**Title**

Leveraging Clinical Imaging and Machine Learning Algorithms to Characterize Acute Ischemic Stroke Patients for Treatment Decision-Making

**Author**

Polson, Jennifer

**Publication Date**

2023

UNIVERSITY OF CALIFORNIA

Los Angeles

# Leveraging Clinical Imaging and Machine Learning Algorithms to Characterize Acute Ischemic Stroke Patients for Treatment Decision-Making

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Bioengineering

by

Jennifer Sara Polson

2023

# ABSTRACT OF THE DISSERTATION

## Leveraging Clinical Imaging and Machine Learning Algorithms to Characterize Acute Ischemic Stroke Patients for Treatment Decision-Making

by

Jennifer Sara Polson

Doctor of Philosophy in Bioengineering

University of California, Los Angeles, 2023

Professor Corey W. Arnold, Chair

For patients diagnosed with acute ischemic stroke, treatments such as thrombolysis and thrombectomy aim to restore blood flow to areas experiencing ischemia. These treatments have vastly improved outcomes, but it is currently unknown why some patients experience unsuccessful reperfusion or hemorrhagic complications. Taking advantage of recent advances in deep learning vision transformers, we developed algorithms for classification and prediction tasks regarding a patient's potential response to therapies using imaging taken at hospital admission. These models achieved higher generalization performance when identifying patients within the treatment window and those that will achieve successful recanalization. Our results illustrate that magnetic resonance (MR) and computed tomography (CT) imaging contains signal that can predict successful treatment response and that deep learning models can localize to salient regions within imaging without requiring time-intensive manual segmentation.

The dissertation of Jennifer Sara Polson is approved.

Peipei Ping

William Hsu

Alex Anh-Tuan Bui

William F. Speier

Corey W. Arnold, Committee Chair

University of California, Los Angeles

2023

## TABLE OF CONTENTS

## LIST OF FIGURES

# LIST OF TABLES

# ACKNOWLEDGMENTS

The work contained in this dissertation includes several previously published works, the authors of which I would like to thank, and re-thank, for their contributions. Chapter 3 contains material from Zhang, H., Polson, J. S., et al. (2021). Intra-domain task-adaptive transfer learning to determine acute ischemic stroke onset time. Computerized Medical Imaging and Graphics, 90, 101926 and Polson, J. S., Zhang, H., et al. (2022). Identifying acute ischemic stroke patients within the thrombolytic treatment window using deep learning. *Journal of Neuroimaging*, *32*(6), 1153-1160. Chapter 4 comprises work published as: Polson, J., Zhang, H., et al. (2021, November). A Semi-Supervised Learning Framework to Leverage Proxy Information for Stroke MRI Analysis. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 2258-2261). IEEE. Chapter 5 is an amalgamation of the following publications: Zhang, H., Polson, J., et al. (2021, July). A machine learning approach to predict acute ischemic stroke thrombectomy reperfusion using discriminative MR image features. In *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)* (pp. 1-4). IEEE; Zhang, H., Polson, J., et al. (2022). Predicting Thrombectomy Recanalization from CT Imaging Using Deep Learning Models. In *Medical Imaging with Deep Learning.* Finally, I would like to acknowledge my gratitude for the UCLA Department of Bioengineering and the NIH for providing financial support of my dissertation work through the following grants: NIBIB T32EB016640 and NINDS R01NS100806.

My graduate experience has been filled with myriad challenges, triumphs, and lessons learned thanks to the community of people that was essential to my progression. First, I express

my deepest gratitude to the members of my committee. Thank you to Dr. Corey Arnold for serving as my advisor; your leadership, precision, and patience enabled me to learn and improve both as a researcher and as a human, and your sense of humor was a crucial source of levity amidst the challenges faced throughout my studies. To the ever-meticulous and compassionate Dr. William Speier, I am grateful for your listening ear and helpful suggestions throughout my research. Thank you to Dr. William Hsu for providing ample programmatic and philosophical foundation upon which I could build out this research. Dr. Alex Bui, I am grateful for your constant encouragement and thoughtful questions to help me expand my way of thinking. And thank you to the inimitable Dr. Peipei Ping, without whom I would not have had the confidence nor ability to pursue this field of study.

My sincerest thanks to Haoyue Zhang for complementing and reinforcing my strengths, helping me grow immensely as a researcher. I would also like to acknowledge members of CDx and MII for their collaborative support, including Karthik Sarma, Johnny Ho, Jaiyun Li, Wenyuan Li, Yiwen Meng, Tianran Zhang, Leihao Wei, Zichen Wang, Carlos Olivares, Saarang Panchavati, Mara Pleasure, Ashwath Radhachandran, Katya Redekop, Nathan Siu, Eric Yang, Ilyass Majji, Simon Han, David Gordon, Yannan Lin, Samir Akre, Henry Zheng, Nova Smedley, Edgar Rios Piedra, and Panayiotis Petousis. Additional thanks to faculty members, Dr. Ricky Tiara, Dr. Denise Aberle, Dr. Kambiz Nael, and Dr. Suzie El-Saden, for providing mentorship during my graduate studies. Our scholarly research would not have been possible without the support of key staff members, including Shawn Chen, Rose Doherty, Stevie Marks, Isabel Rippy, and Denise Luna.

These research efforts would certainly not have come to pass without my personal support. I thank my parents, Tom and Kathy Polson, and my siblings, David and Leigh Polson, for supporting and cultivating my scientific spirit and intellectual curiosity throughout my life. Last and certainly not least, I express my deepest and most heartfelt gratitude to my partner, Dave Anders-Richards, for being a singular embodiment of unconditional love and support. Thank you.

# CURRICULUM VITAE

| | |
|---|---|
| 2010 – 2014 | B.S. Biochemistry, University of California Los Angeles. |
| 2015 – 2017 | M.S. Integrative Biology & Physiology, University of California Los Angeles. |
| 2017 – | Graduate Student, Computational Diagnostics Lab, Department of Radiological Sciences, David Geffen School of Medicine at UCLA. |
| 2021 – 2022 | Data Science Imaging Intern, Genentech. |
| 2021 – | Research & Development Analyst, SimX. |

# SELECTED PUBLICATIONS

Polson, J. S.*, Zhang, H.*, Nael, K., Salamon, N., Yoo, B. Y., El-Saden, S., Starkman, S., Kim, N., Kang, D., Speier, W.F. & Arnold, C. W. (2022). Identifying acute ischemic stroke patients within the thrombolytic treatment window using deep learning. Journal of Neuroimaging, 32(6), 1153-1160.

Wang, D., Lee, J., Polson, J. S., Liem, D. A., & Ping, P. (2022). A Protein-Centric Perspective of Autophagy and Apoptosis Signaling and Crosstalk in Health and Disease. In Biochemistry of Apoptosis and Autophagy (pp. 1-22). Springer, Cham.

Polson, J.S.*, Zhang, H.*, Nael, K., Salamon, N., Yoo, B., Speier, W., & Arnold, C. (2021, November). A Semi-Supervised Learning Framework to Leverage Proxy Information for Stroke

MRI Analysis. In 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE.

Zhang, H.*, Polson, J.S.*, Nael, K., Salamon, N., Yoo, B., Speier, W., & Arnold, C. (2021, July). A Machine Learning Approach to Predict Acute Ischemic Stroke Thrombectomy Reperfusion using Discriminative MR Image Features. In 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI) (pp. 1-4). IEEE.

Li, W., Li, J., Polson, J.S., Wang, Z., Speier, W., & Arnold, C. (2021). High-resolution histopathology image generation and segmentation through adversarial training. Medical Image Analysis, 102251.

Zhang, H.*, Polson, J.S.*, Nael, K., Salamon, N., Yoo, B.Y., El-Saden, S., Scalzo, F., Speier, W., & Arnold, C. W. (2021) Intra-Domain Task-Adaptive Transfer Learning to Determine Acute Ischemic Stroke Onset Time. Computerized Medical Imaging and Graphics, 101926.

Li, W., Wang, Z., Li, J., Polson, J.S., Speier, W., & Arnold, C. W. (2019, May). Semi-supervised learning based on generative adversarial network: a comparison between good GAN and bad GAN approach. In CVPR Workshops.

Ping, P., Hermjakob, H., Polson, J.S., Benos, P. V., & Wang, W. (2018). Biomedical informatics on the cloud: a treasure hunt for advancing cardiovascular medicine. Circulation research, 122(9), 1290-1301.

Garlid, A. O.*, Polson, J.S.*, Garlid, K. D., Hermjakob, H., & Ping, P. (2016). Equipping physiologists with an informatics tool chest: toward an integrated mitochondrial phenome. In Pharmacology of Mitochondria (pp. 377-401). Springer, Cham.

# CHAPTER 1

# Introduction

## 1.1 Motivation

Stroke is a major public health concern, as it is one of the leading causes of morbidity and mortality worldwide. In the United States, stroke is the fifth leading cause of death, and approximately 795,000 people in the United States are estimated to have a stroke each year, with acute ischemic stroke (AIS) accounting for the large majority of cases. Patients with strokes can have significant, detrimental impacts on their lives; stroke can lead to long-term disability, including paralysis, cognitive impairment, and communication difficulties. These disabilities can have a significant impact on the quality of life of the patient and their ability to function independently. Additionally, stroke is one of the leading causes of institutionalization and long-term care, which can have a significant economic impact on healthcare systems and society. Given the high incidence, impact, and cost of stroke, there is significant motivation to reduce incidence and improve patient outcomes. In addition to prevention and recovery efforts, a prompt and accurate assessment of the patient's condition is of paramount importance in the diagnosis and treatment of AIS, as the longer the brain is deprived of blood flow, the more extensive the damage it sustains.

Machine learning (ML) algorithms have been demonstrated to be an effective tool in this regard, providing automated analysis of patient data across multiple modalities, potentially improving clinical decision making in an expedient fashion. In addition to providing a quick and accurate diagnosis, machine learning algorithms can also help doctors better understand the underlying causes of a patient's stroke and predict the likelihood of future strokes. This can be particularly helpful in cases where the underlying cause of stroke is unclear or where a patient has multiple risk factors. By analyzing imaging data and other patient information, machine learning algorithms can provide physicians with insight into the factors that are most likely to contribute to a stroke and help them to develop personalized treatment plans for each patient. The use of machine learning in acute ischemic stroke is motivated by the need for faster and more accurate diagnosis and treatment of this condition. Using the power of machine learning, doctors and researchers can gain insight into the underlying causes of stroke and develop more effective treatments to improve patient outcomes.

### 1.1.1 Identifying Patients Within the Thrombolytic Treatment Window

Thrombolysis is a medical treatment that aims to dissolve blood clots that have formed in blood vessels. It is achieved by administering a thrombolytic agent, such as tissue plasminogen activator (tPA), which promotes the breakdown of the clot, allowing blood to flow normally again. For AIS patients, the benefit of thrombolytic therapy is positively associated with earlier reperfusion time.[1,2] Until recently, thrombolysis was only recommended for AIS patients with a known symptom onset time (i.e. time since stroke or TSS) within 4.5 hours.[2,3] AIS with unknown or unclear TSS has been reported in as many as 35% of patients.[4] In one study, only 6.5% of patients hospitalized for AIS received intravenous thrombolysis, with unknown TSS being the primary reason for treatment exclusion.[3] Many studies have sought clinical factors to assess eligibility and risk for thrombolytics, with a significant focus on neuroimaging.[5–7]

### 1.1.2 Predicting Patient Responses to Endovascular Thrombectomy

Endovascular thrombectomy (EVT) physically removes the clot from the occluded artery and has an entirely different set of risk factors that could influence outcome, such as collateral status. Collateral status refers to the compensatory vascular anatomy that may provide blood perfusion to areas experiencing ischemia.[8] Per the current clinical guidelines, advanced imaging such as perfusion-weighted imaging may be taken at admission to assess for collateral flow from other arteries in the brain.[2] The presence of high collateral status may indicate that a patient will respond well to invasive mechanical thrombectomy therapies.[9–11] As with collateral status, other clinical response variables have been substantially linked to post-thrombectomy outcome. For example, restoration of blood flow to the stroke area is quantified by the change in cerebral infarction (TICI) score.[6,7,12] During EVT, this score is assessed after each attempt to remove the clot, and a final TICI score is assigned to the patient at the end of the procedure to signify how much blood flow has been restored.

## 1.2 Challenges and Objectives

Although recent technological advancements have made machine and deep learning applications more robust and accessible, there remain several challenges to apply these algorithms to clinical tasks. As models increase in complexity, larger amounts of training data are required for them to learn the underlying feature representations that are salient to the pertinent tasks. However, there are substantial regulatory, fiscal, and societal barriers that prevent the sharing and aggregation of disease-specific datasets at a broad scale. Beyond these domain-level concerns, rapid diagnostic and treatment paradigms for suspected stroke patients introduce multifaceted limitations for algorithm development, evaluation, and eventual deployment, particularly in pretreatment and intraprocedural contexts.

### 1.2.1 Stroke Tissue Progression is Complex

The Efficacy and Safety of MRI-Based Thrombolysis in the Wake-Up Stroke (WAKEUP) trial showed that signal mismatch between diffusion-weighted imaging (DWI) and fluid-attenuated inversion recovery (FLAIR) can be used to select AIS patients with unknown TSS for thrombolytic treatment.[13] Accordingly, use of DWI-FLAIR mismatch is now recommended (evidence level IIa) in the updated American Heart Association-American Stroke Association (AHA-ASA) guidelines to identify unwitnessed AIS patients who may benefit from thrombolytic treatment.[2] However, assessing DWI-FLAIR mismatch relies on radiologists who may have differing levels of skill or approaches to image interpretation. Thus, this approach is prone to reader variability that may result in the erroneous exclusion of patients who could benefit from thrombolytic treatment.[14] TSS on the other hand, is an objective surrogate biomarker in clinical settings. Therefore, an automated method that accurately classifies TSS would increase the number of patients eligible for thrombolytic treatment. Machine learning has shown utility for stroke-specific clinical decision support,[15,16] and deep learning specifically has been widely explored for imaging-based tasks.[17,18] However, models often perform more poorly on unseen external datasets, requiring external evaluation of these algorithms.[19]

### 1.2.2 Rapid Acquisition Impacts Image Quality

Substantial efforts have been made to explore additional imaging protocols for the characterization of vascular physiology that assess stroke progression and inform potential treatment response. Several algorithms have been developed using perfusion imaging as input to identify physiological changes in brain tissue.[15,17,20] However, given the time-dependent benefits of swift stroke diagnosis and treatment, current American Heart Association/American Stroke Association (AHA/ASA) guidelines recommend studies be completed within 20 minutes of arrival to the hospital.[2] Clinical stroke imaging acquisition protocols are optimized to gather only the minimum information necessary to diagnose ischemic stroke and rule out contraindicative phenomena such as hemorrhage. Consequently, protocols used in routine clinical practice typically lack advanced perfusion imaging series.

These rapid protocols currently recommended by AHA/ASA generate images of substantially lower quality than those taken with protocols with longer acquisition windows. The slice thickness of CT or MRI acquired for suspected stroke can be up to 5mm, substantially thicker than anatomical scanning protocols for such neurological diseases as glioblastoma, multiple sclerosis, and dementia.[21–23] This also means that it is often not feasible to perform a second study if the image quality is low, and the final image may contain substantial noise or artifacts. These quality control issues can limit the ability of machine learning algorithms to distill information from these images, preventing landmark-based preprocessing or model convergence during training. Finally, for patients with AIS, diagnostic and treatment decision-making takes place on the order of minutes. Thus, models developed with a clinical use case in mind must prioritize inferential efficiency. This can preclude the use of time-intensive manual annotations or preprocessing steps in machine learning pipelines designed for pretreatment classification or prediction tasks.

## 1.3 Contributions

In this work, we seek to develop automated approaches that leverage clinical imaging to characterize patients with acute ischemic stroke before and during treatment, to better inform clinicians. Building on our previous work, our objective is to expand our lightweight medical image architecture to determine the eligibility of a patient for AIS treatment, either by thrombolytics or by thrombectomy. Utilizing novel architectures and training approaches, our objective is to classify or predict a patient's time clock (TSS), tissue clock (DWI-FLAIR mismatch), thrombectomy recanalization (mTICI), and first pass thrombectomy (FPE), and collateral status, to provide eligibility information for clinicians. This work represents efforts across several modalities, with efforts toward classifying vascular characteristics and predicting likelihood of successful therapies. This dissertation can be summarized in three aims:

Aim 1 *To explore the relationship between TSS and DWI-FLAIR mismatch for patients through fully and semi-supervised deep learning methods.* We hypothesize that patients with AIS have individualized responses to ischemic injury that can be characterized by standard pretreatment imaging protocols. We first report on the evaluation of a fully supervised deep learning algorithm to classify TSS on a data set comprising patients from two institutions. Our model achieved higher generalization performance on external evaluation datasets than the current state-of-the-art for TSS classification. We then employ a semi-supervised deep learning framework to classify the DWI-FLAIR mismatch, which utilizes prior information about TSS to generate pseudolabels for training. These results demonstrate the potential for automatic assessment of onset time from imaging without the need for highly trained radiologists. More broadly, these efforts illustrate that incorporating clinical proxy information into semi-supervised learning frameworks can improve model optimization by increasing the fidelity of unlabeled samples included in the training process.

Aim 2 *To develop machine learning frameworks to predict response to endovascular therapy from pretreatment stroke imaging.* We hypothesize that machine learning methods can predict a patient's response to treatment from image series acquired during routine clinical treatment. We first develop an annotation-free radiomics approach that uses atlas-based thresholding to compute features from a region of interest and then feed those into feature selection and machine learning algorithms. We then compare this approach to a deep learning algorithm that utilizes weight sharing and non-local attention modules. We evaluate the models on both MRI and CT images, reporting the cross-validation performance and performance on unseen prospective test sets. The results demonstrate that certain advanced imaging series improve the predictive capacity of the deep learning model and, to our knowledge, represent the first annotation-free prediction models for recanalization and first-pass effect.

Aim 3 *To extract temporal imaging biomarkers from images taken during an EVT procedure via generation of perfusion parameter maps and transformer-based temporal deep learning networks.* We hypothesize that digital subtraction angiography (DSA), taken during EVT, can be leveraged during the procedure to inform procedural decision-making. We first demonstrate the ability of DSA to capture vascular parameters by creating perfusion maps from temporal DSA sequences and comparing these with pre- and post- treatment MRI series. We then implement transformer-based architectures to generate a binary classification of a patient's collateral flow in the affected stroke region. These networks are then compared for their ability to classify collateral flow from a DSA sequence taken during a single pass of EVT, evaluating these methods against a network previously developed on DSA images. The results indicate that transformers can leverage raw unprocessed versions of DSA sequences to provide information on the physiology of a patient and inform treatment decision-making.

## 1.4 Organization of the Dissertation

The rest of the dissertation will proceed as follows. Chapter 2 illustrates the landscape of acute ischemic stroke diagnosis and treatment, provides an overview of machine learning and deep learning, and highlights recent and notable literature in the field of artificial intelligence for ischemic stroke. Chapter 3 summarizes efforts to classify stroke onset time from admission imaging for patients with unknown stroke onset. This chapter is based on a previous publication. Chapter 4 outlines efforts to detect penumbral tissue as targets for thrombolytic therapy, which is a summary of previously published work. Chapter 5 describes efforts to predict clinical response to EVT using pretreatment imaging. Chapter 6 contains a recent proof-of-concept study to classify vascular parameters from DSA images acquired during EVT. Chapter 7 concludes the dissertation, describing the limitations of each study and the utility of the findings, as well as suggesting future directions of study.

# CHAPTER 2

# Background

## 2.1 Acute Ischemic Stroke

Acute ischemic stroke (AIS) occurs when blood flow to a part of the brain is blocked by a clot. Stroke is a leading cause of death and disability worldwide. According to the World Health Organization, stroke is the second leading cause of death globally and is one of the leading causes of disability, with more than 50 million people living with the effects of stroke.[24] The prevalence of stroke varies by region, with higher rates typically seen in low- and middle-income countries. The severity of stroke can also vary, with some people experiencing only mild symptoms, while others may have more severe symptoms that require long-term care. Common symptoms of stroke include sudden weakness or numbness on one side of the body, difficulty speaking or understanding language, difficulty seeing in one or both eyes, and a severe headache.

During a stroke, the lack of blood flow to a part of the brain (ischemia) causes a lack of oxygen and nutrients to brain cells. This leads to a decrease in the production of available cellular energy required for regular function. Cells may also start to produce free radicals, which can damage the cell membrane and other cell structures. As ischemia continues, cells may release neurotransmitters, such as glutamate, that can overstimulate other cells and cause them to die. The cells may also release enzymes, such as proteases and lipases, which can break down the cell's structural proteins and fats. These reactions can further damage the cell and contribute to its death. If ischemia is not reversed and blood flow is not restored, brain cells will eventually experience cell death, known as infarction. The brain has limited ability to recover from stroke, so it is crucial to seek medical attention as soon as possible. Early treatment can help minimize the extent of the damage and improve the chances of a successful recovery.

Substantial clinical research has shown that minimizing the time between the onset of symptoms and thrombolytic administration can improve outcomes. Therefore, prehospital and hospital workflows are designed to minimize the delay between symptom recognition and treatment. Upon admission, hospital provider teams at primary stroke centers perform immediate diagnostics, including an image study, in rapid succession. These diagnostic studies and labs allow the provider to assess patient eligibility for two different treatments: thrombolysis and thrombectomy. The prehospital workflow recommended by the American Heart Association/American Stroke Association is detailed in Figure 2-1.

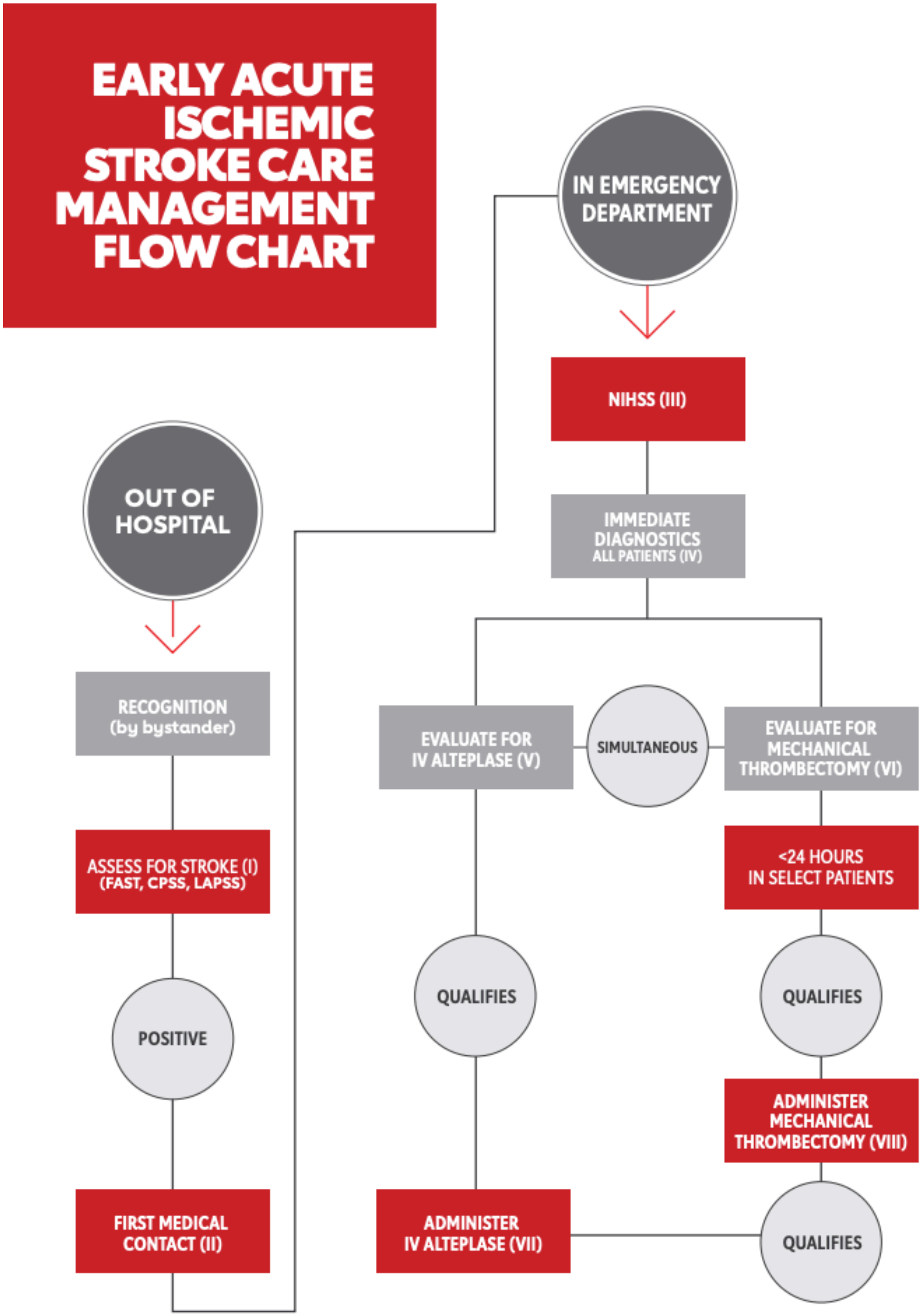


Figure 2-1. Prehospital and admission workflow for patients with suspected stroke. Reproduced from the AIS Toolkit[25]

### 2.1.1 Thrombolysis

Thrombolytics, such as tissue plasminogen activator (tPA), are medications that are used to dissolve blood clots and restore blood flow to the brain. tPA works after injection by binding to fibrin infrastructure of a clot thrombus. Once bound, the enzyme cleaves peptide bonds on plasminogen found in the clot, turning it into plasmin. Plasmin is a serine protease that works directly on fibrin, thereby loosening the fibrous structure and dissolving the clot. For many thrombus types, tPA is effective at restoring blood flow to the brain; ideally, thrombolytics work to reperfuse tissue that is experiencing ischemia but that is not yet infarcted. Accordingly, it must be given within a certain time frame after the onset of stroke symptoms to be effective. tPA also carries a risk of hemorrhagic transformation and thus may be contraindicated for many patients.

Per the current clinical guidelines in the U.S., thrombolysis is only recommended for patients meeting certain clinical criteria and with a known symptom onset within 4.5 hours.[2] Up to 35% of patients were deemed ineligible for this treatment due to unknown time since stroke (TSS).[26] However, the recent WAKEUP trial provided a new avenue for patients to receive thrombolytic treatment. Using MR imaging, a neuroradiologist may assess differences in signal between DWI and FLAIR series when stroke onset time is unknown to determine thrombolytic eligibility.[13] In fact, in the most recent version of the American Heart Association guidelines for treating acute stroke, for patients with unclear time of symptom onset, MRI can be performed to identify areas with DWI-FLAIR mismatch that could benefit from IV tPA. This can offer another therapeutic avenue for these patients; however, this assessment is prone to a large amount of reader variability and may exclude many patients from treatment.[14] At many institutions, the lack of expert level neuroradiologists could pose another problem given the short diagnostic window to evaluate MR images for stroke patients.

As clinical care teams cannot fully control the time between the onset of symptoms and the arrival of the patient to the hospital, substantial effort has been put into reducing door-to-needle time (DNT), defined as the time elapsed between when a patient arrives at the hospital and when receiving thrombolytic therapy. The American Heart Association & American Stroke Association (AHA/ASA) creates clinical treatment guidelines that are followed by hospitals in the U.S. and then adopted by stroke associations across the globe. They have launched a multiphase campaign called the Target: Stroke Initiative that created evidence-based strategies to reduce DNT as well as ideal DNT metrics for stroke centers (tertiary care centers that can provide interventions to patients admitted for stroke) to meet.

### 2.1.2 Endovascular Thrombectomy

Endovascular thrombectomy (EVT) represents another more invasive treatment avenue. This procedure involves using a device to physically remove a blood clot from within a blood vessel in the brain. Current clinical guidelines recommend two techniques: stent retrieval and direct aspiration. In stent retriever thrombectomy, a stent (a small metal mesh tube) is used to physically remove a blood clot from within a blood vessel in the brain. The stent is placed within the vessel and then expanded to capture the clot. The stent and clot are then removed from the vessel together. Direct aspiration thrombectomy uses a catheter with a small suction device to remove a blood clot from within a blood vessel in the brain. The catheter is inserted through a small incision in the skin and guided to the site of the clot using imaging techniques such as X-ray or ultrasound. These techniques represent the current state-of-the-art, but there are other techniques, including balloon-guided catheterization and stenting, that are still used depending on the device and training access.[27]

Through several clinical trials, EVT has been demonstrated to produce superior outcomes with statistically insignificant differences in complications.[28] However, only a limited number of patients are eligible for this treatment. Original trial criteria demonstrated efficacy in patients with a large vessel occlusion (LVO) identified on vessel imaging presenting within 6 hours of symptom onset, a NIH Stroke Scale (NIHSS) score $\geq 6$, an Alberta Stroke Program Early CT Score (ASPECTS) $\geq 6$, and a prestroke modified Rankin Score (mRS) of 0 or 1. Under these criteria, it is estimated that only 10% of patients with AIS would be eligible for this treatment.[29–32]

Procedural advances in neurointerventional radiology have continued to make great strides toward improving the safety and outcomes of patients with AIS. Two trials demonstrated efficacy for patients beyond the 6-hour window under different inclusion parameters. The DWI or CT perfusion (CTP) assessment with clinical mismatch in the Triage of Wake-Up and Late Presenting Stroke Undergoing Neurointervention with Trevo (DAWN) trial, EVT had demonstrated clinical efficacy and improved outcomes for patients within 6 to 24 hours of symptom onset who experienced a large clinical deficit with a small infarct volume, provided that they did not have an infarct covering more than a third of the middle cerebral artery territory or intracranial hemorrhage.[6,33,34] Following this, the DEFUSE 3 trial developed clinical and imaging conditions under which a patient could experience improved outcomes more than 6 hours from symptom onset.[35] Patients included in this study required a level of mismatch between the volume of signal on diffusion-weighted imaging and perfusion-weighted MR or CT imaging taken at admission. These two studies substantially expanded the pool of eligible patients.

The demonstrated success of EVT with DAWN and DEFUSE 3 criteria fomented significant efforts to expand other eligibility criteria. Recent trials have expanded the criteria along most of the axes to include non-proximal occlusions (such as distal branches of the middle cerebral artery and posterior circulation), mild strokes of LVO, large infarct volume, and those presenting beyond the 24-hour window.[36–41]

#### 2.1.2.1. Measuring EVT Success

Successful EVT requires reperfusion of the ischemic tissue, which is achieved through the reopening of blocked blood vessels (know as recanalization). This is measured primarily using the mTICI (modified thrombolysis in cerebral infarction) scoring system, which evaluates the degree of recanalization achieved after thrombolytic treatment or EVT. The TICI scale was first introduced as a modification of the previously used TIMI (thrombolysis in myocardial infarction) scale, which had a range of 0-2 compared to a range of 0-3 in mTICI. The mTICI scale was developed as a more comprehensive and accurate tool to assess recanalization after thrombolytic therapy, as it allowed for a more detailed assessment of the degree of recanalization achieved.

