia has grown tremendously. This burgeoning interest has been met with an expansion of ML algorithms in neurosciences (1, 24).

Oliveira et al. (14) evaluated an unsupervised v-One-Class Support Vector Machine (v-OC-SVM) trained with neuroimaging variables, such as cortical thickness and cerebral volume of the brain, from healthy subjects to calculate an abnormality index and compare it with patients diagnosed with mild cognitive impairment (MCI) and Alzheimer's disease (AD). The method correctly classified AD subjects as outliers with an accuracy of 84.3%, and the brain abnormality index was directly associated with the group diagnosis, clinical data, biomarkers, and risk of future conversion to AD.

In schizophrenia, Greenstein et al. (9) used Random Forest (RF), a machine learning algorithm, to discriminate between childhood-onset schizophrenia and healthy patients based on brain magnetic resonance imaging (MRI) measurements of regions of interest (ROI): left temporal lobes, bilateral dorsolateral prefrontal regions, and left medial parietal lobes. The algorithm correctly classified groups with 73.7% accuracy, and a greater brain-based probability of illness

was associated with a statistically significant worse functioning and fewer developmental delays. Machine learning can also help distinguish between subsets of a certain disease. Bleich-Cohen et al. (7) utilized Searchlight Based Feature Extraction (SBFE), a data-driven multi-voxel pattern analysis (MVPA) approach, to search for activation clusters of cognitive loads in brain functional Magnetic Resonance Imaging (fMRI). This ML method helped to identify the two subgroups of schizophrenic patients with and without Obsessive-Compulsive Disorder (OCD) with a 91% accuracy, successfully delineating between symptom severity and a psychiatric comorbidity.

An et al. (12) compared whole-brain white matter changes in patients with mesial temporal epilepsy and matched healthy controls, evaluating tract-based spatial statistics and fractional anisotropy with an ML approach. This ML-based approach discriminated each group accurately and demonstrated high sensitivity to changes in fractional anisotropy in mesial temporal epilepsy patients, which may be beneficial when no lesion can be identified on neuroimaging. Moghim et al. (13) introduced a predictive model for seizure occurrence in a single patient. This approach was based on a multi-class support vector machine (SVM) and 14 selected features of an electroencephalogram in patients with epilepsy. The predicted time of seizure with a window between 20 and 25 min was reported with an average sensitivity of 90.15, 99.44% specificity, and 97% accuracy.

Lesion burden estimation in traumatic brain injury (TBI), AIS, dementia, and multiple sclerosis serves to identify the affected regions, the extent of damage, and therefore, the functional outcome in such patients. Kaminatas et al. (25) proposed an approach for lesion segmentation using a multi-modal brain MRI based on an 11-layers deep, multi-scale, 3D Convolutional Neural Networks (CNN) called Deep Medic. Their proposed novel training scheme is based on two main components, a 3D CNN that produces accurate soft segmentation maps and a connected Conditional Random Field that imposes regularization constraints on the CNN output and produces the final hard segmentation labels. This allows for a deeper and more discriminative delimitation of lesion burden, with the highest reported accuracy observed in a cohort of patients with severe TBI.

CHALLENGES IN ACUTE ISCHEMIC STROKE

Stroke is the leading cause of serious long-term disability and the fifth leading cause of death in the United States, with its prevalence increasing with advancing age in both males and females, as each year ~795,000 Americans experience a new or recurrent stroke (26). This burden is coupled with a direct medical expense of an estimated \$23.6 billion according to the last annual report of 2014 (26). With the increasing complexity of the acute ischemic stroke therapy and the rising of per-person costs, there is a real and urgent need for a technological solution to aid in the streamlined care of patients and selection of the appropriate therapeutic intervention.

Present treatments for AIS revolve around rapid reperfusion of ischemic tissue, using intravenous (IV) thrombolytic medications such as tissue plasminogen activator (tPA) and/or endovascular techniques to mechanically remove the obstruction to blood flow. Contemporary clinical trials are now implementing a higher complexity of neuroimaging modalities to define treatment standards, resulting in an increased economic as well as logistical burden on healthcare. The WAKE-UP multicenter clinical trial (27) used magnetic resonance imaging (MRI) in patients that presented with an unknown time of onset of symptoms to identify brain regions that exhibited a restricted diffusion on diffusion-weighted imaging (DWI) scan and no T2-signal hyperintensity on fluid-attenuated inversion recovery (FLAIR) sequence, estimating the onset of the infarct to be <4.5 h and thus guiding stroke therapy. Previous to this study, non-contrast head CT, an imaging modality that is widely and readily available, was the only imaging screen used to assess for tPA eligibility. The new, expanded tPA indication requiring MRI poses challenges for a majority of centers, which do not have ready access to this type of imaging emergently and 24/7.

The growing dependence on neuroimaging in determining treatment options for acute ischemic stroke is observed as well for endovascular stroke therapy (EST), which has shown to improve outcome when used in combination with standard medical care (28). In 2015, numerous clinical trials demonstrated a clear benefit of endovascular treatment over medical management alone for a select group of patients with acute ischemic stroke seen within 6 h of the onset of stroke (29–33), and relied on imaging modalities including NCHCT, CT/MR angiography (CTA/MRA) and CT/MRI perfusion (CTP/MRP) scans. Results derived from these trials showed an advantage in using advanced imaging modalities in identifying patients with a higher likelihood of better outcomes from EST. Two additional clinical trials (34), DAWN and DEFUSE3, published in 2018 evaluated a much larger population of stroke patients, those presenting up to 24 h after their symptoms, and required the use of perfusion imaging with CT or MRI.

This increased reliance on neuroimaging has led to a tremendous improvement in our ability to care for patients with AIS but has been coupled with a number of challenges. Specifically, limited availability of these imaging modalities, a shortage of specialists to promptly interpret these studies, as well as inter-observer variability have limited the implementation of the above findings. Indeed, studies evaluating inter-observer performance on Alberta Stroke Program Early CT Score (ASPECTS), a 10-region imaging grading system in stroke, showed significant variability (35–39). Further adding to the complexity of acute stroke treatment is that while the need to perform and interpret advanced neuroimaging has recently increased, the urgency with which such evaluation is being performed has remained the same. For every minute that a patient with a large vessel occlusion fails to be treated, an estimated 1.9 million neurons and 14 billion synapses are lost in the brain (40). Trials evaluating efforts to promptly assess and treat patients with AIS have demonstrated superior outcomes and decreased morbidity. In patients treated with

intravenous thrombolysis, reducing treatment times by 15-min was associated with reduced in-hospital mortality, reduced incidence of symptomatic intracranial hemorrhage, and a greater likelihood of independent ambulation at discharge (41, 42). In patients treated with endovascular therapy, for every 15-min reduction of onset to recanalization of the occluded artery, 34 per 1,000 treated patients had improved disability outcome (43). As such, there is an urgent need for systems to rapidly and precisely interpret neuroimaging data in acute ischemic stroke.

IMPLEMENTATIONS OF MACHINE LEARNING IN ACUTE ISCHEMIC STROKE

Machine learning algorithms have been used to assist in the diagnosis and individualized treatment decisions in acute ischemic stroke. The implementations of machine learning are numerous, from early identification of imaging diagnostic findings (44), estimating time of onset (27, 45), lesion segmentation (46), and fate of salvageable tissue (47, 48), to the analysis of cerebral edema (49, 50), and predicting complications (51–53) and patient outcomes (54–57) after treatment. A summary of the most recent articles investigating the applications of machine learning for automated diagnosis and outcome prediction in acute ischemic stroke is given in **Table 1**.

One of the most relevant clinical criteria to decide if a patient with an acute ischemic stroke is eligible for IV thrombolysis with tPA is a time from symptom onset of <4.5 h, but in medical practice, stroke symptom onset is usually unknown. Ho et al. (45) developed a deep learning algorithm based on an autoencoder architecture to extract imaging features in perfusion-weighted images (PWI) in MRI to determine the time elapsed since stroke onset.

Lesion estimation and identification of salvageable tissue are essential in the acute decision making in stroke, but the expense and resources involved present a challenge for physicians. Chen et al. (46) used a framework with two CNNs to segment stroke lesions using DWI in MRI. One CNN was a combination of two DeconvNets (EDD Net), and the second CNN was a multi-scale convolutional label evaluation net (MUSCLE Net) to help reduce the potential false positives detected by the EDD Net. The dataset was built with clinical acquired DWI from 741 subjects, exhibiting a high lesion detection rate, and accuracy.

Measurement of the perfusion-diffusion mismatch and calculation of infarction probability using MRI-based approaches for tissue-at-risk evaluation can be applied in stroke treatment decisions. Bouts et al. (47) analyzed the ability of five algorithms to depict potentially salvageable tissue using MRI imaging from rats subjected to a right-sided MCA occlusion without subsequent reperfusion, and with spontaneous or thrombolysis-induced reperfusion. The highest accuracy of risk-based identification of acutely salvageable ischemic tissue that could recover on subsequent reperfusion was observed using a generalized linear model (Dice's similarity index = 0.79 ± 0.14). Similarly, Huang et al. (48) used an SVM to predict infarct on a pixel-by-pixel basis using acute cerebral blood flow (CBF) and apparent diffusion coefficient (ADC) on MRI data. Serial images

were collected during the acute phase up to 3 h and again at 24 h from 12 rats in each of the stroke groups exposed to a 30-min, 60-min, or permanent middle cerebral artery (MCA) occlusion. The accuracy observed for this approach was high in all groups and was enhanced by adding neighboring pixel information and spatial infarction incidence.

Takahashi et al. (44) designed a method to identify a hyperdense MCA, also known as the MCA dot sign, an important evaluation in an NCHCT as it represents a thrombus in a vessel. The authors created ROIs around the Sylvian fissure region and identified MCA dots based on the morphologic top-hat transformation, and classified images using an SVM with four features. Two hundred and ninety-seven CT images from seven patients with an MCA dot sign were classified by an SVM system, which exhibited a maximum sensitivity of 97.5% at a false positive rate of 1.28 per image and 0.5 per hemisphere while assessing the MCA dot sign.

Another application of ML in AIS is predicting factors that will contribute to neurological deterioration and increased morbidity, such as cerebral edema. Chen et al. (49) proposed a machine learning algorithm using serial CT scans of stroke patients to delineate and measure cerebrospinal fluid (CSF) volume over time, as it may represent a sensitive biomarker of cerebral edema progression. The initial cohort consisted of 155 subjects and preliminary processing using a generalized estimating equations (GEE) model to calculate CSF volumes over time, adjusting for age, demonstrated that a reduction in CSF volume from baseline to final CT was correlated with infarct volume, the presence of cerebral edema, and the degree of midline shift. Comparatively, Dhar et al. (50) validated an automated technique for intracranial CSF segmentation by an ensemble of RF-based machine learning with a geodesic active contour (GAC) segmentation. CSF spaces were outlined on scans performed within 6 h of stroke onset and then closest to 24 h later in 38 patients. This method accurately tracked changes in CSF volume with an average DSC > 0.7. Pearson correlation coefficients between the changes in CSF and the ground truth were found to be statistically significant. These algorithms represent a potential for future research and may serve as a biomarker of cerebral edema severity.

The outcome of acute ischemic stroke patients is dependent on therapy, and risks for complications should be considered when deciding for stroke therapy. Yu et al. (53) established a method to predict the location and extent of hemorrhagic transformation (HT) in stroke, the most severe complication following reperfusion therapy. PWI and DWI of 165 patients treated with reperfusion therapy in a stroke center were collected and analyzed using five machine learning approaches, with Kernel spectral regression exhibiting an accuracy of $83.7 \pm 2.6\%$. A multi-center retrospective study (52) assessed the predictive power for hemorrhagic transformation of PWI in MRI. Dynamic T2-weighted perfusion MR images from 263 patients from four medical centers were collected and served as input for linear and nonlinear predictive models, the latter having an average accuracy >85% in predicting HT. In one study, Nielsen et al. (54) ran a deep learning convoluted neural network (CNN_{deep}) with 9 biomarkers as input to calculate lesion volume in patients

TABLE 1 | Use of machine learning in stroke diagnosis and outcome prognosis.

References	ML-based approach	Feature	Results*
Asadi et al. (55)	Artificial neural network	Prediction of Dichotomized mRS	70% accuracy
Bentley et al. (51)	Supported vector machine	Prediction of sICH	74.4% accuracy
Bouts et al. (47)	Adaptive boosting	Prediction of infarction volume	89 ± 5% accuracy
Chen et al. (46)	RF + GAC	Relation of CSF shifts and cerebral edema	$r = 0.879$
Forkert et al. (56)	Multi-class supported vector machine	Predicted 30-day post-stroke mRS	
		<i>Multi-value mRS</i>	56% accuracy
		<i>Multi-value mRS ± 1</i>	82% accuracy
		<i>Dichotomized mRS</i>	85% accuracy
Huang et al. (48)	Supported vector machine	Predicted infarct penumbra volume	
		<i>30-min occlusion of MCA</i>	86 ± 2.7% accuracy
		<i>60-min occlusion of MCA</i>	89 ± 1.4% accuracy
		<i>Permanent occlusion of MCA</i>	93% accuracy
Scalzo et al. (52)	Non-linear regression model	Prediction of HT	>85% accuracy
Takahashi et al. (44)	Supported vector machine	Detection of MCA dot sign	97.5% sensitivity
Yu et al. (53)	SR-KDA	Prediction of HT	83.7 ± 2.6% accuracy
Nielsen et al. (54)	Deep features CNN	Prediction of patient outcome after IV thrombolysis	88 ± 0.12% accuracy

*The results displayed for each article are the most accurate or relevant in matter of the machine learning approach utilized according to the author.

CNN, Convolved Neural Network; GAC, Geodesic Active Contour; HT, Hemorrhagic transformation; MCA, Middle Cerebral Artery; mRS, modified Rankin Scale; RF, Rain Forest; sICH, symptomatic Intracranial Hemorrhage; SR-KDA, Spectral Regression Kernel Discriminant Analysis.

treated with IV tPA. Input data from 29 untreated patients and 35 patients that received IV tPA were compared. This model predicted final infarct volume with 88% accuracy, being superior to other models in this study. Bentley et al. (51) predicted the risk of symptomatic intracerebral hemorrhage (sICH) after IV thrombolysis therapy. CT images of 116 patients who were treated with IV tPA, 16 of which had sICH, were entered as inputs into an SVM along with clinical severity. They found a better prognostication of the SVM when compared to the traditional clinician-based prognostication tools such as Hemorrhage after thrombolysis (HAT), and Sugar, Early Infarct signs, Dense cerebral artery sign, Age, and NIHSS scores (SEDAN).

Machine learning algorithms based on structural and functional MR images as input may assist in predicting motor deficits in stroke patients. Forkert et al. (56) applied 12 SVM classification models in calculating the corresponding 30-day mRS score of ischemic stroke patients through parameters including lesion overlap from different brain regions, stroke laterality, and other optional features such as infarct volume, NIHSS at admission, and patient age. Superior mRS prediction was observed by integrating the optional features and providing stroke location information, with a multi-value mRS prediction accuracy of 56%, and a dichotomized mRS (0–2 vs. 3–5) prediction accuracy of 85%. In a study by Rondina et al. (57), a proposed model to predict upper extremity motor deficit in 50 stroke patients was developed from data on structural MRI instead of functional MRI. Lesion probability images were derived using patterns of voxels and was then compared to lesion load per ROI in predicting outcomes, with the former providing better results when multiple regions of interest such as a range of cortical and subcortical motor areas and corticospinal tract were analyzed.

CURRENT CHALLENGES AND FUTURE DIRECTIONS IN MACHINE LEARNING FOR ACUTE ISCHEMIC STROKE

Early promising results have demonstrated that ML techniques may be useful as decision support tools in treatment choices for AIS. To improve the generalizability of the findings discussed above, however, there are a number of limitations in currently existing architectures that need to be addressed. The first limitation is that of sample size. Deep learning algorithms using medical imaging often require datasets of tremendous magnitude, the types of which may not be readily available. For example, an ML algorithm demonstrated superior performance at differentiating skin cancer lesions from their benign corresponding equivalent when compared against 21 board-certified dermatologists, using a dataset of nearly 130,000 images (58). A dataset of this size in AIS for public use does not currently exist. This shortcoming, however, has been recognized as a problem that can and ought to be solved, and multiple calls for the creation of such a repository have been made (59). The obstacles in inter-institutional data sharing, as well as a lack of funding to correctly pre-process and curate these images, along limitations to host such a dataset account for some of the delays in the creation of this repository.

Another limitation encountered in neuroimaging-based ML techniques is the need for labeling regions of interest or “gold standard” findings on the images. That is to say, beyond collecting the images, the images and the findings on the images would need to be identified for the question being evaluated. For example, a study evaluating the presence or absence of a hyperdense MCA would need each image to be tagged with the true result, to train

the algorithm. Without foresight, this degree of manual curating could be required for each individual project.

CONCLUSION

Machine learning applications are expanding in the medical field for diagnostic and therapeutic purposes, and the rapidly expanding and increasingly neuro-imaging reliant field of AIS is proving to be fertile ground. There is a particular need for ML solutions in this field, which is faced with the challenge of increasingly complex data, with limited human

expert resources. Future directions in ML for AIS may require collaborative approaches across multiple institutions to build a robust dataset for efficient training of ML networks.

AUTHOR CONTRIBUTIONS

HK and SS contributed equally to manuscript conception, design, revisions, and approved the submitted version. VL contributed to the manuscript design, revisions and edits and approved the submitted version.

REFERENCES

- Vu MT, Adali T, Ba D, Buzsáki G, Carlson D, Heller K. A shared vision for machine learning in neuroscience. *J Neurosci.* (2018) 38:1601–7. doi: 10.1523/JNEUROSCI.0508-17.2018
- Song X, Chen NK. A unified machine learning method for task-related and resting state fMRI data analysis. *Conf Proc IEEE Eng Med Biol Soc.* (2014) 2014:6426–9. doi: 10.1109/EMBC.2014.6945099.
- Wernick MN, Yang Y, Brankov JG, Yourganov G, Strother SC. Machine learning in medical imaging. *IEEE Signal Process Mag.* (2010) 27:25–38. doi: 10.1109/MSP.2010.936730
- Li GZ, Yang J, Ye CZ, Geng DY. Degree prediction of malignancy in brain glioma using support vector machines. *Comput Biol Med.* (2006) 36:313–25. doi: 10.1016/j.combiomed.2004.11.003
- Zacharak E, Kanas VG, Davatzikos C. Investigating machine learning technique for MRI-based classification of brain neoplasms. *Int J Comput Assist.* (2011) 6:821–8. doi: 10.1007/s11548-011-0559-3
- Zacharak E, Wang S, Chawla S, Soo Yoo D, Wolf R, Melhem ER. Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. *Magn Reson Med.* (2009) 62:1609–18. doi: 10.1002/mrm.22147
- Bleich-Cohen M, Jamshy S, Sharon H, Weizman R, Intrator N, Poyurovsky M, et al. Machine learning fMRI classifier delineates subgroups of schizophrenia patients. *Schizophr Res.* (2014) 160:196–200. doi: 10.1016/j.schres.2014.10.033
- Gould IC, Shepherd AM, Laurens KR, Cairns MJ, Carr VJ, Green MJ. Multivariate neuroanatomical classification of cognitive subtypes in schizophrenia: a support vector machine learning approach. *Neuroimage Clin.* (2014) 6:229–36. doi: 10.1016/j.nicl.2014.09.009
- Greenstein D, Malley JD, Weisinger B, Clasen L, Gogtay N. Using multivariate machine learning methods and structural MRI to classify childhood onset schizophrenia and healthy controls. *Front Psychiatry* (2012) 3:53. doi: 10.3389/fpsy.2012.00053
- Guo H, Cao X, Liu Z, Li H, Chen J, Zhang K. Machine learning classifier using abnormal brain network topological metrics in major depressive disorder. *Neuroreport* (2012) 23:1006–122. doi: 10.1097/WNR.0b013e32835a650c
- Hoexter MQ, Miguel EC, Diniz JB, Shavitt RG, Busatto GF, Sato JR. Predicting obsessivecompulsive disorder severity combining neuroimaging and machine learning methods. *J Affect Disord.* (2013) 150:1213–6. doi: 10.1016/j.jad.2013.05.041
- An J, Fang P, Wang W, Liu Z, Hu D, Qiu S. Decreased white matter integrity in mesial temporal lobe epilepsy: a machine learning approach. *Neuroreport* (2014) 25:788–94. doi: 10.1097/WNR.0000000000000178
- Moghim N, Corne DW. Predicting epileptic seizures in advance. *PLoS ONE* (2014) 9:e99334. doi: 10.1371/journal.pone.0099334
- Andrade de Oliveira A, Carthery-Goulart MT, Oliveira Junior PP, Carrettiro DC, Sato JR. Defining multivariate normative rules for healthy aging using neuroimaging and machine learning: an application to Alzheimer's disease. *J Alzheimers Dis.* (2015) 43:201–12. doi: 10.3233/JAD-140189
- Bhagyasree SIR, Nagaraj K, Prince M, Fall CHD, Krishna M. Application of machine learning methods for diagnosis of dementia based on the 10/66 battery of cognitive function tests in South India. *Soc Psychiatry Psychiatric Epidemiol.* (2018) 53:77–86. doi: 10.1007/s00127-017-1410-0
- Casanova R, Hsu FC, Sink KM, Rapp SR, Williamson JD, Resnick SM, et al. Alzheimer's disease risk assessment using large-scale machine learning methods. *PLoS ONE* (2013) 8:e77949. doi: 10.1371/journal.pone.0077949
- Di Deco J, González AM, Díaz J, Mato V, García-Frank D, Álvarez-Linera J, et al. Machine learning and social network analysis applied to Alzheimer's disease biomarkers. *Curr Top Med Chem.* (2013) 13:652–62. doi: 10.2174/1568026611313050008
- Dyrba M, Ewers M, Wegrzyn M, Kilimann I, Plant C, Oswald A, et al. Robust automated detection of microstructural White matter degeneration in AD using machine learning classification of multicenter DTI data. *PLoS ONE* (2013) 8:e64925. doi: 10.1371/journal.pone.0064925
- Huertas-Fernandez I, Garcia-Gomez FJ, Garcia-Solis D, Benitez-Rivero S, Marin-Oyaga VA, Jesus S, et al. Machine learning models for the differential diagnosis of vascular parkinsonism and Parkinson's disease using [(123)I]FP-CIT SPECT. *Eur J Nucl Med Mol Imaging* (2015) 42:112–924. doi: 10.1007/s00259-014-2882-8
- Pekkala T, Hall A, Lötjönen J, Mattila J, Soininen H, Ngandu T, et al. Development of a late-life dementia prediction index with supervised machine learning in the population-based CAIDE study. *J Alzheimer's Dis.* 3:1055–67. doi: 10.3233/JAD-160560
- Ion-Mărgineanu A, Kocevcar G, Stamile C, Sima DM, Durand-Dubief F, Van Huffel S, et al. Machine learning approach for classifying multiple sclerosis courses by combining clinical data with lesion loads and magnetic resonance metabolic features. *Front. Neurosci.* (2017) 11:398. doi: 10.3389/fnins.2017.00398
- Nedjati-Gilani GL, Schneider T, Hall MG, Wheeler-Kingshott CA, Alexander DC. Machine learning based compartment models with permeability for white matter microstructure imaging. *Med Image Comput Comput Assist Interv.* (2014) 17(Pt3):257–64. doi: 10.1007/978-3-319-10443-0_33
- Zhao Y, Healy BC, Rotstein D, Guttmann CR, Bakshi R, Weiner HL, et al. Exploration of machine learning techniques in predicting multiple sclerosis disease course. *PLoS ONE* (2017) 12:e0174866. doi: 10.1371/journal.pone.0174866
- Oktar N, Oktar Y. Machine learning and neuroimaging. *J Neurol Sci Turk.* (2015) 32:43.
- Kamnitsas K, Ledig C, Newcombe VFH, Simpson JP, Kane AD, Menon DK, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal.* (2017) 36:61–78. doi: 10.1016/j.media.2016.10.004
- Benjamin EJ, Virani SS, Callaway CW, Chamberlain AM, Chang AR, Cheng S, et al. Heart disease and stroke statistics-2018 update: a report from the American Heart Association. *Circulation* (2018) 137:e67–e492. doi: 10.1161/CIR.0000000000000558
- Thomalla G, Simonsen CZ, Boutitie F, Andersen G, Berthezene Y, Cheng B, et al. MRI-guided thrombolysis for stroke with unknown time of onset. *N Engl J Med.* (2018) 379:611–22. doi: 10.1056/NEJMoa1804355
- 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
- Berkhemer OA, Fransen PS, Beumer D, van den Berg LA, Lingsma HF, Yoo AJ, et al. A randomized trial of intraarterial treatment for acute ischemic stroke. *N Engl J Med.* (2015) 372:11–20. doi: 10.1056/NEJMoa1411587

30. Campbell BC, 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
31. Goyal M, Demchuk AM, Menon BK, Eesa M, Rempel JL, Thornton J. Randomized assessment of rapid endovascular treatment of ischemic stroke. *N Engl J Med.* (2015) 372:1019–30. doi: 10.1056/NEJMoa1414905
32. 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
33. Saver JL, Goyal M, Bonafe A, Diener HC, Levy EI, Pereira VM, et al. Stent-retriever thrombectomy after intravenous t-PA vs. t-PA alone in stroke. *N Engl J Med.* (2015) 372:2285–95. doi: 10.1056/NEJMoa1415061
34. Estruch R, Ros E, Salas-Salvado J, Covas MI, Corella D, Arós F, et al. Thrombectomy for stroke with perfusion imaging selection at 6–16 hours. *N Engl J Med.* (2018) 378:708–18. doi: 10.1056/NEJMoa1713973
35. Gupta AC, Schaefer PW, Chaudhry ZA, Leslie-Mazwi TM, Chandra RV, González RG, et al. Interobserver reliability of baseline noncontrast CT Alberta Stroke Program early CT score for intra-arterial stroke treatment selection. *AJNR Am J Neuroradiol.* (2012) 33:1046–9. doi: 10.3174/ajnr.A2942
36. Barber PA, Demchuk AM, Zhang J, Buchan AM. Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy. *Lancet* (2000) 355:1670–74. doi: 10.1016/S0140-6736(00)02237-6
37. Coutts SB, Hill MD, Demchuk AM, Barber PA, Pexman JH, Buchan AM, et al. Aspects reading requires training and experience. *Stroke* (2003) 34:e179. doi: 10.1161/01.STR.0000092221.81498.91
38. Mak HK, Yau KK, Khong PL, Ching AS, Cheng PW, Au-Yeung PK, et al. Hypodensity of >1/3 middle cerebral artery territory versus Alberta Stroke Programme Early CT Score (ASPECTS): comparison of two methods of quantitative evaluation of early CT changes in hyperacute ischemic stroke in the community setting. *Stroke* (2003) 34:1194–96. doi: 10.1161/01.STR.0000069162.64966.71
39. Wardlaw JM, Mielke O. Early signs of brain infarction at CT: observer reliability and outcome after thrombolytic treatment—systematic review. *Radiology* (2005) 235:444–53. doi: 10.1148/radiol.2352040262
40. Saver JL. Time is brain—quantified. *Stroke* (2006) 37:263–6. doi: 10.1161/01.STR.0000196957.55928.ab
41. Saver JL, Fonarow GC, Smith EE, Reeves MJ, Grau-Sepulveda MV, Pan W, et al. Time to treatment with intravenous tissue plasminogen activator and outcome from acute ischemic stroke. *JAMA* (2013) 309:2480–8. doi: 10.1001/jama.2013.6959
42. Saver JL, Gornbein J, Grotta J, Liebeskind D, Lutsep H, Schwamm L, et al. Number needed to treat to benefit and to harm for IV tPA therapy in the 3–4.5 hour window: joint outcome table analysis of the ECASS 3 trial. *Stroke* (2009) 40:2433–7. doi: 10.1161/STROKEAHA.108.543561
43. Sheth SA, Jahan R, Gralla J, Pereira VM, Nogueira RG, Levy EI, et al. Time to endovascular reperfusion and degree of disability in acute stroke. *Ann Neurol.* (2015) 78:584–93. doi: 10.1002/ana.24474
44. Takahashi N, Lee Y, Tsai DY, Matsuyama E, Kinoshita T, Ishii K. An automated detection method for the MCA dot sign of acute stroke in unenhanced CT. *Radiol Phys Technol.* (2014) 7:79–88. doi: 10.1007/s12194-013-0234-1
45. Ho KC, Speier W, El-Saden S, Arnold CW. Classifying acute ischemic stroke onset time using deep imaging features. In: *AMIA Annual Symposium Proceedings* (Washington, DC) (2017). p. 892–901.
46. Chen L, Bentley P, Rueckert D. Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks. *NeuroImage* (2017) 5:633–43. doi: 10.1016/j.nicl.2017.06.016
47. Bouts MJ, Tiebosch IA, van der Toorn A, Viergever MA, Wu O, Dijkhuizen RM, et al. Early identification of potentially salvageable tissue with MRI-based predictive algorithms after experimental ischemic stroke. *J Cereb Blood Flow Metab.* (2013) 33:1075–82. doi: 10.1038/jcbfm.2013.51
48. Huang S, Shen Q, Duong TQ. Quantitative prediction of acute ischemic tissue fate using support vector machine. *Brain Res.* (2011) 1405:77–84. doi: 10.1016/j.brainres.2011.05.066
49. Chen Y, Dhar R, Heitsch L, Ford A, Fernandez-Cadenas I, Carrera C. Automated quantification of cerebral edema following hemispheric infarction: application of a machine-learning algorithm to evaluate CSF shifts on serial head CTs. *NeuroImage* (2016) 2:673–680. doi: 10.1016/j.nicl.2016.09.018
50. 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
51. Bentley P, Ganesalingam J, Carlton Jones AL, Mahady K, Epton S, Rinne P, et al. Prediction of stroke thrombolysis outcome using CT brain machine learning. *NeuroImage* (2014) 4:635–40. doi: 10.1016/j.nicl.2014.02.003
52. Scalzo F, Alger JR, Hu X, Saver JL, Dani KA, Muir KW. Multi-center prediction of hemorrhagic transformation in acute ischemic stroke using permeability imaging features. *Magn Reson Imag.* (2018) 31:6:961–9. doi: 10.1016/j.mri.2013.03.013
53. Yu Y, Guo D, Lou M, Liebeskind D, and Scalzo F. Prediction of hemorrhagic transformation severity in acute stroke from source perfusion MRI. In: *IEEE Transactions on Biomedical Engineering. Vol 65.* (2018). pp. 2058–65.
54. Nielsen A, Hansen MB, Tietze A, Mouridsen K. Prediction of tissue outcome and assessment of treatment effect in acute ischemic stroke using deep learning. *Stroke* (2018) 49:1394–401. doi: 10.1161/STROKEAHA.117.019740
55. Asadi H, Dowling R, Yan B, Mitchell P. Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy. *PLoS ONE* (2014) 9:e88225. doi: 10.1371/journal.pone.0088225
56. Forkert ND, Verleger T, Cheng B, Thomalla G, Hilgetag CC, Fiehler J, et al. Multiclass support vector machine-based lesion mapping predicts functional outcome in ischemic stroke patients. *PLoS ONE* (2015) 10:e0129569. doi: 10.1371/journal.pone.0129569
57. Rondina JM, Filippone M, Girolami M, Ward NS. Decoding post-stroke motor function from structural brain imaging. *Neuroimage Clin.* (2016) 12:372–80. doi: 10.1016/j.nicl.2016.07.014
58. Esteve A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Thru Dermatologist-level classification of skin cancer with deep neural networks. *Nature* (2017) 542:115–8. doi: 10.1038/nature21056
59. Wintermark M, Albers GW, Alexandrov AV, Alger JR, Bammer R, Baron JC, et al. Acute stroke imaging research roadmap. *AJNR Am J Neuroradiol.* (2008) 39:1621–8. doi: 10.1161/STROKEAHA.107.512319

Conflict of Interest Statement: 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.

Copyright © 2018 Kamal, Lopez and Sheth. 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.



Learning to Predict Ischemic Stroke Growth on Acute CT Perfusion Data by Interpolating Low-Dimensional Shape Representations

Christian Lucas^{1,2*}, André Kemmling³, Nassim Bouteldja¹, Linda F. Aulmann⁴, Amir Madany Mamlouk⁵ and Mattias P. Heinrich¹

¹ Institute of Medical Informatics, University of Lübeck, Lübeck, Germany, ² Graduate School for Computing in Medicine and Life Sciences, University of Lübeck, Lübeck, Germany, ³ Department of Clinical Radiology, University Hospital Münster, Münster, Germany, ⁴ Institute of Neuroradiology, University Medical Center Schleswig-Holstein, Lübeck, Germany, ⁵ Institute for Neuro- and Bioinformatics, University of Lübeck, Lübeck, Germany

OPEN ACCESS

Edited by:

Fabien Scalzo,
University of California, Los Angeles,
United States

Reviewed by:

David Robben,
KU Leuven, Belgium
Ulas Bagci,
University of Central Florida,
United States
Aalpen A. Patel,
Geisinger Health System,
United States

*Correspondence:

Christian Lucas
lucas@imi.uni-luebeck.de

Specialty section:

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

Received: 03 May 2018

Accepted: 02 November 2018

Published: 26 November 2018

Citation:

Lucas C, Kemmling A, Bouteldja N, Aulmann LF, Madany Mamlouk A and Heinrich MP (2018) Learning to Predict Ischemic Stroke Growth on Acute CT Perfusion Data by Interpolating Low-Dimensional Shape Representations. *Front. Neurol.* 9:989. doi: 10.3389/fneur.2018.00989

Cerebrovascular diseases, in particular ischemic stroke, are one of the leading global causes of death in developed countries. Perfusion CT and/or MRI are ideal imaging modalities for characterizing affected ischemic tissue in the hyper-acute phase. If infarct growth over time could be predicted accurately from functional acute imaging protocols together with advanced machine-learning based image analysis, the expected benefits of treatment options could be better weighted against potential risks. The quality of the outcome prediction by convolutional neural networks (CNNs) is so far limited, which indicates that even highly complex deep learning algorithms are not fully capable of directly learning physiological principles of tissue salvation through weak supervision due to a lack of data (e.g., follow-up segmentation). In this work, we address these current shortcomings by explicitly taking into account clinical expert knowledge in the form of segmentations of the core and its surrounding penumbra in acute CT perfusion images (CTP), that are trained to be represented in a low-dimensional non-linear shape space. Employing a multi-scale CNN (U-Net) together with a convolutional auto-encoder, we predict lesion tissue probabilities for new patients. The predictions are physiologically constrained to a shape embedding that encodes a continuous progression between the core and penumbra extents. The comparison to a simple interpolation in the original voxel space and an unconstrained CNN shows that the use of such a shape space can be advantageous to predict time-dependent growth of stroke lesions on acute perfusion data, yielding a Dice score overlap of 0.46 for predictions from expert segmentations of core and penumbra. Our interpolation method models monotone infarct growth robustly on a linear time scale to automatically predict clinically plausible tissue outcomes that may serve as a basis for more clinical measures such as the expected lesion volume increase and can support the decision making on treatment options and triage.

Keywords: ischemic, stroke, prediction, growth, learning, shape, CT, perfusion

1. INTRODUCTION

Cerebrovascular diseases, in particular strokes, are one of the leading global causes of death in developed countries (1). Acute stroke, which is usually caused by the blockage of cerebral blood flow due to a blood clot, is often diagnosed through CT or MR perfusion imaging (beside others, such as CTA). The derived perfusion parameter maps, e.g., Cerebral Blood Volume (CBV) or Time To Drain (TTD), provide spatio-temporal distributions of a contrast medium bolus within brain tissue. In contrast to native CT or standard MR sequences, such as T2 or FLAIR, perfusion images with their apparent functional signals enable the delineation of the potential infarct area even in the early acute phase and allow to visually assess the expected stroke severity, which helps the radiologist to come to a final therapy decision as early as possible.

In order to decide for a treatment the doctor has to weigh the risk of a therapy such as thrombolysis or thrombectomy against the expected outcome. For instance, (2) describe hemorrhages, such as symptomatic intracerebral hemorrhage, as a typical risk of intravenous thrombolysis therapy. For large vessel occlusions, mechanical thrombectomy improves functional outcomes but is logistically challenging. It is of major importance to consider the immediate availability of a therapy option, since the expected outcome strongly relies on the onset-to-treatment time (3). Depending on the expected time until revascularization, the radiologist has to estimate if further progression of the stroke can be avoided so that substantial parts of the tissue-at-risk within the penumbra could be salvaged. In the infarct core, however, as evident by a decrease in CBV, severe tissue injury and permanent vascular collapse have occurred.

Since stroke lesions vary widely in shape or size, and also evolve spatially heterogeneously over time, it is challenging for the radiologist to estimate growth or the size of the potentially stroke-affected tissue. For this reason, it is difficult to derive a time window in which a specific therapy path may be beneficial over another. Deep learning with CNNs has become popular in medical image analysis over the past recent years by clearly exceeding the so far state-of-the-art results, potentially capable of modeling this complex relationship.

1.1. Objective

We present a novel tool for automatic stroke tissue outcome estimation using a CNN with a convolutional auto-encoder (CAE) that incorporates learned stroke shapes of core and penumbra. In a proof-of-concept, the trained model is able to predict the stroke lesion growth for patients with successful recanalization based on a given time-to-treatment for the thrombectomy (**Table 1**) and the CTP imaging parameter maps CBV and TTD. An evaluation of the method shows the practicality in principle on a limited dataset and the discussion provides pros and cons that suggest to further investigate this approach for clinical use.

1.2. Outline

In order to gain a fundamental understanding of the method and its design choices, we provide a methodological overview in

TABLE 1 | Inclusion criteria of the dataset for evaluation.

Inclusion criteria
Initial CT perfusion imaging
Thrombectomy
Follow-Up CT within 6–24 h
Age at least 18 years

the following section: First, a review on the established stroke image analysis methods in clinical research literature is given; Second, the foundations of different image representations that can help solving higher-level tasks for image analysis as well as machine learning methods that have been investigated for stroke imaging are described; Third, we explain the use of CAEs for regularization in image segmentation by learning shape representations and how our work is based upon it. Subsequent, a detailed description of the assumptions and components of our method is provided—this third section also explains how to reproduce the method, that is, how to train the shape space and predict follow-up lesions based on noisy shape estimates. The fourth section lists the materials for a comparative evaluation and discusses its results, before we provide a final conclusion in the last section.

2. IMAGE ANALYSIS

Classic thresholding methods for stroke image analysis have the drawback of only modeling a single univariate hard decision border between affected and non-affected tissue. Even when splitting into subgroups of different admission times (4) or distinguishing core against penumbra, the result will be a binary map of the affected vs. unaffected tissue. Further, purely voxel-based methods can produce irregular and physiologically implausible shapes. Statistical models, e.g., a linear regression model as used by Kemmling et al. (5), can cope with the variances in the data according to the complexity of the model and proper parameterization, while simple models will usually show a strong bias when used for high-dimensional problems.

2.1. Representation and Spaces

In general, noisy images make it difficult to operate in image space, for instance, to apply a threshold to images for extracting regions that define the outcome. The input representation is not always suitable to detect the complex input patterns that determine the output. As known from signal processing and analysis, transforming the input into another representation (extract *features*) can often make it easier to perform classification or regression tasks. There are many transformations, e.g., non-linear kernel methods can bring a representation into a higher dimensionality where it may become linear separable. As this is a vast field that shall not be described here in detail, we emphasize that the input image data often needs to be fit into a regularized model or transformed to another representation first, on which the high-level task becomes easier to solve.

Kolouri et al. (6) propose *transport spaces* based on optimal transport theory to model biomedical problems such as tumor growth. The idea of looking at images as mass particle distributions is related to tissue distributions in biology, for example, when learning the transformations of some sample images onto a mean non-pathological image to extract the main modes of variance ideally representing the change from benign to pathological tissue (7). The modes could be extracted by principle component analysis (PCA) or other machine-learning approaches, e.g., auto-encoders (8). The formulation of the transport space and described applications suggest to make use of it for modeling other biological growth processes. However, to our knowledge, this has not yet been investigated for the cerebrovascular domain and applicability remains unclear for stroke tissue prediction.

While for all the above-mentioned methods their transformations to acquire a new representation are predefined, there are other models whose parameters can be estimated from samples. For instance, this could be learning a statistical distribution, like the shape and appearance of point representations (9), where the parameterization of the probabilistic distribution is learned from a training set. With a suitable representation at hand, there are several ways to machine-learn rather than fitting the input representation with its outcome into a statistical model. This often leads to more accurate results: Before Deep Learning has become popular in the last years, the medical image community had investigated Decision and Regression Forest models extensively, and they have shown good performance over statistical linear models or boosting approaches (10). However, these methods rely on previously specified or separately learned feature representations that need to be extracted from the image data first.

Opposite to the prior definition of the representations used, one can also machine-learn the representation *and* the classification (or regression) both at once using non-linear artificial neural networks that are capable of learning sufficiently complex models without the need of tuning the right parameterization by hand. The review paper of Lee et al. (11) shows that the methods used with deep learning are still new to the field of stroke imaging and analysis. Some attempts with other machine learning methods have been made for diagnosis and prognosis, however, those models usually predict disease scores or specific clinical outcomes but not tissue outcome.

2.2. Deep Learning for Stroke Imaging

Deep Learning with artificial neural networks is based on the idea of perceptrons where the output of a perceptron is computed by the weighted sum of its inputs x followed by an activation function σ (e.g., rectifier as in Equation 1). The power of such networks has been proven in the early work of Hornik et al. (12) by the fact that even a single hidden layer perceptron network with a proper activation function is capable of approximating any mathematical function. However, estimating that relation between input and output requires a lot of data and proper regularization since we have an underdetermined system when

learning the coefficients w_i (neuron weights for incoming connections), otherwise.

$$z = \sigma\left(\sum_i w_i x_i\right), \quad \sigma(y) = \max(0, y) \quad (1)$$

Although known for a long time, training and regularization of such networks is difficult. As a consequence of this, there have only been few attempts to utilize them for spatial data such as medical images, e.g., as proposed by Huang et al. (13) for predicting tissue fates of stroke on acute image data, but their performance could not be tweaked to exceed other former state-of-the-art approaches. For image data, the breakthrough came with deep (i.e., many layers) CNNs automatically learned through the back-propagation algorithm. Their layers form a feature hierarchy of increasingly complex features detected by the single layers through convolving the input of shared-weights kernel neurons, which themselves can be simulated by perceptrons. Interspersing pooling layers with spatial strides allows to learn texture or in general global features of the input images. See Schmidhuber (14) for an accurate explanation of these principles.

One of the first approaches modeling stroke tissue outcome with a deep learning CNN has been presented by Stier et al. (15). They trained a 2D-patch-based architecture with respect to the T_{max} feature from MR perfusion observed for acute ischemic stroke patients and a follow-up segmentation on FLAIR about 4 days later. The patch-based method clearly outperformed voxel-based approaches. As with other typical *black box*-like deep learning models, there are no further hyperparameters or constraints that can be set to control the prediction, e.g., for estimating the effect of time.

There are two major challenges in deep learning for stroke analysis tasks with regard to the data: First, there exists a general lack of accessible medical (ground truth) data and, second, the data is of irregular temporal nature. That makes it difficult to apply regular sequence models, such as Markov chains, recurrent neural networks, or the Long-Short-Term-Memory of Hochreiter and Schmidhuber (16). The data is temporally scattered: The points in time t_{Onset} , $t_{Imaging}$, and $t_{Treatment}$ are sampled as patients rush into the hospital's emergency room and cannot be collected in a regular manner.

2.2.1. U-Net Architecture

The U-Net architecture of Ronneberger et al. (17) has been successfully and widely used for biomedical applications by producing semantic segmentations through a fully-convolutional CNN (18) that additionally incorporates skip connections between the context encoding and the refining decoding path for each scale level. The encoding-decoding pattern has established well especially for fully-convolutional networks and is also known from auto-encoders, as used in our proposed method of this paper. Considering different scales is usually a good approach to capture context and details, and this works already well with just two pathways as in the DeepMedic architecture of Kamnitsas et al. (19) who won the sub-acute ischemic stroke lesion segmentation task of the first ISLES challenge in 2015 (20).

At the 2017 edition of the ISLES challenge (21) we presented a robust network on perfusion image data to predict an average lesion outcome and ranked second overall for the binary segmentation output. Many of the top-ranked methods exploited a U-Net architecture, such as the challenge winner (22) who used a 3D U-Net within an ensemble along with other networks and focused on its hyperparameter optimization. In our 2D network instead, we added further skip connections within the encoding path to enhance sensitivity in particular for the difficult smaller lesions in comparison to a standard U-Net (23).