The mTICI scale ranges from 0 to 3, with 0 indicating no recanalization, 1 indicating partial recanalization, 2a-b indicating partial recanalization, and 3 indicating complete recanalization. Detailed criteria for each mTICI score can be seen in Table 2-1. As with any qualitative assessment, mTICI scoring can be prone to inter-reader variability. When studying this variability, Suh et al. obtained an intraclass coefficient (ICC) of 0.67.[42] Mair et al. used Krippendorff's Alpha (K-Alpha) to assess interrater variability in three groups: multicenter angiography panel experts (K-Alpha of 0.60), neuroradiology trainees (0.63), and nonexperts (0.39).[43] Tung et al. also examined interrater variability for and found an ICC of 0.77 and an overall Cohen's Kappa of 0.58 for oTICI (67 percent class boundary between 2a and 2b) and 0.62 for mTICI.[44]

Despite this variability is assessment, clinical trials have illustrated that patients who experience partial and/or full recanalization of the blood vessel typically experience better outcomes, particularly if recanalization is achieved in three attempts or less.[45–49] Moreover, the best long-term clinical outcomes have been associated with patients who experience recanalization after one attempt to remove the clot; this is known as the first pass effect (FPE).[50–55]

Table 2-1. Descriptions of mTICI scoring criteria. Scores are assessed using vascular imaging during treatments.

| TICI Grade | Original TICI | Modified TICI | Modified TICI With 2c |
|---|---|---|---|
| **0/1** | No/minimal reperfusion | No/minimal reperfusion | No/minimal reperfusion |
| **2a** | Partial filling <2/3 territory | Partial filling <50% territory | Partial filling <50% territory |
| **2b** | Partial filling ≥2/3 territory | Partial filling ≥50% territory | Partial filling ≥50% territory |
| **2c** | ... | ... | Near complete perfusion except slow flow or few distal cortical emboli |
| **3** | Complete perfusion | Complete perfusion | Complete perfusion |

### 2.1.2.2. EVT Complications

There are several adverse complications that can arise during EVT. Complications occur in about 20% of cases, and recanalization is not achieved in approximately 10-30% of patients, dependent on technique.[56] Vascular anatomy can affect clot retrieval in even highly proximal occlusions. The vessel may not be accessible because of unforeseen stenosis, elongation of the clot, pre-existing dissection within the vasculature, or inability to guide the device through the clot to deploy the retrieval mechanism.[57] When the clot can be effectively accessed, there can still be complications that prevent full removal of the clot thrombus and therefore successful recanalization. The most common complications arise when the clot is recalcitrant to removal; either peripheral components of the clot remain after EVT, or no material could be removed from the thrombus at all. Clinically ineffective reperfusion refers to a spectrum of phenomena wherein patients are recanalized during the procedure but tissue still experiences a lack of oxygen resulting in tissue death.[58] Recent clinical studies have explored the no-reflow phenomenon, defined as a lack of tissue reperfusion despite angiographic reperfusion.[59–62] Between 6-20% of patients are affected by reocclusion of the treated vessels within 24 hours of treatment.[63,64] Distal embolization can occur when the clot fragments upon contact with retrieval devices and clot fragments proceed further into the vasculature. There may also be damage to the vasculature itself; arterial perforation can cause symptomatic hemorrhage of the cerebrum or subarachnoid space. Even among patients without symptomatic complications, there are post-procedural complications that can affect outcome long-term.[56,65–68] Reperfusion injury is a phenomenon that can occur following the restoration of blood flow to an area of the brain that has been previously deprived of oxygen and nutrients due to an occlusive thrombus. It can occur due to several mechanisms, including the generation of reactive oxygen species, the activation of the complement system, and the influx of inflammatory cells. The consequences of reperfusion injury can include the exacerbation of ischemic injury, the development of edema, and the induction of neuronal death.[69–72]

There are several pathophysiological mechanisms that can influence the probability of recanalization and subsequent reperfusion in a patient. Factors correlated with functional outcomes following EVT include the location of clot, length of EVT procedure, use of statins, remaining thrombus fragments, and stenotic vessels.[64] Additionally, stroke etiology and impact on outcome is a prevalent area of study. For example, intracranial atherosclerosis (ICA), which is the accumulation of plaque in the blood vessels of the brain, can affect the success of EVT. Plaque build-up in blood vessels can narrow the lumen (the opening inside the blood vessel) and make it harder for the EVT device to reach and remove the clot. ICA has been shown to increase the risk of reocclusion and recurrent strokes, causing inflammation and damage to the arterial wall that can lead to further vascular injury or further narrowing of the vessel lumen.[66]

### 2.1.2.3. Adjuvant Therapies

If EVT does not proceed as planned, neurointerventionalists have a variety of adjuvant therapies at their disposal. These therapies are intended to enhance the effectiveness of thrombectomy by improving blood flow restoration, reducing the risk of complications, and mitigating the effects of ischemic injury. The most invasive include rescue stenting, which involves the use of a stent to mechanically open a blocked or narrowed blood vessel in the brain rather than retrieving the clot in standard EVT.[73] Clinical evidence suggests that, among refractory cases of EVT, rescue stenting can achieve recanalization rates similar to those for which EVT is successful.[74,75] Current clinical guidelines do not recommend rescue stenting as a first-line treatment option, given the increased risk of hemorrhage from the use of required antiplatelet therapeutics.[2]

Pharmacological agents that can also be utilized during or following EVT. Thrombolytics such as tPA can be administered intravenously (IV) or intra-arterially (IA) using a microcatheter or a thrombolytic delivery system in conjunction with EVT. The latter avenue has garnered recent interest due to the Chemical Optimization of Cerebral Embolectomy (CHOICE) trial, which compared the safety and effectiveness of the administration of IA-tPA versus placebo in the treatment of patients with acute LVO ischemic stroke who had already undergone a successful EVT. Their preliminary findings indicated that IA-tPA improved neurological status at 90 days compared to placebo with no significant impact on patient safety.[76] Additionally, some studies have suggested that agents such as magnesium, nicorandil, and statins, may have a neuroprotective effect and reduce the risk of reperfusion injury.[77,78] Further clinical trials have also indicated the use of anticoagulants such as heparin, may improve recanalization rates, decrease the risk of reocclusion and reduce the risk of no-reflow phenomenon; however, these are nascent studies that have not been clinically recommended.[79–81]

## 2.2 Stroke Imaging

Imaging plays a crucial role in the diagnostic process of AIS. One imaging technique commonly employed in the diagnosis of AIS is magnetic resonance imaging (MRI). MRI employs a strong magnetic field and radio waves to generate detailed images of the brain. MRI is particularly valuable in identifying the underlying causes of a stroke, such as thrombotic or hemorrhagic events, and in predicting the likelihood of future strokes, allowing for the development of personalized treatment plans for each patient.[5,82] Computed tomography (CT) scans are among the most frequently used imaging techniques in AIS diagnosis. CT scans generate detailed images of the brain utilizing X-ray technology and are performed quickly and efficiently in emergency settings. CT scans aid physicians in identifying the presence and location of a stroke, as well as assessing the extent of any brain damage.

### 2.2.1 Magnetic Resonance Imaging

MRI is a non-invasive imaging modality that uses a combination of a strong magnetic field, radiofrequency (RF) pulses, and a computer to produce detailed images of internal organs and structures within the body. MRI generation is based on the principle of nuclear magnetic resonance. An MRI machine uses a large, cylindrical magnet to generate the main magnetic field, which is usually between 0.5 and 3 Tesla. The patient is placed inside the magnet, and the RF pulses are emitted by a coil or antenna that surrounds the patient. Because of their positive charge, protons in atomic nuclei align when placed in a strong magnetic field and precess, or spin, around the nuclei at a specific frequency. Protons first absorb energy from a RF pulse. As the protons relax back to their original state, they emit energy in the form of RF signals, which can be detected by a coil or antenna. These signals are used to create images of the tissue being imaged.

MRI scanners can be programmed to utilize different pulse sequences and relaxation measurements to generate image series that highlight different types of tissue. T1-weighted images are created by using a short RF pulse and a short recovery time. This results in low signal intensity in tissues with a high concentration of water, such as cerebrospinal fluid, and high signal intensity in tissues with a low concentration of water. T2-weighted images, on the other hand, are created by using a long RF pulse and a long recovery time. This results in a high signal intensity in tissues with a high concentration of water and a low signal intensity in tissues with a low concentration of water. T2-weighted images are useful for visualizing pathology in the brain and spinal cord, such as edema, hemorrhage, and inflammation, as well as for detecting tumors and other lesions in other parts of the body. Both T1 and T2-weighted images can utilize contrast agents, which are paramagnetic compounds that alter the local magnetic field around the protons. This decrease in the longitudinal magnetization of protons results in a decrease in signal intensity on T1-weighted images and therefore increases in contrast between normal and abnormal tissue.

#### 2.2.1.1. Fluid-attenuated Inversion Recovery

Fluid-attenuated inversion recovery (FLAIR) is a pulse sequence that is commonly used in neuroimaging. The basic principle of FLAIR is the inversion of the longitudinal magnetization of fluid-like tissues, such as cerebrospinal fluid (CSF), which have a long T1 relaxation time, while leaving the magnetization of other tissues, such as brain tissue, unaffected. This leads to a suppression of the signal from fluid-like tissues, which results in improved contrast between the brain tissue and CSF. The FLAIR sequence is derived from the inversion recovery (IR) pulse sequence, which is typically used to suppress the signal from tissues with long T1 relaxation times. The IR pulse sequence is composed of a 90-degree RF excitation pulse, followed by a variable inversion time and a 180-degree RF refocusing pulse. The T1 is chosen to match the T1 relaxation time of the tissue being suppressed. In the FLAIR sequence, the T1 is prolonged to match the T1 relaxation time of the CSF. This is achieved by applying a frequency-selective inversion pulse before the 90-degree RF excitation pulse, which inverts the magnetization of the fluid-like tissues with a long T1 relaxation time, such as the CSF. In practice, FLAIR MRI images show hypointense signals in the CSF spaces, such as ventricles, cisterns, and sulci, due to the suppression of the signal from these fluid-like structures. FLAIR is particularly useful in neuroimaging because it allows for better visualization of lesions that have similar signal intensity as CSF, which can help to improve the detection and characterization of lesions, such as tumors and infarcts, that are in the brain's CSF spaces.

#### 2.2.1.2. Diffusion-weighted Imaging

Diffusion-weighted MRI (DWI) is a technique used to visualize the microstructure of biological tissues in the brain. DWI is sensitive to the random Brownian motion of water molecules in a tissue. The underlying theory is that the diffusion of water molecules is restricted by the presence of obstacles, such as cell membranes and fibers. By measuring the diffusion of water molecules in different directions, DWI can provide information about the microstructure of the tissue and its tissue properties. DWI is performed by applying a magnetic field gradient in a specific direction, which causes the water molecules to diffuse preferentially in that direction. The signal intensity in a voxel is then measured using the apparent diffusion coefficient (ADC) of water molecules. The

ADC is defined as the ratio of the diffusion coefficient to the density of the diffusing water. The ADC can be measured using the following formula:

$$ADC = \frac{(S_0 - S_b)}{S_0}$$

Where $S_0$ is the signal intensity without a gradient applied and $S_b$ is the signal intensity with a gradient applied. The $b$ value is the strength of the applied gradient, and it is related to the diffusion coefficient by the formula:

$$b = \gamma^2 * \Delta t * G^2$$

Where $\gamma$ is the gyromagnetic ratio, $\Delta t$ is the duration of the applied gradient and $G$ is the gradient strength. In clinical practice, DWI is used to measure the abnormal diffusion of water, which can highlight areas of restricted blood flow in the brain. During a stroke, the interruption of blood flow leads to an energy crisis in the affected tissue, resulting in the breakdown of the cellular membrane and the subsequent leakage of intracellular contents into the extracellular space. This leads to an increase in the extracellular space and a decrease in the water diffusion. Specifically, a reduction in the ADC values within the affected area is observed. The sensitivity of DWI in detecting acute ischemic injury is due to its ability to detect these microstructural changes within minutes of the onset of ischemia, whereas other MRI sequences such as T2-weighted or FLAIR may not detect these changes until several hours or even days later, when edema or tissue death has occurred.

#### 2.2.1.3. Perfusion-weighted Imaging

Perfusion-weighted imaging (PWI) is an advanced imaging technique based on the measurement of the contrast agent concentration in the blood over time, and it can provide valuable information about the hemodynamic status of the brain in the setting of stroke. Raw PWI data is generated by injecting a contrast agent into the bloodstream, typically a bolus injection of a paramagnetic contrast agent such as gadolinium-based contrast agents, and then measuring the signal intensity of the contrast agent in the brain over time using an MRI sequence such as dynamic susceptibility contrast (DSC) or dynamic contrast-enhanced (DCE) MRI. The DSC sequence is based on the measurement of the signal intensity changes caused by the susceptibility effect of the contrast agent, which is a property of the contrast agent that causes a local distortion of the magnetic field. The DCE sequence is based on the measurement of the signal intensity changes caused by the T1 relaxation effect of the contrast agent. The T1 relaxation effect of the contrast agent is caused by the interaction of the contrast agent with the protons in the tissue, which causes a change in the longitudinal relaxation time.

To derive hemodynamic parameters from serial tissue tracer concentration measurements, it is essential to have a general model that describes the way in which the tracer passes through or distributes in the target organ. This model must be based on an understanding of the administration method of the tracer, whether it be bolus injection or constant infusion, and assumptions about the pharmacokinetic properties of the tracer in the organ of interest. These assumptions include diffusibility from intravascular to extravascular space, volume of distribution, and equilibrium half-life of the tracer. Curves depicting changes in intensity based on the concentration of the contrast agent over time can be generated that are then used for computation of hemodynamic perfusion parameter maps.

The main parameters derived from PWI are time to maximum (Tmax), mean transit time (MTT), cerebral blood volume (CBV), and cerebral blood flow (CBF). Tmax represents the time to peak enhancement of the contrast agent in the blood, MTT represents the mean transit time of the contrast agent through a voxel, CBV represents the volume of the blood in a voxel, and CBF represents the blood flow through a voxel. Tmax and MTT are calculated by measuring the contrast agent concentration over time in the affected area and in a normal area of the brain. Tmax is the time at which the contrast agent concentration is highest in the affected area, and MTT is the average time required for the contrast agent to pass through the area. CBV and CBF are calculated by measuring the contrast agent concentration before and after a bolus injection. The difference in contrast agent concentration between the pre-injection and post-injection images is used to calculate CBV and CBF. CBV is calculated by dividing the difference in contrast agent concentration by the pre-injection contrast agent concentration, and CBF is calculated by dividing the CBV by the MTT.

PWI can be used to detect and localize the area of ischemia or infarction caused by the interruption of blood flow. In the acute phase of stroke, CBF values in the affected area decrease, indicating a decrease in blood flow, consistent with the ischemic injury.

As the tissue recovers or degenerates, the CBF values may increase or decrease accordingly.

### 2.2.2 Computed Tomography

CT imaging is widely used in the diagnosis of AIS due to its ability to provide detailed images of the brain quickly and easily in an emergency setting. During a CT scan, the patient is positioned on a table that moves through a circular opening in the CT scanner. The X-ray tube and detector rotate around the patient, capturing multiple X-ray images from different angles. X-rays are electromagnetic radiation with a short wavelength and high energy, which allows them to penetrate through the body and be absorbed by different tissues to varying degrees. The absorption of X-rays by the body is dependent on the molecular composition and density of the tissue, which allows for the creation of detailed images of the internal structures of the body. To acquire the image, a narrow beam of X-rays is directed through the patient, and the intensity of the X-rays that pass through the patient is measured by an array of detectors on the other side. This raw projection data is a series of X-ray attenuation measurements taken at different angles around the patient. The machine then uses image reconstruction algorithms to convert the projection data into an image that represents the distribution of X-ray attenuation coefficients within the patient's body.

CT images can be generated with or without contrast using an iodine-based agent. While non-contrast CT (NCCT) images are at a lower resolution than those with intravenous contrast, they require a lower radiation dose and can be generated more rapidly. In stroke, the contrast serves to provide more detailed visualization of the vasculature and other anatomical structures within the brain. After the contrast agent is injected, a series of CT images are acquired at different time intervals. The images are then processed using specialized image reconstruction techniques, such as maximum intensity projection (MIP), to create detailed images of the blood vessels. MIP is a 3D image reconstruction technique that uses the most intense voxel within a specific region of interest to create a single image that highlights the blood vessels. Additionally, the use of contrast can also generate CT perfusion (CTP) images. In these protocols, CT scanners rapidly acquire images of the brain at several time points, generating a temporal volume that can trace the contrast agent. As the agent is permeable, these time-series images visualize how the contrast perfuses from blood vessels into the brain tissue. These can be processed in the same fashion as MRI-perfusion to generate perfusion parameter maps.

### 2.2.3 Digital Subtraction Angiography

Digital subtraction angiography (DSA) is a medical imaging technique used to visualize blood vessels in the body. It is commonly used to diagnose and treat conditions that affect the blood vessels, such as blockages or aneurysms. In a DSA exam, a radiosensitive contrast agent is injected via catheter into proximal blood vessels. X-ray images are taken prior to and following contrast injection. The contrast absorbs the radiation, making blood vessels appear with high contrast. This allows the blood vessels to be clearly visualized and any abnormalities to be detected. At the end of the exam, the images are subtracted, removing any tissue that did not receive contrast. The resulting image highlights the vascular structure, which can be useful for diagnosing vessel abnormalities or providing prognostic value during therapeutic procedures.

Within ischemic stroke, DSA can be used to visualize stenosis, or the thinning of artery openings, and characterize the extent of stenosis. It remains the gold standard for visualization of collateral flow. Beyond diagnosis, DSA provides intra-procedural insight into the vasculature during several phases of EVT, as it's used to guide catheter placement and navigation to the thrombus and to confirm successful recanalization or clot retrieval following removal attempts. Modern stent retriever devices have even been manufactured with radiopaque materials to be optimally visualized during the procedure. The use of DSA during EVT has become standard clinical practice among primary stroke institutions, as it's been demonstrated to reduce the rate of vessel injuries.[83–85]

During EVT, DSA images are acquired in two views. Frontal DSA images are taken from a direct front-facing angle, looking at the coronal view. These images provide detailed information about the blood vessels in both hemispheres of the brain, allowing for contralateral visualization of vessels in addition to those on the affected side. Lateral DSA images are taken from a side angle, looking at the blood vessels from a sagittal view. These images can provide more granular visualization of blood vessels in the posterior portion of the brain, such as the vertebral artery and the posterior cerebral artery. Both frontal and lateral views are typically taken during a DSA examination to provide a comprehensive understanding of the blood vessels in the brain and to identify any abnormalities or blockages. The neurointerventionalist will use the information from both views to plan the best approach during the procedure.

## 2.3 Machine Learning

Machine learning (ML) is a subfield of artificial intelligence that involves the development of algorithms and statistical models that enable systems to learn from and make predictions about data. These models can identify patterns and relationships within large and complex datasets and can be used to make predictions and decisions without explicit human intervention. ML methods can have varying degrees of supervision during training and typically belong to one of a few categories: supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, the algorithm is provided with labeled training data and learns to make predictions based on this data. In unsupervised learning, the algorithm is provided with unlabeled data and learns to identify patterns and relationships within the data. In reinforcement learning, the algorithm is trained through trial-and-error interactions with its environment. One of the key benefits of machine learning is its ability to improve the performance of classification or prediction tasks by automatically learning patterns underlying data representations.

### 2.3.1 Radiomics

Radiomics is an emerging field that leverages advanced image processing techniques and machine learning algorithms to extract a wide range of quantitative features and biomarkers from medical images that can then be used for various clinical applications. These features can then be used to classify, predict, or monitor various aspects of disease progression and treatment response.

A radiomics pipeline typically requires image preprocessing, then segmentation of the image into a region of interest (ROI). From an ROI, features can be extracted that are categorized into three types: first-order statistics, shape-based features, and texture-based features. First-order statistics include measures such as mean, standard deviation, and histogram-based features. They provide information on the overall intensity distribution of the ROI. Shape-based features provide information on the geometric properties of the ROI, and may include area perimeter, and compactness. Beyond these features, statistical methods can include texture-based features that provide detail as to the spatial distribution of intensities within the ROI. These features can then be used in any manner of machine learning.

The applications of radiomics are diverse and include the identification of imaging biomarkers for disease diagnosis, prognosis, and treatment response prediction. For example, radiomics has been used to extract features from CT images for the prediction of treatment response in lung cancer, prostate cancer, and head and neck cancer. Radiomics has also been used to extract features from MRI images for the diagnosis of brain tumors and prediction of treatment response in glioma. Radiomics has also been applied in the field of radiation therapy, where it has been used to predict treatment response and to monitor tumor response to radiation therapy.

Despite significant potential to improve the diagnostic accuracy, prognostic power, and treatment response prediction of various diseases, the field of radiomics is still in its infancy and there is a need for further research to establish the clinical utility and generalizability of radiomics features across different imaging modalities, disease types, and populations. To achieve this, it is necessary to develop standardized and validated image processing pipelines and to establish rigorous evaluation criteria for the assessment of radiomics features.

### 2.3.2 Deep Learning

Deep learning is a subfield of ML that involves the use of neural networks to perform tasks that are typically difficult for traditional machine learning algorithms, such as image and speech recognition, natural language processing, and decision-making. Deep learning algorithms are composed of multiple layers, each of which is responsible for a specific feature extraction or transformation. The layers are organized in a hierarchical structure, with the lower layers responsible for simple feature extraction, such as edge detection, and the upper layers responsible for more complex feature extraction, such as object recognition. This hierarchy of layers is what gives deep learning its name, as it allows the algorithm to learn a hierarchy of features, from simple to complex. Deep learning is different from traditional machine learning in a few ways: Deep learning models can process large amounts of data with high dimensionality. When trained, these models can extract features from the input data where traditional machine learning algorithms need to be provided with predefined features. These models can learn non-linear relationships between input and output data. Finally, deep learning algorithms can generalize well to new data, even when it has a different distribution than the training data.

### 2.3.3 Semi-Supervised Learning

Semi-supervised learning (SSL) is a machine learning technique that uses a combination of labeled and unlabeled data to train a model. It is often used when there is a large amount of unlabeled data available, but only a small amount of labeled data. SSL relies on several assumptions. First, if each sample in a set is plotted along feature axes, the set of samples lies in a continuous space (known as the continuity assumption). Second, dimensions of the input space's manifold are significantly higher than that on which the data lies (known as the manifold assumption). Finally, for the cluster assumption, the data has natural predilection to form clusters and that data points that are a part of the same cluster have the same label.

Semi-supervised approaches have utility for medical machine learning use cases. It can be difficult and expensive to obtain a large amount of labeled medical data that has been labeled by clinicians or other medical experts. In this case, semi-supervised learning can be used to make the most of the limited amount of labeled data available by incorporating additional unlabeled data into the training process. A semi-supervised approach can help to improve the performance of medical diagnostic systems. For example, by using semi-supervised learning to train a model on a combination of labeled and unlabeled medical data, it may be possible to develop a diagnostic system that is more accurate and reliable than a system trained only on a small amount of labelled data samples.[86]

## 2.4 Explainable AI

Explainable AI (XAI) refers to the field of artificial intelligence that aims to develop algorithms and models that can be easily understood and explained by humans. The goal of XAI is to provide interpretability and transparency into the decision-making process of AI systems, so that the reasoning behind the predictions made by these systems can be understood and trusted.

In deep learning, interpretability methods enable the examination of the underlying patterns and rules that are learned by an algorithm, as well as the identification of potential biases and areas of failure within the model. These methods can be used to evaluate the model's performance, and improvements can be made to enhance its overall performance. Additionally, interpretability is an essential aspect for ensuring the ethical alignment of the model, such as identifying and mitigating unintended discrimination or bias. Medical deep learning models are often used to make decisions that have a direct impact on patient care and treatment, so it is essential that the decisions made by these models can be understood and explained by medical professionals. Additionally, medical images can be noisy and hard to interpret, interpretability helps to understand the decision-making process and identify the important features in the image.

There are many different techniques and methods that are used in XAI at different stages of model deployment. Pre-modeling explanations can work to describe the data and modeling inputs. During development, models can be chosen that are inherently interpretable, depending on the model inputs and parametric complexity. In deep learning, most XAI methods are implemented post-modeling to generate local or global explanations. These XAI methods include model-specific and model-agnostic interpretability methods that can utilize gradients or activations within the model to generate visualizations.

### 2.4.1 Gradient-Based Saliency Maps

In deep learning, saliency maps are often used to help understand and interpret the decisions made by a neural network when classifying an image. A saliency map is a visual representation of an image that highlights the regions that are most important or most relevant to a particular task. Generating saliency maps relies on backpropagation, computing the gradients of the loss with respect to the learned weights of the network. Given an input image $I$ with dimensions $(x, y)$ and a target class $c$, the image is fed into a network to generate a class score $S_c(I)$ and activation $a$. Backpropagation calculates the error between $S_c(I)$ and the true output. From this loss, the gradient is calculated with respect to the weights of the network:

$$w_{i,j} = \frac{\partial S_c}{\partial I}$$

The gradient $w_{i,j}$is then normalized and rearranged to match the dimensionality of the input. As this is computed for every input feature, the output shares 1:1 dimensionality with an input. For an image, this algorithm computes the gradient of the output prediction, illustrating how much the prediction would change if we perturbed each pixel of the input image. Higher values in the saliency map indicate regions of the input image that are more important for the prediction. Beyond this vanilla implementation of gradient maps, there are many variations and refinements to this algorithm, such as using guided backpropagation or integrating gradients across multiple layers.

### 2.4.2 Grad-CAM

Gradient class activation maps (Grad-CAMs) are another method for highlighting which regions of an image are most important in a deep learning network.[87,88] The Grad-CAM algorithm generalizes class activations maps (CAMs) by using the gradient information and activations from a trained network to find regions of the image that are most relevant to the prediction. Using the same conventions from 2.4.1, $f_k(x, y)$ can represent the activation at the last convolutional layer for a given pixel location, and $\omega_k^c$ can represent the weight corresponding for the class at that convolutional unit. Accordingly, the CAM can be given as follows at each spatial point:

$$A_c(x, y) = \sum_k \omega_k^c \ f_k(x, y)$$

Using these activation maps, the next step is to compute the gradient of the output of the network with respect to the input image from the activations $A^k$ and class $y^c$ , $\frac{\partial y^c}{\partial A_{i,j}^k}$, which are then subjected to global average pooling along the x and y dimensions:

$$\alpha_k^c = \sum_{x,y} \frac{\partial y^c}{\partial A_{i,j}^k}$$

The gradient of the output of the network with respect to the input image is utilized to determine the level of perturbation in the output of the network, resulting from a small change in the input image. Subsequently, the gradients are multiplied by the activations of the final convolutional layer of the network, which serves to accentuate the regions of the input image that have the highest influence on the network's output. The final step involves summing the weighted gradients over all channels and applying a rectified linear unit (ReLU) activation function:

$$L^c_{GradCAM} = \mathrm{ReLU}\left(\sum_k \alpha^c_k A^k\right)$$

Given the dimensionality reduction that occurs with convolutional operations, the map serves as a low-dimensional visual representation of the regions of the input image that are most important for the given classification task.

## 2.5 Attention and Transformer Modules

Attention is a mechanism used in neural networks to allow the model to focus on different parts of the input for different tasks. It is typically used in sequence-to-sequence models and transformer-based architectures.[89] More specifically, a self-attention mechanism computes a weight for each element in the input sequence, which represents the importance of that element with respect to the other elements in the sequence and the target label.[90]

The self-attention mechanism can be mathematically described as follows. First, a set of query, key, and value vectors are computed for each element in the input sequence. These vectors are typically generated by applying linear transformations to the input element. The dot product of the query vector and the key vector of each element is computed. This produces a matrix of dot products, known as the attention matrix. The attention matrix is then passed through a softmax function, which produces a probability distribution over the elements in the input sequence. This distribution represents the importance of each element in the input sequence with respect to the task at hand. The value vectors for each element in the input sequence are then weighted by the probability distribution and summed to produce a weighted sum of the value vectors. Finally, the weighted sum of the value vectors is passed through a linear transformation to produce the final output of the self-attention mechanism.

Recent focus has been shifting to transformer-based architectures, which extend the self-attention mechanism for classification tasks on sequential data. Instead of using a single matrix of weights to compute the attention weights, transformers use multi-head attention to combine the outputs of the different attention heads. In multi-head attention, the context vectors produced by each attention head are concatenated and passed through another set of learned weights to produce a final output. By employing stacked multi-head attention modules, transformers can capture long-range dependencies in sequences, and large-scale self-supervised pretraining methods such as BERT and GPT have been successfully applied to many NLP tasks.[90,91]

Vision transformers (ViTs) are a recent development in computer vision that are based on the original NLP transformer models.[92] Just as transformer models in NLP can capture long-range dependencies in sequences, ViTs can capture long-range dependencies in images, which is important for tasks such as object recognition and segmentation. There are several strengths of vision transformers that make them suitable for medical imaging tasks. ViTs bring versatile modeling capabilities for heterogeneous data, allowing for effective combination and modeling of various data modalities, such as imaging data, clinical data, and text data. Vision transformer models have demonstrated great scalability with respect to model complexity and larger data size. Self-supervised approaches have shown that transformers can capture long-distance dependencies in images, leading to state-of-the-art performance in major natural image tasks.[93–95] With these advantages, ViTs may have better potential than CNNs to handle spatiotemporal data, given their ability to process sequence input.[96–100] It is expected that by adjusting the components of the transformer model, it can capture spatial information while simultaneously incorporating pixel-level features in the time series dimension.

# CHAPTER 3

# Classifying Time Since Stroke Using Admission Imaging

## 3.1 Overview

The treatment of acute ischemic stroke (AIS) patients is a time-sensitive endeavor, as therapies aim to target areas experiencing ischemia to prevent irreversible damage to brain tissue. The administration of thrombolytics, such as tissue plasminogen activator (tPA), is considered a viable treatment option for AIS patients, however, the window of opportunity for administering these therapies is limited. tPA administration carries risks and negative contraindications. For example, tPA is contraindicated in cases of active bleeding, recent surgery, head trauma, or intracranial bleeding. Additionally, tPA should not be administered to patients who have had a stroke within the previous 3 months, or who have high blood pressure. There is also a risk of hemorrhage and other serious side effects with tPA administration. Therefore, it is important that the decision to administer tPA to AIS patients is made on a case-by-case basis, taking into consideration the patient's medical history, current condition, and imaging results.

The earlier treatments like thrombolysis and EVT are administered, the greater chance a patient can reverse the impacts of ischemia and progression into cell death. However, if thrombolytics are administered too far into the tissue progression, there can be further damage to the neurovasculature. Currently, time since stroke (TSS) serves as a proxy for stroke progression, and clinical workflows rely on this to assess a patient’s eligibility for any stroke treatment. Many patients are unable to access these therapies due to an unknown TSS. To expand the number of patients eligible for these lifesaving treatments, many studies and clinical trials have sought to use other clinical factors to assess eligibility and risk. The most common avenue is imaging, which is currently part of the clinically recommended workflow for stroke treatment. Many clinical trials have used various imaging series and parameters to assess at-risk tissue and determine the most effective treatment windows for patients. Recently, the WAKEUP trial indicated that a neuroradiologist may assess differences in signal between DWI and FLAIR when stroke onset time is unknown to determine a proxy for tissue progression. In fact, in the most recent version of the American Heart Association guidelines for treating acute stroke, for patients with unclear time of symptom onset, MRI can be performed to identify areas with DWI-FLAIR mismatch that could benefit from IV alteplase.

There are many algorithmic approaches that could be used to determine patient eligibility for treatments, with the goal of improving the patient outcomes thanks to increased accuracy and speed of assessment. Machine learning has shown utility among clinical imaging tasks.[101–104] Often, the approach involves extracting features from a region of interest (ROI) within an image, and then using an algorithm for classification. For stroke-related models, stroke lesions have been segmented using a variety of automated methods to reduce neuroradiologist annotation requirements. However, these methods have achieved middling performance, and moreover there may be more information outside of these segmented regions that could inform clinical factors, such as TSS. When considering large image inputs such as 3D MRI or CT scans, deep learning may also be applied; it has recently become an area of prolific medical research across multiple diagnostic domains. Convolutional networks have been widely explored as a deep learning approach for imaging-based tasks because they aggregate pixel information in neighboring regions; intuitively, this works well for medical images that contain a high number of input pixels. However, these algorithms require a large volume of data, and the performance on the training set does not translate to good performance on new data. To assess algorithms for practical use, external validation is necessary to bridge this generalization gap.