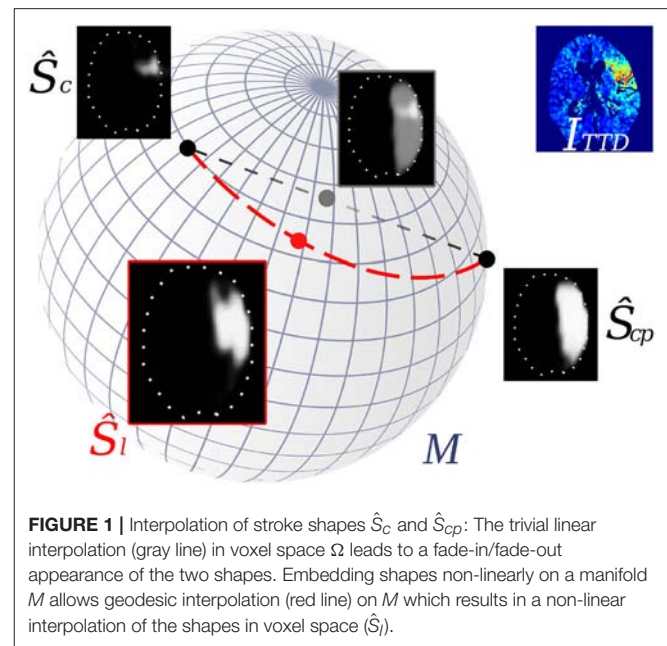
We did not observe advantages when providing clinical variables (e.g., disease scores, time points) as constant input features along with the perfusion images to predict the follow-up lesion, although they are known to be good predictors for the outcome. In fact, the 2D U-Net performance on the ISLES data could not benefit from the additional information and so we only used spatial perfusion maps to train on. The visual results suggested that rather the robust image features for detecting some highly probable necrotic stroke tissue were learned (cf. also our experiments later in this paper: **Table 4**, **Figure 9**). This makes such a network suitable for segmenting present perfusion lesions, but requires a new strategy to make use of clinical variables for predicting follow-up lesions. It remains an open question in literature, how to ideally incorporate clinical data in a U-Net-only architecture.

2.2.2. Biomedical Shape Regularization

Auto-encoders (AE) are one of several unsupervised methods to learn meaningful features from a data representation by typically encoding the input data $x \in \Omega$ into a lower-dimensional representation (*bottleneck*) and decoding this representation to get the reconstruction $z \in \Omega$ of the input x (Equation 2). This can be achieved through classical fully-connected layers or also by shared-weights convolutional layers for image data. In a convolutional auto-encoder (CAE) as introduced by Masci et al. (8) encoder $E(x)$ usually consists of a typical convolutional feature hierarchy (akin to CNNs) that results in a discriminative latent code $y \in M$, which could be a feature vector or map. Decoder $D(y)$ computes a reconstruction z back in input space Ω . During training, the weights of both are optimized such that a loss $L(x, z)$, e.g., the mean squared error $\frac{1}{n} \sum_n (x - z)^2$ for n training samples, is minimized. If used with volumetric segmentations, one can learn shape embeddings on a low-dimensional manifold M (**Figure 1**) with its dimensions representing some main modes of the shapes by optimizing the CAE:

$$z = (D \circ E)(x). \quad (2)$$

The principle of shape-constrained segmentation learning was proposed by Ravishankar et al. (24), whose cascaded architecture includes a U-Net and a CAE for shape regularization. While the U-Net follows the same encoder-decoder principle like the convolutional auto-encoder, it does not learn local geometry and shape but produces rather noisy predictions through its skip connections that *skip* its inner bottleneck. The authors combine both sub-tasks of segmentation and reconstruction in an overall loss to utilize the anatomically regularizing bottleneck of the



auto-encoder for completing noisy kidney segmentations, which improves the segmentations by about 5% compared to U-Net only.

Oktay et al. (25) presented an anatomically constrained neural network (ACNN) approach to also incorporate shape constraints of anatomical labels as prior knowledge. Their generic training scheme can be applied to various image analysis tasks and was documented for image segmentation and super-resolution. By using a CAE that is trained on ground truth shapes, they constrain the predicted image segmentations to lie close to the learned latent representation of the ground truth. In the end the decoder produces an anatomically constrained reconstruction of the segmentation from the learned shape space, because the segmentation has indeed been forced during training to lie close to the anatomy shape ground truth.

2.3. Our Contribution

In this paper – based on the robust results that U-Nets achieve on perfusion imaging data and the shape-constrained network idea of Ravishankar et al. (24)—we present a novel methodology that:

1. Utilizes a 3D U-Net and constrains its segmentations on the CBV and TTD maps through a 3D CAE;
2. Enables continuous interpolation within the non-linear and low-dimensional embedding of both core and core+penumbra segmentations according to the time-to-treatment, that results in a shape-constrained prediction of the final lesion;
3. Empirically demonstrates—quantitatively and visually—the feasibility of predicting the final lesion shape from core and core+penumbra segmentations, as well as the advantage over using an unconstrained CNN or linear interpolations in image space.

3. METHODS

Our idea is to estimate a time-to-treatment-dependent tissue outcome based on CBV and TTD perfusion images. We hypothesize that the minimum and maximum extents of the potential final stroke lesion can be approximated by delineating the core and penumbra area on the perfusion maps. This includes the following assumptions:

1. The CBV is sufficient for segmenting the core area, and TTD for core+penumbra.
2. The final lesion extents will not exceed the area boundaries of the tissue-at-risk outlined by the penumbra, but continuously evolve from the necrotic parts of the core in the direction of the outer penumbral boundary.
3. Growth over time is conducted linearly in the non-linear shape space.
4. We set a limit of 10 h after stroke onset for the infarct progression to reach the final size of core+penumbra. This value is chosen by experience and technically has to contain the maximum time-to-treatment from all training samples (maximum in our evaluation dataset: 7 h after onset).

It should be noted that training a model normalized to the acute stroke phase time range of 24 h is possible (cf. results of our evaluation later in this paper: **Table 3**, **Figure 8**) and recommended, if enough follow-up lesion data is available to sample roughly the entire space between core (0 h) and core+penumbra (24 h) representations to avoid areas of uncertainty.

3.1. Architecture

The method consists of a two-phase neural network that combines three main components for automatic shape-constrained follow-up lesion prediction (**Figure 2**):

1. U-Net estimating core (\hat{S}_c) and core+penumbra (\hat{S}_{cp}) from CBV and TTD maps.
2. CAE transforming the segmentations into a shape space and back.
3. Linear interpolation in the shape space to predict the follow-up lesion (\hat{S}_l).

First, the perfusion images I_{CBV} and I_{TTD} are processed by a U-Net U to compute the segmentation estimates \hat{S}_c and \hat{S}_{cp} . Second, the encoder E of the CAE transforms each segmentation into a low-dimensional shape embedding \hat{y}_c and \hat{y}_{cp} of a shape space that must be learned beforehand. Linear interpolation (\hat{y}_l , Equation 3) between the latent core and core+penumbra codes \hat{y}_c and \hat{y}_{cp} is conducted according to the expected $t_{\text{Imaging} \rightarrow \text{Treatment}}$ time, which must be normalized by the remaining time to reach 10 h after onset (corresponding to the total core+penumbra).

$$\hat{y}_l = \hat{y}_c + \eta(\hat{y}_{cp} - \hat{y}_c), \quad \eta = \frac{t_{\text{Imaging} \rightarrow \text{Treatment}}}{10 - t_{\text{Onset} \rightarrow \text{Imaging}}} \quad (3)$$

This linear interpolation in shape space is crucial, as it corresponds to a non-linear interpolation of the reconstructed

shapes on the manifold (**Figure 1**). The decoder D of the CAE is required to compute that reconstruction of the interpolated code \hat{y}_l in the voxel space with a segmentation \hat{S}_l for the final lesion as result (**Figure 3** illustrates a binarized 3D segmentation).

3.1.1. Cascading Networks for Prediction

The construction of a two-phase network targets different sub-tasks that constrain the learning of the high-dimensional and complex overall task of follow-up lesion prediction. When discriminating only the final lesion binary label from background through high-dimensional multivariate input, it makes it hard for a machine learning algorithm to properly generalize the relation between input and outcome. Instead of directly estimating a follow-up segmentation from the input data, we guide the first sub-network (U-Net) to segment core and penumbra correctly, which is—as explained before—of major importance for predicting final lesion tissue outcome. Once this data is provided, the second sub-network (CAE) can learn the most salient shape features on a rather simplified representation with respect to the task (shape probability maps vs. different physiological CTP parameters) along with clinical data to estimate the follow-up lesion.

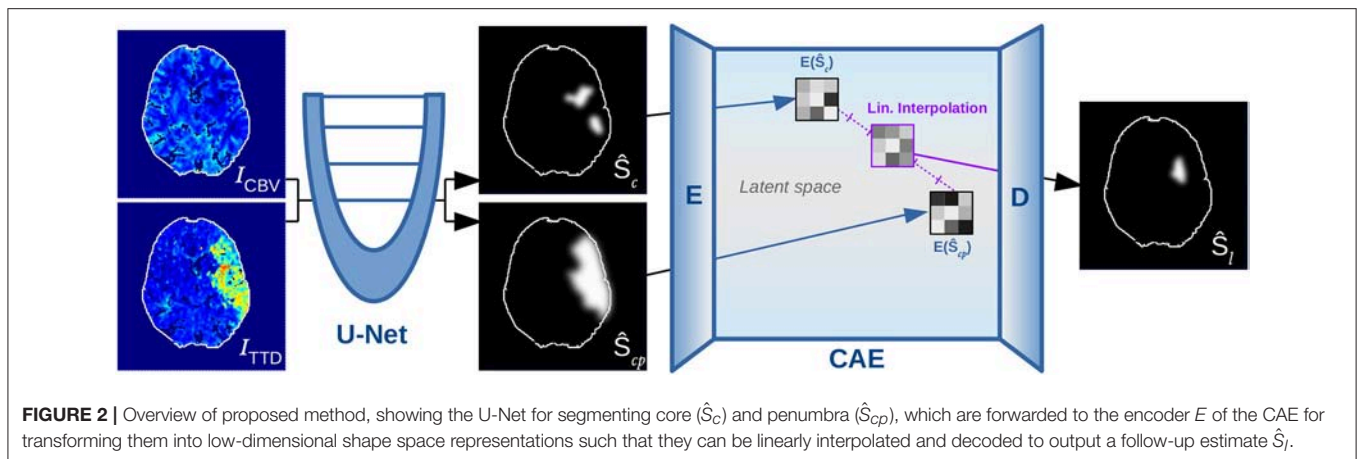
3.1.2. U-Net

Instead of taking the 2D U-Net that we have used before at the ISLES 2017 challenge, we employ a smaller standard 3D U-Net U to reduce computational and memory demand while it can cope well with the three-dimensional nature of stroke volume data. Furthermore, instead of forwarding the full $128 \times 128 \times 28$ input CTP images the U-Net is trained on randomly positioned cubic patches of size $64 \times 64 \times 28$ (with additional padding of 20 voxels in each direction) and thus needs to forward 4 patches for segmentation of one single case. It receives patches from I_{CBV} and I_{TTD} as input and estimates $\hat{S}_c, \hat{S}_{cp} = U(I_{CBV}, I_{TTD})$. The U-Net is build of double-convolutional blocks as known from Ronneberger et al. (17), while each $3 \times 3 \times 3$ convolution is preceded by a batch normalization layer of Ioffe and Szegedy (26) for data whitening. The blocks are spread over three resolution levels (two Max-Poolings) with 16, 32, and 64 channels, respectively. This sums up to a total of about 355.000 network parameters.

3.1.3. Convolutional Auto-Encoder

Focusing on the subsequent CAE, we had to ensure some minimum number of layers (**Figure 4**) to detect the most salient and descriptive abstract stroke shape features while being well-regularized in order to reconstruct a good general estimate of the shape. This requires a bottleneck layer between E and D to produce low-dimensional latent codes that must have a limited but sufficient dimensionality. Consequently, only the main modes should be represented in the code with their major variances describing different stroke shapes without noise or overfitting of training samples. Regarding overfitting, our linear approach for the interpolation on the low-dimensional codes is in principle also robust against noise in time.

The input of the CAE is forwarded akin to the U-Net through double-convolutional blocks including batch normalization



layers. Instead of two Max-Pooling operations, three additional 2-stride convolutional layers intersperse those blocks, while the final block is a single $3 \times 3 \times 3$ convolution that convolves the feature map to the $10 \times 10 \times 1$ bottleneck size. Like for the U-Net, $3 \times 3 \times 3$ filter kernels throughout all convolutional layers of the CAE, as introduced by Simonyan and Zisserman (27), are exclusively used to decrease complexity compared to networks with bigger kernels while the same receptive field sizes can be reached by just stacking more layers of the smaller kernels. This results in more non-linearities and usually generalizes better.

The decoder is a mirrored encoder, with deconvolutional layers replacing the 2-stride convolutions. With this architecture, a single $128 \times 128 \times 28$ shape segmentation image can be encoded into a low-dimensional code representation, and then decoded back to a segmentation image. Note that although the U-Net is fed with both CBV and TTD images as separate channels at once, the CAE's encoder has to forward their segmentations independently to get two separate latent codes, which will be then interpolated and forwarded in one step through the decoder to get the final lesion prediction.

Since there are several global scalar predictors (time, age, sex, clinical scores) that might be required to model the space properly, we also tried to map the shape segmentation input of the CAE to a vector representation (instead of spatial feature maps) in the bottleneck by using a fully-connected or convolutional layer with appropriate kernel size. This would have allowed us to add an arbitrary number of scalar values directly as additional dimensions to the latent code in the bottleneck and to easily quantify the dimensionality of such shape representations in the latent space. However, regularization is difficult – even when using dropout (28)—and the reconstructions are less accurate than with latent spatial feature map representations. Therefore, the CAE remains convolutional-only and no clinical variables other than the combination η of both time predictors for the interpolation are used as per definition in Equation (3).

3.2. Training

The training is conducted in three consecutive steps (see **Figure 5** for illustration of steps 2 and 3) that are characterized by different objectives formulated in their corresponding losses:

1. Training the U-Net with a $L_{SoftDice}$ loss.
2. Training E_1 and D to learn the shape space of the CAE from manual segmentations (L_{Shape})
3. Training E_2 to fit automatic segmentation estimates of U into the shape space ($L_{Prediction}$)

The U-Net is initially trained beforehand using the stochastic gradient variant ADAM of Kingma and Ba (29) for optimization. The ground truth segmentations S_c and S_{cp} of core and core+penumbra are used to penalize their predictions with a bigger loss for less overlap in the SoftDice measure, which is defined for all voxel positions i in a segmentation A with ground truth B and a small constant ϵ as follows:

$$SoftDice(A, B) = \frac{2 \cdot \sum_i (A_i B_i) + \epsilon}{\sum_i (A_i A_i + B_i B_i) + \epsilon} \quad (4)$$

3.2.1. Shape Space Learning

First, a low-dimensional shape space is learned that embeds the ground truth shapes of core and core+penumbra segmentations, S_c and S_{cp} . This is enforced by loss L_{Shape} in Equation (5), that consists of three parts: (1) Reconstruction R_c , R_{cp} , and R_l of core, core+penumbra and final lesion shape, (2) the property of the reconstructed core/lesion volume to still be a subset of the core+penumbra volume, and (3) a $L1$ loss for keeping the latent code y_l of the lesion shape close to the linear interpolation y_i of the core and core+penumbra codes:

$$L_{Shape} = \sum_{s \in \{c, cp\}} L_{SoftDice}(R_s, S_s) + \sum_{s \in \{c, l\}} L_{mono}(R_s, R_{cp}) + \alpha L1(y_l, y_i) \quad (5)$$

For the first 25 epochs $\alpha = 0$ holds, otherwise $\alpha = 1$. We observed that this helps the non-convex optimization function to first learn to reconstruct the shape into the correct brain hemisphere and was found to robustly prevent the network from getting trapped in implausible minima at the beginning of the training. Since $SoftDice \in [0, 1]$ with 0 indicating non-overlap and 1 for full overlap, we need to define

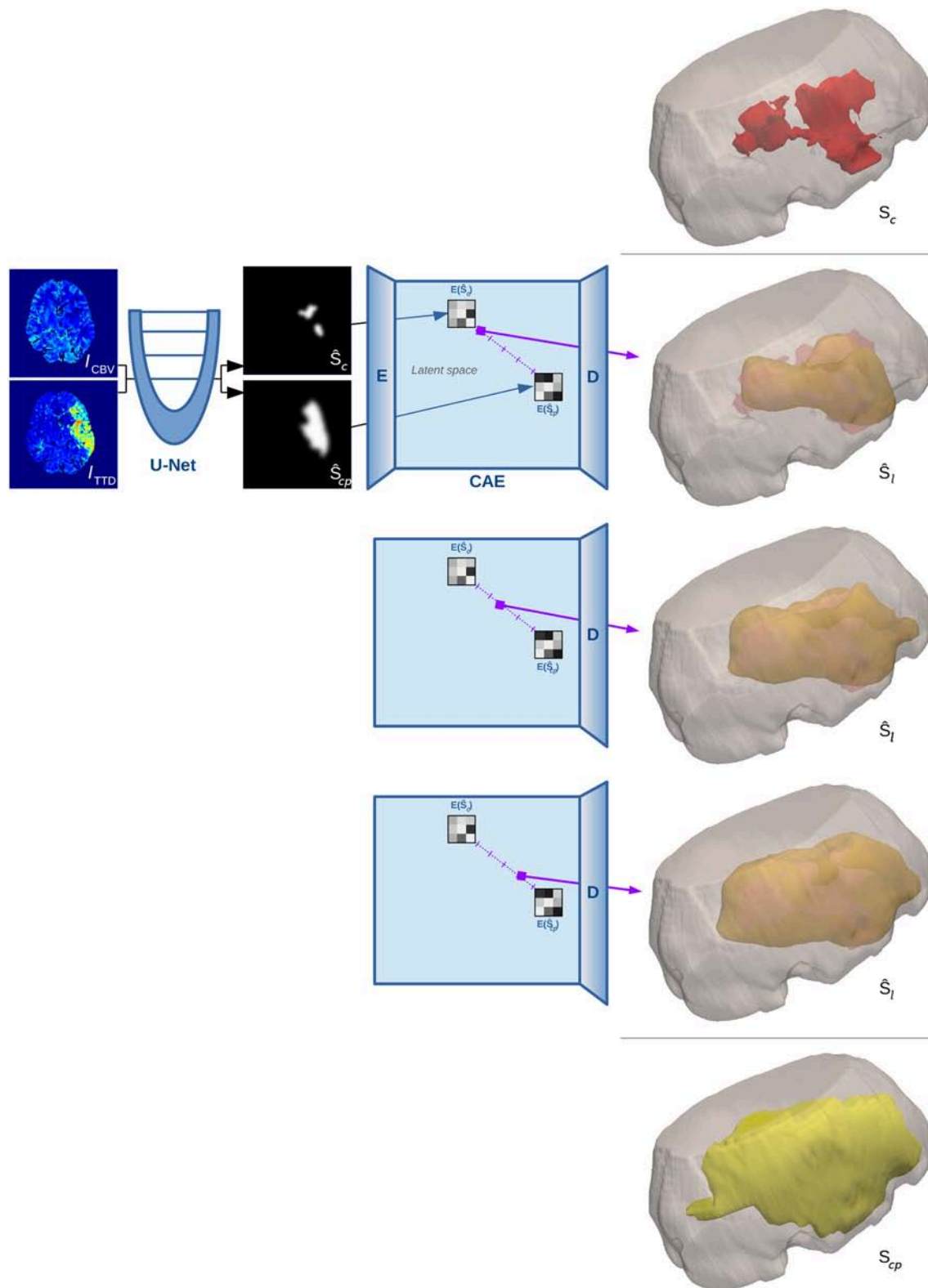
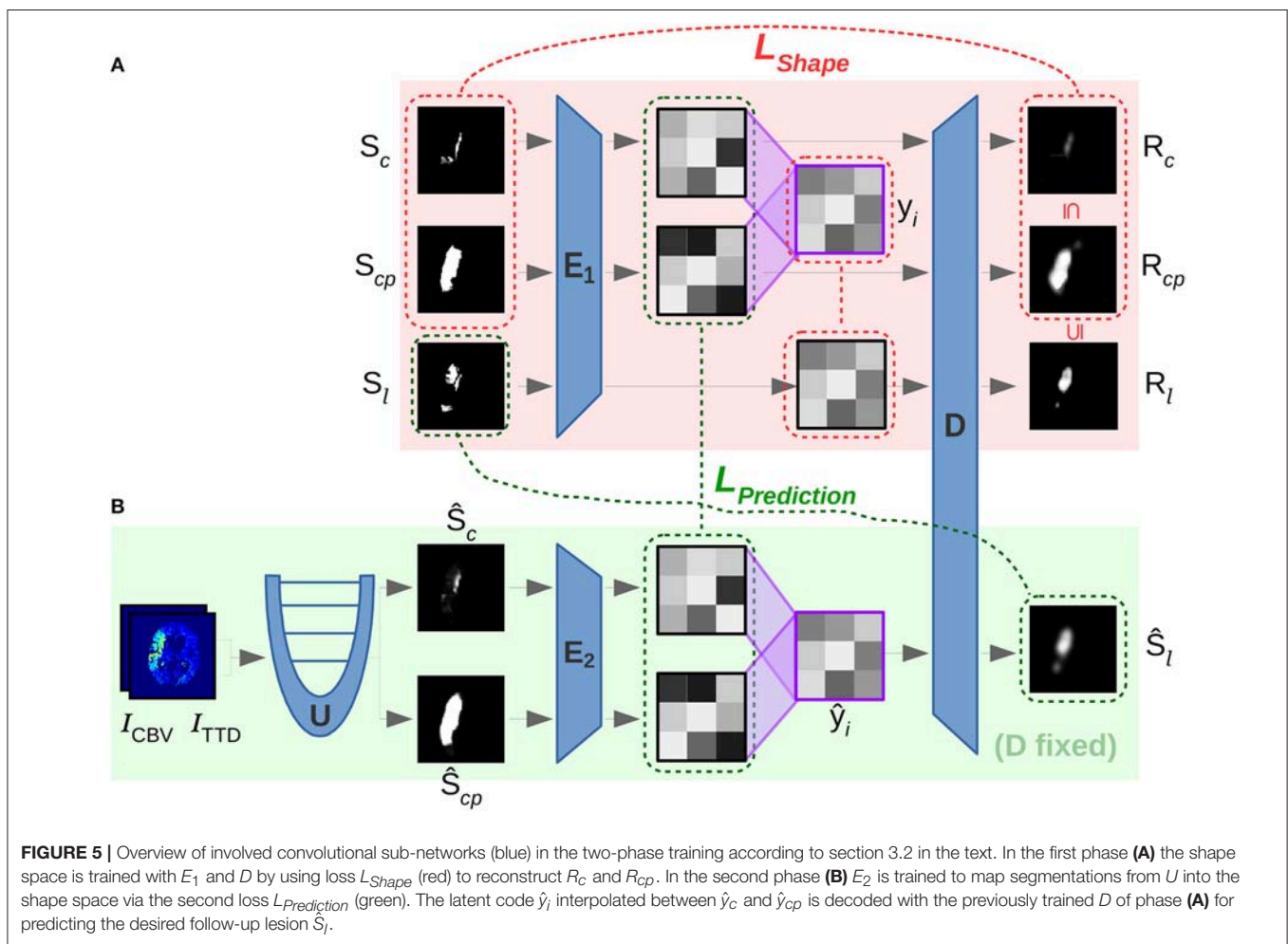
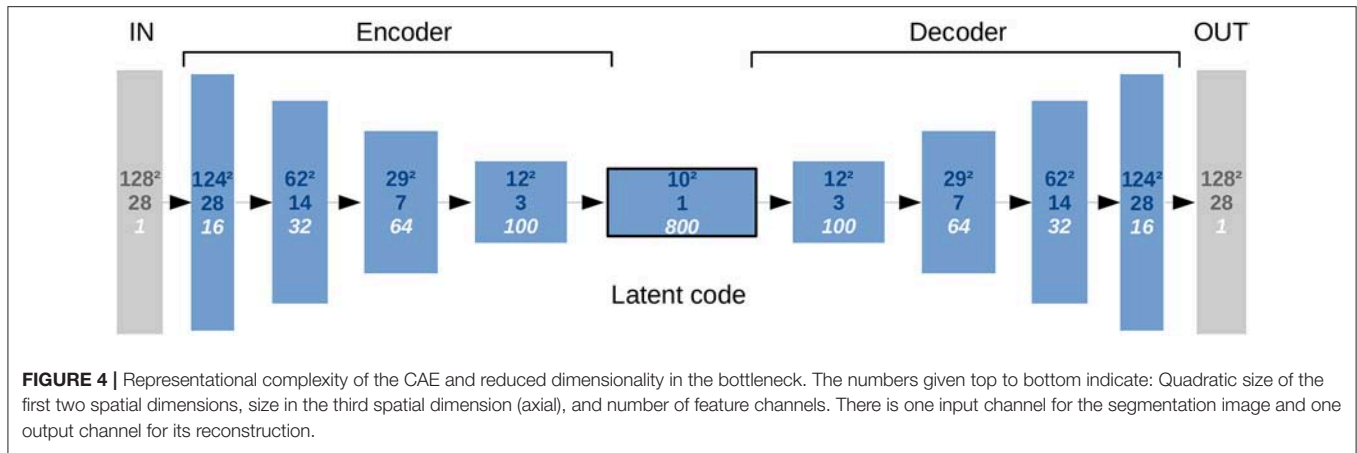


FIGURE 3 | 3D surface illustration for the binarized reconstruction of the interpolation in shape space for a single test case from our experiments. The top image shows the manual segmentation of core in red within a cutout of the brain volume, while the bottom image shows the ground truth segmentation of penumbra in yellow. Three steps of the interpolation between the estimates of both manual segmentations are shown in the middle rows.



$L_{SoftDice}(S, \hat{S}) = 1 - SoftDice(S, \hat{S})$. To force the CAE to learn that the interpolation is only growing when interpolating along the (time) trajectory from core until reaching the core+penumbra segmentation at maximum, we define a constraint L_{mono} for all voxel positions i in two segmentation images A, B so that the reconstructions of the core segmentation and the

intermediate lesion interpolation are monotone increasing to the total core+penumbra segmentation:

$$L_{mono}(A, B) = \sum_i \max(A_i - B_i, 0) \quad (6)$$

3.2.2. Noisy Shape Interpolation

In the second training phase, the encoder and decoder weights from the shape space learning phase before will be fixed. A second encoder is then learned for the U-Net predictions \hat{S}_c and \hat{S}_{cp} of core and core+penumbra to compute latent representations \hat{y}_c and \hat{y}_{cp} that are located close to the shape embeddings y_c and y_{cp} of the corresponding ground truth segmentations in terms of L1 norm. $L_{Prediction}$ in Equation (7) further enforces the monotone properties for the reconstructed segmentations and the main goal of high overlap for the prediction \hat{S}_l decoded from the interpolated code \hat{y}_i to the actual follow-up ground truth S_l :

$$L_{Prediction} = L_{SoftDice}(\hat{S}_l, S_l) + \sum_{s \in \{c, l\}} L_{mono}(\hat{S}_s, \hat{S}_{cp}) + \sum_{s \in \{c, l, i\}} L1(\hat{y}_s, y_s) \quad (7)$$

This way, the decoder D of the first phase ideally decodes an approximate representation from the shape space so that the reconstruction of the core and core+penumbra estimates should be close to the ground truth reconstruction of core and core+penumbra, $(D \circ E_2)(\hat{S}_c) \approx (D \circ E_1)(S_c)$ and $(D \circ E_2)(\hat{S}_{cp}) \approx (D \circ E_1)(S_{cp})$. Moreover, the main goal remains to achieve an interpolation as close as possible to the true lesion segmentation:

$$D\left(\eta \cdot E_2(\hat{S}_{cp}) + (1 - \eta) \cdot E_2(\hat{S}_c)\right) \approx (D \circ E_1)(S_l) \quad (8)$$

The individual loss terms have not been weighted by further manual parameters as we found no benefit for other than the uniformly weighted individual loss parts for both L_{Shape} and $L_{Prediction}$. Further, we tried to apply both losses beforehand in an alternating and joint manner but this did not let the optimizer find proper minima; in particular, learning the exclusive occurrence of a stroke on either hemisphere could not be learned, which is what the network basically learns during the first epochs of training.

4. EXPERIMENTS

We run a 5-fold test on a 29 subjects dataset, because the time demand for a full cross-validation was too high. In order to test each of the 5 folds, we thus had to train 5 models with the remaining 4 folds. Four of the folds consist of 6, and 1 fold consists of 5 cases. About one fourth of the training samples were used as validation set, so that a training set consists of 17 or 16 cases and is validated in each epoch on 6 other cases (disjoint with the test fold).

We prevent overfitting of the model by training until the validation loss converges and choose the model with the lowest validation loss. Since the model is eventually tested on a different fold of patients not used for the training and validation, we also avoid that the evaluation results could be tuned on the validation loss optimum. Due to the huge number of parameters in our 3D sub-networks, the memory demand for gradient computation increases rapidly, so a batch size of 4 had to be chosen to fit the training data into 11 GB of GPU memory.

TABLE 2 | Characteristics for subjects with manually segmented core, penumbra and follow-up lesion of the retrospectively collected data.

Baseline characteristics	Value
Subjects n	29
Male sex, n (%)	16 (55%)
Age, years, median (IQR)	70 (63 – 77)
Admission NIHSS, median (IQR)	15 (12 – 16)
Core volume, ml, median (IQR)	27 (3 – 86)
Penumbra vol., ml, median (IQR)	164 (149 – 199)
Lesion (FU) vol., ml, median (IQR)	31 (18 – 93)
$t_{\text{Onset} \rightarrow \text{Imaging}}$, hours, median (IQR)	1.7 (1.4 – 3.4)
$t_{\text{Imaging} \rightarrow \text{Treatment}}$, hours, median (IQR)	1.7 (1.4 – 2.1)
$t_{\text{Onset} \rightarrow \text{Treatment}}$, hours, median (IQR)	3.9 (3.2 – 4.9)

4.1. Data

We used a dataset of 29 subjects from the Neuroradiology department at the University Hospital Schleswig-Holstein formerly collected for the TRAVESTROKE project for which one rater had created manual segmentations on the CT perfusion (CTP) modalities CBV (Cerebral Blood Volume) and TTD (Time-To-Drain) at the time of admission for core and core+penumbra, as well as a lesion segmentation on follow-up CT after treatment. The data was acquired with a Siemens Somatom Definition AS 40 (Siemens Healthcare GmbH Forchheim, Germany) and the raw data was deconvolved using the vendor algorithm to get CT perfusion parameters such as CBV or TTD. All patients of the dataset had been treated successfully with thrombectomy (TICI score 2b or 3). See baseline characteristics of the subjects included in the evaluation in **Table 2**.

The dataset was pre-processed with FSL-FLIRT (30) for affine registration to correct tilted heads and transform them into common space. A downsampling was applied so that input size of the CBV and TTD maps was $128 \times 128 \times 28$ voxels for reducing the computational demand. Additional clinical data for each subject was given. For the evaluation we only used the two durations $t_{\text{Onset} \rightarrow \text{Imaging}}$ and $t_{\text{Imaging} \rightarrow \text{Treatment}}$, which were normalized according to Equation (3). All shape segmentations have been elastically deformed in each epoch to augment the limited available training data for learning generic features of the CAE such that the representations in shape space can be robustly reconstructed.

4.2. Comparison

In section 2.2.1, we referred to the ISLES 2017 task of predicting a follow-up lesion based on MR perfusion data. We participated in this challenge with a single U-Net directly predicting the final follow-up lesion, as many of the teams were using unconstrained CNNs [as presented in (21)]. Since this does not lead to accurate predictions of a progressed stroke when facing acute image data, we compare our proposed method with this simple U-Net approach with and without clinical time

TABLE 3 | Reconstruction (1. training phase) results: Average Dice values on the test data for the CAE trained on core (S_c) and core+penumbra (S_{cp}) expert segmentations, as well as average Dice overlap of the follow-up expert lesion segmentation (S_l) with the reconstructed interpolation from shape space (R_l).

Method	Method	# Parameters		Dice	
Core/Penumbra	Lesion (Training phase)	of network	Core (R_c)	Core+Penumbra (R_{cp})	Lesion (R_l)
Expert	Oracle CAE 10 h (1. phase)	$4.7 \cdot 10^6$	0.68	0.90	0.53
Expert	CAE 10 h (1.)	$4.7 \cdot 10^6$	0.68	0.90	0.46
Expert	CAE 24 h (1.)	$4.7 \cdot 10^6$	0.70	0.90	0.44

While the first row shows an oracle prediction overlap for the theoretically best-fit per case interpolation within our model (could be before or after the true time-to-treatment), the second row lists the result for our proposed approach using the actual time-to-treatment ground truth to predict the follow-up lesion from ground truth. The third row indicates that the effect of choosing a different normalization value from within the time range of the acute stroke phase can rather be neglected.

TABLE 4 | Experimental results from 5-fold test data based on the U-Net's \hat{S}_c , \hat{S}_{cp} , and lesion estimates: The average values for the Dice overlap with the ground truth segmentations S_c , S_{cp} , and S_l are presented.

Method	Method	# Parameters		Dice	
Core/Penumbra	Lesion (Training phase)	of network(s)	Core	Core+Penumbra	Lesion
U-Net	CAE (2. phase)	$3.6 \cdot 10^5 + 4.7 \cdot 10^6$	0.55	0.81	0.43 (\hat{S}_l)
U-Net	CAE (1. phase)	$3.6 \cdot 10^5 + 4.7 \cdot 10^6$	0.43	0.80	0.40 (\hat{S}_l)
U-Net	Image Interpolation	$3.6 \cdot 10^5$	0.45 (\hat{S}_c)	0.81 (\hat{S}_{cp})	0.36
–	U-Net 2in	$3.6 \cdot 10^5$	–	–	0.34
–	U-Net 4in	$3.6 \cdot 10^5$	–	–	0.22

The baseline U-Net-only approach has the same architecture as used in combination with the CAE above, but instead of predicting core and penumbra (2 channel output), it directly predicts the follow-up lesion (1 channel output) as described in section 4.2. We input CBV and TTD maps only (2in), or additionally both $t_{\text{Onset} \rightarrow \text{Imaging}}$ and $t_{\text{Imaging} \rightarrow \text{Treatment}}$ as constant image channels (4in), whereas the latter performed worse coincident with our previous experience (see section 2.2.1 and **Figure 9**). The highest overlaps for core, core+penumbra, and lesion are highlighted in bold.

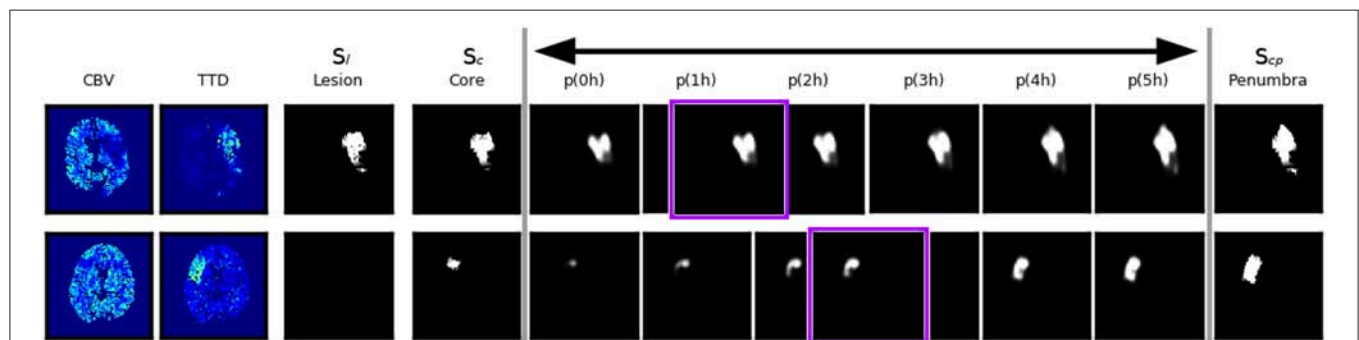


FIGURE 6 | Two example cases with good and bad performance from the validation set at the end of the shape space learning (1. training phase) and their interpolation at different $t_{\text{Imaging} \rightarrow \text{Treatment}} \in [0, 5]$ values. The first row shows an axial slice of an accurate reconstruction of core at 0h as well as of the penumbra, and as a consequence the non-linear interpolation of both gives a reliable estimation of the true follow-up lesion at its actual $t_{\text{Imaging} \rightarrow \text{Treatment}}$ value (outlined in purple). The second row shows a case where the temporal progression of the stroke is different, such that the final lesion was much smaller than an interpolation between core and penumbra would suggest.

points as input. Unfortunately, the ISLES dataset consists of MR perfusion data without appropriate core and penumbra segmentations, so we cannot directly compare on the same dataset.

Our sub-task of linearly interpolating along the trajectory between core and core+penumbra requires representations of such in a suitable non-linear shape space that has to be learned before. In order to show the advantage of conducting this in such a shape space, we compare with the naïve way of linearly interpolating the shape segmentations in Ω . This can be simply

computed with the same η as defined before in Equation (3):

$$\hat{S}_l = \hat{S}_c + \eta(\hat{S}_{cp} - \hat{S}_c) \quad (9)$$

5. RESULTS

The reconstruction results (**Table 3**) demonstrate the capability of our learned model to make time-dependent predictions and present the advantages of our CAE approach using core

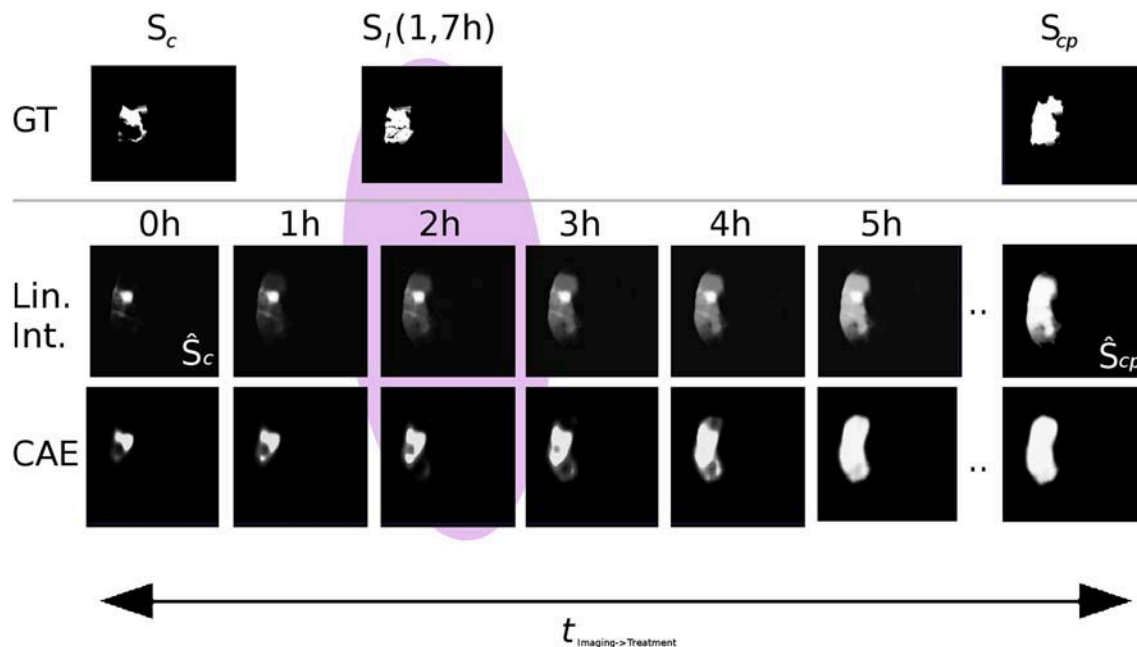


FIGURE 7 | Visual comparison of linear interpolation in image and shape space for an axial slice of a single case from the 5-fold test (Ground truth denoted as GT). Compare the predictions at 2h with the actual follow-up at 1.7h: With linear interpolation of the segmentations, the core area \hat{S}_c remains unchanged while the huge penumbra area is faded in. Note that for a normalization of 24h, the fading over time would progress even slower. With the CAE and its interpolation in shape space, the shape grows non-linearly, first locally, and then quickly into the surrounding tissue-at-risk areas segmented by the U-Net in \hat{S}_{cp} .

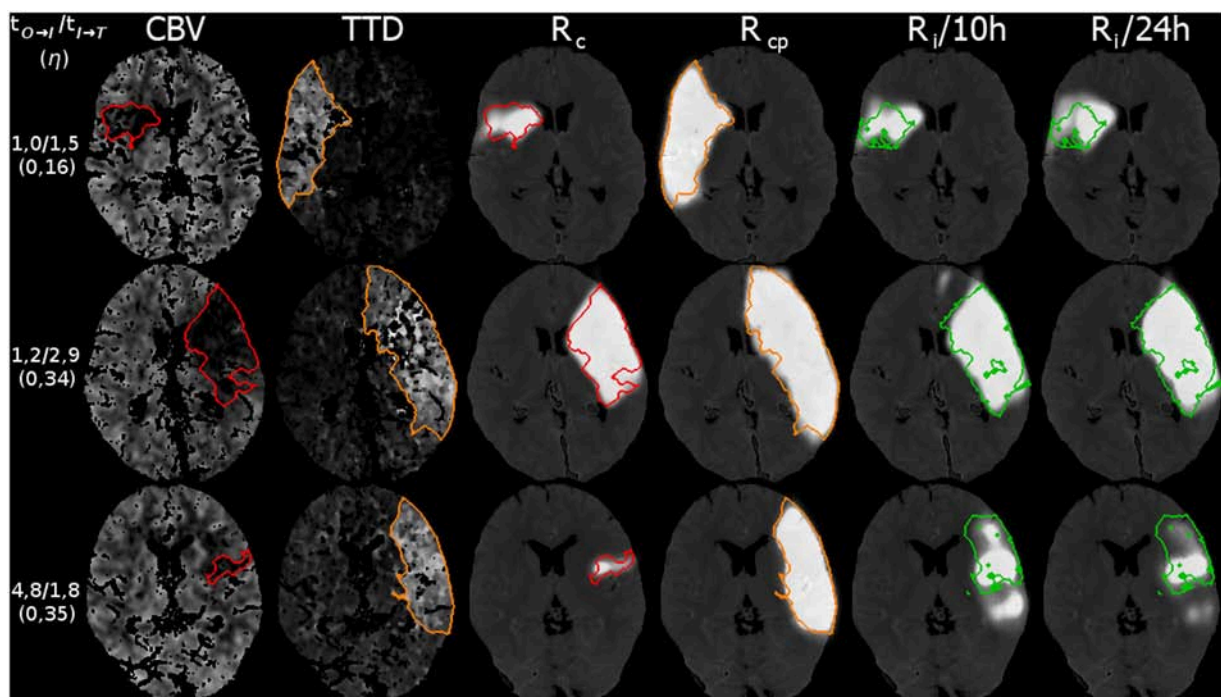


FIGURE 8 | Three example cases (as in Figure 9) from the 5-fold test data with their results for the reconstruction. The top row shows a case with fast admission and treatment times, where the lesion appears about the same size as the initial core. The middle and bottom row are interpolated at η located one third on the trajectory from core to core+penumbra, while one case is an early and the other is a late admission; However, the case with bigger core progresses slowly with respect to the non-linear reconstruction compared to the smaller core case at the bottom. Reference overlays for manual segmentations S_c (red), S_{cp} (orange), and S_i (green) are shown as outlines. Note that models trained with a normalization of 10 or 24 h output similar predictions.

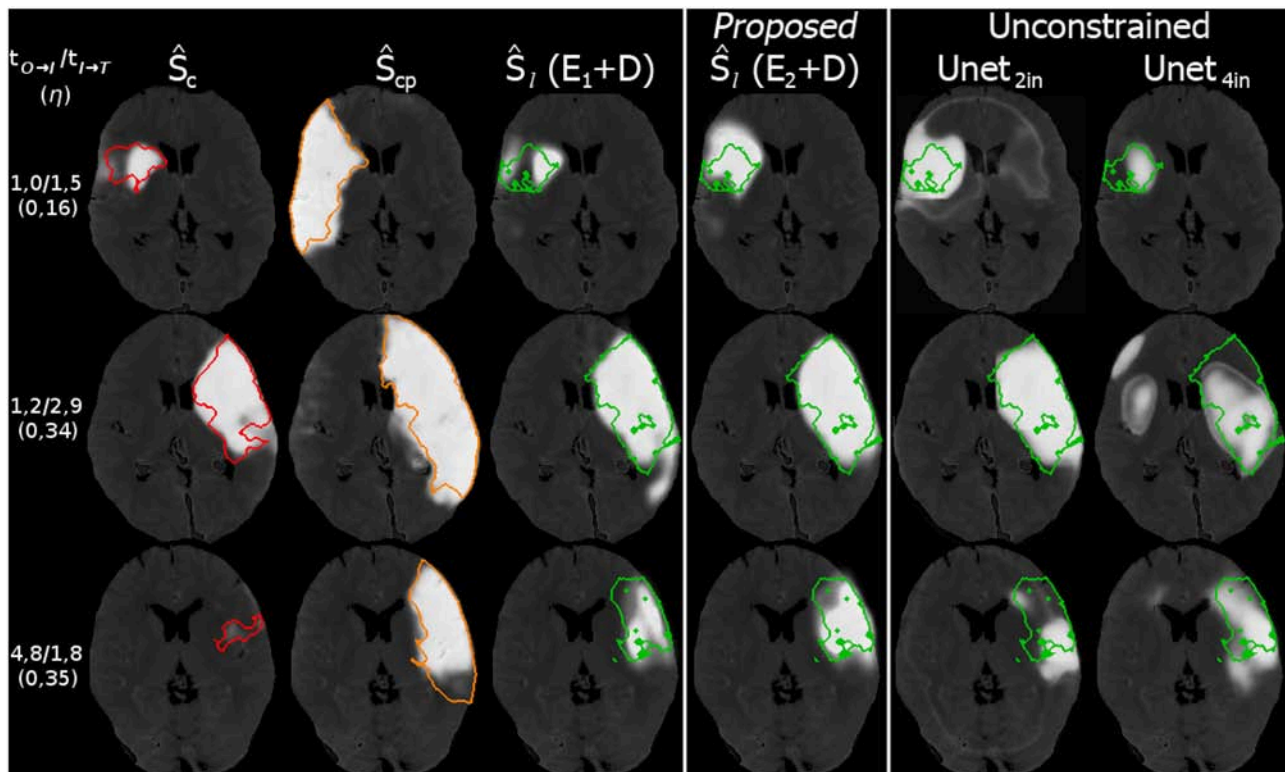


FIGURE 9 | Three example cases (as in **Figure 8**) from the 5-fold test data with their results for the prediction. The top row shows a case with fast admission and treatment times, where the lesion appears about the same size as the initial core. The middle and bottom row are interpolated at η located one third on the trajectory from core to core+penumbra, while one case is an early and the other is a late admission; However, the case with bigger core progresses slowly with respect to the non-linear reconstruction compared to the smaller core case at the bottom. Reference overlays for manual segmentations S_c (red), S_{cp} (orange), and S_l (green) are shown as outlines. Note that prediction with unconstrained U-Net-only variants with (4in) or without (2in) clinical input channels cannot reach the prediction performance of our proposed U-Net + CAE method.

and penumbra segmentations to generalize well even from a small training dataset to estimate non-linear follow-up lesion interpolations. A Dice overlap of 0.46 was achieved in comparison to a manual rater. Considering a reconstruction Dice overlap for the CAE itself of 0.68 and 0.90 for core and core+penumbra, respectively, this represents a good result. In a use case, where a clinical expert manually segments core and penumbra, this can already be a helpful estimate for assessing the expected treatment outcome after thrombectomy.

However, even a very good reconstruction of core and penumbra does not guarantee a good final lesion estimate (**Figure 6**, bottom row). The quality and severity of the stroke in routine clinical data is not always fully encoded in its core and penumbra shape segmentations, and some of the follow-up lesion segmentations are actually smaller than even their corresponding core segmentations contrary to the definition that core should include only necrotic tissue which cannot be recovered. Potential reasons for this include the challenges of the manual annotations based on CBV and TTD alone.

With our trained model we could determine an upper bound of 0.53 for a linear interpolation-based lesion prediction *oracle* (**Table 3**), which does not use the true time-to-treatment but

knows the correct $\eta \in [0, 1]$ that results in the best overlap with the ground truth lesion. Apart from non-linear growth over time that has been observed in literature (31) and the lack of any information from the perfusion signal, the difference between 0.46 and 0.53 could be explained by too much noise in the time data; Especially the determination of $t_{\text{Onset} \rightarrow \text{Imaging}}$ can often be quite inaccurate in clinical practise. While it would be desirable to have the best interpolation reconstructed from times as near as possible to the true time-to-treatment, the linear approach is quite robust against inaccurate times and monotone growth is enforced. The doctor is essentially interested to see if there will be much relative growth and, consequently, how much of the tissue could be salvaged within the next hours.

Given the CAE, the second encoder E_2 learns to map the segmentations from the U-Net with a high quality into the shape space during the second training phase (**Table 4**). This phase requires only 50 epochs for a convergence of the validation set loss, compared to 200 epochs in the first phase when training E_1 and D . Automatic core+penumbra segmentations achieved by our U-Net model are of very high quality (Dice score of 0.81) and close to the optimal reconstruction given the constraints of the CAE (Dice of 0.90). The segmentation of

the core is more challenging yielding Dice scores of only 0.45 which is improved by the second CAE to 0.55. This confirms the results of the shape-constrained network proposed by Ravishankar et al. (24).

Interpolating between the latent shape representations of core and core+penumbra estimated by the U-Net is less accurate than performing this with the latent representations of ground truth segmentations. Nevertheless, the advantages of our proposed U-Net + CAE architecture with a Dice score of 0.43 for the lesion are evident as the result is close to the ground truth interpolation (Dice 0.46). A significant improvement is found in comparison to the two baseline methods (0.36 and 0.34). It can be visually observed in **Figure 7** that our linear interpolation in the shape space leads to a non-linear growth of the infarct shapes: first locally, then into the outer penumbra. Contrary to that, simply interpolating linearly on segmentation predictions leads to implausible fading of the entire tissue-at-risk infarct probabilities in the image voxel space, where the rate of progression is also strictly depending on the normalization value!

Moreover, our subdivided approach clearly reveals the sub-tasks that need to be tuned in our method for improvement of the final prediction, different to a closed unconstrained model like the single U-Net. We observed that there is less overlap for the core than for both core+penumbra. By improving the core reconstruction from shape space, the interpolation trajectory would be closer to the true lesion representation. Thus, prediction for the true time-to-treatment could benefit, for both ground truth and estimated segmentations. If furthermore the core estimate could be more accurate, the closer will the latent code of \hat{y}_c be located to the representation y_c of the ground truth core, and so will the trajectory in shape space be more close to the ideal trajectory.

Compared to the results of the ISLES stroke lesion challenges of the last 2 years on MR perfusion and diffusion data, none of the participating groups has reached a higher overlap of the predicted lesion outcome with the actual follow-up than a Dice of 0.32 (see <https://www.isles-challenge.org>). With respect to the similar task and comparable functional imaging modalities, the results of our method predicting the lesion outcome on core and penumbra estimates are promising.

6. CONCLUSION

In this work we have shown the feasibility of using interpolations between low-dimensional shape embeddings of core and penumbra segmentations for improving the prediction of stroke lesion tissue outcome. First, we could show that a CAE is able to model the main variances of volumetric stroke shapes resulting in good reconstructions on test data. With the latent representation at hand, one can now continuously interpolate along robust linear trajectories in the shape space to obtain non-linear shape growth from the core to the entire penumbral area. Fed with an actual time-to-treatment point, this results in a shape-constrained estimate of the expected final lesion for the given time, making it possible to compute other measures on this

result, such as volume or density, to be of further assistance to the radiologist. Thus, our framework facilitates the assessment of potential infarct growth and possible salvageable tissue to support treatment decisions and prioritization.

With our current interpolation method we have an upper bound for the prediction Dice score of 0.53, which can be achieved on manual expert segmentations as end points for the linear interpolation in shape space when using a time-to-treatment oracle. This is nearly reached with our best performing fully-automatic model based on the actual time-to-treatment (Dice score 0.43). To improve the overall performance for the prediction by interpolating between the shape representations of automated core and penumbra segmentations according to time, we believe that time as a factor for the stroke growth will not always be in fixed linear relationship with the interpolation. First of all, there are other clinical variables that have an (combined) effect on the outcome, such as age or NIHSS (National Institutes of Health Stroke Scale) score, not yet considered in our method. Furthermore, differences in the growth rate even for similar early lesions could be found between patients. In future, we would like to investigate how an integrated approach can also learn non-linear growth over time to further close the gap from 0.43 to 0.53. Nevertheless, our method does not only strive for ideal overlap but rather robust growth over time in a plausible manner.

We observed that there are still cases lowering the overall prediction performance, where the follow-up lesion remains smaller than the core area. This cannot only be explained by different treatment outcomes or a decline in swelling, and requires a review with clinical experts on the dataset (perfusion parameters, manual segmentation protocol) as well as our hypothesis. For instance, if manual segmentations are not consistent throughout the dataset, rejecting data cases, which do not fit the hypothesis and thus make it difficult to train our proposed network according to our preconditions, could show substantial improvements in the results.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the “Antrag an die Ethik-Kommission vom 14. April 2015, Retrospektive Auswertung von CT- und MRT-Datensätzen zur Entwicklung eines Prädiktionsmodells bei Schlaganfallpatienten, Ethik-Kommission ÄK Hamburg,” with reference number AZ 15-113 (23 Apr 2015). The protocol was approved by the Ethik-Kommission ÄK Hamburg. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

DATA AVAILABILITY STATEMENT

The datasets for this manuscript are not publicly available because they were collected for institutional usage only. Requests to access the datasets should be directed to André Kemmling

(andre.kemmling@uksh.de). The latest source code is available www.github.com/multimodallearning/stroke-prediction.

AUTHOR CONTRIBUTIONS

CL conceived network architectures, interpolation and training scheme, conducted the literature review, prepared the manuscript, and partly preprocessed the dataset. MH and NB conceived the idea of auto-encoders for the core and penumbra interpolation and noted on the scheme proposed by CL. MH suggested literature for anatomical constraints in image analysis and commented on the work steadily. NB further suggested architecture variants for the interpolation problem. AK proposed the clinical task, defined the needs, and collected the dataset. LA conducted the manual segmentations and parts of the dataset preprocessing. AM gave valuable

input on the perfusion segmentation via CNNs. All authors reviewed and approved the submission of this version of the manuscript.

FUNDING

This work was partly supported by the Graduate School for Computing in Medicine and Life Sciences funded by Germany's Excellence Initiative [DFG GSC 235/2] and by DFG project 320997906. The work was also partly funded by the Lübeck Medical School within the TRAVE Stroke project.

ACKNOWLEDGMENTS

We would like to thank Nvidia Corporation for providing us with a Titan Xp graphics card.

REFERENCES

- Allen CL, Bayraktutan U. Risk factors for ischaemic stroke. *Int. J. Stroke* (2008) 3:105–16. doi: 10.1111/j.1747-4949.2008.0187.x
- El Tawil S, Muir KW. Thrombolysis and thrombectomy for acute ischaemic stroke. *Clin Med.* (2017) 17:161–5. doi: 10.7861/clinmedicine.17-2-161
- Khatiri P, Abruzzo T, Yeatts S, Nichols C, Broderick J, Tomsick T, et al. Good clinical outcome after ischemic stroke with successful revascularization is time-dependent. *Neurology* (2009) 73:1066–72. doi: 10.1212/WNL.0b013e3181b9c847
- d'Esperre CD, Boesen ME, Ahn SH, Pordeli P, Najm M, Minhas P, et al. Time-dependent computed tomographic perfusion thresholds for patients with acute ischemic stroke. *Stroke* (2015) 46:3390–7. doi: 10.1161/STROKEAHA.115.009250
- Kemmling A, Flottmann F, Forkert ND, Minnerup J, Heindel W, Thomalla G, et al. Multivariate dynamic prediction of ischemic infarction and tissue salvage as a function of time and degree of recanalization. *J Cereb Blood Flow Metab.* (2015) 35:1397–405. doi: 10.1038/jcbfm.2015.144
- Kolouri S, Park SR, Rohde GK. The Radon cumulative distribution transform and its application to image classification. *Trans Image Process.* (2016) 25:920–34. doi: 10.1109/TIP.2015.2509419
- Basu S, Kolouri S, Rohde GK. Detecting and visualizing cell phenotype differences from microscopy images using transport-based morphometry. *Proc Natl Acad Sci USA.* (2014) 111:3448–53. doi: 10.1073/pnas.1319779111
- Masci J, Meier U, Cireşan D, Schmidhuber J. Stacked convolutional auto-encoders for hierarchical feature extraction. In: Honkela T, Duch W, Girolami M, Kaski S, editors. *Artificial Neural Networks and Machine Learning – ICANN 2011*. Berlin; Heidelberg: Springer (2011). p. 52–59.
- Ehrhardt J, Krüger J, Handels H. Statistical shape and appearance models without one-to-one correspondences. In: *Proceedings of SPIE Medical Imaging*. Vol. 9034. San Diego, CA (2014).
- Maier O, Schröder C, Forkert ND, Martinetz T, Handels H. Classifiers for ischemic stroke lesion segmentation: a comparison study. *PLoS ONE* (2015) 10:e0145118. doi: 10.1371/journal.pone.0145118
- Lee EJ, Kim YH, Kim N, Kang DW. Deep into the brain: artificial intelligence in stroke imaging. *J Stroke* (2017) 19:277–85. doi: 10.5853/jos.2017.02054
- Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. *Neural Netw.* (1989) 2: 359–66.
- Huang S, Shen Q, Duong TQ. Artificial neural network prediction of ischemic tissue fate in acute stroke imaging. *J Cereb Blood Flow Metab.* (2010) 30:1661–70. doi: 10.1038/jcbfm.2010.56
- Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw.* (2015) 61:85–117. doi: 10.1016/j.neunet.2014.09.003
- Stier N, Vincent N, Liebeskind D, Scalzo F. Deep learning of tissue fate features in acute ischemic stroke. In: *2015 International Conference on Bioinformatics and Biomedicine (BIBM)*. Washington, DC: IEEE (2015). p. 1316–1321.
- Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput.* (1997) 9:1735–80.
- Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Cham: Springer (2015). p. 234–41.
- Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015). p. 3431–40.
- Kamnitsas K, Ledig C, Newcombe VF, Simpson JP, Kane AD, Menon DK, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Anal.* (2017) 36:61–78. doi: 10.1016/j.media.2016.10.004
- Maier O, Menze BH, von der Gabelntz J, Häni L, Heinrich MP, Liebrand M, et al. ISLES 2015-A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. *Medical Image Anal.* (2017) 35:250–69. doi: 10.1016/j.media.2016.07.009
- Winzeck S, Hakim A, McKinley R, Pinto JAADS, Alves V, Silva C, et al. ISLES 2016 and 2017-benchmarking ischemic stroke lesion outcome prediction based on multispectral MRI. *Front Neurol.* (2018) 9:679. doi: 10.3389/fneur.2018.00679
- Choi Y, Kwon Y, Lee H, Kim BJ, Paik MC, Won JH. Ensemble of deep convolutional neural networks for prognosis of ischemic stroke. In: *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Cham: Springer (2016). p. 231–43.
- Lucas C, Kemmling A, Madany Mamlouk A, Heinrich MP. Multi-scale neural network for automatic segmentation of ischemic strokes on acute perfusion images. In: *International Symposium on Biomedical Imaging (ISBI) 2018*. Washington, DC: IEEE (2018).
- Ravishanker H, Venkataramani R, Thiruvankadam S, Sudhakar P, Vaidya V. Learning and incorporating shape models for semantic segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Cham: Springer (2017). p. 203–11.
- Oktay O, Ferrante E, Kamnitsas K, Heinrich M, Bai W, Caballero J, et al. Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation. *Trans Med Imaging* (2018) 37:384–95. doi: 10.1109/TMI.2017.2743464

26. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: *International Conference on Machine Learning* Lille (2015). p. 448–56.
27. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556* (2014).
28. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res.* (2014) 15:1929–58.
29. Kingma DP, Ba J. Adam: a method for stochastic optimization. *arXiv:1412.6980* (2014).
30. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. Fsl. *Neuroimage* (2012) 62:782–90. doi: 10.1016/j.neuroimage.2011.09.015
31. Wheeler HM, Mlynash M, Inoue M, Tipirnini A, Liggins J, Bammer R, et al. The growth rate of early DWI lesions is highly variable and associated with

penumbral salvage and clinical outcomes following endovascular reperfusion. *Int J Stroke* (2015) 10:723–9. doi: 10.1111/ijis.12436

Conflict of Interest Statement: 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.

Copyright © 2018 Lucas, Kemmling, Bouteldja, Aulmann, Madany Mamlouk and Heinrich. 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.



Stroke Lesion Outcome Prediction Based on MRI Imaging Combined With Clinical Information

Adriano Pinto^{1,2*}, Richard Mckinley³, Victor Alves², Roland Wiest³, Carlos A. Silva^{1*} and Mauricio Reyes^{4*}

¹ CMEMS-UMinho Research Unit, University of Minho, Guimarães, Portugal, ² Centro Algoritmi, University of Minho, Braga, Portugal, ³ Support Center for Advanced Neuroimaging, University Institute for Diagnostic and Interventional Neuroradiology, Inselspital, Bern University Hospital, Bern, Switzerland, ⁴ Institute for Surgical Technology and Biomechanics, University of Bern, Bern, Switzerland

OPEN ACCESS

Edited by:

Fabien Scalzo,
University of California, Los Angeles,
United States

Reviewed by:

Ivana Galinovic,
Centrum für Schlaganfallforschung
Berlin, Germany
Henry Ma,
Monash University, Australia

*Correspondence:

Adriano Pinto
id6376@alunos.uminho.pt
Carlos A. Silva
csilva@dei.uminho.pt
Mauricio Reyes
mauricio.reyes@istb.unibe.ch

Specialty section:

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

Received: 12 April 2018

Accepted: 21 November 2018

Published: 05 December 2018

Citation:

Pinto A, Mckinley R, Alves V, Wiest R, Silva CA and Reyes M (2018) Stroke Lesion Outcome Prediction Based on MRI Imaging Combined With Clinical Information. *Front. Neurol.* 9:1060. doi: 10.3389/fneur.2018.01060

In developed countries, the second leading cause of death is stroke, which has the ischemic stroke as the most common type. The preferred diagnosis procedure involves the acquisition of multi-modal Magnetic Resonance Imaging. Besides detecting and locating the stroke lesion, Magnetic Resonance Imaging captures blood flow dynamics that guides the physician in evaluating the risks and benefits of the reperfusion procedure. However, the decision process is an intricate task due to the variability of lesion size, shape, and location, as well as the complexity of the underlying cerebral hemodynamic process. Therefore, an automatic method that predicts the stroke lesion outcome, at a 3-month follow-up, would provide an important support to the physicians' decision process. In this work, we propose an automatic deep learning-based method for stroke lesion outcome prediction. Our main contribution resides in the combination of multi-modal Magnetic Resonance Imaging maps with non-imaging clinical meta-data: the thrombolysis in cerebral infarction scale, which categorizes the success of recanalization, achieved through mechanical thrombectomy. In our proposal, this clinical information is considered at two levels. First, at a population level by embedding the clinical information in a custom loss function used during training of our deep learning architecture. Second, at a patient-level through an extra input channel of the neural network used at testing time for a given patient case. By merging imaging with non-imaging clinical information, we aim to obtain a model aware of the principal and collateral blood flow dynamics for cases where there is no perfusion beyond the point of occlusion and for cases where the perfusion is complete after the occlusion point.

Keywords: stroke, machine learning, deep learning, MRI, prediction

1. INTRODUCTION

Stroke ranks second as leading cause of death worldwide (1), with ischemic stroke being the most common type (2). Ischemic stroke arises from an artery occlusion caused by local thrombolysis, hemodynamic factors or embolic causes. Due to artery occlusion, the surrounding area suddenly suffers a blood flow reduction, leading the cells to a transient state slightly above cell death. The hypo-perfused area concerns the tissue at risk, also known as salvageable tissue, that can

eventually reach a non-viable point of failure even after flow restoration (3, 4). Therefore, stroke lesion can be characterized by a core tissue, encompassed by brain dead tissue, and a penumbra tissue corresponding to the salvageable tissue. The temporal evolution of a stroke lesion can be characterized into four main phases: hyper-acute (initial event), acute (6 h after event), sub-acute (from 24 h) and chronic phase (from 2 weeks) (5).

Neuroimaging plays an essential role in the diagnosis and treatment of stroke, where Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) are the preferred imaging modalities. However, MRI provides a better detection and assessment of potentially salvageable tissue, due to its multi-spectral property (6). After diagnosing and evaluating the stroke lesion through neuroimaging acquisitions, the clinicians need to plan the treatment phase. Such phase encompasses either mechanical thrombectomy or thrombolysis (7, 8) to revascularize the hypo-perfused tissue, which is only viable for the sub-acute phase. Therefore, in a short period of time, expert physicians must carefully evaluate the associated risks and benefits of the clinical intervention, namely the volume of hypo-perfused tissue potentially salvageable vs. the risk of causing haemorrhage or other complications (7, 9). If performed, the reperfusion success is assessed via the standardized Thrombolysis in Cerebral Infarction (TICI) scale (9).

Predicting stroke lesion outcome (i.e., at 3-month follow-up), and the potential efficacy of the treatment according to the nature of the lesion, has a great potential to guide the decision making of physicians. An automatic stroke tissue outcome prediction method would help the physician in such time-critical decision-making process (10). In this paper, we propose a novel end-to-end deep learning architecture that combines imaging information with clinical meta-data, the TICI scale. Our method incorporates clinical meta-data at two levels. First, at the population level, which implicitly encodes expected correlations between tissue loss and the TICI score into a custom loss function of the network. Second, at a patient level, which explicitly encodes the TICI score of each patient as an extra input channel of the network. To evaluate our proposal, we used the publicly available ISLES 2017 dataset, where we show the potential value of incorporating imaging and clinical meta-data for stroke tissue outcome prediction at a 3-month follow-up.

1.1. Previous Work

Several methods have been proposed for stroke lesion segmentation (11). However, only recently approaches based on machine learning have been proposed for ischemic stroke lesion outcome prediction. These proposals are based on multivariate linear regression models (12–14), decision trees (15), and CNN-based deep learning architectures (16, 17).

Scalzo et al. (12) proposed a framework to predict stroke tissue outcome, 4 days after clinical intervention (thrombectomy), based on Fluid Attenuation Inversion Recovery (FLAIR) MRI sequence, and Apparent Diffusion Coefficient (ADC) and Time-to-Maximum (Tmax) maps, if available. Tissue outcome prediction was achieved through a regression model that learns the behavior of neighbouring voxels within a cuboid. Kemmling et al. (14) used CT and MRI perfusion maps alongside clinical

information, encompassing the reperfusion success. The authors used a generalized linear model to consider the effect of multiple clinical variables when performing the stroke lesion outcome prediction, however, each voxel is considered independently, disregarding spatial context. Rose et al. (13) proposed a two-stage approach for stroke lesion outcome prediction based on perfusion maps, Cerebral Blood Flow (CBF), Cerebral Blood Volume (CBV), Mean Transit Time (MTT), and Diffusion-Weighted Imaging (DWI) maps. Initially, the method defines a region of interest (ROI) from the intensity signal of the perfusion and diffusion maps. Afterwards, a Gaussian mixture model, trained in different sets of MRI maps, performs stroke outcome prediction. McKinley et al. (15) also used a two-stage classification, where each stage comprehends two Random Forests (RFs). In the first stage, the method focusses on lesion delineation, through the definition of a ROI, where each classifier considers features extracted from different sets of MRI maps. After defining the hypo-perfused ROI, a second set of two RFs performs a precise prediction of the stroke lesion. These classifiers are trained on different sets of patients. One classifier is trained with patients with no reperfusion, to obtain worst case scenarios, whereas a second classifier is trained in patients with good reperfusion, therefore predicting scenarios where hypo-perfused tissue has higher chances of being salvaged. Afterwards, the final prediction is obtained by combining the results of both classifiers, using a logistic regression model.

Most recent methods are based on deep learning. Choi et al. (17) employed an ensemble of 12 deep learning methods, divided in two different groups. One group performs voxel-wise segmentation, based on the U-net architecture (18) adapted for 3D data, totalling four models. The other group encompasses Fully Connected Networks architectures with different patch sizes, to perform classification. The final prediction results from a weighted merging.

In previous approaches, the clinical information related to the success of reperfusion (TICI scale) has either been used within multivariate linear regression models (14), or to dichotomize the training data to train specific RFs models (15). Nonetheless, non-imaging clinical information has up to our knowledge not been integrated in deep learning architectures to predict stroke lesion outcome.

1.2. Contributions

In this paper, we propose an automatic method for stroke lesion outcome prediction, whose main contributions are:

1. The combination of imaging and non-imaging clinical data in an end-to-end deep learning architecture.
2. The development of a customized loss function to incorporate clinical information during the learning phase. Therefore, learning relationships between imaging and non-imaging information at a population level.
3. The inclusion of clinical information during the prediction phase at a patient-specific level, allowing us to perform predictions of different outcome scenarios in clinical environment.

The following sections are organized as follows: section 2 describes the proposed method. Section 3 details the database used and evaluation methods. Section 4 presents the results and its discussion. Finally, section 5 summarizes up the main aspects of the proposal.

2. METHODS

Stroke lesion outcome prediction consists of characterizing follow-up changes in location and extension of lesions over time from multi-sequence MRI and clinical information. In our proposal, to perform tissue outcome prediction, the method assigns to each voxel of the MRI volume one out of two classes, healthy tissue or stroke lesion. The following subsections describe the main steps of our proposal.

2.1. Pre-processing

Our proposal uses diffusion and perfusion maps, adding up to six MRI parametric maps: diffusion ADC map, and perfusion relative Cerebral Blood Flow (rCBF), relative Cerebral Blood Volume (rCBV), Mean Time to Transit (MTT), Time-to-Peak (TTP), and Tmax maps. **Figures 1, 2** show two cases of MRI maps with different TICI scores, alongside the manual segmentation (ground truth) obtained from a T2 sequence at a 90 day follow-up.

ISLES 2017 dataset provides MRI acquisitions from different centers (19). So, the perfusion and diffusion maps were acquired with different sets of configurations. Therefore, for each patient we first resized all maps to a common volume of dimension of $256 \times 256 \times 32$. Afterwards, the ADC maps were clipped between $[0, 2, 600] \times 10^{-6} \text{mm}^2/\text{s}$ and the Tmax maps were clipped to $[0, 20\text{s}]$, since values beyond these ranges are known to be biologically meaningless (15). As a final step of pre-processing, we applied a linear scaling across all maps transforming them to the range $[0, 255]$.

2.2. Deep Learning Architecture

Deep learning encompasses a variety of representation learning techniques capable of automatically learning hierarchical and complex features from the data. This property grants various levels of abstraction, translating to higher discriminative features, when comparing to hand-crafted features. In imaging processing, the most common techniques of deep learning are the Convolutional Neural Networks (CNNs) (20–22) and the Recurrent Neural Networks (RNNs) (22, 23).

CNNs have recently achieved remarkable success in well-known computer vision challenges (21). CNNs convolve a set of kernels over an input (image or image patches) obtaining a new feature space that characterizes local interactions in the input data.

Gated RNNs, which achieved success in the biomedical imaging field (23), provide a tighter notion of context. Initially proposed for the analysis of discrete sequences, their architecture contains gates that learn to store and read information from linear units. Due to this property, Gated RNNs, namely Long-Short Term Memory (24) and Gated Recurrent Unit (25), can process inputs and outputs of varying lengths and retain

information over long time-steps. When applied to computer vision, the memory capability of multi-dimensional gated RNNs allows us to model interactions among all the input data, which translates to a higher notion of context regardless of the receptive field.

Our proposal is inspired by the fully convolutional U-net architecture (18), which has proved to be competitive in many biomedical image segmentation applications. In addition, we combined the U-net with a 2D-dimensional Gated Recurrent Unit (GRU) layer (25) to obtain smoother and structured predictions. **Figure 3** shows the proposed architecture. The convolutional layers are responsible for the generation of discriminative feature vectors. Afterwards, the feature maps are fed into the GRU layer to enforce the spatial context of the network. Finally, a convolutional layer of 1×1 reduces the feature space to combine it with the clinical information.

2.3. Combining Imaging With Non-imaging Data

Besides MRI imaging data, non-imaging clinical information is also gathered during the acute phase of stroke, such as the Time Since Stroke (TSS), Time to Treatment (TTT), modified Ranking Scale (mRS) score, and TICI score. TSS and TTP are time measures that mark the time-points when the stroke incident was diagnosed and when clinical intervention was performed, respectively. The mRS score characterizes the degree of disability 90 days after a stroke incidence. However, the most relevant factor is the TICI score (9), which indicates the degree of success of the mechanical thrombectomy, based on cerebral angiography. Low scores ($\text{TICI} \in \{0, 1\}$) describe cases with minimal perfusion or no perfusion at all. Mid-range scores ($\text{TICI} \in \{2a, 2b\}$) characterize cases with progressively better partial perfusion. The highest score ($\text{TICI} = 3$) characterizes a complete flow-restoration (9). Consequently, it is expected that higher TICI scores naturally lead to increased levels of tissue being salvaged, and conversely, lower TICI scores might indicate increased levels of tissue loss. In our proposal, we aim to incorporate this information into a deep learning architecture, to relate imaging (e.g., stroke location, extension) with clinical information. In our proposal we aim to include such knowledge during the learning and testing phases of the system. To do so, our method considers the TICI scale at two levels: population-level and patient-level.

2.3.1. Population-Level

Incorporating clinical information at a population-level is achieved through a custom loss function, which drives the learning process to solutions conditioned to the clinical TICI score. Due to the presence or absence of perfusion beyond the location of the occlusion, stroke lesion extension can present changes between the TSS and the follow-up acquisitions. For cases with no perfusion, it is expected that the lesion grows between the two exams, while cases with existent perfusion should present a shrinkage of the lesion volume. In our proposal, we aim to model such lesion dynamics when predicting the lesion progression from the MRI parametric maps at the first exam to a future time. To do so, the training procedure is performed based on the MRI sequences from the first exam and the manual

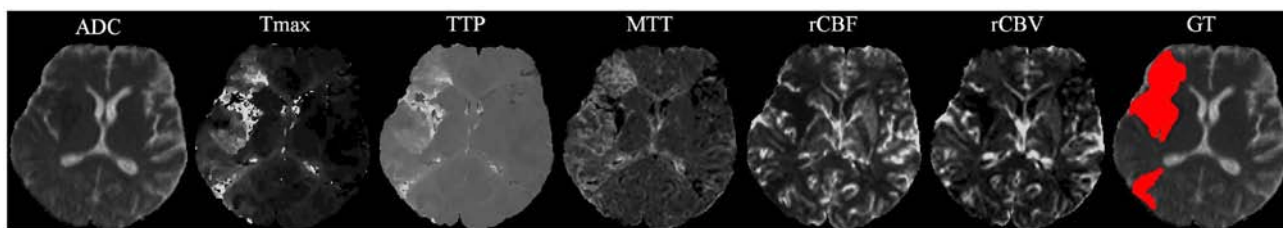


FIGURE 1 | MRI parametric maps of a stroke patient with TICl score 0, and the respective manual segmentation. Only one class is defined, describing simultaneously the infarct core and the penumbra regions.

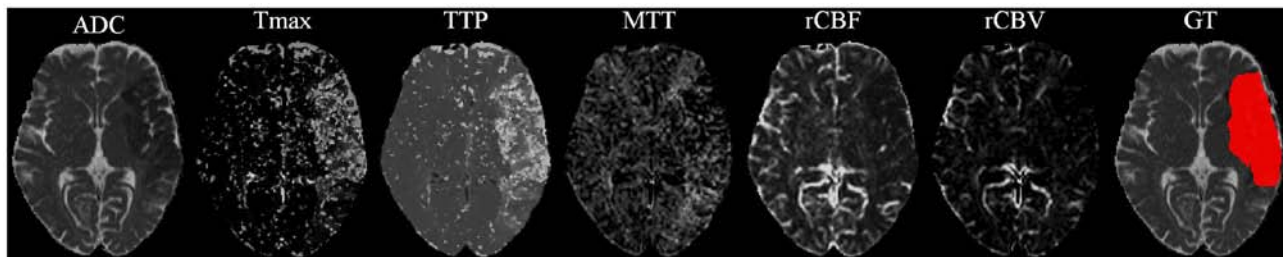


FIGURE 2 | MRI parametric maps of a stroke patient with TICl score 3, and the respective manual segmentation.

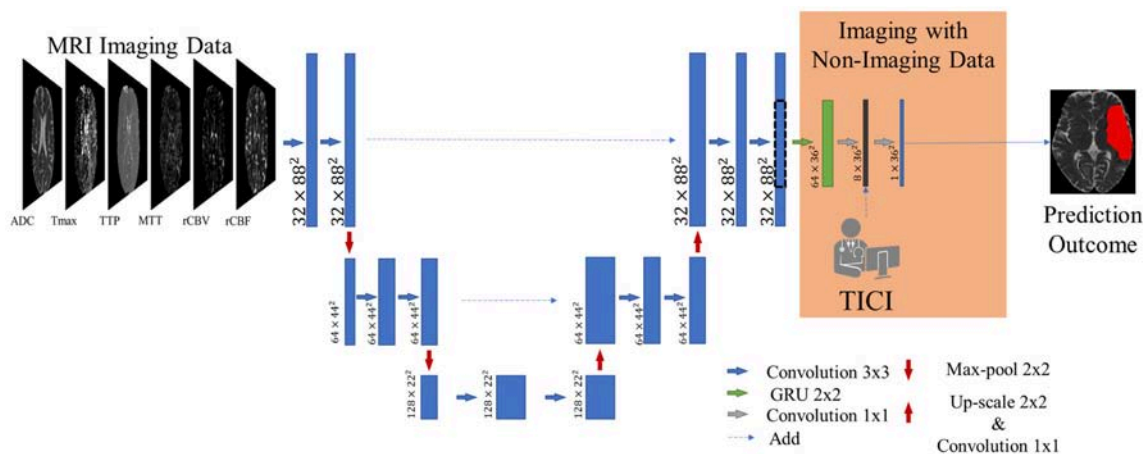


FIGURE 3 | Overview of the proposed architecture. Blue feature maps result from 2D-dimensional convolutions. The green feature maps represent the 2D-dimensional GRU layer. The first dimension corresponds to the number of feature maps. The dashed line consists of a cropping step to connect the U-Net with the GRU layer. The prediction is provided by the last layer, corresponding to a SoftMax activation.

segmentation of the lesion at the follow-up acquisition. When the lesion shrinks, our system must learn that although the lesion presents a larger extension in the MRI sequences, it should produce a smaller segmentation, and when the lesion grows, it should learn to predict a larger segmentation, although the information provided by the MRI sequences indicates it is smaller. We may model this dynamic by interpreting the growth as oversegmentation, and the shrinkage as undersegmentation in relation to the information supported by the MRI sequences in the present time. We may interpret the oversegmentation

as an increase in false positives (FP) and the shrinkage as an increase in false negatives (FN), since these are not supported by the information in the MRI sequences, acquired at the first medical exam. Such dynamic in our proposal is modeled by the F_β score that combines the Precision and Recall scores as follows:

$$F_\beta = (1 + \beta^2) \frac{\text{precision} \times \text{recall}}{(\beta^2 \times \text{precision}) + \text{recall}}. \quad (1)$$

The Precision score, defined as $Precision = \frac{TP}{TP+FP}$, measures the presence of false positives (FP), while the Recall, given by $Recall = \frac{TP}{TP+FN}$, considers the presence of false negatives (FN) (TP corresponds to the number of true positives). As shown in Equation (1), the relation between these two scores is controlled by β , which in our proposal encodes the TICI score. To be applicable to a supervised learning approach, F_β needs to relate the predictions with the ground truth, which is defined in the following way:

$$F_\beta = (1 + \beta^2) \frac{\sum_i^N p_i g_i}{\sum_i^N \beta^2 p_i^2 + \sum_i^N g_i^2}. \quad (2)$$

The sum is performed for the N voxels of the patch in the prediction, $p_i \in P$, and the ground truth, $g_i \in G$. The gradient of the F_β score for the j^{th} voxel prediction is computed as:

$$\frac{\delta F_\beta}{\delta p_j} = (1 + \beta^2) \left(\frac{g_j (\sum_i^N \beta^2 p_i^2 + \sum_i^N g_i^2) - (2\beta^2 p_j) \sum_i^N p_i g_i}{(\sum_i^N \beta^2 p_i^2 + \sum_i^N g_i^2)^2} \right). \quad (3)$$

2.3.2. Patient-Level

The inclusion of the TICI score at a patient level is achieved by an extra channel before the final layer of the architecture (see **Figure 3**). By combining the feature set extracted from imaging data and the respective TICI score, we aim to drive the learning process to search for correlations among them. With this approach we hypothesize that the model should be aware that different TICI scores should predict different lesion outcomes, during the estimation phase. Therefore, our proposal would be capable of predicting the amount of salvageable tissue loss in the presence and absence of recanalized perfusion.

2.4. Post-processing

As post-processing step, we performed simple morphological filtering. Stroke lesions vary significantly in size. The post-processing should take this variation into account to avoid the complete removal of stroke lesions; therefore, a threshold to remove only connected components with less than 25 voxels was defined using cross-validation.

3. EXPERIMENTAL SETUP

We evaluated our proposal on the ISLES 2017 training and testing datasets, where the online platform also includes an automated evaluation of prediction results submitted to the system. In this work, we compared the performance of our proposal with and without using clinical meta-data.

3.1. Dataset

ISLES 2017 dataset comprises a total of 75 ischemic stroke patients divided into two groups: training ($n = 43$) and testing ($n = 32$), who underwent mechanical thrombectomy. For each subject a total of six MRI acquisitions are provided: ADC, TTP, Tmax, rCBV, and rCBF. All image modalities are already co-registered and skull-stripped (16). Alongside the diffusion

TABLE 1 | TICI distribution for ISLES 2017 training and testing datasets.

	TICI 0	TICI 1	TICI 2a	TICI 2b	TICI 3
Training	6 (14%)	3 (7%)	3 (7%)	11 (26%)	20 (46%)
Testing	3 (9%)	2 (6%)	4 (13%)	6 (19%)	17 (53%)

and perfusion parametric MRI maps, each patient has a lesion outcome manually segmented by a clinician on a 90-day follow-up T2 MRI. The ground truth was provided only for the training dataset, since the test set is evaluated by the online platform. Alongside the imaging information, each patient is also characterized by the TICI score, TSS, TTT, and mRS Score. Although other clinical information is available, only the TICI scores were used in this study. **Table 1** describes the distribution of TICI score for each available dataset.

3.2. Evaluation

The performance of each method was evaluated using five metrics: Dice Similarity Score (DSC), Precision, Recall, Hausdorff Distance and Average Symmetric Surface Distance (ASSD). DSC measures the similarity between two volumes and is defined by $DSC = \frac{2TP}{FP+2TP+FN}$. As for the distance metrics, Hausdorff Distance denotes the maximum distance between two volumes surface points, capturing outliers. It is defined as: $HD(A, B) = \max\{\max_{a \in A} \min_{b \in B} d(a, b), \max_{b \in B} \min_{a \in A} d(b, a)\}$. Finally, ASSD describes the average distance between the volumes surface points defined as: $ASSD(A, B) = \frac{\sum_{a \in A} \min_{b \in B} d(a, b)}{|A|}$.

3.3. Setup

The validation set comprised seven cases, while the testing set of 36 cases from ISLES 2017 training set. To assess the added value of our contributions, we perform a 7-fold-cross-validation scheme within the training set. We compare our proposal with a baseline architecture, which does not encompass any clinical meta-data. In addition, we changed the loss function to the soft dice (26), which is a standard loss function for segmentation tasks.

3.4. Hyper-parameters

For each subject, around 500 patches of size 88×88 were extracted, using a uniform random sampling scheme. The network was trained with ADAM optimizer (27) (learning rate of 1×10^{-5}) using a mini-batch size of 4. The implementation was based on Keras (28) with Theano backend. All tests were conducted on a workstation equipped with a GeForce GTX 1070 with 8 GB. For each patient, prediction took around 15s.

3.4.1. Inclusion of Clinical Information

When considering cases with low TICI score, predicting the maximal extent of tissue loss eases the clinical decision-making process, therefore decreasing the chances of tissue death by hypoperfusion. In such circumstances, with the inclusion of the TICI score we aim to drive the model to predict the worst-case scenario of stroke lesion outcome. Conversely, in a case with a high TICI score we would prefer a prediction where the recovered

TABLE 2 | Results obtained through cross-validation in ISLES 2017 training dataset for the baseline method and our proposal. Each metric contains the average \pm standard deviation.

	Dice	Hausdorff distance	ASSD	Precision	Recall
Baseline	0.34 \pm 0.22	35.09 \pm 17.27	6.08 \pm 5.27	0.37 \pm 0.29	0.54 \pm 0.26
Proposal	0.35 \pm 0.22	31.38 \pm 15.81	5.55 \pm 5.00	0.41 \pm 0.30	0.47 \pm 0.24

hypo-perfused tissue due to reperfusion is achieved with success, holding on the same principles as before. It is worth mentioning that such relationship is further affected by several other clinical and patient-specific pathophysiological aspects, such as collateral flood, onset time of the stroke, etc.

Giving the available number of cases per TICI in ISLES 2017 dataset, we merged TICI scores, increasing the number of cases per score. Therefore, at a population level, β in Equation (4) encodes the TICI score as follows:

$$\beta = \begin{cases} 2, & \text{if } TICI \in \{0, 1\} \\ 1, & \text{if } TICI \in \{2, 2a, 2b\} \\ 0.5, & \text{if } TICI = 3 \end{cases} \quad (4)$$

In this way, for $TICI = 3$ (i.e., complete perfusion) we defined $\beta = 0.5$, so recall is weighted four times less than precision. Hence, we drive the model to give higher importance to the expression of false positives rather than false negatives, preferring scenarios with low tissue loss. Conversely, for $TICI \in \{0, 1\}$ (i.e., poor recanalization), we defined a $\beta = 2$, where recall is weighted four times higher than precision. For such cases, the motivation is to give preference to high tissue loss. Finally, for $TICI \in \{2a, 2b\}$ the value of $\beta = 1$, obtaining the Dice Score, where precision and recall are equally taken into consideration. Such scale of β was defined through cross-validation.

4. RESULTS AND DISCUSSION

In this section, we first evaluate the main contribution of our proposal in the training set. Using cross-validation we compare the performance of the baseline method without non-imaging clinical information against our proposal. Afterwards, we present the results obtained in ISLES 2017 testing dataset, performing a comparison against state-of-the-art methods.

4.1. Incorporation of Non-imaging Clinical Information

Due to the large diversity of appearance, size and shape, the tissue outcome prediction presents as a challenging task (10). In this study, we show the importance of having non-imaging clinical information in a neural network, to characterize principal and collateral blood flow hemodynamic and obtain better prediction outcomes. The results obtained for the training set are shown in Table 2.

When comparing with the baseline, our proposal is capable of achieving higher DSC and lower Hausdorff Distance, showing

the added value of incorporating the TICI score into the neural network. Considering the precision and recall metrics, our proposal achieved higher precision but lower recall. This suggests a higher capability to perform stroke lesion outcome prediction, by depicting gradual changes in the hypo-perfused tissue. We hypothesize that making the model aware to intrinsic biological phenomena of lesion growth or shrinkage (TICI dependent) lead to more precise predictions, which is sustained by the lower values of distance metrics and higher DSC score.

However, in clinical practice the TICI score is only obtained after recanalization. Being so, predicting the stroke lesion at a 90 day follow up, during the sub-acute phase, needs to consider different reperfusion scenarios. In our proposal, we grant such property at patient-level domain. By adding an extra input channel that contains the TICI score, we aim to obtain tissue outcome predictions with successful and unsuccessful reperfusions. When accessing both case scenarios, during the decision-making process, our method could provide to clinicians additional information on the salvaged tissue if mechanical thrombectomy was performed. In Figures 4, 5 we show the added value of incorporating clinical information on two patients with different TICI scores: one with an unsuccessful reperfusion ($TICI = 0$), and one with a successful reperfusion ($TICI = 3$).