In this work, we evaluate and compare three methods to assess TSS: the current clinical standard of measuring DWI-FLAIR mismatch by neuroradiologists, a previously published machine learning method involving radiomic feature extraction from a region of interest, and a novel deep learning algorithm with a soft attention module. We train both models and report the bootstrapped performance on internal validation datasets; we compare this performance to the radiologist-generated mismatch assessments. We use gradient-based visualizations to gain insight into model behavior and inspect these visualizations as weakly supervised localization to the stroke area. Using an external validation set of patient images, we explore the generalizability of the algorithms by varying the amounts of new data and augmentations used for model refinement and retraining. Given the recent research in DWI-FLAIR mismatch and the updated clinical guideline for treating AIS patients, we also explore the performance of our proposed deep learning model refined on DWI-FLAIR mismatch labels and evaluated in 5-fold cross-validation. Our contributions are the following: 1) performance improvement compared to previous state-of-the-art in TSS classification, 2) performance validation on an external clinical dataset, and 3) an exploration of clinical factors that may underlie model performance.

## 3.2 Materials and Methods

### 3.2.1 Datasets

This study utilized a retrospective, observational dataset comprising patients from two institutions. Individuals were included in the cohorts based on the following inclusion criteria: 1) diagnosis with AIS, 2) received pretreatment MRI protocol with DWI, FLAIR, and apparent diffusion coefficient (ADC) series without motion degradation, and 3) known TSS within 24 hours of image acquisition. The internal cohort (Figure 3-2) comprised 417 patients treated from 2011-2019. The second dataset, published by Lee et al., totaled 355 patients, with more extensive exclusion criteria previously described.[16] A summary of the preprocessing pipeline is seen in Figure 3-1. T2w(DWI b0), DWI(DWI b1000), and FLAIR imaging sequences were obtained from the institutional image archiving and communication system for each patient (PACS). All patients had MRIs conducted on a 1.5 T or 3 T echo-planar Siemens MR imaging machine with 12-channel head coils. The DWI b0 sequence was employed as a T2w proxy because it represents the initial step of DWI acquisition with no diffusion attenuation, and the DWI represents the sequence with a b value of 1000. To ensure consistency across both datasets, images were subjected to a preprocessing pipeline.[105] Images had the neck and skull removed via the Brain Extraction Tool and underwent N4-bias field correction.[106,107] The T2 series was registered to the Montreal Neurological Institute (MNI-152) T2 atlas and then served as the fixed volume for co-registration. Finally, series underwent z-score intensity normalization and histogram matching.[108]

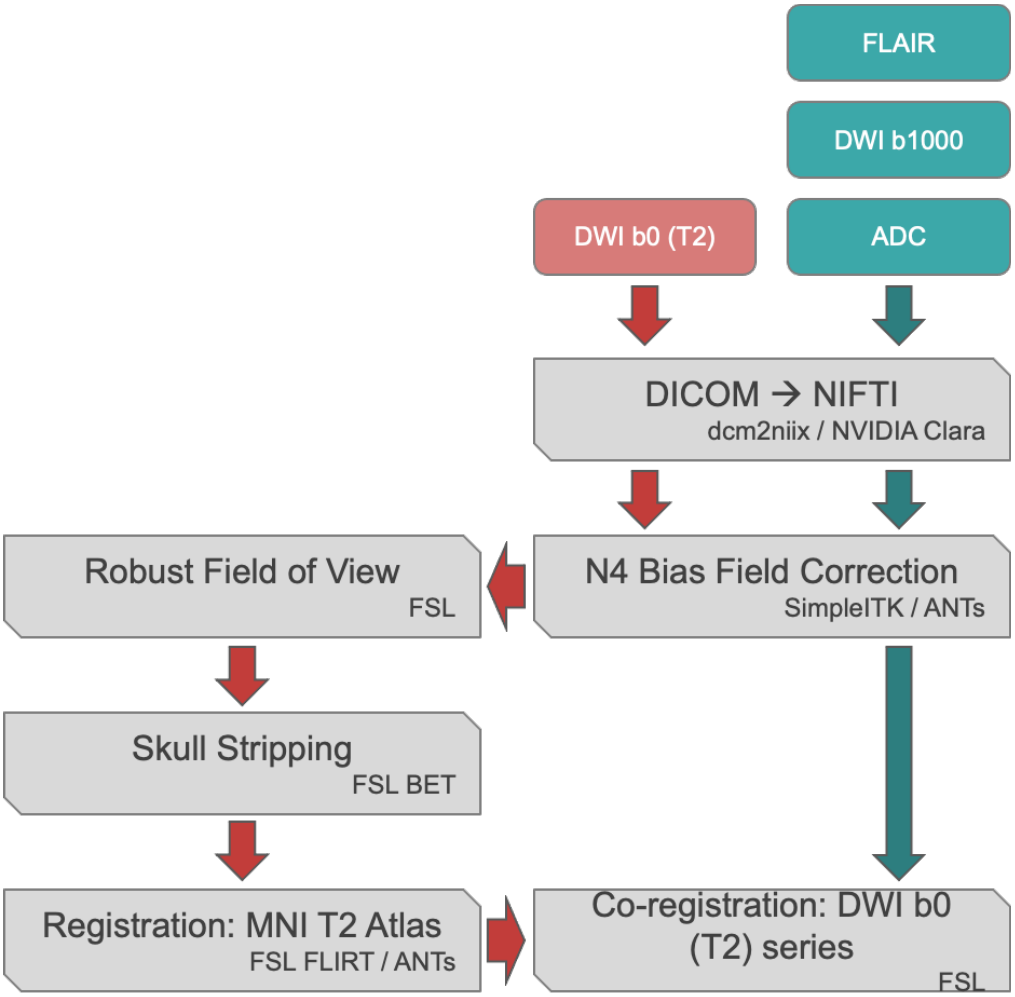


Figure 3-1. Preprocessing Pipeline. The sequences were loaded into our automated preprocessing workflow after picture retrieval. First, all sequences were subjected to N4 bias field correction. The picture series were then reoriented to the T2w MNI-152 atlas. FSL BET was then used to remove the neck and skull. The T2w sequence was registered to a version of the T2w MNI-152 atlas that was enlarged to 224 224 26 using linear interpolation to match the z dimension of the stroke sequences using FSL FLIRT. Following a second run of FSL BET to eliminate any lingering artifacts, the remaining sequences were co-registered to the T2w volume. Finally, the intensity was normalized, and histogram matching was done with the help of a reference research.

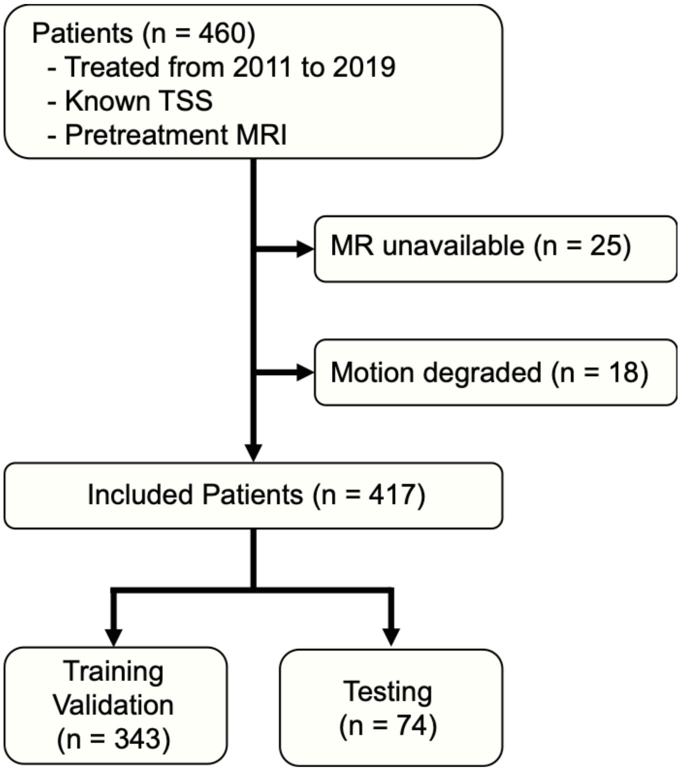


Figure 3-2. Patient flowchart illustrating inclusion criteria for this study. n, number of patients.

### 3.2.2 Classification Task

The primary goal of our model was to classify patients within the thrombolytic treatment window (4.5 hours) based on their pretreatment imaging. To determine patient label, three time points were collected from the electronic health record and image metadata. From the patient record, last known well and symptom onset time were gathered, if they were varied. Additionally, the image study end time was extracted from the PACS image metadata. Patients were binarized into two groups. A positive label was given to individuals who underwent imaging within 4.5 hours of their last known well time and symptom onset, and a negative label was assigned to those who underwent imaging outside this window. Patients at exactly 4.5 hours were deemed outside the window. Any patient with an unknown last known well time, were excluded from the dataset. Each dataset was then divided into development and evaluation sets to be used for training and testing following an 80/20 random stratified split with respect to the target label as well as clinical parameters.

### 3.2.3 DWI-FLAIR Mismatch Assessments

For each patient in the evaluation sets, three neuroradiologists independently assessed mismatch between the DWI and FLAIR series. These labels served as a proxy for TSS, as mismatch indicates that a stroke occurred recently enough such that there are regions that have experienced ischemia (visible on DWI) but are not yet infarcted (visible on FLAIR). Thus, cases with mismatch were assigned labels of $<4$ hours and cases without mismatch were labeled $\geq 4.5$ hours. Radiologists performed these assessments on workstations within the same facility, and they were blinded to model classifications and electronic health record data. Final assessments were determined by a majority vote among the three experts.

### 3.2.4 Deep Learning Model

As seen in Figure 3-3, our deep learning model utilized DWI, ADC, and FLAIR volumes. Model input encompassed three corresponding MRI slices, one from each series, of a single hemisphere of the brain. We designed a multi-slice model that utilizes weight sharing to extract spatial information from neighboring slices. Image series were stacked as channels and fed into a network architecture based on ResNet34, which uses residual "skip connections" between ResBlocks (two convolutional layers each followed by an activation function) to learn image representations.[109] Each convolutional block is modeled after ResNet residual blocks, consisting of the following sequence: convolutional kernel, batch normalization, rectified linear unit activation function, second convolutional kernel and second batch normalization. After feeding the images through a shared convolutional layer and ResBlock, intermediate features of neighboring slices were grouped and fed through five individual weight sharing neighborhood subnetworks. The outputs were fed into a trainable aggregation layer to fuse the features across subnetworks, enabling the model to learn the importance of certain subnetworks over others. Channel and spatial attention modules, which enable learning salience across channels and slices, were attached to the last two ResBlocks to extract multi-scale features.[110] Outputs were then fused with backbone features and fed through a fully connected layer to generate a patient-level TSS classification. Following previous work, we first trained the model to perform stroke detection and then further trained it for our primary task of TSS classification.[105] The model was trained for 100 epochs with early stopping if validation metrics did not improve for five epochs. Binary cross-entropy was used as the loss function. The Adam optimizer was used with weight decay, which optimizes objective functions using adaptive estimates of gradients.[111] Other hyperparameters included a learning rate of 0.0001 and a batch size of 12.

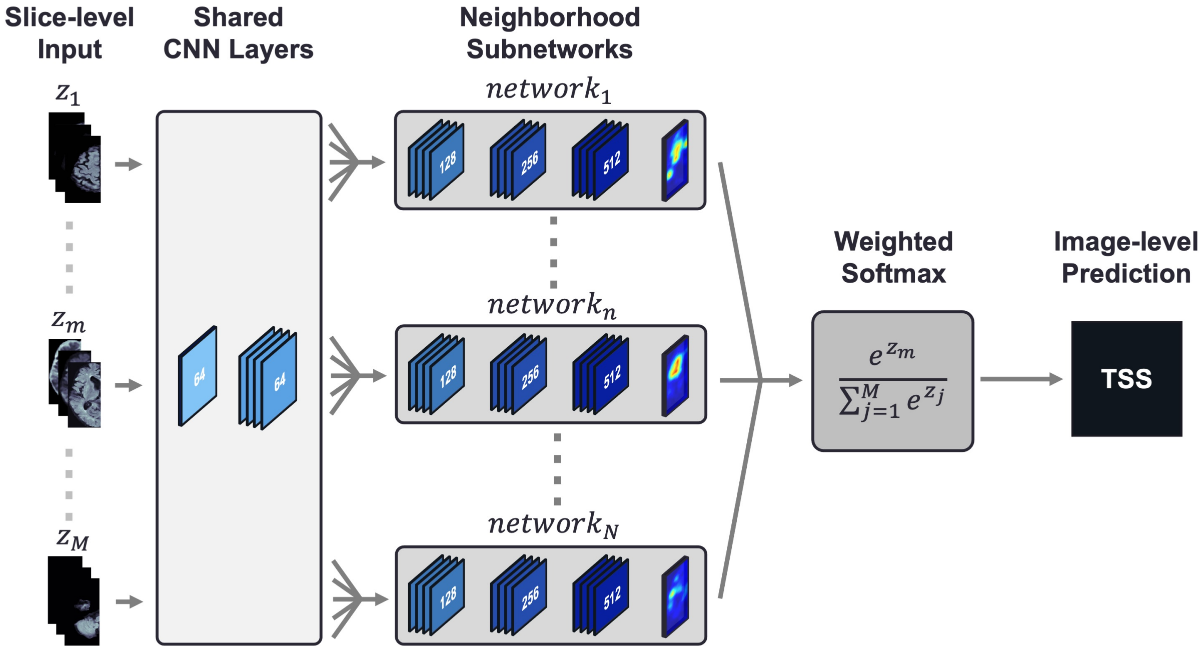


Figure 3-3. Convolutional neural network architecture with shared weights used to classify time since stroke (TSS). The deep learning architecture used DWI, FLAIR, and advanced diffusion coefficient series as input. The model split the volume into slices $z_1, \dots, z_M$ and stacks the image series as channels. Each slice $z_m$ was fed into a shared set of convolutional layers. Intermediate output features from groups of adjacent slices were then propagated through five neighborhood subnetworks $network_1, \dots, network_N$, where weights are shared among the slice neighborhoods. Each subnetwork contained convolutional ResBlocks as well as convolutional attention modules to assist the model with localization. The resulting outputs from each subnetwork are aggregated using a weighted softmax function to generate a TSS classification for the image. CNN, convolutional neural network.

### 3.2.5 Training and Evaluation

To train models, the development sets were split into five folds for cross-validated hyperparameter tuning, and the chosen parameters were used to train a model on the entire development set. Training was run in replication across ten random seeds. These trained models were run on the evaluation data, and metrics were computed and aggregated to generate confidence intervals. Metrics included sensitivity, specificity, accuracy, and receiving-operator characteristic area-under-the-curve (ROC-AUC). The AUC analysis threshold was determined utilizing Youden's Index on the training data.

These statistics were compared to those calculated using radiologist-assessed mismatch as a surrogate for TSS.

## 3.3 Experiments

### 3.3.1 Baseline Comparison

The proposed method, a weight sharing deep learning model, was compared to two other methods: neuroradiologist assessed DWI-FLAIR mismatch and a threshold-based radiomics machine learning (ML) model.[16] A summary of these models can be seen in Table 3-1. The ML method began with infarct segmentation via thresholding of the normalized patient ADC map using a previously determined absolute value. These regions of interest (ROIs) were used as the basis for radiomic feature extraction, using DWI and FLAIR series and a FLAIR-ADC ratio map. These features were subjected to univariate *t*-tests to select the most informative features that were used as input for the following algorithms: random forest classification, support vector machine classification, and logistic regression.

Table 3-1. Summary of proposed model and baseline comparisons. ML, machine learning; DL, deep learning; DWI, diffusion-weighted imaging; FLAIR, fluid-attenuated inversion recovery.

| | **Model Name** | **Type** | **Description** |
|---|---|---|---|
| **Proposed** | DL Model | Deep Learning | Neighborhood network utilizing weight sharing across slices and attention networks. |
| **Baseline Comparisons** | ML Model | Machine Learning | Radiomic pipeline using thresholding to generate infarct ROI, extraction of features from FLAIR regions, and statistical t-tests to select the most relevant features for classification algorithms. |
| | DWI-FLAIR Mismatch | Neuroradiologist Assessment | Majority vote from three neuroradiologists to assess signal mismatch between DWI b1000 and FLAIR series. |

### 3.3.2 Incorporation of External Data

To evaluate the clinical utility of the machine learning and deep learning algorithms, we considered the following experiments: (1) training and testing on data from the same institution, (2) training and testing on data from different institutions, (3) training on a combination of data from both institutions. For the last experiment, we considered two different types of experiments, whereby we either added incremental amounts of external data, which increased the size of the training set, or we replaced increasing amounts of training samples from the internal dataset with training samples from the external set. For each experiment, we report the performance on each test set.

### 3.3.3 Visualizations for Interpretability

We implemented three visualization methods used for model interpretability: occlusion sensitivity, class activation maps (CAMs), and integrated gradients. Each method provides unique feature importance maps for a given input. Occlusion sensitivity involves perturbing patches of input images and calculating the effect each perturbation has on the target class classification.[112] To generate CAMs, an activation map is computed using the output from the last convolutional layer of the network; this serves to identify regions of the image that provide the greatest discrimination for the correct label. Finally, outputs are backpropagated through the network to create pixelwise maps of network gradients for individual input images. We also visualize the gradients generated via guided backpropagation.[113]

## 3.4 Results

### 3.4.1 Cohort Characteristics

Our study utilized two datasets. Of the internal set, 222 patients had a TSS under 4.5 hours, while the remaining 195 patients had a TSS over 4.5 hours. For external evaluation, we utilized a dataset totaling 355 patients, of which 182 underwent MRI within 4.5 hours of onset and 173 after 4.5 hours of onset.[16] Clinical characteristics of these datasets are summarized in Table 3-2.

Table 3-2. Distribution of clinical demographics for the datasets: internal (left) and external (right, reproduced from Lee et al.[16]). Represented as n (%) or median (interquartile range). n, number of patients; NIHSS, NIH Stroke Scale.

| | **Internal Dataset** | | **External Dataset** | |
|---|---|---|---|---|
| | Train (n = 343) | Test (n = 74) | Train (n = 299) | Test (n = 56) |
| **Age (years)** | 70 (55-80) | 68 (57-79) | 63 (55–73) | 67 (55–71) |
| **Female** | 176 (52%) | 46 (56%) | 86 (34%) | 20 (36%) |
| **Admission NIHSS** | 8 (4-16) | 6.5 (2-18) | 4 (2–10) | 5 (2–12) |
| **Onset to MRI (minutes)** | 210 (105-683) | 230 (107-661) | 270 (152–715) | 240 (142–448) |
| **Within 4.5-h Window (%)** | 185 (54%) | 37 (50%) | 153 (58%) | 24 (43%) |

### 3.4.2 Baseline Comparison

The majority radiologist assessment of mismatch for the internal evaluation set, when compared to EHR-derived TSS, had low sensitivity (0.622) with high specificity (0.865). The aggregate assessment achieved higher accuracy (0.743) compared to the average accuracy of any individual radiologist (0.658). The mismatch assessments for the external evaluation set had higher sensitivity (0.743) while maintaining a high specificity (0.800).

The performance results of the DL and ML methods trained on the internal, external, and combination training sets are summarized in Table 3-3. As a result of the thresholding technique applied by the ML method, 204 patients out of 417 patients from the internal dataset had an extracted ROI, and 343 out of 355 patients from the external dataset had an extracted ROI. Additionally, the ML model selected different radiomics features depending on the dataset. In applying univariate *t*-tests to 89 radiomics features, 37 features were selected for the internal training set, and 35 were selected for the external training set with only seven features overlapping. When compared to the radiologist assessments, both the ML and DL models had higher sensitivity, though lower specificity. The average rate of agreement between the DL classifications and radiologist assessments was 0.411 (0.01), indicating a similar level of agreement as among the radiologists for the internal evaluation set.

The internally trained model achieved an AUC of 0.768 (0.03), with an accuracy of 0.726 (0.02), a sensitivity of 0.712 (0.08), and a specificity of 0.741 (0.09). On the external dataset, the model achieved an AUC of 0.737 (0.03), an accuracy of 0.724

(0.04), a sensitivity of 0.757 (0.04), and a specificity of 0.679 (0.07). When trained using both internal and external data, performance on both evaluation sets improved, achieving an average AUC of 0.840 (0.03) on the internal test dataset and 0.814 (0.01) on the external test dataset. This aggregate model yielded an average accuracy of 0.794 (0.04), surpassing the accuracy of the aggregate neuroradiologist assessment.

Table 3-3. Performance Metrics for TSS Classification. Performance metrics for the deep learning (DL) machine learning (ML)[16] models trained on the internal, external, and combination training sets, and tested on each separate test set. All the data represent mean ± standard deviation. n, number of image studies used in the training set.

| | **Train Set** | **Test Set** | **AUC** | **Accuracy** | **Sensitivity** | **Specificity** |
|---|---|---|---|---|---|---|
| **Deep Learning Model** | UCLA (n = 340) | UCLA | 0.768±0.03 | 0.726±0.02 | 0.712±0.08 | 0.741±0.09 |
| | | External | 0.737±0.03 | 0.724±0.04 | 0.757±0.04 | 0.679±0.07 |
| | External (n = 299) | UCLA | 0.732±0.02 | 0.707±0.03 | 0.716±0.09 | 0.687±0.08 |
| | | External | 0.772±0.02 | 0.767±0.03 | 0.850±0.08 | 0.648±0.09 |
| | Both (n = 639) | UCLA | 0.840±0.03 | 0.789±0.04 | 0.777±0.06 | 0.802±0.07 |
| | | External | 0.814±0.01 | 0.800±0.04 | 0.850±0.08 | 0.727±0.08 |
| **Machine Learning Model** | UCLA (n = 164) | UCLA | 0.730±0.07 | 0.675±0.07 | 0.405±0.07 | 0.811±0.08 |
| | | External | 0.680±0.15 | 0.653±0.10 | 0.714±0.15 | 0.500±0.13 |
| | External (n = 284) | UCLA | 0.698±0.08 | 0.625±0.09 | 0.297±0.08 | 0.865±0.10 |
| | | External | 0.780±0.05 | 0.735±0.05 | 0.657±0.05 | 0.800±0.08 |
| | Both (n = 448) | UCLA | 0.783±0.03 | 0.750±0.04 | 0.405±0.03 | 0.892±0.03 |
| | | External | 0.795±0.03 | 0.735±0.03 | 0.686±0.03 | 0.750±0.04 |

### 3.4.3 Incorporation of External Data

The impact of external training data on model AUC is summarized in Figure 3-4. The proposed deep-learning model achieved lower performance on the external evaluation set when no refinement is performed; however, the model achieved comparable performance for both evaluation sets when as few as 40 external samples were introduced into training, and better performance when 160 external patients were used. Intuitively, this corroborates the idea that deep learning algorithms achieve higher performance when trained on larger amounts of data and is illustrated in the second panel of Figure 3-4, where the performance on both cohorts did not improve with the replacement of internal data with external data.

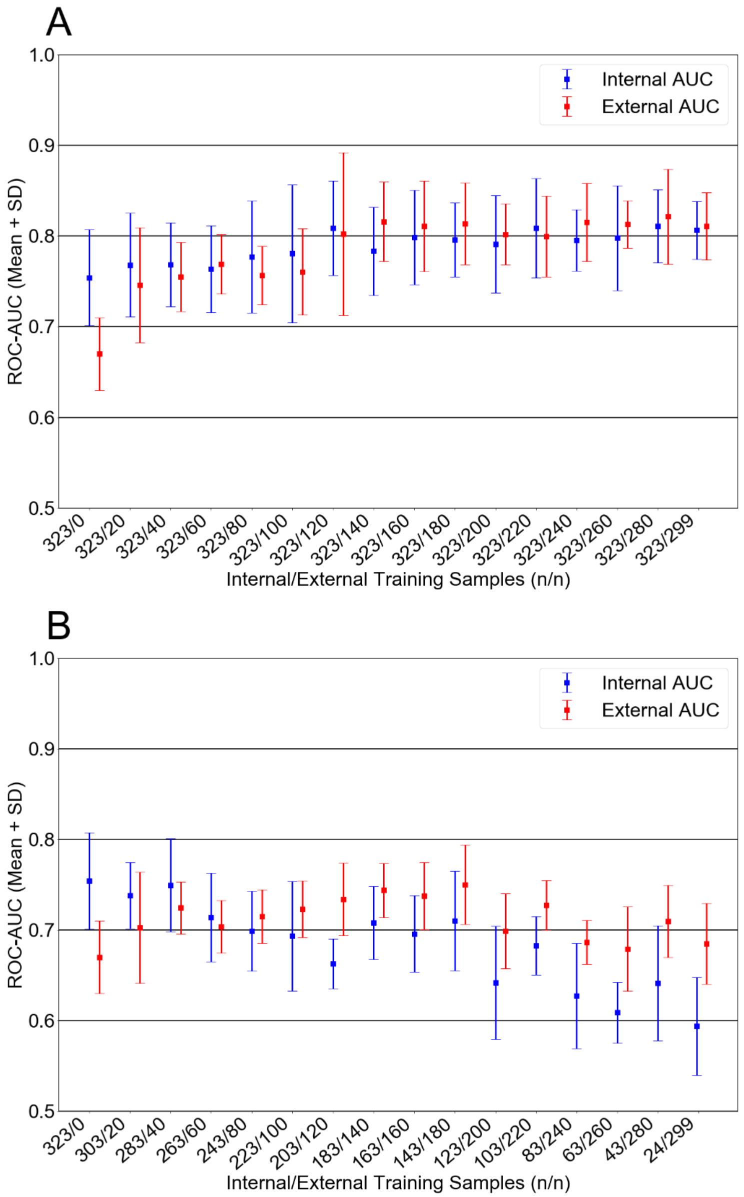
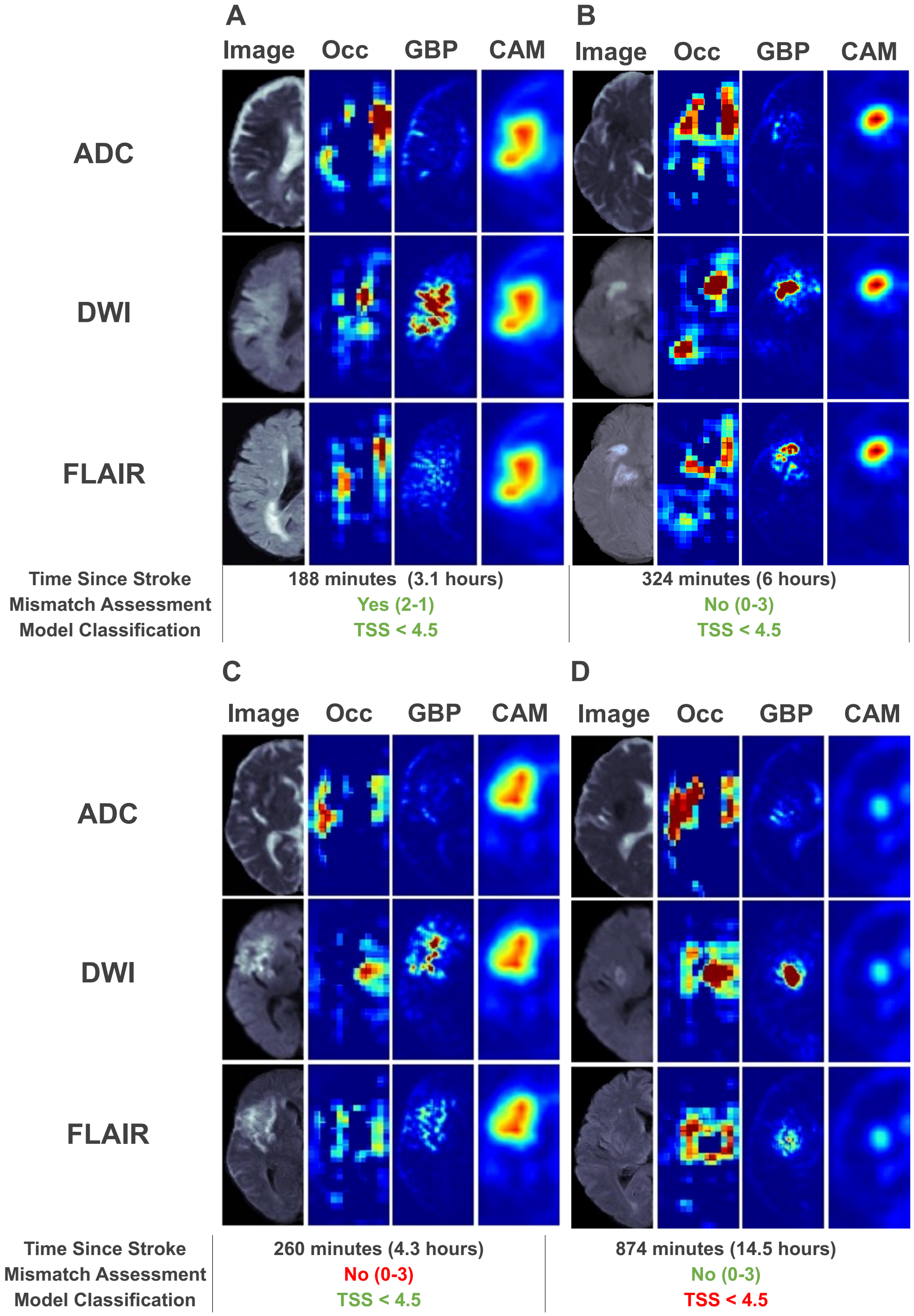


Figure 3-4. Receiving-operator characteristic area under curve (ROC-AUC) of models with varying amounts of external training data, both when added to (A) or replacing (B) samples in the internal training set. Performance on both internal and external test sets are reported, in blue and pink, respectively, with 95% confidence intervals. Numbers on the x-axis indicate the number of internal/external samples used for training. SD, standard deviation.

### 3.4.4 Visualizations for Interpretability

Visualizations were generated to reveal image regions on which the model focused. Four representative cases are shown in Figure 3-5. Figure 3-5A shows a patient with DWI-FLAIR mismatch that is within the treatment window that the model correctly classifies. The model does not focus solely on areas of high imaging signal contrast, suggesting that there is valuable signal in less prominent regions. Figure 3-5B shows a patient outside the treatment window and without signal mismatch. The DL model also correctly classified this image, and gradient visualizations indicate that the model was able to detect tissue changes on both FLAIR and DWI sequences. Figure 3-5C shows a case with a stroke onset time just under the 4.5-hour threshold that the neuroradiologists agreed contained no signal mismatch. In this instance, the model's classification was inside the window. The gradients and CAM localize to the stroke lesion, while the occlusion method shows that areas outside the stroke volume were most salient to the classification. Finally, Figure 3-5D shows a case well over the window for tPA. The radiologists agree that there is no mismatch, yet our model classified this case as within the window. The occlusion-based visualization shows that the model is unable to localize the stroke on either the ADC or FLAIR series. The CAM highlights that there is not a strong region of activation. Notably, the signal intensity of the stroke is relatively low, which may account for the model's behavior. Changes to the preprocessing protocol may better distinguish the lesion and improve model performance for such cases.



| | | |
|---|---|---|
| Time Since Stroke | 188 minutes (3.1 hours) | 324 minutes (6 hours) |
| Mismatch Assessment | Yes (2-1) | No (0-3) |
| Model Classification | TSS < 4.5 | TSS < 4.5 |

| | | |
|---|---|---|
| Time Since Stroke | 260 minutes (4.3 hours) | 874 minutes (14.5 hours) |
| Mismatch Assessment | No (0-3) | No (0-3) |
| Model Classification | TSS < 4.5 | TSS < 4.5 |

Figure 3-5. Deep learning visualizations for four patients (a-d) generated from a model trained to classify patient images as within or outside the thrombolytic window. For each patient, three visualizations were generated: occlusion, guided backpropagation (GBP) and class activation maps (CAM). The table below lists the TSS, age, radiologist-assessed mismatch, and classification yielded by the model. ADC, apparent diffusion coefficient.