For each case, we present the tissue outcome predictions with and without non-imaging clinical information. In the absence of the TICI score, the tissue outcome prediction performs worse than our proposal, for both cases. Our proposal is capable of employing the TICI score to yield better predictions, which are corroborated by higher Dice scores, but also provides a result that physiologically is more plausible. Observing the stroke lesion outcome predictions of our proposal against the baseline, it is noticeable the presence of physiologically infeasible isolated regions in the latter. Additionally, we also tested if our method was capable of predicting different lesion outcomes by changing the TICI score. When changing the TICI score, we obtained different lesion outcomes for each patient. Furthermore, such scenarios agreed with the expected outcome describe for each TICI score (e.g., by changing from a TICI score of 3 to 0 it was observed a larger lesion outcome volume). From the latter study, we show that our proposal gained awareness to scenarios of no-perfusion and complete perfusion. Such capability could provide the clinicians useful insight on the benefits and risks associated to the mechanical thrombectomy. Moreover, it can also be used to forecast recovery, which is important for patient treatment and the complete standard care associated to patient recovery. To corroborate our qualitative analysis, Table 3 contains the ground-truth lesion volume for each case, alongside the predicted volume outcome for the original TICI score and for the opposite case scenario, respectively.

On Table 3, we show the effect of the TICI score in our proposal. When changing the TICI score we observe different stroke lesion outcome predictions, in agreement to the reperfusion success. When increasing the TICI score the volume of salvaged hypo-perfused tissue becomes higher, which corresponds to a stroke lesion shrinkage. Case 24, with $TICI = 0$, shows such behavior. After increasing the TICI score to $TICI = 3$, we obtain a smaller stroke lesion volume. As

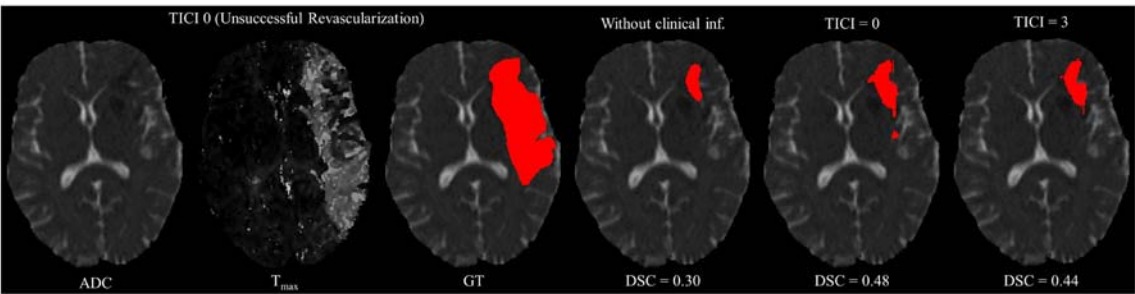


FIGURE 4 | Example case of stroke lesion outcome prediction, with and without non-imaging clinical information in a patient with unsuccessful reperfusion. For sake of description we present the ADC and T_{max} maps and the GT. In the presence of clinical information, we show the two possible outcomes: unsuccessful (TICI = 0) and successful reperfusion (TICI = 3), respectively.

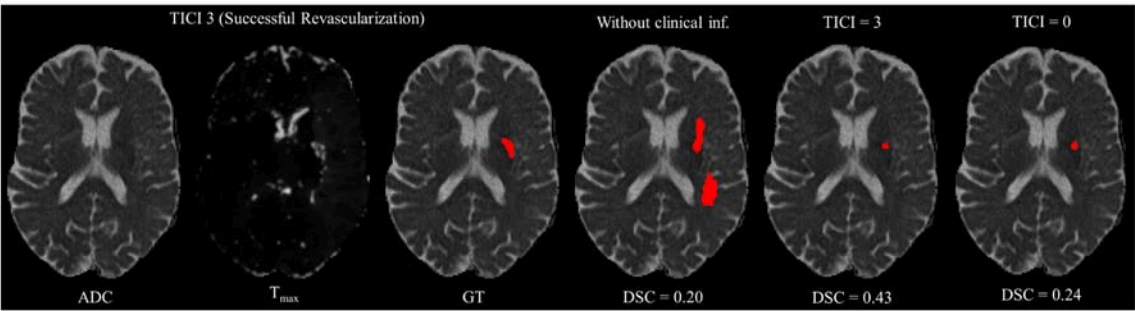


FIGURE 5 | Example case of stroke lesion outcome prediction, with and without non-imaging clinical information in a patient with successful reperfusion. We also present the ADC and T_{max} maps and the GT. In the presence of clinical information, we show the two possible outcomes: successful (TICI = 3) and unsuccessful reperfusion (TICI = 0), respectively.

TABLE 3 | Results obtained by our proposal on two patient cases with different TICI scores, alongside the obtained result after changing the original TICI score to its opposite (marked with a *).

Case	GT volume (voxels)	TICI	Dice	Precision	Recall	Predicted volume (voxels)
24	21,310	0	0.48	0.87	0.33	8170
		3*	0.44	0.90	0.29	6840
42	288	3	0.43	0.59	0.33	163
		0*	0.24	0.17	0.39	651

for case 42 with $TICI = 3$, when we decrease the TICI score from $TICI = 3$ to $TICI = 0$ the prediction volume characterized the opposite phenomena. With $TICI = 0$ there is higher hypo-perfused tissue loss, and the tissue outcome prediction volume is larger. From both case scenarios, the observed changes in the tissue outcome prediction volume shows that the TICI score was capable of driving the tissue outcome prediction scenario, and simultaneously grant a lesion growth or shrinkage in accordance with the physiological dynamics of each TICI score and without infeasible isolated regions.

4.2. ISLES 2017 Testing Set

In **Table 4**, we compare our proposal with methods from ISLES 2017 testing dataset, evaluated by the online platform (29) and ordered decreasingly by the DSC score. To reinforce our analysis, we also included the baseline method.

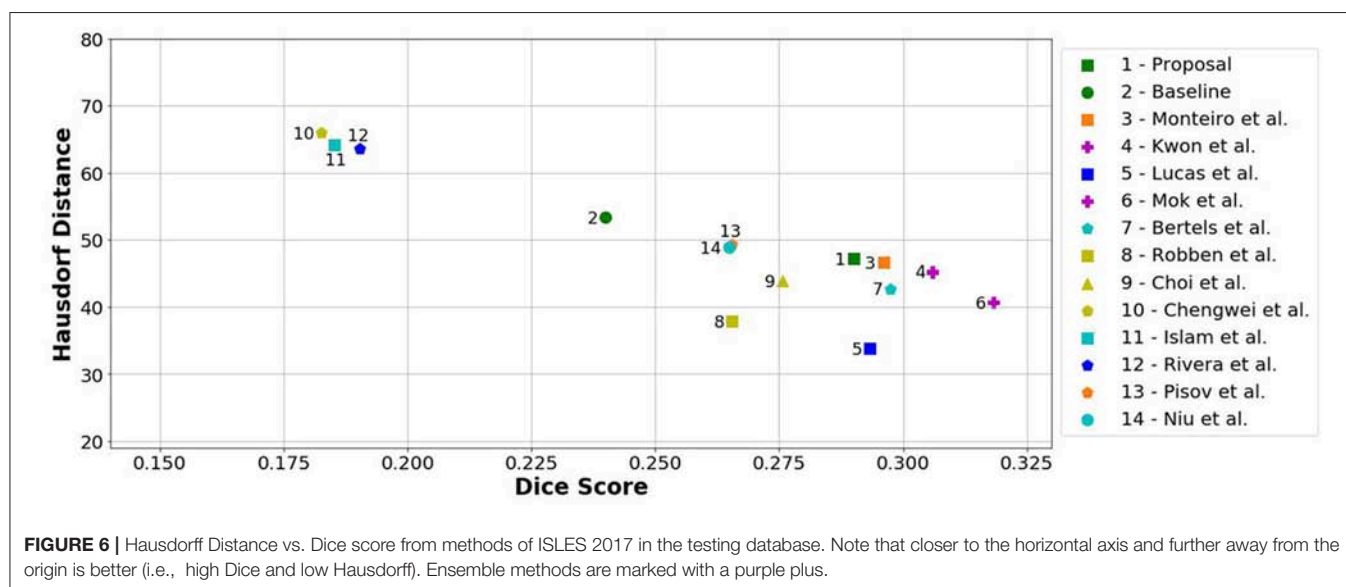
Incorporating clinical information through the proposed custom loss function and the extra TICI channel resulted in a higher performance, in comparison to the baseline. Our proposal was able extract information from non-imaging data and to drive its training and testing phases toward better predictions. Therefore, the simultaneous incorporation of the reperfusion status, as an additional feature and in the loss function, improved performance of the classifier. In addition, we show the higher generalization capability of our proposal, since the performance metrics of our proposal for both datasets present less variation.

Although a previous work (15) had investigated the use of non-imaging clinical information to conduct the training of machine learning methods, such information has not been evaluated directly in the context of deep learning methods. The results on the ISLES2017 indicate the benefits of incorporating non-imaging clinical information in deep learning architectures implicitly during the training phase, and explicitly by extra channels, incorporating patient-specific information.

TABLE 4 | Results of ISLES 2017 testing dataset, alongside our baseline method and proposal. Each metric contains the average \pm standard deviation.

		Dice	Hausdorff distance	ASSD	Precision	Recall
Challenge	Mok et al. *	0.32 \pm 0.23	40.74 \pm 27.23	8.97 \pm 9.52	0.34 \pm 0.27	0.39 \pm 0.27
	Kwon et al. *	0.31 \pm 0.23	45.26 \pm 21.04	7.91 \pm 7.31	0.36 \pm 0.27	0.45 \pm 0.30
	Bertels et al. *	0.30 \pm 0.21	33.85 \pm 16.82	6.81 \pm 7.18	0.34 \pm 0.26	0.51 \pm 0.32
	Monteiro et al. *	0.30 \pm 0.22	46.60 \pm 17.50	6.31 \pm 4.05	0.34 \pm 0.27	0.51 \pm 0.30
	Lucas et al. *	0.29 \pm 0.21	33.85 \pm 16.82	6.81 \pm 7.18	0.34 \pm 0.26	0.51 \pm 0.32
	Choi et al. *	0.28 \pm 0.22	43.89 \pm 20.70	8.88 \pm 8.19	0.36 \pm 0.31	0.41 \pm 0.31
	Robben et al. *	0.27 \pm 0.22	37.84 \pm 17.75	6.72 \pm 4.10	0.44 \pm 0.32	0.39 \pm 0.31
	Pisov et al. *	0.27 \pm 0.20	49.24 \pm 32.15	9.49 \pm 10.56	0.31 \pm 0.27	0.39 \pm 0.29
	Niu et al. *	0.26 \pm 0.20	48.88 \pm 11.20	6.26 \pm 3.02	0.28 \pm 0.25	0.56 \pm 0.26
	Sedlar et al. *	0.20 \pm 0.19	58.30 \pm 20.02	11.19 \pm 9.10	0.23 \pm 0.24	0.40 \pm 0.29
	Rivera et al. *	0.19 \pm 0.16	63.58 \pm 18.58	11.13 \pm 7.89	0.27 \pm 0.25	0.21 \pm 0.17
	Islam et al. *	0.19 \pm 0.18	64.15 \pm 28.51	14.17 \pm 15.80	0.29 \pm 0.28	0.25 \pm 0.25
	Chengwei et al. *	0.18 \pm 0.17	65.95 \pm 25.94	9.22 \pm 6.99	0.37 \pm 0.30	0.21 \pm 0.23
	Yoon et al. *	0.17 \pm 0.16	45.23 \pm 19.14	12.43 \pm 11.01	0.23 \pm 0.27	0.36 \pm 0.32
	Baseline	0.24 \pm 0.20	53.29 \pm 26.95	10.59 \pm 4.98	0.27 \pm 0.27	0.50 \pm 0.35
	Proposal	0.29 \pm 0.22	47.17 \pm 22.13	7.20 \pm 4.14	0.26 \pm 0.23	0.61 \pm 0.28

* Static results in Ischemic Stroke Lesion Segmentation Challenge (29).



When comparing to the state-of-the-art methods, our proposal can reach competitive results, being placed among top scoring methods. With single model method, our proposal yields results within the top five methods, alongside ensemble approaches [e.g., Choi et al. (17)]. In the same group, our method achieved the highest recall metric, with lower precision score. As for the distance metrics, our proposal can provide competitive ASSD score, with low standard deviation, and a Hausdorff Distance among of top methods. We emphasize that, as post-processing step, our method only applies a simple morphological removal of small connected components. Therefore, elaborate schemes of post-processing such as Conditional Random Fields or even weighted schemes of ensemble can boost the performance of such approaches. Even in such cases, our approach provides a good robustness and precision in stroke lesion outcome

delineation. To enforce such analysis in **Figure 6**, we show the average DSC score and the Hausdorff Distance obtained by each state-of-the-art method in ISLES 2017 testing dataset. Besides our proposal, we included the baseline method.

From **Figure 6**, we can observe the performance boost of our proposal over the baseline method, placing it within the group of top scoring methods.

5. CONCLUSIONS

Prediction of stroke lesion outcome has the potential to assist interventionists when assessing the risks and benefits associated to mechanical thrombectomy. Therefore, having such tool can provide useful information during the clinical decision process.

In this work, we propose a novel deep learning architecture that beyond previously proposed architectures incorporates clinical information in a principled way. To do so, our proposal integrates clinical information at two different levels of the architecture. The first level considers the population domain-knowledge, achieved through the development of a custom loss function, to depict relationships between the TICI score and the tissue outcome prediction. The second level considers the patient-specific domain, where the TICI is encoded into an input channel of the architecture. From the latter level, we showed that our proposal was able to characterize different outcome scenarios of successful and unsuccessful reperfusion. Such methodology presents itself as a ground-breaking tool with potential to access the risks and benefits associated to the mechanical thrombectomy. The evaluation of our proposal was conducted on the publicly available ISLES 2017 dataset. We observe that the proposed method has benefited from the combination of imaging and non-imaging information. In addition, when comparing to the state-of-the-art methods, we observed that a single architecture with fewer parameters, such as ours, yields competitive performance metrics similar to more elaborate and/or ensemble methods.

However, there is still room from improvement since none of the current state-of-the-art methods, provides the robustness and accuracy needed for clinical practice, and are currently bellow the inter-rater performance of expert radiologists (DSC=0.58) (19). In the future, we would like to investigate on adding other clinical information, such as TTT and TSS. We esteem that the proposed approach can be further applied to other diseases where clinical information complements imaging information.

ETHICS STATEMENT

The study utilizes anonymized data from the Bernese stroke registry, a prospectively collected database approved by

the Kantonale Ethikkommission Bern. All patients were treated for an acute ischemic stroke at the University Hospital of Berne between 2005 and 2013. The study was performed according to the ethical guidelines of the Canton of Bern (Swiss Humanforschungsgesetz) with approval of our institutional review board (Kantonale Ethikkommission Bern). Some cases were supplied by the University Medical Center Schleswig-Holstein in Lübeck, Germany. They were acquired in diagnostic routine with varying resolutions, views, and imaging artifact load. A smaller group of cases were scanned at the Department of Neuroradiology at the Klinikum rechts der Isar in Munich, Germany. Both centers are equipped with 3T Phillips systems. The local ethics committee approved their release under Az.14-256A. Full data anonymization was ensured by removing all patient information from the files and the facial bone structure from the images.

AUTHOR CONTRIBUTIONS

AP is the main author of the research presented in the manuscript, being supervised by RM during an internship at Bern. CS, VA, RW and RM gave thoughtful insights during this research.

FUNDING

AP was supported by a scholarship from the Fundação para a Ciência e Tecnologia (FCT), Portugal (scholarship number PD/BD/113968/2015). This work is supported by FCT with the reference project UID/EEA/04436/2013, by FEDER funds through the COMPETE 2020 Programa Operacional Competitividade e Internacionalização (POCI) with the reference project POCI-01-0145-FEDER-006941. We acknowledge support from the Swiss National Science Foundation –DACH320030L_163363.

REFERENCES

- World Health Organization. *Global Status Report on Noncommunicable Diseases 2014*. World Health Organization (2014).
- Mathers CD, Boerma T, Ma Fat D. Global and regional causes of death. *Brit Med. Bull.* (2009) 92:7–32. doi: 10.1093/bmb/ldp028
- Wardlaw J. Neuroimaging in acute ischaemic stroke: insights into unanswered questions of pathophysiology. *J Intern Med.* (2010) 267:172–90. doi: 10.1111/j.1365-2796.2009.02200.x
- Memezawa H, Smith ML, Siesjö BK. Penumbra tissues salvaged by reperfusion following middle cerebral artery occlusion in rats. *Stroke* (1992) 23:552–9.
- Gonzalez R, Hirsch J, Koroshetz W, Lev M, Schaefer P. Acute ischemic stroke: imaging and intervention. *Am J Neuroradiol.* (2007) 28:1622–3. doi: 10.3174/ajnr.A0628
- Butcher K, Emery D. Acute stroke imaging part II: the ischemic penumbra. *Can J Neurol Sci.* (2010) 37:17–27. doi: 10.1017/S0317167100009604
- El Tawil S, Muir KW. Thrombolysis and thrombectomy for acute ischaemic stroke. *Clin Med.* (2017) 17:161–65.
- Sandercock P, Willems H. Medical treatment of acute ischaemic stroke. *Lancet* (1992) 339:537–9.
- Higashida RT, Furlan AJ, Roberts H, Tomsick T, Connors B, Barr J, et al. Trial design and reporting standards for intraarterial cerebral thrombolysis for acute ischemic stroke. *J Vasc Interv Radiol.* (2003) 14:E1–31. doi: 10.1161/01.STR.0000082721.62796.09
- Maier O, Wilms M, von der Gablentz J, Krämer UM, Münte TF, Handels H. Extra tree forests for sub-acute ischemic stroke lesion segmentation in MR sequences. *J Neurosci Methods* (2015) 240:89–100. doi: 10.1016/j.jneumeth.2014.11.011
- Rekik I, Allassonnière S, Carpenter TK, Wardlaw JM. Medical image analysis methods in MR/CT-imaged acute-subacute ischemic stroke lesion: segmentation, prediction and insights into dynamic evolution simulation models. A critical appraisal. *Neuroimage Clin.* (2012) 1:164–78. doi: 10.1016/j.nicl.2012.10.003
- Scalzo F, Hao Q, Alger JR, Hu X, Liebeskind DS. Regional prediction of tissue fate in acute ischemic stroke. *Ann Biomed Eng.* (2012) 40:2177–87. doi: 10.1007/s10439-012-0591-7
- Rose SE, Chalk JB, Griffin MP, Janke AL, Chen F, McLachan GJ, et al. MRI based diffusion and perfusion predictive model to estimate stroke evolution. *Magn Reson Imaging* (2001) 19:1043–53. doi: 10.1016/S0730-725X(01)00435-0

14. Kemmling A, Flottmann F, Forkert ND, Minnerup J, Heindel W, Thomalla G, et al. Multivariate dynamic prediction of ischemic infarction and tissue salvage as a function of time and degree of recanalization. *J Cereb Blood Flow Metab.* (2015) 35:1397–405. doi: 10.1038/jcbfm.2015.144
15. McKinley R, Häni L, Gralla J, El-Koussy M, Bauer S, Arnold M, et al. Fully automated stroke tissue estimation using random forest classifiers (FASTER). *J Cereb Blood Flow Metab.* (2016) 37:2728–41. doi: 10.1177/0271678X16674221
16. Maier O, Menze BH, von der Gablentz J, Häni L, Heinrich MP, Liebrand M, et al. ISLES 2015-A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. *Med Image Anal.* (2017) 35:250–69. doi: 10.1016/j.media.2016.07.009
17. Choi Y, Kwon Y, Lee H, Kim BJ, Paik MC, Won JH. Ensemble of deep convolutional neural networks for prognosis of ischemic stroke. In: *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Athens: Springer (2016). p. 231–43.
18. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Munich: Springer (2015). p. 234–41.
19. Winzeck S, Hakim A, McKinley R, Pinto JA, Alves V, Silva C, et al. ISLES 2016 and 2017-benchmarking ischemic stroke lesion outcome prediction based on multispectral MRI. *Front Neurol.* (2018) 9:679. doi: 10.3389/fneur.2018.00679
20. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. Imagenet large scale visual recognition challenge. *Int J Comput Vis.* (2015) 115:211–52. doi: 10.1007/s11263-015-0816-y
21. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Bartlett PL, Pereira FCN, Burges CJC, Bottou L, Weinberger KQ, editors. *Advances in Neural Information Processing Systems*. Lake Tahoe, NV: Curran Associates Inc. (2012). p.1097–105.
22. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* (2017) 42:60–88. doi: 10.1016/j.media.2017.07.005
23. Stollenga MF, Byeon W, Liwicki M, Schmidhuber J. Parallel multi-dimensional LSTM, with application to fast biomedical volumetric image segmentation. In: Cortes C, Sugiyama M, Lee DD, and Garnett R, editors *Advances in Neural Information Processing Systems* Montreal, QC (2015). p. 2998–3006.
24. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput.* (1997) 9:1735–80.
25. Cho K, Van Merriënboer B, Bahdanau D, Bengio Y. On the properties of neural machine translation: encoder-decoder approaches. arXiv preprint *arXiv:1409.1259* (2014).doi: 10.3115/v1/W14-4012
26. 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)*. California, CA: IEEE (2016). p. 565–71.
27. Kingma D, Ba J. Adam: a method for stochastic optimization. arXiv preprint *arXiv:1412.6980* (2014).
28. Chollet F. Keras (2015). Available online at: <https://keras.io>
29. Ischemic Stroke Lesion Segmentation Challenge. Available online at: <http://www.isles-challenge.org> (Accessed March 29, 2018) (2017).

Conflict of Interest Statement: 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.

Copyright © 2018 Pinto, McKinley, Alves, Wiest, Silva and Reyes. 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.



Convolutional Neural Networks for Direct Inference of Pharmacokinetic Parameters: Application to Stroke Dynamic Contrast-Enhanced MRI

Cagdas Ulas^{1*}, Dhritiman Das^{1,2}, Michael J. Thrippleton³, Maria del C. Valdés Hernández³, Paul A. Armitage⁴, Stephen D. Makin³, Joanna M. Wardlaw³ and Bjoern H. Menze^{1,5}

¹ Department of Computer Science, Technische Universität München, Munich, Germany, ² GE Global Research, Munich, Germany, ³ Department of Neuroimaging Sciences, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, United Kingdom, ⁴ Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, Sheffield, United Kingdom, ⁵ Institute of Advanced Study, Technische Universität München, Munich, Germany

OPEN ACCESS

Edited by:

Fabien Scalzo,
University of California, Los Angeles,
United States

Reviewed by:

Bin Jiang,
Beijing Neurosurgical Institute, Beijing
Tiantan Hospital, Capital Medical
University, China
Ruogu Fang,
University of Florida, United States
Matthias Günther,
University of Bremen, Germany

*Correspondence:

Cagdas Ulas
cagdas.ulas@tum.edu

Specialty section:

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

Received: 05 May 2018

Accepted: 11 December 2018

Published: 08 January 2019

Citation:

Ulas C, Das D, Thrippleton MJ, Valdés Hernández MdC, Armitage PA, Makin SD, Wardlaw JM and Menze BH (2019) Convolutional Neural Networks for Direct Inference of Pharmacokinetic Parameters: Application to Stroke Dynamic Contrast-Enhanced MRI. *Front. Neurol.* 9:1147. doi: 10.3389/fneur.2018.01147

Background and Purpose: The T1-weighted dynamic contrast enhanced (DCE)-MRI is an imaging technique that provides a quantitative measure of pharmacokinetic (PK) parameters characterizing microvasculature of tissues. For the present study, we propose a new machine learning (ML) based approach to directly estimate the PK parameters from the acquired DCE-MRI image-time series that is both more robust and faster than conventional model fitting.

Materials and Methods: We specifically utilize deep convolutional neural networks (CNNs) to learn the mapping between the image-time series and corresponding PK parameters. DCE-MRI datasets acquired from 15 patients with clinically evident mild ischaemic stroke were used in the experiments. Training and testing were carried out based on leave-one-patient-out cross-validation. The parameter estimates obtained by the proposed CNN model were compared against the two tracer kinetic models: (1) Patlak model, (2) Extended Tofts model, where the estimation of model parameters is done via voxelwise linear and nonlinear least squares fitting respectively.

Results: The trained CNN model is able to yield PK parameters which can better discriminate different brain tissues, including stroke regions. The results also demonstrate that the model generalizes well to new cases even if a subject specific arterial input function (AIF) is not available for the new data.

Conclusion: A ML-based model can be used for direct inference of the PK parameters from DCE image series. This method may allow fast and robust parameter inference in population DCE studies. Parameter inference on a 3D volume-time series takes only a few seconds on a GPU machine, which is significantly faster compared to conventional non-linear least squares fitting.

Keywords: dynamic contrast enhanced MRI, pharmacokinetic parameter inference, convolutional neural networks, ischaemic stroke, tracer kinetic modeling, contrast agent concentration, loss function

1. INTRODUCTION

Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is an effective dynamic imaging technique that can be used to study microvascular structure *in vivo* by tracking the diffusion of a paramagnetic contrast agent such as gadopentate dimeglumine (Gd-DTPA) over time (1). By collecting a series of T₁-weighted MR images at intervals of a few seconds, the uptake and washout of the administered contrast agent can be observed in the imaged tissue, resulting in characteristic intensity-time curves across different tissues (2). Vascular and cellular regularities in human body usually have a strong impact on the local vascular perfusion and permeability. To this end, DCE imaging has been used as a promising tool for clinical diagnostics of brain tumors, multiple sclerosis lesions, and several neurological disorders that lead to disruption and breakdown of blood-brain barrier (BBB) (3–6). In DCE-MRI, changes in contrast agent concentration are determined from changes in signal intensity over time, and then regressed through the use of tracer kinetic (TK) models to estimate pharmacokinetic (PK) parameters which characterizes the vascular permeability and tissue perfusion (7, 8).

One of the key limitations of TK modeling methods is that they are simply based on the fitting of voxelwise PK parameters to contrast agent concentration-time curves (9). The fitting is usually performed using a nonlinear least squares (NLS) approach. However, the acquired voxelwise concentration-time curves are generally very noisy and involve only a small number of sampling points, hence the model fitting may yield parameter estimates with large variance as well as considerable bias (see **Figure 1** for an exemplary representation of this limitation). Moreover, an iterative NLS solver may converge to erroneous solutions since the NLS objective is not convex and can have multiple local minima (10). Another major drawback is that the voxelwise model fitting is computationally demanding considering the thousands of voxels in a single MR slice (11). More sophisticated approaches (10, 12) were also proposed based on Bayesian theory of statistical inference of the DCE parameters for the fitting of nonlinear models. Unlike the standard NLS regression, these approaches exploit the spatial information of the neighboring voxels and provide reduce variability of parameters in local homogeneous regions. However, the bottleneck is their drastically increased computation time, usually taking hours for the estimation of parameters on a single DCE scan.

Machine learning (ML) methods have been extensively used in the medical imaging community for several tasks (13) such as parameter estimation, disease classification, segmentation, so on. Recently, a random forest regression based method (14) was proposed to estimate accurate spectral parameters in MR spectroscopy. Deep learning methods (15–17) have recently gained large popularity and achieved predominantly state-of-the-art results in the medical imaging field including various image-to-image translation tasks (18–20). A deep neural network based approach for perfusion parameter estimation (21) was first proposed for dynamic susceptibility contrast (DSC) MRI without requiring a standard deconvolution process.

To alleviate the aforementioned limitations in DCE-MRI, we present a direct and fast PK parameter estimation method which introduces several concepts from machine learning. Our proposed approach can directly infer the PK parameters from the observed signal intensity over time. In order to achieve this, we first train a deep convolutional neural network (CNN) to learn the underlying mapping – or relation – between intensity image-time series and PK parameters using a large training data consisting of millions of voxels taken from the brain DCE dataset. In our method, the target PK parameters used in training step can be either estimated by any existing tracer kinetic models, or can be defined with reference values depending on a specific biomarker or disease that has been built on one specific type of model. Our method can intrinsically provide the following advantages over the conventional model fitting based parameter estimation approaches:

- The proposed method can directly estimate the corresponding physiological perfusion parameters when only observed signal intensities over time given, which eliminates several intermediate computation steps of the conventional pipeline as illustrated in **Figure 2**.
- Our method serves as a high-level parameter estimation model such that we can train a network from which we expect to yield parameter estimates as close as the target values that are obtained using any optimization approach, e.g., standard NLS fitting, regularized Bayesian estimation methods, etc.
- Due to its strong generalization ability, this method shows increased robustness to signal noise and outliers, and it can significantly mitigate the effect of irregularity and discontinuity problem which is quite apparent in the parameter maps estimated by conventional NLS fitting.
- The parameter estimates obtained by the proposed approach yields improved statistically significant differences between different tissue types, which can ultimately allow better discrimination of normal and pathological regions in stroke analysis.
- Compared to conventional fitting methods, the PK parameter inference with our ML based approach is computationally faster, taking only seconds on an entire 3D DCE-MRI volume.

2. MATERIALS AND METHODS

2.1. Dataset and Preprocessing

2.1.1. Patients

Fifteen patients were recruited for this study. The patient cohort presents first clinically evident mild (i.e., expected to be non-disabling) ischaemic stroke from the local stroke service. The patients were over 18 years old and had a definite diagnosis of ischaemic stroke. They were able to consent themselves, had an MRI scan at diagnosis and were medically stable enough to return for a DCE-MRI scan at between 1 and 3 months post-stroke and a follow-up after 1 year. All patients underwent clinical assessment by a stroke physician, diagnostic MR imaging and cognitive testing at presentation. An expert panel of stroke physicians and neuro-radiologists assessed each case in order to confirm the diagnosis of ischaemic stroke and classify the ischaemic stroke

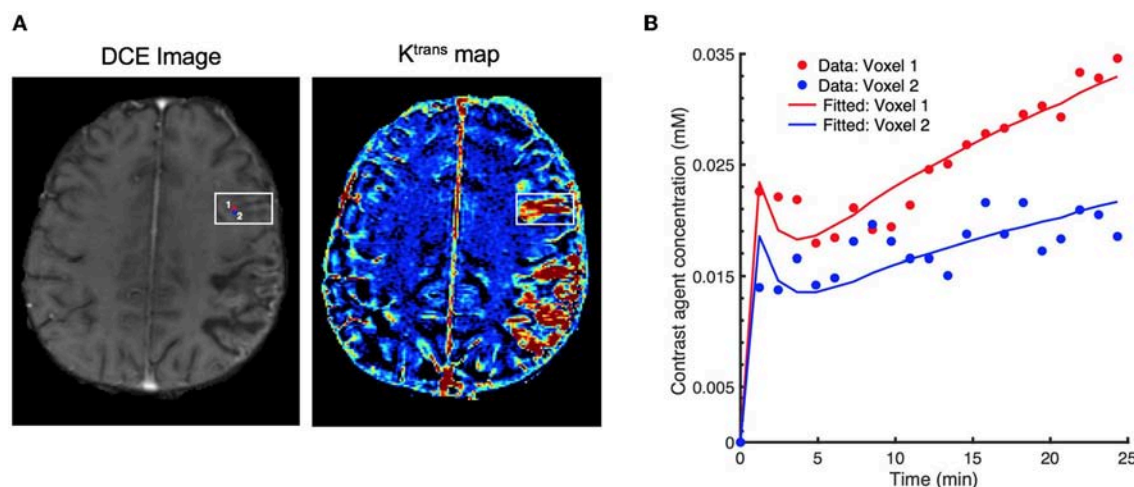


FIGURE 1 | Effect of signal noise on the resulting parameters with conventional fitting models. **(A)** an exemplary DCE image (left) displaying two neighboring voxels (marked by red and blue circles) in the stroke region, and the corresponding K^{trans} maps (right), **(B)** resulting fitted contrast agent concentration curves for these two voxels using Extended Tofts model. Although the neighboring voxels are spatially very close to each other (only 1-pixel away), the observed concentration data are different due to the excessive signal noise. Eventually, there is a substantial difference in the fitted concentration curves and parameter values ($K^{\text{trans}} = 6.18 \times 10^{-3} \text{ min}^{-1}$ for voxel 1, and $K^{\text{trans}} = 2.48 \times 10^{-3} \text{ min}^{-1}$ for voxel 2).

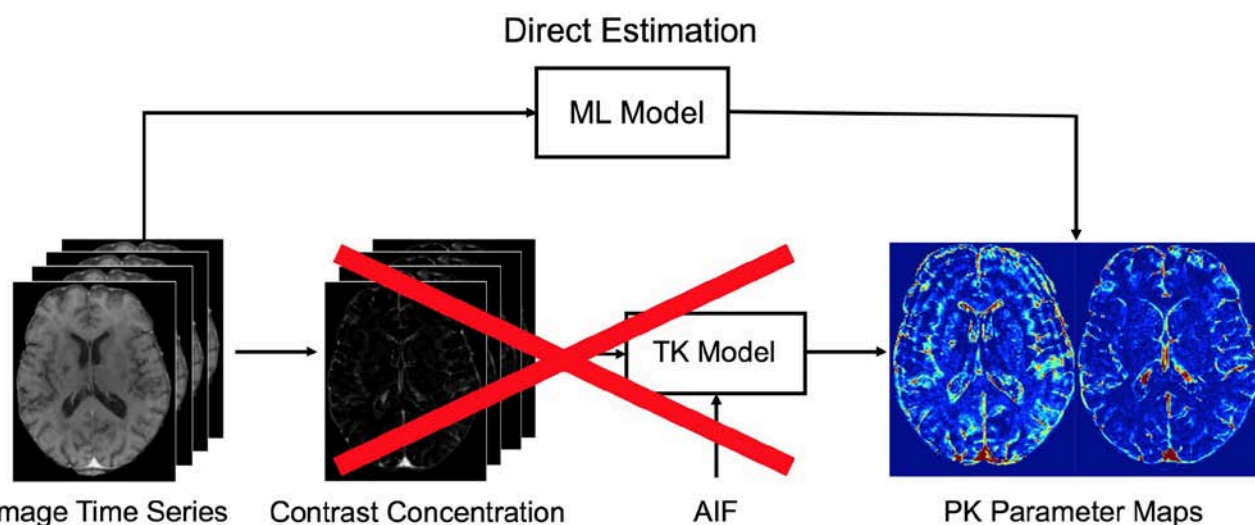


FIGURE 2 | The conventional pipeline of pharmacokinetic parameters for DCE-MRI. Our proposed machine learning (ML) model allows to directly infer the parameters from the acquired DCE image time series. To this end, the intermediate computational steps—i.e., conversion to contrast concentration, extraction of AIF, and fitting to a tracer kinetic (TK) model—can be eliminated when applied on a test data. We note that in our approach a specific TK model can be still used to estimate target parameter values during training.

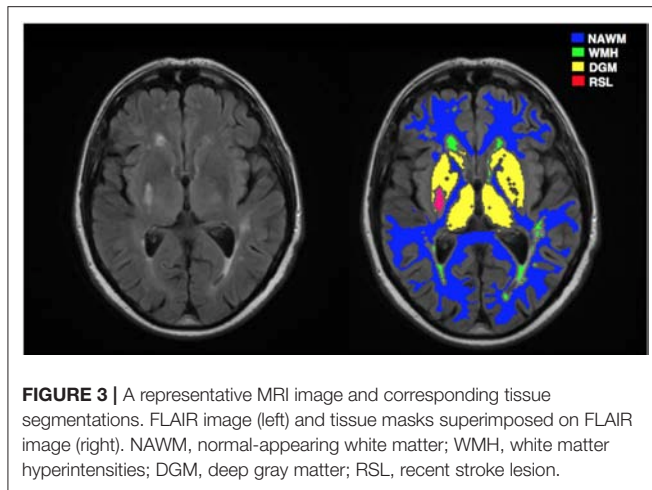
subtype. DCE-MRI was performed a minimum of 1 month after the stroke in order to avoid acute effects of the stroke on the local BBB (22). This study was approved by the Lothian Ethics of Medical Research Committee (REC 09/81101/54) and the NHS Lothian R + D Office (2009/W/NEU/14), and all patients gave written informed consent.

2.1.2. MRI Acquisition

MR imaging was performed on a 1.5 T MRI scanner (Signa HDxt, General Electric (GE), Milwaukee, WI) using an 8-channel

phased-array coil. Structural MR images for diagnostic purpose were acquired at first including axial T2-weighted (T2W; TR/TE = 6000/90 ms, FoV = 240×240 mm, acquisition matrix = 384×384 , 1.5 averages, 28×5 mm slices, 1 mm slice gap), and axial fluid-attenuated inversion recovery (FLAIR; TR/TE/TI = 9000/153/2200 ms, FoV = 240×240 mm, acquisition matrix = 384×224 , 28×5 mm slices, 1 mm slice gap).

DCE image series were acquired using a 3D T1W spoiled gradient echo sequence (TR/TE = 8.24/3.1 ms, flip angle = 12° , FoV = 240×240 mm, acquisition matrix = 256×192 , slice



thickness = 4 mm, 42 slices). Two pre-contrast acquisitions were carried out at flip angles of 2° and 12° to calculate pre-contrast longitudinal relaxation times (T_{10}). An intravenous bolus injection of 0.1 mmol/kg of gadoterate meglumine (Gd-DOTA, Dotarem, Guerbet, France) was administered simultaneously with the start of 20 acquisitions with 12° flip angle and a temporal resolution of 73 seconds. The total acquisition time for DCE-MRI was approximately 24 minutes.

2.1.3. Image Processing

For image preprocessing, we mainly followed the steps described in Heye et al. (22). First, all structural and DCE MR images were coregistered to the 12° pre-contrast image using rigid-body registration to correct for bulk patient movement. All small vessel features were determined according to agreed STRIVE standards (23). We employed a multispectral MRI data fusion and minimum variance quantization method (24) for the segmentation of white matter hyperintensities (WMH) and normal-appearing white matter (NAWM), and the resulting masks were manually refined. We used the “Region of Interest” tool of Analyze 11.0TM (AnalyzeDirect, KS) to semi-automatically outline the old stroke lesions and recent stroke lesion (RSL) boundaries separately. Stroke lesion masks were checked for precision by a neuroradiologist; all other tissue masks were checked visually for accuracy and manually edited by an expert if necessary. Moreover, subcortical/deep gray matter (DGM) masks were generated automatically using a software pipeline as described in Heye et al. (22). In order to minimize any residual contamination of the DGM, the resulting mask was eroded by one voxel. **Figure 3** depicts a representative FLAIR image and corresponding tissue segmentation.

2.2. DCE-MRI Analysis

Data collected at multiple flip angles were first used to calculate the T_{10} map based on the variable flip angle method proposed in Brookes et al. (25), given by

$$\frac{1}{T_{10}} = \frac{1}{T_R} \ln \left(\frac{S_R \sin \alpha_b \cos \alpha_a - \sin \alpha_a \cos \alpha_b}{S_R \sin \alpha_b - \sin \alpha_a} \right), \quad (1)$$

where $S_R = S_a/S_b$ with S_a and S_b denoting the signal intensities of the two pre-contrast acquisitions with flip angles $\alpha_a = 2^\circ$ and $\alpha_b = 12^\circ$, and T_R is the repetition time.

Dynamic DCE image series $S(t)$ are converted to contrast agent concentration $C_t(t)$ through the steady-state spoiled gradient echo (SGPR) signal equation (26),

$$S(t) = \frac{M_0 \sin \alpha_b (1 - e^{-(K+L)})}{1 - \cos \alpha_b e^{-(K+L)}} + \left(S(0) - \frac{M_0 \sin \alpha_b (1 - e^{-K})}{1 - \cos \alpha_b e^{-K}} \right), \quad (2)$$

where $K = T_R/T_{10}$, $L = r_1 C_t(t) T_R$, r_1 is the contrast agent relaxivity taken as $4.2 \text{ s}^{-1} \text{ mM}^{-1}$, $S(0)$ is the baseline (pre-contrast) image intensity, and T_{10} and M_0 are respectively the T_1 relaxation and equilibrium longitudinal magnetization that are calculated from a pre-contrast T_1 mapping acquisition.

For each subject, we extracted a vascular input function (VIF) from a region located on the superior sagittal sinus (SS) because partial volume effects and inflow artifact were reduced at this location compared to obtaining the arterial input function (AIF) from a feeding artery (22); the delay between arterial and venous responses is expected to be very small compared with the temporal resolution of our acquired data. Instead of selecting only a single voxel, we determined a 3×3 patch inside the SS region and estimated the VIF by averaging the time-signal intensities over the voxels within the patch. This enabled us to obtain more smooth variations in the DCE-MRI time course. We converted the whole-blood concentration $C_b(t)$ measured in the SS to plasma concentration using the formula $C_p(t) = C_b(t)/(1 - \text{Hct})$ where Hct is the blood hematocrit measured in large arteries and assumed to be $\text{Hct} = 0.45$ as previously used in literature (22, 26, 27).

2.2.1. Tracer Kinetic Models

Tracer kinetic modeling (28) is applied in DCE-MRI to provide a link between the contrast agent concentration and the physiological or so-called pharmacokinetic parameters, including the fractional plasma volume (v_p), the fractional interstitial volume (v_e), the volume transfer rate (K^{trans}) at which contrast agent (CA) is delivered to the extravascular extracellular space (EES) from plasma space.

In this study, we fitted the following two models to the tissue concentration curves $C_t(t)$: (i) the extended Tofts model, (ii) the Patlak model. A schematic overview of the two models and their relationship is illustrated in **Figure 4**.

The extended Tofts (eTofts) model (29) mainly describes a highly perfused ($F_p = \infty$) two-compartment tissue model considering bidirectional transport between the blood plasma and EES. The concentration of contrast agent in the tissue is determined by,

$$C_t(t) = v_p C_p(t) + K^{\text{trans}} \int_0^t C_p(\tau) e^{-k_{ep}(t-\tau)} d\tau, \quad (3)$$

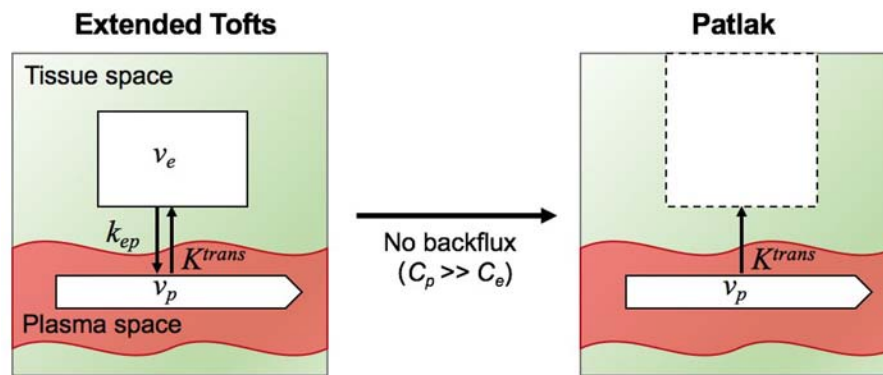


FIGURE 4 | Illustration of two tracer kinetic models: Extended Tofts (left) and Patlak (right) model. Target parameters of DCE-MRI modeling are the contrast agent transfer rate from plasma space to tissue space K^{trans} , the fractional plasma volume v_p , the fractional interstitial volume v_e , and transfer constant from the tissue space to the blood plasma k_{ep} . Patlak model is related to Extended Tofts model through the assumption $C_p \gg C_e$ such that the backflux from the EES into the plasma space is negligible.

where $k_{ep} = K^{\text{trans}}/v_e$ represents the transfer constant from the EES back to the blood plasma. For the fitting of eTofts model, we used limited-memory Broyden-Fletcher Goldfarb-Shannon (LBFGS) method for nonlinear minimization of the sum of squared residuals. The algorithm was run till convergence for a maximum of 30 iterations.

The Patlak Model (30) can be considered as a special case of the eTofts model, where the backflux from the EES into the blood plasma compartment is negligible. To this end, this model only allows measurement of the two parameters K^{trans} and v_p given by,

$$C_t(t) = v_p C_p(t) + K^{\text{trans}} \int_0^t C_p(\tau) d\tau, \quad (4)$$

An attractive feature of Patlak model is that the model equation in (4) is linear and model parameters can be fitted using linear least squares which has a closed-form solution, hence parameter estimation is fast (9).

2.3. Deep Learning for Pharmacokinetic Parameter Estimation

In this study, we consider the PK parameter inference in DCE-MRI as a mapping problem between intensity image-time series and parameter maps where the underlying mapping can be efficiently learned using deep CNNs. The proposed CNN aims at learning data-driven features with the use of convolutional feature filters to effectively detect the local spatio-temporal characteristics of the DCE time series. The extracted spatio-temporal features are desired to represent the underlying relation between the input and output of the network as much as possible.

Specifically, our CNN is trained to learn a mapping between $S(t)$ and θ to output an estimate of PK maps $\hat{\theta}$; $\hat{\theta} = f(S(t)|\mathbf{w})$, where f denotes the forward mapping of the CNN with the learned set of filter weights \mathbf{w} . We note that set of parameters are represented by $\theta = \{K^{\text{trans}}, v_p\}$ for Patlak model and $\theta = \{K^{\text{trans}}, k_{ep}, v_p\}$ for eTofts model.

2.3.1. Loss Function

To learn the network weights (\mathbf{w}) during training, we need to define an objective function (or loss function) to be minimized. In addition to the standard mean squared error (MSE) loss between the true PK parameter values θ and the estimated values $\hat{\theta}$ which enforces high fidelity in parameter reconstruction, we simultaneously seek the fitted contrast agent concentrations of the PK parameters to be sufficiently close to the observed concentrations, $C_t(t)$. To this end, we formulate a new loss function which jointly incorporates these two loss criteria. Given a large number of training samples \mathcal{D} of input-output pairs $(S(t), \theta)$, we train a CNN model that minimizes the following loss,

$$\mathcal{L}(\mathbf{w}) = \sum_{(S(t), \theta) \in \mathcal{D}} \left(\|\theta - \hat{\theta}\|_2^2 + \|C_t(t) - f_{ik}(\hat{\theta})\|_2^2 \right), \quad (5)$$

where f_{ik} is the tracer kinetic model equation of either eTofts or Patlak model as formulated by Equation (3) or Equation (4), respectively.

2.3.2. Network Architecture

We illustrate the network structure used in this study in Figure 5. The network takes DCE image-time series as input with a patch size of $24 \times 24 \times 21$, where time frames are stacked as input channels. The first convolutional layer applies 2D filters to each channel individually to extract low-level temporal features which are aggregated over frames via learned filter weights to produce a single output per voxel. Inspired by the work on brain segmentation (31) and denoising in arterial spin labeling (32), our network consists of parallel dual pathways to efficiently capture multi-scale information after the first layer. The local pathway focuses on extracting details from the local image regions while the global pathway is designed to incorporate more contextual global information. The global pathway consists of 3 dilated convolutional layers with dilation factors of 2, 4, and 8, indicating increased receptive field sizes. Zero-padding is applied before every convolution operation to keep the spatial

dimensions of the output equal to the input. The filter size of each convolutional layer including dilated convolutions is chosen as 4×4 . The rectified linear units (ReLU) activation function ($f(x) = \max(0, x)$) is applied after each convolution to introduce non-linearity into the mapping. Local and global pathways are then concatenated to form a multi-scale feature set. Following this, two fully-connected layers of 256 and 128 hidden nodes are used to determine the best possible feature combination that can accurately map the input to output of the network. Finally, the last fully-connected layer outputs the parameter estimates of a patch size $24 \times 24 \times n$, where n is the number of kinetic model parameters. We emphasize that as our proposed network was structured to estimate outputs for every single voxel of the input patch, it is essential to keep the spatial dimensions of the input and output same throughout the network. Therefore, in our network we can consider a fully-connected (FCN) layer as a convolutional (CONV) layer with 1×1 convolutions.

2.3.3. Network Training

Among all the follow-up scans we only selected one DCE-MRI scan per subject in our experiments. All these scans were acquired at between 1-3 months post-stroke. For each patients data, we neglected the first and last 5 image slices due to insufficient brain coverage. Among the remaining slices of each patient we randomly selected 20 slices to be considered in analysis. We note that these are the central 20 slices that contain most of the brain regions in overall. Following to this, each 2D DCE image slice was divided into overlapping patches of size 24×24 voxels with step size of 6 voxels. This resulted in a collection of approximately 12,000 patches for every patients data. We applied the same procedure on contrast agent concentration data and target parameter maps required for network training.

All experiments were performed in a leave-one-subject-out fashion, i.e., 30 different networks were trained in total based on both Patlak and eTofts model parameters. Randomly chosen 10,000 overlapping patches of each subject were split into training (80%) and validation (20%) sets. The networks were trained using the Adam optimizer with a learning rate of 10^{-3} and a decay rate of 10^{-4} for maximum number of 200 epochs and a mini-batch size of 1000 patches. Early stopping was applied to prevent poor generalization performance when the validation loss stopped improving within consecutive 15 epochs. In **Figure 6** we provide two exemplary plots depicting the changes in training and validation loss over epochs for CNN trained on Patlak and eTofts models. Both losses show a decreasing trend and converge to a minimum. We implemented our code using Keras library with TensorFlow (33) backend, and experiments were run on a NVIDIA GeForce Titan Xp GPU with 12 GB RAM.

2.3.4. Testing

Once the network is trained and network parameters are learned, DCE image-time series data of a test subject can be fed into the network to directly predict the PK parameters. Since the predictions are processed in a patch-wise manner, all overlapping 16 predictions of a neighborhood are averaged to obtain a final value for every individual voxel.

3. RESULTS

3.1. Comparison of Pharmacokinetic Maps

We compare the qualitative PK parameter maps obtained by Patlak model fitting, eTofts model fitting and CNN model trained by either Patlak or eTofts model. **Figures 7B,C** shows PK parameter maps of an exemplary slice of a patients data. In overall, the parameter maps by CNN model looks very similar with the Patlak model fitting. However, the CNN model produces higher estimates of K^{trans} in especially small RSL region as marked on the DCE image in **Figure 7A**. Moreover, the RSL region is more distinctive and can be discriminated well with respect to other tissues in both the parameter maps of CNN model. For numerical evaluation of output parameter maps, we used two evaluation metrics calculated within the entire brain region: Structural similarity index (SSIM) and normalized root mean square error (nRMSE). These values were calculated by considering the output maps of Patlak model as reference, shown in **Figures 7B,C**. For K^{trans} , we obtain a high SSIM of 0.991 and a low nRMSE of 0.0144. For v_p , SSIM is calculated as 0.973 and nRMSE is 0.0168.

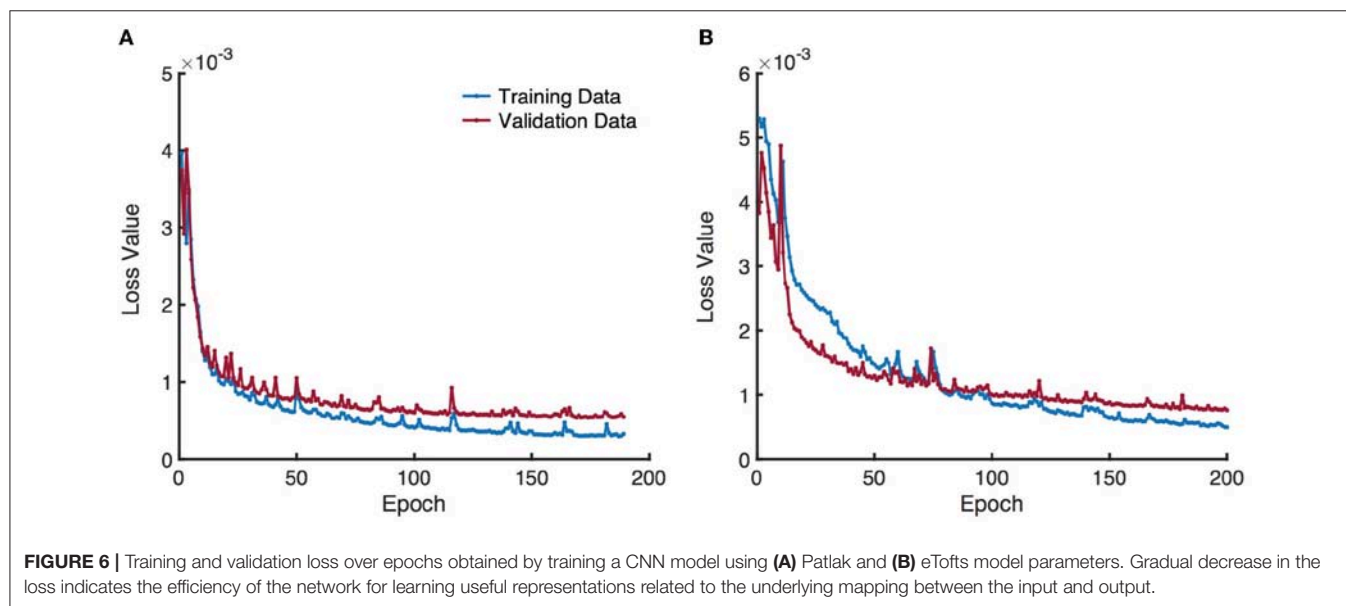
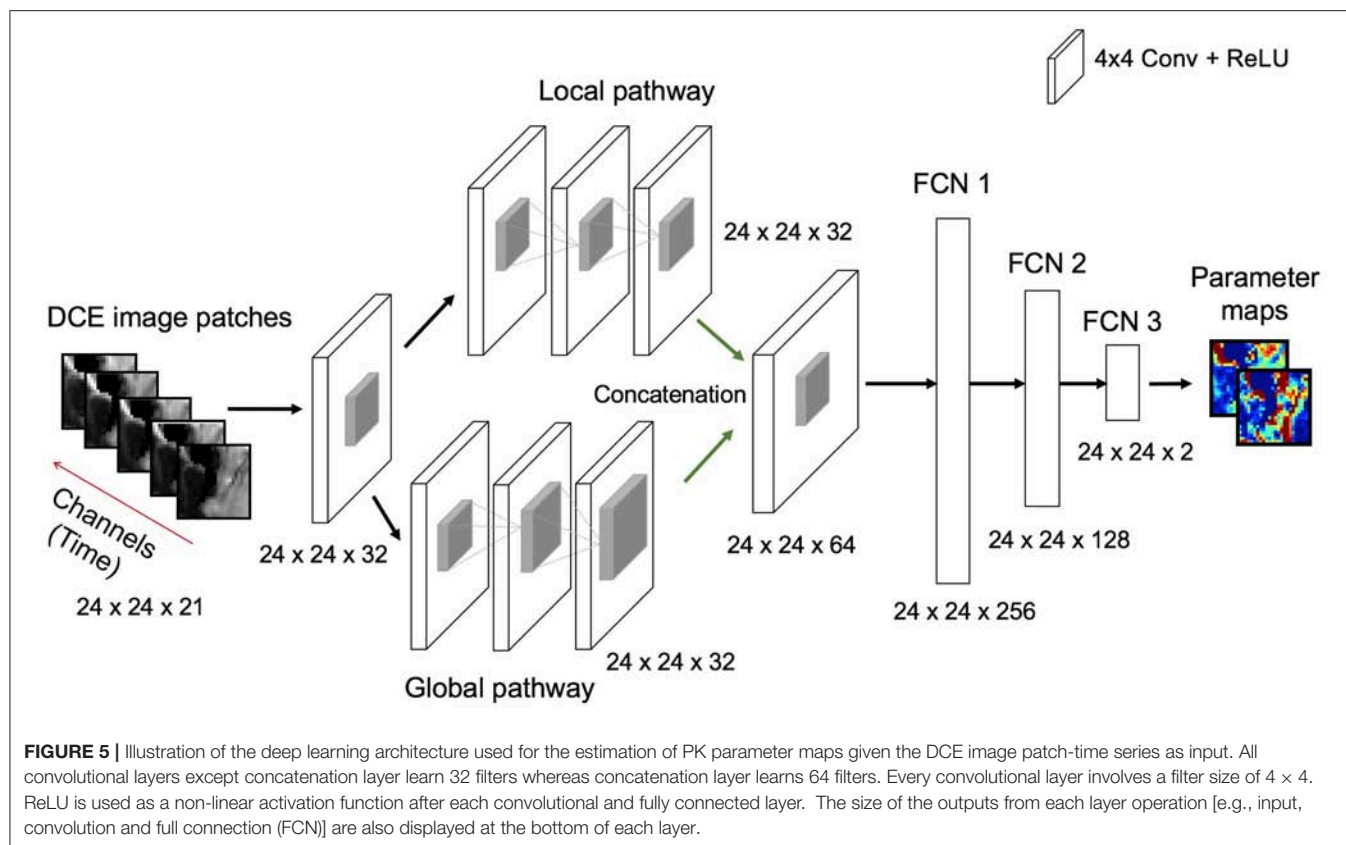
Figures 8B,C demonstrates PK parameter maps of an exemplary slice of an another patients data fitted by eTofts model. The parameter estimates significantly match each other (for CNN and eTofts) in many of the tissue regions except NAWM as depicted on the DCE image in **Figure 8A**. As shown in **Figure 8C**, CNN model yields lower v_p values in comparison to eTofts model in NAWM. Hence, the discrimination of the NAWM with respect to WMH is more prominent. Quantitatively, when compared against the parameter maps obtained by eTofts model, CNN maps yield a SSIM score of 0.998 and 0.961 for K^{trans} and v_p , respectively, while nRMSE is 0.0073 and 0.0156.

3.2. Fitting to the Observed Concentration -Time Series

We evaluate the accuracy of the fitting to the observed concentration-time series data. The fitted contrast agent concentration-time series were estimated via (3) and (4) by using the parameter estimates of Patlak, eTofts, and CNN models separately.

Table 1 demonstrates the quantitative comparison of the fitting to the observed contrast agent concentration time series data for different models in terms of nRMSE and SSIM. The metric values were calculated for every 2D slice of a subject's volume, and statistical values (mean \pm std) were obtained using all 15 subject's data. The results indicate that standard Patlak and eTofts model can fit the data better compared to the CNN model trained with these models separately. However, the difference is not substantial that CNN model still achieves high accuracy with less than an average %2 fitting error.

Figures 9A,B shows the fitting of contrast concentration (in mM) for the NAWM and RSL regions in a single patient data. In general, the CNN model trained by either Patlak or eTofts model parameters can fit the data similarly



well when compared with Patlak and eTofts model. An interesting observation in **Figure 9B** is that the eTofts model does not fit the observed data well whereas the fitting obtained by CNN model trained on eTofts parameters is more accurate.

3.3. Statistical Analysis of PK Parameter Estimation

We perform statistical analysis of the parameter estimates on different tissues. A comparison between tissue types is shown in **Figure 10**. We assessed the statistical significance of the

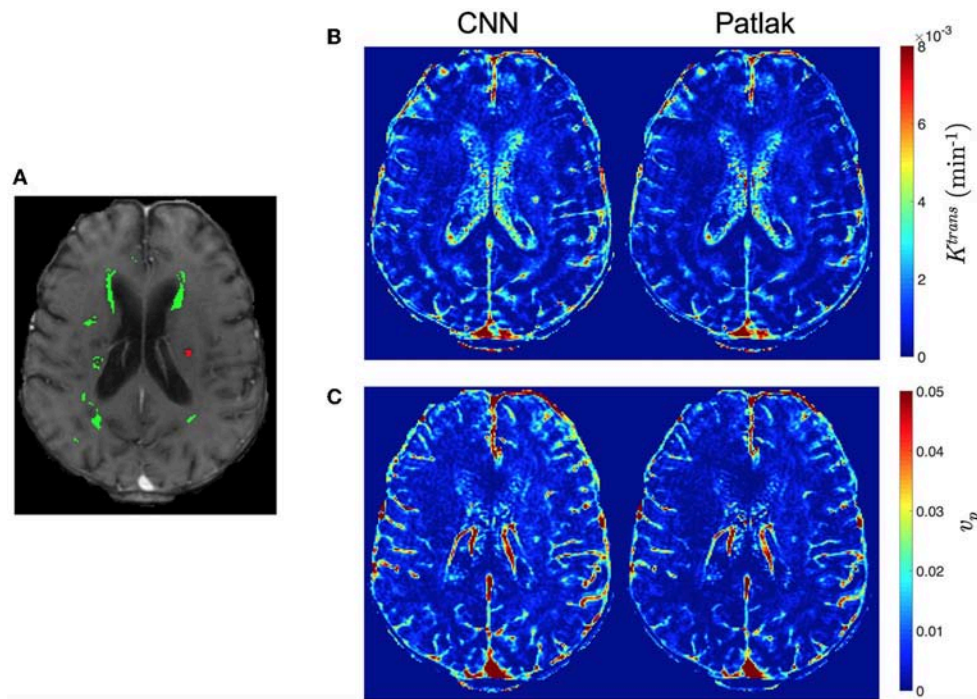


FIGURE 7 | Comparison of qualitative PK parameter maps from a slice of a stroke patient data. **(A)** a DCE image slice on which the tissue masks are superimposed (WMH: green, RSL: red), **(B)** K^{trans} and **(C)** v_p parameter maps obtained by CNN model and Patlak fitting.

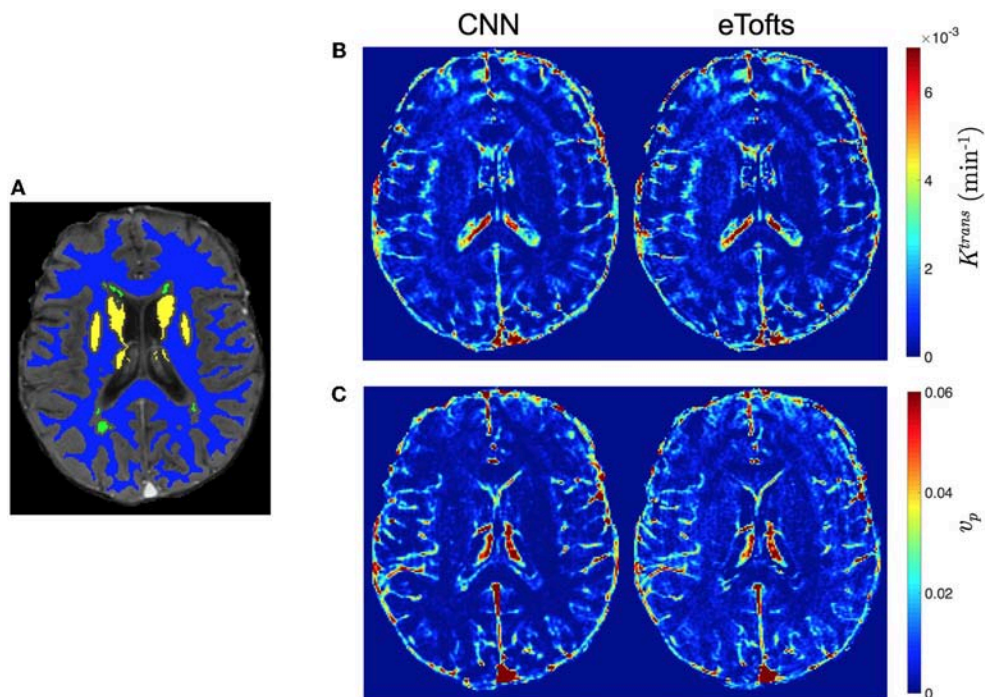


FIGURE 8 | Comparison of qualitative PK parameter maps from a slice of a patient data with white matter hyperintensities. **(A)** a DCE image slice on which the tissue masks are superimposed (NAWM: blue, WMH: green, DGM: yellow), **(B)** K^{trans} and **(C)** v_p parameter maps obtained by CNN model and eTofts fitting. We remark that WMH represents the WM tissue associated with increased risk of dementia and cognitive decline.

TABLE 1 | nRMSE (%) and SSIM statistics (mean \pm std) obtained from concentration-time series data fitting. The SSIM value can vary between -1 and 1 , where 1 indicates perfect similarity.

Metric	Models			
	Patlak	eTofts	CNN: Patlak trained	CNN: eTofts trained
nRMSE (%)	1.1200 \pm 0.5225	1.0575 \pm 0.5744	1.6398 \pm 0.6878	1.7360 \pm 0.7408
SSIM	0.9812 \pm 0.0141	0.9835 \pm 0.0127	0.9750 \pm 0.0138	0.9719 \pm 0.0162

differences using the paired Wilcoxon signed rank test. For Patlak and eTofts model, all differences between tissue types were significant ($p < 0.001$) except for K^{trans} in DGM and WMH, and v_p in WMH and RSL. For CNN model trained on Patlak model parameters, all differences of K^{trans} between tissue types were significant including the difference between WMH and DGM ($p = 3.4 \times 10^{-4}$). The difference between WMH and RSL for v_p is again statistically significant with $p = 1.6 \times 10^{-5}$. The CNN model trained on Patlak generally tends to overestimate the K^{trans} and v_p parameters compared to either Patlak or eTofts model. The difference between them are significant with $p < 0.001$, and this is valid for all tissue types except DGM ($p = 0.021$ for K^{trans}). On the other hand, the CNN model trained with eTofts parameters yield underestimated K^{trans} and overestimated v_p values when compared against either Patlak or eTofts model. The underestimation of K^{trans} by CNN is statistically significant for all tissue types except WMH ($p = 0.317$). The overestimation of v_p by CNN is significant for all tissue types ($p < 0.001$).

Figure 11 depicts the Bland-Altman plots of K^{trans} values in three different tissues (DGM, WMH, RSL) obtained from a patient's data. As can be observed in **Figure 11A**, when compared against the Patlak model, CNN model trained with Patlak tends to slightly underestimate the K^{trans} in DGM and overestimate the values in WMH and RSL. **Figure 11B** indicates that K^{trans} are underestimated by CNN trained with eTofts in DGM and RSL. The values in WMH highly match with Patlak fitting showing no systematic difference. In general, the results in Bland-Altman plots agree with the statistical results as shown in **Figure 10**, meaning that systematic differences are observable between the estimates of CNN and model fitting although concordance correlation coefficients (CCCs) indicate a strong agreement.

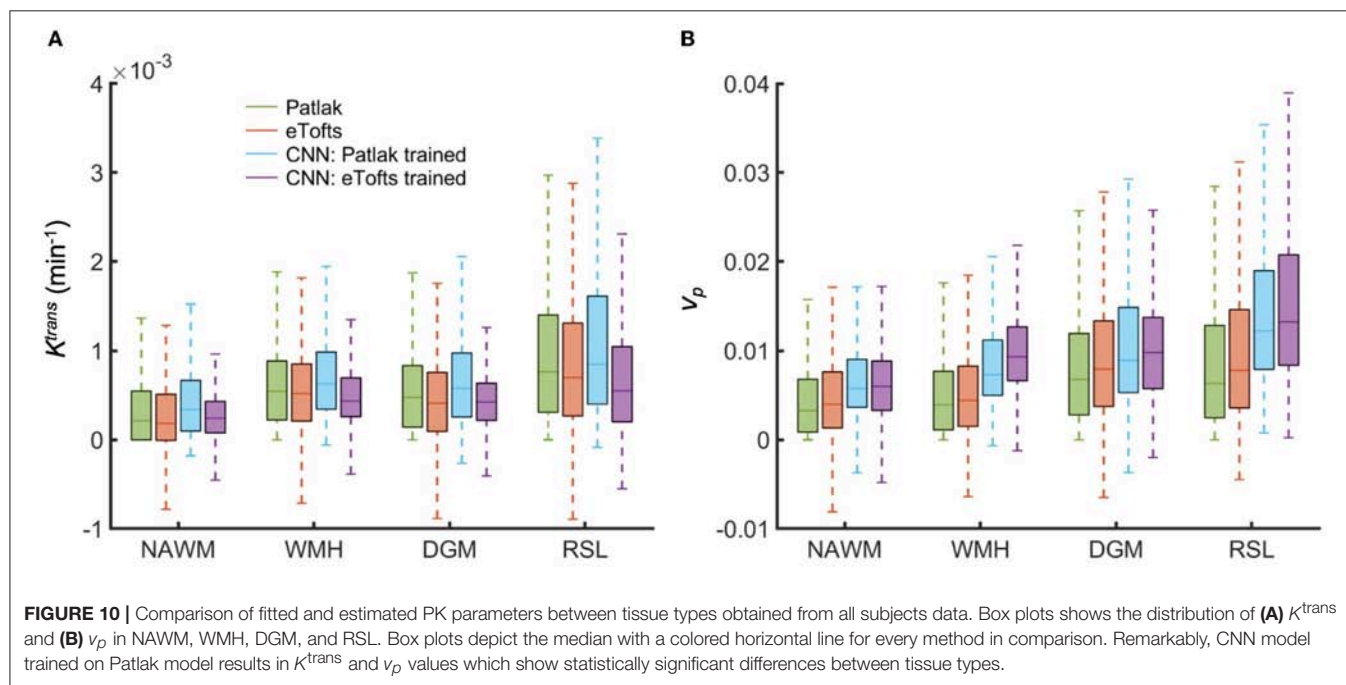
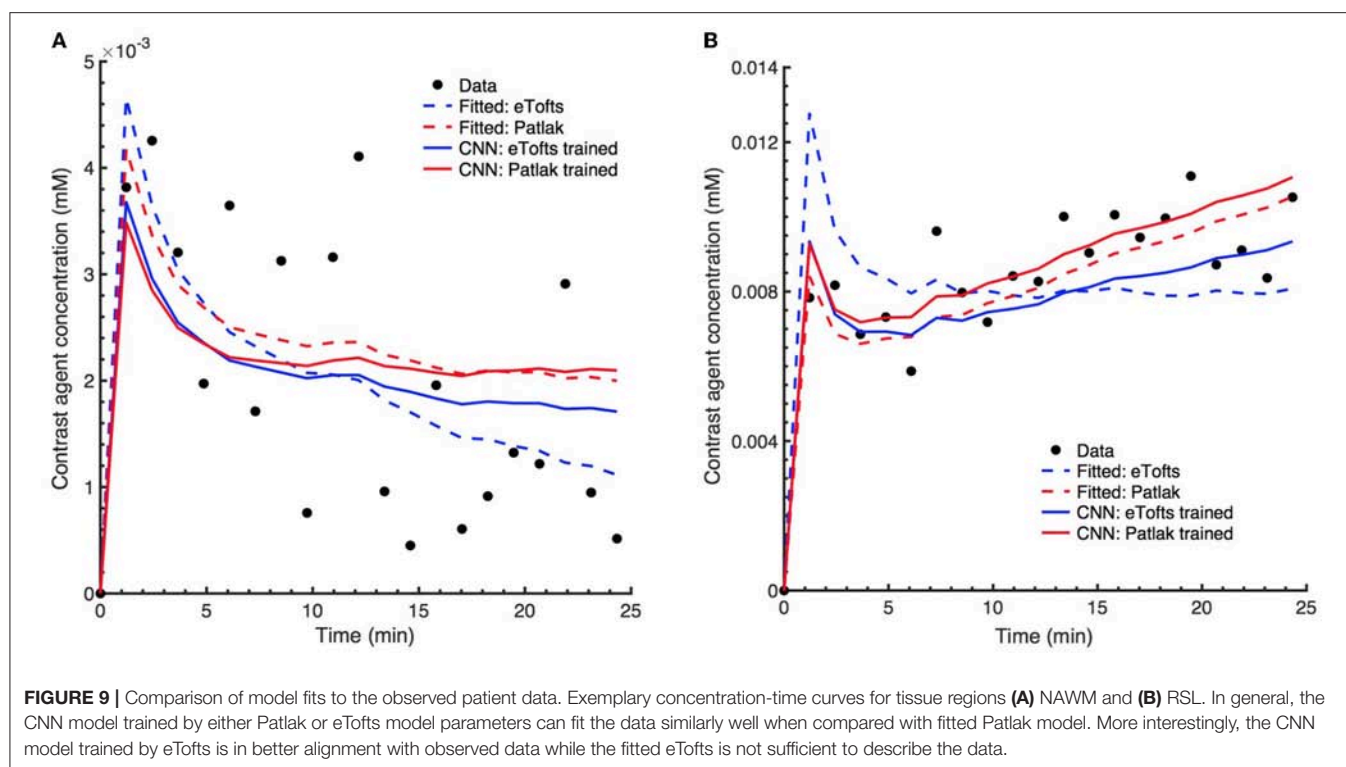
4. DISCUSSION

The results of this study show that a CNN based ML model can yield PK parameter estimates that are comparable to traditional model fitting. As depicted in **Figures 7, 8**, the qualitative parameter maps estimated by CNN models match highly with the ones obtained by conventional TK model fitting methods. Moreover, ML based models can enable better discrimination of different brain tissues. As can be seen in **Figure 7**, small stroke lesion is more visible with higher K^{trans} values assigned to this region. In addition to this, the discontinuities of parameter values arising especially at highly perfused regions (i.e., vessels) can be mitigated by CNN model, and more smoother local areas are produced in these regions as shown in **Figures 7, 8**.

Statistical analysis in **Figure 10** indicate that significant differences between tissue types can be achieved by CNN model whereas both Patlak and eTofts model fail in quantitatively differentiating some of the tissues pairwise including WMH-DGM. Especially higher K^{trans} values are generally assigned to stroke regions i.e., RSL, allowing better discrimination of these areas against non-stroke regions. To this end, the proposed ML model can be an appropriate parameter inference model for quantification of subtle BBB disruption where measuring low-level BBB permeability is vital in several diseases, including cerebral small vessel disease, lacunar stroke and vascular dementia. Another interesting observation is that the plasma volume v_p values estimated by CNN model in WMH are considerably greater than in normal-appearing WM areas. This may result in improved identification of the hyperintensity areas from the surrounding normal appearing WM tissue. WM hyperintensities are usually regarded as surrogates of small vessel disease and frequently seen in elderly people (34).

The major advantage of ML based model is that the parameter inference of a voxel belonging to a specific tissue type is performed by taking into account many other training samples, or voxels, of the same tissue type. Therefore, if the signal time series of a target voxel is subject to high noise, it is likely that a parameter value associated with the voxels that show similar signal trends and located in the same tissue type can be assigned to the target voxel. One example relevant to this observation can be seen in **Figure 9B**, where the fitted concentration time curves are provided for a ROI inside the RSL region of a patients data. Here, the eTofts model does not provide a good fit to the measured signal and the fitted concentration-time curve describes more a vascular region (i.e., blood vessel). However, the fit of the CNN model trained with eTofts model parameters can produce significantly better fit to the observed data, and the fit resembles more an RSL region, which is highly similar with the fits by Patlak and CNN model trained by Patlak model parameters. These findings reveal better generalization ability of ML models (35) which can extract and learn important tissue specific features from a large cohort of training examples. However, it should be noted that the correction of misfit of concentration time curves in **Figure 9B** does not point out an unique feature of our CNN based approach, but rather shows a specific case. The avoidance of a misfit with the CNN network primarily depends upon the model and optimization approach on which the network is trained.

Another observation from **Figure 11** also signifies the tendency of CNN model to produce parameter estimates close to a mean value of parameter distribution learned from



many training voxels within a specific tissue. Here, when compared to the standard Patlak model parameters, we observe overestimated values in especially WMH and RSL region where the K^{trans} usually has higher values. The overestimation in some of the voxels within these tissues is presumably caused by the relatively lower parameter values estimated by

Patlak model due to significant signal noise and fitting to the local minima. In this regard, systematic differences between CNN model estimates and standard NLS fitting are inevitable because the parameter estimates by NLS fitting is not optimal and usually produces a parameter distribution from a high range of values within the voxels of a specific tissue, as it

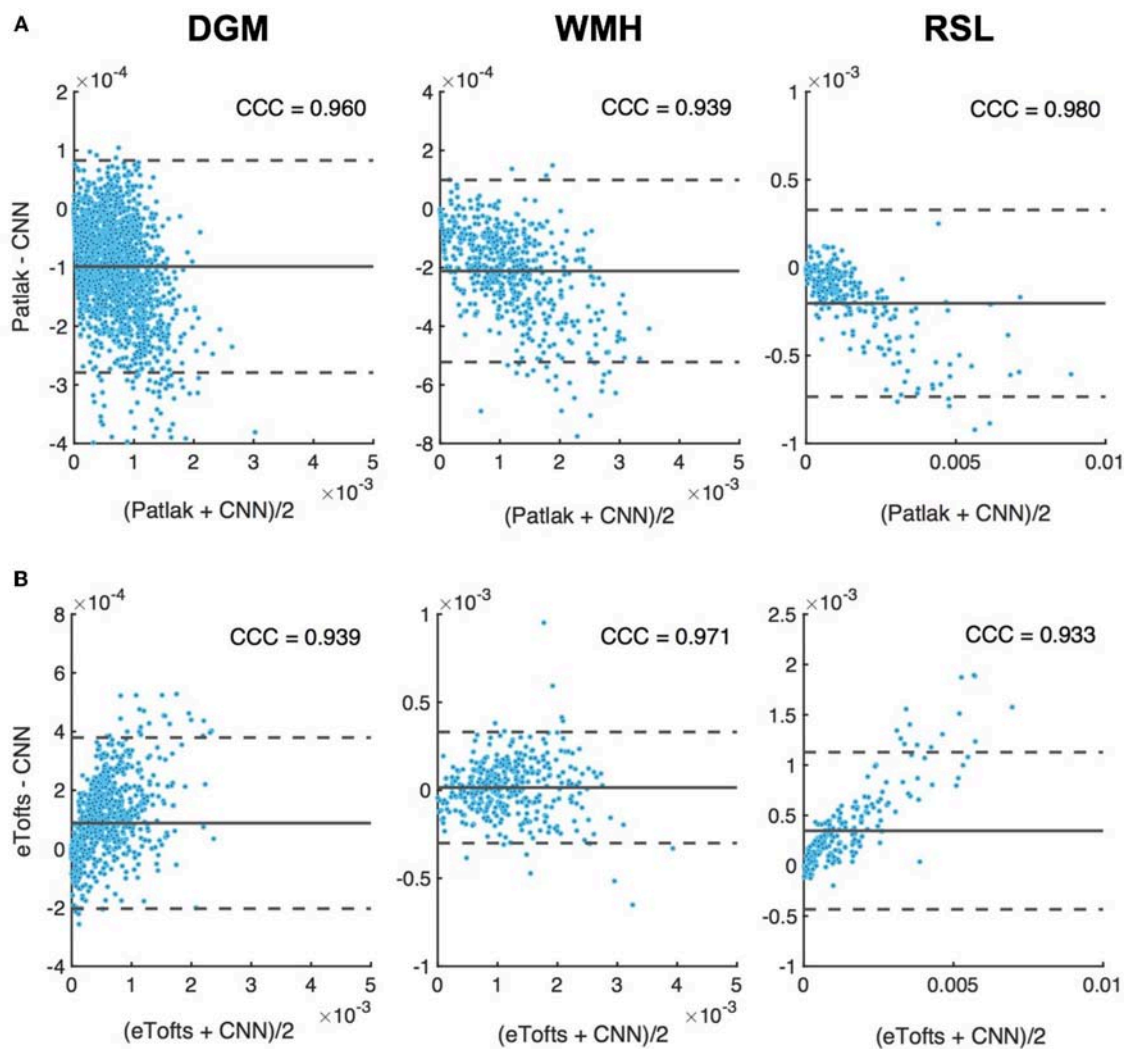


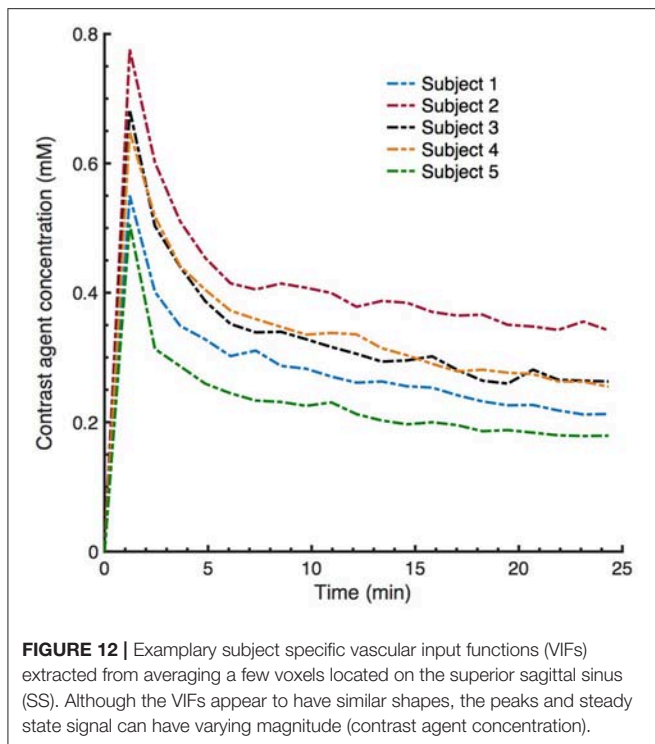
FIGURE 11 | Bland-Altman plots of permeability K^{trans} parameters in different tissues, DGM (left), WMH (middle), RSL (right). Difference (y-axis) in K^{trans} values between CNN model and (A) Patlak, and, (B) eTofts fitting is plotted against the mean values of the two (x-axis). Solid gray line indicates mean difference (*mdiff*). Top and bottom gray dashed lines correspond to upper and lower margins of 95% limits of agreement estimated by $\text{mdiff} \pm 1.96 \times \text{SD}$, SD = standard deviation. Units for horizontal and vertical axes are in min^{-1} . The computed Lin's concordance correlation coefficient (CCC) values are displayed at top-right corner of each plot.

can be seen in **Figure 10**. We anticipate that more accurate evaluation of systematic differences can be obtained using the synthetic DCE dataset where the ground truth parameters are known.

As mentioned before, one of the key advantages of our method is its utility to avoid intermediate computation steps of parameter inference in DCE-MRI by replacing it with a direct inference model. Although we use two existing TK models to estimate the reference parameters, based on the specific DCE application, one can use different TK models in literature (9) to infer the PK parameters to be used during training of the CNN network. If available, the network can be also trained using ground truth parameter values. In addition to this, as previously done in Banerji et al. (36) and Bosca and Jackson (37), synthetic DCE phantom data can be generated by simulating the signal equation

and TK model equations with the PK parameters estimated from real patient's data, and a CNN model can be trained based on the synthetic data and corresponding parameter maps. With this approach, more realistic synthetic DCE datasets can be generated by taking into account the acquisition noise and motion artifacts. The generated synthetic datasets may be utilized to train a network which can be later tested on *in vivo* DCE dataset to obtain less noise-sensitive parameter estimates.

In conventional DCE-MRI analysis pipeline, subject-specific AIF extraction from a ROI of a feeding artery is one of the essential steps for the estimation of kinetic parameters (28, 38). In this study, we demonstrate that CNN based ML model can estimate PK parameters by no need of subject-specific AIF of the test subject without introducing any significant bias in the parameter estimation. Although this can be seen as one of the



benefits of our model, we should remark that the data used in this work is a part of a population study where the temporal resolution and other parameters related to DCE acquisition and contrast injection are fixed in all subjects. However, as can be clearly seen in **Figure 12**, the subject-specific AIFs of our dataset usually have varying magnitudes of the peak and steady-state signal even though the time point where the signal reaches the peak is similar for all subjects. The signal pattern of the AIF curves are directly related to signal time intensities through Equation (2), hence the trained network can intrinsically learn the relation between the AIF and target parameters via the mapping between the input and output of the network and the designed loss function which takes into account the underlying TK model through its equation. On the other hand, the performance of proposed model on a mixed data—ideally involving DCE image series acquired with different acquisition parameters and protocols—can be subject to further investigation. For parameter estimation with a model trained on a mixed data, we anticipate that a bi-CNN input model similar to as proposed for DSC-MRI (21) might be a good approach to avoid bias and error in parameter estimation. In that setting, the DCE image time series and other acquisition parameters—including AIF—of both training and test subjects can be given to the network as two separate inputs.

We emphasize that our CNN model is not trained on a entire brain basis, but on individual time series. Out of the 15 patient datasets we extract more than 160 million training samples, i.e., number of total voxels in the training dataset. Moreover, our network architecture is not very deep and we demonstrate that this huge number of training samples is sufficient to train a network that generalizes well, where the inability to

generate reproducible results is not an issue. Nevertheless, a wider sampling of pathological cases and MRI artifacts in training data is highly desirable and is one of the major direction for improvements of the proposed approach. The proposed model can—even should—be updated accordingly when applied to a larger pool of patient datasets. In general, based on the literature in ML, we anticipate that CNN-based ML models perform better when there is a high correlation and similarity between the training and test data. The dataset used in this study for both training and testing involve voxels from different type of tissues, e.g., healthy and pathological tissues, containing a good mixture of different tissue characteristics. There is a high similarity between the temporal profiles of training and test image patches, hence, the performance of CNN is very stable and robust. However, a poor generalization issue may usually arise in a scenario that the training data only consists of healthy tissue voxels whereas the unseen test data with pathological tissues is tested using the trained model. In this scenario, since the model is not trained with sufficient number of pathological samples, it is quite likely that the CNN model shows a poor performance on these test data comprised of non-healthy tissues. In principle, in order to obtain a stable CNN model, it is necessary to constitute a training data pool according to the demands or expectation from such a prediction model in our specific clinical applications. For instance, if we aim to discriminate well the acute/post-acute stroke regions, our training data should contain high number of voxels from both stroke and non-stroke regions.

Nevertheless, we should discuss the several limitations of this study. First, although ML based methods can have strong generalization ability, the bias is also inevitable when tested on an unseen data because the model is always trained using other subject's data without any access to test data. Second, the performance of our method may be improved depending on the input patch size and filter size of the network. Moreover, we only considered 2D convolution operations, however, 3D convolutions may produce better results when more spatial context information are extracted. Third, further investigation on synthetic data is required to perform accurate assessment of error and bias when the ground truth parameter values are known. Lastly, our current approach is sensitive to variation in acquisition parameters, especially temporal resolution, i.e., number of time points in DCE data. One feasible solution to the variations in temporal resolution across multiple datasets is to apply interpolation on time. In practice, we may interpolate all training data acquired with various temporal resolutions to a common temporal resolution so that a test data with completely different temporal resolution can be also fed into the trained network to produce parameter estimates.

In conclusion, this study shows that a ML based direct inference approach can estimate PK parameters that are comparable to the conventional model fitting in DCE-MRI. Our results, based on a sample of mild ischaemic stroke patients, demonstrate the efficiency of CNN model to enable better discrimination of brain tissue types. Specifically, our proposed ML based method has the potential to improve the current quantitative analysis of DCE-MRI studies due to its increased robustness to noise. Significant difference of

permeability parameters between stroke and non-stroke regions may ultimately effect the stroke medical decision process. Finally, parameter inference of the proposed model on a 3D brain volume is considerably faster than the standard NLS fitting, demonstrating the applicability of such models in clinical practice. Considering such faster computation time, clinical experts may perform parameter inference using various TK models in parallel to benefit from making more detailed analysis between different models.

AUTHOR CONTRIBUTIONS

CU study concept, data analysis, experimental design, writing of manuscript. DD experimental design, writing of manuscript. MT study concept, experimental design. MV image preprocessing. PA and SM data collection. JW funding. BM study concept, writing of manuscript, funding. All authors contributed to reviewing and editing the final manuscript.