## 3.5 Discussion

The relationship between TSS and imaging features has been studied extensively; nevertheless, it remains unclear which signal patterns accurately capture the time course of ischemic tissue. DWI-FLAIR mismatch is one eligibility measure for thrombolysis in the most recent treatment guidelines. In our study, the inter-reader agreement for DWI-FLAIR mismatch aligns with that found in previous studies.[114,115] Despite an average of 12 years' experience among the neuroradiologists, variability among their assessments implies that a patient's treatment options and therefore potential outcomes are reader dependent. Using TSS as the eligibility metric ("time clock"), the radiologist assessments identified 62% (23/37) of evaluation set patients who were within the 4.5-hour window of stroke onset. The DL model, by contrast, identified 76% (29/37) of patients within the window of eligibility. The lack of agreement among radiologists for the DWI-FLAIR mismatch assessments, along with the discrepancy between "tissue clock" and "time clock," illustrate the need for more research into this relationship.

Our study reports the average performance of ten replicates and evaluates two methods on the same datasets, revealing insight into the generalizability of these algorithms. When evaluated on external data, our model was able to achieve higher performance than the current state-of-the-art ML method. This could be due to a few reasons, including inclusion of potentially informative penumbral brain tissue, and more inclusive feature selection. Previous models, including the ML model evaluated in this study, have utilized segmentation models that identify the stroke region of interest from DWI.[16,116] When compared to expert segmentation, performance of these methods has been moderate, primarily under-segmenting the stroke lesion. Moreover, these methods fail to incorporate penumbral regions that could inform vascular stroke progression status.[117] In contrast, the DL model utilizes the ipsilateral brain hemisphere, thereby including information from both the ischemic core and the penumbral tissue outside diffusion-weighted lesions that may provide key insights into the tissue clock. Additionally, ROI extraction methods such as thresholding may exclude cases when a lesion is not identified, which is not a limitation in our brain hemisphere based approached. For the ML model, the selection of statistically significant radiomics features may induce bias into the model that favors the training data. Our DL model, in contrast, distills features from the entire input iteratively. The DL model also carries advantages over the previous deep learning models,[105] likely as the model uses attention modules to focus on pertinent channel and brain regions as well as the integration of information from neighboring slices. Despite these advantages, the DL model does have some drawbacks. The model has more input parameters than a standard radiomics-based ML model, requiring larger datasets and more computational time. This computation time is negligible for inference but should be considered for updating models when training. Additionally, there was still a performance gap for the DL model between the internal and external datasets, which motivated our external evaluation experiments. Aggregating the training datasets improves performance on both evaluation cohorts, indicating that DL classifiers improve synergistically when exposed to diverse training data.

Our study has several areas of potential improvement. While our dataset comprises the largest used for TSS classification from two cohorts, it cannot fully represent all patients seen in practice. Our preprocessing ideally minimizes dataset variation, but further analysis is needed to assess applicability to cohorts from other institutions. With recent advances in deep learning for medical image processing, multi-site harmonization could further minimize batch effects and scanner variations across sites, provided there is sufficient data to train these models.[118] Second, we were only able to evaluate this model for a small set of patients for which the radiologists assessed mismatch. A common bottleneck when using machine learning for medical image tasks is that acquiring labels (e.g., having multiple neuroradiologists assess images for DWI-FLAIR mismatch) is labor-intensive and may not be feasible on a large scale. Third, TSS is not a perfect surrogate biomarker as it does not always correlate to underlying tissue changes informing ischemia.[13] Nonetheless, given the low inter-reader agreement of DWI-FLAIR mismatch, a TSS classification using an automated method may aid the radiologist in clinical decision-making. This model predicted a binarized stroke onset due to the variations in last known well and symptom onset time; with more discrete TSS determinations, predictions could be made on a continuous or ordinal scale to prove more granular patient classifications.

## 3.6 Conclusion

Our proposed DL model allowed classification of TSS based on MR images and achieved higher AUC than the current best-in-class ML model when external data was introduced, showing a more robust automated algorithm to determine stroke onset time. The results of this study indicate that a small amount of external data can improve generalized performance across patients from multiple institutions. These findings support the future study of the implementation of a deep learning algorithm for clinical decision support in the setting of AIS treatment. This highlights the need for more accurate and reliable methods for determining the optimal treatment window for AIS patients, and further research is needed to better understand the underlying conditions that lead to optimal treatment outcomes.

# CHAPTER 4

# Combining Proxy Information with Semi-supervised Learning for Thrombolytic Candidate Identification

## 4.1 Overview

The age of a stroke lesion, or the length of time that a particular area of the brain has been affected by ischemia, is an important consideration in the management of stroke. In the acute phase of ischemia, the brain's response is primarily focused on preserving cellular function. This phase lasts for several hours after the onset of symptoms, during which time interventions such as thrombolytic therapy and endovascular therapy may be effective in restoring blood flow and preserving brain function. The ischemic penumbra is the area around the core of the infarction where brain cells are still alive but have not yet experienced irreversible cell death infarction. However, as the ischemia persists and the infarction becomes more established, the brain's response shifts towards a process of cellular death. The infarct core is characterized by the death of brain cells and likely irreversible tissue damage. The mechanistic underpinnings that correlate tissue progression, cell death, and time clock are an area of active study. While tPA is generally considered for patients with a known onset within 4.5 hours, recent clinical trials have demonstrated that patients with signal intensity differences between diffusion-weighted imaging (DWI) and fluid-attenuated inversion recovery (FLAIR) sequences can benefit from thrombolytic therapy.[33,82,115,116,119] However, this intensity mismatch is subject to high inter-reader variability, and current clinical guidelines still primarily rely on known stroke onset times.[14,114,120] An automated method of detecting this signal mismatch could provide utility for neuroradiologists when assessing patients for AIS treatments.

Deep learning techniques have demonstrated great potential in the field of medical imaging diagnosis. However, one of the primary challenges that hinder the widespread application of this approach is the acquisition of high-quality annotations. The process of acquiring annotations, which are vital for training models, necessitate assessments from multiple domain experts and is both costly and labor-intensive. To address this challenge, various approaches have been proposed in the literature to tackle classification tasks with manually intensive labels, particularly when the sample size is small.[86] Semi-supervised learning (SSL) is one such approach that has gained popularity across multiple medical domains. This approach involves assigning pseudolabels to unlabeled datasets and incorporating them into the training dataset.[121,122] These approaches can leverage unlabeled data to improve the performance of models even when the amount of labeled data is limited. However, it must be noted that this approach carries the risk of confirmation bias, which occurs when the model incorrectly assigns pseudolabels and influences the loss function to optimize to the incorrect minimum.

There are also machine learning paradigms in which there are annotations that are weakly predictive of the target label; in these instances, weakly supervised learning (WSL) can be utilized to evaluate samples with unverified labels in relation to their proximity to the fully annotated label. [123] This proximity is commonly measured via comparison of the features of the samples to the features of the fully annotated samples, with samples subsequently being assigned a label based on their degree of similarity to the fully annotated samples. This approach can be applied to various tasks such as object detection, image classification, and semantic segmentation. In the realm of medical imaging, WSL has been utilized for object detection, and semantic segmentation, both to generate localizations or region masks while only using image-level labels or bounding boxes rather than pixel-level annotations. WSL similarly has the risk of annotation errors, particularly when utilizing imprecise annotations. Additionally, the performance of models trained via WSL may not match that of models trained via supervised learning utilizing fully annotated data. Despite these limitations, WSL has demonstrated promise in a variety of medical imaging applications and may prove to be a valuable tool in scenarios where fully annotated data is scarce.

In this study, we sought to detect DWI-FLAIR mismatch in an automated fashion, and we compared this to assessments done by three expert neuroradiologists. Our approach involved training a deep learning model on MRI to classify tissue clock and leveraging time clock as a weak proxy label to supplement training in an SSL framework. We evaluate our deep learning model by testing it on an unseen dataset from an external institution. In total, our proposed framework was able to improve detection of DWI-FLAIR mismatch, achieving a top ROC-AUC of 74.30%. From a technical perspective, this study demonstrated that incorporating clinical proxy information into SSL can improve model optimization by increasing the fidelity of unlabeled samples included in the training process.

## 4.2 Materials and Methods

### 4.2.1 Dataset and Problem Formulation

Our dataset contained two distinct labels - the time clock (TSS) and the tissue clock (DWI-FLAIR mismatch). On one hand, the TSS label was available for all patients in both datasets, however, it does not provide an entirely accurate representation of the underlying tissue changes that lead to ischemic tissue that has not yet experienced infarction. In contrast, the DWI-FLAIR mismatch label is a more precise representation of the salvageable tissue targeted by thrombolytic treatments. However, the process of generating these labels across the entire dataset is quite labor-intensive and there is a degree of variability that can occur between different readers. This makes it a challenge to ensure consistency and accuracy when generating these labels.

We will consider $X$ as the total set of samples that are available for our semi-supervised framework. Within this, there are two subsets: $X_F$, which are fully labeled samples, and $X_W$, which are samples that only have the weak label. For this study, we can assume that each sample in $X_F$ has both the weak proxy and manually acquired full label. Thus, we will consider TSS as a weak proxy label, and DWI-FLAIR mismatch as a full label. Each fully labeled patient can fall into one of four categories, as illustrated in Figure 4-1. The goal of the model is to detect DWI-FLAIR mismatch, i.e., classify a set of patient images as containing mismatch or not. Accordingly, we formulate this as a binary classification task to classify images as mismatch positive (1) or mismatch negative (0).

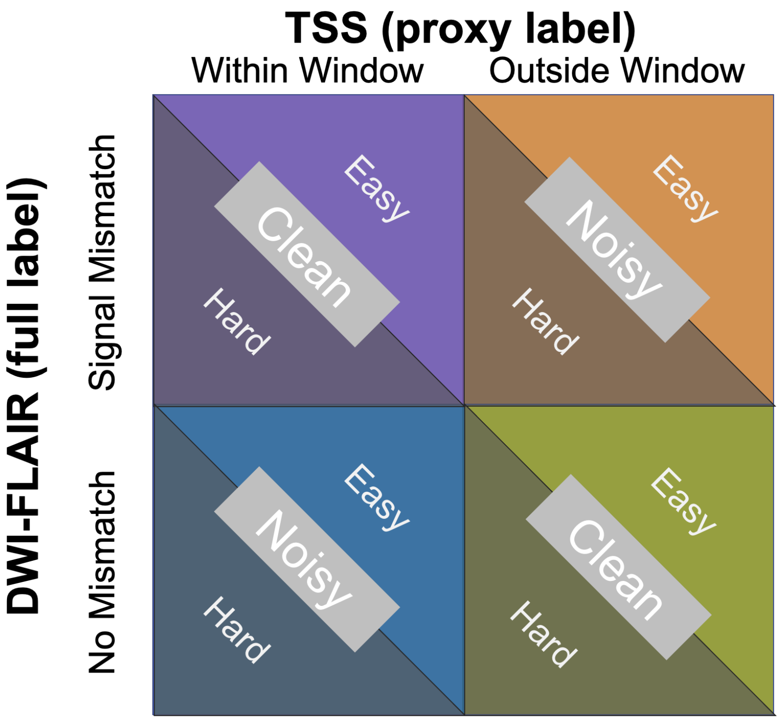


Figure 4-1. Categories of fully labeled dataset. Patients can be categorized as Clean, where proxy and true label align, or Noisy, where the two labels do not match. In addition, within each of these four categories, a patient can also be categorized as easy or hard based on the level of inter-reader agreement for the DWI-FLAIR mismatch assessment.

### 4.2.2 Classification Models

Our framework utilized two models that were trained in a decoupled fashion. The first was a deep learning model $g_\theta \to \mathbb{R}^d \to P$ that served as both a feature extractor and target label classifier, where $\mathbb{R}^d$ represents extracted features and $P$ indicates the final binary classification for the target variable. The convolutional backbone was based on ResNet-18, as that was determined to be the optimal architecture from previous stroke MRI classification studies.[105] The second model was a discriminator $D_S \to L$ that classified samples based on the relationship between their weak proxy and target labels. That is, the model determined a sample's likelihood that the weak proxy matched the target. The model took features extracted from $g_\theta$ as input and computed cosine similarity to those features extracted for each of the four data categories.

### 4.2.3 SSL Framework

Our framework consisted of two stages. In the first stage, we trained our deep learning model. Once the deep learning architecture was sufficiently trained, we froze the network for it to serve as a feature extractor. In the second phase, we extracted features for both $X_F$ and $X_W$. Using these extracted features, we trained $D_S$ to classify samples in $X_W$based on their cosine similarity to samples in $X_F$. We then used these classifications to infer labels for samples in $X_W$for which there was high confidence that the pseudolabel was correct. To account for the risk of incorrect pseudolabeling, we included two confidence criteria: 1) Does the model output for the unlabeled sample match that of the cluster, and 2) Does the proxy (TSS) for the unlabeled sample match that of the cluster? In this respect, we only included samples for which the weak proxy label for $X_W^i$ matched the classification determined by $D_S$. These high confidence samples were then incorporated into the training set, and the process iterates again starting at stage 1. Our framework algorithm is depicted in Figure 4-2.

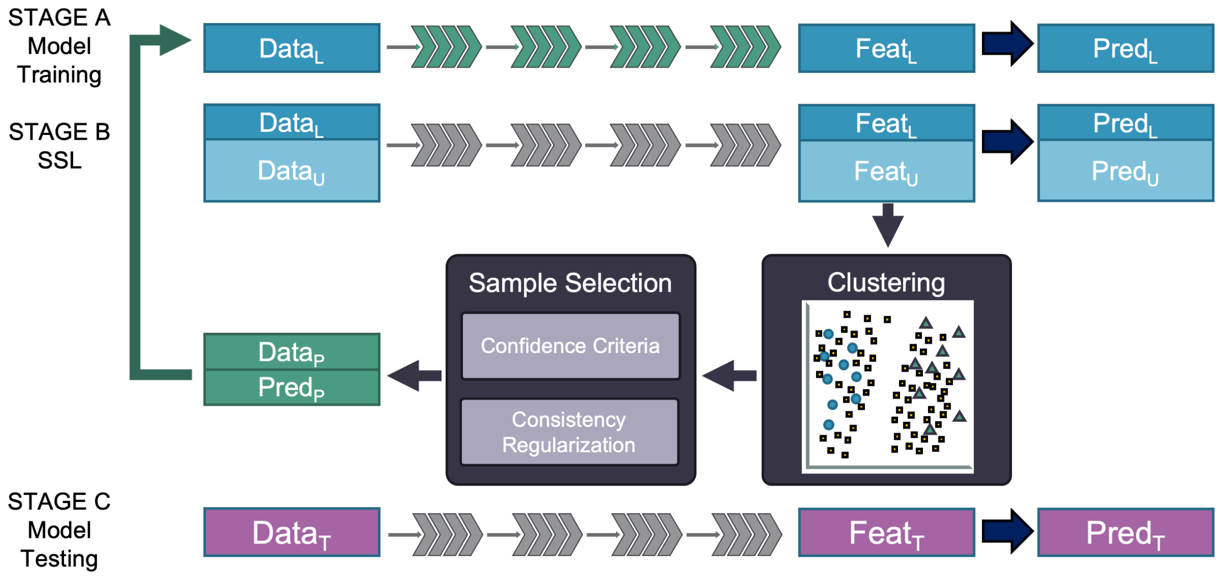


Figure 4-2. Outline of the SSL framework. In Stage A, the model is trained on the fully labeled data Data$_\mathrm{L}$ to generate predictions Pred$_\mathrm{L}$. In Stage B, the model generates features Feat$_\mathrm{L}$ and Feat$_\mathrm{U}$ for the labeled and unlabeled Data$_\mathrm{L}$ and Data$_\mathrm{U}$, respectively. These features are then fed into a clustering and sample selection pipeline to generate a set of pseudolabeled data (Data$_\mathrm{P}$ and Pred$_\mathrm{L}$). These pseudolabels are then used to repeat Stage A, model training, on the augmented dataset. After a set of iterations between Stage A and Stage B, the model enters Stage C, model evaluation. Here, the final trained model is used to generate predictions Pred$_\mathrm{T}$ on the external test set Data$_\mathrm{T}$.

## 4.3 Experiments

### 4.3.1 Baseline Comparisons

Assessments were performed by three radiologists with 11, 16, and 29 years of experience. Fleiss' kappa was calculated to measure the level of agreement among the three radiologist assessments. We also report the inter-label agreement between the clinically recorded TSS and the DWI-FLAIR mismatch. We also sought to assess the concordance of a TSS classification model with one that evaluates DWI-FLAIR mismatch. Accordingly, we trained a DL model using the majority radiologist assessment of mismatch as training labels, and we tested this model on the test set from the other institution.

### 4.3.2 Semi-Supervised Learning

For all experiments, we used the same initial training set $X_F$, comprising 72 patients from UCLA with both proxy and target labels. A total of 345 patients $X_W$, all from the UCLA Stroke dataset, were used as potential pseudo-labeled data in the semi-supervised stage of training. All models and experiments were tested on a set of patients $X_E$, which comprised 56 patients from Asan Medical Center. For the deep learning model $g_\theta$, hyperparameters were tuned in accordance with previous experiments for this architecture.[105,124]

We hypothesize that proxy information can supplement the semi-supervised framework training process in the following ways: using proxy labels to select weakly labeled samples for inclusion in the training set and incorporating samples of varying noisiness at different training iterations. We compared our proposed framework to a baseline, fully supervised network trained on fully supervised samples alone. We also compare to the current state-of-the-art in SSL, a method performed by Berthelot et al. that involves consistency regularization on augmented samples.[125] To test the effects of our proposed techniques, we also completed ablation studies for each methodological adaptation implemented in our framework. The primary metric used to perform this study was receiver operator characteristic area under the curve (ROC-AUC), though we also report detection sensitivity and specificity. Each model was run ten times to report mean performance across metrics.

## 4.4 Results

### 4.4.1 Mismatch Assessments

Among the 130 test patients within the internal (UCLA) and external (AMC) datasets, 37.8% (28/74) and 55% (31/54) of patients were found to have DWI-FLAIR mismatch, respectively. Inter-reader agreement among the radiologists as pairs and collectively are summarized in Table 4-1. Fleiss' was 0.460 for the internal dataset and 0.575 for the external dataset, representing moderate levels of agreement. Performance of the human readers, compared to time clock assessment, is illustrated in Table 4-2 for the internal and external datasets.

Table 4-1. Inter-rater Agreement for the Internal and External Datasets. Calculated using Cohen's kappa, except for All Radiologists, which is computed Fleiss' kappa. Rad, radiologist; Agg, aggregate reading representing majority vote by three radiologists; TSS, time since stroke.

| **Site** | **Rad** | **Agreement ($\varkappa$)** |
|---|---|---|
| **Internal** | Rad 1 – Rad 2 | 0.3677 |
| | Rad 1 – Rad 3 | 0.5264 |
| | Rad 2 – Rad 3 | 0.4879 |
| | All Radiologists | 0.4600 |
| | Agg – TSS | 0.4430 |
| **External** | Rad 1 – Rad 2 | 0.5893 |
| | Rad 1 – Rad 3 | 0.6306 |
| | Rad 2 – Rad 3 | 0.5086 |
| | All Radiologists | 0.5755 |
| | Agg – TSS | 0.5208 |

Table 4-2. Radiologist Performance Metrics. Performance metrics for individual and aggregate radiologist assessments for the internal and external datasets. Rad, Individual Radiologist; Agg, Aggregate reading by radiologists; n, number of patients assessed.

| **Site** | **Reader** | **Mismatch Positive** | **Accuracy** | **Sensitivity** | **Specificity** |
|---|---|---|---|---|---|
| **Internal (n=74)** | Rad 1 | 38 | 0.608 | 0.568 | 0.649 |
| | Rad 2 | 19 | 0.676 | 0.432 | 0.919 |
| | Rad 3 | 28 | 0.689 | 0.541 | 0.838 |
| | Agg | 28 | 0.743 | 0.622 | 0.865 |
| **External (n=56)** | Rad 1 | 31 | 0.691 | 0.686 | 0.700 |
| | Rad 2 | 35 | 0.836 | 0.857 | 0.800 |
| | Rad 3 | 24 | 0.636 | 0.543 | 0.750 |
| | Agg | 31 | 0.764 | 0.743 | 0.800 |

### 4.4.2 Baseline Model

The results of retraining the model and evaluating on internal cross validation, using the mismatch evaluations as labels, are summarized in Table 4-3 and Figure 4-3.

Table 4-3. Summary statistics for 5-fold cross-validation performance of model refined to classify DWI-FLAIR mismatch.

| | Average (SD) |
|---|---|
| AUC | 0.754 (0.07) |
| Sensitivity | 0.651 (0.12) |
| Specificity | 0.967 (0.07) |
| Accuracy | 0.769 (0.07) |

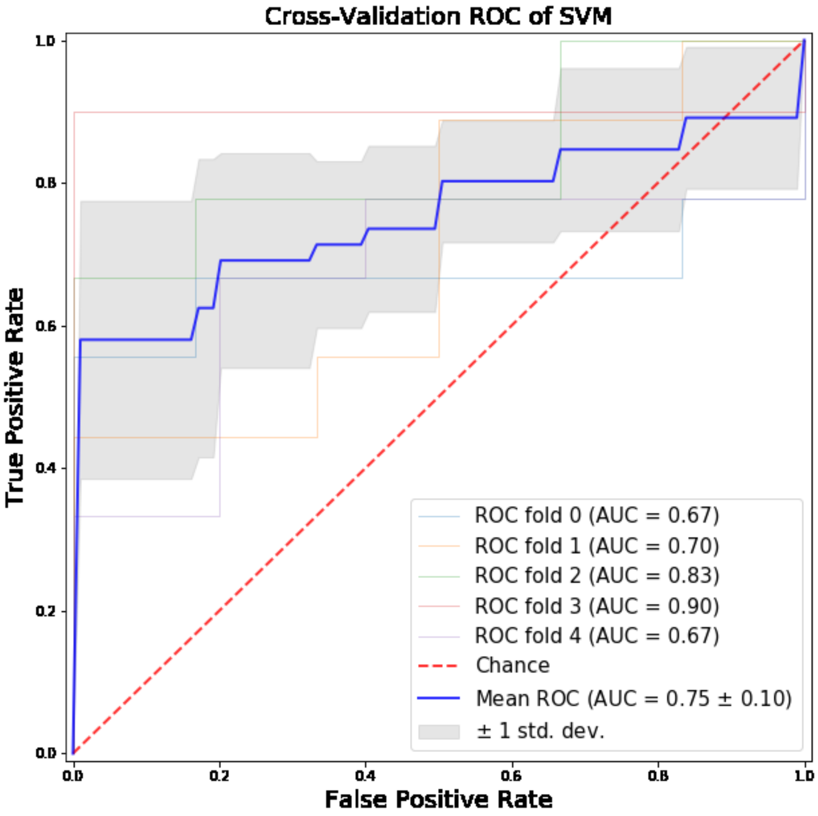


Figure 4-3. ROC curve with confidence interval for 5-fold cross-validation performance of the model refined to classify the DWI-FLAIR mismatch.

### 4.4.3 Semi-Supervised Method

Our experimental results are summarized in Figure 4-4 and Table 4-4. When tested on an unseen external validation dataset, our semi-supervised framework demonstrated outstanding performance, achieving an average ROC-AUC of 74.30±1.9%. This model not only outperformed the current state-of-the-art in SSL for the DWI-FLAIR mismatch detection task, but it also showed the lowest variance when run in replicate. Our ablation experiments further illustrated that both methods used to leverage proxy information were effective in enhancing model performance.

Table 4-4. Performance metrics across experiments and our proposed framework.

| Model | ROC-AUC | Sensitivity | Specificity |
|---|---|---|---|
| Fully Supervised | 0.6436±0.066 | 0.6415±0.298 | 0.4516±0.277 |
| Baseline SSL | 0.6711±0.035 | 0.6097±0.145 | 0.7045±0.142 |
| + Noise Selection | 0.6735±0.382 | 0.7065±0.111 | 0.6091±0.108 |
| + Proxy Selection | 0.7082±0.035 | 0.7581±0.094 | 0.6136±0.089 |
| + Our Model | 0.7430±0.019 | 0.7387±0.107 | 0.6908±0.088 |
| MixMatch125 | 0.6390±0.024 | 0.7241±0.054 | 0.7012±0.103 |

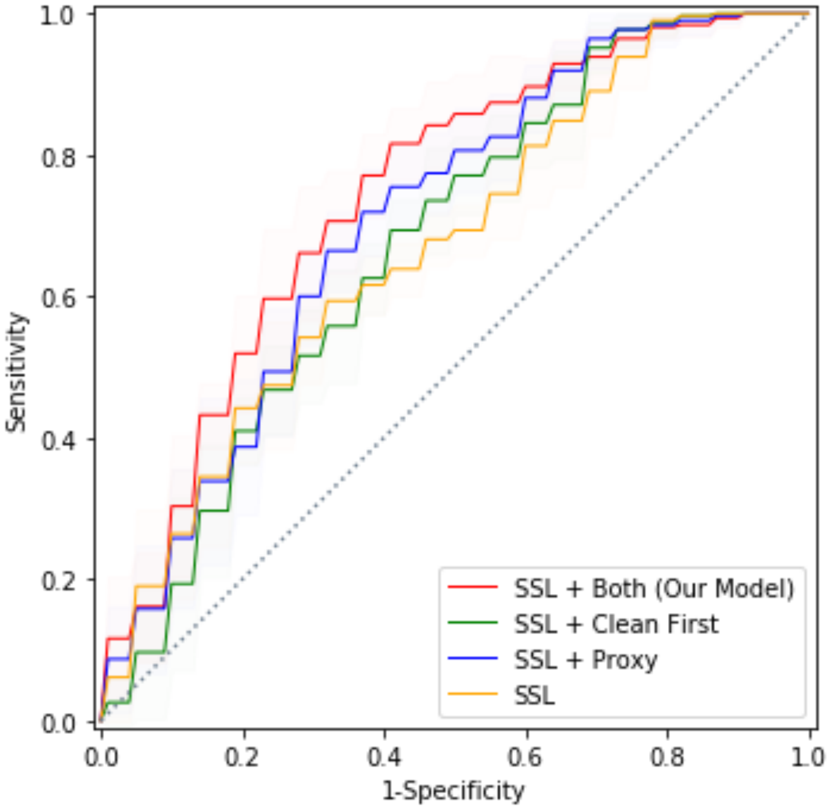


Figure 4-4. ROC curves that illustrate the average performance of our proposed method alongside ablation experiments.

## 4.5 Discussion

This study sought to evaluate a novel proxy learning framework for detection of imaging mismatch in AIS patients. The use of a semi-supervised learning framework allowed us to utilize information from the TSS label, which was available for all patients in our dataset, as a proxy for the more complex and resource-intensive DWI-FLAIR mismatch label. By incorporating this proxy information, we were able to improve the performance of our model and achieve more accurate and reliable results. Furthermore, our study provides evidence of the potential benefits of using SSL techniques for medical imaging tasks such as AIS, where labels are often expensive and time-consuming to acquire. This approach could be extended to other medical tasks where expert annotation is required and could help to overcome the limitations of small dataset sizes and inter-reader variability. Moreover, the low variance in our results indicate that our semi-supervised learning framework is robust and reliable, which is essential for its translation to the clinical setting.

Determining the age of an AIS lesion is critical for informing stroke treatments and making the most effective and timely decisions for patient care. The relationship between the time clock and tissue clock has been the subject of extensive research for many years, with TSS being identified as a surrogate proxy for the progression of ischemic tissue. Clinical imaging allows for the identification of the underlying tissue changes that occur as the progression of ischemic tissue takes place. One of the most widely used methods for identifying patients within 3 hours of stroke onset is using signal intensity differences between DWI and FLAIR imaging.[115] Recently, this imaging biomarker has been incorporated into stroke treatment guidelines with the presence of DWI-FLAIR mismatch being sufficient justification to give patients thrombolytics when the onset time is unknown. The use of DWI-FLAIR mismatch has also been clinically correlated with better outcomes for other stroke treatments such as mechanical thrombectomy.[119,120] At the same time, there are two major challenges when it comes to detecting DWI-FLAIR mismatch. Firstly, it requires assessment by an expert neuroradiologist, and secondly, it is an inherently subjective assessment that is prone to inter-reader variability.[114] Machine learning provides a path to reap the benefits of this imaging while reducing these factors.

This proxy-target paradigm of using a readily available clinical variable as a surrogate proxy for an underlying label that requires expert annotation is not limited to just stroke treatments but can be extended to other medical tasks as well. A common problem among medical tasks is that labels are expensive to acquire, leading to training on small datasets with questionable generalization when evaluated solely on data from one institution. To address this, one possible solution is to evaluate generalizability using an unseen, external dataset. This allows for a more comprehensive understanding of how well a model can generalize to different populations and settings.

Many areas of machine learning research have explored methods to incorporate prior knowledge into their models, as this can be particularly informative for medical image detection and segmentation tasks. Our approach utilizes prior information in a semi-supervised framework in two ways: to stratify unlabeled examples into clinically meaningful categories, and to classify samples according to the level of noise. Both methods have the goal of enhancing confidence in pseudo-labeled samples. Combining these two strategies yielded both higher detection performance and lower variability across replicates, the latter addressing instability of machine learning models trained on small datasets. Moreover, our proposed method outperformed the current state-of-the-art in semi-supervised learning, which involves calculating the consistency of predictions on augmented samples.[126,127] Additionally, the image registration pipeline creates spatially aligned images such that stroke location is information contained within the image. This allows for greater utility in a model where spatial location is informative and reduces the need for other forms of augmentation used in natural image datasets (e.g., flipping, rotation, and translation).[128] Given that medical datasets, even without labels, are often small, even a good feature representation that is reliant on further data augmentation may not be as effective. The types of augmentation used in natural image datasets, e.g., flipping, rotation, and translation, may be less useful in a model where spatial location could be informative.

## 4.6 Conclusion

Our study aimed to utilize TSS as a surrogate marker to detect DWI-FLAIR mismatch in MRI performed on patients suffering from AIS. Our experiments demonstrate that using proxy information within an SSL framework can significantly improve performance in terms of both classification accuracy and model stability. However, our study also has several limitations, including the relatively small sample size for both the development and evaluation datasets, as well as the limited number of experts who generated the annotated labels.

To address these limitations, future research directions could include evaluating this approach on a larger, multi-institutional dataset. Such an evaluation could enable the stratification of patients according to various clinical factors, such as demographics and medical history, which could provide new insights into the underlying mechanisms of tissue progression during AIS. Moreover, a wider evaluation of this approach could potentially lead to the development of a model that is capable of automatically detecting DWI-FLAIR mismatch from MRI scans taken at the time of imaging. This could be of great value in informing treatment options for AIS patients, as it could provide a more accurate and reliable assessment of salvageable tissue and help to optimize treatment decisions.