REFERENCES

- O'Connor JP, Jackson A, Parker GJ, Roberts C, Jayson GC. Dynamic contrast-enhanced MRI in clinical trials of antivasular therapies. *Nat Rev Clin Oncol*. (2012) 9:167–77. doi: 10.1038/nrclinonc.2012.2
- Yankeelov TE, Gore JC. Dynamic contrast enhanced magnetic resonance imaging in oncology: theory, data acquisition, analysis, and examples. *Curr Med Imaging Rev*. (2007) 3:91–107. doi: 10.2174/157340507780619179
- Collins DJ, Padhani AR. Dynamic magnetic resonance imaging of tumor perfusion. *IEEE Eng Med Biol Magazine* (2004) 23:65–83. doi: 10.1109/EMEMB.2004.1360410
- Farrall AJ, Wardlaw JM. Blood–brain barrier: ageing and microvascular disease-systematic review and meta-analysis. *Neurobiol Aging* (2009) 30:337–52. doi: 10.1016/j.neurobiolaging.2007.07.015
- Wardlaw JM, Smith C, Dichgans M. Mechanisms of sporadic cerebral small vessel disease: insights from neuroimaging. *Lancet Neurol*. (2013) 12:483–97. doi: 10.1016/S1474-4422(13)70060-7
- Heye AK, Culling RD, del C Valds Hernandez M, Thrippleton MJ, Wardlaw JM. Assessment of blood–brain barrier disruption using dynamic contrast-enhanced MRI. A systematic review. *Neuroimage* (2014) 6:262–74. doi: 10.1016/j.neuroimage.2014.09.002
- Parker GJM, Buckley DL. Tracer Kinetic Modelling for T1-Weighted DCE-MRI. In: Jackson A, Buckley DL, Parker GJM, editors. *Dynamic Contrast-Enhanced Magnetic Resonance Imaging in Oncology*. Berlin/Heidelberg: Springer (2005). pp. 81–92.
- Lebel RM, Jones J, Ferre JC, Law M, Nayak KS. Highly accelerated dynamic contrast enhanced imaging. *Magn Reson Med*. (2014) 71:635–44. doi: 10.1002/mrm.24710
- Sourbron SP, Buckley DL. Classic models for dynamic contrast-enhanced MRI. *NMR Biomed*. (2013) 26:1004–27. doi: 10.1002/nbm.2940
- Kelm BM, Menze BH, Nix O, Zechmann CM, Hamprecht FA. Estimating kinetic parameter maps from dynamic contrast-enhanced MRI using spatial prior knowledge. *IEEE Trans Med Imaging* (2009) 28:1534–47. doi: 10.1109/TMI.2009.2019957
- Debus C, Floca R, Nrenberg D, Abdollahi A, Ingris M. Impact of fitting algorithms on errors of parameter estimates in dynamic contrast-enhanced MRI. *Phys Med Biol*. (2017) 62:9322. doi: 10.1088/1361-6560/aa8989
- Schmid VJ, Whitcher B, Padhani AR, Taylor NJ, Yang G. Bayesian methods for pharmacokinetic models in dynamic contrast-enhanced magnetic resonance imaging. *IEEE Trans Med Imaging* (2006) 25:1627–36. doi: 10.1109/TMI.2006.884210
- Wang S, Summers RM. Machine learning and radiology. *Med Image Anal*. (2012) 16:933–51. doi: 10.1016/j.media.2012.02.005
- Das D, Coello E, Schulte RF, Menze BH. Quantification of metabolites in magnetic resonance spectroscopic imaging using machine learning. In: *Medical Image Computing and Computer Assisted Intervention-MICCAI 2017-20th International Conference*. Quebec City, QC (2017) pp. 462–470.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* (2015) 521:436–44. doi: 10.1038/nature14539
- Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems* (Lake Tahoe, NV) (2012). pp. 1097–105.
- Szegedy C, Toshev A, Erhan D. Deep neural networks for object detection. In: *Advances in Neural Information Processing Systems* (Lake Tahoe, NV) (2013). pp. 2553–61.
- Golkov V, Dosovitskiy A, Sperl JI, Menzel MI, Czisch M, Smann P, et al. q-Space deep learning: twelve-fold shorter and model-free diffusion MRI scans. *IEEE Trans Med Imaging* (2016) 35:1344–51. doi: 10.1109/TMI.2016.2551324
- Yang G, Yu S, Dong H, Slabaugh G, Dragotti PL, Ye X, et al. DAGAN: deep De-Aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Trans Med Imaging* (2018) 37:1310–21. doi: 10.1109/TMI.2018.2858779
- Benou A, Veksler R, Friedman A, Raviv TR. Ensemble of expert deep neural networks for spatio-temporal denoising of contrast-enhanced MRI sequences. *Med Image Anal*. (2017) 42(Suppl. C):145–59. doi: 10.1016/j.media.2017.07.006
- Ho KC, Scalzo F, Sarma KV, El-Saden S, Arnold CW. A temporal deep learning approach for MR perfusion parameter estimation in stroke. In: *23rd International Conference on Pattern Recognition (ICPR)* (Cancun) (2016). pp. 1315–20.
- Heye AK, Thrippleton MJ, Armitage PA, Valdés Hernández MDC, Makin SD, Glatz A, et al. Tracer kinetic modelling for DCE-MRI quantification of subtle blood–brain barrier permeability. *Neuroimage* (2016) 125:446–55. doi: 10.1016/j.neuroimage.2015.10.018
- Wardlaw JM, Smith EE, Biessels GJ, Cordonnier C, Fazekas F, Frayne R, et al. Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *Lancet Neurol*. (2013) 12:822–38. doi: 10.1016/S1474-4422(13)70124-8
- Valdés Hernández MdC, Ferguson KJ, Chappell FM, Wardlaw JM. New multispectral MRI data fusion technique for white matter lesion segmentation: method and comparison with thresholding in FLAIR images. *Eur Radiol*. (2010) 20:1684–91. doi: 10.1007/s00330-010-1718-6
- Brookes JA, Redpath TW, Gilbert FJ, Murray AD, Staff RT. Accuracy of T1 measurement in dynamic contrast-enhanced breast MRI using two- and three-dimensional variable flip angle

FUNDING

The research leading to these results has received funding from the European Unions H2020 Framework Programme (H2020-MSCA-ITN-2014) under grant agreement no 642685 MacSeNet and German Research Foundation (DFG)-Project number 326824585 (Personalized treatment of stroke: Improvement of diagnosis through a computer-aided selection of treatment). We acknowledge Wellcome Trust (Grant 088134/Z/09/A) for recruitment and MRI scanning costs. MV is funded by Row Fogo Centre for Research into Aging and the Brain, MT is funded by NHS Lothian Research and Development Office.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the GeForce Titan Xp GPU used for this research.

- fast low-angle shot. *J Magn Reson Imaging*. (1999) 9:163–71. doi: 10.1002/(SICI)1522-2586(199902)9:2<163::AID-JMRI3>3.0.CO;2-L
26. Guo Y, Lingala SG, Zhu Y, Lebel RM, Nayak KS. Direct estimation of tracer-kinetic parameter maps from highly undersampled brain dynamic contrast enhanced MRI. *Magn Reson Med*. (2017) 78:1566–78. doi: 10.1002/mrm.26540
 27. Calamante F, Ahlgren A, van Osch MJ, Knutsson L. A novel approach to measure local cerebral haematocrit using MRI. *J Cereb Blood Flow Metabolism* (2016) 36:768–80. doi: 10.1177/0271678X15606143
 28. Sourbron SP, Buckley DL. Tracer kinetic modelling in MRI: estimating perfusion and capillary permeability. *Phys Med Biol*. (2012) 57:R1–33. doi: 10.1088/0031-9155/57/2/R1
 29. Tofts PS, Brix G, Buckley DL, Evelhoch JL, Henderson E, Knopp MV, et al. Estimating kinetic parameters from dynamic contrast-enhanced t1-weighted MRI of a diffusible tracer: standardized quantities and symbols. *J Magn Reson Imaging* (1999) 10:223–32. doi: 10.1002/(SICI)1522-2586(199909)10:3<223::AID-JMRI2>3.0.CO;2-S
 30. Patlak CS, Blasberg RG, Fenstermacher JD. Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data. *J Cereb Blood Flow Metab*. (1983) 3:1–7. doi: 10.1038/jcbfm.1983.1
 31. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal*. (2017) 36:61–78. doi: 10.1016/j.media.2016.10.004
 32. Kim KH, Choi SH, Park SH. Improving arterial spin labeling by using deep learning. *Radiology* (2018) 287:658–66. doi: 10.1148/radiol.2017171154
 33. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. (2015). Available online at: <https://www.tensorflow.org/>
 34. Wardlaw JM, Valdés Hernández MC, Muñoz-Maniega S. What are white matter hyperintensities made of?: relevance to vascular cognitive impairment. *J Am Heart Assoc*. (2015) 4:e001140. doi: 10.1161/JAHA.114.001140
 35. Bishop CM. *Pattern Recognition and Machine Learning*. New York, NY: Springer (2006).
 36. Banerji A, Caunce A, Watson Y, Rose C, GA B, Parker G. A flexible software phantom for generating realistic dynamic contrast-enhanced MR images of abdominal tumours. In: *Proceedings of International Society of Magnetic Resonance Medicine* (Toronto, ON) (2008). p. 493.
 37. Bosca RJ, Jackson EF. Creating an anthropomorphic digital MR phantom: an extensible tool for comparing and evaluating quantitative imaging algorithms. *Phys Med Biol*. (2016) 61:974. doi: 10.1088/0031-9155/61/2/974
 38. Cheng HM. T1 measurement of flowing blood and arterial input function determination for quantitative 3D T1-weighted DCE-MRI. *J Magn Reson Imaging* (2007) 25:1073–8. doi: 10.1002/jmri.20898

Conflict of Interest Statement: DD is affiliated to GE Global Research as a doctoral student.

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.

Copyright © 2019 Ulas, Das, Thrippleton, Valdés Hernández, Armitage, Makin, Wardlaw and Menze. 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.



Early Identification of High-Risk TIA or Minor Stroke Using Artificial Neural Network

Ka Lung Chan¹, Xinyi Leng^{1,2*}, Wei Zhang³, Weinan Dong³, Quanli Qiu³, Jie Yang⁴, Yannie Soo¹, Ka Sing Wong¹, Thomas W. Leung¹ and Jia Liu^{3*}

¹ Department of Medicine and Therapeutics, Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong, China, ² Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China, ³ Shenzhen Institutes of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen, China, ⁴ Department of Neurology, The Second Affiliated Hospital of Guangzhou Medical University, Guangzhou, China

OPEN ACCESS

Edited by:

Fabien Scalzo,
University of California, Los Angeles,
United States

Reviewed by:

Ruogu Fang,
University of Florida, United States
Maurizio Acampa,
Azienda Ospedaliera Universitaria
Senese, Italy

*Correspondence:

Xinyi Leng
xinyi_leng@cuhk.edu.hk
Jia Liu
jia.liu@siat.ac.cn

Specialty section:

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

Received: 01 April 2018

Accepted: 08 February 2019

Published: 01 March 2019

Citation:

Chan KL, Leng X, Zhang W, Dong W, Qiu Q, Yang J, Soo Y, Wong KS, Leung TW and Liu J (2019) Early Identification of High-Risk TIA or Minor Stroke Using Artificial Neural Network. *Front. Neurol.* 10:171. doi: 10.3389/fneur.2019.00171

Background and Purpose: The risk of recurrent stroke following a transient ischemic attack (TIA) or minor stroke is high, despite of a significant reduction in the past decade. In this study, we investigated the feasibility of using artificial neural network (ANN) for risk stratification of TIA or minor stroke patients.

Methods: Consecutive patients with acute TIA or minor ischemic stroke presenting at a tertiary hospital during a 2-year period were recruited. We collected demographics, clinical and imaging data at baseline. The primary outcome was recurrent ischemic stroke within 1 year. We developed ANN models to predict the primary outcome. We randomly down-sampled patients without a primary outcome to 1:1 match with those with a primary outcome to mitigate data imbalance. We used a 5-fold cross-validation approach to train and test the ANN models to avoid overfitting. We employed 19 independent variables at baseline as the input neurons in the ANN models, using a learning algorithm based on backpropagation to minimize the loss function. We obtained the sensitivity, specificity, accuracy and the c statistic of each ANN model from the 5 rounds of cross-validation and compared that of support vector machine (SVM) and Naïve Bayes classifier in risk stratification of the patients.

Results: A total of 451 acute TIA or minor stroke patients were enrolled. Forty (8.9%) patients had a recurrent ischemic stroke within 1 year. Another 40 patients were randomly selected from those with no recurrent stroke, so that data from 80 patients in total were used for 5 rounds of training and testing of ANN models. The median sensitivity, specificity, accuracy and c statistic of the ANN models to predict recurrent stroke at 1 year was 75%, 75%, 75%, and 0.77, respectively. ANN model outperformed SVM and Naïve Bayes classifier in our dataset for predicting relapse after TIA or minor stroke.

Conclusion: This pilot study indicated that ANN may yield a novel and effective method in risk stratification of TIA and minor stroke. Further studies are warranted for verification and improvement of the current ANN model.

Keywords: transient ischemic attack, minor stroke, artificial neural network, risk stratification, prognosis

INTRODUCTION

The prevalence of transient ischemic attack (TIA) is estimated to be 103.3 per 100,000 in the Chinese population (1). Although TIA may be regarded as a “benign” cerebrovascular event, subsequent stroke could be disabling. Studies conducted over 15 years ago reported that 12–20% of TIA or minor stroke patients would have a recurrent stroke within 3 months (2). The risk of stroke recurrence in such patients has been declining over the past decade partly due to advances in the stroke service system—5.1% of TIA or minor stroke patients in hospitals with dedicated TIA/minor stroke service systems had a recurrent stroke within 1 year in the recently published TIAregistry.org Project, but there have been concerns that the recurrent rate is probably higher in routine clinical practice (3). In the Clopidogrel in High-Risk Patients with Acute Nondisabling Cerebrovascular Events (CHANCE) trial, 10.0% of 5,170 minor stroke patients or TIA patients with an ABCD² score ≥ 4 recruited from 114 hospitals of different levels in China had a recurrent stroke at 3 months, despite of early-initiated dual or mono antiplatelet treatment (4).

A few risk scores have been developed to identify high-risk TIA or minor stroke patients, for instance, the ABCD² score, which has been commonly used in research and in clinical practice. Patients with an ABCD² score ≥ 4 are generally considered as high-risk patients (5, 6). However, recent studies indicated that ABCD² score may not reliably differentiate TIA or minor stroke with mimics, or those at high or low risk of recurrent stroke (7). Moreover, those with a ABCD² score < 4 and ≥ 4 could have similar risks of recurrent stroke at 3 months (8). Other factors have been considered to supplement the ABCD² score, for instance, presence of new infarct(s) and carotid arterial stenosis and dual TIA, to form the ABCD³-I score (9). These new scoring systems have been well validated in other populations, reporting the c-statistics of 0.60–0.64 in predicting the recurrent stroke within 3 month following TIA. However, new scores such as ABCD³-I score have not been recommended for risk stratification in such patients by the guidelines by far (10, 11).

In the current study, we aimed to use a novel method to predict the risk of stroke recurrence in TIA or minor stroke patients, which was the artificial neural network (ANN) technique. It is a commonly used machine learning algorithm to form a diagnostic or risk prediction model, which typically consists of three layers of neurons, an input layer of independent variables, a hidden layer with no real-world meaning but allowing nonlinear interactions among the input variables, and an output layer for the probability of an outcome (Figure 1). Advantages of ANN over conventional statistical methods in forming a risk prediction model lie in that it detects interactions between the input variables that commonly exist in clinical studies, and that it takes into account the weights of input variables in their correlations with the outcome (12, 13).

Previous studies have used ANN to diagnose acute myocardial infarction and stroke, and to predict mortality in ischemic stroke, intracerebral hemorrhage, traumatic brain injury, etc., which demonstrates improved accuracy against conventional methods in most circumstances (14–18). However, to our knowledge,

ANN had not been applied in predicting the risk of recurrence following a TIA or minor stroke. Therefore, in this pilot study, we developed and tested ANN models for risk stratification of TIA or minor stroke. In addition, other algorithms for machine learning such as Support Vector Machine (SVM) and Naïve Bayes classifiers (NBC) have also been utilized in medical research (19, 20). For instance, SVM has been extensively applied in diagnosis of a disease or classification of groups with certain features based on imaging data (20). The NBC algorithm was able to diagnose carpal tunnel syndrome with the highest detection rate among four machine learning methods (21). Thus, in the current cohort, we also compared the performance of ANN with SVM and NBC in risk stratification of TIA and minor stroke patients.

MATERIALS AND METHODS

Study Design and Subjects

Consecutive patients with acute TIA or minor ischemic stroke presenting at a tertiary hospital between January 2004 and December 2005 were recruited. TIA was defined as a transient episode of neurological dysfunction caused by focal brain ischemia, which completely resolved within 24 h. A minor ischemic stroke was diagnosed as sudden onset of neurological deficits caused by brain ischemia lasting longer than 24 h, with admission NIHSS score of 0–3. Diagnosis of TIA and minor ischemic stroke was made by the neurologists in charge. Stroke or TIA mimics such as toxic metabolic syndrome, seizure, migraine, demyelinating disorders, drug ingestion were excluded (22). We collected patients' characteristics at baseline as detailed below, including well-established factors readily available in clinical practice that might be associated with stroke recurrence. TIA and minor stroke patients were regularly followed up at the outpatient clinic, when recurrent cerebral ischemic events and other events were recorded. The primary outcome was defined as recurrent ischemic stroke within 1 year, as confirmed with CT or MR imaging, or diagnosed by the neurologist in charge. We developed and tested ANN models based on patients' characteristics to predict the primary outcome. We also conducted conventional statistical analyses for independent predictors for the primary outcome. The study was approved by the Joint Chinese University of Hong Kong–New Territories East Cluster Clinical Research Ethics Committee (The Joint CUHK-NTEC CREC).

Data Collection

We collected demographic characteristics (sex, age) and vascular risk factors (smoking, hypertension, diabetes mellitus, dyslipidemia, prior TIA or ischemic stroke, atrial fibrillation, ischemic heart disease) at baseline. We also collected certain clinical characteristics including systolic and diastolic blood pressure at admission, National Institutes of Health Stroke Scale (NIHSS) score at admission, premorbid modified Rankin Scale (mRS) score, symptom duration (for transient deficits) and symptom type (unilateral weakness and slurring speech). We also gathered medications prescribed at discharge, such

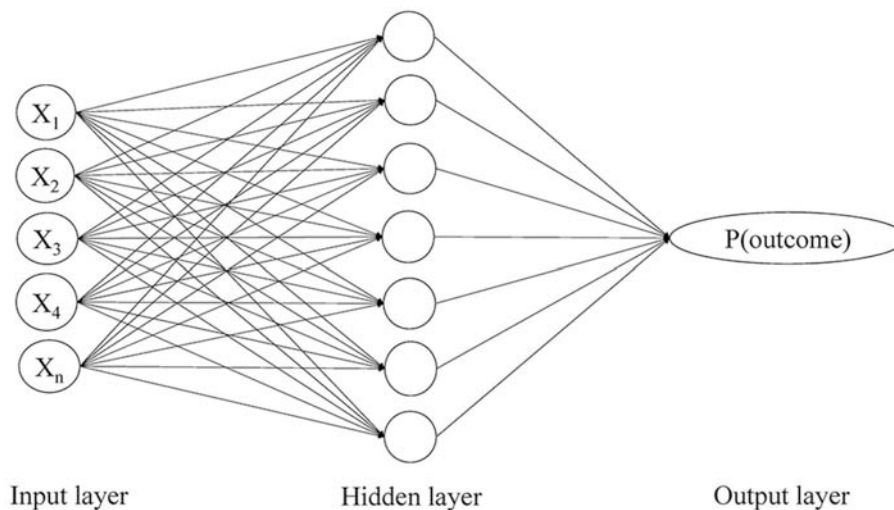


FIGURE 1 | The three-layer perceptron artificial neural network model showing input, hidden and output layers and nodes with feed forward links. P(outcome) refers to the probability of an outcome.

as antiplatelets, antihypertensive, anticoagulants, antidiabetics and statins.

We also collected neuroimaging features including new infarct(s) on brain CT or magnetic resonance imaging (MRI), and findings by cerebrovascular workup including presence of extra-/intra-cranial arterial stenosis. Extracranial arterial stenosis was defined as at least 50% narrowing of the internal carotid artery or vertebral artery lumen on carotid duplex ultrasound, CT or MR angiography, or digital subtraction angiography, by the NASCET method (23). Intracranial artery stenosis was defined as at least 50% narrowing of middle cerebral artery, anterior cerebral artery, posterior cerebral artery, intracranial segment of internal carotid artery and vertebrobasilar arteries in transcranial Doppler, cerebral CT or MR angiography, or digital subtraction angiography, using WASID method (24). We defined large artery stenosis as either extracranial arterial stenosis or intracranial arterial stenosis.

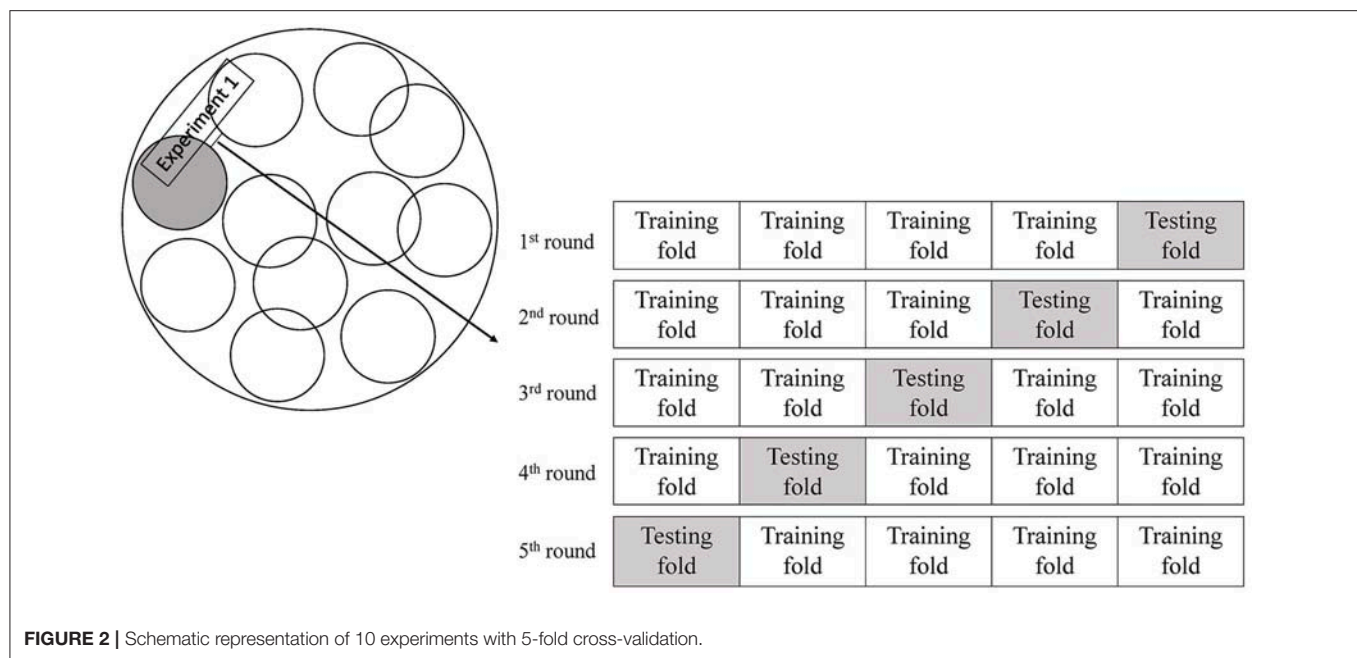
Training and Testing in ANN

We employed a three-layer multilayer perceptron (MLP) model in this study, a most common type of ANN. It was composed of an input layer of 19 independent variables (**Supplemental Table S1**), a hidden layer with a certain number of neurons that were adjusted through training, and an output layer representing the probability of the primary outcome (**Figure 1**). A backpropagation algorithm was used to minimize the loss function by iteratively updating the weights between the neurons and thus maximize the predictive power of the ANN model for the primary outcome. Loss function represented the inconsistency between the predictive and actual values. Within each iteration of the backpropagation algorithm, the partial derivatives of the loss function with respect to each weight were propagated backward from the output layer and passed through

the hidden layer, which eventually adjusted all the weights back to the input neurons.

The study cohort was imbalanced in the numbers of patients with or without a primary outcome. This may cause biased prediction toward the no-recurrence group. Therefore, we randomly selected the same number of patients without a primary outcome as patients with a primary outcome for the training and testing (1:1 matched). In view of the relatively small sample size, we used a 5-fold cross-validation approach to train and test the ANN models (**Figure 2**), to avoid overfitting of the models (25, 26). The dataset was randomly partitioned into 5 folds, and we performed 5 rounds of training and testing of the ANN models. In each round of the experiments, 4 folds were the training subsets and the remaining subset was retained to test the ANN model. Each of the 5 subsets was only used once as the testing set in the 5-fold cross-validation process. We implemented the training and testing procedures for the ANN models in Matlab. We trained and tested ANN models with 4, 6, 8, 10, and 12 hidden neurons, respectively, and finalized the number of hidden neurons when the ANN model reached a minimal loss function in each round of the cross-validation experiments. We repeated such cross-validation procedures for 10 times (10 Experiments, **Figure 2**), each time with a new group of randomly selected patients without a primary outcome 1:1 matched with patients with a primary outcome.

For continuous independent variables, data normalization was required to speed up the gradient, thus to find the optimal solution. Shapiro-Wilk test was performed to determine whether a continuous independent variable was normally distributed. If not normally distributed, we scaled the data to a range of 0–1 before entering an ANN model. To be efficient in achieving the minimization of the loss function, we employed an adaptive moment estimation (Adam) rate rather than the constant learning rate. In addition, we assigned a weight between



−1 and +1 for the independent variables to speed up the learning process and escape local minima. We obtained the sensitivity, specificity, accuracy, and the c statistic of the ANN models developed in each of the 5 rounds of training and testing, in each of the 10 cross-validation experiments; we then computed the overall medians (interquartile range, IQR) of sensitivity, specificity, accuracy and c statistic of the models in all of the 50 rounds of training and testing.

Training and Testing in SVM and NBC

For training the SVM models, we tried a few standard kernel functions, such as sigmoid, Gaussian, polynomial and linear kernel, and adjusted different values of the parameter C in the range of 0.1–10. We finally determined the kernel function and the value of the parameter C by the highest accuracy in predicting the test set.

Additionally, using the Naïve Bayesian equation to calculate the posterior probability for each class, we first computed the prior probability of each class $P(c)$, of each predictor $P(x)$, and the likelihood which is the probability of predictor given class $P(x|c)$. The outcome of prediction was the class with the highest posterior probability, using the Maximum A Posteriori (MAP) estimation.

In accordance with the ANN, we also applied the 5-fold cross-validation approach for both SVM and NBC and repeated it for 10 times with randomly selected patients without a primary outcome to 1:1 match the patients with a primary outcome. The overall medians (IQR) of sensitivity, specificity, accuracy and c statistic were also calculated for SVM models and NBC. We used Kruskal–Wallis test to compare the overall medians of sensitivity, specificity, accuracy and c statistic among the three machine learning algorithms and *post-hoc* comparisons between any two of these algorithms.

Other Statistical Analyses

In addition, we also conducted conventional statistical analyses for predictors for the primary outcome in the study cohort. Continuous variables were presented as medians (interquartile range [IQR]), whilst categorical variables were presented as numbers (percentage). For univariate comparisons between patients with and without the primary outcome, continuous variables were analyzed with independent *t*-tests or Mann–Whitney *U*-test, whilst categorical variables were analyzed with χ^2 -test or Fisher's exact test. To identify factors independently associated with the primary outcome, variables with $p < 0.1$ in univariate analyses were entered in a multivariate logistic regression model for further analysis. Odds ratio (OR) and 95% confidence interval (CI) were calculated. $P < 0.05$ were considered statistically significant. All the conventional statistical analyses were conducted in IBM SPSS Statistics version 22.0 (SPSS Inc., Chicago, IL, United States).

RESULTS

In total, 451 patients were recruited; 201 (44.6%) patients had a TIA and the remaining had a minor stroke as an index ischemic event. The median NIHSS was 1 (IQR 0–2). Forty (8.9%) patients had the primary outcome of recurrent ischemic stroke within 1 year. Twelve patients died within 1 year, three of which developed recurrent ischemic stroke before death; and the remaining nine patients died from other reasons. More of the patients with a primary outcome had a history of TIA (15 vs. 5.4%, $p = 0.038$), and extra- and/or intra-cranial large artery stenosis (62.5 vs. 35.8%, $p = 0.001$), compared with patients without a primary outcome event (**Table 1**). Other baseline characteristics or medications prescribed at discharge were not significantly different between those with and without

TABLE 1 | Characteristics of patients with or without recurrent ischemic stroke within 1 year.

Variables	Recurrence (n = 40)	No recurrence (n = 411)	P-value
Age, years	71 (60–74)	65 (55–73)	0.116
Male	20 (50.0)	231 (56.2)	0.451
History of hypertension	29 (72.5)	256 (62.3)	0.201
History of diabetes mellitus	13 (32.5)	123 (29.9)	0.735
History of dyslipidemia	21 (52.5)	252 (61.3)	0.276
History of AF	7 (17.5)	37 (8.9)	0.084
History of ischemic stroke	10 (25.0)	68 (16.5)	0.177
History of TIA	6 (15.0)	22 (5.4)	0.038
History of ischemic heart disease	4 (10.0)	39 (9.5)	0.916
Smoker	8 (20.0)	142 (34.5)	0.062
Unilateral weakness	25 (62.5)	226 (55.0)	0.213
Slurring speech	15 (37.5)	107 (26.0)	0.361
Symptom duration			0.355
≤10 min	3 (7.5)	63 (15.3)	
11–60 min	4 (10)	50 (12.2)	
>60 min	33 (82.5)	298 (72.5)	
Systolic blood pressure, mmHg	166 (146–182)	160 (143–180)	0.453
Diastolic blood pressure, mmHg	80 (71–97)	83 (74–94)	0.949
NIHSS at admission	1 (0.25–2)	1 (0–2)	0.893
Premorbid mRS	0 (0–0)	0 (0–0)	0.125
Large artery stenosis	25 (62.5)	147 (35.8)	0.001
New infarct	11 (27.5)	102 (24.8)	0.709
Antiplatelet(s)	29 (72.5%)	320 (77.9%)	0.916
Anticoagulant	0 (0.0%)	9 (2.2%)	1.000
Antihypertensives	19 (47.5%)	209 (50.9%)	0.851
Antidiabetics	6 (15.0%)	92 (22.4%)	0.409
Statins	16 (40.0%)	182 (44.3%)	0.984

AF indicates atrial fibrillation; TIA indicates transient ischemic attack; NIHSS indicates National Institute of Health Stroke Scale; mRS indicates modified Rankin Scale.

a primary outcome in the study cohort (Table 1). History of TIA, large artery stenosis, atrial fibrillation and smoking were further analyzed in multivariate logistic regression to predict the primary outcome. Only presence of large artery stenosis (OR: 2.87; 95% CI: 1.45–5.67; $p = 0.002$) was significantly associated with recurrent ischemic stroke within 1 year following a TIA or minor stroke in multivariate analysis.

In each of the 10 experiments of developing and testing the ANN models with 5-fold cross-validation, 40 patients were randomly selected from those without a primary outcome to 1:1 match with the 40 patients with a primary outcome (Figure 2). The number of neurons in the hidden layer was finalized as 10, after testing 4, 6, 8, 10, and 12 hidden neurons in the models. The median sensitivity, specificity and accuracy of the ANN models was 75% (63.3–83.3%), 75% (62.5–83.3%), and 75% (68.8–76.6%), respectively. The median c statistic was 0.77 (0.68–0.84) (Table 2).

After testing several kernel functions and values of parameter C, we found that the SVM model with the linear kernel

and parameter C equaling to 1 was optimal among all the others. The median sensitivity, specificity and accuracy of the SVM models was 62.5% (50–62.5%), 75% (50–87.5%), and 62.5% (56.3–68.8%).

Moreover, we also calculated the posterior probability for each class based on the Naïve Bayesian equation, and selected the outcome with highest probability. The median sensitivity, specificity and accuracy of the NBC was 62.5% (50–75%), 75% (62.5–75%), and 62.5% (56.3–68.8%). The performance of ANN models in identifying patients with recurrent ischemic stroke was better than that of SVM and NBC algorithm (Table 2).

DISCUSSION

This pilot study demonstrated the feasibility of using ANN to predict the risk of recurrent stroke within 1 year after a TIA or minor stroke, based on parameters that are readily available in clinical practice. With a relatively small sample size and a smaller number of the primary outcome event, conventional univariate and multivariate analyses only identified the presence of cervico-cerebral large artery stenosis as an independent predictor for stroke recurrence within 1 year. However, the ANN models developed based on this study cohort showed moderate-to-good accuracy in predicting the primary outcome in comparison with SVM model and Naïve Bayes classifier, which suggested ANN as an alternative or even more effective approach for risk stratification of TIA or minor stroke.

Despite of the small sample size, presence of extra- and/or intra-cranial large artery stenosis was identified as an independent risk factor of recurrent stroke in the current study. This was consistent with relevant findings in the TIAregistry.org project and other previous studies. For instance, in the TIAregistry.org project, major brain imaging findings, including ≥ 1 acute ischemic lesion and ≥ 1 intra- and/or extra-cranial stenosis $>50\%$, were associated with increased risk of stroke recurrence at 3 months or 1 year after a TIA or minor stroke (3). Particularly in subjects recruited from Asia in the TIAregistry.org project, presence of intracranial stenosis tended to increase the 1-year stroke risk, independent of other confounding factors ($p = 0.09$) (27). Therefore, the current study reinforced intracranial stenosis as a strong risk factor for stroke recurrence in Asians.

ANN models developed in the current study showed moderate-to-good accuracy in predicting the primary outcome, while we used a 5-fold cross-validation approach to avoid overfitting of the models. Previous studies also found ANN models accurate and effective in differentiating cerebral ischemia from stroke mimics, and in predicting mortality in patients with intracerebral hemorrhage, etc. (14, 17). As mentioned above, ANN possessed advantages over conventional statistical methods in forming a risk prediction or diagnostic model. ANN could detect complex nonlinear relationships between independent and dependent variables and assign weights to the independent variables in their associations with the outcome, thus enhancing the model fit as compared with logistic regression methods (13).

TABLE 2 | Predictive performance of ANN, SVM and NBC models.

Statistics*	ANN	SVM	P-value (ANN vs. SVM)	Naïve Bayes	P-value (ANN vs. NBC)	P-value (overall comparison)
Sensitivity	75% (63.3–83.3%)	62.5% (50–62.5%)	<0.001	62.5% (50–75%)	0.001	<0.001
Specificity	75% (62.5–83.3%)	75% (50–87.5%)	0.081	75% (62.5–75%)	0.121	0.151
Accuracy	75% (68.8–76.6%)	62.5% (56.3–68.8%)	<0.001	62.5% (56.3–68.8%)	<0.001	<0.001
C statistic	0.77 (0.68–0.84)	0.63 (0.56–0.69)	<0.001	0.63 (0.56–0.69)	<0.001	<0.001

*presented with medians (IQR).

ANN indicates artificial neural network; SVM indicates support vector machine; NBC indicates Naïve Bayes classifier.

ANN could also take into account possibly complex interactions between the independent variables, (28) which commonly exist in clinical scenarios, e.g., interactions between age and presence of the vascular risk factors. Moreover, in the era of precision medicine, simple dichotomization of a factor as a continuous variable in nature (e.g., age and blood pressure) in conventional scoring systems may not accurately reflect the effects of these variables in determining the risk of stroke recurrence, while the ANN approach could accommodate variables as they are in the risk prediction models.

Our results showed that the ANN outperformed the SVM and NBC. For the SVM, we tested the standard kernel functions and the best accuracy was achieved with merely 62.5% by a linear kernel. This suggests that the standard nonlinear kernel functions we tested might not be appropriate for projecting the data into a space where they can be classified by an SVM. Finding a better kernel function for this particular problem is however not intuitive and is not in the scope of this study. In contrast, the major advantage of the proposed ANN is that projecting the data into a space for classification is driven by the data. Thus, there is no need to pre-define a kernel function. For the NBC, the assumption of independence between the input variables might not be well satisfied for this study. This can negatively affect the accuracy of the NBC. Though this can possibly be improved by carefully selecting the input variables, we however did not try this procedure in order to show the ANN is an end-to-end approach that does not require data pre-selection.

The present study had several limitations. It was a retrospective single-center study with data collected years ago, but this pilot study indicated potential application of ANN in risk stratification of TIA and minor stroke patients. In addition, the study cohort was inevitably imbalanced in view of the numbers of patients with and without the primary outcome. We mitigated the imbalance between the two groups by randomly down-sampling the no-recurrence cases. However, useful information may be discarded by such resampling, and the cases randomly selected may not represent accurately the rest of the patients. We are currently collecting recent data with a larger sample size to further validate and improve the current models. Last but not least, in the current ANN models, we only employed clinical factors that are readily available in clinical practice and imaging features that could be reliably identified with routine imaging exams, while subsequent relevant studies could accommodate more clinical and imaging factors that

might influence the risk of recurrence in TIA and minor stroke patients. Automatic image analysis and image feature extraction by methods such as convolutional neural networks would help in establishing more intelligent models for risk stratification of such patients.

CONCLUSION

Under the modern stroke service system, timely attention and management for TIA and minor stroke patients are becoming more readily available, which has significantly reduced the risk of stroke relapse in these patients. However, certain subgroups of patients are still at a high risk of subsequent disabling stroke, who may not be accurately identified with conventional risk predicting scores. Therefore, a more accurate and intelligent risk prediction strategy is needed. The ANN approach has advantages over conventional statistical methods or risk prediction scores that it could account for relationships between the independent variables, reflect complex relationships between continuous and categorical independent variables and the outcome, and quantify the weights of independent variables regarding their impact upon the outcome. The current pilot study indicated that ANN may yield a novel and effective method in risk stratification of TIA or minor stroke patients. Further studies are warranted for verification and improvement of such ANN models.

AUTHOR CONTRIBUTIONS

XL and JL made substantial contributions to the conception and design of the study. JY, YS, KW, and TL made substantial contributions to the acquisition of data. WZ, KC, WD, and QQ made substantial contributions to the analysis of data. KC, XL, WZ, and WD contributed to the interpretations of data. KC drafted the first version of the manuscript and XL made valuable revisions. All the authors revised the draft for intellectual content, gave their final approval of the final version for publication, and agreed to be accountable for all aspects of the work.

FUNDING

The Young Elite Scientist Sponsorship Program 2017–2019 (Reference No. 2017QNRC001), the China Association

for Science and Technology; the National Natural Science Foundation of China / Research Grants Council Joint Research Scheme (Reference No. 81661168015); Young Scientists Fund (Reference No. 81601000), National Natural Science Foundation of China.

REFERENCES

- Jiang B, Sun H, Ru X, Sun D, Chen Z, Liu H, et al. Prevalence, incidence, prognosis, early stroke risk, and stroke-related prognostic factors of definite or probable transient ischemic attacks in China, 2013. *Front Neurol.* (2017) 8:309. doi: 10.3389/fneur.2017.00309
- Coull AJ, Lovett JK, Rothwell PM, Oxford Vascular S. Population based study of early risk of stroke after transient ischaemic attack or minor stroke: implications for public education and organisation of services. *BMJ.* (2004) 328:326. doi: 10.1136/bmj.37991.635266.44
- Amarenco P, Lavallee PC, Labreuche J, Albers GW, Bornstein NM, Canhao P, et al. One-year risk of stroke after transient ischemic attack or minor stroke. *N Engl J Med.* (2016) 374:1533–42. doi: 10.1056/NEJMoa1412981
- Wang Y, Wang Y, Zhao X, Liu L, Wang D, Wang C, et al. Clopidogrel with aspirin in acute minor stroke or transient ischemic attack. *N Engl J Med.* (2013) 369:11–9. doi: 10.1056/NEJMoa1215340
- Johnston SC, Rothwell PM, Nguyen-Huynh MN, Giles MF, Elkins JS, Bernstein AL, et al. Validation and refinement of scores to predict very early stroke risk after transient ischaemic attack. *Lancet.* (2007) 369:283–92. doi: 10.1016/S0140-6736(07)60150-0
- Lemmens R, Smet S, Thijs VN. Clinical scores for predicting recurrence after transient ischemic attack or stroke: how good are they? *Stroke.* (2013) 44:1198–203. doi: 10.1161/STROKEAHA.111.000141
- Wardlaw JM, Brazzelli M, Chappell FM, Miranda H, Shuler K, Sandercock PA, et al. ABCD2 score and secondary stroke prevention Meta-analysis and effect per 1,000 patients triaged. *Neurology.* (2015) 85:373–80. doi: 10.1212/WNL.0000000000001780
- Amarenco P, Labreuche J, Lavallee PC. Patients with transient ischemic attack with ABCD2 <4 can have similar 90-day stroke risk as patients with transient ischemic attack with ABCD2 ≥4. *Stroke.* (2012) 43:863–5. doi: 10.1161/STROKEAHA.111.636506
- Merwick Á, Albers GW, Amarenco P, Arsava EM, Ay H, Calvet D, et al. Addition of brain and carotid imaging to the ABCD2 score to identify patients at early risk of stroke after transient ischaemic attack: a multicentre observational study. *Lancet Neurol.* (2010) 9:1060–9. doi: 10.1016/S1474-4422(10)70240-4
- Kiyohara T, Kamouchi M, Kumai Y, Ninomiya T, Hata J, Yoshimura S, et al. ABCD3 and ABCD3-I scores are superior to ABCD2 score in the prediction of short- and long-term risks of stroke after transient ischemic attack. *Stroke.* (2014) 45:418–25. doi: 10.1161/STROKEAHA.113.003077
- Mayer L, Ferrari J, Krebs S, Boehme C, Toell T, Matosevic B, et al. ABCD3-I score and the risk of early or 3-month stroke recurrence in tissue- and time-based definitions of TIA and minor stroke. *J Neurol.* (2018) 265:530–4. doi: 10.1007/s00415-017-8720-8
- Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol.* (2017) 2:230–43. doi: 10.1136/svn-2017-000101
- Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol.* (1996) 49:1225–31. doi: 10.1016/S0895-4356(96)00002-9
- Abedi V, Goyal N, Tsvigoulis G, Hosseinichimeh N, Hontecillas R, Bassaganya-Riera J, et al. Novel screening tool for stroke using artificial neural network. *Stroke.* (2017) 48:1678–81. doi: 10.1161/STROKEAHA.117.017033
- Baxt WG. Use of an artificial neural network for the diagnosis of myocardial infarction. *Ann Intern Med.* (1991) 115:843–8. doi: 10.7326/0003-4819-115-11-843
- Celik G, Baykan OK, Kara Y, Tireli H. Predicting 10-day mortality in patients with strokes using neural networks and multivariate statistical methods. *J Stroke Cerebrovasc Dis.* (2014) 23:1506–12. doi: 10.1016/j.jstrokecerebrovasdis.2013.12.018
- Edwards DF, Hollingsworth H, Zazulia AR, Diringer MN. Artificial neural networks improve the prediction of mortality in intracerebral hemorrhage. *Neurology.* (1999) 53:351–7. doi: 10.1212/WNL.53.2.351
- Shi HY, Hwang SL, Lee KT, Lin CL. In-hospital mortality after traumatic brain injury surgery: a nationwide population-based comparison of mortality predictors used in artificial neural network and logistic regression models. *J Neurosurg.* (2013) 118:746–52. doi: 10.3171/2013.1.JNS121130
- Langarizadeh M, Moghbeli F. Applying naive Bayesian networks to disease prediction: a systematic review. *Acta Inform Med.* (2016) 24:364–9. doi: 10.5455/aim.2016.24.364-369
- Orru G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci Biobehav Rev.* (2012) 36:1140–52. doi: 10.1016/j.neubiorev.2012.01.004
- Sayin R, Keskin S, Hamamci M. Evaluation of several classification methods in carpal tunnel syndrome. *J Pak Med Assoc.* (2017) 67:1654–7.
- Yang J, Fu JH, Chen XY, Chen YK, Leung TW, Mok V, et al. Validation of the ABCD2 score to identify the patients with high risk of late stroke after a transient ischemic attack or minor ischemic stroke. *Stroke.* (2010) 41:1298–300. doi: 10.1161/STROKEAHA.110.578757
- Moneta GL, Edwards JM, Chitwood RW, Taylor LM, Lee RW, Cummings CA, et al. Correlation of North American Symptomatic Carotid Endarterectomy Trial (NASCET) angiographic definition of 70% to 99% internal carotid artery stenosis with duplex scanning. *J Vasc Surg.* (1993) 17:152–9. doi: 10.1016/0741-5214(93)90019-1
- Lynn M, Chimowitz M, Howlett-smith H, Frankel M, Hertzberg V, Chaturvedi S, et al. Warfarin vs aspirin for symptomatic intracranial stenosis: subgroup analyses from WASID. *Stroke.* (2005) 36:457. doi: 10.1212/01.wnl.0000238506.76873.2f
- Bentley P, Ganesalingam J, Jones ALC, Mahady K, Epton S, Rinne P, et al. Prediction of stroke thrombolysis outcome using CT brain machine learning. *Neuroimage Clin.* (2014) 4:635–40. doi: 10.1016/j.nicl.2014.02.003
- Zhou Z, Folkert M, Cannon N, Iyengar P, Westover K, Zhang Y, et al. Predicting distant failure in early stage NSCLC treated with SBRT using clinical parameters. *Radiother Oncol.* (2016) 119:501–4. doi: 10.1016/j.radonc.2016.04.029
- Hoshino T, Uchiyama S, Wong LKS, Sissani L, Albers GW, Bornstein NM, et al. Differences in characteristics and outcomes between Asian and non-Asian patients in the TIAregistry.org. *Stroke.* (2017) 48:1779–87. doi: 10.1161/STROKEAHA.117.016874
- Basheer IA, Hajmeer M. Artificial neural networks: fundamentals, computing, design, and application. *J Microbiol Methods.* (2000) 43:3–31. doi: 10.1016/S0167-7012(00)00201-3

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2019.00171/full#supplementary-material>

Conflict of Interest Statement: 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.

Copyright © 2019 Chan, Leng, Zhang, Dong, Qiu, Yang, Soo, Wong, Leung and Liu. 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.



STIR-Net: Deep Spatial-Temporal Image Restoration Net for Radiation Reduction in CT Perfusion

Yao Xiao¹, Peng Liu¹, Yun Liang¹, Skylar Stolte¹, Pina Sanelli^{2,3,4,5}, Ajay Gupta², Jana Ivanidze² and Ruogu Fang^{1*}

¹ J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL, United States, ² Department of Radiology, Weill Cornell Medical College, New York, NY, United States, ³ Imaging Clinical Effectiveness and Outcomes Research, Department of Radiology, Northwell Health, Manhasset, NY, United States, ⁴ Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY, United States, ⁵ Center for Health Innovations and Outcomes Research, Feinstein Institute for Medical Research, Manhasset, NY, United States

OPEN ACCESS

Edited by:

Jean-Claude Baron,
University of Cambridge,
United Kingdom

Reviewed by:

Kyung Hyun Sung,
UCLA Health System, United States
William Farran Speier,
University of California, Los Angeles,
United States
Fabien Scalzo,
University of California, Los Angeles,
United States

*Correspondence:

Ruogu Fang
ruogu.fang@bme.ufl.edu

Specialty section:

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

Received: 04 June 2018

Accepted: 03 June 2019

Published: 26 June 2019

Citation:

Xiao Y, Liu P, Liang Y, Stolte S,
Sanelli P, Gupta A, Ivanidze J and
Fang R (2019) STIR-Net: Deep
Spatial-Temporal Image Restoration
Net for Radiation Reduction in CT
Perfusion. *Front. Neurol.* 10:647.
doi: 10.3389/fneur.2019.00647

Computed Tomography Perfusion (CTP) imaging is a cost-effective and fast approach to provide diagnostic images for acute stroke treatment. Its cine scanning mode allows the visualization of anatomic brain structures and blood flow; however, it requires contrast agent injection and continuous CT scanning over an extended time. In fact, the accumulative radiation dose to patients will increase health risks such as skin irritation, hair loss, cataract formation, and even cancer. Solutions for reducing radiation exposure include reducing the tube current and/or shortening the X-ray radiation exposure time. However, images scanned at lower tube currents are usually accompanied by higher levels of noise and artifacts. On the other hand, shorter X-ray radiation exposure time with longer scanning intervals will lead to image information that is insufficient to capture the blood flow dynamics between frames. Thus, it is critical for us to seek a solution that can preserve the image quality when the tube current and the temporal frequency are both low. We propose STIR-Net in this paper, an end-to-end spatial-temporal convolutional neural network structure, which exploits multi-directional automatic feature extraction and image reconstruction schema to recover high-quality CT slices effectively. With the inputs of low-dose and low-resolution patches at different cross-sections of the spatio-temporal data, STIR-Net blends the features from both spatial and temporal domains to reconstruct high-quality CT volumes. In this study, we finalize extensive experiments to appraise the image restoration performance at different levels of tube current and spatial and temporal resolution scales. The results demonstrate the capability of our STIR-Net to restore high-quality scans at as low as 11% of absorbed radiation dose of the current imaging protocol, yielding an average of 10% improvement for perfusion maps compared to the patch-based log likelihood method.

Keywords: CT perfusion image, radiation reduction, image restoration, deep learning, brain hemodynamics

1. INTRODUCTION

Acute stroke has high mortality and severe long-term disability rates worldwide. In the United States, more than 795,000 people have a stroke annually, and about 140,000 of them lose their lives, accounting for 5% of all deaths (1). Someone develops a stroke approximately every 40 s, and nearly every 4 min, someone loses he or her life because of stroke. Stroke can occur at any age, and it increases in likelihood with age. In 2009, two-thirds of people who had been hospitalized for stroke were older than 65 years old (2). The estimated cost related to stroke in the United States is about 34 billion dollars each year (3).

Acute stroke is an emergency, and successful patient outcomes require accurate diagnosis and prompt treatment. It is critical for someone to receive treatments for stroke within three hours from when he or she presents initial symptoms, as the disability rate measured three months after the stroke is generally high in those who did not receive timely treatments (4). There are two types of stroke: hemorrhagic and ischemic stroke. Hemorrhagic stroke occurs when a fragile blood vessel ruptures, while ischemic stroke is caused by thrombosis or embolism. Due to different etiologies and therapies, it is essential for patients to get timely diagnoses and treatments.

Computed Tomography (CT) scanning is a widely used imaging modality for rapid and detailed evaluation of the brain and cerebral vasculature; it is particularly valuable in the triage of acute stroke patients. CT can provide a rapid diagnosis of ischemic or hemorrhage stroke. It is clinically meaningful as rapid diagnosis enables clinicians to initiate optimized treatment for each of these two major categories of stroke. Patients with ischemic stroke often benefit from further characterization of brain tissue hemodynamics, and as such, often go through CT Perfusion (CTP) for further diagnosis and to guide treatment planning such as thrombolytic therapy. As CTP imaging can promptly offer an active view of cerebrovascular physiology, doctors can acquire CTP to evaluate cerebral blood flow status.

Obtaining a comprehensive visualization of blood flow dynamics and a clear brain anatomic structure requires contrast dose injection and repeated CT scanning. Under the acute stroke protocol, X-ray radiation from a 40-s CTP scan is comparable to a year's worth of radiation exposure from natural surroundings (5, 6). The CTP/CT Angiography (CTA) data acquisition process on a whole brain has a mean dose level of 6.8 mSv (7), which is two times more than that from natural background radiation sources; in comparison, the annual radiation exposure from the natural background is around 2.4 mSv (8). Moreover, repetitively scanning brain regions leads to accumulative radiation exposure to patients that may increase health risks such as skin irritation/erythema, hair loss/epilation (9), cataract formation (10), and even the induction of cancer (11, 12). In the US, about 80 million CT scans are performed annually. Therefore, seeking solutions to reduce the radiation dose that is associated with CT scans draws many researchers' attention.

Many researchers have attempted to seek practical solutions for radiation dose reduction in CT imaging. Solutions for reducing radiation exposure include two primary directions: optimizing CT systems and reducing contrast dose. Typical

optimization of CT systems comprises shortening temporal sampling frequency and reducing radiation sources such as the tube current/voltage and the number of beams and receptors. However, a simple reduction by the methods above will increase image noise and artifacts. In order to reduce CTP radiation exposure and maintain high diagnostic image quality, we integrate a deep learning approach with CT imaging to carry out this study.

In this paper, we propose an end-to-end Spatial-Temporal Image Restoration Net (STIR-Net) for CTP image restoration. This structure consists of two main components: Super-Resolution Denoising Nets (SRDNs) and a multi-directional conjunction layer which addresses image super-resolution (SR) and denoising in both spatial and temporal cross-sections. The contributions of this work are five-fold:

- 1) SRDN's patch representation layer extracts features from both the spatial and temporal dimensions of the CTP volume as cross-sections, which allows our model to present spatial-temporal details at the same time.
- 2) SRDN has the ability to perform image SR and denoising individually and simultaneously. It also can handle multi-level noise and multi-scale resolution and sampling.
- 3) We integrate multiple SRDNs based on different cross-sections into a multi-directional network, which can boost the performance further than individual cross-sections.
- 4) The results of the experiments demonstrate the effectiveness of STIR in the recovery of low radiation dose CTP images. STIR-Net can provide practical solutions for radiation dose reduction from three aspects (low tube current, decreased temporal sampling rate, and poor spatial resolution) with comparable image quality to the standard dose protocol.
- 5) We also provide the comparisons of Cerebral Blood Flow (CBF) and Cerebral Blood Volume (CBV), these maps attest that our proposed method can provide comparable results to the existing methods.

It is important to point out that no work has addressed low tube current, decreased temporal sampling rate, and poor spatial resolution simultaneously with a single deep learning structure. Through extensive experiments, our results demonstrate that STIR-Net has the capability of image restoration from these three types of data limitations simultaneously. Compared to low-dose scans using conventional methods, our network yields an average of 21% improvement of peak signal-to-noise ratio (PSNR) at around 21% to 42% low tube currents for the CTP sequences and an average of 10% improvement for the calculated perfusion maps. Hence, STIR-Net is a promising method for reducing radiation exposure in CTP imaging.

2. RELATED WORK

It is necessary to develop low-dose CTP protocols to reduce the risks associated with excessive X-ray radiation exposure. Different acquisition parameters such as tube current, temporal sampling frequency, and the spatial resolution are meticulously related to the quality of the reconstructed CTP images,

especially for generating perfusion maps that will be directly used by doctors to make treatment decisions. Related work includes radiation dose reduction approaches with respect to image processing strategies, deep learning approaches, image SR methods, and denoising methods. The previous work of our spatio-temporal architecture is introduced at the end of this section.

2.1. Radiation Dose Reduction Approaches

Radiation dose reduction approaches include reducing tube current, temporal sampling frequency, and beam number. There is a linear relationship between radiation dose and the tube current. For example, lowering 50% of the tube current will lead to a 50% reduction in radiation dose. However, image noise and the square root of tube current have an inverse proportional relationship. Simply reducing the tube current will deteriorate the CTP image quality with increased noise and artifacts. Current simulation studies demonstrate the possibility and the effectiveness of maintaining image quality at reduced tube current (13, 14). Reducing temporal sampling frequency is the same as the increment of time intervals between acquiring two CTP slices in the same CT study. Similar to the decrement of the tube current, the reduction in temporal sampling frequency will reduce radiation correspondingly, as the total amount of scanning period is fixed and the time interval has been increased. However, current research (15–17) shows that the reductions in sampling interval yield little advantages when the time intervals are greater than 1 s.

2.2. Image-Based Radiation Dose Reduction Approaches

Acquiring CT scans at low-dose and long scanning intervals will result in noisy and low-resolution (LR) images, with insufficient hemodynamic information. It is important to obtain higher quality CT images from limited data. Therefore, we address this problem of CT radiation reduction as image-based dose reduction. Recent work shows that an image-based dose reduction approach is a promising way for CT radiation reduction. For example in Yu et al. (18), a study of pediatric abdomen, pelvis, and chest CT examinations demonstrate that a 50% dose reduction can still maintain diagnostic quality. The image-based approaches include iterative reconstruction algorithm, sparse representation and dictionary learning, and example-based restoration methods. We review the relevant work as follows.

The iterative reconstruction (IR) algorithm is a promising approach for dose reduction. It produces a set of synthesized projections by meticulously modeling the data acquisition process in CT imaging. For example, adaptive statistical iterative reconstruction (ASIR) algorithm (19) was the first IR algorithm to be used in the clinic. By modeling the noise distribution of the acquired data, ASIR can provide clinically acceptable image quality at reduced doses. Many CT systems apply ASIR as an assuring radiation dose reduction approach because it can reduce image noise and provide dose-reduced clinical images with preserved diagnostic value (20). Another IR algorithm is called model-based iterative reconstruction, which is more

complicated and accurate than ASIR, as it models photons and system optics jointly.

Sparse representation and dictionary learning describe data as linear combinations of several fundamental elements from a predefined collection called a dictionary. In the computer vision and medical image analysis domains, sparse representation and dictionary learning have shown promising results in various image restoration applications. Such applications include sparsity-based simultaneous denoising and interpolation (21) for optical coherence tomography images reconstruction, dictionary learning with group sparsity and graph regularization (22) for medical image denoising and fusion, and (23) for magnetic resonance image reconstruction.

The example-based restoration approach is another popular method for image restoration. It extracts and stores patch pairs from both low-quality images and high-quality images in a database as prior knowledge. At the restoring phase, it learns a model that can synthesize high-quality images by searching the best-matched paired patches. Applications in image restoration (24–26) show the promising performance by using prior knowledge.

2.3. Deep Learning

In recent years, deep learning methods have emerged in various computer vision tasks, including image classification (27) and object detection (28), and have dramatically improved the performance of these systems. These approaches have also achieved significant improvement in image restoration (29, 30), super-resolution (31), and optical flow (32). The reason for the significant performance is due to the advanced modeling capabilities of the deep structure and the corresponding non-linearity combined with discriminative learning on large datasets.

Convolutional Neural Network (CNN), as one of the most renowned deep learning architectures, shows promising results for image-based problems. CNN structures are usually composed of several convolutional layers with activation layers, followed by one or more fully connected layers. The CNN architecture design utilizes image structures via local connections, weights sharing, and non-linearity. Another benefit of CNN is that they are easier to train and have fewer parameters than fully connected networks with the same number of hidden units. CNN structures allow automatic feature extraction and learning from limited information to reconstruct high-quality images.

2.4. Image Super-Resolution

Image super-resolution aims at restoring HR images from the observed LR images. SR methods use different portions of LR images, or separate images, to approximate the HR image. There are two types of SR algorithms: frequency domain-based and spatial domain-based. Initially, SR methods were mostly for problems in the frequency domain (33, 34). Algorithms addressed in the frequency domain using a simple theoretical basis for observing the relationships between HR and LR images. Though these algorithms show high computational efficiency, they are limited due to sensitivity to model errors and difficulty in managing complex motion models. Algorithms for the spatial domain then became the main trend by

overcoming the drawbacks of the frequency domain algorithms (35). Predominate spatial domain methods include non-uniform interpolation (36), iterative back-projection (37), projection onto convex sets (38), regularized methods (39), and a number of hybrid algorithms (40).

Deep learning is a popular approach for image SR problems, and it has achieved significant performance (31, 41–43). However, most SR frameworks focus on 2D images, as involving the temporal dimension is more challenging, especially in CTP imaging. In this work, we propose to overcome the difficulties involving spatial dimension and to prove the feasibility of our framework in cerebral CTP image restoration.

2.5. Image Denoising

Image denoising tasks aim at recovering a clean image from an observed noisy image, whereas the observed image is intruded by additive Gaussian noise. One of the main challenges for image denoising is to accurately identify the noise and remove it from the observed image. Based on the image properties being used, existing methods can be classified as prior-based (44), sparse coding based (25), low-rank-based (45), filter-based (46), and deep learning based (47, 48). The filter-based approach (46) methods are classical and fundamental, and many subsequent studies are developed from it (49).

Numerous works have reconstructed clean CT images that can preserve the image quality of perfusion maps successfully; these works include methods such as bilateral filtering, non-local mean (50), nonlinear diffusion filter (51), and wavelet-based methods (52). The oscillatory nature of the truncated singular value decomposition (TSVD)-based method has initiated research that incorporates different regularization methods to stabilize the deconvolution. This research has shown varying degrees of success in stabilizing the residue functions by enforcing both temporal and spatial regularization on the residue function (53, 54). However, prior studies have focused exclusively on regularizing the noisy low-dose CTP, without considering the corpus of high-dose CTP data and the multi-dimensional data properties of CT images.

Recently, deep learning based methods (47, 48) have shown many advantages in learning the mapping of the observed low-quality images to the high-quality ones. These methods use CNN models that are trained on tens of thousands of samples; however, paired training data is usually scarce in the medical field. Hence, an effective learning based model is desired. In this work, we utilize data extracted from different cross-sections of the CTP volume to achieve better performance in image SR and denoising. The experiment result shows that the proposed network can handle various noise and image degradation levels.

2.6. Spatial-Temporal Architecture

In our previous work, we proposed Spatio-Temporal Architecture for Super-Resolution (STAR) (55) for low-dose CTP image super-resolution. It is an end-to-end spatio-temporal architecture that preserves image quality at reduced scanning time and radiation that has been reduced to one-third of its original level. This is an image-based dose reduction approach that focuses on super-resolution only. STAR is inspired by the

work in Kim et al. (31) and is extended to three-dimensional volumes by conjoining multiple cross-sections. Through this work, we found that features extracted from both spatial and temporal directions are helpful to improve SR performance. The integration of multiple single-directional networks (SDNs) can boost the performance of SR for the spatio-temporal CTP data. The experimental results show that the proposed basic model of SDN improves both spatial and temporal resolution, while the multi-directional conjoint network further enhances the SR results—comparing favorably with only temporal or only spatial SR. However, this work only addresses low spatial and temporal resolution; it misses the important noise issue in low dose CTP.

In this paper, we propose STIR-Net, an end-to-end spatio-temporal image restoration net for CTP radiation reduction. We compose and integrate several SRDNs instead of SDNs at different cross-sections for both image super-resolution and denoising simultaneously. The STIR-Net structure is explained in section 3. In section 4, we provide the experiment platform setup and describe the data acquisition method and the preprocessing procedures. In section 5, we detail the experiments and results. Finally, section 6 concludes the paper.

3. METHODOLOGY

In this section, we first introduce the patch representation schema for generating 2D spatio-temporal input patches for STIR-Net. Then, we describe how to synthesize the multi-directional spatio-temporal image restoration network by joint super-resolution and denoising at various cross-sections.

3.1. Patch Representation

Three types of patches serve as inputs in this work, consisting of the following: patches for image SR tasks, for denoising tasks, and for conjoint SR and denoising tasks. All the 2D LR patches are generated from the 3D CTP volumes. We use $X \times Y \times T$ to indicate the three dimensions of the volume, where X and Y are spatial dimensions and T is the temporal dimension. We extract 2D patches along the $X \times Y$ direction as well as along one of the spatial directions with temporal T dimension: $X \times T$ and $Y \times T$. We create 2D LR patches by down-sampling the cross-sectional images in the spatial direction, temporal direction, or both spatial and temporal directions. For instance, using $X \times T$ and $Y \times T$ cross-sections, we remove every other pixel along the T direction to simulate scanning intervals which are two times longer. This corresponds to two times less X-ray radiation exposure in the resulting images. For the denoising task, we simulate the low tube current images by adding spectrum Gaussian noise on the entire CTP volume, with more details in section 4.3. The 2D patches for denoising are generated based on the noisy volumes along the $X \times T$, $Y \times T$, and $X \times Y$ cross-sections. For joint SR and denoising tasks, we apply the same scaling strategies that we use to create LR patches, but we apply them on top of noisy volumes. After feeding these LR and/or noisy patches with their labels (the patches extracted from the standard dose) into convolution

layers for learning the spatio-temporal details, HR and/or denoised outputs will be generated in the testing stage based on the captured features.

3.2. STIR-Net: Spatial-Temporal Image Restoration Net

Our proposed STIR-Net is a CNN-based end-to-end spatial-temporal architecture for image restoration. To begin, we describe the fundamental SRDN structure—super-resolution denoising networks for cross-section images. Then, we explain in detail the composition of STIR-Net.

3.2.1. SRDN: Super-Resolution Denoising Structure

The usage of kernel combination strategy in GoogLeNet (56) shows that a creative structuring of layers can lead to improved performance and computationally efficiency. Inception modules place various sizes of kernels in parallel. This can extract fine-grain details in volume, while the broader kernel can cover a large receptive field of the input. Extracting diverse information can help with the prediction in classification tasks; however, image denoising poses different challenges.

SRDN is an end-to-end structure that learns from pair-wise LR/noisy patches with their original clean images and outputs high-quality CT images based on low-quality input images while testing. The structure of SRDN is shown in **Figure 1**. The main functional part of SRDN is built by stacking four modularized Kernel Regulation Blocks (KR-Block). KR-Blocks are inspired by GoogLeNet (56), which has a combination of kernels of varying sizes. Specifically, each block comprises of two 1×1 convolutional layers, one 7×7 convolutional layer, and one 3×3 convolutional layer for regulating the features extracted by the 7×7 convolutional layer. The combination of large and small filters is to balance extraction of subtle and edge features. Moreover, each block is embedded with a skip-connection, which allows reference to the feature mapping from previous layers and boosts the network performance.

- **Serial connections.** Image classification needs to summarize diverse information to a linear classifier. On the contrary, image denoising needs to find the most prominent features for a progressive transformation. Therefore, we adopt three kernel sizes (e.g., 1×1 , 3×3 , and 7×7) in the KR-block module. Kernels of each size are placed in series to allow the small kernels to regulate the features extracted by the large one.
- **Small behind large.** Large kernels (e.g., 7×7) can extract certain features by observing a local region with more statistical pixel information. The small kernels (e.g., 3×3) are primarily used for exploiting deeper prior information from the underlying feature-maps obtained by large preceding kernels. The subtle textures are especially highlighted during this regularization procedure. Large kernels excel in noise removal but may also smooth the whole image irrespective of its edges or details. Small kernels can preserve subtle textures, but noise pixels may detract from the information attained. Therefore, placing a small kernel behind the large one is a straightforward strategy to enhance the denoiser regularization.

- **Feature blending.** The features extracted by large kernels contain both actual pixel values and noise, whereas the small kernel can capture real pixels while simultaneously ignoring much of the noise. At the end of a KR-block, features captured by small kernels are blended with the features extracted from large kernels. To allow the locally highlighted features to be shared across neighboring KR-blocks, feature-blending is processed by pixel-wise summation (see **Figure 1**-top) rather than concatenation (e.g., in GoogLeNet). This helps with finding the most prominent features for a forward transformation. Eventually, the output of a KR-block contains more accurate pixel information with less noise.
- **1×1 convolution.** The special usage of 1×1 convolution in KR-block is for two purposes: first, it reduces the dimensions inside KR-block modules, such as the first 1×1 convolution layer; second, it adds more non-linearity by having PReLU immediately after every 1×1 convolution and suffers less from over-fitting due to smaller kernel size.

3.2.2. SRDN Architecture

Convolutional networks learn a mapping function between a corrupted image input and a corresponding noise-free image. The network contains L convolution layers (Conv), each of which implements a feature extraction procedure. To ensure our network has rich feature representations, we use a considerable amount of large filters in the first two convolutional layers (57) to extract diverse and representative features for feature mapping and spatial transformation. We define densely convolutional features extracted from the l th layer as

$$x_l = \text{Conv}(y_l, f_l, n_l, c_l)_{f \geq 7 \times 7, n \geq 128} \quad (1)$$

where $l = 1 \dots L$ indexes the layer, y_l , f_l , n_l , and c_l represent the l 's input, the filter size, filter number, and channel number, respectively. x_l are the feature maps extracted from y_l by $\text{Conv}(\cdot)$, which denotes convolution. As the top and bottom layers have different functional attentions (57), the network can be decomposed into three parts (the bottom part is shown in **Figure 1**): feature extraction, feature regulation and mapping, and image reconstruction. In the proposed SRDN, the first two layers have the same volume: $(f_l, n_l, c_l) = (7, 128, 1)$.

Several KR-blocks are cascaded to perform feature regulation, mapping, and transformation. Also, residual learning is performed here by skip-connection, which connects the outputs of two adjacent KR-blocks. The use of skip connection between KR-blocks leads to faster and more stable training. The purpose of using a shortcut between the input and the end of the network is to incorporate more information from the original input into image reconstruction. This strategy helps relax the network interference difficulty because input data contains much real pixel information that can be taken as a prior. To make SRDN more compact, we introduce two 1×1 composite units, referred as “Shrinking” and “Expanding,” shown in **Figure 1**. After densely convolutional feature-extraction layers, we reduce the number of feature maps by “Shrinking.” After feature regulation and mapping, we expand feature maps such that there are sufficient various features that can be provided for image

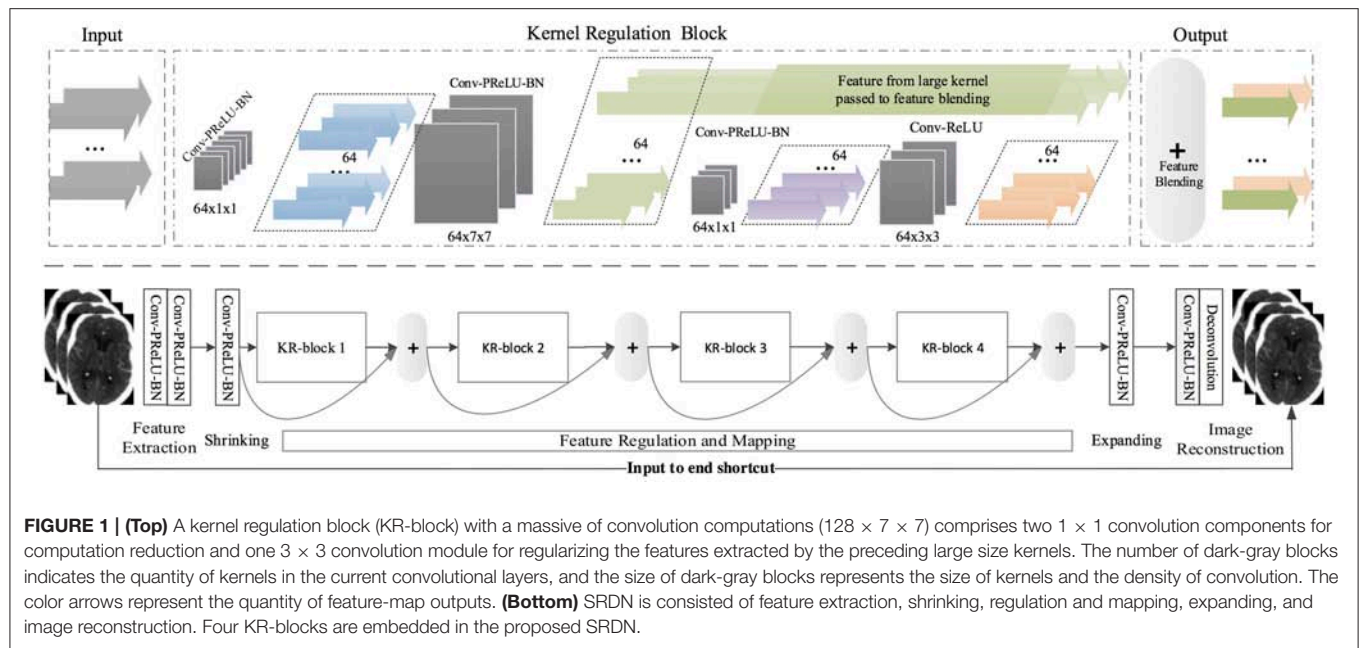


FIGURE 1 | (Top) A kernel regulation block (KR-block) with a massive of convolution computations ($128 \times 7 \times 7$) comprises two 1×1 convolution components for computation reduction and one 3×3 convolution module for regularizing the features extracted by the preceding large size kernels. The number of dark-gray blocks indicates the quantity of kernels in the current convolutional layers, and the size of dark-gray blocks represents the size of kernels and the density of convolution. The color arrows represent the quantity of feature-map outputs. **(Bottom)** SRDN is consisted of feature extraction, shrinking, regulation and mapping, expanding, and image reconstruction. Four KR-blocks are embedded in the proposed SRDN.

reconstruction. The convolutional layer before the last layer has the volume: $(f_l, n_l, c_l) = (3, 128, 1)$. We utilize a deconvolutional layer with the volume: $(f_l, n_l, c_l) = (3, 1, 1)$ as our last layer.

3.2.3. STIR-Net Structure

The combination of the various features extracted from multi-directional data enhances the network's capability for inference and generality. Since multi-directional inputs provide different perspectives of the 3D volume data, they cannot merely be regarded as feeding more training data into multi-networks. Instead, they complement each other nicely to encode the sparse features through the network.

Dense convolutions and kernel regulation strategy ensure diverse features from multi-directional brain CT images, which can be encoded as network representations. In this paper, we adopt three SRDNs to cope with three directional extracted data respectively: $Y \times T$, $X \times T$, and $X \times Y$ to form our STIR-Net. The structure of STIR-Net is shown in **Figure 2**. During training, the input and output layers are matched with pair-wise noisy and label patches. The label here refers to the patches extracted from the original high radiation dose CTP volume ($X \times Y \times T$). Each SRDN contains 4 KR-blocks that can fully encode the features from each directional data without overfitting. For the testing stage, the outputs of the three SRDN nets assemble into a conjoint learning layer. This layer blends various features from all SRDN nets together to be one spatio-temporal volume by calculating the mean of the three outputs.

4. PLATFORM AND DATA ACQUISITION

4.1. Computational Platform

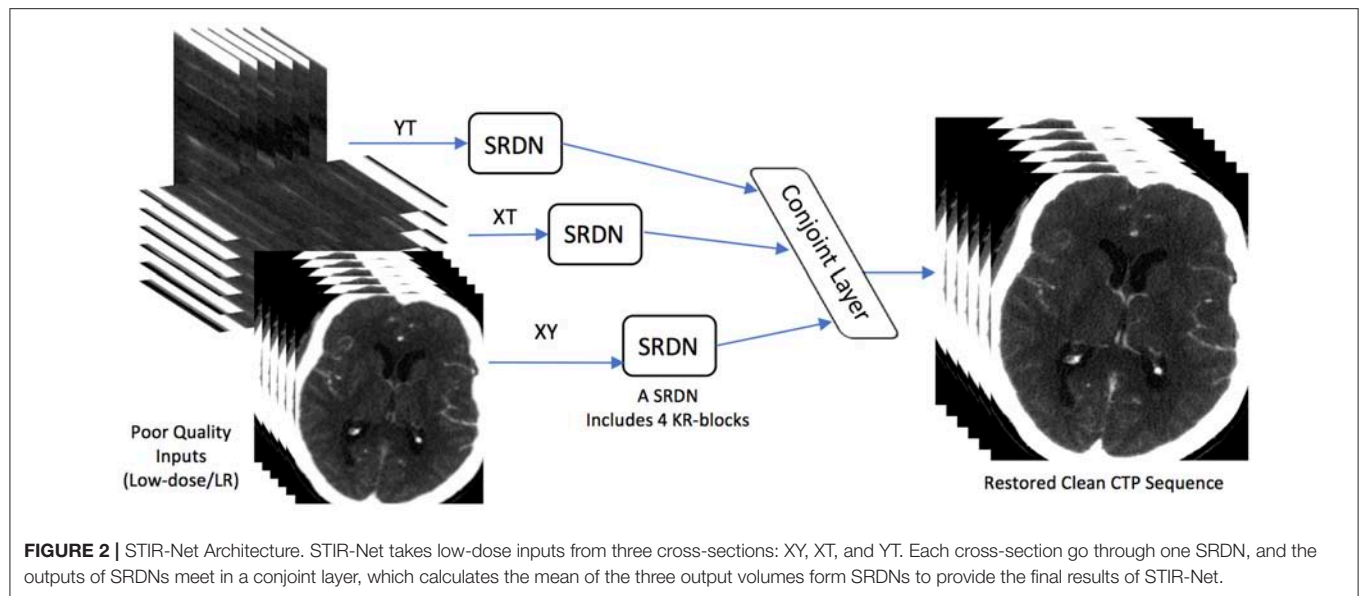
We use the deep learning framework Caffe (58) for constructing the proposed STIR-Net. All experiments are conducted by a GPU workstation that contains four NVIDIA PASCAL xp GPUs. For

data preprocessing and post analysis, we use MATLAB (Version R2016b) as it is an efficient programming language for matrix-based image processing.

4.2. Datasets

We evaluate the proposed method on 23 stroke patients' CTP sequences. All CTP sequences are scanned using the same acute stroke protocol for patients from August 2007 to June 2010 using GE Lightspeed or Pro-16 scanners (General Electric Medical Systems, Milwaukee, WI). The scanners are in cine 4i scanning mode and perform 45 s acquisitions at one rotation per second using 80 kVp and 190 mAs. Approximately 45 mL of non-ionic iodinated contrast was administered intravenously at 5 mL/s using a power injector with a 5 s delay. The thickness of the brain region at the z-axis is 20 mm for each sequence, and each sequence has four slices along the z-axis where each slice is 5 mm thick (cross-plane resolution). The brain region has 0.43 spatial resolution (in-plane resolution) on the xy-plane. The slices within one CTP sequence are intensity normalized and co-registered over time. The entire volume size of one patient is $512 \times 512 \times 4 \times 119$, where 512 is the height and width of each CT slice, 4 is the number of slices on the z-axis, and 119 is the number of frames in the CTP sequence. In this paper, we only select one slice along the z-axis, thus the size of resulting the CTP volume is $512 \times 512 \times 119$, denoted as $X \times Y \times T$.

We randomly split the patients into three groups: 12 patients for training, four patients for validation, and seven patients for testing. As each patient has 119 slices, the training, validation, and testing set resulted in 1,428, 476, and 833 images in XY cross-section (the spatial direction), respectively. We only maintain brain regions in the images for the other two cross-sections, XT and YT, or about 300 pixels for the X and the Y directions. Therefore for these cross-sections, we estimate that we have 3,600 images for training, 1,200 for validation, and 2,100 for testing.



We use the patch-based method in this paper, so the images are further cropped into patches of size 41×41 with a stride of 21. This resulted in 822,528 and 274,176 patches in XY cross-section, 75,600 patches in XT cross-section, and 25,200 patches in YT cross-section, respectively for training and validation.

4.3. Low Radiation Dose Simulation and Data Preprocessing

To simulate low radiation dose CTP images, we address three generation approaches: reducing the tube current, shortening X-ray radiation exposure time, and lowering spatial resolution. We detail each criterion as below.

- **Low Tube Current.** We followed the same steps described in Britten et al. (59) to simulate the low-dose CT images by adding spatially correlated statistical noise (spectrum Gaussian noise). The generated noise is directly added on the original high-dose images, where the high-dose volumes are scanned at tube current $I_0 = 190\text{mAs}$. Based on Britten et al. (59), the noise model is built on the inverse relationship between the tube current I and the noise standard deviation σ in CT images. The noise level σ (the standard deviation of Gaussian noise that we want to add to the original images) is adjusted based on tube current I that we want to simulate according to equation

$$\sigma = K \times \sqrt{\frac{1}{I} - \frac{1}{I_0}} \quad (2)$$

where $K = 103.09\text{mA}^{\frac{1}{2}}$ is computed based on phantom studies. We simulate four levels of noisy images in this paper at different tube currents: 20, 40, 60, and 80 mAs.

- **Low Temporal Sampling Rate.** To reduce the temporal sampling rate for shorter X-ray radiation exposure time, we simulate longer scanning intervals by removing frames between specific time intervals. For example, we remove every

other frame from the CTP volume to generate the down-sampled volume that is two times shorter on the temporal dimension than the original length. In this way, we skip frames with two scales S_i : two times shorter S_2 and three times S_3 shorter than the original time. We also keep the original length S_1 for comparison. For all down-sampled volumes, we scale them back to the original size via bicubic interpolation for deep learning experiments.

- **Low Spatial Sampling Rate.** We lower the CT spatial sampling rate to mimic the low spatial resolution images that are produced by a limited amount of beams and receptors. For instance, we create the down-sampled images by skipping every other pixel (scaling rate of two) along the X and Y directions in the original high radiation dose images respectively (so-called grid-wise). We simulate the LR images by skipping pixels grid-wise with two scales S_i : two times down-sampled S_2 , and three times down-sampled S_3 . We set S_1 as no down-sampling for comparison. Then, we interpolate the down-sampled images by the bicubic method to scale them back to the original image size.

Based on different patch representations that are described in section 3.1, we preprocess the data subsequently. We have three combinations of directional cross-sections XY, XT, and YT for STIR-Net. For each individual denoising and super-resolution case, we add Gaussian noise to the high-dose images and apply spatio/temporal down-sampling, respectively. For the combination of super-resolution and de-noising, we add the noise first and then apply spatial/temporal down-sampling depending on different scaling factors.

5. EXPERIMENTS AND RESULTS

The experiments of this work are carried out in three steps: image super-resolution, image denoising, and image super-resolution

with denoising. In the first two steps, we want to show that the proposed STIR-Net is capable of different image restoration tasks independently. Further, in the third step, we want to demonstrate that our STIR-Net can tackle super-resolution and denoising simultaneously. We train the STIR-Net structure from scratch using low-quality images from different cross-sections, then we test each of the cross-sections as spatial-only, temporal-only, and spatial and temporal combined. The performance is computed based on the average result from seven patients' 119 slices. As cross-sections (XT and YT) are trained and tested in a 2D circumstance that combined temporal dimension with one spatial dimension, we concatenate the resulted 2D images into 3D volumes and recalculate the performance based on XY direction.

5.1. Evaluation Metrics

The experiment performance is evaluated based on two evaluation metrics: structural similarity (SSIM) index and PSNR. SSIM is used for measuring the similarity between two images based on the computation of luminance term $l(x, y)$, the contrast term $c(x, y)$, and the structural term $s(x, y)$, where x and y are two images. We calculate SSIM based on the following equations

$$SSIM(x, y) = [l(x, y) \cdot c(x, y) \cdot s(x, y)] \quad (3)$$

$$l(x, y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}, c(x, y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}, \\ s(x, y) = \frac{2\sigma_{xy} + c_3}{\sigma_x^2 + \sigma_y^2 + c_3} \quad (4)$$

where $\mu_x, \mu_y, \sigma_x, \sigma_y, \sigma_{xy}$ are the local means, standard deviations, and cross-covariance for images x and y . The value of c_1, c_2 , and c_3 are set as 6.5025, 58.5225, and 29.26125, where the values are calculated based on the dynamic range L of the pixel-values (here is 255) in function $c_1 = (0.01 \cdot L)^2$, $c_2 = (0.03 \cdot L)^2$, and $c_3 = c_2/2$. PSNR defines the ratio between the maximum intensity value in the ground truth image I_{max} and the power of corrupting noise σ (root mean square error between the ground truth and enhanced image) that affects representation fidelity.

$$PSNR = 20 \log_{10} \frac{I_{max}}{\sigma} \quad (5)$$

5.2. Image Super-Resolution

The first experiment is image super-resolution, which is independently conducted on three cross-sections ($Y \times T, X \times T$, and $X \times Y$) at two sampling rates (S_2 : down-sampling to 1/2, S_3 : down-sampling to 1/3). We want to evaluate whether the proposed STIR-Net is capable of achieving a stable performance in different cross-sections at different levels of scaling. For the XY cross-section, we down-sample along the spatial directions to create low-resolution images. For the XT and YT cross-sections, we down-sample on the temporal direction only to simulate scanning in a shorter X-ray radiation exposure time. The experimental results of STIR-Net are shown in **Table 1**. We calculate SSIM and PSNR values for LR inputs, SR outputs, and the improvements of SR from LR. The greatest improvements

for both SSIM and PSNR are in the XY direction, while the XT and YT directions have achieved similar improvements. When the sampling rate is high, the improvements compared to the lower sampling rate are higher in almost all cross-sections. The improvements of SSIM and PSNR are highly stable and follow the same trend in different conditions. A one-tailed paired t -test was conducted to compare the performance improvements of PSNR and SSIM values. There was a significant difference in the scores for PSNR (Mean = 37.623, SD = 10.955) and SSIM (Mean = 0.950, SD = 0.001) before and after using the proposed method; where $p = 0.0003$ for PSNR and $p = 0.0004$ for SSIM show that the improvements are significant as $p < 0.05$. These results suggest that PSNR and SSIM do improve significantly after applying our model in this experiment. This experiment indicates that STIR-Net has the potential to address low spatial and temporal resolution in CTP image volumes.

5.3. Image Denoising

In this experiment, we explore different levels of low tube current for training STIR-Net. We added the spectrum Gaussian noise to simulate four low tube currents: 20, 40, 60, and 80 mAs, which are 11, 21, 32, and 42% of the original 190 mAs tube current. We train the proposed STIR-Net by mixing together the different tube currents - it is more difficult to restore high-dose images at lower tube current, as shown in **Table 2**. This table shows that the SSIM and PSNR performances for the XY direction when STIR-Net is trained and tested with mixed levels of tube currents, which are at a fixed spatial/temporal sampling rate of S_2 . The improvement of SSIM increases as tube currents decrease, while the improvement of PSNR remains in a similar range. We show that STIR-Net is a general solution for different tube currents, as the PSNR improvements for different test cases are all higher than 5 dB. In this experiment, we demonstrate that STIR-Net can tackle denoising problems as well, even for mixed noise levels. The improvements are very stable for different tube current levels.

5.4. Spatial-Temporal Super-Resolution and Denoising

In addition to the encouraging individual experiment results for image super-resolution and denoising, the experiment results in both spatial and temporal super-resolution with denoising have also achieved great enhancements. We evaluate the resulted images based on two aspects in this section: the analysis on the resulted CTP sequence and the analysis on the generated perfusion maps.

5.4.1. CTP Sequence Analysis

Table 3 shows the PSNR comparison of the resulted CTP sequence among Multi-Scale Expected Patch Log Likelihood (MS-EPLL) (60) method, our previously proposed method STAR, and the current method STIR-Net. The test results are displayed as an average value over seven test patients' 833 slices output. The STAR and STIR-Net methods both contain three scenarios: spatial SR only, temporal SR only, and joint spatial and temporal SR. In both methods, the temporal SR includes two cross-sections (the XT and YT directions).

TABLE 1 | Average SSIM and PSNR (dB) performance of seven patients' 833 CTP slices between different sampling scales for STIR-Net image super-resolution at different spatio-temporal cross-sections.

Direction	Scale	SSIM			PSNR		
		LR	STIR-Net	Improvement	LR	STIR-Net	Improvement
XY	S ₂	0.954	0.987	0.033	32.348	43.195	10.846
	S ₃	0.855	0.929	0.074	26.958	35.317	8.359
YT	S ₂	0.939	0.971	0.032	33.354	38.961	5.607
	S ₃	0.887	0.929	0.042	29.240	35.167	5.926
XT	S ₂	0.929	0.967	0.038	32.965	38.580	5.615
	S ₃	0.866	0.915	0.048	28.831	34.513	5.681

TABLE 2 | Average SSIM and PSNR (dB) performance of seven patients' 833 slices for XY direction when STIR-Net is trained and tested with mix levels of tube currents, where at a fixed spatial/temporal sampling scale S₂.

mAs	SSIM			PSNR		
	LR	STIR-Net	Improvement	LR	STIR-Net	Improvement
20	0.778	0.859	0.080	25.227	30.445	5.217
40	0.830	0.896	0.065	26.663	32.349	5.686
60	0.860	0.909	0.049	27.323	33.062	5.739
80	0.879	0.915	0.036	27.702	33.409	5.706

We show that STIR-Net is a general solution for different tube currents as the PSNR improvements for different test cases are all higher than 5 dB. mAs is the unit for tube current-time product.

TABLE 3 | Average PSNR comparison of seven patients' 833 CTP slices for different conditions.

mAs	Scale	LR	MS-EPLL	STAR			STIR-Net		
			Spatial	Spatial	Temporal	Conjoint	Spatial	Temporal	Conjoint
20	S ₁	23.438	28.566	31.734	32.550	33.273	31.856	32.456	33.189
	S ₂	23.032	26.022	28.705	31.006	31.017	28.847	30.774	30.919
	S ₃	21.993	22.517	26.636	29.990	29.682	26.861	30.405	30.096
40	S ₁	26.764	31.176	33.763	34.511	35.452	33.848	34.685	35.402
	S ₂	24.698	27.357	29.899	31.841	32.214	30.092	32.471	32.659
	S ₃	23.246	23.576	27.289	31.020	30.609	27.555	31.272	30.909
60	S ₁	29.079	33.030	35.336	36.145	37.254	35.415	36.064	37.195
	S ₂	25.498	28.161	30.705	33.269	33.489	30.956	33.544	33.913
	S ₃	23.801	24.689	27.694	31.673	31.207	27.785	31.867	31.450
80	S ₁	31.025	34.607	36.769	37.664	38.967	36.854	36.724	38.469
	S ₂	25.971	28.608	31.505	34.123	34.464	31.719	34.324	34.728
	S ₃	24.116	24.658	28.027	32.268	31.706	27.688	32.438	31.687
Avg		25.222	27.748*	30.672*	33.005*	33.278*	30.790*	33.086*	33.385*

In this table, three methods are compared: MS-EPLL, STAR, and STIR-Net. The conditions include four types of tube current (20, 40, 60, and 80 mAs) and three kinds of SR scales (S₁: no down-sampling, S₂: down-sampling to 1/2, and S₃: down-sampling to 1/3). LR means the PSNR value for the noise image after down-sampling. S₁ is image denoising only. The best values are highlighted for different scenarios. The average value is listed at the bottom of the table. The asterisk symbol denotes the result of the current method achieves significant higher PSNR value than the LR images at $\alpha = 0.05$ when performing the one-tailed paired t-tests, and the star symbol denotes the comparison between MS-EPLL method.