# CHAPTER 5

# Predicting Recanalization in Acute Ischemic Stroke Patients Undergoing Endovascular Therapy from Pretreatment Imaging

## 5.1 Overview

For acute ischemic stroke (AIS) patients, treatments such as thrombolysis and thrombectomy aim to restore blood flow to areas experiencing ischemia. Endovascular thrombectomy (EVT) is recommended for patients up to 24 hours from onset.[7,24] EVT is considered successful if blood flow is deemed completely or near-completely restored to the brain region affected by the stroke. This restoration is quantified by the modified treatment in cerebral infarction (mTICI) score, which is assessed both during the EVT procedure and upon completion.[45,47,129] Clinical trials have illustrated that patients who experience significant and/or full recanalization of the blood vessel typically experience better outcomes, particularly if recanalization is achieved on the first attempt – known as the first pass effect (FPE).[49,55,130,131]

Evidence indicates patients with similar clinical history, stroke characteristics, and procedural factors experience varied recanalization outcomes. Several studies have sought to elucidate the mechanisms underlying a patient's likelihood of successful recanalization, many of which have used imaging as a non-invasive method of determining clinical correlates. Like assessments for thrombolytic therapy, TSS onset is positively associated with long-term clinical outcomes post-EVT; identification of penumbral tissue via MR or CT imaging can inform treatment outcomes. Additionally, compensatory flow from pial collateral circulation has a strong correlation with prognosis post-EVT.[9,11,132] The current AHA/ASA stroke guidelines weakly recommend advanced imaging to assess a patient's collateral status.[2]

Thus, predicting a patient's post-treatment TICI and potential for FPE can inform interventional neuroradiologists about the potential benefit of thrombectomy. Imaging has already proven to contain signal that can elucidate individual patient characteristics, informing how one might respond to thrombolytic, or the more invasive thrombectomy, treatment. Developing methods to predict automatic classification of these key clinical variables from admission imaging alone can help clinicians at both community hospitals and trauma centers assess the most efficient course of treatment. Some imaging biomarkers have been statistically linked to successful recanalization, but they are not perfect predictors, and some are prone to inter-reader assessment variability. [36,133,134] Machine learning models have been developed to predict recanalization, achieving varied prediction performance. Every model with imaging inputs has utilized non-contrast CT (NCCT) and/or CT-Angiography (CTA). While NCCT can be useful to identify intracranial hemorrhage, diffusion-weighted MRI can more precisely visualize the infarct core volume for patients whose onset time is unknown outside a 6-hour window of symptom onset. Automated prediction methods have only achieved moderate performance and required manual segmentation of the clot by an expert neuroradiologist.[135–138] Even with the help of segmentation by experts, the time-intensive task of manual segmentation across a volume may not be compatible with current clinical guidelines such as the Target: Stroke quality improvement Phase III campaign, which recommends a door-to-needle time for EVT within 90 minutes of direct admission and 60 minutes of transfer patients.[139,140]

In this study, we develop and evaluate two methods – a region-based radiomics model and an end-to-end deep learning network – for their ability to predict the likelihood of a patient experiencing successful recanalization from their admission imaging. We report performance metrics for two cohorts of patients: those who underwent CT and MRI before treatment. The paper's contributions include a proof-of-concept radiomics model in MRI, which establishes the feasibility of using MRI as an input modality for EVT success prediction. Additionally, the work demonstrates that region-based, annotation-free models can match and outperform previous models requiring annotation in CT, achieving matching moderate performance in MRI. The lightweight convolutional architecture that uses spatial and cross attention outperform radiomics-based machine learning models across both modalities.

## 5.2 Materials and Methods

### 5.2.1 Data

The consecutive, retrospective cohort used for this study was treated at UCLA Ronald Reagan Medical Center for acute ischemic stroke from 2014-2021. Patients were included if they were diagnosed with a large vessel occlusion (LVO) stroke, had an adequate-quality MRI or CT acquired upon admission or suspected stroke, and received EVT treatment. Exclusion criteria were as follows: the presence of significant hemorrhage, and image registration errors resulting from significant midline shift or motion artifacts. As part of the EVT protocol at UCLA, mTICI is assessed during the procedure after each clot retrieval pass. For this study, successful recanalization was defined as an mTICI of 2b, 2c, or 3. Baseline features such as age, sex, NIH Stroke Scale (NIHSS) at admission, and time since stroke, were compared between patients who did or did not experience FPE using the chi-square test for categorical features and Student's t-test for continuous features. The cohort's clinical, imaging, and procedural characteristics are listed in Table 3-1. The tasks were binary mTICI classification and FPE classification.

### 5.2.2 Image Acquisition and Processing

Patient MR imaging acquisition was performed on 1.5T and 3T echo-planar MR scanners with 12-channel head coils (Siemens, Germany). In the stroke MRI brain imaging admission protocol, the diffusion-weighted imaging (DWI) and fluid-attenuated inversion recovery (FLAIR) sequences were acquired using the following parameters: DWI: TR 4000-9000, TE 78-122ms, corresponding pixel dimensions 0.859x0.859x6.000 to 1.850x1.850x6.500 mm; FLAIR: TR 8000-9000ms, TE 88-134ms, corresponding pixel dimensions 0.688x0.688x6.000 to 0.938x0.938x6.500 mm. Patients underwent MRI using a 1.5T or 3T echo-planar Siemens MR imaging scanner, performed with 12-channel head coils. DWI, FLAIR and apparent diffusion coefficient (ADC) sequences were used in this study. The DWI images were acquired using a TR range of 4,000-9,000ms and a TE range of 78-122ms. The pixel dimension for DWI varied from 0.859x0.859x6.000mm to 1.850x1.850x6.500mm. FLAIR images were acquired using a TR range of 8,000-9,000ms and a TE range of 88-134ms. The corresponding pixel dimension varied from

0.688x0.688x6.000mm to 0.938x0.938x6.500mm. ADC maps were calculated from DWI b0 and DWI b1000 using the following formula, where $S_{b1000}$ and $S_{b0}$ are the intensity values of DWI b1000 and DWI b0 images.

$$ADC = -\frac{\ln\left(\frac{S_{b1000}}{S_{b0}}\right)}{1000}$$

After image retrieval, the sequences were fed into our previously published automated preprocessing pipeline to ensure data consistency across individuals and reduce noisy information.[141] First, N4 bias field correction was applied to all sequences. Then, DWI images were skull-stripped and registered to a T2w MNI-152 atlas. FLAIR and ADC were co-registered afterwards. Next, intensity normalization and histogram matching were performed using a reference case. Lastly, a vascular territory template was mapped on the registered images to extract ROI where the stroke lesion was located.[142] Instead of manually segmenting the stroke lesion or thrombus, or using unvalidated segmentation algorithms, our detection method extracted the affected brain region automatically.

Two CT scanners, a Lightspeed VCT (GE Health Care, Milwaukee, USA) and a SOMATOM Definition (Siemens, Forchheim, Germany), were used for CT imaging. After administering 50 mL of contrast agent intravenously at 5 mL/second, a single-phase CT angiography (CTA) was obtained (120 kV, 120 reference mAs, 0.3 second rotation time, 0.6 pitch, effective dose of about 3 mSv). Following intravenous injection of contrast agent, totaling 50 mL at a rate of 5 mL/second, CTP included 30 successive spiral acquisitions (80 kV, 150 mA, effective dose = 3.3mSv, 100 mm in the z-axis) in a total of 60s acquisition. Saline was used after each contrast agent injection, with 30mL being used for each injection. Both non-contrast CT (NCCT) and CTA image series were included as inputs for the imaging-based models. The preprocessing protocol for CT images included field-of-view removal, skull stripping, and registration to MNI space. The ROI was segmented using the same method from the MRI pipeline, utilizing an atlas-based intensity method.

### 5.2.3 Radiomics Model

Radiomic features were extracted from the ROI for DWI, FLAIR and ADC sequences separately using the Python library pyradiomics.[143] The features included a) 13 3D shape features and 19 first-order features, b) 24 texture features computed from the Gray Level Co-occurrence Matrix (GLCM), c) 14 features from Gray Level Dependence Matrix (GLDM), d) 16 features from Gray Level Size Zone Matrix (GLSZM), e) 16 features from Gray Level Run Length Matrix (GLRLM), and f) five features from Neighboring Gray Tone Difference Matrix (NGTDM).

Many extracted radiomic features have a high degree of correlation, and they may provide overlapping information to our machine learning model. To minimize information overlap, we implemented three feature selection methods and compared the performance of each using different classifiers. Two supervised methods, Least Absolute Shrinkage and Selection Operator (LASSO), Random Forest (RF), and one unsupervised method, Principal Component Analysis (PCA), were selected due to their popularity and efficiency.[144] Briefly, LASSO applies a regularization process that penalizes the coefficient of regression features to minimize the prediction error and the features with non-zero coefficient after shrinking process are selected. For RF, features are selected by calculating each feature's contribution to the decrease of the weighted impurity of a tree. Feature importance is calculated by averaging the decrease of impurity across trees and ranking the features according to this measure, only keeping the top 50th percentile of features in this rank. For PCA, the current features are transformed to a representation with fewer new features (principal components) by a dimensionality reduction process that involves orthogonal linear transformation while preserving the variance presented in the data.

To predict mTICI, the following classifiers were compared: Support Vector Machines (SVM), K-Nearest Neighbor (KNN), Logistic Regression (LR), and RF. SVMs separate data into target classes by finding the hyperplane that maximizes the margin between the classes, while minimizing the classification error. The hyperplane is found by identifying the support vectors, which are the data points closest to the decision boundary and optimizing the parameters of the hyperplane using various optimization techniques. KNN classifies data based on the class of their k-nearest neighbors in the training data set, where k is a user-specified parameter. It works by measuring the distance between the new data point and all the data points in the training set and selecting the k data points that are closest in distance to the new data point. The class of the new data point is then determined based on the class of the majority of its k-nearest neighbors. LR models the probability of a binary outcome using the logistic function, which is a mathematical function that relates the probability of the outcome to a linear combination of predictor variables. The coefficients of the predictor variables are estimated using maximum likelihood estimation, and the resulting model can be used to make predictions for new values of the predictor variables. Classifiers were trained using 5-fold cross-validation in the training set. The best model hyper-parameters were selected via grid search. For example, different kernel functions (radial basis, linear, sigmoid), gamma and C values were examined for SVM to generate best combination of hyper-parameters. The selected features were scaled using min-max normalization before being used in classification training. All feature selection and classification models were evaluated using receiving-operator characteristic area-under-the-curves (ROC-AUC), as well as sensitivity and specificity, which were calculated using optimal Youden's J statistic. Performance ranges for a held-out validation set were calculated for each combination of feature selection method and classifier. This model is summarized in Figure 5-1.

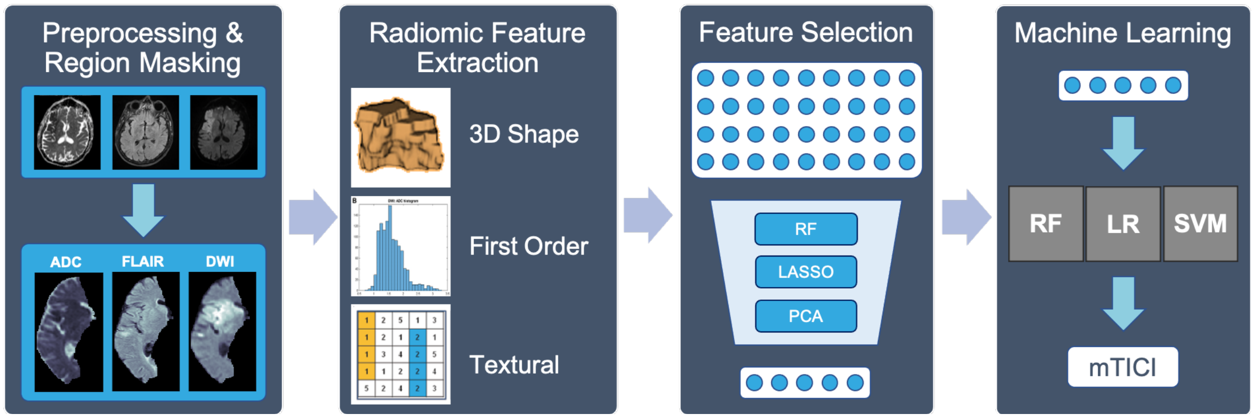


Figure 5-1. The radiomics pipeline used for prediction of recanalization. RF: Random Forest, LASSO: Least Absolute Shrinkage and Selection Operator, PCA: Principal Component Analysis, LR: Logistic Regression, SVM: Support Vector Machine.

### 5.2.4 Deep Learning Model Architecture

The proposed deep learning model is an end-to-end trainable network consisting of both convolutional and attention-based components.[145] This network incorporates modifications and enhancements to the widely used ResNet backbone. The first component is a global feature extractor, which leverages residual convolutional blocks to extract features from each slice. These slice-level features are then fed into local networks, which learn representations of adjacent slices and share weights during training. Within this local network, a non-local attention module determines the salient regions within each slice. Using the intermediate features as input, the non-local attention module uses a 1x1 convolution to generate single-head attention for each patch within the image as computed to all other patches. These are aggregated using matrix multiplication and softmax activation. The non-local attention module was included in the network for its self-contained nature, meaning it can be inserted into existing architectures without substantial increases in computation.[146]

Following the local networks, outputs were fed into the volumetric classifier consisting of two modules. The first is a cross-attention module utilizing recent advances in vision transformers.[147] The low-level features from every slice are fed into the module, which utilizes multi-head attention operations to generate slice-level importance. Resembling other attention modules including non-local attention, multi-head attention consists of a linear layer to generate attention across several scales of the image volume. The attention operations are fused using cross-attention, wherein the features from each scale are exchanged via layer normalization and residual connection. The use of this module in the network enables the model to weigh the most salient slices more heavily for the final prediction, while adding limited computational complexity. This attention output, along with the output from the local networks, are fed into a linear layer that serves as the final classifier, generating the volume-level prediction.

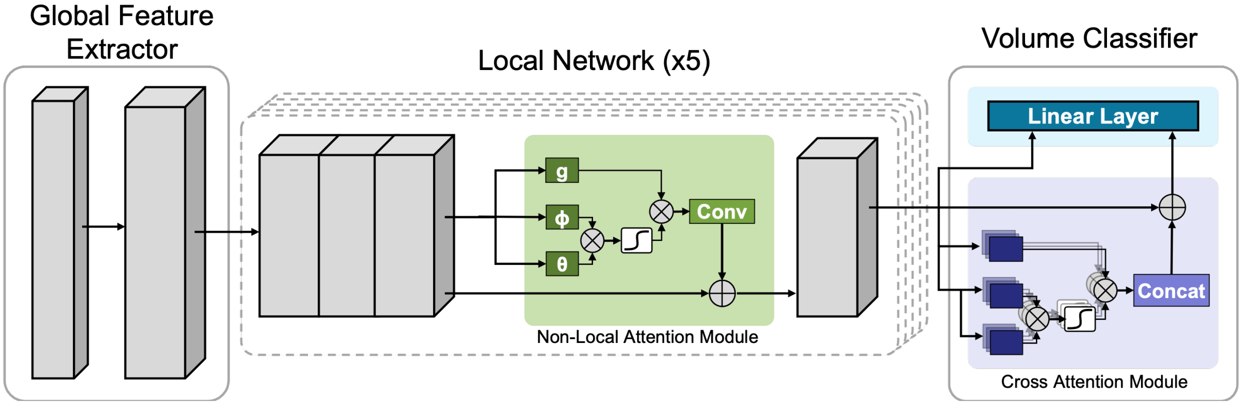


Figure 5-2. Overview of the proposed deep learning architecture. Utilizing a ResNet backbone, MR and CT images were used as a slice-wise input (input size 26x224x224) for a global 2D convolutional block, chosen because of the large slice thickness found in stroke protocols. The outputs from neighboring slices were then fed into ResNet34-based branches that shared weights across slice neighborhoods. The model leveraged two versions of transformer attention modules. A non-local attention module utilized multi-head attention on each slice to focus on salient regions. Within each neighborhood branch, a cross attention transformer identified important slices. Finally, the branch outputs were subjected to a weighted softmax layer to ultimately generate binary predictions.

### 5.2.5 Model Training and Evaluation

Models were evaluated for their ability to predict a binarized label for each patient. A patient was given a positive label if they had an mTICI score of 2b, 2c, or 3 after one pass during EVT. Patients that achieved recanalization in several attempts, or who did not achieve successful recanalization, were assigned a negative label. The MRI and CT cohorts were segmented into retrospective development and prospective evaluation groups. Patient images were included in the prospective cohort if they underwent EVT in 2020 or later. The development groups were each split into five folds for cross-validation. The model was trained for 200 epochs with early stopping, using the Adam optimizer with weight decay, and a batch size of 12. The learning rate was set to 0.0005 and the weight decay was set to 0.05. The training was implemented using PyTorch 1.9.0 on an NVIDIA DGX-2.

Following the development and hyperparameter tuning, algorithms were evaluated on the corresponding prospective evaluation cohort. Receiving-operator characteristic area-under-the-curves (ROC-AUC) were reported accordingly. Sensitivity, specificity, and accuracy were calculated using Youden's J statistics from the ROC curve. All metrics were reported as mean (standard deviation) on the evaluation set for each cohort.

### 5.2.6 Impact of Imaging Data

The first set of experiments involved assessing the ability for MRI to predict treatment response variables, as compared to clinical data extracted from the health record. Clinical variables extracted aligned with those used in previous studies, including demographic, historical, and stroke parameters. Demographic variables included patient age and sex. Patient variables also included history of prior stroke, hypertension, hyperlipidemia, atrial fibrillation, diabetes mellitus, and coronary artery disease. Stroke parameters extracted were the NIH Stroke Scale (NIHSS) taken at admission and the location of the stroke. These variables were utilized as input into the previously described machine learning pipeline to generate a final classification; this model was used as the baseline. This was compared to the previous radiomics model, as well as an aggregate model that used both the imaging and clinical variables.

In addition, experiments were conducted that evaluated the impact of using different imaging series for the predictive model. The series used were those generated during standard acquisition protocols for stroke patients upon admission. MRI acquisition protocols included diffusion-weighted and perfusion-weighted imaging (DWI and PWI, respectively). Experiments included evaluation of the use of DWI alone, PWI alone, and the image series together. The CT acquisition protocol involves NCCT and CTA. Similar to the MRI experiments, the models were evaluated for their ability to predict FPE using NCCT, CTA, and both.

## 5.3 Results

### 5.3.1 Patient Characteristics

This cohort included 379 patients who met the criteria; of these, 76 patients were excluded due to missing image series (29) or degraded image quality preventing preprocessing (24). From this final cohort of 326 patients, 149 underwent MRI, and 177 underwent CT before EVT. The cohort had an average age of $71.45 \pm 16.04$ years and was 55.73% female. Of this cohort, 98 patients experienced a stroke within 24 hours of the last-known well time but had an indeterminable onset time. Among patients with known onset time, 189 (57.98%) received imaging within the 4.5-hour window and 202 (61.96%) underwent contrast MRI or CT within 6 hours. Median NIHSS upon admission was 16 (IQR 10-20). Prior to EVT, 93 patients (28.79%) received intravenous thrombolytic therapy. Additional clinical variables as well as differences between the MRI and CT cohorts are summarized in Table 5-1.

Table 5-1. Demographics of patients included in model development. N, number of patients; SD, standard deviation, IQR, interquartile range; NIHSS, National Institutes of Health stroke scale; mTICI, modified treatment in cerebral infarction score.

| **Parameters** | **Measure** | **Total (N=326)** | **MRI (N=149)** | **CT (N=177)** | **p-value** |
|---|---|---|---|---|---|
| Female | *N* (%) | 180 (55.21%) | 85 (57.05%) | 95 (53.67%) | 0.1713 |
| Age | Mean ± SD | 70.7 ± 16.11 | 70.7 ± 16.11 | 72.08 ± 16.01 | 0.4432 |
| NIHSS | Median (IQR) | 16 (10-20) | 15 (8-19) | 16 (11-21) | 0.1144 |
| Received Intravenous Thrombolysis | *N (%)* | 96 (29.45%) | 35 (23.49%) | 61 (34.46%) | <0.0001 |
| **Stroke Onset Time** | | | | | 0.6013 |
| Stroke Onset to Image Time (min) | Median (IQR) | 54 (15-15) | 167 (123-255) | 113 (83-191) | -- |
| Unknown Onset | *N (%)* | 97 (29.75%) | 54 (36.24%) | 43 (24.29%) | -- |
| Onset < 4.5 hours | *N* (%) | 189 (57.98%) | 72 (48.32%) | 117 (66.1%) | -- |
| Onset < 6 hours | *N* (%) | 202 (61.96%) | 75 (50.34%) | 127 (71.75%) | -- |
| **Thrombectomy Outcome** | | | | | 0.4009 |
| Unsuccessful | *N (%)* | 58 (17.79%) | 28 (18.79%) | 30 (16.95%) | -- |
| mTICI 0 \| 1 \| 2a | *N / N / N* | 20 \| 4 \| 34 | 11 \| 2 \| 15 | 9 \| 2 \| 19 | -- |
| Successful, 2+ Passes | *N (%)* | 129 (39.57%) | 55 (36.91%) | 74 (41.81%) | -- |
| mTICI 2b \| 2c \| 3 | *N / N / N* | 73 \| 31 \| 25 | 36 \| 11 \| 8 | 37 \| 20 \| 17 | -- |
| Successful, First Pass | *N (%)* | 136 (41.72%) | 66 (44.3%) | 70 (39.55%) | -- |
| mTICI 2b \| 2c \| 3 | *N / N / N* | 59 \| 34 \| 43 | 31 \| 13 \| 22 | 28 \| 21 \| 21 | -- |

### 5.3.2 Radiomics Model Performance

A total of 321 features were extracted for each patient with DWI, FLAIR, and ADC sequences. 112 patients were assigned to the training set and 29 patients were assigned to the validation set following a 4-1 split, where the validation set was never seen by neither the feature selection, grid search nor classifier training processes. The demographic distribution for both training and validation sets remained the same for all experiments. To examine the stability of model performance, we trained and validated the results 100 times by changing the random seed to shuffle the training stage cross-validation. The results were reported as the mean ±95% confidence interval. All model training was performed using scikit-learn in Python 3. The best parameters for each model from grid search were: Random Forest with 80 max depth of the tree, max features 3, minimum number of samples required at a leaf node 3, minimum number of samples required to split an internal node 12, number of trees 200; SVM with radial basis kernel, penalty parameter C 10 and gamma 0.001; LR with L2 penalty and inverse regularization C 0.1; KNN with ten neighbors.

The RF feature selection determined 138 features were informative for mTICI classification. The LASSO feature selection kept 36 important features. We implemented PCA at 0.99 explained variance cutoff with five principal components for modeling. ROC curves on the validation set for each of the top models with best feature selection was reported in Figure 5-3. RF feature selection and LASSO feature selection yielded comparable performance across different models where the combination of RF feature selection and RF classification model achieved best ROC-AUC 74.29±0.68%. On the other hand, PCA feature selection yielded lower performance across models. Moreover, models with PCA features tended to be less stable with a larger variance except for LR model, showing that the PCA method may exclude much important information during dimension reduction step. Although the RF model achieved the highest performance, the LR model achieved 72.91±0.84% using the RF feature selector and 72.29±0.85% using LASSO feature selector. SVM achieved $>$70% ROC-AUC for RF features but lower for LASSO and PCA features. In general, KNN model achieved lower ROC-AUC values for all three feature selectors with higher variance across experiments.

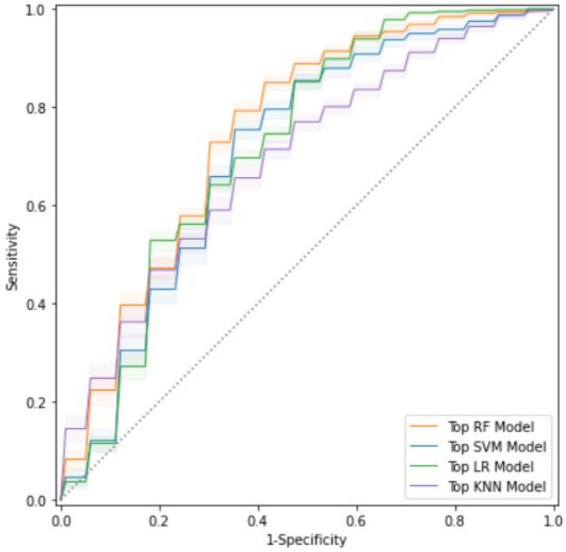


Figure 5-3. ROC curves - best feature selection method for each model. Top RF model = RF feature + RF model, top SVM model = RF feature + SVM model, top LR model = RF feature + LR model, top KNN model = LASSO feature + KNN model.

### 5.3.3 Impact of Series Selection

The results of imaging input experiments are summarized in Table 5-2 and Table 5-3 for MRI and Table 5-4 and Table 5-5 for CT. Across both modalities, the combination of imaging modalities improved predictive performance. In MRI, the deep learning model performed best when using both diffusion-weighted and perfusion-weighted images. The performance of the model was not significantly different in the cross-validation set, but it achieved higher ROC-AUC (0.9100 ± 0.0583) and accuracy (0.8889 ± 0.0703) in the prospective test set. For the CT cohort, the combination of NCCT and CTA achieved the highest FPE prediction performance, with substantial increases in ROC-AUC (0.8120 ± 0.0650) and accuracy (0.8233 ± 0.0582) in the test set. These findings suggest that the combination of images may provide more comprehensive information for accurate prediction of perfusion deficits.

Table 5-2. Deep learning model performance on MRI cross-validation folds.

| Series | ROC-AUC | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|
| **PWI** | 0.8450 (0.0720) | 0.8347 (0.0584) | 0.8639 (0.1084) | 0.8089 (0.1248) |
| **DWI** | 0.7725 (0.0510) | 0.7770 (0.0612) | 0.7400 (0.1008) | 0.8057 (0.1264) |
| **PWI + DWI** | 0.8364 (0.0900) | 0.8320 (0.0987) | 0.7500 (0.1470) | 0.9111 (0.0831) |

Table 5-3. Deep learning model performance on prospective MRI test set.

| Series | ROC-AUC | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|
| **PWI** | 0.8960 (0.0543) | 0.8600 (0.0490) | 0.8000 (0.1265) | 0.9200 (0.0980) |
| **DWI** | 0.8519 (0.0559) | 0.8000 (0.0567) | 0.8857 (0.1400) | 0.7455 (0.1564) |
| **PWI + DWI** | 0.9100 (0.0583) | 0.8889 (0.0703) | 0.8000 (0.1265) | 1.0000 (0.0000) |

Table 5-4. Deep learning model performance on CT cross-validation folds.

| Series | ROC-AUC | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|
| **NCCT** | 0.6925 (0.1340) | 0.7260 (0.1053) | 0.7533 (0.0859) | 0.7118 (0.1577) |
| **CTA** | 0.6704 (0.0684) | 0.6650 (0.0539) | 0.5905 (0.1871) | 0.7474 (0.1610) |
| **NCCT + CTA** | 0.7928 (0.0894) | 0.8430 (0.0571) | 0.8333 (0.1394) | 0.8442 (0.0885) |

Table 5-5. Deep learning model performance on prospective CT test set.

| Series | ROC-AUC | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|
| **NCCT** | 0.7154 (0.0890) | 0.7314 (0.0596) | 0.7470 (0.1566) | 0.7200 (0.1492) |
| **CTA** | 0.7347 (0.0325) | 0.7378 (0.0356) | 0.7868 (0.0913) | 0.6993 (0.1120) |
| **NCCT + CTA** | 0.8120 (0.0650) | 0.8233 (0.0582) | 0.7861 (0.1210) | 0.8339 (0.0830) |

### 5.3.4 Deep Learning Model Performance

The classification performance of the deep learning model as compared to radiomics methods are summarized in Table 5-6. This includes comparison to the aforementioned MRI-based radiomics model and an annotation-based radiomics model proposed by Hofmeister et al.[148] When applied to MRI series, the deep learning model achieved an average ROC-AUC of 0.9100, with an accuracy of 0.8889 on the prospective test set. The model outperformed the previous method, notably achieving near-perfect specificity across experimental replicates while maintaining high sensitivity. In the prospective CT evaluation set, the deep learning method performed slightly worse, yielding a mean ROC-AUC of 0.8120 and accuracy of 0.8233. When compared to literature, this model achieved slightly lower average accuracy, though with a substantially smaller confidence interval. While the accuracy was slightly lower, the model achieved a balanced sensitivity and specificity of 0.7861 and 0.8339, respectively, compared to the previous model that achieved high specificity at the expense of very low sensitivity.

Table 5-6. Performance of our current model benchmarked against results from literature as well as previously published models applied to this cohort.

| Modality | ROC-AUC | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|
| **MRI** | **0.9100 (0.0583)** | **0.8889 (0.0703)** | **0.8000 (0.1265)** | **1.000 (0.0210)** |
| **Zhang et al.**[149] | 0.7429 (0.0680) | 0.7533 (0.1516) | 0.6992 (0.0253) | 0.8059 (0.0229) |
| **CT** | **0.8120 (0.0650)** | **0.8233 (0.0582)** | **0.7861 (0.1210)** | **0.8339 (0.0830)** |
| **Hofmeister et al.** [135] | 0.88 | 0.8510 (0.717–0.938) | 0.5000 (0.211–0.789) | 0.9710 (0.851–0.999) |

## 5.4 Discussion

Successful recanalization has been shown to correlate with better outcomes for AIS patients. Establishing a reliable predictive relationship between pretreatment imaging and final mTICI/FPE is crucial for better EVT strategy planning. In this study, we aimed to explore the capacity of pretreatment imaging to predict the likelihood of recanalization or FPE during EVT using a deep learning approach. To our knowledge, this study represents the first attempt to explore automated prediction of FPE across multiple imaging modalities and methods. The use of deep learning algorithms in this study provides several advantages over traditional methods. Firstly, our approach does not require manual segmentation of the clot, which is a time-consuming process and can delay valuable treatment time. Instead, our model automatically learns to identify relevant features from the input images without requiring manual intervention. Secondly, our models do not require advanced imaging techniques, such as perfusion imaging, to achieve high performance in predicting successful recanalization. Such advanced imaging has a long acquisition time and may not be available in all stroke triage settings.

One interesting finding is that the models using MRI consistently achieve higher performance than both previous methods as well as the same model applied to CT series. This is unsurprising, given that MRI series contain more substantial and complementary information. However, the CT model was still able to achieve balanced performance without advanced imaging, which may be more relevant for the vast majority of healthcare workflows that utilize CT for AIS patient admission. There is important information from standard diffusion and perfusion MR images before treatment that is directly related to EVT recanalization, leading to a potential new path of investigation in pre-treatment MR imaging and thrombectomy outcome.

Our machine learning performance results produce a few findings. First, we demonstrated that the RF feature selector combined with the RF model achieved the best performance. Based on the 95% Confidence Interval for 100 times repeated experiments, RF and LR models both show stable performance with minimal variation when shuffling the training data. RF models have been illustrated to be robust across many iterations, and moreover, have proven to achieve high classification performance for other radiomics tasks. Other tasks have found optimal performance using other classifiers, however, so more testing is needed to evaluate these models at a broader scale. A second finding is that our automated region extraction method yielded benchmark performance comparable to those generated from manually segmented regions on CT images. A machine learning model that doesn't require manual segmentation can provide significantly faster inference time, as manual annotation can be labor intensive, particularly to segment small regions within the brain such as a clot thrombus. Given the time-sensitive nature of stroke treatment decisions, this fully automatic method can quickly provide a model with the relevant stroke region without sacrificing performance. Additionally, a model that requires no segmentation is not prone to subjectivity of manual segmentations or variability among segmentation models.

Several studies have sought to predict recanalization for AIS patients, which achieved moderate performance using clinical variables,[36,150] while others relied on handcrafted or statistical features extracted from manually segmented regions on CT imaging.[134,151–153] For example, Hofmeister et al utilized radiomics features in a machine learning model to predict FPE from CT, achieving high specificity but low sensitivity.[154] The method proposed here, in contrast, requires no manual segmentation and achieves balanced sensitivity and specificity with comparable accuracy. The radiomics and deep learning approaches proposed in this work were developed and evaluated on cohorts who received either MR or CT imaging before treatment, thereby carrying wider applicability to clinical stroke workflows.

Limitations of this study include study cohort bias, imaging limitations due to retrospective collection, and small sample size. As patients were only included in the cohort if they underwent EVT as part of the study design, this model may be subject to treatment bias introduced during treatment decision-making. An additional source of bias is that the target variables rely solely on the assessment of the neurointerventionalist performing the procedure. There is substantial discourse surrounding the use of TICI scores and correlations with outcome, which undoubtedly introduces variability in the expert assessment depending on their training. This cohort was assessed using the mTICI score. This evaluation is inherently subjective; while there is a high degree of reliability for patients scored mTICI 2c and 3, there is high inter-reader variability for patients scored mTICI 2b.[155] This is likely due to the large range of patients within a score class, as patients with 2b can experience anywhere from 50%-89% recanalization. This scoring metric has undergone several augmentations since its proposal in 2005, both because of this variability and poor correlation with functional outcomes. Even with the proposed new categories, mTICI 2b comprises a large spectrum of recanalization rates. The Highly Effective Reperfusion Evaluation in Multiple Endovascular Stroke (HERMES) trial sought to further delineate these categories.[47] The results from the HERMES trial established a 7-point scale known as the extended thrombolysis in cerebral infarction (eTICI) score.[156] Even so, this scale provides moderate correlations with functional outcome.[157] Future studies could involve reassessing these patients using their procedural imaging taken during EVT to stratify the mTICI 2b patients into more granular quantification of recanalization.

There are further data biases introduced due to routine clinical imaging protocols that preclude a direct comparison of imaging series inputs across modalities. The current

MR stroke protocol generates an angiogram with partial coverage of the brain, excluding lateral portions of the head in the field-of-view. Similarly, the CT protocol optimized for patients presenting within 18 hours of onset acquires perfusion imaging with an incomplete field-of-view in along the superior/inferior axis. Both series were unable to be registered with the image processing pipeline that was optimized for speed, given current treatment paradigms. There are more advanced image registration techniques, including those that leverage deep learning, that may be able to perform partial registration. However, preliminary experiments indicated that these operations would perform at a time scale that would not be feasible in a real-world use case. Scanners at other institutions may be able to capture full field-of-view angiography and perfusion imaging for both MR and CT studies. Cohorts with full coverage across all series would be able to demonstrate the utility of standard and advanced imaging for prediction of EVT success. Finally, this is a proof-of-concept study from one institution, and the architecture has many parameters. The training and evaluation cohorts were split to maximize evaluation capacity, and two cohorts with different imaging modalities were used, but they still only comprise one institutional dataset. External validation is required to determine the applicability of these models to other hospitals and institutions.

## 5.5 Conclusion

We have presented a fully automatic, end-to-end method to predict treatment response to EVT. On a dataset of patients who received either MR or CT prior to treatment, we have demonstrated that the volume-based deep learning network can distinguish whether a patient will be successfully recanalized in one attempt or fewer, achieving peak accuracies of 88.80% using MR imaging and 82.33% using CT image series. This method outperformed previously published methods without requiring manual thrombus segmentation, illustrating the capability of deep learning algorithms to inform treatment planning for AIS patients.

There are a few areas of future research direction this study provides. Our analysis and modeling were conducted from a single center with images acquired retrospectively over eight years. A multi-center prospective study with a larger sample size is needed in the next step. In this study, we only examined several popular feature selection methods and machine learning algorithms. Many other feature selection methods and ML classification should be examined in future study. A reliable automated stroke lesion or thrombus segmentation algorithm is expected to provide more accurate radiomics features for use. By incorporating more accurate 3D image features, stratification by the type of EVT technique, and clinical factors as features into the machine learning model, the model performance can be improved.

# CHAPTER 6

# Leveraging Intra-Procedural Imaging During Endovascular Therapy to Inform Treatment Decision-Making

## 6.1 Overview

Endovascular therapy (EVT) has become the gold standard treatment for eligible patients presenting with acute ischemic stroke (AIS). Despite recent advancements in thrombectomy tools and protocols, there remains a risk of procedural complications and poor functional outcomes, even among patients deemed eligible under the strictest clinical trial criteria.[67,68] While a first pass retrieval is the optimal treatment target, if recanalization is not achieved in one attempt, the stroke team must decide the next step in a short time window. Options may be to proceed with another retrieval attempt, perform an adjunct therapy, or stop the procedure altogether. While there remains debate about pretreatment decision-making, a neurointerventional radiologist can deploy adjunctive techniques in real time, such as rescue stenting or intra-arterial thrombolysis (IA-tPA) to decrease the likelihood of adverse outcomes, though both come with substantial risk of post-procedural complications.

Among patients for whom EVT was truncated due to procedural difficulties, 20% still experience moderate outcomes and limited disability.[66] The decision to pursue adjuvant options is reliant on several factors, many of which cannot be ascertained prior to the procedure. Imaging plays an essential role in decision-making during EVT. Digital subtraction angiography (DSA) enables neurointerventionalists to guide devices to the site of occlusion and attempt to remove the clot through either retrieval or direct aspiration techniques.[37] It is also used to assess recanalization after every retrieval attempt and to quantify compensatory collateral flow.

Collateral flow refers to the flow of blood through alternative vessels in the brain that can compensate for a blocked or narrowed vessel. In the context of ischemic stroke, collateral flow can provide an alternative supply of blood to the affected brain tissue, thus mitigating the damage caused by the blocked vessel. Intrinsic collaterals refer to the blood vessels within the brain that can provide alternative blood flow, such as the pial, leptomeningeal, and deep penetrating collaterals. Extrinsic collaterals refer to blood vessels outside the brain that can provide alternative blood flow, such as the circle of Willis and the external carotid artery collaterals.[132,158] The presence of adequate collateral flow can have significant clinical implications in stroke patients. Patients with good collateral flow have been shown to have better outcomes following treatment, including higher rates of recanalization, lower rates of symptomatic intracerebral hemorrhage, and better chances of functional recovery.[9,11] However, the volume and extent of collateral flow varies from person to person and is impacted by a variety of parameters such as age, blood pressure, and the existence of underlying cardiovascular disorders.

Previous work has demonstrated the value in using imaging informatics and machine learning techniques to extract insights from DSA series, but these studies have been preliminary in nature.[159–161] Recently, vision transformers have been shown to be effective for a wide range of medical image understanding tasks, including disease medical image registration, disease diagnosis, and semantic segmentation.[99,162] One potential advantage of vision transformers for medical image analysis is their ability to process images of arbitrary size without the need for resizing or cropping. This ability is useful as medical images often have different sizes and aspect ratios depending on the modality and the specifics of the imaging acquisition hardware.

In this work, we apply deep learning to automatically quantify collateral flow during EVT using DSA imaging. We developed an automatic arterial input function (AIF) localization algorithm for DSA images and used it to estimate hemodynamic perfusion parameter maps. Additionally, we implemented a time-series transformer network to classify collateral status. We compare this algorithm to previous work and demonstrate improved classification performance.

## 6.2 Materials and Methods

### 6.2.1 Patient Cohort

This study was based on a retrospective, single-cohort dataset of patients with AIS treated with EVT at the University of California, Los Angeles between 2011 and 2021. EVT was conducted by neuroradiologists with more than five years of expertise. The study's inclusion criteria were (1) the availability of biplane projection data (lateral and frontal, as per mTICI scoring recommendations), (2) a minimum DSA series length of five frames, (3) no radiopaque foreign body projections over the middle cerebral artery territory, and (4) no additional pathology such as dissection, subarachnoid hemorrhage, or aneurysm.[163]

### 6.2.2 Collaterals Assessment

On baseline pretreatment DSA runs, two interventional neuroradiologists with over six years of post-fellowship experience assessed the collaterals using the American Society of Interventional and Therapeutic Neuroradiology/Society of Interventional Radiology (ASITN/SIR) collateral flow rating method summarized in Table 6-1.[164] The clinical information for each case was hidden from the neuroradiologists; if there was disagreement about the collateral assessment, the neuroradiologists established consensus via discussion. Patients were divided into two groups: those with adequate collaterals (ASITN/SIR grades 3 and 4) and those with insufficient collaterals (ASITN/SIR grades 0, 1, and 2).

Table 6-1. ASITN/SIR collateral grading scale.

| Score | Definition |
|---|---|
| **0** | No presence of collaterals |
| **1** | Slow collateral flow to the peripheral vasculature surrounding the ischemic site |
| **2** | Rapid collaterals with persistence of some flow defects |
| **3** | Slow but complete angiographic blood flow to the ischemic bed in later venous phase |
| **4** | Complete and rapid collateral blood flow to the entire ischemic territory |

### 6.2.3 Image Acquisition and Preprocessing

A regular timed contrast-bolus passage approach was used to generate DSA images using a Philips biplane. Manual injection of Omnipaque 300 was conducted at a dilution of 70% (30% saline) such that 10 cc of contrast was delivered intravenously at a rate of about 5 $\mathrm{cm}^3$/s. Image acquisition settings differ across individuals. Frames are acquired in an interleaved form at two common views in the biplane acquisition setting: coronal (AP) and sagittal (lateral). The image sizes were all 1024x1024, however they were taken with various fields of view.

Patient images were queried from the institutional picture archiving and communications system (PACS). Images were converted from the Digital Imaging and Communications in Medicine (DICOM) to Neuroimaging Informatics Technology Initiative (NIFTI) formats, using series descriptions from the DICOM header to determine the temporal order of lateral and frontal DSA series.

### 6.2.4 Perfusion Angiography Map Generation

Five perfusion parameter maps were generated for each lateral and frontal DSA pair: cerebral blood flow (CBF), cerebral blood volume (CBV), time to peak (TTP), mean transit time (MTT), and time to maximum (Tmax). Taken together, these hemodynamic calculations characterize the level of perfusion or lack thereof for a time-series image. These operations rely on the initial assumption that the signal intensity $I(t)$ in a DSA frame is inversely proportional to the amount of contrast within that location. Therefore, the concentration of contrast $C(t)$ at any given pixel $(x, y)$ can be represented as:

$$C(t)_{(x,y)} = \frac{k}{I(t)_{(x,y)}\mu\rho_{(x,y)}}$$

Where $k$ and $\mu$ represent attenuation coefficients and $\rho_{(x,y)}$ denotes the thickness of the vessel wall at that location. This was used as the tissue level concentration curve $C_u$ at any given tissue point, as well as the AIF $C_a$. Their relationship can be represented as a convolution $\bigotimes$ using the transit time distribution $h(t)$ of the residue function $R(t)$.

$$C_u(t) = C_a(t) \bigotimes h(t)$$

$$R(t) = 1 - \int_{\tau=0}^{t} h(\tau)d\tau$$

This is reliant on localizing a region of interest (ROI) to the artery within the image and computing the tissue concentration curve.[165] The identification of the arterial ROI is based on the intuition that the AIF is most clearly represented in large cerebral vessels because there is minimal partial volume effect from surrounding tissues.[166] To identify the largest vessel, we utilize the heuristics that the injection will pass through the arterial vessel in the first half of the DSA timesteps for a given series, and that the vessel can be identified on the image by finding the largest single contour with high contrast. ROI extraction for AIF localization was automatically performed via the following process:

1. Crop each video to remove border and label artifacts and remove the second half of timesteps (1024x1024xT → 900x900x0.5T)
2. Perform Otsu thresholding to identify the areas of highest concentration within the series, then binary opening on the thresholded masks.
3. Find the timestep with the highest amount of contrast concentration.
4. Within this timestep frame, identify the largest single region of high contrast.
5. Perform binary opening and region growing to generate a final ROI.

With this ROI, the tissue concentration curve was calculated by measuring the average amount of contrast concentration at each given timestep. Graphical representations of $C_u$, $C_a$, and $R(t)$ are summarized in Figure 6-1. This average contrast concentration curve was then fit to a mixture gamma distribution to account for multiple contrast passages at any given spatial pixel (x, y). The resulting AIF function was used to compute the residual function and the subsequent five perfusion parameter maps. To generate $R(t)$, a Toeplitz matrix was computed for the AIF sorted in descending order along the timepoints. The identity of the single value decomposition matrix was used as $R(t)$.

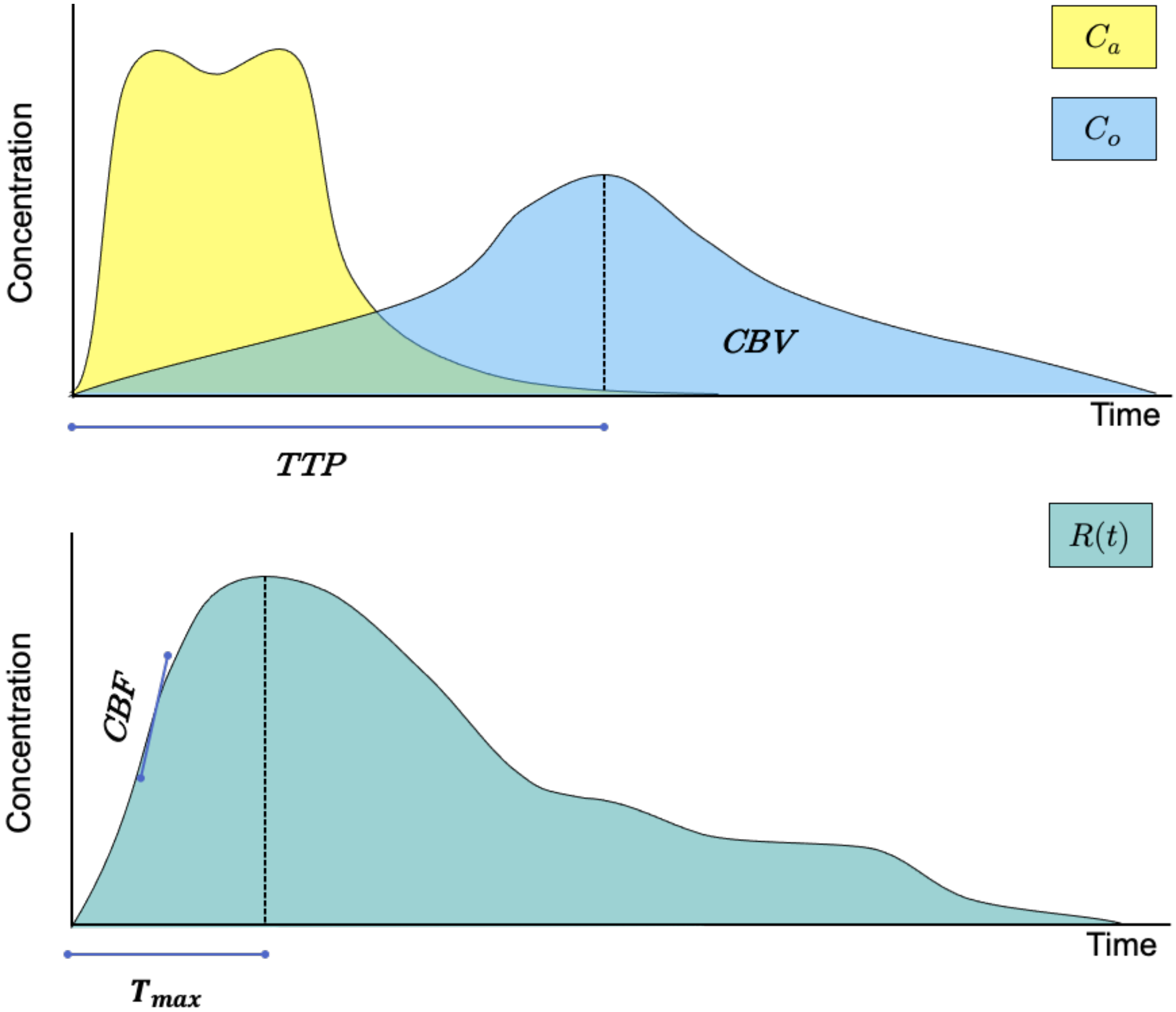


Figure 6-1. Relationship between concentration functions and perfusion parameters. Performing deconvolutions between $C_u$ and the AIF $C_a$ (top) produces residue function $R(t)$ (bottom) Cerebral blood volume (CBV) is calculated as the integral of the tissue concentration curve Cu, and time to peak (TTP) is the time taken to reach the maximum value. The cerebral blood flow (CBF) is the maximum gradient of $R(t)$, and time to maximum (Tmax) is the time required for $R(t)$ to reach maximum concentration.

The resulting $C_u$ , $C_a$ , and $R(t)$ curves were used to generate hemodynamic parameters; the formulae used for each can be seen in Table 6-2. Briefly, CBF was calculated as the peak of $R(t)$ , and CBV was calculated as the quotient of the concentration curve integrals. MTT is the quotient of CBV divided by CBF. Finally, the TTP was computed as the time when peak concentration was achieved in tissue, and Tmax was determined to be the time at which the residual function reached its peak.

Table 6-2. Perfusion parameter calculations. CBF, cerebral blood flow; CBV, cerebral blood volume; MTT, mean transit time; TTP, time to peak; Tmax, time to maximum.

| Parameter | Definition | Equation |
|---|---|---|
| **CBF** | Cerebral blood flow denoting the flux at a particular pixel; computed by identifying the maximum level of contrast identified in the residual function. | $rCBF = \max R(t)$ |
| **CBV** | Amount of blood contained within a given amount of brain tissue; calculated via integration of concentration curves and typically reported as a local, relative metric. | $rCBV = \frac{\int_{t=0}^{\infty} C_u(t)dt}{\int_{t=0}^{\infty} C_a(t)dt}$ |
| **MTT** | Ratio of blood volume to blood flow; local, pixel-level metric that refers to the average time a red blood cell requires to traverse through a vessel. | $MTT = \frac{CBV}{CBF}$ |
| **TTP** | Time taken for the tissue to reach the maximum concentration of contrast tracer agent. Reflects the time required for the contrast to be delivered to the tissue; does not depend on AIF. | $TTP = \ argmax_t C_u$ |
| **Tmax** | Time at which the concentration of a tracer or contrast agent reaches its maximum value in a particular ROI. Delay is associated with acute ischemic stroke infarct. | $TMax = \ argmax_t R(t)$ |

### 6.2.5 Deep Learning Models

For our subject cohort, the model task was to binarily classify collateral status. In this study, we utilized three different temporal models to analyze the given time series data. The first model was a Gated Recurrent Unit (GRU)-based encoder, which consists of multiple GRU layers that process temporal input data and encode it into a compact representation, capturing the underlying temporal dependencies in the data. The second model was the Video Swin Transformer, which is an extension of the Transformer architecture designed specifically for temporal data using non-overlapping multi-head self-attention. We also employed the TimeSformer architecture that leverages a Transformer-based encoder-decoder architecture, where the encoder processes the input data to extract relevant features, and the decoder uses this information to make predictions for future time steps.

All three models were randomly initialized and trained for 200 epochs with preset hyperparameters outlined in their original publications on the first biplane pair of images acquired prior to the first EVT pass. For ease of training, the most lightweight variation of each (i.e., that with the fewest parameters) was chosen for comparison. The binary classification performance was measured using five metrics: ROC-AUC, accuracy, sensitivity, specificity, and precision. The models were evaluated in five-fold stratified cross-validation.

### 6.2.5.1. Gated Recurrent Unit-based Encoder

To evaluate the temporal transformer networks for binary collateral classification, we generated baseline performance metrics using a method already applied to DSA images.[159,167] Nielsen et al. proposed a spatiotemporal network to automatically score thrombolysis in cerebral infarction (mTICI). This model utilized both planes of the DSA image series as input. For each stack of a sagittal and coronal frame, the model utilized a pretrained EfficientNet-B0 as a feature extractor and encoder. A gated recurrent unit (GRU) generates frame-level predictions, which are then aggregated into a final series-level prediction via a classification head containing swish and linear layers. A GRU is a type of recurrent neural network (RNN) that uses two gate mechanisms, the update gate and the reset gate, to regulate information flow in the network. A GRU operates on an input sequence $x_t$ and the hidden state $h_{t-1}$ at the previous time step. The update gate $z_t$ and the reset gate $r_t$ are computed as follows:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z)$$
$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r)$$

$W$ and $U$ are weight matrices, $b$ is a bias vector, and $\sigma$ is the sigmoid activation function. At each step, $z_t$ determines the proportion of information to be incorporated from $x_t$ and $r_t$determines the proportion of information to be preserved from $h_{t-1}$. The next hidden state, $h_t$ is then calculated:

$$h_t = (1 - z_t)h_{t-1} + z_t * \tanh[x_t + U_h(r_t \bigodot h_{t-1}) + b_h]$$

The combination of the update and reset gates enables the GRU to effectively manage the balance between the preservation of information from the previous hidden state and the incorporation of new information from the current input.

### 6.2.5.2. Vision Transformers

Two transformer-based architectures were chosen for their suitability for time-series data. A vision transformer architecture typically consists of multiple layers of self-attention, which allow the model to focus on different parts of the input data at different times, allowing it to effectively process input sequences of any length.[92,168] An image is divided into a grid of small image patches, which are then treated as a sequence of tokens transformed into a set of "queries," "keys," and "values," which are used to compute the attention mechanism.[90,169] These linear transformations are achieved through matrix multiplication. The attention mechanism itself is then computed using dot products between the queries, keys, and values. Specifically, the dot product between the query and key matrices is computed, and this is used to compute weights that determine how much each element of the value matrix should be weighted. The weighted sum of the value matrix is then computed using these weights, resulting in a new matrix that represents the attended-to elements of the input data. The output of the attention mechanism is then typically passed through a classification network, which outputs a final segmentation or classification.

These models process video frames in a sequential manner, using the attention layers to focus on relevant parts of the frames as they progress through a sequence of data. Proposed by Bertasius et al., TimeSFormers involve computing global self-attention in the space and time dimensions in immediate series, generating three-dimensional embeddings that represent both dimensions.[170] This divided space-time attention works to compute self-attention, first in the temporal dimension and then the spatial dimension to reduce computational complexity. The approach proceeds as follows. Patches are generated in the temporal dimension, creating key/query/value sets $\left\{W_{K^{time}}^{(\ell,a)}, W_{Q^{time}}^{(\ell,a)}, W_{V^{time}}^{(\ell,a)}\right\}$. After linear embedding into a learnable matrix, the temporal self-attention $\alpha_{(p,t)}^{(\ell,a)}$ can be computed for a single patch $(p,t)$:

$$\alpha_{(p,t)}^{(\ell,a)time} = softmax\left(\frac{q_{(p,t)}^{(\ell,a)}}{\sqrt{D/_{\mathcal{A}}}}^{\top} \cdot \left[k_{(0,0)}^{(\ell,a)} \cdot \left\{k_{(p\prime,t\prime)}^{(\ell,a)}\right\}_{t'=1,\ldots,T}\right]\right)$$

After computing self-attention for all temporal patches, the resulting attention vectors are fed into a multi-layer perceptron (MLP) layer to generate a new set of keys, queries and values for the spatial dimension $\left\{W_{K^{space}}^{(\ell,a)}, W_{Q^{space}}^{(\ell,a)}, W_{V^{space}}^{(\ell,a)}\right\}$. In this way, the embeddings learned from the temporal patches are then used as patches for the self-attention computations in the spatial dimension:

$$\alpha_{(p,t)}^{(\ell,a)space} = softmax\left(\frac{q_{(p,t)}^{(\ell,a)}}{\sqrt{D/_{\mathcal{A}}}}^{\top} \cdot \left[k_{(p,t)}^{(\ell,a)} \cdot \left\{k_{(p\prime,t\prime)}^{(\ell,a)}\right\}_{p\prime=1,\ldots,S}\right]\right)$$

The resulting sequential attention embeddings are then fed into a final classification head to generate labels for the task of interest. This architecture carries several advantages over previous iterations: it can be applied to longer temporal data and is relatively faster than applying a space-time attention module in parallel.

Global self-attention can be highly effective at localizing to salient regions but requires colossal amounts of computing time. Considering this weakness, Liu et al. proposed a local transformer mechanism based on the imaging-based nonlocal Swin Transformer,[97,171] which has shown utility for several medical deep learning tasks.[172–175] The architecture works by partitioning an input into non-overlapping 3D windows, before applying multi-head self-attention (MSA) modules in sequence. Following a feed-forward network consisting of two MLP layers and non-linear activation, this process then repeats, first merging the patches and then downsampling embeddings to create a new set of partitions with a shifted window. The entire network utilizes four sets of MSA and feed-forward networks in series, with the learned representations being fed into a classification head. Compared to TimeSFormers, this approach requires fewer attention computations and therefore requires less training time. For example, given an input $T \times W \times H$ with dimensions $8 \times 8 \times 8$, and a window size of $4 \times 4 \times 4$, attention would be computed for $2 \times 2 \times 2 = 8$ windows; this contrasts with TimesFormer, which would require $8 \times 2 \times 2 = 32$ self-attention modules under the same paradigm. Video Swin transformers also inherently contain relative position embedding through cross-window connections, which has been shown to improve performance on video classification and segmentation tasks.[90,176]

## 6.3 Results

### 6.3.1 Patient Characteristics

A total of 122 patients were included based on treatment and imaging criteria. Clinical data is summarized in Table 6-3. This cohort was 59.02% female, with an average age of 71.71 years. Median NIHSS at admission was 15. While most patients had a known onset time, with a median onset time of 165 minutes, about 40% of patients had an unknown onset time, though they were known to be within 24 hours as is required by EVT inclusion criteria. One quarter of patients (31) received intravenous thrombolysis prior to undergoing EVT. Patients had on average 30 DSA images associated with the study in AP and lateral views. During the EVT procedure, 22% did not achieve successful recanalization, and 37% achieved successful recanalization in one attempt.

Table 6-3. Demographic data of the patient cohort.

| Parameter | Measure | (N = 122) |
|---|---|---|
| ASITN/SIR Collateral Score - Good \| Bad | N \| N | 43 \| 81 |
| Female | N (%) | 72 (59.02%) |
| Age | Mean ± SD | 71.71 ± 15.18 |
| NIHSS | Median (IQR) | 15 (8-19) |
| Received Intravenous thrombolysis | N (%) | 31 (25.41%) |
| Stroke Onset Time | | |
| Stroke onset to Image time (min) | Median (IQR) | 165 (120-244) |
| Unknown | N (%) | 50 (40.98%) |
| ≤ 4.5 h | N (%) | 59 (48.36%) |
| ≤ 6 h | N (%) | 63 (51.64%) |
| Thrombectomy Outcome | | |
| Unsuccessful | N (%) | 27 (22.13%) |
| mTICI 0 \| 1 \| 2a | N \| N \| N | 6 \| 5 \| 13 |
| Successful, 2+ Passes | N (%) | 45 (36.89%) |
| mTICI 2b \| 2c \| 3 | N \| N \| N | 30 \| 9 \| 6 |
| Successful, First Pass | N (%) | 52 (42.62%) |
| mTICI 2b \| 2c \| 3 | N \| N \| N | 26 \| 10 \| 16 |

### 6.3.2 Hemodynamic Perfusion Parameters

The hemodynamic estimation framework effectively processed 92% of the DSA images; instances were rejected during processing if there was inadequate acquisition time (i.e., the DSA acquisition did not cover the whole injection cycle) or insufficient number of frames. This workflow is visualized in Figure 6-2, and more representative patient images can be seen in Figure 6-3.

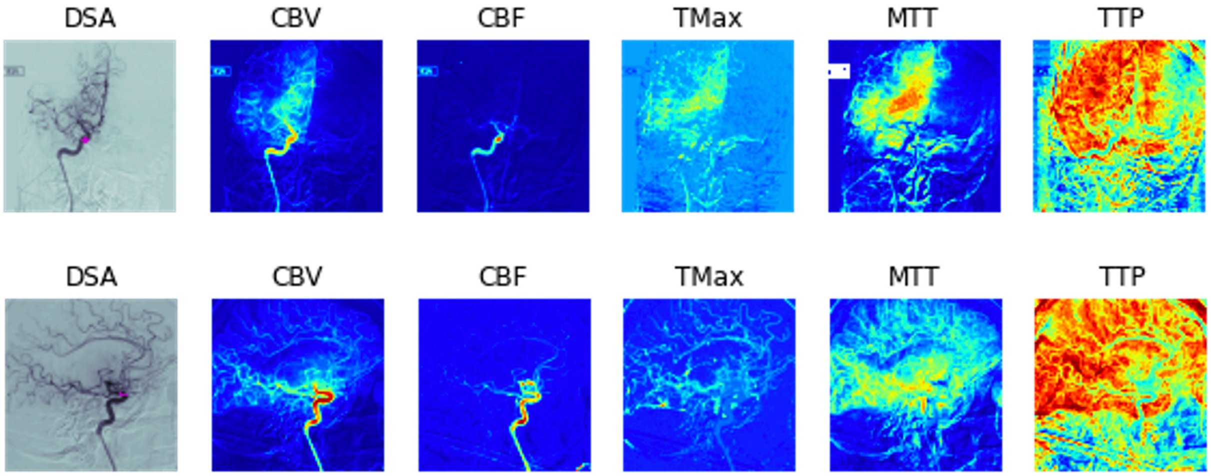


Figure 6-2. Example of a processed set of DSA images in the frontal view. The original DSA (top left) underwent the AIF localization method to create a ROI, seen in pink. Following this, the tissue concentration curve (CTC) was generated to create the arterial input function (AIF) seen in panel 3. From these, the CBV, CBF, Tmax, MTT, and TTP maps were generated (panels 4-8).

### 6.3.3 Classification Performance

The performance metrics of each time-series model are summarized in Table 6-4. The previous model tailored to DSA classification based on GRU modules achieved high sensitivity at the expense of specificity and precision, overall attaining moderate classification accuracy of 0.6927 (0.0402). In contrast, the two transformer-based architectures achieved better performance both with respect to overall accuracy as well as balanced sensitivity and specificity. In particular, the Video Swin model achieved the highest accuracy and ROC-AUC of 0.7507 (0.0622), and the TimeSformer achieved higher sensitivity of 0.7444 (0.1938) and specificity of 0.7537 (0.1885), as well as better precision of 0.6969 (0.1879). However, these performance metrics all had substantial variation across folds, preventing further statistical conclusions from being made about the increased performance.

Table 6-4. Classification performance of temporal deep learning architectures.

| | ROC-AUC | Accuracy | Sensitivity | Specificity | Precision |
|---|---|---|---|---|---|
| **Gated Recurrent Unit** | 0.7286 (0.0479) | 0.6927 (0.0402) | 0.9333 (0.0889) | 0.5684 (0.1226) | 0.5429 (0.0655) |
| **Video Swin Transformer** | 0.7341 (0.0459) | 0.7507 (0.0622) | 0.7222 (0.1097) | 0.7412 (0.1061) | 0.6119 (0.0696) |
| **TimeS-former** | 0.7083 (0.0287) | 0.7343 (0.0382) | 0.7444 (0.1938) | 0.7537 (0.1885) | 0.6969 (0.1879) |

Three representative patients are shown in Figure 6-3. These three images show the angiographic variance of patients who experience partial recanalization (mTICI 2b), as well as the variance in model performance. The images in Figure 6-3A are from a patient who was correctly classified as having bad collateral flow. The parameter maps produced by our method demonstrate restoration of blood flow following the successful EVT pass. Despite this poor compensatory flow, this patient was successfully being recanalized after four attempts, did not experience large infarct growth, and had a favorable mRS score of 4 at 90 days post-discharge. Figure 6-3B shows a case that was also successfully classified by all three models. Here, the perfusion parameter maps indicate a successful restoration of partial flow, aligning with the mTICI 2b score assessed by the radiologist after one attempt. The model was able to correctly classify this patient despite motion artifacts seen on the perfusion maps. However, the model was not always successful in classifying collateral flow in images with motion artifacts; as seen in Figure 6-3, these images were acquired from a patient with a poor collateral score who experienced poor outcomes. The models classified this patient as having strong collaterals despite motion interfering with the visualization noted by the neuroradiologists during assessments. It is possible that the motion artifacts in this image series interfered with the model's ability to localize to areas of high contrast across time points.

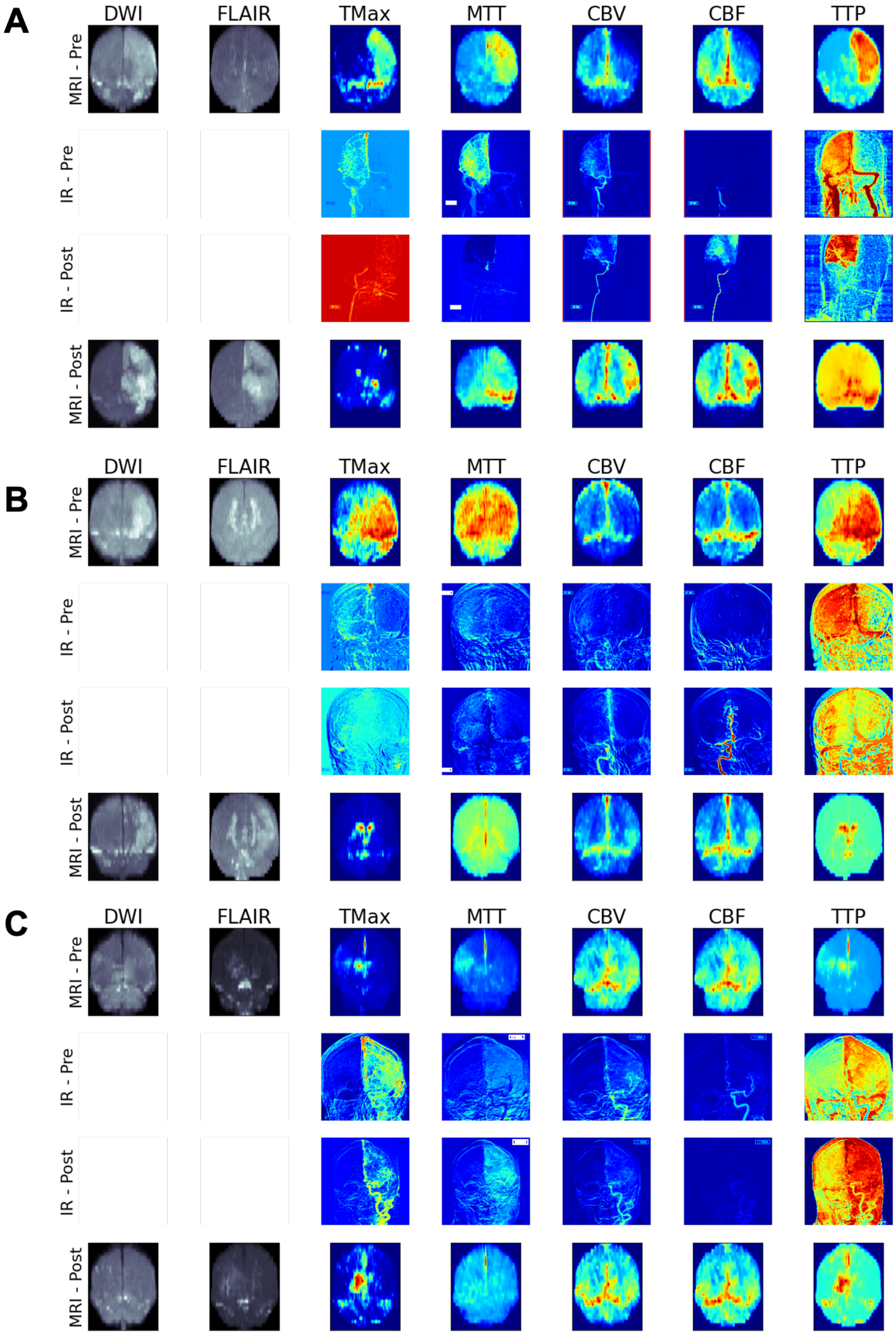


Figure 6-3. Perfusion parameters maps generated from DSA images. For each patient A-C, MRI – Pre and MRI – Post images represent a maximum intensity projection (MIP) of series taken pre- and post-EVT, respectively: DWI, FLAIR, Tmax, MTT, CBV, CBF, and TTP. IR – Pre and IR Post are the first and last biplane pair of images taken during EVT, processed to generate the perfusion maps.

## 6.4 Discussion

We found that the perfusion parameter maps provided utility when inspecting patients who experienced successful recanalization during EVT but went on to have large regions of infarcted tissue. Qualitative inspection revealed substantial variation between patients who were sedated and those who were not. Among patients kept awake during EVT, motion sometimes degraded the images such that the vessels could not be visualized on the parameter maps. Motion correction algorithms could be applied to these images prior to inference. Recent advances in video motion correction across other domains could increase the quality of these images, which could benefit both the interventionalist during the procedure and any downstream algorithms. For example, work by Ueda et al. utilized a U-net based deep learning model to perform motion correction and remove misregistration artifacts by generating a synthetic angiogram from dynamic DSA images.[177] Further research on the validation of perfusion angiography for estimation of hypoperfusion volume and degrees of recanalization/reperfusion during endovascular interventions would be valuable.

Our classification experiments also found that transformers are suited to utilizing DSA images for classification. There have been a small number of studies using EVT-derived DSA image series in deep learning and machine learning applications. Nielsen et al. developed the GRU-based network to automatically classify EVT success from DSA during treatment; this model achieved moderate agreement and was used as the baseline comparison model for this study.[159] Compared to this baseline, the Video Swin Transformer achieved better performance and required less computation time; in particular, the model achieved higher accuracy and a balanced sensitivity and specificity. The lowered computation time can be explained by the reduction in memory requirements used by Swin. Many RNN-based architectures including GRU-units require memory gate mechanisms, which must be computed in series (rather than in parallel). One of the significant advantages of Transformer architectures is that self-attention mechanisms can be computed in parallel, which offers two benefits: parallel processing significantly reduces both training and inference time, and the non-recurrent nature of the attention mechanism allows the model to maintain encodings of all time steps during the computation. This design helps to prevent information loss across time steps, which could enhance model performance. Additionally, the transformer allows for concurrent consideration of both future and past elements, combining the benefits of bidirectional RNNs, without incurring the additional computational cost of duplicated computation.

Neuroimaging plays a crucial role in providing neurologists and neurointerventionalists with information for guidance in clinical decision-making for stroke patients. These images are used before treatment to classify the stroke, during treatment for decision-making and beyond the acute phase for recovery and management. In the past year, DSA has been used for inputs for stroke detection and landmark detection models.[160,178] To our knowledge, our work represents the first attempt to automatically classify collateral flow from DSA images. While the models achieve moderate accuracy utilizing these DSA images as input, there is largely room for improvement in performance. While a model with balanced sensitivity and specificity may more accurately identify patients' collateral statuses, clinical paradigms may shift to favor a sensitive model (identifying those with strong collaterals) or a specific one. The latter may be most useful when considering patients for adjuvant and/or bridging therapies during EVT.

Our approach was developed with several considerations toward the eventual clinical use case. The models and workflows were designed to require limited processing and inference time, to mimic the temporal requirements of neurointerventionalists during EVT. Additionally, the model relies solely on DSA imaging, not utilizing clinical parameters or even pretreatment imaging. This was done for two reasons: to minimize the parameters for an input modality with high dimensionality, and to reduce reliance on import/export features that may not be tenable in several EVT workflows. With these considerations in mind, it is still possible that inference time may still be too slow depending on the computation resources allotted to an angiography suite.

There are several limitations to this study. The cohort used for these models was small and comprised patients from one institution. This means that the results may not be generalizable to other populations or institutions and may not be as accurate for patients who do not fit the characteristics of the patients in the study. The cohort is inherently biased towards those who qualify for EVT, as a patient would not undergo vascular imaging like DSA during a stroke unless their clinical treatment plan required such imaging. Finally, the motion degradation of several images limited the ability to evaluate this method.

## 6.5 Conclusion

Our results indicate that DSA can be used as an input modality for deep learning tasks to aid in EVT decision support. Future technical adaptations could include adaptive learning algorithms that could integrate new information as the procedure progresses. These algorithms could also be refined to include multi-view encodings, which explicitly model the spatial information extracted from frontal and lateral DSA series. From a clinical perspective, further study is needed to assess the utility of such algorithms and how they would be deployed in a clinical workflow, which may involve reader studies during procedural simulations.

# CHAPTER 7
# Conclusion

## 7.1 Summary of Contributions

The objective of this dissertation was to investigate the use of imaging techniques taken during routine clinical diagnosis and treatment of acute ischemic stroke (AIS) to characterize and predict patient response to various treatment modalities. The research focused on developing techniques that can be easily incorporated into the workflow of AIS, specifically, methods that utilize inputs acquired during standard clinical diagnosis and treatment, do not require time-intensive manual annotation or a significant amount of modeling or inferential time. The technical chapters of this dissertation present four scientific endeavors that have been undertaken to achieve the research goal. These chapters detail the contributions to science that have been made through these efforts.

In Chapter 3, we evaluated an automatic technique for determining TSS from imaging that does not require sub-specialist radiology expertise. Using an international dataset comprising patients from two institutions, we developed and externally evaluated a deep learning network for classifying TSS from MR images and compared algorithm predictions to neuroradiologist assessments of DWI-FLAIR mismatch. Models were trained to classify TSS within 4.5 hours and performance metrics with confidence intervals were reported on both internal and external evaluation sets. The deep learning method performed similarly to radiologists and outperformed previously reported methods. Our model achieved higher generalization performance on external evaluation datasets than the current state-of-the-art for TSS classification. This represents the largest evaluation of any TSS classification algorithm to date. These results demonstrate the potential of automatic assessment of onset time from imaging without the need for expertly trained radiologists.

In Chapter 4, we developed a novel method for detecting DWI-FLAIR mismatch in MRI scans of patients with AIS using TSS as a surrogate marker. The tailored semi-supervised method was evaluated on an unseen dataset from an external institution. The framework was found to significantly improve the performance of the classifier in terms of both accuracy and model stability, and the method outperformed other state-of-the-art semi-supervised learning methods. The findings suggest that incorporating proxy information within a semi-supervised learning framework can significantly enhance the performance. With more rigorous evaluation, this model could be beneficial for AIS patients as it would provide an objective assessment of salvageable ischemic tissue. In the face of changing clinical guidelines, this detection model could provide verification of image assessments performed by neuroradiologists.

In Chapter 5, we created and assessed algorithms for their ability to predict a patient's potential response to endovascular thrombectomy (EVT). Specifically, we evaluated predictions of recanalization and first pass effect (FPE) from pretreatment imaging. We employed radiomics pipelines and deep learning networks with non-local and cross-attention modules to predict FPE on MR and CT imaging series. These results suggest that non-perfusion MRI and CT contain signals that can predict successful EVT FPE, and that time-consuming manual segmentation is not necessary to develop an automated method. This study represents the first classification of FPE from MRI alone and the first automated FPE classification method in CT; moreover, this effort is the first to evaluate prediction methods across both MRI and CT. This model could be useful for neurologists when deciding to perform EVT on patients when they are admitted for AIS.

In Chapter 6, we assessed the potential of intra-procedural imaging to inform neurointerventionalist decision-making during EVT by developing visualizations and classification models informative to patient response. We utilize digital subtraction angiography (DSA) images to generate perfusion parameter maps. The distribution perfusion angiographic patterns following recanalization may offer additional clinical benefits beyond the evaluation of source angiographic images solely through visual assessment. We also evaluated transformer-based models for their ability to classify patient's collateral flow from DSA. We demonstrated that these transformer-based models outperform previous temporal models applied to DSA, presenting methods to rapidly quantify collaterals during EVT. These visualizations and classification models, taken together, supplement information gathered during EVT to provide neurointerventionalists with clinical biomarkers to provide the most optimal treatments possible to the patient in real time.

These efforts sought to aid decision-making across a patient journey when being treated for AIS. The clinical use-case for each task influenced several aspects of the experimental design. The considerations toward minimizing time-to-treatment influenced the design of the models and pipelines for specific tasks. Given the low data and annotation paradigms faced, these approaches also demonstrated the utility of lightweight models with relatively small numbers of trainable parameters. We leveraged several common low-data techniques to improve model performance such as transfer and semi-supervised learning. With the former, we found that pretraining on out-of-domain images negatively impacted performance; in several cases, transfer learning from natural images (on which most off-the-shelf architectures are trained) was similar to or worse than random initialization. Models pretrained on medical images did improve performance, with the highest performance seen when a model was pretrained on neuroimaging datasets. We found that this only improved performance to a certain extent, likely due to variations in neuroimaging series used for other disease domains.

An additional consideration was towards the training and evaluation of models for specific tasks. For some tasks, such as identifying patients within the thrombolytic treatment window, sensitivity was the preferred metric as the clinical application was to expand the pool of eligible patients. In contrast, when predicting successful EVT response, neurointerventionalists sought to focus on specificity, that is, identifying patients where the treatment would not be successful. This was based on the current treatment paradigm where those performing EVT tend to pursue treatment if inclusion criteria are met, and they would want a discerning predictor of unsuccessful recanalization. That said, evaluations placed highest emphasis on accuracy given the balance of labels within these tasks, both in the clinical setting and for our datasets specifically. Accordingly, models were trained to optimize for accuracy and ROC-AUC, where the evaluation focused on the specific metrics outlined by clinicians.

A final design principle was that the model outputs needed to be readily understood by clinicians. We implemented several interpretability visualizations, ranging from posthoc visualizations to trainable attention maps. We found that simple, singular visualizations proved the most utile for clinicians to accept the model predictions, rather than an expansive set of analyses, but that this suite of visualizations proved useful for dissecting model behavior during the development and evaluation phases of our study.

Our studies demonstrated the effectiveness of lightweight models with low-data techniques, such as transfer and semi-supervised learning, to assist in decision-making during the treatment of AIS patients. We designed our models and pipelines with the goal of minimizing time-to-treatment in mind and optimized them to align with clinicians' preferred evaluation metrics. Additionally, we prioritized interpretability and included a range of visualization techniques to help clinicians comprehend our model outputs. These results show promise for enhancing the clinical management of AIS patients and offer a potential basis for future investigations.

## 7.2 Future Work

Machine learning has the potential to revolutionize the field of stroke management by providing faster, more accurate, and more efficient diagnosis and treatment of AIS. As machine learning algorithms continue to develop, they have the potential to transform the field of stroke imaging by enabling more precise and personalized diagnosis and treatment. With the ability to process vast amounts of data quickly and accurately, these algorithms can identify stroke patterns that may be difficult for human observers to detect.

The primary research questions put forth in this study provided preliminary data supporting the use of imaging biomarkers for stroke triage and treatment decision-making. There were several barriers encountered that provide the foundation for future study. First, the small amount of data that was able to be collected limited the ability to provide more granular model outputs; all of the tasks were trained for binary classification or prediction. For example, the potential extension of the time clock window for thrombolytic treatment is currently under active study through clinical trials, and thus, a model that could more effectively delineate between patients that are within three to 4.5 hours of onset vs. patients within 4.5 to six hours of onset could provide clinicians with information about their eligibility for treatment. In a similar vein, being able to discretely predict if a patient will experience unsuccessful recanalization, successful recanalization in several attempts, or first pass effect could inform either the neurointerventionalist's decision to perform EVT or other treatment. Larger datasets spanning multiple institutions with annotations from several experts could enable these types of predictions, as the current datasets were too small to achieve model stability on such tasks.

Beyond label granularity, there are remaining questions about the utility of imaging biomarkers in conjunction with other standard clinical data. The technical advancements in the broader fields of machine and deep learning provide ample fodder for technological innovation, leveraging techniques most suited to medical data or clinical tasks. For example, the nascent field of modality fusion in deep learning refers to the integration of multiple data sources or modalities into a single model for better performance. For clinical tasks, different modalities such as imaging, molecular, and activity tracking data can provide complementary information that can improve the accuracy of diagnosis and treatment planning. Modality fusion in deep learning has the potential to enhance the effectiveness of medical imaging by combining information from multiple modalities to create a more comprehensive representation of the underlying anatomy or pathology. One promising approach is the use of machine learning to analyze pathology data from the retrieved clot, which can provide valuable information about the underlying cause of the stroke. By accurately identifying the underlying cause of the stroke, clinicians can choose appropriate treatments that are targeted to the specific subtype. Additionally, AI can be used to monitor and manage stroke patients after treatment. AI algorithms can analyze patient data, such as vital signs and medication adherence, and identify patients who are at high risk of complications or readmission. This information can be used to guide follow-up care and reduce the risk of adverse events.

This dissertation focused on augmenting certain decision points in the patient workflow, but there are several potential research questions at other decision points both within and beyond the acute setting. For example, the use of imaging algorithms for prehospital detection and characterization of AIS is a promising approach for early triage and treatment planning. Recent work in AI for prehospital detection of stroke has focused on developing machine learning algorithms that can accurately and rapidly identify potential stroke patients based on prehospital data such as vital signs, symptoms, and medical history. In particular, the use of portable imaging units can provide imaging at the site of discovery. These algorithms can be used to alert clinicians and emergency responders of a potential stroke patient before they arrive at the hospital, enabling them to prepare for the patient's arrival and provide timely and appropriate care. Rapid inference machine learning algorithms could identify EVT candidates for prehospital providers, thereby enabling more efficient routing to primary stroke centers and reducing time to treatment. There is also potential in utilizing imaging and other biomarker data to model long term outcomes. Once treated, AIS patients are at risk of not experiencing full functional recovery or experiencing a recurrent stroke within a year of their original incident. The mechanisms informing these kinds of responses remain unclear and should be an active field of study moving forward. Research questions include further elucidating the relationship between time clock and tissue clock to rapidly identify salvageable penumbral tissue in order to more rapidly route patients to facilities that can perform EVT. An additional mechanism worth further analysis is the reperfusion of tissue following recanalization. Up to 15% of patients do not experience restoration of oxygen to ischemic areas following treatment despite removal dissolution of clot thrombi. Understanding the biochemical and pathological changes in vessel architecture and inflammatory responses that could lead to this phenomenon could pave the path forward to algorithms that predict the likelihood of this and other complications that may preclude invasive treatments such as EVT.

While there have been many promising developments in this field, the vast majority of studies and efforts (including those described in this dissertation) have been single pilot studies with small sample sizes. This has been a common issue in many fields of AI and machine learning, where the results of initial studies may not necessarily generalize well to larger and more diverse datasets. It is crucial to evaluate the performance of these methods in large, multi-center studies with diverse patient populations. Such studies can provide a more robust evaluation of the effectiveness and generalizability of machine learning methods across different medical settings and patient populations. The lack of standardized reporting makes it difficult to compare and evaluate the performance of different models, which can hinder the adoption and development of AI in various fields. Due to growing recognition of the need for standardized reporting of AI models in scientific publications, several initiatives have been launched to establish reporting standards such PRISM for diagnostic tests, CONSORT for clinical trials, and TRIPOD for diagnosis and prognosis. These efforts have created checklists of items that should be reported in scientific publications, such as the study design, the patient population, the reference standard used, and the statistical methods used to evaluate the model's performance. These initiatives aim to promote transparency, reproducibility, and comparability of AI and machine learning studies; many journal publications now require self-reporting of the applicable checklist. In theory, these checklists will enable secondary analyses about the applicability of such models beyond the initial development and evaluation contexts.

Deploying ML algorithms in the healthcare setting can be challenging due to several barriers. One of the main barriers is the need for large amounts of high-quality data to train and validate ML models. To achieve high performance, ML models need to be trained on large, diverse, and representative datasets that accurately reflect the population of patients for which the model will be used. ML models that perform well on small, homogeneous datasets may not perform well on larger, more diverse datasets. This is because the models may not be able to generalize to new patients or may be biased towards certain subgroups of patients. Therefore, it is important to evaluate ML models on large, diverse, and representative datasets that accurately reflect the population of patients for which the model will be used. Complicating this is the fact that healthcare data is often siloed and not easily accessible, making it difficult to obtain the large amounts of data needed to train and validate ML models. Privacy and security concerns also pose a significant challenge in the deployment of ML algorithms in healthcare as the data contains sensitive information about patients.

Additionally, the algorithms must be designed to integrate into existing clinical workflows and infrastructure, which can be complex and time-consuming. Furthermore, the performance of the algorithms must be rigorously tested and validated to ensure that they are accurate, reliable, and generalizable. Ethical challenges are related to the use of machine learning algorithms in medical decision-making. For example, there may be concerns about the potential for bias in the algorithms, particularly if they are trained on data that is not representative of the patient population being served. There may be concerns about the potential for the algorithms to be used to automate decisions that should be made by human physicians, or about the potential for the algorithms to be misused or abused. Regulatory challenges include ensuring that the algorithms are developed and validated in accordance with relevant regulations and guidelines, obtaining appropriate approvals and certifications, and addressing issues related to data privacy and security.

Despite these challenges, the potential benefits of using machine learning in stroke imaging are immense. Machine learning has the potential to reduce physician workload by automating many of the tedious and time-consuming tasks associated with stroke imaging, such as image processing and analysis. This can free up physicians to focus on more complex and nuanced aspects of stroke diagnosis and treatment, improving the quality of care that they are able to provide. Moreover, machine learning has the potential to improve patient outcomes by enabling earlier and more accurate diagnosis of stroke, which can be critical for successful treatment. By providing more personalized treatment plans based on patient data, machine learning can also help to improve treatment outcomes and reduce the risk of complications.

# References

1. Lees KR, Bluhmki E, von Kummer R, et al. Time to treatment with intravenous alteplase and outcome in stroke: an updated pooled analysis of ECASS, ATLANTIS, NINDS, and EPITHET trials. *The Lancet*. 2010;375(9727):1695-1703. doi:10.1016/S0140-6736(10)60491-6
2. Powers WJ, Rabinstein AA, Ackerson T, et al. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke a guideline for healthcare professionals from the American Heart Association/American Stroke A. *Stroke*. 2019;50(12):E344-E418. doi:10.1161/STR.0000000000000211
3. Demaerschalk BM, Kleindorfer DO, Adeoye OM, et al. Scientific Rationale for the Inclusion and Exclusion Criteria for Intravenous Alteplase in Acute Ischemic Stroke A Statement for Healthcare Professionals from the American Heart Association/American Stroke Association. *Stroke*. 2016;47(2):581-641. doi:10.1161/STR.0000000000000086
4. Urrutia VC, Faigle R, Zeiler SR, et al. Safety of intravenous alteplase within 4.5 hours for patients awakening with stroke symptoms. *PLoS One*. 2018;13(5):e0197714. doi:10.1371/journal.pone.0197714
5. Thomalla G, Simonsen CZ, Boutitie F, et al. MRI-guided thrombolysis for stroke with unknown time of onset. *New England Journal of Medicine*. 2018;379(7):611-622. doi:10.1056/nejmoa1804355
6. Nogueira RG, Jadhav AP, Haussen DC, et al. Thrombectomy 6 to 24 Hours after Stroke with a Mismatch between Deficit and Infarct. *New England Journal of Medicine*. 2018;378(1):11-21. doi:10.1056/nejmoa1706442
7. Albers GW, Marks MP, Kemp S, et al. Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging. *New England Journal of Medicine*. 2018;378(8):708-718. doi:10.1056/nejmoa1713973
8. Bang OY, Saver JL, Kim SJ, et al. Collateral flow predicts response to endovascular therapy for acute ischemic stroke. *Stroke*. 2011;42(3):693-699. doi:10.1161/STROKEAHA.110.595256

9. Shuaib A, Butcher K, Mohammad AA, Saqqur M, Liebeskind DS. Collateral blood vessels in acute ischaemic stroke: A potential therapeutic target. *Lancet Neurol*. 2011;10(10):909-921. doi:10.1016/S1474-4422(11)70195-8
10. Bang OY, Saver JL, Buck BH, et al. Impact of collateral flow on tissue fate in acute ischaemic stroke. *J Neurol Neurosurg Psychiatry*. 2008;79(6):625-629. doi:10.1136/jnnp.2007.132100
11. Bang OY, Saver JL, Kim SJ, et al. Collateral flow predicts response to endovascular therapy for acute ischemic stroke. *Stroke*. 2011;42(3):693-699. doi:10.1161/STROKEAHA.110.595256
12. Tomsick T. *TIMI, TIBI, TICI: I Came, i Saw, i Got Confused*. Vol 28.; 2007. Accessed March 17, 2021. www.ajnr.org
13. Thomalla G, Fiebach JB, Østergaard L, et al. A multicenter, randomized, double-blind, placebo-controlled trial to test efficacy and safety of magnetic resonance imaging-based thrombolysis in wake-up stroke (WAKE-UP). *International Journal of Stroke*. 2014;9(6):829-836. doi:10.1111/ijs.12011
14. Odland A, Særvoll P, Advani R, Kurz MW, Kurz KD. Are the current MRI criteria using the DWI-FLAIR mismatch concept for selection of patients with wake-up stroke to thrombolysis excluding too many patients? *Scand J Trauma Resusc Emerg Med*. 2015;23(1). doi:10.1186/s13049-015-0101-7
15. Ho KC, Speier W, Zhang H, Scalzo F, El-Saden S, Arnold CW. A machine learning approach for classifying ischemic stroke onset time from imaging. *IEEE Trans Med Imaging*. 2019;38(7):1666-1676. doi:10.1109/TMI.2019.2901445
16. Lee H, Lee EJ, Ham S, et al. Machine learning approach to identify stroke within 4.5 hours. *Stroke*. 2020;51(3):860-866. doi:10.1161/STROKEAHA.119.027611
17. Ho KC, Speier W, El-Saden S, Arnold CW. Classifying Acute Ischemic Stroke Onset Time using Deep Imaging Features. *AMIA Annu Symp Proc*. 2017;2017:892-901.
18. Winzeck S, Hakim A, McKinley R, et al. ISLES 2016 and 2017-benchmarking ischemic stroke lesion outcome prediction based on multispectral MRI. *Front Neurol*. 2018;9(SEP):679. doi:10.3389/fneur.2018.00679
19. Balachandar N, Chang K, Kalpathy-Cramer J, Rubin DL. Accounting for data variability in multi-institutional distributed deep learning for medical imaging.

*Journal of the American Medical Informatics Association.* 2020;27(5):700-708. doi:10.1093/jamia/ocaa017

20. D'Esterre CD, Trivedi A, Pordeli P, et al. Regional Comparison of Multiphase Computed Tomographic Angiography and Computed Tomographic Perfusion for Prediction of Tissue Fate in Ischemic Stroke. *Stroke.* 2017;48(4):939-945. doi:10.1161/STROKEAHA.116.015969
21. Thust SC, Heiland S, Falini A, et al. Glioma imaging in Europe: A survey of 220 centres and recommendations for best clinical practice. *Eur Radiol.* 2018;28(8):3306-3317. doi:10.1007/S00330-018-5314-5/FIGURES/3
22. Rovira Á, Wattjes MP, Tintoré M, et al. MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis—clinical implementation in the diagnostic process. *Nature Reviews Neurology 2015 11:8.* 2015;11(8):471-482. doi:10.1038/nrneurol.2015.106
23. Atri A. The Alzheimer's Disease Clinical Spectrum: Diagnosis and Management. *Medical Clinics.* 2019;103(2):263-293. doi:10.1016/J.MCNA.2018.10.009
24. Benjamin EJ, Muntner P, Alonso A, et al. Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association. *Circulation.* 2019;139(10):e56-e528. doi:10.1161/CIR.0000000000000659
25. *Guidelines for the Early Management of Patients with Acute Ischemic Stroke: 2019 Update to the 2018 Guidelines for the Early Management of Acute Ischemic Stroke* .; 2020. Accessed January 17, 2023. https://www.stroke.org/-/media/Stroke-Files/Ischemic-Stroke-Professional-Materials/AIS-Toolkit/Guidelines-for-Mangaging-Patients-with-AIS-2019-Update-to-2018-Guidelines.pdf
26. Etherton MR, Barreto AD, Schwamm LH, Wu O. Neuroimaging paradigms to identify patients for reperfusion therapy in stroke of unknown onset. *Front Neurol.* 2018;9(MAY):327. doi:10.3389/fneur.2018.00327
27. Papanagiotou P, Ntaios G. Endovascular Thrombectomy in Acute Ischemic Stroke. *Circ Cardiovasc Interv.* 2018;11(1). doi:10.1161/CIRCINTERVENTIONS.117.005362
28. Lapergue B, Blanc R, Gory B, et al. Effect of Endovascular Contact Aspiration vs Stent Retriever on Revascularization in Patients With Acute Ischemic Stroke and

Large Vessel Occlusion: The ASTER Randomized Clinical Trial. *JAMA*. 2017;318(5):443-452. doi:10.1001/JAMA.2017.9644

29. Campbell BCV, Donnan GA, Lees KR, et al. Endovascular stent thrombectomy: the new standard of care for large vessel ischaemic stroke. *Lancet Neurol*. 2015;14(8):846-854. doi:10.1016/S1474-4422(15)00140-4
30. Furlan AJ. Endovascular therapy for stroke--it's about time. *N Engl J Med*. 2015;372(24):2347-2349. doi:10.1056/NEJME1503217
31. Cohen DL, Kearney R, Griffiths M, Nadesalingam V, Bathula R. Around 9% of patients with ischaemic stroke are suitable for thrombectomy. *BMJ*. 2015;351. doi:10.1136/BMJ.H4607
32. 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(5):1377-1380. doi:10.1161/STROKEAHA.116.013165
33. Jovin TG, Saver JL, Ribo M, et al. 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 (DAWN) trial methods. *Int J Stroke*. 2017;12(6):641-652. doi:10.1177/1747493017710341
34. Marks MP, Heit JJ, Lansberg MG, et al. Endovascular treatment in the DEFUSE 3 study. *Stroke*. 2018;49(8):2000-2003. doi:10.1161/STROKEAHA.118.022147
35. Albers GW, Lansberg MG, Kemp S, et al. A multicenter randomized controlled trial of endovascular therapy following imaging evaluation for ischemic stroke (DEFUSE 3). *Int J Stroke*. 2017;12(8):896-905. doi:10.1177/1747493017701147
36. Alexandre AM, Valente I, Consoli A, et al. Posterior circulation endovascular thrombectomy for large-vessel occlusion: Predictors of favorable clinical outcome and analysis of first-pass effect. *American Journal of Neuroradiology*. 2021;42(5):896-903. doi:10.3174/AJNR.A7023
37. Yeo LLL, Jing M, Bhogal P, et al. Evidence-Based Updates to Thrombectomy: Targets, New Techniques, and Devices. *Front Neurol*. 2021;12:1477. doi:10.3389/FNEUR.2021.712527/BIBTEX
38. Wollenweber F, Tiedt S, Alegiani A, Stroke BA, 2019 undefined. Functional outcome following stroke thrombectomy in clinical practice. *Am Heart Assoc*.

Accessed January 17, 2023. https://www.ahajournals.org/doi/abs/10.1161/STROKEAHA.119.026005

39. Kang DH, Jung C, Yoon W, et al. Endovascular thrombectomy for acute basilar artery occlusion: A multicenter retrospective observational study. *J Am Heart Assoc.* 2018;7(14). doi:10.1161/JAHA.118.009419
40. Liu X, Dai Q, Ye R, et al. Endovascular treatment versus standard medical treatment for vertebrobasilar artery occlusion (BEST): an open-label, randomised controlled trial. *Elsevier.* Accessed January 17, 2023. https://www.sciencedirect.com/science/article/pii/S1474442219303953
41. Bendszus M, Bonekamp S, Berge E, et al. A randomized controlled trial to test efficacy and safety of thrombectomy in stroke with extended lesion and extended time window. *International Journal of Stroke.* 2019;14(1):87-93. doi:10.1177/1747493018798558/ASSET/IMAGES/LARGE/10.1177_1747493018798558-FIG1.JPEG
42. Suh SH, Cloft HJ, Fugate JE, Rabinstein AA, Liebeskind DS, Kallmes DF. Clarifying differences among thrombolysis in cerebral infarction scale variants: Is the artery half open or half closed? *Stroke.* 2013;44(4):1166-1168. doi:10.1161/STROKEAHA.111.000399
43. Mair G, von Kummer R, Adami A, et al. Observer reliability of CT angiography in the assessment of acute ischaemic stroke: data from the Third International Stroke Trial. *Neuroradiology.* 2015;57(1):1-9. doi:10.1007/S00234-014-1441-0/FIGURES/2
44. Tung EL, McTaggart RA, Baird GL, et al. Rethinking Thrombolysis in Cerebral Infarction 2b: Which Thrombolysis in Cerebral Infarction Scales Best Define Near Complete Recanalization in the Modern Thrombectomy Era? *Stroke.* 2017;48(9):2488-2493. doi:10.1161/STROKEAHA.117.017182
45. Dargazanli C, Consoli A, Barral M, et al. Impact of modified TICI 3 versus modified TICI 2b reperfusion score to predict good outcome following endovascular therapy. *American Journal of Neuroradiology.* 2017;38(1):90-96. doi:10.3174/ajnr.A4968
46. Goyal N, Tsivgoulis G, Frei D, et al. Comparative Safety and Efficacy of Modified TICI 2b and TICI 3 Reperfusion in Acute Ischemic Strokes Treated with

Mechanical Thrombectomy. *Clin Neurosurg.* 2019;84(3):680-686. doi:10.1093/neuros/nyy097

47. D.S. L, T.G. J, C.B. M, et al. TICI reperfusion in HERMES: Success in endovascular stroke therapy. *Stroke.* 2017;48(Supplement 1). Accessed March 17, 2021. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed18&NEWS=N&AN=617461696

48. Volny O, Cimflova P, Szeder V. Inter-Rater Reliability for Thrombolysis in Cerebral Infarction with TICI 2c Category. *Journal of Stroke and Cerebrovascular Diseases.* 2017;26(5):992-994. doi:10.1016/j.jstrokecerebrovasdis.2016.11.008

49. Dargazanli C, Fahed R, Blanc R, et al. Modified Thrombolysis in Cerebral Infarction 2C/Thrombolysis in Cerebral Infarction 3 reperfusion should be the aim of mechanical thrombectomy: Insights from the ASTER trial (contact aspiration versus stent retriever for successful revascularization). *Stroke.* 2018;49(5):1189-1196. doi:10.1161/STROKEAHA.118.020700

50. di Maria F, Kyheng M, Consoli A, et al. Identifying the predictors of first-pass effect and its influence on clinical outcome in the setting of endovascular thrombectomy for acute ischemic stroke: Results from a multicentric prospective registry. *International Journal of Stroke.* 2021;16(1):20-28. doi:10.1177/1747493020923051

51. Ducroux C, Piotin M, Gory B, et al. First pass effect with contact aspiration and stent retrievers in the Aspiration versus Stent Retriever (ASTER) trial. *J Neurointerv Surg.* 2020;12(4):386-391. doi:10.1136/neurintsurg-2019-015215

52. Zaidat OO, Castonguay AC, Linfante I, et al. First pass effect: A new measure for stroke thrombectomy devices. *Stroke.* 2018;49(3):660-666. doi:10.1161/STROKEAHA.117.020315

53. Jindal G, Carvalho HDP, Wessell A, et al. Beyond the first pass: Revascularization remains critical in stroke thrombectomy. *J Neurointerv Surg.* 2019;11(11):1095-1099. doi:10.1136/neurintsurg-2019-014773

54. Zaidat OO, Castonguay AC, Linfante I, et al. First pass effect: A new measure for stroke thrombectomy devices. *Stroke.* 2018;49(3):660-666. doi:10.1161/STROKEAHA.117.020315

55. Flottmann F, Brekenfeld C, Broocks G, et al. Good Clinical Outcome Decreases with Number of Retrieval Attempts in Stroke Thrombectomy: Beyond the First-Pass Effect. *Stroke*. 2021;52(2):482-490. doi:10.1161/STROKEAHA.120.029830

56. Krishnan R, Mays W, Elijovich L. Complications of Mechanical Thrombectomy in Acute Ischemic Stroke. *Neurology*. 2021;97(20 Supplement 2):S115-S125. doi:10.1212/WNL.0000000000012803

57. Heider DM, Simgen A, Wagenpfeil G, et al. Why we fail: mechanisms and co-factors of unsuccessful thrombectomy in acute ischemic stroke. *Neurological Sciences*. 2020;41(6):1547-1555. doi:10.1007/S10072-020-04244-5/TABLES/3

58. Nie X, Leng X, Miao Z, Fisher M, Liu L. Clinically Ineffective Reperfusion After Endovascular Therapy in Acute Ischemic Stroke. *Stroke*. Published online December 7, 2022. doi:10.1161/STROKEAHA.122.038466

59. Schiphorst A ter, Charron S, Hassen W ben, et al. Tissue no-reflow despite full recanalization following thrombectomy for anterior circulation stroke with proximal occlusion: A clinical study. *Journal of Cerebral Blood Flow and Metabolism*. 2021;41(2):253-266. doi:10.1177/0271678X20954929

60. Ng FC, Coulton B, Chambers B, Thijs V. Persistently elevated microvascular resistance postrecanalization a clinical marker of no-reflow phenomenon. *Stroke*. 2018;49(10):2512-2515. doi:10.1161/STROKEAHA.118.021631

61. Kloner RA, King KS, Harrington MG. No-reflow phenomenon in the heart and brain. *Am J Physiol Heart Circ Physiol*. 2018;315(3):H550-H562. doi:10.1152/AJPHEART.00183.2018

62. Ng FC, Churilov L, Yassi N, et al. Prevalence and Significance of Impaired Microvascular Tissue Reperfusion Despite Macrovascular Angiographic Reperfusion (No-Reflow). *Neurology*. 2022;98(8):e790-e801. doi:10.1212/WNL.0000000000013210

63. Santana D, Laredo C, Renú A, et al. "Incidence and Clinico-Radiological Correlations of Early Arterial Reocclusion After Successful Thrombectomy in Acute Ischemic Stroke." *Transl Stroke Res*. 2020;11(6):1314-1321. doi:10.1007/S12975-020-00816-X/FIGURES/3

64. Li W, Ding J, Sui X, et al. Prognosis and risk factors for reocclusion after mechanical thrombectomy. *Ann Clin Transl Neurol.* 2020;7(4):420-428. doi:10.1002/ACN3.50999

65. Flottmann F, Broocks G, Faizy TD, et al. Factors Associated with Failure of Reperfusion in Endovascular Therapy for Acute Ischemic Stroke: A Multicenter Analysis. *Clin Neuroradiol.* 2021;31(1):197-205. doi:10.1007/S00062-020-00880-8/TABLES/3

66. Kim BM. Causes and solutions of endovascular treatment failure. *J Stroke.* 2017;19(2):131-142. doi:10.5853/jos.2017.00283

67. Kaesmacher J, Gralla J, Mosimann PJ, et al. Reasons for reperfusion failures in stent-retriever-based thrombectomy: registry analysis and proposal of a classification system. *AJNR Am J Neuroradiol.* 2018;39(10):1848-1853. doi:10.3174/ajnr.a5759

68. Leischner H, Flottmann F, Hanning U, et al. Reasons for failed endovascular recanalization attempts in stroke patients. *J Neurointerv Surg.* 2018;11(5):439-442. doi:10.1136/neurintsurg-2018-014060

69. Nour M, Scalzo F, neurology DLI, 2012 undefined. Ischemia-reperfusion injury in stroke. *karger.com.* Accessed January 17, 2023. https://www.karger.com/Article/Abstract/353125

70. Gauberti M, Lapergue B, de Lizarrondo SM, et al. Ischemia-reperfusion injury after endovascular thrombectomy for ischemic stroke. *Stroke.* 2018;49(12):3071-3074. doi:10.1161/STROKEAHA.118.022015

71. Mizuma A, You JS, Yenari MA. Targeting Reperfusion Injury in the Age of Mechanical Thrombectomy. *Stroke.* 2018;49(7):1796-1802. doi:10.1161/STROKEAHA.117.017286

72. Baracchini C, Farina F, Palmieri A, et al. Early hemodynamic predictors of good outcome and reperfusion injury after endovascular treatment. *AAN Enterprises.* Published online 2019. doi:10.1212/WNL.0000000000007646

73. Kim JH, Choi J il. Feasibility of rescue stenting technique in patients with acute ischemic stroke due to middle cerebral artery occlusion after failed thrombectomy: A single-center retrospective experience. *PLoS One.* 2022;17(9 September). doi:10.1371/JOURNAL.PONE.0274842

74. Chang Y, Kim BM, Bang OY, et al. Rescue stenting for failed mechanical thrombectomy in acute ischemic stroke a multicenter experience. *Stroke.* 2018;49(4):958-964. doi:10.1161/STROKEAHA.117.020072

75. Tschoe C, Coffman S, Kittel C, et al. Outcomes After Intracranial Rescue Stenting for Acute Ischemic Stroke. *Am Heart Assoc.* 2022;2(4). doi:10.1161/SVIN.121.000129

76. Renú A, Millán M, Román LS, et al. Effect of Intra-arterial Alteplase vs Placebo Following Successful Thrombectomy on Functional Outcomes in Patients With Large Vessel Occlusion Acute Ischemic Stroke: The CHOICE Randomized Clinical Trial. *JAMA.* 2022;327(9):826-835. doi:10.1001/JAMA.2022.1645

77. Martin RH, Yeatts SD, Hill MD, Moy CS, Ginsberg MD, Palesch YY. ALIAS (Albumin in Acute Ischemic Stroke) Trials: Analysis of the Combined Data From Parts 1 and 2. *Stroke.* 2016;47(9):2355-2359. doi:10.1161/STROKEAHA.116.012825

78. Saver JL, Starkman S, Eckstein M, et al. Prehospital use of magnesium sulfate as neuroprotection in acute stroke. *N Engl J Med.* 2015;372(6):528-536. doi:10.1056/NEJMOA1408827

79. Sandercock PAG. The International Stroke Trial (IST): A randomised trial of aspirin, subcutaneous heparin, both, or neither among 19 435 patients with acute ischaemic stroke. *Lancet.* 1997;349(9065):1569-1581. doi:10.1016/S0140-6736(97)04011-7

80. CAST: randomised placebo-controlled trial of early aspirin use in 20,000 patients with acute ischaemic stroke. CAST (Chinese Acute Stroke Trial) Collaborative Group. *Lancet.* (9066). doi:10.1002/CENTRAL/CN-00140599

81. Mokin M, Kass-Hout T, Kass-Hout O, et al. Intravenous heparin for the treatment of intraluminal thrombus in patients with acute ischemic stroke: a case series. *J Neurointerv Surg.* 2013;5(2):144-150. doi:10.1136/NEURINTSURG-2011-010134

82. Sobesky J. Refining the mismatch concept in acute stroke: Lessons learned from PET and MRI. *Journal of Cerebral Blood Flow and Metabolism.* 2012;32(7):1416-1425. doi:10.1038/jcbfm.2012.54

83. Pilgram-Pastor SM, Piechowiak EI, Dobrocky T, et al. Stroke thrombectomy complication management. *J Neurointerv Surg.* 2021;13(10):912-917. doi:10.1136/NEURINTSURG-2021-017349
84. Casetta I, Fainardi E, Saia V, et al. Endovascular Thrombectomy for Acute Ischemic Stroke beyond 6 Hours from Onset: A Real-World Experience. *Stroke.* Published online 2020:2051-2057. doi:10.1161/STROKEAHA.119.027974/FORMAT/EPUB
85. Nogueira RG, Diogo ;, Haussen C, et al. Stroke Imaging Selection Modality and Endovascular Therapy Outcomes in the Early and Extended Time Windows. Published online 2021. doi:10.1161/STROKEAHA.120.031685
86. Cheplygina V, de Bruijne M, Pluim JPW. Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Med Image Anal.* 2019;54:280-296. doi:10.1016/J.MEDIA.2019.03.009
87. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. *Int J Comput Vis.* 2020;128(2):336-359. doi:10.1007/s11263-019-01228-7
88. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning Deep Features for Discriminative Localization. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* 2015;2016-December:2921-2929. doi:10.48550/arxiv.1512.04150
89. Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need. *Adv Neural Inf Process Syst.* 2017;2017-December:5999-6009. doi:10.48550/arxiv.1706.03762
90. Shaw P, Uszkoreit J, Vaswani A. Self-attention with relative position representations. *NAACL HLT 2018 - 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference.* 2018;2:464-468. doi:10.18653/v1/n18-2074
91. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference.* 2018;1:4171-4186. doi:10.48550/arxiv.1810.04805

92. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. Published online October 22, 2020. doi:10.48550/arxiv.2010.11929

93. Bao H, Dong L, Piao S, Wei F. BEiT: BERT Pre-Training of Image Transformers. Published online June 15, 2021. doi:10.48550/arxiv.2106.08254

94. He K, Chen X, Xie S, Li Y, Dollar P, Girshick R. Masked Autoencoders Are Scalable Vision Learners. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2021;2022-June:15979-15988. doi:10.48550/arxiv.2111.06377

95. Xie Z, Zhang Z, Cao Y, et al. SimMIM: A Simple Framework for Masked Image Modeling. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2021;2022-June:9643-9653. doi:10.48550/arxiv.2111.09886

96. Arnab A, Dehghani M, Heigold G, Sun C, Lučić M, Schmid C. ViViT: A Video Vision Transformer. *Proceedings of the IEEE International Conference on Computer Vision*. Published online March 29, 2021:6816-6826. doi:10.48550/arxiv.2103.15691

97. Liu Z, Ning J, Cao Y, et al. Video Swin Transformer. Published online June 24, 2021:3192-3201. doi:10.48550/arxiv.2106.13230

98. Yan S, Xiong X, Arnab A, et al. Multiview Transformers for Video Recognition. Published online January 12, 2022:3323-3333. doi:10.48550/arxiv.2201.04288

99. Chen J, Frey EC, He Y, Segars WP, Li Y, Du Y. TransMorph: Transformer for unsupervised medical image registration. *Med Image Anal*. 2021;82. doi:10.1016/j.media.2022.102615

100. Wen Q, Zhou T, Zhang C, et al. Transformers in Time Series: A Survey. Published online February 15, 2022. doi:10.48550/arxiv.2202.07125

101. Jamthikar A, Gupta D, Khanna NN, et al. A low-cost machine learning-based cardiovascular/stroke risk assessment system: integration of conventional factors with image phenotypes. *Cardiovasc Diagn Ther*. 2019;9(5):420. doi:10.21037/CDT.2019.09.03

102. Deo RC. Machine learning in medicine. *Circulation*. 2015;132(20):1920-1930. doi:10.1161/CIRCULATIONAHA.115.001593

103. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. *New England Journal of Medicine*. 2019;380(14):1347-1358. doi:10.1056/nejmra1814259

104. Kamal H, Lopez V, Sheth SA. Machine learning in acute ischemic stroke neuroimaging. *Front Neurol*. 2018;9(NOV):945. doi:10.3389/fneur.2018.00945

105. Zhang H, Polson JS, Nael K, et al. Intra-domain task-adaptive transfer learning to determine acute ischemic stroke onset time. *Computerized Medical Imaging and Graphics*. 2021;90:101926. doi:10.1016/j.compmedimag.2021.101926

106. Smith SM, Jenkinson M, Woolrich MW, et al. Advances in functional and structural MR image analysis and implementation as FSL. In: *NeuroImage*. Vol 23. Academic Press; 2004:S208-S219. doi:10.1016/j.neuroimage.2004.07.051

107. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL - review. *Neuroimage*. 2012;62(2):782-790. doi:10.1016/j.neuroimage.2011.09.015

108. Mintorovitch J, Moseley ME, Chileuitt L, Shimizu H, Cohen Y, Weinstein PR. Comparison of diffusion- and T2-weighted MRI for the early detection of cerebral ischemia and reperfusion in rats. *Magn Reson Med*. 1991;18(1):39-50. doi:10.1002/mrm.1910180106

109. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2016;2016-Decem:770-778. doi:10.1109/CVPR.2016.90

110. Woo S, Park J, Lee JY, Kweon IS. CBAM: Convolutional block attention module. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2018;11211 LNCS:3-19. doi:10.1007/978-3-030-01234-2_1

111. Kingma DP, Ba JL. Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. Published online December 22, 2015. doi:10.48550/arxiv.1412.6980

112. Zhao ZQ, Zheng P, Xu ST, Wu X. Object Detection with Deep Learning: A Review. *IEEE Trans Neural Netw Learn Syst*. 2019;30(11):3212-3232. doi:10.1109/TNNLS.2018.2876865

113. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int J Comput Vis.* 2020;128(2):336-359. doi:10.1007/s11263-019-01228-7

114. Ziegler A, Ebinger M, Fiebach JB, Audebert HJ, Leistner S. Judgment of FLAIR signal change in DWI-FLAIR mismatch determination is a challenge to clinicians. *J Neurol.* 2012;259(5):971-973. doi:10.1007/s00415-011-6284-6

115. Thomalla G, Cheng B, Ebinger M, 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(11):978-986. doi:10.1016/S1474-4422(11)70192-2

116. Zhu H, Jiang L, Zhang H, Luo L, Chen Y, Chen Y. An automatic machine learning approach for ischemic stroke onset time identification based on DWI and FLAIR imaging. *Neuroimage Clin.* 2021;31:102744. doi:10.1016/j.nicl.2021.102744

117. Neumann-Haefelin T, Wittsack HJ, Wenserski F, et al. Diffusion- and perfusion-weighted MRI: The DWI/PWI mismatch region in acute stroke. *Stroke.* 1999;30(8):1591-1597. doi:10.1161/01.STR.30.8.1591

118. Bashyam VM, Doshi J, Erus G, et al. Deep generative medical image harmonization for improving cross-site generalization in deep learning predictors. *Journal of Magnetic Resonance Imaging.* 2021;55(3):908-916. doi:10.1002/jmri.27908

119. Escalard S, Gory B, Kyheng M, et al. Unknown-onset strokes with anterior circulation occlusion treated by thrombectomy after DWI-FLAIR mismatch selection. *Eur J Neurol.* 2018;25(5):732-738. doi:10.1111/ENE.13580

120. Fahed R, Lecler A, Sabben C, et al. DWI-ASPECTS (Diffusion-Weighted Imaging-Alberta Stroke Program Early Computed Tomography Scores) and DWI-FLAIR (Diffusion-Weighted Imaging-Fluid Attenuated Inversion Recovery) Mismatch in Thrombectomy Candidates: An Intrarater and Interrater Agreement Study. *Stroke.* 2018;49(1):223-227. doi:10.1161/STROKEAHA.117.019508

121. van Engelen JE, Hoos HH. A survey on semi-supervised learning. *Mach Learn.* 2020;109(2):373-440. doi:10.1007/S10994-019-05855-6/FIGURES/5

122. Arazo E, Ortego Di, Albert P, O'Connor NE, McGuinness K. Pseudo-Labeling and Confirmation Bias in Deep Semi-Supervised Learning. *Proceedings of the*

*International Joint Conference on Neural Networks*. Published online August 8, 2019. doi:10.48550/arxiv.1908.02983

123. Zhou ZH. A brief introduction to weakly supervised learning. *Natl Sci Rev*. 2018;5(1):44-53. doi:10.1093/NSR/NWX106

124. Polson JS, Zhang H, Nael K, et al. Identifying acute ischemic stroke patients within the thrombolytic treatment window using deep learning. *Journal of Neuroimaging*. 2022;32(6):1153-1160. doi:10.1111/JON.13043

125. Berthelot D, Research G, Carlini N, et al. MixMatch: A Holistic Approach to Semi-Supervised Learning. *Adv Neural Inf Process Syst*. 2019;32. Accessed December 11, 2022. https://github.com/google-research/mixmatch

126. Abuduweili A, Li X, Shi H, Xu CZ, Dou D. *Adaptive Consistency Regularization for Semi-Supervised Transfer Learning.*; 2021. doi:10.1109/CVPR46437.2021.00685

127. Polson J, Zhang H, Nael K, et al. A Semi-Supervised Learning Framework to Leverage Proxy Information for Stroke MRI Analysis. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. 2021;2021:2258-2261. doi:10.1109/EMBC46164.2021.9631098

128. Perez L, Wang J. The Effectiveness of Data Augmentation in Image Classification using Deep Learning. Published online December 13, 2017. doi:10.48550/arxiv.1712.04621

129. Fugate JE, Klunder AM, Kallmes DF. What is meant by "TICI"? *AJNR Am J Neuroradiol*. 2013;34(9):1792-1797. doi:10.3174/AJNR.A3496

130. Leischner H, Flottmann F, Hanning U, et al. Reasons for failed endovascular recanalization attempts in stroke patients. *J Neurointerv Surg*. 2019;11(5):439-442. doi:10.1136/neurintsurg-2018-014060

131. Blanc R, Redjem H, Ciccio G, et al. Predictors of the Aspiration Component Success of a Direct Aspiration First Pass Technique (ADAPT) for the Endovascular Treatment of Stroke Reperfusion Strategy in Anterior Circulation Acute Stroke. *Stroke*. 2017;48(6):1588-1593. doi:10.1161/STROKEAHA.116.016149

132. Piedade GS, Schirmer CM, Goren O, et al. Cerebral Collateral Circulation: A Review in the Context of Ischemic Stroke and Mechanical Thrombectomy. *World Neurosurg.* 2019;122:33-42. doi:10.1016/J.WNEU.2018.10.066

133. Cerebral vessel anatomy as a predictor of first-pass effect in mechanical thrombectomy for emergent large-vessel occlusion in: Journal of Neurosurgery Volume 134 Issue 2 (2020). Accessed March 16, 2021. https://thejns.org/view/journals/j-neurosurg/134/2/article-p576.xml

134. Velasco Gonzalez A, Görlich D, Buerke B, et al. Predictors of Successful First-Pass Thrombectomy with a Balloon Guide Catheter: Results of a Decision Tree Analysis. *Transl Stroke Res.* 2020;11(5):900-909. doi:10.1007/s12975-020-00784-2

135. Hofmeister J, Bernava G, Rosi A, et al. Clot-Based Radiomics Predict a Mechanical Thrombectomy Strategy for Successful Recanalization in Acute Ischemic Stroke. *Stroke.* 2020;51(8):2488-2494. doi:10.1161/STROKEAHA.120.030334

136. Qiu W, Kuang H, Nair J, et al. Radiomics-based intracranial thrombus features on CT and CTA predict recanalization with intravenous alteplase in patients with acute ischemic stroke. *American Journal of Neuroradiology.* 2019;40(1):39-44. doi:10.3174/ajnr.A5918

137. van Os HJA, Ramos LA, Hilbert A, et al. Predicting outcome of endovascular treatment for acute ischemic stroke: Potential value of machine learning algorithms. *Front Neurol.* 2018;9(SEP):784. doi:10.3389/fneur.2018.00784

138. Hilbert A, Ramos LA, van Os HJA, et al. Data-efficient deep learning of radiological image data for outcome prediction after endovascular treatment of patients with acute ischemic stroke. *Comput Biol Med.* 2019;115. doi:10.1016/j.compbiomed.2019.103516

139. Fonarow GC, Smith EE, Saver JL, et al. Improving door-to-needle times in acute ischemic stroke: The design and rationale for the American Heart Association/American Stroke Association's target: Stroke initiative. *Stroke.* 2011;42(10):2983-2989. doi:10.1161/STROKEAHA.111.621342

140. Xian Y, Xu H, Lytle B, et al. Use of Strategies to Improve Door-to-Needle Times with Tissue-Type Plasminogen Activator in Acute Ischemic Stroke in Clinical

Practice: Findings from Target: Stroke. *Circ Cardiovasc Qual Outcomes.* 2017;10(1). doi:10.1161/CIRCOUTCOMES.116.003227

141. Zhang H, Polson JS, Nael K, et al. Intra-domain task-adaptive transfer learning to determine acute ischemic stroke onset time. *Computerized Medical Imaging and Graphics.* 2021;90. doi:10.1016/j.compmedimag.2021.101926
142. Fransen PSS, Beumer D, Berkhemer OA, et al. MR CLEAN, a multicenter randomized clinical trial of endovascular treatment for acute ischemic stroke in the Netherlands: study protocol for a randomized controlled trial. *Trials.* 2014;15(1). doi:10.1186/1745-6215-15-343
143. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* 2017;77(21):e104-e107. doi:10.1158/0008-5472.CAN-17-0339/SUPPLEMENTARY-VIDEO-S2
144. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine Learning methods for Quantitative Radiomic Biomarkers. *Scientific Reports 2015 5:1.* 2015;5(1):1-11. doi:10.1038/srep13087
145. Zhang H, Polson JS, Yang EJ, et al. Predicting Thrombectomy Recanalization from CT Imaging Using Deep Learning Models. Published online June 4, 2022.
146. Wang X, Girshick R, Gupta A, He K. Non-local Neural Networks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* Published online November 21, 2017:7794-7803. doi:10.48550/arxiv.1711.07971
147. Hou R, Chang H, Ma B, Shan S, Chen X. *Cross Attention Network for Few-Shot Classification.* Vol 32.; 2019.
148. Waqas M, Li W, Patel TR, et al. Clot imaging characteristics predict first pass effect of aspiration—first approach to thrombectomy. *https://doi.org/101177/15910199211019174.* 2021;28(2):152-159. doi:10.1177/15910199211019174
149. Zhang H, Polson J, Nael K, et al. A machine learning approach to predict acute ischemic stroke thrombectomy reperfusion using discriminative MR image features. *BHI 2021 - 2021 IEEE EMBS International Conference on Biomedical and Health Informatics, Proceedings.* Published online August 10, 2021:1-4. doi:10.1109/BHI50953.2021.9508597

150. Velagapudi L, Mouchtouris N, Schmidt RF, et al. A Machine Learning Approach to First Pass Reperfusion in Mechanical Thrombectomy: Prediction and Feature Analysis. *Journal of Stroke and Cerebrovascular Diseases.* 2021;30(7). doi:10.1016/j.jstrokecerebrovasdis.2021.105796
151. di Maria F, Kyheng M, Consoli A, et al. Identifying the predictors of first-pass effect and its influence on clinical outcome in the setting of endovascular thrombectomy for acute ischemic stroke: Results from a multicentric prospective registry. *Int J Stroke.* 2021;16(1):20-28. doi:10.1177/1747493020923051
152. Rohan V, Baxa J, Tupy R, et al. Length of occlusion predicts recanalization and outcome after intravenous thrombolysis in middle cerebral artery stroke. *Stroke.* 2014;45(7):2010-2017. doi:10.1161/STROKEAHA.114.005731
153. Srivatsa S, Duan Y, Sheppard JP, et al. Cerebral vessel anatomy as a predictor of first-pass effect in mechanical thrombectomy for emergent large-vessel occlusion. *J Neurosurg.* 2020;134(2):576-584. doi:10.3171/2019.11.JNS192673
154. Hofmeister J, Bernava G, Rosi A, et al. Clot-Based Radiomics Predict a Mechanical Thrombectomy Strategy for Successful Recanalization in Acute Ischemic Stroke. *Stroke.* 2020;51(8):2488-2494. doi:10.1161/STROKEAHA.120.030334
155. Volny O, Cimflova P, Szeder V. Inter-Rater Reliability for Thrombolysis in Cerebral Infarction with TICI 2c Category. *Journal of Stroke and Cerebrovascular Diseases.* 2017;26(5):992-994. doi:10.1016/J.JSTROKECEREBROVASDIS.2016.11.008
156. Liebeskind DS, Bracard S, Guillemin F, et al. eTICI reperfusion: defining success in endovascular stroke therapy. *J Neurointerv Surg.* 2019;11(5):433-438. doi:10.1136/NEURINTSURG-2018-014127
157. Behme D, Tsogkas I, Colla R, et al. Validation of the extended thrombolysis in cerebral infarction score in a real world cohort. *PLoS One.* 2019;14(1). doi:10.1371/JOURNAL.PONE.0210334
158. Kauw F, Dankbaar JW, Martin BW, et al. Collateral Status in Ischemic Stroke: A Comparison of Computed Tomography Angiography, Computed Tomography Perfusion, and Digital Subtraction Angiography. *J Comput Assist Tomogr.* 2020;44(6):984-992. doi:10.1097/RCT.0000000000001090

159. Nielsen M, Waldmann M, Frölich AM, et al. Deep Learning-Based Automated Thrombolysis in Cerebral Infarction Scoring: A Timely Proof-of-Principle Study. *Stroke*. 2021;52(11):3497-3504. doi:10.1161/STROKEAHA.120.033807

160. Mittmann BJ, Braun M, Runck F, et al. Deep learning-based classification of DSA image sequences of patients with acute ischemic stroke. *Int J Comput Assist Radiol Surg*. 2022;17(9):1633-1641. doi:10.1007/S11548-022-02654-8/FIGURES/5

161. Su R, Cornelissen SAP, van der Sluijs M, et al. AutoTICI: Automatic Brain Tissue Reperfusion Scoring on 2D DSA Images of Acute Ischemic Stroke Patients. *IEEE Trans Med Imaging*. 2021;40(9):2380-2391. doi:10.1109/TMI.2021.3077113

162. Park S, Kim G, Oh Y, et al. Multi-task vision transformer using low-level chest X-ray feature corpus for COVID-19 diagnosis and severity quantification. *Med Image Anal*. 2022;75:102299. doi:10.1016/J.MEDIA.2021.102299

163. Zaidat OO, Yoo AJ, Khatri P, et al. Recommendations on Angiographic Revascularization Grading Standards for Acute Ischemic Stroke. *Stroke*. 2013;44(9):2650-2663. doi:10.1161/STROKEAHA.113.001972

164. Tsui B, Nour M, Chen I, et al. MR Angiography in Assessment of Collaterals in Patients with Acute Ischemic Stroke: A Comparative Analysis with Digital Subtraction Angiography. *Brain Sciences 2022, Vol 12, Page 1181*. 2022;12(9):1181. doi:10.3390/BRAINSCI12091181

165. Tönnes C, Janssen S, Golla AK, et al. Deterministic Arterial Input Function selection in DCE-MRI for automation of quantitative perfusion calculation of colorectal cancer. *Magn Reson Imaging*. 2021;75:116-123. doi:10.1016/J.MRI.2020.09.009

166. Calamante F. Arterial input function in perfusion MRI: A comprehensive review. *Prog Nucl Magn Reson Spectrosc*. 2013;74:1-32. doi:10.1016/J.PNMRS.2013.04.002

167. Nielsen M, Waldmann M, Sentker T, Frölich A, Fiehler J, Werner R. Time Matters: Handling Spatio-Temporal Perfusion Information for Automated TICI Scoring. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2020;12266 LNCS:86-96. doi:10.1007/978-3-030-59725-2_9/TABLES/2

168. Chen CF (Richard), Fan Q, Panda R. CrossViT: Cross-Attention Multi-Scale Vision Transformer for Image Classification. Published online 2021:357-366.
169. Zhao H, Jiaya C, Cuhk J, Koltun V. Exploring Self-Attention for Image Recognition. Published online 2020:10076-10085.
170. Bertasius G, Wang H, Torresani L. Is Space-Time Attention All You Need for Video Understanding? Published online February 9, 2021. doi:10.48550/arxiv.2102.05095
171. Liu Z, Lin Y, Cao Y, et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. *Proceedings of the IEEE International Conference on Computer Vision.* Published online March 25, 2021:9992-10002. doi:10.48550/arxiv.2103.14030
172. Liu Q, Xu Z, Jiao Y, Niethammer M. iSegFormer: Interactive Segmentation via Transformers with Application to 3D Knee MR Images. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* 2022;13435 LNCS:464-474. doi:10.1007/978-3-031-16443-9_45/TABLES/4
173. Lin A, Chen B, Xu J, Zhang Z, Lu G, Zhang D. DS-TransUNet: Dual Swin Transformer U-Net for Medical Image Segmentation. *IEEE Trans Instrum Meas.* 2022;71. doi:10.1109/TIM.2022.3178991
174. Tang ˚ YT, Yang D, Li W, et al. Self-Supervised Pre-Training of Swin Transformers for 3D Medical Image Analysis. Published online 2022:20730-20740. Accessed January 17, 2023. https://www.synapse.org/#!Synapse:syn3193805/
175. Hatamizadeh A, Nath V, Tang Y, Yang D, Roth HR, Xu D. Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* 2022;12962 LNCS:272-284. doi:10.1007/978-3-031-08999-2_22/TABLES/4
176. Wei X, Zhang T, Li Y, Zhang Y, Wu F. *Multi-Modality Cross Attention Network for Image and Sentence Matching.*; 2020. doi:10.1109/CVPR42600.2020.01095
177. Ueda D, Katayama Y, Yamamoto A, et al. Deep learning–based angiogram generation model for cerebral angiography without misregistration artifacts.

*Radiology.* 2021;299(3):675-681. doi:10.1148/RADIOL.2021203692/RY203692SUPPM.MP4

178. Khankari J, Yu Y, Ouyang J, et al. Automated detection of arterial landmarks and vascular occlusions in patients with acute stroke receiving digital subtraction angiography using deep learning. *J Neurointerv Surg.* Published online April 28, 2022. doi:10.1136/NEURINTSURG-2021-018638