Table 3 focuses on the comparison of four levels of tube current (20, 40, 60, 80 mAs) and three SR scales (S₁: no down-sampling, S₂: down-sampling to 1/2, S₃: down-sampling to 1/3). The down-sample rates are applied based on different methods: spatial-only models are scaled down on the spatial dimensions, temporal-only models are scaled down on the temporal dimension, and the conjoint models are scaled down on both spatial and temporal dimensions (depending on different cross-sections). In this table, LR refers to the PSNR value for the noise image after down-sampling. We highlighted the best values for different scenarios. From this table, we can see STAR achieves higher PSNR for denoising than STIR-Net, while STIR-Net performs better for mixed noise and down-sampling scenarios. Moreover, both STAR and STIR-Net methods outperform the

MS-EPLL method. For all tube currents, the PSNR value follows the trend of better image restoration results at higher tube currents. Similarly, a lower down-sampling rate leads to better reconstruction performance. The conjoint of spatial and temporal directions of STAR gives the best results for all four tube current levels. When the low dose CT images have poor spatial or temporal resolutions, it is usually more difficult to tackle both denoising and SR problems; however, our STIR-Net net is more favorable for these situations. Its conjoint model gives us an average 32% improvement from the LR inputs. The experiment results indicate that most mixed low dose and low-resolution scenarios can achieve the best performances, especially for the temporal directions. This means that for the temporal directions, there is more related information that can be used for reconstructing CT frames that are nearby the down-sampled slices. The average performance improvement for STIR-Net net is about 8.08 dB from the LR inputs and around 4 dB compared to the MS-EPLL method. We perform one-tailed paired *t*-tests in **Table 3** to compare PSNR values at different mAs and super-resolution scales using $\alpha = 0.05$. All three types of STIR-Net perform significantly better than LR and MS-EPLL, especially the conjoint model achieves the best performance among all methods.

5.4.2. Perfusion Maps Analysis

We compare the perfusion maps (CBF and CBV) based on which physicians make the clinical decision, as the perfusion maps can show the hemodynamic changes of blood flow. Therefore, achieving higher accuracy in restoration in perfusion maps is critical for clinical diagnosis.

Visual Comparison: The visual comparisons of the generated perfusion maps (CBF and CBV) are presented in **Figures 3–6** for patient # 18, # 19, and # 21 in the case of scale level S_2 and S_3 with 40 mAs. We enlarge the region of interest for each image to check the details, and we highlight the details by using white arrows. From these figures, the edges in the LR images are distorted compared to the original images, and MS-EPLL restores the detail information incorrectly. The resulting images of the STIR-Net models are much closer to the ground truth images compare to MS-EPLL and STAR. The boundaries and details of the features in STIR-Net results are well-preserved, and the figures are less blurry than other methods. In sum, the proposed STIR-Net gives us much accurate perfusion maps compare to MS-EPLL and STAR methods as it restores the edge information much closer to the ground truth images.

Quantitative Comparison: We calculate the CBF and CBV values based on the CTP sequences resulted from different methods. Then, we use PSNR and SSIM as evaluation metrics. As the proposed method STIR-Net is designed for CTP image super-resolution and denoising simultaneously, we show the results of 40 mAs at the down-sample scale of S_2 and S_3 . **Tables 4, 5** provide the PSNR and SSIM comparisons of CBF and CBV maps in the case of scale level S_2 with 40 mAs and **Tables 6, 7** are for scale level S_3 . In general, STIR-Net models achieve the best performance, and the temporal model is usually the top performer.

We perform one-tailed paired *t*-tests for each table to compare PSNR and SSIM of the restored images with LR images

and restored images using MS-EPLL and STAR models. The hypothesis for all *t*-tests is: after using the proposed method, we can achieve significant improvements in PSNR and SSIM values from the images of LR, MS-EPLL method, or STAR models. The results show that our proposed STIR-Net models not only significantly improve the PSNR and SSIM values from the LR images but also achieves significantly higher PSNR and SSIM values than the MS-EPLL method, especially for the temporal models and the conjoint models. For comparison with STAR model, **Table 4** shows that at S_2 and 40 mAs, CBF's SSIM values using the STIR-Net temporal model is significantly ($p = 0.002067$) better than the STAR temporal model, similar for CBV ($p = 0.01554$). STIR-Net's conjoint model is also significantly better than the STAR conjoint model ($p = 0.00994$) in terms of SSIM. In **Table 7**, for the case of S_3 and 40 mAs, similar observations are made. STIR-Net temporal model is significantly ($p = 0.03521$) better than the STAR temporal model and conjoint model in terms of both PSNR and SSIM.

Overall, the test results demonstrate the advantage of our STIR-Net to restore high-quality scans at as low as 11% of absorbed radiation dose of the current imaging protocol, yielding an average of 17% improvement in PSNR and SSIM values for perfusion maps including CBF and CBV compared to LR images and 10% improvements compared to MS-EPLL method. For the comparison of STIR-Net and STAR, we calculate the improvements by averaging out all three models including the spatial model, temporal model, and the conjoint model. Our proposed STIR-Net method achieves an average of 0.2% improvements in PSNR and SSIM values for perfusion maps than STAR models.

6. CONCLUSION

This paper presents a novel deep learning-based multi-directional spatio-temporal framework to recover the low radiation dose CTP images of acute stroke patients by addressing both denoising and super-resolution problems simultaneously. Our proposed framework, called STIR-Net, is an end-to-end image restoration network that is capable of recovering images scanned at low tube current, short X-ray radiation exposure time, and low spatial resolution jointly. We emphasize the characteristic of our proposed STIR-Net in CTP image super-resolution and denoising jointly, which directs prior and data fidelity terms with two insights: First, a well-trained CNN-based denoiser can be regarded as a sequence of filter-based denoisers. Second, each component of a CNN-based denoiser has the capacity of jointly dealing with image denoising and super-resolution problems. By combining the cross-sectional features in the spatio-temporal domain, our STIR-Net achieves to better reconstruction results, especially for mixed low-resolution and noise cases. After inputting low dose and low-resolution patches at different cross-sections of the spatio-temporal data simultaneously, STIR-Net blends the features from both spatial and temporal domains to reconstruct high-quality CT volumes. The experimental results indicate that

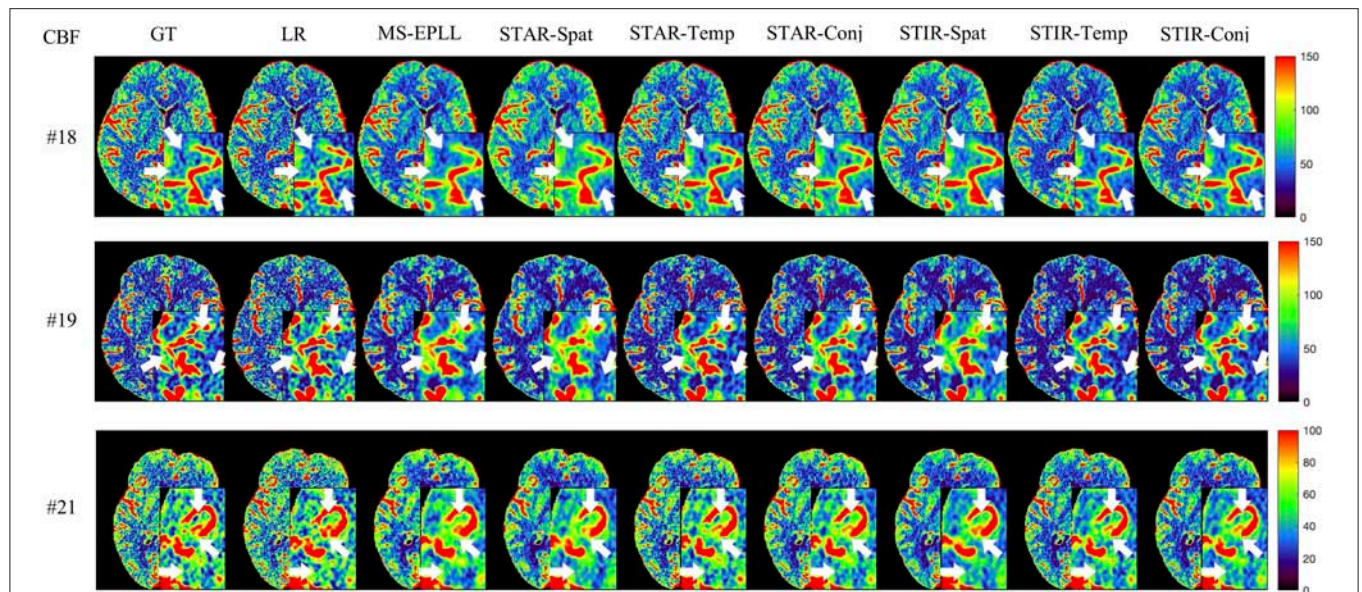


FIGURE 3 | Visual comparison of CBF for three test patients: #18, #19, and #21, when reducing the tube current to 40 mAs with a down-sample ratio of two (two times low spatial and two time low temporal resolutions). The notation for each column is: GT, Ground truth image; LR, Low-Resolution input; MS-EPLL, MS-EPLL restoration result; STAR-Spat, STAR reconstruction result (spatial only); STAR-Temp, STAR reconstruction result (temporal only); STAR-Conj, STAR reconstruction result (spatial + temporal); STIR-Spat, STIR-Net reconstruction result (spatial only); STIR-Temp, STIR-Net reconstruction result (temporal only); STIR-Conj, STIR-Net reconstruction result (spatial + temporal). All figures are displayed by using the same colormap and the color range for each patient is shown in the colorbar on the rightmost of each row. We use white arrows to compare the details in the region of interests.

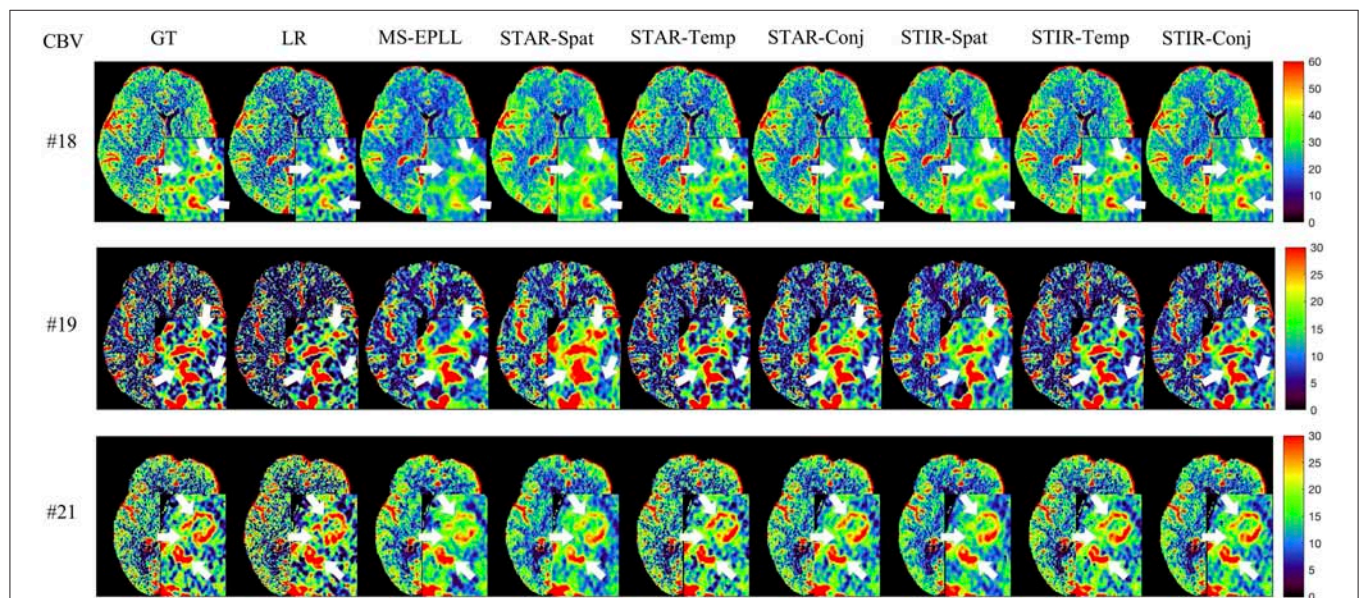


FIGURE 4 | Visual comparison of CBV for three test patients: #18, #19, and #21, when reducing the tube current to 40 mAs with a down-sample ratio of two (two times low spatial and two time low temporal resolutions). The notation for each column is: GT, Ground truth image; LR, Low-Resolution input; MS-EPLL, MS-EPLL restoration result; STAR-Spat, STAR reconstruction result (spatial only); STAR-Temp, STAR reconstruction result (temporal only); STAR-Conj, STAR reconstruction result (spatial + temporal); STIR-Spat, STIR-Net reconstruction result (spatial only); STIR-Temp, STIR-Net reconstruction result (temporal only); STIR-Conj, STIR-Net reconstruction result (spatial + temporal). All figures are displayed by using the same colormap and the color range for each patient is shown in the colorbar on the rightmost of each row. We use white arrows to compare the details in the region of interests.

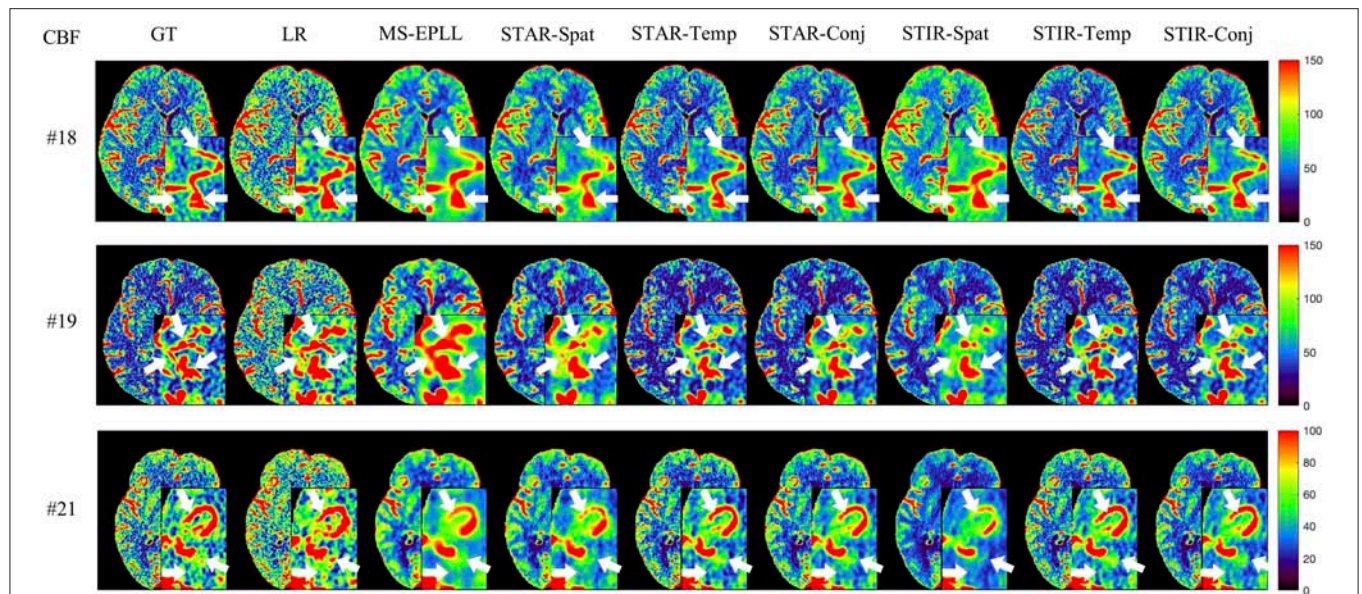


FIGURE 5 | Visual comparison of CBF for three test patients: #18, #19, and #21, when reducing the tube current to 40 mAs with a down-sample ratio of three (three times low spatial and two time low temporal resolutions). The notation for each column is: GT, Ground truth image; LR, Low-Resolution input; MS-EPLL, MS-EPLL restoration result; STAR-Spat, STAR reconstruction result (spatial only); STAR-Temp, STAR reconstruction result (temporal only); STAR-Conj, STAR reconstruction result (spatial + temporal); STIR-Spat, STIR-Net reconstruction result (spatial only); STIR-Temp, STIR-Net reconstruction result (temporal only); STIR-Conj, STIR-Net reconstruction result (spatial + temporal). All figures are displayed by using the same colormap and the color range for each patient is shown in the colorbar on the rightmost of each row. We use white arrows to compare the details in the region of interests.

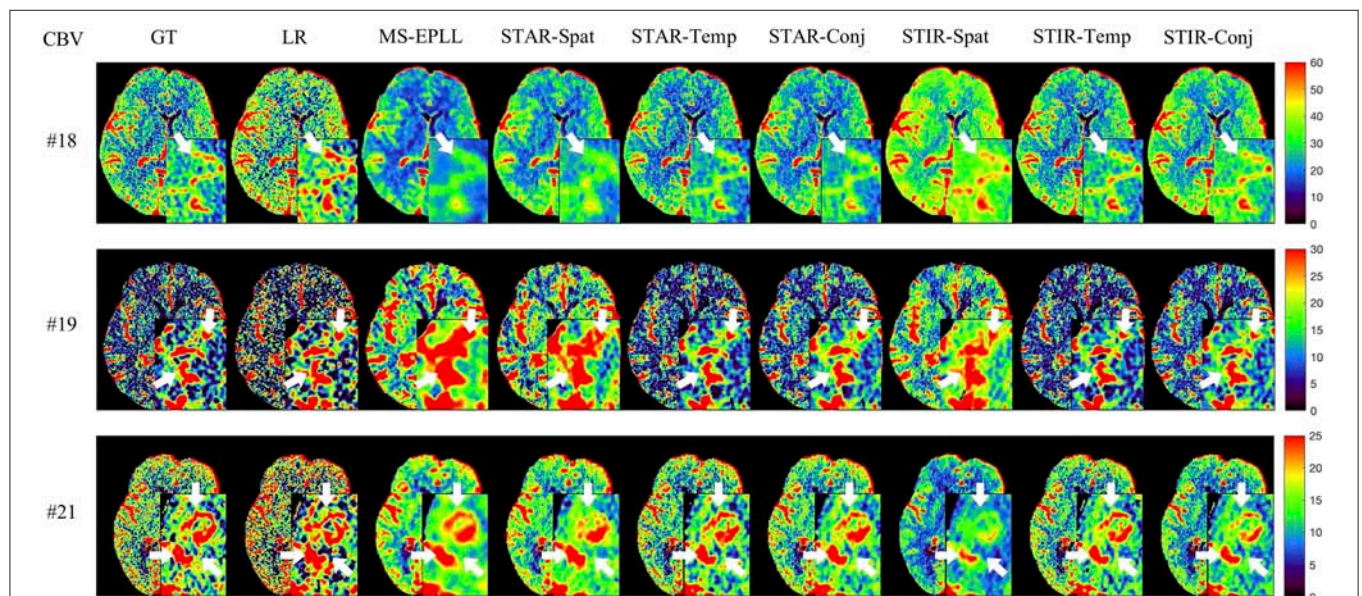


FIGURE 6 | Visual comparison of CBV for three test patients: #18, #19, and #21, when reducing the tube current to 40 mAs with a down-sample ratio of three (three times low spatial and two time low temporal resolutions). The notation for each column is: GT, Ground truth image; LR, Low-Resolution input; MS-EPLL, MS-EPLL restoration result; STAR-Spat, STAR reconstruction result (spatial only); STAR-Temp, STAR reconstruction result (temporal only); STAR-Conj, STAR reconstruction result (spatial + temporal); STIR-Spat, STIR-Net reconstruction result (spatial only); STIR-Temp, STIR-Net reconstruction result (temporal only); STIR-Conj, STIR-Net reconstruction result (spatial + temporal). All figures are displayed by using the same colormap and the color range for each patient is shown in the colorbar on the rightmost of each row. We use white arrows to compare the details in the region of interests.

TABLE 4 | PSNR and SSIM value comparison of seven patients' CBF maps calculated at scale S_2 with tube current 40 mAs.

CBF	Patient	LR	MS-EPLL	STAR			STIR		
			Spatial	Spatial	Temporal	Conjoint	Spatial	Temporal	Conjoint
PSNR	#18	24.57	25.74	27.83	30.54	31.30	28.06	29.07	29.79
	#19	23.87	25.18	26.31	28.91	28.81	26.25	27.83	27.84
	#20	23.07	26.09	29.30	30.94	31.63	24.16	32.02	32.31
	#21	25.34	26.64	28.02	31.70	31.59	27.86	31.26	30.70
	#22	23.25	26.13	27.90	30.46	30.86	28.26	30.24	31.14
	#23	23.84	25.16	27.37	27.26	28.99	27.87	26.24	27.11
	#24	18.23	21.59	23.55	26.62	26.61	24.68	25.89	26.30
	Avg	23.15	25.21*	27.18*	29.49*	29.97*	26.73*	28.94*	29.31*
	Var	5.28	2.80	3.34	3.77	3.59	2.95	5.74	5.09
SSIM	#18	0.84	0.85	0.88	0.93	0.93	0.88	0.94	0.93
	#19	0.77	0.80	0.82	0.89	0.89	0.81	0.90	0.90
	#20	0.75	0.84	0.86	0.91	0.92	0.78	0.94	0.93
	#21	0.78	0.82	0.83	0.90	0.91	0.83	0.92	0.91
	#22	0.78	0.83	0.86	0.92	0.92	0.86	0.93	0.93
	#23	0.82	0.84	0.86	0.89	0.92	0.87	0.89	0.89
	#24	0.72	0.77	0.80	0.88	0.89	0.82	0.90	0.89
	Avg	0.78	0.82*	0.84*	0.90*	0.91*	0.83*	0.92*	0.91*
	Var	0.0016	0.0008	0.0008	0.0003	0.0002	0.0014	0.0004	0.0004

In this table, three methods are compared: MS-EPLL, STAR, and STIR-Net. LR means the PSNR value for the noise image after down-sampling. The best values are highlighted for different patients. The average value and the variance are listed at the bottom of the table. The asterisk symbol denotes the result of the current method achieves significant higher PSNR value than the LR images at $\alpha = 0.05$ when performing the one-tailed paired t-tests, and the star symbol denotes the result is significantly higher than MS-EPLL method.

TABLE 5 | PSNR and SSIM value comparison of seven patients' CBV maps calculated at scale S_2 with tube current 40 mAs.

CBV	Patient	LR	MS-EPLL	STAR			STIR		
			Spatial	Spatial	Temporal	Conjoint	Spatial	Temporal	Conjoint
PSNR	#18	28.00	28.62	31.78	34.24	35.22	31.68	34.66	34.17
	#19	32.32	34.05	33.79	37.52	38.11	35.15	38.75	38.63
	#20	30.62	32.83	37.53	38.30	39.69	34.80	40.15	39.77
	#21	32.20	33.38	34.95	37.67	37.98	35.09	38.07	38.27
	#22	31.60	32.76	32.70	37.80	37.57	33.15	38.94	38.50
	#23	30.55	31.51	32.03	35.87	36.46	34.76	35.37	35.97
	#24	26.42	29.37	31.37	32.53	33.28	30.74	35.37	33.55
	Avg	30.22	31.76*	33.44*	36.26*	36.89*	33.60*	37.08*	36.97*
	Var	4.87	4.16	4.77	4.67	4.47	3.18	6.11	5.82
SSIM	#18	0.86	0.87	0.90	0.94	0.94	0.90	0.95	0.94
	#19	0.84	0.87	0.87	0.92	0.93	0.88	0.94	0.94
	#20	0.89	0.92	0.93	0.96	0.96	0.93	0.97	0.97
	#21	0.84	0.88	0.89	0.92	0.93	0.89	0.94	0.93
	#22	0.85	0.88	0.89	0.93	0.94	0.89	0.95	0.95
	#23	0.85	0.88	0.89	0.92	0.93	0.90	0.93	0.93
	#24	0.77	0.82	0.84	0.88	0.89	0.84	0.90	0.90
	Avg	0.84	0.88*	0.89*	0.93*	0.93*	0.89*	0.94*	0.94*
	Var	0.0015	0.0008	0.0008	0.0005	0.0004	0.0007	0.0004	0.0004

In this table, three methods are compared: MS-EPLL, STAR, and STIR-Net. LR means the PSNR value for the noise image after down-sampling. The best values are highlighted for different patients. The average value and the variance are listed at the bottom of the table. The asterisk symbol denotes the result of the current method achieves significant higher PSNR value than the LR images at $\alpha = 0.05$ when performing the one-tailed paired t-tests, and the star symbol denotes the result is significantly higher than MS-EPLL method.

TABLE 6 | PSNR and SSIM comparison of seven patients' CBF maps calculated at scale S_3 with tube current 40 mAs.

CBF	Patient	LR	MS-EPLL	STAR			STIR		
			Spatial	Spatial	Temporal	Conjoint	Spatial	Temporal	Conjoint
PSNR	#18	21.68	23.89	24.92	27.88	27.30	24.64	28.09	27.92
	#19	19.95	20.55	23.39	26.51	26.60	23.71	27.09	26.59
	#20	22.41	26.23	25.57	29.84	29.62	26.14	29.51	29.04
	#21	23.35	24.85	25.11	28.10	27.54	23.18	29.36	27.77
	#22	21.97	23.95	24.23	23.97	28.04	24.83	28.54	28.29
	#23	22.50	24.03	23.95	26.05	26.35	23.85	24.85	25.84
	#24	17.21	21.45	20.66	24.53	23.91	20.91	24.04	23.49
	Avg	21.29	23.56*	23.97*	26.70*	27.05*	23.89*	27.35*	26.99*
	Var	4.30	3.77	2.67	4.31	3.07	2.66	4.66	3.51
SSIM	#18	0.74	0.75	0.77	0.88	0.87	0.79	0.88	0.88
	#19	0.64	0.69	0.71	0.83	0.83	0.71	0.83	0.82
	#20	0.70	0.76	0.73	0.87	0.87	0.77	0.87	0.86
	#21	0.68	0.70	0.73	0.82	0.82	0.70	0.85	0.84
	#22	0.69	0.75	0.76	0.83	0.87	0.77	0.87	0.87
	#23	0.73	0.74	0.76	0.85	0.85	0.76	0.84	0.85
	#24	0.63	0.68	0.69	0.82	0.82	0.69	0.82	0.81
	Avg	0.69	0.72*	0.74*	0.85*	0.85*	0.74*	0.85*	0.85*
	Var	0.0016	0.0009	0.0008	0.0006	0.0007	0.0015	0.0005	0.0006

In this table, three methods are compared: MS-EPLL, STAR, and STIR-Net. LR means the PSNR value for the noise image after down-sampling. The best values are highlighted for different patients. The average value and the variance are listed at the bottom of the table. The asterisk symbol denotes the result of the current method achieves significant higher PSNR value than the LR images at $\alpha = 0.05$ when performing the one-tailed paired t-tests, and the star symbol denotes the result is significantly higher than MS-EPLL method.

TABLE 7 | PSNR and SSIM comparison of seven patients' CBV maps calculated at scale S_3 with tube current 40 mAs.

CBV	Patient	LR	MS-EPLL	STAR			STIR		
			Spatial	Spatial	Temporal	Conjoint	Spatial	Temporal	Conjoint
PSNR	#18	24.94	27.53	29.48	32.33	31.44	27.43	34.13	31.43
	#19	30.71	26.58	28.50	36.15	35.01	29.86	36.39	35.48
	#20	28.40	34.42	31.90	37.98	38.11	34.22	38.49	38.17
	#21	30.45	31.87	32.83	35.86	35.52	31.36	35.97	34.87
	#22	29.28	30.51	29.96	36.00	35.83	28.05	36.00	35.77
	#23	28.83	31.31	29.08	35.08	33.86	27.33	34.35	34.23
	#24	24.96	27.44	28.68	30.22	30.76	28.53	30.87	31.01
	Avg	28.21	29.93	30.03*	34.80*	34.35*	29.52	35.16*	34.42*
	Var	5.60	8.15	2.61	6.93	6.59	6.25	5.67	6.30
SSIM	#18	0.77	0.81	0.83	0.90	0.90	0.83	0.91	0.90
	#19	0.77	0.76	0.77	0.89	0.89	0.78	0.90	0.89
	#20	0.81	0.89	0.85	0.94	0.94	0.89	0.94	0.94
	#21	0.77	0.84	0.85	0.90	0.90	0.83	0.90	0.89
	#22	0.76	0.82	0.82	0.91	0.91	0.80	0.91	0.91
	#23	0.76	0.84	0.83	0.90	0.90	0.82	0.90	0.90
	#24	0.69	0.76	0.77	0.84	0.85	0.77	0.85	0.85
	Avg	0.76	0.82*	0.82*	0.90*	0.90*	0.82*	0.90*	0.90*
	Var	0.0013	0.0022	0.0012	0.0009	0.0008	0.0016	0.0007	0.0008

In this table, three methods are compared: MS-EPLL, STAR, and STIR-Net. LR means the PSNR value for the noise image after down-sampling. The best values are highlighted for different patients. The average value and the variance are listed at the bottom of the table. The asterisk symbol denotes the result of the current method achieves significant higher PSNR value than the LR images at $\alpha = 0.05$ when performing the one-tailed paired t-tests, and the star symbol denotes the result is significantly higher than MS-EPLL method.

our framework has the potential to maintain the diagnostic image quality not only for reducing the tube current down to 11% of the commercial standard but also for 1/3 X-ray radiation exposure time and 1/3 spatial resolution. Hence, our approach is an efficient and effective solution for radiation dose reduction in CTP imaging. In the future, we will extend the work into multimodal imaging radiation dose reduction by combining low-dose non-contrast CT, CTA, and CTP images holistically.

AUTHOR CONTRIBUTIONS

YX drafted the manuscript, designed the STIR-Net architecture and the experiments, carried out the experiments and analysis. PL designed the SRDN deep learning network structure, drafted

the manuscript SRDN section. YL assisted the generation of the perfusion maps and the related analysis. SS, PS, AG, and JI revised the manuscript critically for important intellectual content. RF designed and directed the project.

ACKNOWLEDGMENTS

This work is partially supported by the National Science Foundation under Grant No. IIS-1564892, the University of Florida Informatics Institute SEED Funds and the UF Clinical and Translational Science Institute, which is supported in part by the NIH National Center for Advancing Translational Sciences under award number UL1 TR001427. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

- Yang Q, Tong X, Schieb L, Vaughan A, Gillespie C, Wiltz JL, et al. Vital signs: recent trends in stroke death rates—United States, 2000–2015. *Morb Mortal Wkly Rep.* (2017) 66:933–9. doi: 10.15585/mmwr.mm6635e1
- Hall MJ, Levant S, DeFrances CJ. Hospitalization for stroke in US hospitals, 1989–2009. *Diabetes.* (2012) 18:23.
- Benjamin EJ, Blaha MJ, Chiuve SE, Cushman M, Das SR, Deo R, et al. Heart disease and stroke statistics-2017 update: a report from the American Heart Association. *Circulation.* (2017) 135:e146–603. doi: 10.1161/CIR.0000000000000485
- Centers for Disease Control and Prevention. Awareness of stroke warning symptoms—13 States and the District of Columbia, 2005. *Morb Mortal Wkly Rep.* (2008) 57:481. doi: 10.1001/jama.300.3.274
- Mettler FA Jr, Bhargavan M, Faulkner K, Gilley DB, Gray JE, Ibbott GS, et al. Radiologic and nuclear medicine studies in the United States and worldwide: frequency, radiation dose, and comparison with other radiation sources—1950–2007. *Radiology.* (2009) 253:520–31. doi: 10.1148/radiol.2532082010
- Takei Y, Miyazaki O, Matsubara K, Shimada Y, Muramatsu Y, Akahane K, et al. Nationwide survey of radiation exposure during pediatric computed tomography examinations and proposal of age-based diagnostic reference levels for Japan. *Pediatr Radiol.* (2016) 46:280–5. doi: 10.1007/s00247-015-3474-x
- Thierfelder KM, Sommer WH, Baumann AB, Klotz E, Meinel FG, Strobl FF, et al. Whole-brain CT perfusion: reliability and reproducibility of volumetric perfusion deficit assessment in patients with acute ischemic stroke. *Neuroradiology.* (2013) 55:827–35. doi: 10.1007/s00234-013-1179-0
- Cho G, Kim JH, Park TS, Cho K. Proposing a simple radiation scale for the public: radiation index. *Nucl Eng Technol.* (2017) 49:598–608. doi: 10.1016/j.net.2016.10.005
- Wintermark M, Lev M. FDA investigates the safety of brain perfusion CT. *Am J Neuroradiol.* (2010) 31:2–3. doi: 10.3174/ajnr.A1967
- Chodick G, Bekiroglu N, Hauptmann M, Alexander BH, Freedman DM, Doody MM, et al. Risk of cataract after exposure to low doses of ionizing radiation: a 20-year prospective cohort study among US radiologic technologists. *Am J Epidemiol.* (2008) 168:620–31. doi: 10.1093/aje/kwn171
- de González AB, Mahesh M, Kim KP, Bhargavan M, Lewis R, Mettler F, et al. Projected cancer risks from computed tomographic scans performed in the United States in 2007. *Arch Intern Med.* (2009) 169:2071–7. doi: 10.1001/archinternmed.2009.440
- Journey NM, Lee C, Harbron RW, McHugh K, Pearce MS, de González AB. Projected cancer risks potentially related to past, current, and future practices in paediatric CT in the United Kingdom, 1990–2020. *Br J Cancer.* (2017) 116:109. doi: 10.1038/bjc.2016.351
- Juluru K, Shih J, Raj A, Comunale J, Delaney H, Greenberg E, et al. Effects of increased image noise on image quality and quantitative interpretation in brain CT perfusion. *Am J Neuroradiol.* (2013) 34:1506–12. doi: 10.3174/ajnr.A3448
- Murphy A, So A, Lee TY, Symons S, Jakubovic R, Zhang L, et al. Low dose CT perfusion in acute ischemic stroke. *Neuroradiology.* (2014) 56:1055–62. doi: 10.1007/s00234-014-1434-z
- Ng CS, Chandler AG, Wei W, Anderson EF, Herron DH, Kurzrock R, et al. Effect of sampling frequency on perfusion values in perfusion CT of lung tumors. *Am J Roentgenol.* (2013) 200:W155–62. doi: 10.2214/AJR.12.8664
- Othman AE, Afat S, Brockmann MA, Nikoubashman O, Brockmann C, Nikolaou K, et al. Radiation dose reduction in perfusion CT imaging of the brain: a review of the literature. *J Neuroradiol.* (2016) 43:1–5. doi: 10.1016/j.neurad.2015.06.003
- Abels B, Klotz E, Tomandl B, Villablanca J, Kloska S, Lell M. CT perfusion in acute ischemic stroke: a comparison of 2-second and 1-second temporal resolution. *Am J Neuroradiol.* (2011) 32:1632–9. doi: 10.3174/ajnr.A2576
- Yu L, Fletcher JG, Shiung M, Thomas KB, Matsumoto JM, Zingula SN, et al. Radiation dose reduction in pediatric body CT using iterative reconstruction and a novel image-based denoising method. *Am J Roentgenol.* (2015) 205:1026–37. doi: 10.2214/AJR.14.14185
- Yu Z, Thibault JB, Bouman CA, Sauer KD, Hsieh J. Fast model-based X-ray CT reconstruction using spatially nonhomogeneous ICD optimization. *IEEE Trans Image Process.* (2011) 20:161–75. doi: 10.1109/TIP.2010.2058811
- Singh S, Kalra MK, Do S, Thibault JB, Pien H, Connor OO, et al. Comparison of hybrid and pure iterative reconstruction techniques with conventional filtered back projection: dose reduction potential in the abdomen. *J Comput Assist Tomogr.* (2012) 36:347–53. doi: 10.1097/RCT.0b013e31824e639e
- Fang L, Li S, McNabb RP, Nie Q, Kuo AN, Toth CA, et al. Fast acquisition and reconstruction of optical coherence tomography images via sparse representation. *IEEE Trans Med Imaging.* (2013) 32:2034–49. doi: 10.1109/TMI.2013.2271904
- Li S, Yin H, Fang L. Group-sparse representation with dictionary learning for medical image denoising and fusion. *IEEE Trans Biomed Eng.* (2012) 59:3450–9. doi: 10.1109/TBME.2012.2217493
- Huang J, Zhang S, Metaxas D. Efficient MR image reconstruction for compressed MR imaging. *Med Image Anal.* (2011) 15:670–9. doi: 10.1016/j.media.2011.06.001
- Hamdan S, Fukumizu Y, Izumi T, Yamauchi H. Example-based face image super-resolution taking into consideration correspondence of facial parts. *IEEE Trans Electr Electron Eng.* (2017) 12:917–24. doi: 10.1002/tee.22483
- Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans Image Process.* (2006) 15:3736–45. doi: 10.1109/TIP.2006.881969
- Trinh DH, Luong M, Dibos F, Rocchisani JM, Pham CD, Nguyen TQ. Novel example-based method for super-resolution and denoising of medical images. *IEEE Trans Image Process.* (2014) 23:1882–95. doi: 10.1109/TIP.2014.2308422

27. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV (2016). p. 770–8. doi: 10.1109/CVPR.2016.90
28. Erhan D, Szegedy C, Toshev A, Anguelov D. Scalable object detection using deep neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH (2014). p. 2147–54. doi: 10.1109/CVPR.2014.276
29. Burger HC, Schuler CJ, Harmeling S. Image denoising: can plain neural networks compete with BM3D? In: *IEEE Conference on Computer Vision and Pattern Recognition*. Providence, RI: IEEE (2012). p. 2392–9. doi: 10.1109/CVPR.2012.6247952
30. Xie J, Xu L, Chen E. Image denoising and inpainting with deep neural networks. In: *Advances in Neural Information Processing Systems*. Lake Tahoe (2012). p. 341–9.
31. Kim J, Kwon Lee J, Mu Lee K. Accurate image super-resolution using very deep convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas (2016). p. 1646–54. doi: 10.1109/CVPR.2016.182
32. Bailer C, Taetz B, Stricker D. Flow fields: dense correspondence fields for highly accurate large displacement optical flow estimation. In: *Proceedings of the IEEE International Conference on Computer Vision*. Santiago (2015). p. 4015–23. doi: 10.1109/ICCV.2015.457
33. Kim S, Bose NK, Valenzuela H. Recursive reconstruction of high resolution image from noisy undersampled multiframes. *IEEE Trans Acoust Speech Signal Process.* (1990) 38:1013–27. doi: 10.1109/29.56062
34. Aizawa K, Komatsu T, Saito T. Acquisition of very high resolution images using stereo cameras. In: *Visual Communications and Image Processing: Visual Communication*. Vol. 1605. Boston, MA: International Society for Optics and Photonics (1991). p. 318–29.
35. Park SC, Park MK, Kang MG. Super-resolution image reconstruction: a technical overview. *IEEE Signal Process Mag.* (2003) 20:21–36. doi: 10.1109/MSP.2003.1203207
36. Nguyen N, Milanfar P, Golub G. A computationally efficient superresolution image reconstruction algorithm. *IEEE Trans Image Process.* (2001) 10:573–83. doi: 10.1109/83.913592
37. Irani M, Peleg S. Improving resolution by image registration. *Graphical Models Image Process.* (1991) 53:231–9. doi: 10.1016/1049-9652(91)90045-L
38. Stark H, Oskoui P. High-resolution image recovery from image-plane arrays, using convex projections. *JOSA A.* (1989) 6:1715–26. doi: 10.1364/JOSAA.6.001715
39. Ng MK, Shen H, Lam EY, Zhang L. A total variation regularization based super-resolution reconstruction algorithm for digital video. *EURASIP J Adv Signal Process.* (2007) 2007:074585. doi: 10.1155/2007/74585
40. Elad M, Feuer A. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Trans Image Process.* (1997) 6:1646–58. doi: 10.1109/83.650118
41. Dong C, Loy CC, He K, Tang X. Learning a deep convolutional network for image super-resolution. In: *European Conference on Computer Vision*. Zürich: Springer (2014). p. 184–99. doi: 10.1007/978-3-319-10593-2_13
42. Shi W, Caballero J, Huszar F, Totz J, Aitken AP, Bishop R, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas (2016). p. 1874–83. doi: 10.1109/CVPR.2016.207
43. Oktay O, Bai W, Lee M, et al. Multi-input cardiac image super-resolution using convolutional neural networks. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Athens: Springer (2016). p. 246–54. doi: 10.1007/978-3-319-46726-9_29
44. Chen F, Zhang L, Yu H. External patch prior guided internal clustering for image denoising. In: *Proceedings of the IEEE International Conference on Computer Vision*. Santiago (2015). p. 603–11. doi: 10.1109/ICCV.2015.76
45. Gu S, Zhang L, Zuo W, Feng X. Weighted nuclear norm minimization with application to image denoising. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH (2014). p. 2862–9. doi: 10.1109/CVPR.2014.366
46. Dabov K, Foi A, Katkovnik V, Egiazarian K. BM3D image denoising with shape-adaptive principal component analysis. In: *Signal Processing with Adaptive Sparse Structured Representations*. Saint Malo (2009).
47. Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a gaussian denoiser: residual learning of deep CNN for image denoising. *IEEE Trans Image Process.* (2017) 26:3142–55. doi: 10.1515/9783110524116
48. Mao XJ, Shen C, Yang YB. Image restoration using convolutional auto-encoders with symmetric skip connections. *arXiv [preprint]*. arXiv:160608921 (2016).
49. Baxes GA. *Digital Image Processing: Principles and Applications*. New York, NY: Wiley (1994).
50. Mendrik AM, Vonken EJ, van Ginneken B, de Jong HW, Riordan A, van Setters T, et al. TIPS bilateral noise reduction in 4D CT perfusion scans produces high-quality cerebral blood flow maps. *Phys Med Biol.* (2011) 56:3857. doi: 10.1088/0031-9155/56/13/008
51. Salinas HM, Fernández DC. Comparison of PDE-based nonlinear diffusion approaches for image enhancement and denoising in optical coherence tomography. *IEEE Trans Med Imaging.* (2007) 26:761–71. doi: 10.1109/TMI.2006.887375
52. Lin JW, Laine AF, Bergmann SR. Improving PET-based physiological quantification through methods of wavelet denoising. *IEEE Trans Biomed Eng.* (2001) 48:202–12. doi: 10.1109/10.909641
53. Fang R, Zhang S, Chen T, Sanelli PC. Robust low-dose CT perfusion deconvolution via tensor total-variation regularization. *IEEE Trans Med Imaging.* (2015) 34:1533–48. doi: 10.1109/TMI.2015.2405015
54. Fang R, Chen T, Sanelli PC. Towards robust deconvolution of low-dose perfusion CT: Sparse perfusion deconvolution using online dictionary learning. *Med Image Anal.* (2013) 17:417–28. doi: 10.1016/j.media.2013.02.005
55. Xiao Y, Gupta A, Sanelli PC, Fang R. STAR: spatio-temporal architecture for super-resolution in low-dose CT perfusion. In: *International Workshop on Machine Learning in Medical Imaging*. Springer (2017). p. 97–105. doi: 10.1007/978-3-319-67389-9_12
56. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston, MA (2015). p. 1–9. doi: 10.1109/CVPR.2015.7298594
57. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: *European Conference on Computer Vision*. Zürich: Springer (2014). p. 818–33. doi: 10.1007/978-3-319-10593-2_53
58. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, et al. Caffe: convolutional architecture for fast feature embedding. In: *Proceedings of the 22nd ACM International Conference on Multimedia*. Orlando, FL: ACM (2014). p. 675–8. doi: 10.1145/2647868.2654889
59. Britten A, Crotty M, Kiremidjian H, Grundy A, Adam E. The addition of computer simulated noise to investigate radiation dose and image quality in images with spatial correlation of statistical noise: an example application to X-ray CT of the brain. *Br J Radiol.* (2004) 77:323–8. doi: 10.1259/bjr/78576048
60. Papayan V, Elad M. Multi-scale patch-based image restoration. *IEEE Trans Image Process.* (2016) 25:249–61. doi: 10.1109/TIP.2015.2499698

Conflict of Interest Statement: 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.

Copyright © 2019 Xiao, Liu, Liang, Stolte, Sanelli, Gupta, Ivanidze and Fang. 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.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

[@frontiersin](https://twitter.com/frontiersin)



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership