Automatic Evaluation of Collaterals in Ischemic Stroke

Mumu Aktar

A Thesis
in
The Department
of
Computer Science and Software Engineering

Presented in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy (Computer Science) at
Concordia University
Montréal, Québec, Canada

October 2023

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: **Mumu Aktar**
Entitled: **Automatic Evaluation of Collaterals in Ischemic Stroke**

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Computer Science)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.
Signed by the Final Examining Committee:

_____ Chair
Dr. Adel M. Hanna

_____ External Examiner
Dr. Kayhan Batmanghelich

_____ External Examiner
Dr. Claudine Gauthier

_____ Examiner
Dr. Thomas Fevens

_____ Examiner
Dr. Adam Krzyzak

_____ Supervisor
Dr. Marta Kersten-Oertel

_____ Co-supervisor
Dr. Hassan Rivaz

Approved by _____
Leila Kosseim, Graduate Program Director
Department of Computer Science and Software Engineering

_____ 2023

_____
Mourad Debbabi, Dean
Gina Cody School of Engineering and Computer Science

Abstract

Automatic Evaluation of Collaterals in Ischemic Stroke

Mumu Aktar, Ph.D.
Concordia University, 2023

Ischemic stroke, caused by blocked arteries in the brain, is one of the leading causes of death and disability worldwide. Endovascular thrombectomy treatment (EVT) is one of the best treatment strategies for restoring blood flow through blocked arteries, but its success rate depends on a number of factors, including the extent of a patient's collateral circulation. Collateral circulation is a subsidiary vascular network that gets activated when the main conduits fail due to ischemic stroke. It helps viable brain tissues to get oxygen and nutrients temporarily.

Evaluation of collaterals by visual inspection of radiologists is time-consuming and prone to inter and intra-rater variability. Thus, computer-aided systems can provide more consistent and reliable assessments of collaterals. Four-dimensional computed tomography angiography (4D CTA) is a reliable method for detailed cerebral vasculature imaging, preventing inaccurate collateral estimation compared to single-phase CTA. Alongside 4D CTA, readily available non-contrast computed tomography (NCCT) serves as a frontline diagnostic tool, free from contrast agents' potential adverse effects. Hence, we propose computer-aided systems for automatic collateral evaluation in ischemic stroke using 4D CTA and NCCT imaging.

We propose an automatic quantification method considering low-rank decomposition, a classic machine learning (ML) method as well as deep learning (DL) methods for the automatic evaluation of collaterals. DL models, while capable of automatic feature extraction unlike classic ML models, face challenges due to limited ischemic stroke data. To overcome data scarcity and class imbalance, we employ transfer learning with focal loss and Siamese network. Furthermore, for efficient 3D vasculature segmentation without extensive slice annotation, we introduce few-shot learning for cerebral blood vessel segmentation which can be a preprocessing step to collateral evaluation.

Acknowledgments

I would like to express my deepest gratitude to my supervisors, Professor Marta Kersten-Oertel and Professor Hassan Rivaz, for their invaluable guidance and unwavering support throughout the entirety of my research journey. Their expertise, encouragement, and mentorship have been instrumental in shaping this thesis and my academic growth. Your constructive feedback and willingness to share your knowledge have always been helpful in enhancing the quality and depth of my research. Dr. Marta Kersten-Oertel and Dr. Hassan Rivaz are very friendly, patient, encouraging, and supportive on the personal side. Not only did they support me in my research, but they also provided valuable mental support during challenging times. This work would not be possible without their help and support.

I am grateful to Professor Yiming Xiao for his invaluable mentorship and guidance throughout my academic journey. Dr. Yiming Xiao's wisdom and support greatly enriched my research experience.

I am grateful to my colleagues from the Applied Perception Lab and IMPACT Lab, including Ashikuzzaman Rasel, Jonatan Reyes, and Ali K. Z. Tehrani for their valuable suggestions, ideas, and collaborations on various thesis projects. Special thanks to Negar Haghbin for her immense support during my challenging transition to Montreal.

I thank my thesis committee members Dr. Thomas Fevens, Dr. Adam Krzyzak, Dr. Claudine Gauthier, Dr. Kayhan Batmanghelich for their constructive comments.

I want to deeply thank my parents for their unwavering love, support, sacrifices, and belief in me. Without them, I could not have achieved this academic milestone. Their love and support mean everything to me. I cannot thank my husband, Md Shamimur Rahman Shuvo, enough for his continuous support and help. Thank you for being my strength several times when I felt I should quit. This journey could be the hardest one without your support. I want to express my immense love to my precious daughter, Samayra Rahman Manha, who came into this world during this challenging PhD journey. I am sorry that you had to spend much of your early years in daycare when you might have wished to be with your mummy. But I am confident that as you grow up, you will be proud of your mother. You have always been my inspiration and the source of encouragement that helped me face all my struggles. Mummy loves you deeply.

I would also like to thank my younger brother, Iftee, and my elder sister, Mitu Apu for their support and understanding during my studies and for taking pride in my achievements which always inspired me. Their encouragement and companionship made this journey more meaningful. Additionally, I cannot ignore the adorable presence of my nephew, Prottoy, who is the love of my life. I am grateful to my brother-in-law who helped me make the right decisions for my career path and always being inspiring and feeling proud of my achievements. Last but not least, I want to extend my heartfelt gratitude to my father and mother-in-law for their love and support.

Author Contributions

As the first author of all the five manuscripts included in this dissertation, I played a central role in designing the research framework, writing codes, conducting experiments, analyzing results, and writing manuscripts. The co-authors were involved through supervision, providing ideas, discussing problem settings and technical issues, and reviewing the manuscripts.

1. **Chapter 3: Automatic Collateral Circulation Scoring in Ischemic Stroke using 4D CT Angiography with Low-Rank and Sparse Matrix Decomposition.** International Journal of Computer-Assisted Radiology and Surgery (2020). https://doi.org/10.1007/s11548-020-02216-w.

    - Authors: Aktar M.*, Tampieri D., Rivaz, H., Kersten-Oertel, M., Xiao Y.
    - Contributions: guarantors of the integrity of the study: all authors; study and design concept: Aktar M.*, Rivaz, H., Kersten-Oertel, M., Xiao Y. ; software development: Aktar M.* ; data provider: Tampieri D.; supervision: Rivaz, H., Kersten-Oertel, M., Xiao Y. manuscript preparation: Aktar M.* and manuscript revision: Aktar M.* , Rivaz, H., Kersten-Oertel, M., Xiao Y.

2. **Chapter 4: Deep learning for collateral evaluation in ischemic stroke with imbalanced data.** International Journal of Computer-Assisted Radiology and Surgery. 2023 Apr;18(4):733-40.

    - Authors: Aktar M.*, Reyes J., Tampieri D., Rivaz, H., Xiao Y., Kersten-Oertel, M.
    - Contributions: guarantors of the integrity of the study: all authors; study and design concept: Aktar M.*, Rivaz, H., Xiao Y., Kersten-Oertel, M.; software development: Aktar M.*, Reyes J.; data provider: Tampieri D.; supervision: Rivaz, H., Xiao Y., Kersten-Oertel, M. . manuscript preparation: Aktar M.* and manuscript revision: Aktar M.*, Reyes J., Rivaz, H., Kersten-Oertel, M., Xiao Y.

3. **Chapter 5: A Radiomics-based Machine Learning Approach to Assess Collateral Circulation in Ischemic Stroke on Non-contrast Computed Tomography** Multimodal Learning for Clinical Decision Support and Clinical Images-Based Procedures (CLIP2020).

    - Authors: Aktar M.*, Xiao Y., Tampieri D., Rivaz, H., Kersten-Oertel, M.,
    - Contributions: guarantors of the integrity of the study: all authors; study and design concept: Aktar M.*, Rivaz, H., Kersten-Oertel, M., Xiao Y.; software development: Aktar M.*; data provider: Tampieri D.; supervision: Rivaz, H.,

Kersten-Oertel, M., Xiao Y. manuscript preparation: Aktar M.* and manuscript revision: Aktar M.* , Rivaz, H., Kersten-Oertel, M., Xiao Y.

4. **Chapter 6: SCANED: Siamese Collateral Assessment Network for Evaluation of Collaterals from Ischemic Damage.** Submitted to Computerized Medical Imaging and Graphics, 2023

    - Authors: Aktar M.*, Xiao Y., Tampieri D., Rivaz, H., Kersten-Oertel, M.,
    - Contributions: guarantors of the integrity of the study: all authors; study and design concept: Aktar M.*, Xiao Y., Rivaz, H., Kersten-Oertel, M.; software development: Aktar M.*, Tehrani A. K. Z.; data provider: Tampieri D.; supervision: Rivaz, H., Xiao Y., Kersten-Oertel, M. . manuscript preparation: Aktar M.* and manuscript revision: Aktar M.*, Rivaz, H., Kersten-Oertel, M., Xiao Y.

5. **Chapter 7: VesselShot: Few-shot learning for cerebral blood vessel segmentation.** Machine Learning in Clinical Neuroimaging workshop in MICCAI. 2023

    - Authors: Aktar M.*, Rivaz, H., Kersten-Oertel, M., Xiao Y.,
    - Contributions: guarantors of the integrity of the study: all authors; study and design concept: Aktar M.*, Rivaz, H., Kersten-Oertel, M., Xiao Y.; software development: Aktar M.*; supervision: Rivaz, H., Kersten-Oertel, M., Xiao Y.; manuscript preparation: Aktar M.* and manuscript revision: Aktar M.*, Rivaz, H., Kersten-Oertel, M., Xiao Y.

Contents

List of Figures

List of Tables

List of Abbreviations

EVT	Endovascular thrombectomy treatment
4D CTA	4-dimensional computed tomography angiography
NCCT	Non-contrast computed tomography
IV-tPA	Intravenous tissue plasminogen activator
tPA	Tissue plasminogen activator
ACA	Anterior cerebral artery
PCA	Posterior cerebral artery
MCA	Middle cerebral artery
sHT	Symptomatic hemorrhagic transformation
IVT	Intravenous thrombolysis
DSA	Digital subtraction angiography
T1-CTA	Timing-invariant CTA
ASPECTS	Alberta Stroke Program Early CT Score
ASITN/SIR	Society of NeuroInterventional Surgery (formerly the American Society of Interventional and Therapeutic Neuroradiology)/ Society of Interventional Radiology
dCTA	Dynamic CTA
ROC	Receiver operating characteristics
AUC	Area under the curve
TIA	Transient ischemic attack
CS	Collateral score
CBS	Clot burden score
CTP	CT perfusion
MTT	Mean-transit-time
CBV	Cerebral blood volume
CC	Collateral circulation
mCTA	Multiphase CTA
mMCAi	Malignant middle cerebral artery infarction
NIHSS	National Institutes of Health Stroke Scale
rLMC	Regional leptomeningeal collateral
CI	Confidence interval
CGS	Collateral grading score
SVM	Support vector machine
DWI	Diffusion-weighted imaging
RF	Random forest
CNN	Convolutional neural network
KNN	K-nearest neighbor

Chapter 1

Introduction

1.1 Ischemic Stroke

Stroke is one of the leading causes of disability as well as death worldwide. Statistics from the World Heart Federation show that each year, 15 million people suffer from stroke, among which 5 million become permanently disabled and 6 million people die [1]. People can suffer from two kinds of stroke: ischemic and hemorrhagic. Ischemic stroke occurs due to blood clots that are either formed in a cerebral artery of the brain or travel from somewhere else in the blood system. In contrast, in hemorrhagic stroke, a cerebral artery ruptures and bleeds into the brain.

Ischemic stroke is much more frequent, with 8 out of 10 stroke cases suffering from it. When diagnosis and treatment of stroke are not performed in time, patients become disabled due to a lack of blood and oxygen flow, which causes neuronal cell death in the affected part of the brain. Recanalization, which restores blood flow to the affected part of the brain, is the most important modifiable predictor in ischemic stroke for positive clinical outcome [5].

1.2 Treatment Strategies of Ischemic Stroke

Ischemic stroke treatments vary between intravenous tissue plasminogen activator (IV-tPA) and endovascular thrombectomy treatment (EVT). The treatment strategy is based on the time window and the patient's condition. A review of ischemic treatment strategies starting from thrombolysis to thrombectomy was presented in the study of Sharma *et al.* [6]. Clinical studies led by the National Institute of Neurological Disorders and Stroke in 1995 provided convincing evidence that recombinant tissue plasminogen activator (tPA) can treat patients and achieve neurological recovery [7]. However, recombinant tPA often fails to treat patients with ischemic stroke because it needs to be administrated within 3 hours of symptom onset. Patients often fail to meet this requirement, resulting in recombinant tPA failing to break large clots and sometimes causing bleeding into the brain. The Third European Cooperative Acute Stroke Study III [8] in 2008 and the American Heart Association/American Stroke Association [9] in 2009 have reported that treatment with IV-tPA can be extended to 4.5h of stroke symptoms onset. Unfortunately, most patients do not reach hospitals within 4.5h of onset and thus are not eligible for the IV-tPA as it can lead to hemorrhagic transformation.

[1] http://www.world-heart-federation.org/cardiovascular-health/stroke/

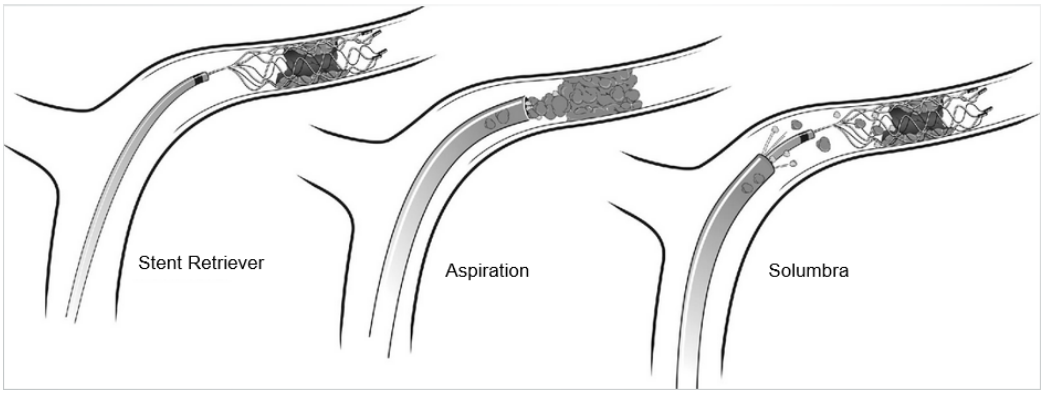

Figure 1.1: Techniques for mechanical thrombectomy. A stent retriever alone (left) is deployed within the clot, then retrieved. With the direct aspiration technique (center), an aspiration catheter is brought to the clot interface, and suction is initiated. The "Solumbra" technique (right) involves the use of a stent retriever with concomitant aspiration [11]. This image is from Wolters Kluwer Health, Inc. publisher under the license number: 5670830072701.

In 2017, a multi-centered clinical trial, the "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," or simply the DAWN trial [10], found that selected patients can be treated within 6h-24h of symptom onset by endovascular treatment. The endovascular treatment uses a catheter with a mechanical device attached to the tip to remove the clot. Different mechanical thrombectomy techniques can be applied in this strategy, as shown in Fig. 1.1.

Mechanical thrombectomy techniques allow blood, as well as oxygen and nutrients that flow through the blood system, to be restored and reduce disability quickly. However, not all stroke patients are suitable candidates for endovascular treatment due to its associated risks. A subgroup of patients with specific indications, including small infarct volume, large penumbra size, and sufficient *collateral circulation*, should undergo such interventions [5].

1.2.1 Collateral Circulation

Collateral circulation is a dynamic vascular network activated as an alternative blood flow path when an occlusion or clot hinders primary blood circulation. It provides nutrients and the oxygen necessary for the survival of neurons in the failure of the primary conduits due to ischemic stroke. Collateral circulation in the brain can be distinguished into two routes: the primary route is the Circle of Willis, which anastomoses the anterior cerebral artery (ACA) with the posterior cerebral artery (PCA) and the nearby main cerebral arteries to each other; the secondary route includes all external to internal carotid artery connections. One of the essential pial arterioles of the latter route is the leptomeningeal collaterals, which connect the middle cerebral artery (MCA) territories with ACA and PCA [12]. The extent of collateral flow through these routes is proved as a radiologic surrogate predicting the response of revascularization therapy. Fig. 1.2 shows the Circle of Willis along with other main cerebral arteries.

[2]https://creativecommons.org/licenses/by-nc-nd/4.0/

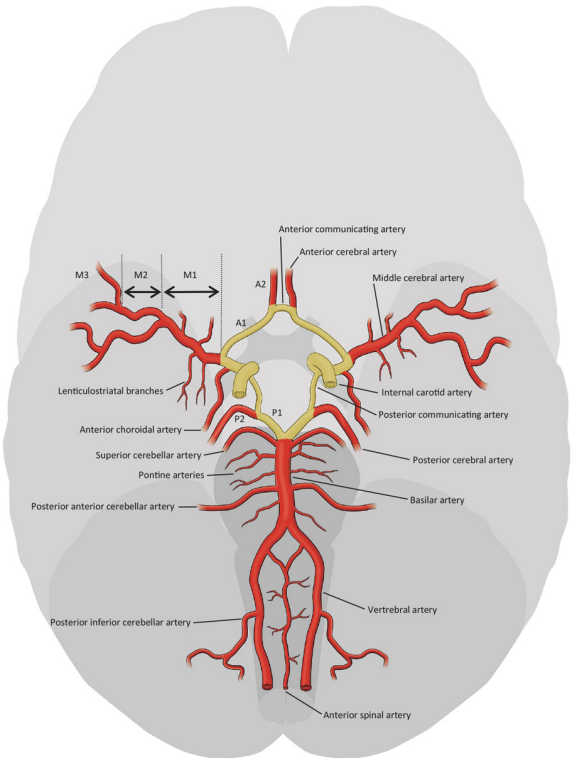

Figure 1.2: Intracranial vessels. Yellow: Circle of Willis. This image is adapted from [12] which is under the license[2].

Fig. 1.3 shows the collateral circulation network with good and poor collaterals. As described above, patients with good collateral circulation are strong candidates for revascularization in acute ischemic stroke because collateral circulation allows for tissue viability and leads to a positive response and favorable functional outcomes in reperfusion therapies. In contrast, patients recanalized with poor collaterals can be affected by symptomatic hemorrhagic transformation (sHT), which may lead to death caused by reperfusion injury [13]. EVT is one of the most potent treatments for ischemic stroke, which can provide improved functional outcomes and reduce mortality, but only in patients with good and sometimes moderate collateral status. The benefits are less or sometimes even absent in the case of patients with poor collaterals [14][15][16].

1.2.2 The Ischemic Penumbra

In ischemic stroke, there are two zones of injury: (1) the core zone where a severe injury occurs and blood and nutrient flow are below 10% to 20%, and (2) the penumbra which is mild to moderately ischemic (i.e. there is a deficient supply of blood) and lies in the area between the normally perfused tissue and the area of infarction (dead tissue due to oxygen and nutrients lacking) (see Fig. 1.4). While approximately two million neurons die each minute around the occluded artery, which enlarges the infarct growth, the penumbral zone can remain viable for several hours while collateral circulation supplies the necessary nutrients and oxygen [12]. Ischemic stroke reperfusion therapies by EVT and intravenous

[3]http://creativecommons.org/licenses/by/4.0/

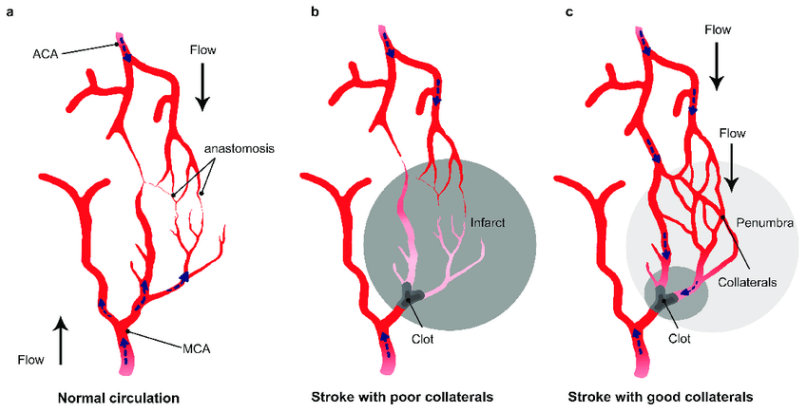

Figure 1.3: **Left:** Schematic drawing of the collateral network showing anastomoses between the MCA and ACA. **Middle:** In a stroke patient with a poor collateral network, the collaterals fail to fill and insufficiently compensate for the flow reduction after arterial occlusion, leading to neuronal death. **Right:** A collateral enhancement is occurring in patients showing a good collateral network. The flow in the collaterals changes direction and allows the thrombolytic to reach the drug from different sides. Image is taken from [17] which is under the license[3].

thrombolysis (IVT) are based on this penumbral zone. But if the timely revascularization of the occluded artery is not performed by EVT or IVT, the penumbra will be at risk of irreversible infarction as collaterals cannot flow indefinitely [18].

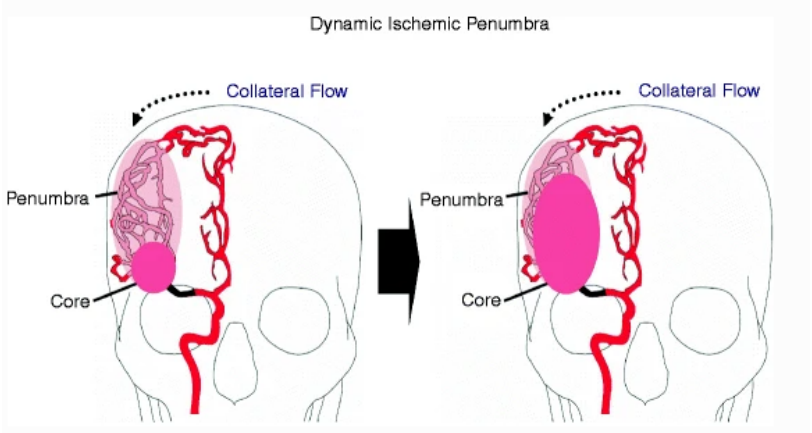

Figure 1.4: The dynamic ischemic penumbra. Shortly after the MCA occlusion (left) two regions of brain become manifest: the small core and the larger penumbra. With the passage of time (right), there is shrinkage of the ischemic penumbra and a corresponding growth of the core[4]. This image is used through the agreement to Springer Nature with the license number: 5670840527687.

The whole penumbra concept and reperfusion therapy depend on the energy supply by the collateral network, which gets activated quickly after occlusion in brain ischemia. Although the extent of collaterals varies significantly between individuals, Fig. 1.5 depicts

[4]https://sbrsport.me/2017/08/01/hyperbaric-therapy-stroke-recover

(using digital subtraction angiography (DSA) imaging) how, in general, collateral networks start working in ischemic stroke. After introducing the contrast agent (Fig. 1.5 A1) into the internal carotid artery, the proximal (closer to the occlusion) MCA occlusion marked by a red arrow can be seen. The activation of leptomeningeal collaterals (i.e., collaterals within the two innermost layers of the meninges) can be seen after 2.5 seconds in Fig. 1.5 A2, indicating it is fed by the anterior (above and right of the blue line) and posterior circulation (below the blue line). Collateral flow fills the superior (red dots) and inferior branch (green dots) of the occluded MCA after 5.2 sec (Fig. 1.5 A3 and A3 zoomed). This alternate flow direction is named retrograde filling. The MCA retrograde filling is shown in Fig. 1.5 (B). After reperfusion therapy, the branches filled retrogradely with collaterals previously are filled anterogradely (the regular direction) (Fig. 1.5 C).

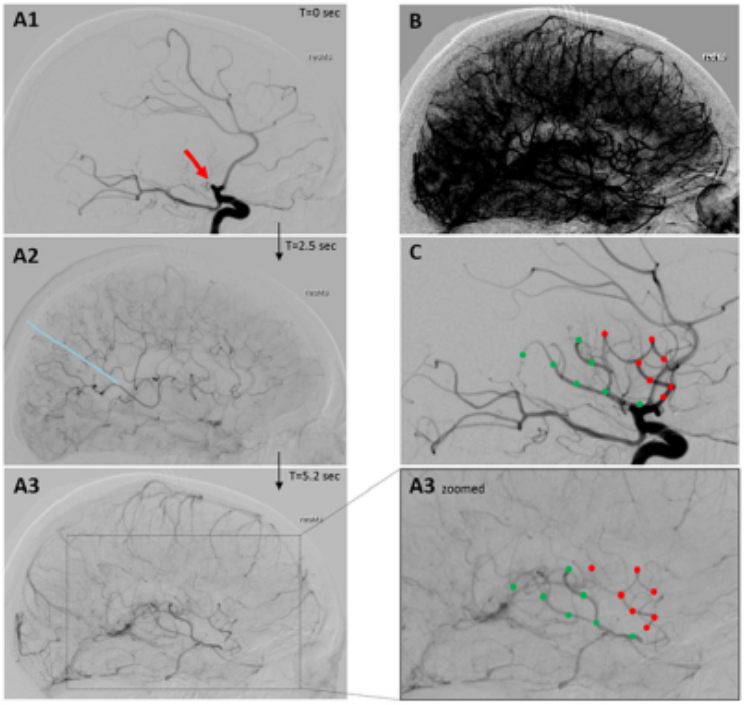

Figure 1.5: Visualization of leptomeningeal collaterals in the occluded MCA. This image is taken from [12] which is under the license[5].

1.3 4D Computed Tomography Imaging

Diversified imaging methods have been used for the visualization and assessment of collaterals like transcranial doppler (TCD), transcranial color-coded duplex sonography (TCCD), traditional single-phase computed tomography angiography (CTA), timing-invariant CTA (TI-CTA) and multiphase CTA (mCTA) or dynamic CTA (dCTA), triphase CT perfusion imaging (CTP), magnetic resonance imaging (MRA), phase-contrast MRA, quantitative MRA (QMRA), and digital subtraction angiography (DSA), in both clinical practices and relevant research [19]. CT imaging is the combination of X-rays and computed tomography, which can be captured without contrast or with contrast, in a single-phase or

[5]https://creativecommons.org/licenses/by-nc-nd/4.0/

multi-phases, capturing the peak, venous, and late venous phases. CTA gives detailed pictures of the brain's blood vessels through contrast injection and CT scanning. DSA is post-processed from conventional angiographic images, where a pre-contrast mask image is subtracted from the post-contrast images to obtain only the blood vessels removing the extraneous structures. Although DSA is considered as a gold standard for the assessment of collaterals, it is avoided as it is invasive and costly. Four-dimensional computed tomography angiography (4D CTA) is a non-invasive method, and it gives detailed hemodynamic information. 4D CTA is advantageous over single-phase CTA because it captures the delayed blood flow along with the regular flow by performing scanning over multiple volumes from the arterial to the slower venous phase. Because of the advantages of 4D CTA over the other imaging methods, we focused on it for our research.

4D CTA is also known as TI-CTA for the combination of its characteristics like non-invasive conventional CTA with a dynamic nature like catheter DSA. To capture the magnitude as well as the directional flow of vessels through 4D CTA, various techniques, such as toggling-table, shuttle, or volume mode scanning, can be used to obtain the whole brain coverage. Although the decision of choosing acquisition mode to ensure adequate brain coverage depends on the CT detector width, volume mode is the most versatile option to allow complete or partial coverage of the whole brain during one rotation [20]. This volume mode acquisition of dynamic 4D CTA can be continuous or discontinuous. The continuous acquisition scanning is performed continuously during a pre-specified period of time and can be reconstructed at any time interval, whereas for discontinuous acquisition, the time interval typically ranges from 1 to 4sec, and finally, all the acquisitions are overlaid to obtain the 4D CTA images. To image the vasculature, a contrast bolus is delivered intravascularly and then imaged in real-time when flowing through the arterial area of interest. Although it depends on the acquisition protocols, normally, 19 volumes are captured while scanning a subject to obtain a 4D CTA, with the first volume obtained before the contrast agent (which is non-contrast computed tomography (NCCT)).

4D CTA can be reconstructed from the CTP dataset, which captures multiple CT datasets in different time intervals following the injection of intravenous contrast [21]. Thus, 4D CTA contains both the vasculature and cerebral perfusion information. MIPs can also be constructed from multiple volumes of the 4D CTA data to provide an overview of the whole temporal information of the vasculature in one 3D volume. A disadvantage of 4D CTA is radiation dose, which is accumulated for each additional volume compared to a single-phase CTA. However, this can be reduced by faster gantry rotation time at identical milliampere-second settings, which leads to the increase of noise [20] but can be mitigated using filtering approaches that reduce noise.

1.4 Motivation

EVT is one of the best treatment strategies for restoring blood flow through blocked arteries. Still, its success rate depends on a number of factors, including the extent of a patient's collateral circulation. As described above, collateral circulation works as a radiologic surrogate predicting the response of revascularization therapy, which temporarily helps viable brain tissues get oxygen and nutrients. Assessment of collateral circulation in ischemic stroke, which can identify patients for the most appropriate treatment strategies, is currently conducted with visual inspection by a radiologist. Yet, numerous studies have shown that visual inspection suffers based on the experience, training, and specialty of

radiologists, as well as, intra- and inter-rater variability. This leads to time-consuming, inefficient results, which might affect the treatment decision adversely. Intra- and inter-raters' variability is shown to be an issue in several studies (e.g. [22, 23, 24]), among which Grotta *et al.* [22] concluded that it is difficult to get the agreement in recognizing and quantifying early ischemic changes even by experienced clinicians. A recent study by Grunwald *et al.* [25] showed that between individual neuroradiologists, the intraclass correlation coefficient ranges from 0.42 to 0.86 and score agreements range from 36.2% to 81.6%.

In recent years, automatic computer-based collateral grading techniques have become an active area of research to mitigate the shortcomings of visual assessment of collateral evaluation. Automatic approaches analogous to the radiologists' grading criteria can evaluate collaterals more robustly than human-rater scoring and are more easily understandable in the clinical environment. Computer-aided decision support systems by deep learning can improve performance while a large dataset is available. Unfortunately, ischemic stroke datasets are rare and imbalanced, affecting the classification performance. The objective of our research is thus to develop systems for automatic assessment of collateral scoring in ischemic stroke using computer-based approaches while mitigating the issues of small and imbalanced datasets.

1.4.1 Contributions

Our research aims to develop novel automated quantitative approaches, specifically machine learning and deep learning-based approaches for collateral evaluation in ischemic stroke, which mitigate the problems that arise through visual assessment. Traditional machine learning (ML) methods have proven effective in achieving generalizability and exemplary performance in collateral assessment. However, they require extensive manual feature engineering, which can be time-consuming and demanding in domain expertise. In contrast, deep learning (DL) models can automatically extract relevant features from data, reducing the need for manual intervention. However, one of the primary challenges with DL is the requirement for abundant training data. Unfortunately, there is a scarcity of open and large labeled stroke datasets, especially those focusing on collaterals.

In this dissertation, we developed computer-aided decision support algorithms for collateral evaluation in ischemic stroke. Our contributions are centered around the same 4D CTA dataset, considering two key phases: (1) 2D images from 3D MIPs of the 4D CTA and (2) NCCT extracted from the 4D CTA before the contrast agent.

1. Collateral evaluation based on 4D CTA:
    - We have devised a novel method for quantitative collateral scoring, emphasizing the radiological assessment of filled vessels in the affected area versus the unfilled vessels over time, using 4D CTA. The ratio of filled to unfilled collaterals constitutes the final score in this approach. (Chapter 3)
    - Recognizing the potential of deep learning in automating feature extraction, we have implemented a deep learning-driven automatic evaluation system using 4D CTA. This approach leverages knowledge transfer from a pre-trained network to alleviate the substantial manual engineering typically associated with classical ML and quantitative methods. (Chapter 4)
2. Collateral evaluation based on NCCT:

- Given that radiologists compare an ischemic patient's affected and unaffected sides to determine collateral scores, we have introduced an approach that enhances the efficacy of the automated evaluation through machine learning. This method focuses on the radiomic features extracted from ischemic damage using NCCT images of both sides of the brain. (Chapter 5)
- We have developed a technique employing Siamese networks to validate further our previous method, which addresses the challenges of small and imbalanced NCCT datasets in collateral evaluation. This approach enhances adaptability to data-scarce medical tasks. (Chapter 6)

3. Segmentation of cerebrovasculature:

    - The extraction of brain vasculature can be an important prerequisite for improved collateral evaluation. However, accurate segmentation typically demands extensive label annotation, especially in 3D scenarios, which is both time-consuming and labor-intensive and requires domain expertise. Furthermore, pre-trained weights for deep learning models in the medical domain are often limited.
    - To address these challenges, we have introduced a 3D blood vessel segmentation approach using few-shot learning. This method requires only a few annotated samples to segment the entire 3D brain vasculature. It can be utilized as a pre-processing step to enhance collateral evaluation. (Chapter 7)

In summary, our work presents novel automated collateral evaluation methods, effectively mitigating inter and intra-rater variability issues. While existing methods primarily focus on automating collateral assessment, our contributions stand out by introducing innovative techniques for identifying collateral scores from ischemic damage, whether observed in NCCT or 4D CTA. These approaches exhibit robustness even with limited and imbalanced data, and they harness the power of transfer learning to make use of pre-trained weights. Additionally, applying Siamese networks extends our methodology to the broader context of similarity-based problem-solving with minimal data requirements. Furthermore, our cerebrovasculature segmentation technique offers versatility beyond collateral scoring, contributing to performance enhancement in various applications.

1.5 Thesis Overview

The thesis organization is as follows. Chapter 2 highlights the existing methods of collateral evaluations performed both manually and automatically. Further, in Chapter 3, 4, 5, 6 and 7, the developed methods are described. Chapter 3 presents an automatic collateral evaluation approach using the low-rank and sparse decomposition method with 4D CTA. In Chapter 4, the deep learning-based method utilizing transfer learning to handle imbalanced data in automatic collateral evaluation is described. Chapter 5 and chapter 6 represent sequentially the radiomic-based machine learning approach and Siamese deep learning-based method for automatic collateral evaluation using NCCT with small imbalanced data. Chapter 7 describes the few-shot learning-based cerebrovasculature segmentation approach with a few annotated 3D data, which is an important prerequisite in collateral evaluation. Chapter 8 summarizes the thesis and comments on future research directions.

Chapter 2

Related Work

In the following chapter, we describe the methodologies used in assessing collaterals as discussed in the existing literature. Our primary focus is a comprehensive exploration of manual collateral grading systems. We give details of existing methods, their associated evaluation tools, and the merits and drawbacks of each. We categorize the described techniques into two groups: manual methods, which involve human raters utilizing visual inspection, and automatic methods, which use computer algorithms. Detailed discussions on automated methods will be elaborated on in subsequent chapters as new approaches emerge that are relevant to our research.

2.1 Clinical Techniques for Collateral Grading

As EVT and IVT treatments for ischemic stroke only result in positive outcomes for a subgroup of patients with good to moderate collateral scores, grading the extent of collateral circulation is an important factor before choosing a treatment strategy. Different reliable approaches are used for the grading, e.g., ASPECTS [26], the collateral score of the Society of NeuroInterventional Surgery (formerly the American Society of Interventional and Therapeutic Neuroradiology (ASITN)/Society of Interventional Radiology (SIR) based on conventional angiography [27], the scores of Christoforidis *et al.* [28], the Miteff System [29], the Mass System [30], the modified Tan Scale [31], regional leptomeningeal collateral (rLMC) score [32], and ASPECTS on collaterals (not to be confused with ASPECTS on ischemic parenchymal changes in non-enhanced CT) [33].

Among all these approaches, ASPECTS is one of the most reliable, systematic, and robust methods to have a favorable clinical outcome. Several studies have examined ASPECTS for scoring collaterals because of their reliability. It considers 10 regions in the MCA territory, and the ischemic change is considered in each region, and one point is subtracted based on focal swelling or parenchymal hypoattenuation compared to the relative region of the contralateral side (See Figure 2.1). An ASPECTS score of 7 or less defines the disability or death.

[1]Micheau A, Hoa D, e-Anatomy, www.imaios.com, DOI: 10.37019/e-anatomy

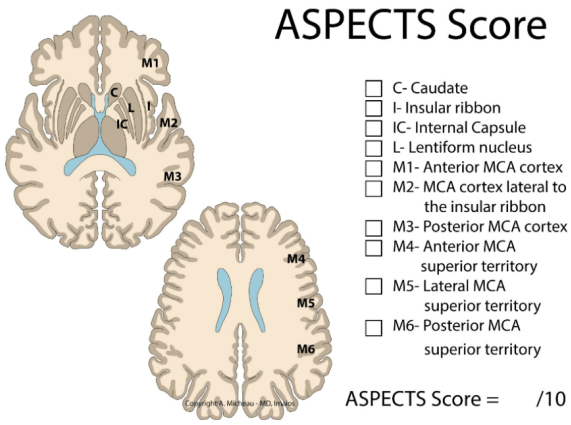

Figure 2.1: ASPECTS Score in acute stroke[1]

2.2 Collateral Assessment by Visual Inspection

Collateral circulation has played an important role in clinical trials where it has been used to determine treatment strategies as well as predict cerebral infarction growth and extent. Kim *et al.* [34] graded collaterals in acute symptomatic MCA occlusion using 15-point regional grading criteria guided by ASPECTS [26] to predict the anatomic infarction. This angiography-based collateral grading system was highly predictive of anatomic infarction using a receiver operating characteristic (ROC) curve; the area under the curve (AUC) was found to be 0.87 (95% CI: 0.83 to 0.91) for 42 patients. The authors concluded that a higher extent of collaterals results in lower infarction and higher clinical outcomes. Cerebral collaterals can alter the risk of stroke and transient ischemic attack (TIA) (i.e., temporary period of similar symptoms to stroke) in symptomatic severe internal carotid artery (ICA) stenosis due to improved cerebral perfusion where the collaterals are graded as present to indicate the appearance of collateral pathways and absent if they are not visualized on the angiogram [35].

Along with collateral score (CS) (scores of 0-3 based on the collateral extent), Tan *et al.* [31] showed that clot burden score (CBS), which is a number from 0-10 that is based on the extent of thrombus (i.e., blood clot) found in the proximal anterior circulation, can be a significant predictor of clinical outcome. This study achieved a good inter-observer correlation for CBS (ICC, 0.97; 95% CI, 0.95–0.98) and CS (ICC, 0.87; 95% CI, 0.80–0.91) and found that both collateral circulation and clot extent can help to predict stroke outcome. A recent study by Alves *et al.* [36] also stated that a higher CBS ($B = 0.063; 95\%$ confidence interval (CI), 0.008˘0.118) as well as higher thrombus attenuation increase ($B = 0.014; 95\%$ CI, 0.003˘0.026) (i.e., the permeability of letting more fluids passing through the thrombus) are associated with higher collateral score (based on a study with 192 subjects).

Miteff *et al.* [29] proved the influence of collateral status in penumbra tissue, delineated by CT perfusion scanning (CTP), to maintain the penumbra until reperfusion takes place. In a study on collateral grading with two observers who had an agreement in 88 out of 92 patients ($k = 0.93$), they found that although a large CTP mismatch ratio (mean-transit-time (MTT) lesion/cerebral blood volume (CBV) lesion) can be a pre-requisite for clinical outcome, only collateral status can determine functional outcomes. The independent predicting power of collateral circulation (CC) was further evaluated by Flores *et al.* [37] using multiphase CTA (mCTA). This study showed a relationship between malignant

middle cerebral artery infarction (mMCAi) and poor CC (using a univariate analysis). Furthermore, a multivariate analysis showed a relationship between mMCAi and age, vessel occlusion, baseline National Institutes of Health Stroke Scale (NIHSS), and recanalization. Both investigations proved that only poor collateral circulation can predict mMCAi with an odds ratio of 9.72 (95% CI, 1.387˜92.53; $P = 0.048$).

Although multiple studies have shown that collateral circulation can be an independent predictor of functional outcomes, the study by Saarinen *et al.* [38] emphasized thrombus location along with collateral circulation ($P = 0.003$ and $P = .001$). This study used 5-scale grading to assess collaterals and hypothesized that proximal occlusions result in poorer collateral filling compared to distal occlusions. Not only does thrombus location play a significant role in the functional outcome of ischemic stroke, but also the origin of leptomeningeal collateral flow. The study by Menon *et al.* [39] suggested that the interterritorial leptomeningeal collaterals originated from PCA-MCA resulted in better functional outcomes than those originating from ACA-MCA territories. This study, which performed collateral grading based on dCTA, considered the anatomical extent and prominence of pial arteries compared to the opposite side and the time of retrograde filling of the collateral circulation from the two origins, PCA-MCA and ACA-MCA. The significance of leptomeningeal collaterals maintaining tissue viability has also been proved in the study of Menon *et al.* [40], which has shown collaterals as the significant predictor of clinical outcomes ($p < .05$) and a strong correlation (interrater reliability with an intraclass correlation coefficient of 0.87, (95%CI, $0.77\% - 0.95\%$) is obtained between the functional outcome in endovascular treatments with good and moderate collateral score using the collateral grading based on regional leptomeningeal collateral (rLMC) scores [32].

In a later study by Menon *et al.* [33], ASPECTS was used for evaluating collaterals on mCTA, where six-point grading was employed to score collaterals based on the comparison between ipsilateral and contralateral hemispheres. The vasculature captured in the peak arterial phase, peak venous, and late venous phase in the mCTA was used to score the collaterals and showed modest performance (C statistic = 0.56, 95% CI: 0.52, 0.63 for $\geq$50% decrease in NIHSS over 24 hours; C statistic = 0.6, 95% CI: 0.53, 0.68 for 90-day mRS score of $0-2$) in predicting clinical outcome, but better performance than single-phase CTA and CTP. An automatic mCTA tool [2] is available for this collateral scoring approach which obtained excellent interrater reliability ($n = 30$, $\kappa = 0.81$, $P < .001$).

To measure the correlation between early infarct core and mismatch ratio, Sekar *et al.* [41] performed a study with dynamic as well as single-phase CTA using the four different scoring methods of the ASITN/SIR [27], Christoforidis *et al.* [28], ASPECTS on collaterals [33] and the Miteff System [29]. An excellent cross-correlation ($rho = 0.901$, $p < 0.001$) was obtained between the scores of ASITN/SIR and ASPECTS, which focused on evaluating the extent and delay of vascular enhancement in the affected territory rather than the Christoforidis *et al.* [28] method and Miteff System [29] which considered backflow of contrast medium to the occlusion in scoring collaterals. Table 2.1 summarizes the scoring terminologies used by different researchers to evaluate collaterals in clinical uses.

[2] http://aspectsinstroke.com/casepacs/assess-mcta-collateral-score

Table 2.1: Commonly used collateral scoring terminologies with CTA

Approach	**Scoring Criteria**
ASPECTS [26]	10 ROIs in the MCA territory are scored based on the extent of ischemic changes compared to the contralateral side.
ASITN/ SIR [27]	Scoring from 0-4: 0: non-existent or barely visible pial collaterals on the ischemic site at any time point 1: partial collateralization of the ischemic site until the late venous phase 2 : partial collateralization of the ischemic site before venous phase 3: complete collateralization of the ischemic site by the late venous phase 4: complete collateralization of the ischemic site before the venous phase.
Regional grading [34]	Grades of collaterals based on a scale from 0 to 3 with: 0: no collaterals visible to the ischemic site (absence of any capillary blush) 1: collaterals to the periphery of the ischemic site 2: complete irrigation of the ischemic bed via collateral flow 3: normal antegrade flow.
Evaluation of mCTA [37]	On a scale of 0-5 and graded as poor collateral circulation (0-3) and good collateral circulation (4-5).
Semi-quantitative rLMC [32]	Scores collaterals in a scale of $0-2$: 0 : artery not seen 1 : less prominent 2 : equal or more prominent Scored by comparing with a matching region in the opposite hemisphere based on the extent of contrast opacification in the arteries distal to an M1 MCA+− ICA occlusion.
mCTA-based collateral evaluation [33]	Collaterals are graded as: 0 : for no vessels visibility in any phase of the occluded vascular territory compared to the asymptomatic contralateral hemisphere. $1-4$: based on the extent and phase delays of their prominence in the occluded site compared to the healthy vascular site. 5 : for the normal extent of collaterals without any delay in the occluded site compared to the asymptomatic site.
Pial collateral score [28]	Scoring between $0-4$ based on the retrograde reconstitution of collaterals in the MCA segments.
Miteff System [29]	Collaterals are good or reduced based on areas of MCA reconstituted by contrast. Good collaterals: MCA reconstitution along with branches with a small abruption in the reconstituted vessels at the distal end within M1 or proximal of M2. Reduced collaterals: moderate vessels visible at Sylvian fissure and poor ones reconstituting only superficial branches of MCA.

Mass System [30]	Compares symptomatic hemisphere against the contralateral hemisphere to score collaterals into five categories: 1: absence of collateral 2: for lower extent from the normal contralateral side 3: equal amount to contralateral 4: greater than contralateral 5: for exuberant collaterals.
Tan System [31]	0%: no collateral in the occluded MCA territory, score 0 > 0 and $< 50\%$: collateral score is 1 $> 50\%$: collateral score is 2 100%: collateral score is 3.

2.3 Automatic Collateral Grading Methods

The terminologies of collateral grading in ischemic stroke highlighted in Table 2.1 have been used by different studies described in Section 2.2 to score collateral circulation and obtain functional clinical outcomes. Automated scoring systems aim to provide robust methods that do not suffer from inter- and intra-rater inconsistencies. This section highlights the computer-aided automatic and semi-automatic systems proposed in studies for collateral grading in ischemic stroke treatment.

A number of automatic and semi-automatic methods [42, 43, 44, 45, 13] have been developed to facilitate treatment decisions in ischemic stroke.

Machine learning excels over traditional quantitative approaches in classification tasks by enabling automatic feature selection, handling high-dimensional data, and uncovering non-linear relationships that can capture complex patterns, resulting in more accurate and adaptable classification models. Due to the advantages of machine learning over traditional methods, several studies [5, 46, 25, 47], The e-ASPECTS® software from Brainomix Ltd. (Oxford, UK) and RAPID ASPECTS® by iSchemaView (Menlo Park, USA)] focused on the automatic evaluation of collaterals with ML methods.

The proposed collateral assessment and ASPECTS automation methods helped achieve generalizability to new data and good performance. However, one potential drawback of machine learning is the extensive hand engineering required, which can be a time-consuming and challenging process requiring significant domain expertise.

Recently, DL has become a popular and powerful tool for solving complex problems in medical imaging. Unlike traditional machine learning methods, DL models can automatically learn and extract relevant features from data, reducing the need for manual feature engineering. Therefore, current studies (e.g., [4, 48, 49, 50, 51, 52] used DL models to evaluate collaterals automatically.

In the greater context of stroke, DL models have been used in automating ASPECTS score based on ischemic damage on NCCT [53, 54, 55, 56] since NCCT is more commonly used in these studies, being widely used in emergent settings and, thus, more accessible.

The issues that arise with deep learning are when available training data are scarce. There are very limited open and large labeled datasets in the case of stroke data. Transfer learning [57] can be used to mitigate the lack of training data.

EfficientNet [58] outperformed state-of-the-art CNNs in transfer learning. This architecture scales the model's depth, width, and resolution in a balanced way through

a compound coefficient, enhancing the network's performance. Since EfficientNet [58] outperforms the existing transfer learning network, we considered this model to achieve our goal of automatic evaluation mitigating the issues of small imbalanced datasets.

The challenge with transfer learning in the medical domain stems from the limited availability of pre-trained weights, which leads to performance deficiencies. The Siamese network, introduced by Bromley *et al.* [59], is a potential alternative to transfer learning with more robustness as it needs a minimal training set.

Based on the concept of the Siamese network, calculating the similarity with good performance and the robust power of handling imbalanced issues, we incorporated the network into one of our proposed methods for the automatic evaluation of collaterals.

Table 2.2 summarizes the main highlights of the existing approaches (based on quantitative measurement, classic machine learning, and deep learning) for collateral grading. Details about all the studies are discussed in the following manuscript-based chapters.

2.4 Summary

This chapter delineates manual and automated collateral evaluation approaches, employing various scoring terminologies. The comprehensive overview highlights the superior efficacy of automated methods compared to manual processes. Furthermore, we explore the advantages and limitations of computerized techniques, including quantitative, traditional machine learning, and deep learning-based approaches.

Table 2.2: State-of-the-art methods for automatic collateral evaluation

Types of methods	Ref	Data	Scoring	Cases	Methods and results
Quantitative methods	Sheih *et al.* [44]	NCCT	ASPECTS	103	Automatic contralateral comparative method; AUC of 90.2%
	Frolich *et al.* [45]	4D CTA	rLMC	82	tMIPs best for collateral prediction; intra-class correlation of 0.78
	Zhang *et al.* [13]	4D CTA	rLMC	80	CGS used by combining velocity & extent of collaterals; AUC of 0.80
	Kersten *et al.* [42]	4D CTA	ASPECTS	29	Intensity differences between left & right hemisphere; Correlation of method to radiologist had $r^2 = 0.71$
	Boers *et al.* [43]	CTA	Tan [31]	422	Vascular ratio between left & right hemispheres; Correlation of 0.75 between visual & quantitative score
Classic ML methods	Xiao *et al.* [5]	4D CTA	ASPECTS	37	Machine learning method based on SVM; Overall accuracy of 82.2%
	Kuang *et al.* [46]	NCCT	ASPECTS	257	Random forest classifier; ICC between method & experts was 0.76
	Grunwald *et al.* [25]	CTA	Tan [31]	98	Automated e-CTA; 90% agreement with radiologist, ICC=0.93
	Su *et al.* [47]	CTA	4-grade Tan [31]	269	Multi-class classification; Overall accuracy 0.80 & dichotomized collateral accuracy 0.90

DL methods	Tetteh *et al.* [48]	3D MR Perfusion	Direct and cascaded 3-class collateral scoring	183	ROI extraction followed by feature extraction and classification with CNN, SVM, RF, and KNN; an overall accuracy of 0.72 with CNN+MLP
	Ali *et al.* [51]	Cone beam CT (CBCT)	collateral scoring by radiologists into 3-classes	30 (4368 CBCT images)	2D ResNet-50 as a classifier to evaluate good and poor collaterals; Average sensitivity and specificity of 0.79 and 0.96 respectively
	Rava *et al.* [50]	Peak arterial CT perfusion volumes	ESCAPE trial and Tan score [31]	200	A CNN model to evaluate collaterals automatically; Average sensitivity of 0.88 and 0.80 for dichotomized and multi-class collaterals respectively
	Huang *et al.* [52]	Multiphase CTA	Positive (good + intermediate) and negative (poor) collaterals	82	CNN model; Accuracy of 0.75 in the validation group with an AUC of 0.70
	Tan *et al.* [49]	Multiphase CTA	Tan score [31]	173	Feature fusion from four phases with hybrid attention mechanism; Accuracy of 90.43%

Chapter 3

Automatic collateral circulation scoring in ischemic stroke using 4D CT angiography with low-rank and sparse matrix decomposition

Preface

This chapter is based on our published paper [60].

This chapter aims to achieve our first goal of evaluating collaterals automatically using the same visual indications (e.g., filled versus unfilled vessels) that radiologists use. This is important as it allows clinicians to better understand the computerized methods, thus enabling clinical acceptance. The proposed method, ACCESS, estimates the extent of unfilled cerebrovasculature in CT angiogrpahy scans by identifying areas lacking contrast agents due to clotting. We apply the fast Robust Matrix Completion (fRMC) algorithm with in-face extended Frank-Wolfe optimization to a group of healthy subjects and a target patient. This enables us to model the patient's unfilled vessels as sparse components and the estimated complete vasculature as low-rank components. The collateral score is then determined as the ratio of unfilled vessels to the full vasculature, mirroring established clinical protocols. This work has been published in the International Journal of Computer-Assisted Radiology and Surgery (IJCARS), and the GitHub code is available at: https://github.com/mumuaktar/ACCESS

Abstract

Sufficient collateral blood supply is crucial for favorable outcomes with endovascular treatment. The current practice of collateral scoring relies on visual inspection and thus can suffer from inter and intra-rater inconsistency. We present a robust and automatic method to score cerebral collateral blood supply to aid ischemic stroke treatment decision-making. The developed method is based on 4D CTA and the ASPECTS scoring protocol. The proposed method, ACCESS (**A**utomatic **C**ollateral **C**irculation **E**valuation in i**S**chemic **S**troke) estimates a target patient's unfilled cerebrovasculature in contrast-enhanced CTA using the lack of contrast agent due to clotting. To do so, the fast Robust Matrix Completion (fRMC) algorithm with in-face extended Frank-Wolfe optimization is applied to a cohort of healthy subjects and a target patient, to model the patient's unfilled vessels and the estimated full vasculature as sparse and low-rank components, respectively. The collateral score is computed as the ratio of the unfilled vessels to the full vasculature, mimicking existing clinical protocols. ACCESS was tested with 46 stroke patients and obtained an overall accuracy of 84.78%. The optimal threshold selection was evaluated using a receiver operating characteristics (ROC) curve with the leave-one-out approach and a mean area under the curve (AUC) of 85.39% was obtained. ACCESS automates collateral scoring to mitigate the shortcomings of the standard clinical practice. It is a robust approach, which resembles how radiologists score clinical scans and can be used to help radiologists in clinical decisions of stroke treatment.

3.1 Introduction

Stroke is one of the leading causes of disability and death worldwide. Statistics from the World Heart Federation show that each year 15 million people suffer from stroke among which 5 million become permanently disabled and 6 million people die [1]. There are two kinds of stroke: ischemic, where a blood clot forms in a cerebral artery, and hemorrhagic, where a cerebral vessel ruptures and bleeds into the brain. Ischemic stroke is much more frequent with 8 out of 10 stroke-affected patients suffering from it.

When diagnosis and treatment of stroke are not performed in time, patients become disabled due to a lack of blood and oxygen, which causes neuronal cell death in the affected part of the brain. Treatment strategies are chosen based on a number of factors, including time window, infarct volume, penumbra size, and collateral circulation. A patient can be treated after 6 hours of symptoms onset with endovascular treatment, where a catheter with a mechanical device attached to the tip is used to remove the clot. This mechanical intervention allows blood flow to quickly be restored. However, not all stroke patients are suitable candidates for endovascular treatment, due to the risks associated with it. One important indication for successful endovascular treatment is the presence of sufficient collateral circulation (i.e., *collaterals*) [61].

Grading the extent of collateral circulation is an important factor for treatment decision-making, and a number of approaches have been developed to visually quantify collateral circulation, including ASITN/SIR Collateral Score, Miteff System, Mass System, modified Tan Scale, and ASPECTS (Alberta Stroke Program Early CT Score) [62]. With these approaches, performance depends on the experience, training, and specialty of radiologists, and thus can result in inter- and intra-rater inconsistency which have been shown to

[1] http://www.world-heart-federation.org/cardiovascular-health/stroke/.

be an issue in a number of studies [22, 23, 24]. Grotta *et al.* [22] concluded that it is difficult to get agreement in recognizing and quantifying early ischemic changes even by experienced clinicians. A recent study by Grunwald *et al.* [25] showed that between individual neuroradiologists, the intraclass correlation coefficient ranges from 0.42 to 0.86 and score agreements range from 36.2% to 81.6%. Automated scoring systems aim to provide robust methods that do not suffer from inter- and intra-rater inconsistencies.

3.2 Related Work

A number of automatic and semi-automatic methods have been developed to facilitate treatment decisions in ischemic stroke. Kersten-Oertel *et al.* [42] developed a method that considered differences of mean intensities on the left and right hemispheres. The results of this method showed a good correlation ($r^2 = 0.71$) between the radiologist and computed score but the method itself had difficulty dealing with individual variations, e.g. from calcification, as well as, normal vasculature asymmetry between hemispheres. Boers *et al.* [43] also considered the ratio between left and right hemispheres for quantitative measurement of collateral status and obtained a good correlation, ρ of 0.75 ($p < .001$) between visual and quantitative collateral score. In their work, multiscale segmentation was done on baseline CTA with better result for arteriovenous acquisition phase to obtain the vasculature before the collateral evaluation. In the work by Xiao *et al.* [5], support vector machines (SVM) were used to score collateral supply after extracting blood vessels automatically using low-rank decomposition. The results of this method showed good separation between good and intermediate versus poor collaterals with an overall accuracy of 82.2%. The drawback of this machine learning-based method is that performance is affected by limited training data. A random-forest-based classifier was developed by Kuang *et al.* [46], where non-contrast CT (NCCT) was used to automate the ASPECTS assessment. The intra-class correlation coefficient between the automated ASPECTS method and the DWI ASPECTS score by experts was found to be 0.76, but NCCT may not be sensitive enough for those with good collaterals. Shieh *et al.* [44] developed a computer-aided decision system for thrombolysis therapy using NCCT. Their scoring based on a contralateral comparative method is independent of ground truth and obtained an area under the curve (AUC) of 90.2%. Collateral assessment with 4D CTA was performed by Frolich *et al.* [45] using the semi-quantitative regional leptomeningeal collateral score (rLMC), proving that temporally fused maximum intensity projections (tMIPs) can better depict the collateral flow. Their study obtained an inter-rater agreement with an intraclass correlation coefficient of 0.78. However, their experiments were limited to certain time points rather than the entire 4D CTA series and confined to only subjects with good collaterals. Zhang *et al.* [13] have integrated the velocity and extent of collaterals in the peak phase and tMIPs to obtain a collateral grading score (CGS) using 4D CTA. Using the rLMC semi-quantitative approach to set the CGS cutoff, the method resulted in an AUC of 0.80. Table 3.1 summarizes the main highlights of the existing approaches for collateral grading.

In this chapter, we describe an automated image-processing approach ACCESS (**A**utomatic **C**ollateral **C**irculation **E**valuation in i**S**chemic **S**troke) for evaluating collateral circulation with the ASPECTS protocol as the reference, which has been shown to be a reliable, systematic and robust approach. The ASPECTS score is based on the extent of contrast opacification in arteries distal to the occlusion clot [26]. Our goal is to use robust low-rank and sparse decomposition to obtain unenhanced collaterals in a patient from the

Table 3.1: A brief survey of existing collateral scoring techniques using CTA.

Ref	Data	Scoring	Cases	Method and Results
Shieh *et al.* [44]	NCCT	ASPECTS	103	Contralateral automatic comparative method; AUC of 90.2%
Frolich *et al.* [45]	4D CTA	rLMC	82	Manual grading; tMIP best for collateral prediction; intra-class correlation of 0.78
Zhang *et al.* [13]	4D CTA	rLMC	80	CGS used by combining velocity & extent of collaterals; AUC of 0.80
Kersten *et al.* [42]	4D CTA	ASPECTS	29	Intensity differences between left & right hemisphere; Correlation of method to radiologist had $r^2 = 0.71$
Xiao *et al.* [5]	4D CTA	ASPECTS	37	Machine learning method based on SVM; Overall accuracy of 82.2%
Kuang *et al.* [46]	NCCT	ASPECTS	257	Random forest based classifier; Intra-class correlation coefficient between proposed method & experts was 0.76
Boers *et al.* [43]	CTA	Tan System [31]	422	Vascular ratio between left & right hemispheres; Correlation of 0.75 between visual and quantitative score
Grunwald *et al.* [25]	CTA	Tan System [31]	98	Automated e-CTA score; 90% agreement with radiologist, intra-class correlation coefficient of 0.93
ACCESS (proposed method)	4D CTA	ASPECTS	54	Automatic collateral circulation scoring; An average AUC of 85.39%

group behavior of normal controls. The developed model is based on the assumption that from the group of normal controls and one target patient taken as columns in a low-rank matrix completion framework, the unfilled collaterals of a stroke patient can be reconstructed in the sparse component whereas the unchanged full vasculature appears in the low-rank component. Based on this concept, we developed a novel automated approach for collateral scoring which considers the ratio of unenhanced collaterals to the full vasculature and determines the collateral score using this ratio. ACCESS uses the fast robust matrix completion (fRMC) method [63] to extract blood vessels benefiting from the in-face extended Frank-Wolfe algorithm [64], a method for solving a defined convex optimization problem.

3.3 Materials and Methods

The ACCESS pipeline is shown in Fig. 3.1 and described in detail in the following section.

3.3.1 Scanning Protocol

Eight healthy subjects were used as reference scans, and 46 subjects with ischemic stroke were used to evaluate our method. All subjects underwent imaging at the Montreal Neurological Hospital (Montreal, Canada). The 4D CTA images were captured on

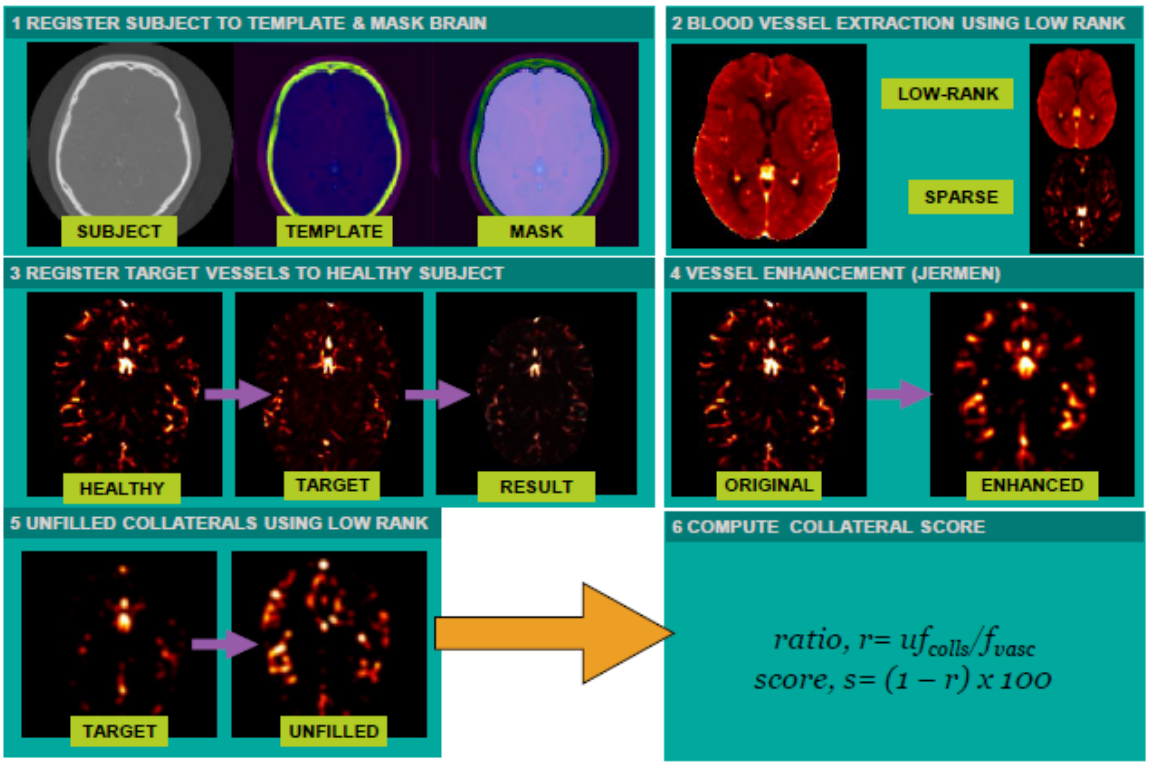

Figure 3.1: The overall workflow of the ACCESS method.

a Toshiba's Aquilion ONE 320-row detector 640-slice cone-beam CT (Toshiba Medical Systems, Tokyo, Japan). The scanner provides whole-brain perfusion and dynamic vasculature information in one single examination with a single rotation of the gantry. The routine stroke protocol performs a series of intermittent volume scans over a period of 60 seconds with a scanning speed of 0.75 s/rotation. A total of 19 volumes are captured for each patient with low-dose scanning for every 2s during the arterial phase and 5s during the venous phase. Isovue-370 (Iopamidol) was used as a non-ionic and low osmolar contrast medium (Iodine content, 370 mg/ml).

3.3.2 Pre-processing

Prior to evaluating collateral supply for each stroke patient, we followed a number of pre-processing steps to (1) register all the subjects (healthy and with stroke) to a standard template space, (2) extract blood vessels (3) refine group-wise blood vessel alignment and (4) enhance vessels using probabilistic segmentation.

3.3.2.1 Image Registration

To process all the subjects in the same space we performed image registration in two stages. In the first stage, all 18 CTA volumes of an individual subject are rigidly registered to that of the first time point. Next, the first volume is registered non-linearly to a CTA brain template using the symmetric image normalization method (SyN) [65] from ANTs (Advanced Normalization Tools)[2]. Then, the non-linear transformation is applied to all

[2]stnava.github.io/ANTs

other volumes of that subject. Thus all subject volumes are registered first to each other and then to the template. The CTA brain template was created following the unbiased group-wise registration approach [66] using 12 healthy subjects' brains. A brain mask was created from the template using active contour segmentation in ITK-SNAP (www.itksnap.org) and used to remove the skulls of all subjects for further analysis. In the second registration stage, we further refine the vessel alignment, in order to estimate the unfilled blood vessels in the patient through low-rank and sparse decomposition. Therefore, additional nonlinear registration using SyN is performed in this stage between a randomly selected healthy subject's temporal 3D average taken as a template and the rest of the 3D subjects participated in further experiments. Note that the temporal 3D average of each subject (averaging along multiple time points) for vessel alignment is obtained after blood vessel extraction described in Section 3.3.2.2.

3.3.2.2 Blood Vessel Extraction

To evaluate collaterals, it is necessary to determine how blood flows over time for each subject. Thus, the static background with grey and white matter, as well as, any calcification, which can affect the scoring are not considered. Similar to the approach proposed by Xiao *et al.* [5], the flow of the contrast agent in the blood vessels are separated from the static background with calcification. However, rather than using the augmented Lagrange multiplier method by Lin *et al.* [67] to recover low-rank and sparse components, we use the more robust fRMC method [63]. This method does not require any parameter tuning and converges very quickly [68] whereas [67] is sensitive to parameter tuning and has a slower convergence rate. For applying low-rank decomposition to a 4D CTA scan, a matrix, $D = [C_s^1, C_s^2, \ldots..C_s^{19}]$ is considered with all the volumes taken as columns of the matrix. The low-rank representation is as below:

$$\min \quad rank(B) \text{ s.t.} \quad D = B + V \tag{1}$$

The minimization of ranks of background, B as in equation (1) for separating the correlated and static background features from the dynamic blood vessels, V is performed by,

$$\min \|B - D\|_F^2 \quad \text{s.t.} \quad \|B\|_* < \delta \tag{2}$$

where $\|.\|_F$ indicates the Frobenius norm, $\|.\|_*$ represents the nuclear norm of a matrix and δ is the constraining upper bound for the nuclear norm of low-rank matrix, B. We can see from Equation 2 that there is no tunable parameter to obtain the low-rank and sparse matrix for fRMC. Since D is a large non-singular matrix with multiple volumes as columns, the square of its Frobenius norm is greater than both its nuclear norm and the nuclear norm of its component, B. Therefore, δ can be comfortably set to any value greater or equal to the square of the Frobenius norm of D. Similar to the study of Ashikuzzaman *et al.* [68] and the reference study of fRMC by Rezaei *et al.* [63] for background subtraction, we set it to ten times the Frobenius norm of D. Thus, we do not need to set any parameters manually for extracting blood vessels and further unfilled collaterals using fRMC. The rank minimization in fRMC is solved using the extended Frank-Wolfe optimizer [64] which requires a lower number of iterations and less computation in each iteration that makes the fRMC method fast. The choice of its convergence parameters, γ_1 and γ_2 mostly affect the convergence speed and rank of the matrix with $0 \leq \gamma_1 \leq \gamma_2 \leq 1$ [63]. We provide further insight of these parameters in Section 3.5. Finally, we extracted the sparse matrix containing blood vessels

by subtracting the background from the original data matrix. The columns obtained in the sparse matrix represent the blood flow in each volume over time. Given this, we can then take the average of the 19 volumes containing blood flow over time to perform collateral scoring in 3D.

3.3.2.3 Enhancement of Vascular Structures

Vessel enhancement is an important prerequisite for computer-aided clinical procedures to highlight the blood vessels and suppress noise and other non-vascular structures. There is much literature on vessel segmentation. Here, we review a few recent techniques, but interested readers can refer to [69]. Yang *et al.* [70] developed a vessel segmentation technique following contrast enhancement, boundary refinement, and content-aware regions of interest adjustment by checking shape consistency and connectivity. Rather than considering a region-based method that may be sensitive to unnatural intensity variations, Meijs *et al.* [71] segmented full cerebral vasculatures in 4D CT using weighted temporal variance and local histogram features as inputs to a random forest classifier and obtained an overall accuracy of 0.995. Vessel segmentation by Jin *et al.* [72] utilized the low-rank and sparse decomposition technique to segment vessels from group behavior of the sequence of XCA images and further removed spatially varying noisy residuals through local-to-global adaptive threshold filtering.

In our approach, we used the vessel enhancement by Jermen *et al.* [73] to increase the visibility of the blood vessels in MIPs as well as to make the contrast agent response uniform. This enhancement method outperforms traditional vesselness filtering approaches by enhancing rounded structures along with elongated ones. The filter allows local structures to be distinguished by analyzing the eigenvalues of the Hessian matrix at each point in the image. Let, $\lambda_i, i = 1, 2, 3$ denote the three eigenvalues of the Hessian matrix of a 3D image with the ideal eigenvalue relationship $\lambda_2 \approx \lambda_3 \wedge |\lambda_{2,3}| >> |\lambda_1|$. This relation, however, can't be maintained if the magnitudes of λ_2, λ_3 are very low. So, to ensure robustness in case of lower eigenvalues, a regularization on the value of λ_3 at multiple scales is done by:

$$\lambda_\rho(s) = \begin{cases} \lambda_3 & \text{if} \quad \lambda_3 > \tau max_x \lambda_3(x, s), \\ \tau max_x \lambda_3(x, s) & \text{if} \quad 0 < \lambda_3 \leq \tau max_x \lambda_3(x, s), \\ 0 & \text{otherwise} \end{cases}$$

where s is the vessel scale and τ is the cutoff threshold (value between 0 to 1), which results in a uniform response. Finally, the elliptic cross-section structures are confined to the ratio $\lambda_2 \geq \lambda_\rho/2 > 0$ and the vessel enhancement function is defined as:

$$V_p = \begin{cases} 0 & \text{if} \lambda_2 \leq 0 \vee \lambda_\rho \leq 0, \\ 1 & \text{if} \lambda_2 \geq \lambda_\rho/2 > 0, \\ \lambda_2^2(\lambda_\rho - \lambda_2)\left[\frac{3}{\lambda_2 + \lambda_\rho}\right]^3 & \text{otherwise} \end{cases}$$

V_p can be computed for both bright and dark structures and the filter response is between 0 and 1 but ideally 0 for non-vascular and 1 for vascular structures.

Using this vessel enhancement method, we segmented vessels in all subjects before using them to estimate unfilled collaterals via the second low-rank and sparse decomposition.

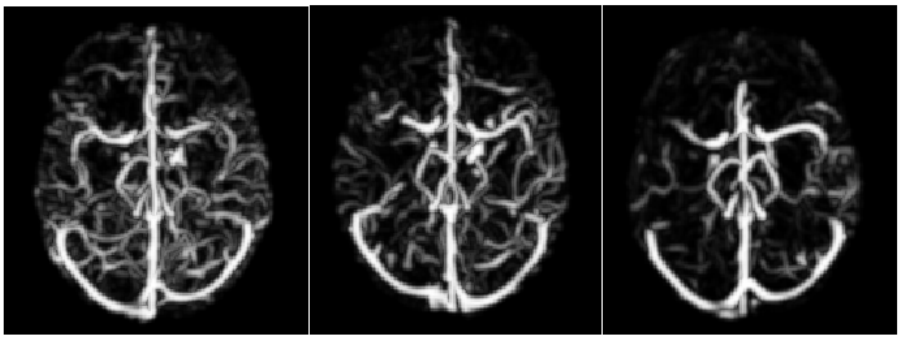

Figure 3.2: Example MIP CTAs of different collateral circulation scores.

3.3.3 Collateral Circulation Evaluation

For evaluating collateral circulation, the score is categorized into three types: good, intermediate and poor. These scores are defined based on the collateral supply in the occluded MCA territory according to ASPECTS [26]. The ASPECTS score, which we used as ground truth in our experiments, is based on the agreement of the visual assessment of the acquired 4D CTA scans by two radiologists. A score of good means 100% collaterals, intermediate means greater than 50% and lower than 100% and poor means below 50% and greater than zero (Fig. 3.2). Since fRMC uses an optimizer to minimize the rank, more variability between individuals can affect the results. To mitigate individual variability, we blurred the data with a Gaussian kernel ($\sigma = 2mm$). Among our 54 subjects, we had 8 normal controls, 14 poor, 17 intermediate, and 15 good subjects.

To measure collateral supply, we compare filled vessels (by contrast agent) with unfilled ones. To obtain the unfilled vessels in a patient, a group of 8 normal subjects and a target subject is created. Normal subjects are considered as healthy with 100% collateral supply. The target collateral score can then be defined using the normal controls. We used the same robust approach fRMC with extended Frank-Wolfe solver in order to obtain the unfilled vessels of a target case into a sparse matrix from the group behavior. Since all the normal controls contain very similar vasculature, the full vasculature is obtained into the low-rank matrix. The data matrix here is defined as, $D = [C_s^1, C_s^2, \ldots..C_s^8, C_s^9]$ where columns, C_s^1 to C_s^8 are the normal subjects and C_s^9 is the test case. Next, low-rank minimization is performed as in Equation (2) which is defined here as:

$$\min \quad \|f\_vasc - D\|_F^2 \quad \text{s.t.} \quad \|f\_vasc\|_* < \delta \tag{3}$$

where f_vasc stands for full vasculature from where we obtain the unfilled collaterals by, uf_colls=D-f_vasc. Finally, the collateral score is measured as below:

$$Ratio, r = uf\_colls/f\_vasc \tag{4}$$

and collateral score in the target subject,

$$Score, s = (1 - r) \quad \times \quad 100. \tag{5}$$

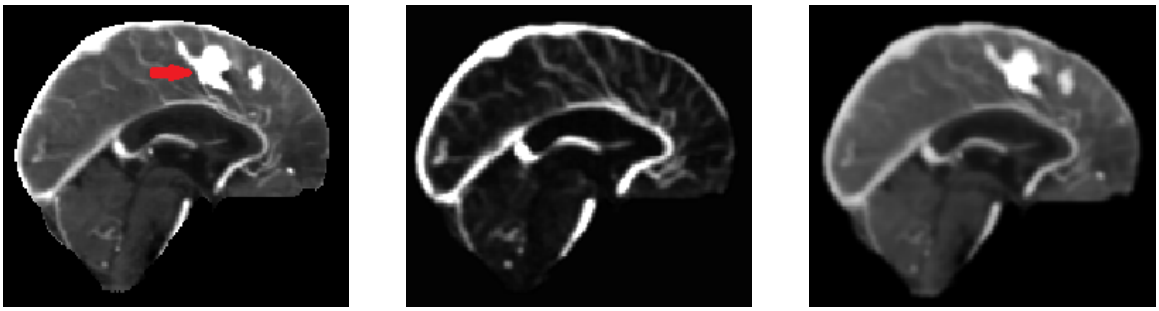

Figure 3.3: Example of low-rank decomposition. *Left:* original sagittal view of a subject (red arrow points to a calcification). *Middle:* sparse image containing blood vessels. *Right:* low-rank image with calcification.

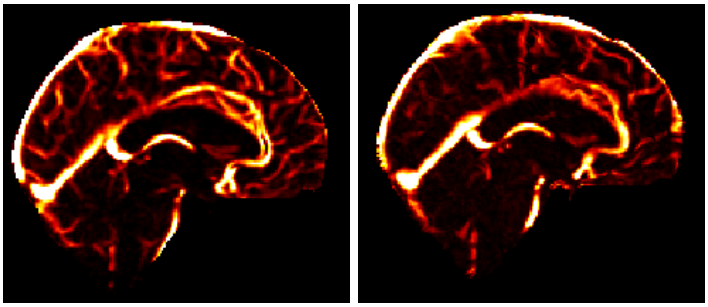

Figure 3.4: *Left:* a healthy subject used as a reference image. *Right:* an individual registered to the reference image.

3.4 Experimental Results

3.4.1 Blood Vessel Extraction

To overcome the inter-volume intensity differences within a scan session, the intensity profiles of all the volumes of a subject are normalized (using Minmax) with respect to the first volume. The fRMC approach is then applied to separate blood vessels from the background which results in the removal of any calcification (e.g. Fig. 3.3).

3.4.2 Image Registration

A two-stage registration was performed by Huck *et al.* [74] to create a cerebral vascular atlas, which used standard parameters from the ANTs tool to align segmented vessels in fine details. Similar to their work, we found that the standard parameters for the SyN algorithm from ANTs worked well for registration to the template as well as the alignment of blood vessels. To align the blood vessels, a SyN deformation on 4 scale levels was done (with iterations of 100x100x50x20). We evaluated the alignment by checking the overlap of blood vessels in multiple subjects extracted by applying low-rank and sparse decomposition. Areas with fewer than 40 connected pixels were ignored to avoid the smallest vessels, which are quite different in each individual and can cause scoring error. The registration performance of an individual subject's blood vessel to a healthy subject is shown in Fig. 3.4.

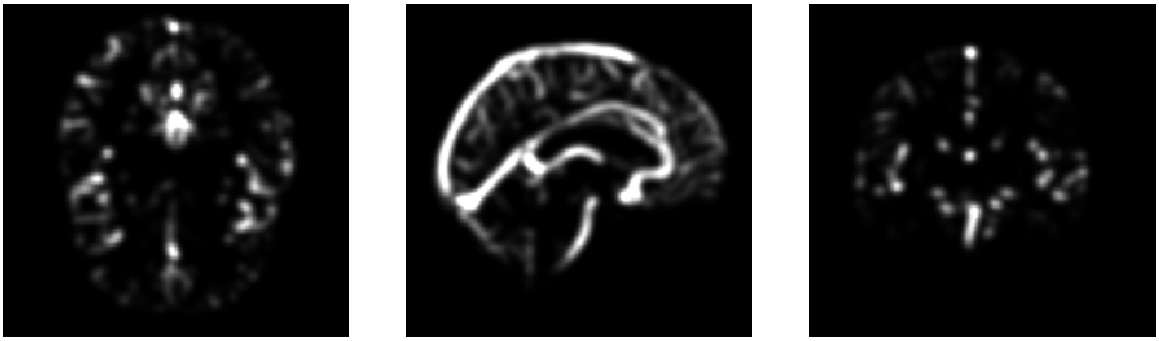

Figure 3.5: Vessel enhancement of a subject (axial, sagittal and coronal view).

3.4.3 Vessel Enhancement

The blood vessels of an example subject segmented by scaling responses between 0 to 1 using Jermen vessel enhancement function [73] is shown below in (Fig. 3.5). The scale, s, ranging from 0.5mm to 2.5mm with a 0.5mm interval was chosen based on the work of Jermen *et al.* [73] on 3D DSA cerebral vasculature segmentation. The regularization parameter, τ, is varied over the chosen scales from 0.5 to 1 to show the effect on the segmentation outcome in Section 3.4.5; τ= 0.5 resulted in a uniform response in our dataset.

3.4.4 Automatic Collateral Circulation Evaluation

To reduce computational complexity and for better visibility, the 2D MIP of the sparse and low-rank matrices in axial directions were taken. Demonstration of low-rank matrix completion to obtain the unfilled vessels in an individual with respect to the cohort of normal subjects is shown in Fig. 3.6.

From the sparse images, we can see that there are a lot of unenhanced vessels obtained. This is due to the variability of the small vessels in individual subjects which appear as changes in the sparse matrix. Post-processing is performed with thresholding in order to ignore the small vessels' variability, as well as some unwanted portions obtained in the sparse component due to contrast variations. To remove very small vessels, all connected components with fewer than 40 pixels were removed from the binary images obtained by thresholding. To overcome the effect of manual thresholding and make ACCESS more robust, we performed a sensitivity analysis in Section 3.4.5 to obtain this optimal threshold. Furthermore, the main sinus and arteries are removed before collateral scoring.

The images in Fig. 3.7 show the binary of the sparse and low-rank components after thresholding.

Finally the collateral score is calculated according to the formula of Equations 4 and 5 using the optimal parameters obtained by the sensitivity analysis. Since there is variability between vessels of individual subjects, which cannot be registered perfectly and the final operation is performed on 2D, the radiologists' scores can be conflicting to use here directly. Thus the scoring performance of ACCESS was evaluated after determining the optimal threshold for each class by computing ROC curves [75] (Fig. 3.8). A ROC curve is drawn with the experimental scores and the true class labels. The sensitivity and specificity for different threshold settings, which are varied between 0-100, based on the scores were calculated. Choosing for the optimal points on the curve for the thresholds to define the good, intermediate and poor collaterals, we found scores under 55.45% should be considered

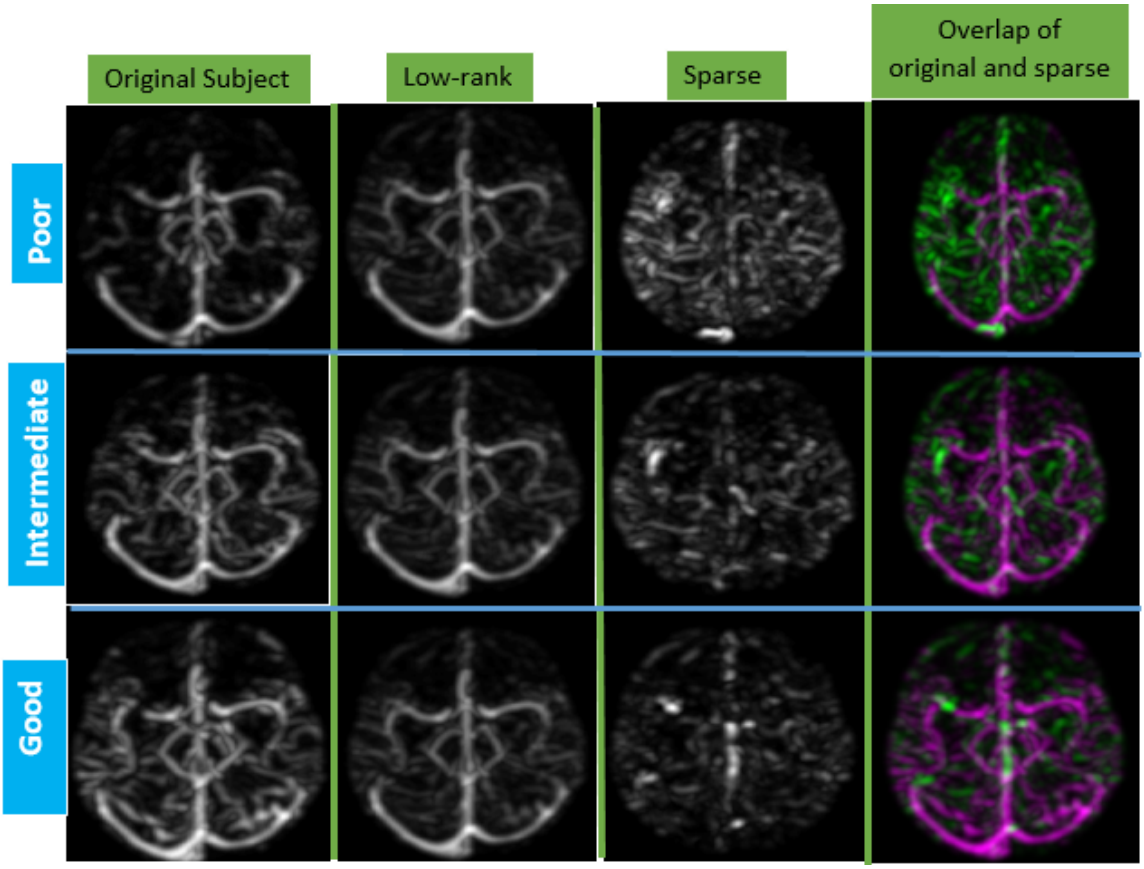

Figure 3.6: 2D MIP representation for Low-rank and sparse decomposition for a poor, intermediate and good subject. The overlaid image is shown for better visibility of unfilled vessels (green) with the original collaterals (pink).

as poor, between 55.45 and 70.5% intermediate and above 70.5% good. To overcome the effect of over-estimation, we performed a leave-one-out approach to draw the ROC curves and the final curve is drawn from the average of 46 iterations of true positive and false positive rates. An AUC of 85.39% was obtained from the ROC, with AUC of 90.95%, 83.53% and 81.70% for good, intermediate and poor classes respectively. Table 3.2 shows the confusion matrix. An overall accuracy of 84.78% is obtained from the true positive and true negative results.

Table 3.2: Confusion Matrix showing ACCESS Results.

Radiologist score / **Automatic score**	**Good**	**Intermediate**	**Poor**
Good	13	2	0
Intermediate	1	14	2
Poor	1	1	12

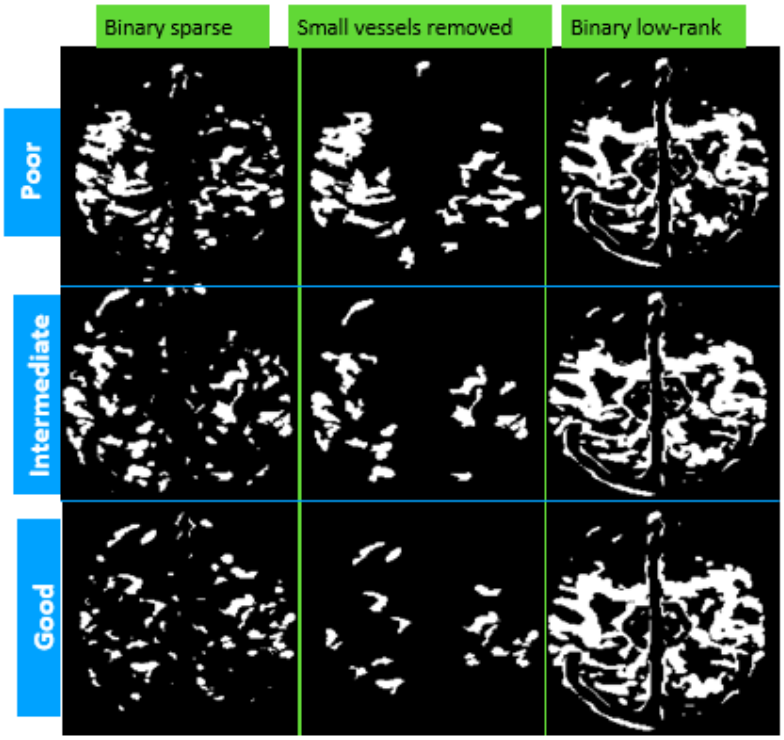

Figure 3.7: Post-processing results of 2D MIP of poor, intermediate and good collaterals.

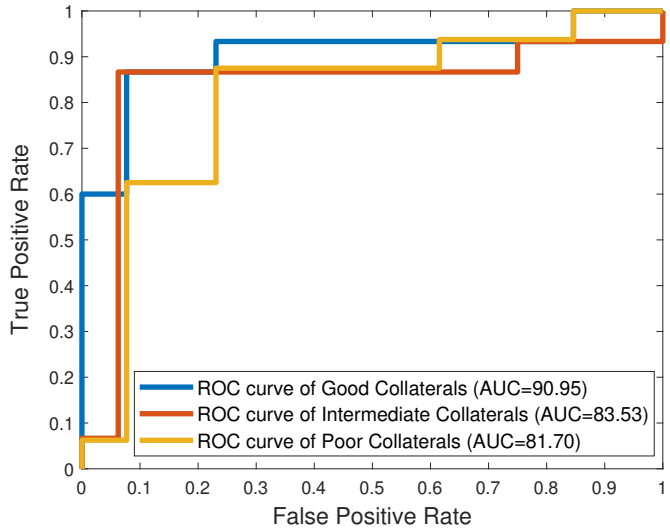

Figure 3.8: ROC curves for Good, Intermediate and Poor Collateral Scores Evaluation.

3.4.5 Sensitivity Analysis

In this section, we perform sensitivity analysis for our pipeline on parameter choices of vessel enhancement. Since we do not have any ground truth data for direct segmentation performance analysis, we assessed the impact of parameters in vessel enhancement (regularization parameter, τ and thresholding value) on the overall performance of ACCESS.

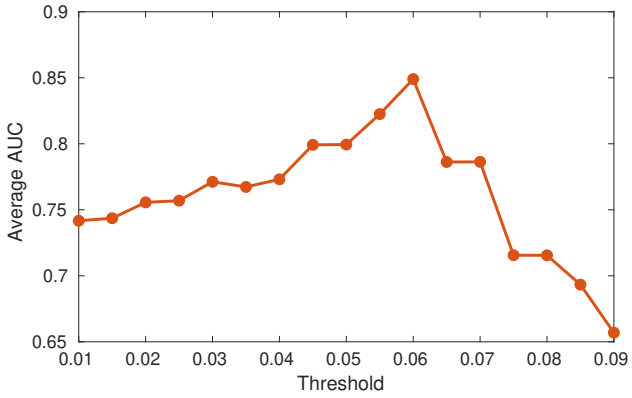

Figure 3.9: Average AUC of three classes for varying threshold values.

The reference study by Jermen's enhancement filtering [73] has shown that probabilistic segmentation of blood vessels can be obtained by setting τ to a value between 0.5 to 1 for a uniform response. Therefore, we tested the scoring results for multiple τ values. We achieved the best result for τ=0.5. In every experiment with different τ, we varied the threshold value within 0.01 to 0.09 (chosen based on the mean intensity of low-rank and sparse images) to obtain the best sensitivity and specificity with the highest AUC in the ROC. Table 3.3 shows the AUC for each τ with the optimum threshold value.

Table 3.3: Influence of τ on the Final Evaluation.

τ	**AUC Good**	**AUC Intermediate**	**AUC Poor**	**AUC Average**
0.5	0.909	0.823	0.816	0.849
0.6	0.875	0.798	0.839	0.837
0.7	0.914	0.656	0.843	0.804
0.8	0.909	0.641	0.843	0.798
0.9	0.923	0.637	0.816	0.792
1	0.904	0.660	0.825	0.797

Since τ=0.5 shows the best performance, we achieved the segmentation with this value. Fig. 3.9 shows how the scoring is sensitive to the varied threshold values. Note that the final ROC curve shown in Fig. 3.8 from the cross-validation results is done using the optimum threshold value.

3.4.6 Inter and Intra-rater Variability Analysis

A subset of 27 test cases from the data were rated individually by the two separate radiologists. To show the effect of visual inspection and human rater's variability, we used the consensus ground truth and the separate ratings by the radiologists as well as one of the authors (MA), who served as the third independent rater. The rating is performed based on ASPECTS by assessing the degree of collaterals visually. The subjects' collaterals are scored as "good" if both sides have equal extent of collaterals in any of the phases from arterial to venous with contrast. The same criteria are followed for intermediate and poor subjects with medium and very low extent of collaterals in the affected side of MCA territory compared to the healthy side respectively.

To evaluate the inter-rater variability of the two radiologists and the author in the subset of 27 cases, we computed Fleiss' Kappa (κ) statistics [76], which ranges between 0 and 1, (with values from 0.0 to 0.2 indicating slight agreement, 0.21 to 0.40 indicating fair agreement, 0.41 to 0.60 indicating moderate agreement, 0.61 to 0.80 indicating substantial agreement, and 0.81 to 1.0 indicating almost perfect or perfect agreement based on the guidelines of Landis *et al.* [77]) for three raters. With an overall κ=0.455 ($p < 0.0005$), the result represents moderate strength of agreement among the raters. The two radiologists' rating variability on the same subset of cases is obtained by computing a paired Cohen's kappa (κ) statistics [78], yielding κ=0.471 ($p < 0.001$). It should be noted that the collateral score influences the clinical outcome significantly. For example, a "poor" case misclassified as being "good" can cause excessive bleeding leading to hemorrhagic stroke with EVT while a "good" case being misclassified as "poor" ignores a patient from EVT. Due to the raters' variability, poor treatment decisions might be made, which would adversely impact patients. In the subset, the two radiologists have disagreement in scoring poor vs. intermediate in 10% of cases, 11% for poor vs. good and 35% for good vs. intermediate cases. These findings (56% of cases misclassified) fall in the middle of the range of what Grunwald *et al.* [25] found where score agreements range from 36.2% to 81.6%. Further, to show the agreement between ACCESS and radiologists' scores, we computed Cohen's kappa [78] with κ=0.771 ($p < .0005$) and obtained a substantial agreement between the automated approach and radiologists' score. To verify the scoring quality of the independent rater (MA), we computed Cohen's kappa coefficient [78] between the independent rater and the consensus of two expert raters and obtained a substantial agreement with κ=0.649 ($p < 0.0005$). Finally, to assess intra-rater variability, two separate ratings were performed with an interval of 5 days by the same rater MA to compute kappa statistics with κ=0.530 ($p < 0.0005$). Based on the statistics, we can see how the inter- and intra-rater variability between human raters potentially affect the collateral scoring.

3.4.7 Computation Times

Processing was done on a Windows 7 machine equipped with an Intel(R) Core(TM) i7-4770 CPU @ 3.40GHz and 28 GB of RAM. Registration using ANTS took an average of 19 minutes to register an individual to the template and a further 16 minutes for vessel alignment. The other processing steps are faster, calcification removal takes under 2 minutes, vessel enhancement for individual 3D scans takes 10 seconds and the overall scoring from the sequence of 8 patients is completed within 25 seconds. In order to reduce the processing time of registration, in the future, we will port the pipeline to the GPU.

3.5 Discussion

The novelty of ACCESS is using the group behavior of normal controls to score the collaterals in ischemic stroke patients. We used the robust fRMC approach to obtain the sparsity and low-rank metrics of blood vessels in 4D CTA. Most previous methods used other imaging techniques or did not consider the fRMC approach to score collaterals, and thus, direct comparisons with the state-of-the-art methods are beyond the scope of the work.

Our proposed method has several advantages over previous techniques. First, very few automated techniques rely on 4D CTA which gives detailed and dynamic filling information of collaterals. The *automatic* approaches that do exist for collateral supply evaluation use

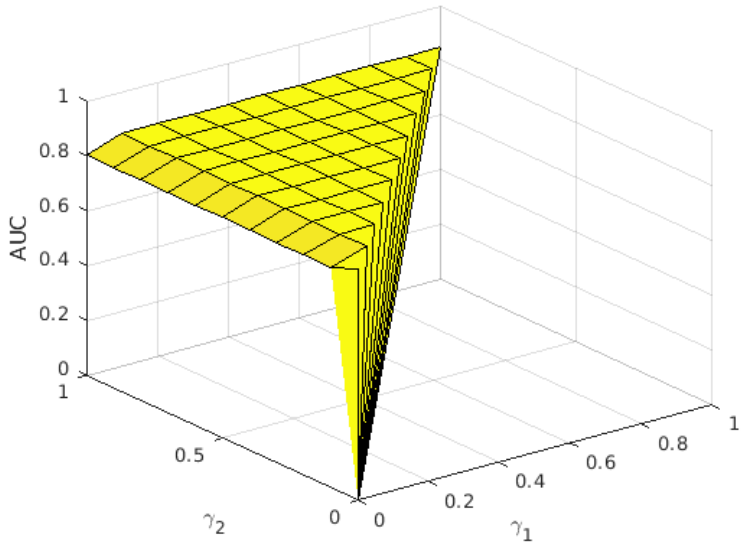

Figure 3.10: Demonstration of γ_1 and γ_2 choices on AUC of the overall system.

single-phase CTA, which may result in inaccurate estimation of collaterals due to suboptimal selection of a time point for scanning. Second, our approach is less dependent on feature selection and training data, potentially making it more robust in practice. Specifically, machine learning-based automatic approaches can be reliable only when there is a large dataset to overcome overfitting. In contrast, the developed ACCESS method is reliable as it is less independent of training. Thus not only does our method take advantage of full dynamic flow information from 4D CTA but it is also automatic and yields results inline or better than previous methods. Furthermore, the assessment of the final score resembles the definition of collateral circulation status, which is more intuitive for physicians to employ.

Experiments as well as existing studies [74] have proved that the reference parameters from the state-of-the-art medical imaging registration toolkit, ANTs, used in the registration blocks of the pipeline are robust for our applications (i.e., registration to the template and fine alignment of vessels). To test the impact of vessel segmentation parameters and the threshold used in post-processing, a sensitivity analysis was performed to obtain the optimal parameters for final evaluation. Based on the optimal parameters, collateral grading performance is analyzed using an ROC curve to find cutoffs for scores from different cutoff value settings in a leave-one-out approach. The experimental design of parameter choices with the best sensitivity and specificity gives a satisfactory AUC for scoring collateral circulation. This also makes our method more robust.

As noted previously, fRMC is independent of tunable parameters. To assess the impacts of the convergence parameters, γ_1 and γ_2 in the extended Frank-Wolfe optimizer, we evaluate the performance of ACCESS varying γ_1 and γ_2 in the step size of 0.1 following the range of $0 \leq \gamma_1 \leq \gamma_2 \leq 1$ with best τ obtained from previous sensitivity analysis. It can be seen in Fig. 3.10 that ACCESS performance in terms of AUC is fairly stable across different γ_1 and γ_2. Note that when only γ_1 is set to 0, the AUC decreases slightly (from 0.85 to 0.81).

Although our method does not require much data for the evaluation, in the future, the cohort of normal subjects can be enriched for better performance. A more thorough validation of ACCESS is still needed, with a larger dataset from different CT scanners and

acquisition protocols, and this will be explored in future work.

The extent of collaterals in intermediate and good subjects was very similar in some cases. Since inter- and intra-rater variability still remains for scoring those subjects, we will seek further validation of our method with ground truth labels collected from the consensus of more raters to improve data annotation quality.

3.6 Conclusion

In this work, we proposed the ACCESS method for automatic scoring of collateral circulation in the context of treatment decision-making in ischemic stroke. To the best of our knowledge, it is the first approach with low-rank and sparse decomposition for collateral score evaluation in ischemic stroke using 4D CTA. With an analog to existing collateral scoring protocols and being less reliant on machine learning methods that require large amounts of training data, the approach may be more robust than human-rater scoring and more easily comprehensible in the clinical environment.

Chapter 4

Deep Learning for Collateral Evaluation with Imbalanced Data

Preface

This chapter is based on our published paper [79].

In the previous chapter 3, we performed an automatic assessment of collaterals following radiologists' criteria of scoring using a quantification method. Recognizing the inherent benefits of deep learning, which excels in automatic feature extraction, our goal was to integrate it into the development of our collateral scoring process. Therefore, in this chapter, we extend our efforts by using deep learning to automatically extract features from the 4D CTA. More specifically, our objective is to assess collaterals from 4D CTA scans by harnessing the power of deep-learning networks, which have gained significant popularity due to their ability to tackle complex problems and offer advantages in automatic feature learning. However, a notable scarcity of large, publicly available databases containing data from ischemic stroke patients has constrained the effective utilization of deep learning techniques in making treatment decisions. Our contribution here lies in addressing the challenges of limited dataset size through applying transfer learning.

To achieve this, we have employed transfer learning by leveraging a pre-trained EfficientNet B0 network to enhance the evaluation of collaterals, employing both slice-based and subject-level classification. Our approach involves the stacking and overlapping of 2D slices extracted from a patient's 4D computed tomography angiography (CTA) scans, and we determine the patient's final collateral grade through a majority voting scheme based on the classification results of all these 2D Maximum Intensity Projections (MIPs).

Furthermore, we have incorporated a mechanism to handle class imbalance during the evaluation process. Specifically, we utilize focal loss with class weights, which penalizes the majority class appropriately, ensuring a more balanced and accurate assessment. This work has been published in the International Journal of Computer-Assisted Radiology and Surgery (IJCARS), and the GitHub code is available at: https://github.com/mumuaktar/DL_imbalanced

Abstract

Collateral evaluation is typically done using visual inspection of cerebral images and thus suffers from intra- and inter-rater variability. Large open databases of ischemic stroke patients are rare, limiting the use of deep learning methods in treatment decision-making. We adopted a pre-trained EfficientNet B0 network through transfer learning to improve collateral evaluation using slice-based and subject-level classification. Our method uses stacking and overlapping of 2D slices from a patient's 4D computed tomography angiography (CTA) and a majority voting scheme to determine a patient's final collateral grade based on all classified 2D MIPs. Class imbalance is handled in the evaluation process by using the focal loss with class weight to penalize the majority class. We evaluated our method using a 9-fold cross-validation performed with 83 subjects. A mean sensitivity of 0.71, specificity of 0.84, and a weighted F1 score of 0.71 in multi-class (good, intermediate, poor) classification were obtained. Considering the treatment effect, a dichotomized decision is also made for collateral scoring of a subject based on two classes (good/intermediate and poor), which achieves a sensitivity of 0.89, and specificity of 0.96 with a weighted F1 score of 0.95. An automatic and robust collateral assessment method that mitigates the issues with the small imbalanced dataset was developed. Computer-aided evaluation of collaterals can help decision-making of ischemic stroke treatment strategy in clinical settings.

4.1 Introduction

Ischemic stroke, caused by blockage of an artery that supplies blood to the brain, is one of the leading causes of disability and death worldwide. To limit neuronal cell death due to a lack of oxygen and nutrient flow and avoid patient disability, diagnosis and treatment of stroke must be expedient. Restoring blood flow to the blocked arteries, known as recanalization through endovascular thrombectomy, is one of the best treatment options for patients that have a small infarct volume, large penumbra size, and sufficient collateral circulation (a secondary vascular network activated when the primary arteries fail to supply sufficient blood flow to the ischemic area of the brain) [5]. Patients with good collateral circulation are strong candidates for revascularization as collateral circulation allows for tissue viability and leads to favorable functional outcomes in reperfusion therapies [13]. Conversely, patients recanalized with poor collaterals can be affected by symptomatic hemorrhagic transformation (sHT), which may lead to death caused by reperfusion injury [13]. Thus, evaluating the extent of collateral circulation is an important factor in ischemic stroke treatment decision-making.

The evaluation of collaterals is most often performed by radiologists using visual inspection of computed tomography (CT) images and thus has several drawbacks, including intra- and inter-rater variability [22]. A recent study by Grunwald *et al.* [25] showed that between individual neuroradiologists, the intraclass correlation coefficient ranges from 0.42 to 0.86 and score agreements range from 36.2% to 81.6%. Furthermore, the impact of intra- and inter-raters' variability through statistical analysis with moderate to substantial agreement between raters has adverse effects on clinical outcomes, particularly when misclassifying good and poor categories [60].

We aim to develop an automatic and efficient assessment of collateral scoring in ischemic stroke using a CNN with transfer learning to mitigate the issues of intra- and inter-rater variability and small and imbalanced datasets. To mitigate the issue of data deficit, we use:

(1) a pre-trained network with fine-tuning to classify collateral scores using 4D computed tomography angiography (CTA) data and (2) boosting of data samples by using 2D axial maximum intensity projections (MIPs) obtained by stacking and overlapping 2D slices from the patient's 4D CTA.

4.2 Related Work

With the growing popularity of deep learning, recent studies [53, 54, 47, 80, 48, 81, 4] have used convolutional neural networks (CNNs) in collateral evaluation, ischemic damage detection, as well as in stroke classification. Do *et al.* [54] proposed to evaluate ASPECTS [26] based on DWI by using a recurrent residual convolutional neural network (RRCNN) and obtained an accuracy of 87.3%. This approach outperforms their previous approach [53] based on a 3D CNN with an overall accuracy of 81%. More recently, Su *et al.* [47] performed deep-learning-based vessel extraction from 3D CTA and automatic collateral scoring with the best overall accuracy obtained at 80% in 4-grade scoring and 90% with dichotomized classification using a random forest classifier. The study by Tetteh *et al.* [48] proposed an automatic collateral grading approach distinguishing no flow, moderate, and good collateral flow in 183 patients using CNN with better performance over the classic machine learning approaches with an overall accuracy of 0.72. A 3D CNN was used by Neethi *et al.* to classify ischemic, and hemorrhagic stroke patients and normal subjects using 234 NCCT brain images. Kim *et al.* [4] used a deep CNN model to evaluate collaterals automatically into good or poor categories using a larger dataset of 327 patients. The study obtained an accuracy of 0.78, a sensitivity of 0.68, and a specificity of 0.84 along with a good agreement with experts' manual grading evaluating 72 patients.

Deep learning methods require a large training set, but unfortunately, large CTA-labeled databases of ischemic stroke patients are not publicly available. Transfer learning [57] can be used to mitigate the lack of training data. The capability of transfer learning to mitigate the issue of overfitting due to limited medical data was shown in the study by Swati *et al.* [82] who classified brain tumors and obtained an average accuracy of 94.82% using pre-trained VGG19 network. Jung *et al.* [80] proposed an adaptive transfer learning approach using a SEResNext model to classify ischemic stroke in NCCT. With the proposed pre-processing and adaptive transfer learning, they achieved an improvement of 18.72% in performance over an existing approach using a dataset of 356 patients. Recently, EfficientNet [58] outperformed state-of-the-art CNNs in transfer learning. This architecture scales the model's depth, width and resolution in a balanced way through a compound coefficient, enhancing the network's performance. Since EfficientNet [58] outperforms the existing transfer learning network, in this study, this model has been utilized to evaluate collaterals automatically mitigating the issues of small imbalanced dataset.

4.3 Materials and Method

4.3.1 CTA Data

The dataset includes 83 patients with 4D CTA imaging captured on a Toshiba Aquilion ONE 320-row detector 640-slice cone-beam CT (Toshiba medical systems, Tokyo, Japan). Each 4D CTA case consists of 19 intermittent volume scans performed for 60 seconds with a scanning speed of 0.75 s/rotation. Following the ASPECTS [26] protocol, two radiologists

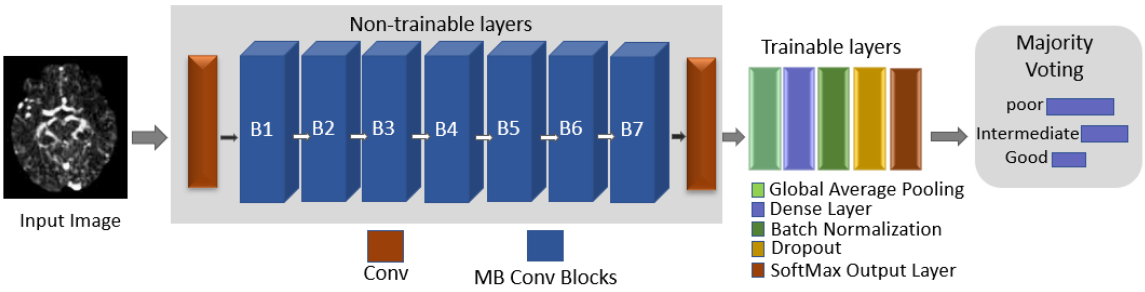

Figure 4.1: Our network architecture of EfficientNet B0 with transfer learning. MBConv Blocks(B1-B7) are the basic building blocks of EfficientNet B0 which stand for inverted residual blocks originally applied in MobileNetV2 [1].

graded collaterals visually using the 4D CTAs and considering the middle cerebral artery (MCA) territory. Based on the extent of collateral circulation, each patient was labeled as having good (100%), intermediate ($> 50\%$ & $< 100\%$) or poor ($< 50\%$) collaterals. We had a total of 53 good, 21 intermediate, and 9 poor cases.

4.3.2 Pre-processing

All images were first registered non-linearly to a template space which was created following an unbiased group-wise registration approach [66] with 12 healthy subjects' CTAs using the symmetric image normalization method (SyN) [65]. Low-rank decomposition was used to remove any vessel calcification and white and grey matter and to extract blood vessels from the background as in [60]. After low-rank decomposition, the volumes with visible vasculature filling in the MCA territory have a transition of total non-zero pixel counts compared to volumes without contrast agents in these areas (As shown in Fig. S1 of supplemental Section 4.6). For each subject, the relevant (i.e. those that have flow information) CTA volumes were extracted automatically. From the filtered 4D volumes, a 3D maximum intensity projection (MIP) blue created, and 2D MIPs obtained by stacking and overlapping 2D slices from the 3D MIP in the axial plane were considered as independent samples during the training and testing of the network.

4.3.3 Network Architecture

We used EfficientNet B0, which was pre-trained on the ImageNet dataset [83]. Since the last layers of the network are trained specifically for the 1000 classes of the ImageNet dataset, those were replaced with a new dense layer. Adding an extra layer with an optimal number of filters helped avoid the misclassification of intermediate cases. The output size of the last layer was changed to classify good, intermediate, and poor cases of collaterals in the place of 1000 units from ImageNet (see Fig. 4.1).

4.3.4 Training Details

As in [84], the images were resized to 224×224 with 3 channels to be compatible with the EfficientNet architecture. We performed 9-fold stratified cross-validation to confirm the same distribution of data from each class in every fold. The splitting was performed at the subject level before the model training with slices to avoid any data leakage in testing new

cases. Before running cross-validation to evaluate all the subjects' classification accuracy, the parameter set for training the model was evaluated through an ablation study performed at the slice level using a subset of the whole dataset in both dichotomized and multi-class classification. A batch size of 16, filter size of 128 for multi-class, and 256 for dichotomized were identified through the ablation study (additional details available in supplemental Section 4.6 in Fig. S2 and Fig. S3, respectively). The Adam optimizer with a learning rate scheduler was used. The learning rate was reduced from 0.01 by a factor of 0.5 if there were no improvement after 7 epochs (Example curve in Fig. S4 in supplemental Section 4.6). Early stopping and dropout (rate=0.2) were used to prevent overfitting, and vertical flipping, random rotations, translations, and contrast changes were applied during training for data augmentation.

4.3.5 Effect of Loss Function

Focal loss, which reshapes the standard cross-entropy loss as in Equation (10) [85], was used to weigh down the contribution of the good class in training the model as the data is highly imbalanced:

$$\text{Focal loss, } F_l = -(1-p_b)^{\gamma} log(p_b) \tag{6}$$

where the term $(1-p_b)^{\gamma}$ works as a modulating factor to adjust the cross-entropy loss using γ=2 based on the higher or lower value of probability, p_b. Along with this, higher class weights (calculated based on the total number of samples) were assigned to poor and intermediate class samples compared to the good class to prevent the easy examples from oppressing the model training. The final loss function considering the class weight is expressed as follows, akin to the approach in the reference study [85], where the authors introduced an α-balanced variant of the focal loss to enhance accuracy:

$$\text{Focal loss, } F_l = -\alpha_t(1-p_b)^{\gamma} \log(p_b) \tag{7}$$

Here, α_t represents the class weight under consideration in our specific case.

An example is used (See Fig. S5 in the supplemental Section 4.6) to demonstrate that for both types of classification, the focal loss is minimized more than the traditional cross-entropy loss.

4.3.6 Collateral Circulation Evaluation

Collateral scoring was performed in two stages: (1) slice-based classification with the EfficientNet B0 using transfer learning and (2) subject-level classification with majority voting. Collateral assessment was performed using both multi-class classification as good, intermediate, and poor collaterals and dichotomized classification combining good and intermediate to one class and poor as another class of collaterals. This dichotomized evaluation was performed similarly to the study of Su *et al.* [47] where the Tan score [31] is converted into two groups: 0-1 (no collaterals-poor collaterals) and 2-3 (moderate-good collaterals) since the first group did not affect treatment whereas the second group showed a substantial effect in treatment in the MR CLEAN trial. We thus looked at both a multi-class and dichotomized assessment using the same settings of model and parameters.

4.3.6.1 Slice-based Classification

For slice-based classification, 2D slices were used from 4D CTA using a stacking and overlapping strategy to increase the data samples. Previous studies have followed various

slicing techniques, for example, strided slicing by Neethi *et al.* [81]). Among the total 160 2D axial slices of the dimension of 221×221 from the 3D MIP of a 4D CTA subject, slices 1 to 20 and 146 to 160, which contain little to no vascular information, were discarded similar to the study by Sarraf *et al.* [86]. For slices between 21 and 145, 2D MIPs of the slices being stacked and overlapped were used so that features related to blood flow through the vessels could be better captured. The final strategy for obtaining 2D MIPs was decided through sensitivity analysis (Further details available in supplemental Section 4.6 Table S1).

4.3.6.2 Subject-level Classification

To classify each subject as good, intermediate, or poor, the best model was applied to all the 2D MIPs of that subject. To determine the class of the test subject, a majority voting approach was used, as described in Sarraf *et al.* [86] which works as follows: considering a test subject, T_s with a total number of N 2D MIPs which are classified at first by the proposed model as good, intermediate or poor classes. The probability of a class being good, intermediate, or poor blue calculated based on the total number of 2D MIPs that belong to that class, N_c such that $P_c = \frac{N_c}{N}$. The final class for each subject, T_s is thus the class with the highest probability.

4.3.7 Transfer Learning Strategy

To perform slice-based classification with the EfficientNet B0, we used the weights from the pre-trained ImageNet dataset. Only the weights in the top layers (layers close to the output) were tweaked to extract specific features from our dataset. Fine-tuning performance varies based on the dataset and problem domain as shown in the studies by [87] and [88] where layer-wise tuning and fine-tuning a pre-trained network rather than the full network respectively showed more robustness. According to the study by [89], the earlier layers of the pre-trained network mainly concentrate on image edge and color information whereas the top layers focus on details of the specific class label which needs more fine-tuning over the pre-trained weights. Similar to these studies, in our case, rather than fine-tuning the convolutional layers' weights, adding a dense layer along with batch normalization and dropout and tweaking the weights of the dense layer with the output layer helped balance the performance of each class given the few examples in the poor class.

4.3.8 Performance Metrics

As the dataset is imbalanced (typical in the domain), accuracy is not an appropriate metric to evaluate performance; therefore, sensitivity, specificity, and weighted F1 score are used (Table 4.1). Specificity is reported to show the true negative rate which has a significant impact on treatment decision-making for ischemic stroke.

4.3.9 Experimental Setup

Training and testing were done with an Intel(R) Core(TM) i5-7500 CPU @ 3.40GHz and 48 GB of RAM processor with an NVIDIA TITAN Xp GPU with 12 GB of RAM. The whole processing of training and testing a subject took around 1.5 hours for dichotomized and around 2 hours for multi-class classification.

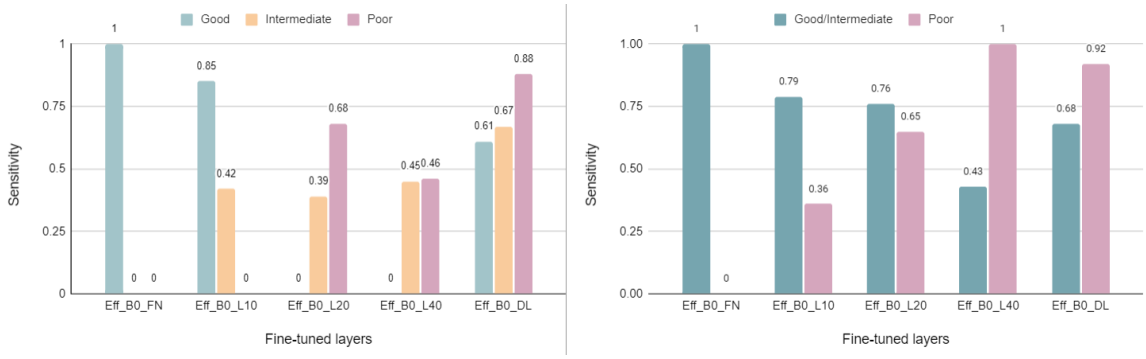

Figure 4.2: Fine-tuning experiments. Left: Multi-class, Right: Dichotomized. FN: Full network; L_10: Last 10 layers; L_20: Last 20 layers; L_40: Last 40 layers; DL: Dense layer. For multi-class, unfreezing any layer except the added dense layer failed to classify any sample from the poor class. In dichotomized, the poor cases are classified unfreezing different blocks of layers with best performance obtained by tuning dense layer only.

4.4 Results

4.4.1 Fine-tuning

We ran five experiments unfreezing different blocks of layers of the EfficientNet B0 network. This was done by tuning only the added dense layer, training the last 10 layers along with the added dense layer, making the last 20 layers trainable, making 40 layers trainable along with the added dense layer, and finally, training the whole EfficientNet B0 model. We have shown the performance of fine-tuning experiments unfreezing different numbers of layers of the EfficientNet B0 network at the slice level (Fig. 4.2) using a subset of the dataset. In the case of both classifications (Fig. 4.2), unfreezing the full network failed to classify any sample of poor cases. This is because the full network training caused the loss of the general image features obtained from the pre-trained convolutional layers, thereby leading to overfitting the small number of poor cases. The maximum number of slices from good, intermediate, and poor classes are classified, tuning only the added dense layer for both types of classification. To obtain the optimal result at the subject level, we have analyzed the performance of slice-based classification over 9-folds with a mean sensitivity of 0.66 for multi-class and 0.71 for dichotomized classification. The mean sensitivity of individual classes was 0.69 (and SD= 0.11), 0.64 (and SD=0.10), and 0.64 (and SD=0.17) for good, intermediate, and poor classes respectively (with a minimum average sensitivity of 0.57 and a maximum average sensitivity of 0.74) for multi-class classification. For dichotomized classification, the mean sensitivity of 0.67 (and SD=0.1) for the good class and 0.75 (and SD= 0.15) for the poor class were obtained with a minimum average sensitivity value of 0.60 and a maximum average sensitivity value of 0.80. Although the slice-based classification showed moderate performance with a slightly higher deviation over 9-folds in case of poor cases (due to lack of training samples) by using majority voting, a good performance in subject-level is achieved.

Since our main target is to evaluate individual subjects with majority voting based on slice training, the fine-tuning result was checked at the subject level as well. Fine-tuning in the dichotomized classification performed on the last 20 layers of the network gave better results for the good (sensitivity = 0.81) class but caused overfitting of the poor

cases (sensitivity = 0.56) due to a lack of samples.

4.4.2 Performance Comparison

We compared our model to a pre-trained VGG16 network and a recent study using a CNN by Kim *et al.* [4] applied to our dataset. To evaluate the VGG16 model's performance for both dichotomized and multi-class classification of collaterals, the best parameter settings obtained from the EfficientNet B0 hyperparameter tuning were used. The last 2-fully connected layers of VGG16 were replaced by a dense layer of filter_size=128 for multi-class and filter_size=256 for dichotomized classification following a dropout layer. The study of automatic grading into good and poor collaterals by Kim *et al.* [4] utilized 255 patients' data to train and 72 external data for validation. To prepare data for training their method, arterial, capillary, and venous phase images were extracted using the scanning protocol of our 4D dataset (peak phase of volumes 5-10 for the arterial, 11-14 for the capillary and 15-19 for the venous phase) to make the 3 channel image. The DL architecture used in the study of Kim *et al.* [4] comprised of 4 convolutional layers with ReLu activation following max pooling layers and 2 fully connected layers. Since the number of filters used in each layer of the study is not mentioned, we have fine-tuned the network and obtained a best model with 16,32,64,128,256 and 512 filters sequentially for the 6 layers. Using similar data augmentation and learning rate as in [4], the model did not perform well, and thus to match with our proposed model for comparison, the same augmentation and learning rate scheduler were applied as our model.

Table 4.1 shows the results of all classification models in the subject level. As can be seen, VGG16 has a reasonable performance in the case of dichotomized classification with a sensitivity of 0.91 (the proposed method's sensitivity is 0.96) while for multi-class classification, it has an average sensitivity of 0.58 (our proposed method has an average sensitivity of 0.71). The performance of the proposed model as well as VGG16 for classifying intermediate cases in multi-class classification has poor sensitivity compared to other classes. This is due to the less evident visual difference between good and intermediate cases which leads the model to classify intermediate cases as good. It is evident that our proposed approach has shown better performance (weighted F1-score=0.95) over the CNN Proposed by Kim *et al.* [4] (weighted F1-score=0.72) in dichotomized classification and therefore, for all cases, our proposed method has the highest sensitivity, specificity and weighted F1-score over the other compared methods. The method of Kim *et al.* [4] performed worse due to the lack of necessary training samples for both classes as the DL method needs a large amount of data such as was used in their study [4].

4.5 Discussion and Conclusions

We proposed an automatic approach to collateral scoring using the pre-trained EfficientNet B0 network and transfer learning with a limited and imbalanced training dataset. This is a novel method for automatic collateral evaluation using 4D CTA, which overcomes the problem of the unavailability of large stroke datasets. The concept of considering 2D MIPs following the overlapping and stacking strategy of 2D axial slices also helps to attenuate the effect of lack of data as well as provide significant sequential distinguishing features.

Deep networks have an advantage over traditional machine learning approaches in that specific features are extracted automatically rather than being hand-crafted. Extracting generic image features from the pre-trained EfficientNet B0 network can also speed up

Table 4.1: Performance comparison of the proposed model at subject-level with pre-trained VGG16 network and recent study of collateral grading by Kim *et al.* [4]

Type of collateral Grading	**Network**	**Collateral Grades**	**Sensitivity**	**Specificity**	**Weighted F1-Score**
Multi-class classification	EfficientNet B0	Good	0.72	0.80	0.71
		Intermediate	0.62	0.82	
		Poor	0.78	0.89	
	VGG16	Good	0.68	0.63	0.61
		Intermediate	0.38	0.79	
		Poor	0.67	0.88	
Dichotomized classification	EfficientNet B0	Good/ intermediate	0.96	0.89	0.95
		Poor	0.89	0.96	
	VGG 16	Good/ intermediate	0.91	0.67	0.89
		Poor	0.67	0.91	
	Proposed CNN by Kim *et al.* [4]	Good/ intermediate	0.66	0.55	0.72
		Poor	0.55	0.66	

convergence. Although the study by Do *et al.* [53] considered 3D CNN models to assess ASPECTS automatically with DWI, the study needed a larger dataset of 312 subjects to detect early ischemic changes whereas our proposed approach can evaluate collaterals using a small number of samples with 2D MIPs by taking advantage of transfer learning. The performance of our study is also comparable to the study by Su *et al.* [47] where collaterals were scored automatically using random forest after vessel extraction through DL. However, their approach used hand-engineered features based on the ratios of vessel length and volume. The recent study by Kim *et al.* [4] to grade collaterals into good and poor categories with which the proposed model is compared, performed the classification with CNN using a large dataset of 327 patients and it failed to distinguish the two classes with less training samples.

Our method has shown that tuning only the output layer along with a new dense layer, which mostly helped in evaluating intermediate cases, yields better performance over tweaking weights of the convolutional layers. This is because the convolutional layers of EfficientNet B0 mostly extract low-level features [89] whereas fine-tuning the last layers helps extract the distinguishing features of the 3 classes. This finding matches the outcome by the study of Tajbakhsh *et al.* [87]. Although layer-wise tuning showed better performance in that study, tuning the last fully connected layers can be sufficient when using transfer learning since those layers can extract problem-specific distinguishing high-level features. Starting from the pre-trained weights with general low-level features can save the network from the undesirable local minimum which is caused by iterative random weight updates in a small labeled dataset [87]. The robustness of the fine-tuned pre-trained network to the size of the training dataset over the full-trained network was also supported in the study of Mehra *et al.* [88] where the pre-trained VGG16 yielded the best performance with the highest accuracy as a feature extractor.

The developed method is able to handle extremely imbalanced class samples through the use of focal loss where the majority class is penalized. This allowed us to classify most cases correctly in the dichotomized experiment. The dichotomized classification shows better performance over the multi-class and clinically is more important in terms of endovascular treatment decision-making. Further, we can utilize temporal information to better distinguish good and intermediate cases to enhance the performance of multi-class collateral evaluation, too.

Although 2D MIPs from 4D CTA helped overcome the lack of training data and stacking strategy provided significant vascular flow, considering the 3D MIPs further from 4D CTA with 3D deep transfer learning networks should be considered in the future work. Since some important vascular features can be missing while using 2D samples from 4D temporal data, 3D Siamese network as few-shot learning can be an interesting extension to this research to handle small and imbalanced dataset similar to the study of Osama *et al.* [90] and Liu *et al.* [91].

Our method has shown to be effective in the case of small databases, which is common in some medical imaging domains. The developed method shows better generalization ability due to using data augmentation and an extra dropout layer on top of the EfficientNet architecture. Thus, the developed method has the potential to mitigate the issues in current visual inspection-based assessments, potentially benefiting both the patients and clinical workflow.

4.6 Supplementary material for Chapter 4

4.6.1 Extraction of relevant CTA volumes

Fig 4.3 refers to the automatic extraction of significant volumes from a 4D CTA.

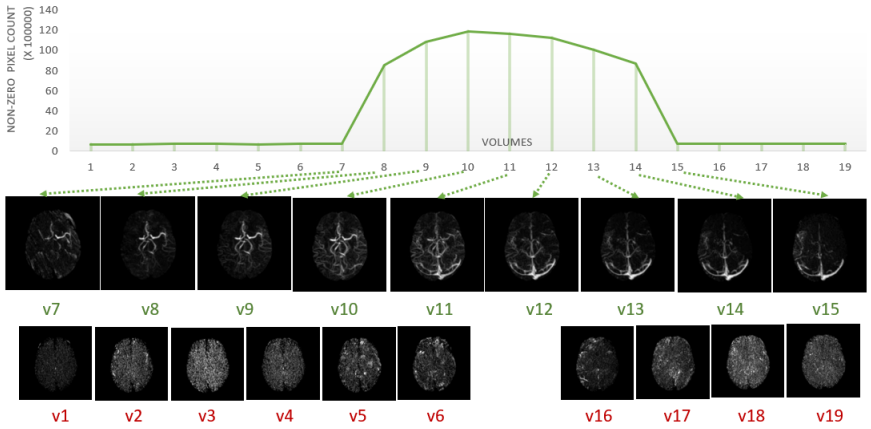

Figure 4.3: Curve showing the continuous flow of contrast agent in informative volumes (v7-v15) and separated non-informative volumes (v1-v6, v16-v19) of a subject extracted automatically based on non-zero pixel count.

Since the timing in contrast filling to MCA varies between subjects, the number of selected relevant volumes varied (e.g. for one subject, 7-15 (total: 9 volumes) were the relevant volumes whereas for another subject it was volumes 8-13 (total: 6 volumes)).

4.6.2 Ablation study

A subset of the training set is at first considered to identify the batch size for both types of classification. As seen in Fig. 4.4, although the performance fluctuates with batch size (and more so in the case of dichotomized classification), for a batch size 16 both types of classification have the highest performance.

we ran several experiments for both dichotomized and multi-class classification to identify the filter size for the dense layer. The result can be found in Fig. 4.5 which shows

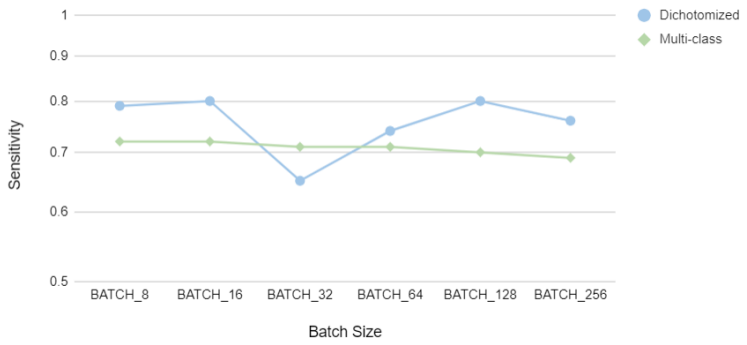

Figure 4.4: Determination of final batch size based on sensitivity obtained from the experiment of a small subset for both types of classification.

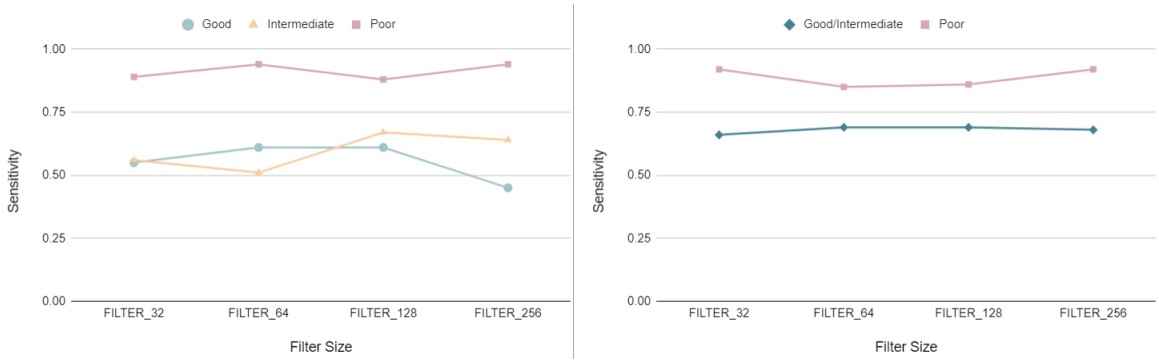

Figure 4.5: Sensitivity for varying filter sizes (*Left:* multi-class with filter_size=128 and *Right:* dichotomized with filter_size=256 shows best average performance for classes).

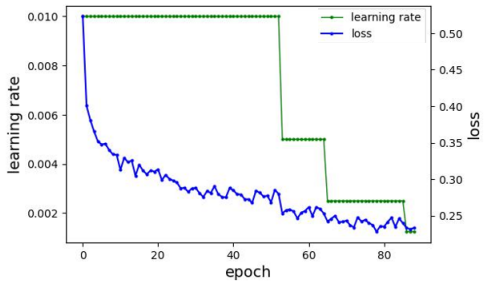

Figure 4.6: Effect of learning rate scheduler in loss minimization.

that best performance is obtained for multi-class and dichotomized classification with 128 and 256 filters respectively.

4.6.3 Selection of Learning rate

Fig 4.6 shows the changes of loss with varying learning rate. As epoch increases, loss is minimized. In epoch around 55, there is a transition of learning rate reduction to improve loss minimization further. It can be seen that at around epoch 90, with the reduced learning rate, the loss is minimized to 0.23.

4.7 Advantage of focal loss over cross-entropy loss

A curve 4.7 is drawn to see the performance of focal loss over cross-entropy loss. For both multi-class and dichotomized classification, focal loss starts from a lower value and minimizes more than the cross-entropy loss

4.7.1 Sensitivity analysis of slicing strategy

We evaluated four different methods (Table 4.2) for deciding the final slice stacking strategy. Each method uses slices 71 to 126 individually as this is the area that contains the important contrast filling information. On top of this we consider (1) using only single slices throughout, (2) merging every 10 sequential slices, (3) merging 10 sequential slices along with overlapping using an interval of 5 slices to increase the number of samples, and

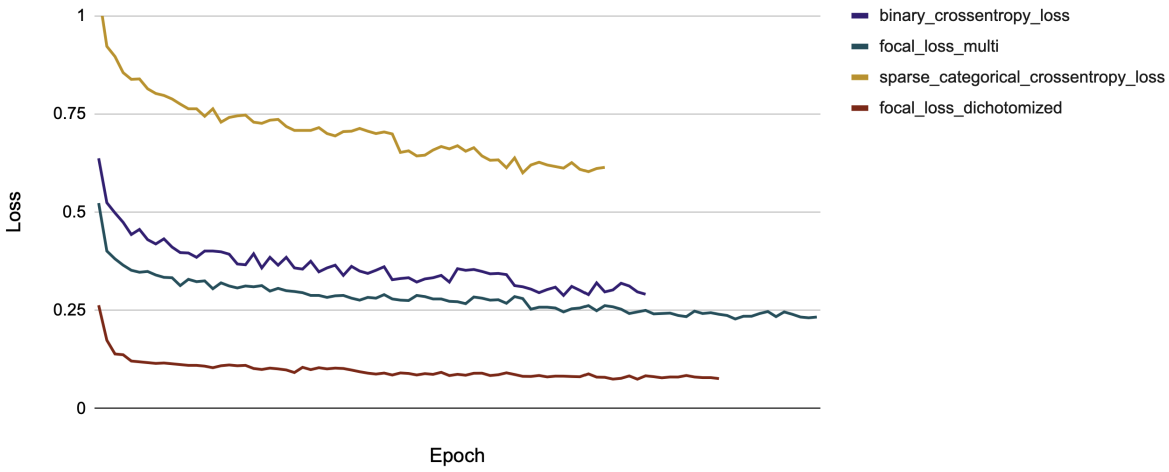

Figure 4.7: Example curve showing performance of focal loss over cross-entropy loss. For both cases, focal loss starts from a lower value and minimizes more than the cross-entropy loss.

Table 4.2: Slicing for 2D MIPs based on stacking and overlapping strategy

2D MIPs	Collateral Grading	Number of Samples	Sensitivity
Single Slices from 3D MIPs	Good	6625	0.63
	Intermediate	2625	0.51
	Poor	1125	0.39
Slices (10) merging; No overlapping	Good	2968	0.56
	Intermediate	1176	0.70
	Poor	504	0.56
Slices (7) merging with overlapping (5)	Good	3621	0.65
	Intermediate	1504	0.74
	Poor	721	0.69
Slices (10) merging with overlapping (5)	**Good**	**3498**	**0.68**
	Intermediate	**1386**	**0.77**
	Poor	**598**	**0.78**

(4) merging 7 slices with the same overlapping to check the performance. Based on the analysis and performance, the final slicing strategy of merging 10 slices and overlapping with an interval of 5 was done as follows:

- Slices 21 to 70 & 127 to 145: Every 10 slices stacked together into one MIP with an overlap of 5 slices (e.g. 21 to 30, 26 to 35, 127 to 136, etc.)
- Slices 71 to 126: used following the same overlapping strategy as well as individually (these contain the most distinguishing features between the three classes with the contrast agent filling the MCA territory.)

Based on this stacking, from a total of 83 subjects, there were 3498 2D axial MIPs for good, 1386 for intermediate, and 594 for poor cases.

Table 4.2 shows that although single slices helped increase the total number of training and testing samples, this has poor performance in collateral grading. We posit this is because the single slices provide fewer vasculature features. The best performance was obtained with merging and overlapping slices which helps increase data samples as well as provide better 3D vasculature information. Although 7 slices merging increased the data samples, it failed to supply more valuable vasculature features.

Chapter 5

A Radiomics-based Machine Learning Approach to Assess Collateral Circulation in Ischemic Stroke on Non-contrast Computed Tomography

Preface

This chapter is based on our published paper [92].

Our previous methods described in chapters 3 and 4 implemented automatic collateral evaluation approaches using 4D CTA, whereas this chapter demonstrates the use of non-contrast computed tomography (NCCT) scans, which are readily available and safe for clinical use for the collateral scoring with machine learning. The focus is on assessing collaterals from ischemic damage in stroke using the ASPECTS terminology. Collaterals were assessed using 4D CTA as a ground truth with radiologist observations. Radiomic features were extracted separately from the left and right hemispheres to capture non-symmetry between them and to categorize collaterals as good, intermediate, or poor using support vector machines (SVM). The method leverages NCCT to detect tissue degeneration and identify regions with insufficient collateral circulation. This work was presented in the Multimodal Learning for Clinical Decision Support and Clinical Image-Based Procedures (CLIP) MICCAI workshop in 2020. The Github link for the implementation code is available at: https://github.com/mumuaktar/MICCAI-CLIP

Abstract

Assessment of collateral circulation in ischemic stroke, which can identify patients for the most appropriate treatment strategies, is currently conducted with visual inspection by a radiologist. Yet numerous studies have shown that visual inspection suffers from inter and intra-rater variability. We present an automatic evaluation of collaterals using radiomic features and machine learning based on the ASPECTS scoring terminology with non-contrast computed tomography (NCCT). The method includes ASPECTS region identification, extraction of radiomic features, and classification of collateral scores with support vector machines (SVM). Experiments are performed on a dataset of 64 ischemic stroke patients to classify collateral circulation as good, intermediate, or poor and yield an overall area under the curve (AUC) of 0.86 with an average sensitivity of 80.33% and specificity of 79.33%. Thus, we show the feasibility of using automatic evaluation of collateral circulation using NCCT when compared to the ASPECTS score by radiologists using 4D CT angiography as ground truth.

5.1 Introduction

Acute ischemic stroke (AIS) caused by blocked arteries in the brain is one of the leading causes of death worldwide. Treatment strategies of AIS vary from intravenous tissue plasminogen activator (IV-tPA) to endovascular thrombectomy treatments (EVT) based on the time window and patients' conditions. EVT is one of the best treatments for restoring blood flow through blocked arteries but its success rate depends on the extents of a patient's collateral circulation. Collateral circulation is a secondary vascular network that is recruited temporarily that allows for the survival of viable brain tissues when the main conduits fail due to ischemic stroke. The extent of leptomeningeal collateral flow from the middle cerebral artery (MCA) flowing to the anterior cerebral artery (ACA) and posterior cerebral artery (PCA) has been shown to be a radiologic surrogate predicting the response of revascularization therapy [12]. However, scoring of the collaterals manually following conventional radiologic strategy suffers from the intra- and inter-rater variability [60, 22, 25], less reliable results, and is time-consuming. Some studies have compared the automated approach with visual inspection based on ASPECTS evaluation having greater agreement ($\kappa = 0.90$) than neuroradiologists [93] or slightly worse agreement than human expert ratings [94]. Therefore, developing automatic and robust approaches to collateral evaluation in AIS based on systematic radiologic criteria and methods is important. A number of different approaches have been proposed to score collaterals, for example, ASPECTS (Alberta Stroke Program Early CT Score) [26], the collateral score of the Society of NeuroInterventional Surgery/Society of Interventional Radiology (ASITN/SIR) based on conventional angiography [27] which is adapted to be applicable to dynamic computed tomography angiography (CTA) further in the study of Sekar *et al.* [41], the scores of Christoforidis *et al.* [28], Miteff System [29], Mass System [30], modified Tan Scale [31], regional leptomeningeal collateral score (rLMC) [32], collateral evaluation with 4D-CTA based on ASPECTS [42], [5], and ACCESS [60].

ASPECTS is one of the most reliable, systematic and robust approaches shown to have positive clinical outcomes in ischemic damage detection in many studies (e.g. [95, 96, 41, 97, 98, 99]) with baseline CTA source images(CTA-SI), CT perfusion images(CTP), contrast-enhanced CT (CECT), non-contrast CT (NCCT), and timing invariant CTA (TiCTA)

modalities respectively. Although some studies [100, 101, 95] show that contrast-enhanced CT has superior performance delineating brain vasculature, these are limited to manual intervention or semi-quantitative approaches.

Different studies (Brainomix: e-ASPECTS, e-CTA, iSchemaView: Rapid CTA, Rapid ASPECTS, Syngo.via Frontier ASPECT Score Prototype V2: not clinically approved) have focused on automating ASPECTS using artificial intelligence and feature-based machine learning. The e-ASPECTS® software from Brainomix Ltd. (Oxford, UK) and RAPID ASPECTS® by iSchemaView (Menlo Park, USA) are the only two certified clinical software for ischemic damage detection using ASPECTS on NCCT. Although they are not intended yet to be used as stand-alone diagnostic tools, both suggested NCCT as an alternative to CTP for ischemic damage quantification [102]. The ability of NCCT to work as the stand-alone diagnostic tool extracting much clinical information was shown by Sheih *et al.* [98]. In their work, the authors compared diffuse hypoattenuation and focal hypoattenuation on contralateral hemispheres in 10-ASPECTS regions and obtained an area under the curve (AUC) of 90.2% for a total of 103 subjects. The study by Kuang *et al.* [97] also performed contralateral analysis using a machine learning-based approach to assess early ischemic changes by classifying the 10-ASPECTS regions based on the differences of contralateral texture features. Taking diffusion-weighted imaging (DWI) as ground truth, this study with 257 patients obtained an AUC of 0.89 between the proposed method and experts' reading. Our work resembles Kuang *et al.* [97] in that it uses contralateral radiomic features, however unlike their work we evaluate collateral circulation rather than early ischemic damage.

Since NCCT is easily available and used as a front-line diagnostic tool in clinical settings, also being free from contrast agent that can cause adverse effects to some patients, we used NCCT to automatically assess scoring in acute ischemic stroke based on the ASPECTS terminology. Unlike most of the state-of-the-art methods of automating ASPECTS to obtain ischemic damage by assessing hypoattenuation using DWI as ground truth, we used 4D CTA as ground truth to evaluate collaterals scored through observing multiple phases by radiologists. We aim to evaluate collaterals using NCCT with radiomic feature extraction in the MCA territory of left/right hemispheres and classify them into good, intermediate, or poor categories with support vector machines (SVM). Since brain collaterals vary between individuals and represent symmetric characteristics between left/right hemispheres of the same individual, we extracted radiomic features from each side of the hemispheres of a subject separately and took the difference between them to obtain the non-symmetry. Fig. 5.1 shows the different collateral categories in contrast-enhanced CTA. To the best of our knowledge, this is the first study using NCCT to evaluate the collateral circulation. The underlying assumption of our novel approach is that we can identify regions with insufficient collateral circulation using radiomic features based on tissue degeneration which may be captured on NCCT and score the extent of collaterals using these features.

5.2 Materials and Methods

5.2.1 Scanning Protocols

We have evaluated our method with 8 poor, 17 intermediate, and 39 patients with good collaterals. All 64 subjects underwent imaging at the Montreal Neurological Hospital (Montreal, Canada). A Toshiba Aquilion ONE 320-row detector 640-slice cone-beam CT (Toshiba medical systems, Tokyo, Japan) scanner, which provides whole-brain perfusion and dynamic vasculature information in one single examination with a single rotation of

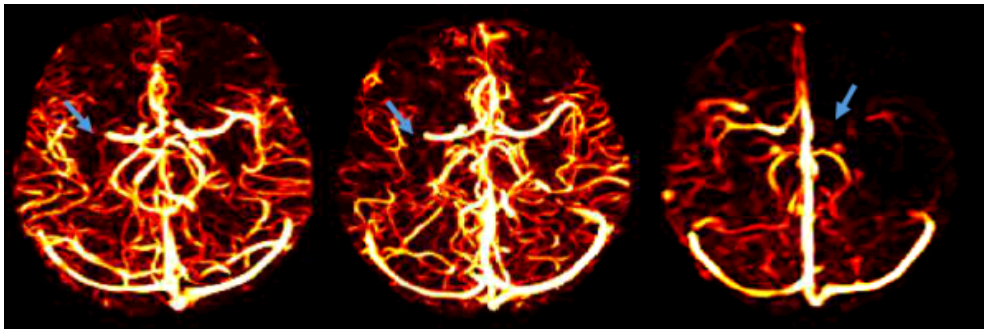

Figure 5.1: From left to right, an example of good, intermediate and poor collaterals on contrast-enhanced CTA. The blue arrow indicates the occlusion on the MCA.

the gantry, was used to capture the 4D CTA. A series of intermittent volume scans are performed for 60 seconds with a scanning speed of 0.75 s/rotation to capture a total of 19 volumes for each patient with low-dose scanning for every 2s during the arterial phase and 5s during the venous phase. The 19 volumes are divided into 5 groups according to the tube current. The first volume, which we used in our method, was captured before contrast arrival. Isovue-370 (Iopamidol) was used as the non-ionic and low osmolar contrast medium to visualize the vessels in the remaining volumes.

5.2.2 Ground Truth Labels

The ground truth for collateral circulation was based on scoring by two radiologists examining the 4D CTA images. Following the ASPECTS scoring terminology, the radiologists' defined patients with 0-50% collaterals as poor, greater than 50% and less than 100% as intermediate and 100% collaterals as good.

5.2.3 Mapping of ASPECTS Regions

Prior to the collateral evaluation, we employed atlas registration to align the 10-ASPECTS regions in all patients. The atlas is generated by extracting the ASPECTS regions from the MNI structural atlas and Harvard Oxford atlas using FSL. Following the ASPECTS score in acute stroke[1] and the study of Pexman *et al.* [26], we extracted the insular ribbon (I), caudate (C), Lentiform nucleus (L), internal capsule (IC), M1 (anterior to the anterior end of the Sylvian fissure including frontal operculum), M2(anterior temporal lobe), M3(posterior temporal lobe), M4 (Anterior MCA territory), M5 (Lateral MCA territory) and M6 (Posterior MCA territory). These are rostral to basal ganglia and approximately 2cm superior to M1, M2 and M3, respectively.

Three steps were performed to map the atlas to the subject brain. First, as the atlas we extracted is in MNI template space, it was mapped onto an average CT template (created using 12 healthy subjects' brains following the unbiased group-wise registration approach by Fonov *et al.* [66]) using affine registration. Next, the CT template was registered to the subjects' native space using symmetric normalization (SyN). Finally, the transformations obtained from the previous steps were used to map the atlas to each subject to delineate specific ASPECTS regions in all subjects. All registration steps were done using ANTs

[1] http://aspectsinstroke.com/

(Advanced Normalization Tools)[2]. We extracted the brain from each subject using FSL following the study of Muschelli *et al.* [103]. Fig. 5.2 shows the ASPECTS regions mapped onto a subject.

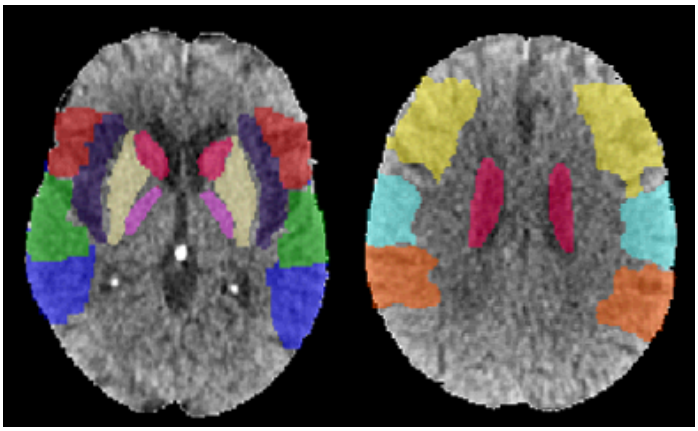

Figure 5.2: The 10-ASPECTS regions mapped to an individual patient's brain

5.2.4 Pre-processing

The skull was removed from the NCCT images following the study of Muschelli *et al.* [103]. In this work each image is thresholded within the brain tissue range of 0-100 Hounsfield units (HU) before skull stripping thus removing any calcifications with very high-intensity values, which were present in some of our patients. Further, a Gaussian pyramid from the multi-scale image representation approach using the kernel from the study of Burt *et al.* [104] was applied to perform smoothing and sub-sampling by one level to all subjects.

5.2.5 Image Features

The study of Shieh *et al.* [98] shows that ischemic damage can be identified through focal and diffuse hyper attenuation occurring in any ASPECTS region. Following the study of Shieh *et al.*, a deviation and a contrast map were extracted from each side of the brain in order to highlight the areas with insufficient collaterals due to ischemic damage. The deviation and contrast degradation in the areas with less collaterals can be obtained comparing the difference between affected and unaffected sides of the brain. Therefore, the deviation map, D_{map} is generated from each side of brain by subtracting each voxel's intensity, $V(x, y, z)$ from the mean voxel intensity, V_μ and normalizing it with the standard deviation, V_σ using the equation 10.

$$D_{map} = \frac{V(x, y, z) - V_\mu}{V_\sigma} \tag{8}$$

On the other hand, a contrast map is obtained from each side choosing the edges with maximum gradient using the Sobel operator. Rather than comparing the deviation and contrast map between both sides using threshold as the study of Shieh [98], radiomic features are extracted automatically from the maps separately to obtain the spatial relation of voxels and finally the difference between features from each hemisphere is taken (Fig. 5.3 shows

[2]stnava.github.io/ANTs

the feature maps with the radiomic features). The feature classes used include radiomic features for the 3D subjects from the study of Van *et al.* [105], as described below.

5.2.5.1 Gray Level Co-occurrence Matrix (GLCM):

The co-occurrence of voxels for specific values are used to examine the textures of the images by statistical measurement of energy, contrast, entropy, homogeneity, correlation, variance, sum average, autocorrelation, and dissimilarity.

5.2.5.2 Gray Level Size Zone Matrix (GLSZM):

The gray level zone of each subject is quantified by computing the number of voxels that share the same gray level intensity in a 26-connected region.

5.2.5.3 Global Features:

Along with GLCM and GLSZM, the global features of mean, energy, entropy, variance, skewness, and kurtosis of the entire MCA territory from both sides were considered.

In total, 56 features are extracted from the deviation and contrast maps from each side of the brain.

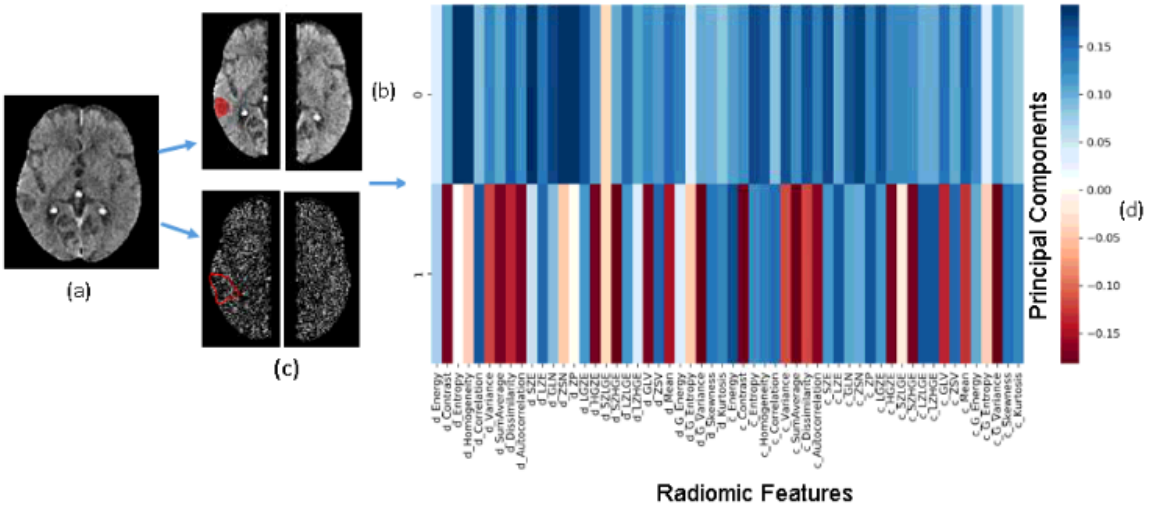

Figure 5.3: Feature maps with occlusion and radiomic features (a) Original brain image (b) Deviation maps of left and right hemispheres with highlighted occlusion (c) Contrast maps of both sides with the degradation shown in a polygon (d) Radiomic features

5.2.6 Classification of Collaterals

Before feeding the radiomic features obtained from the difference of the hemispheres into a classifier, they were ranked using Principal Component Analysis (PCA) with 97% variance. Fig. 5.3 (d) shows the radiomic features' correlation with the first two principal components where the color bar represents the weight of correlation. For example, the GLCM features: *d_Entropy*, *d_Homogeneity* from the first principal component have the highest correlation of 0.19 with the principal component which is also visible from the color bar range. The

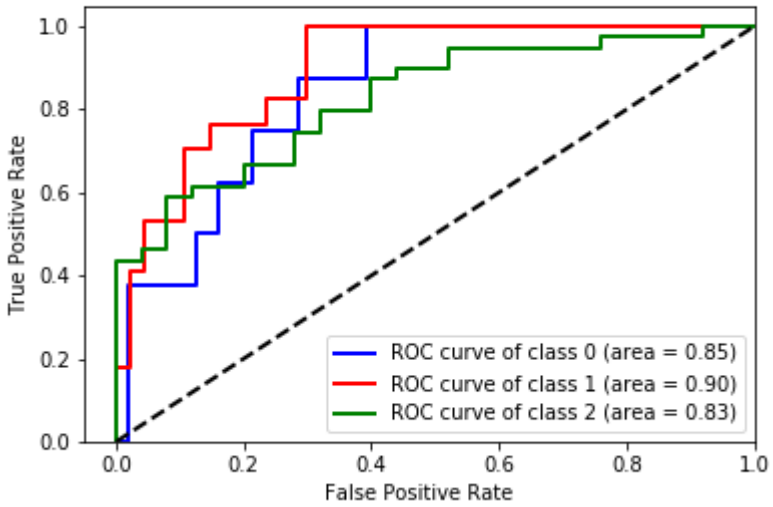

Figure 5.4: ROC curve showing classification performance of good, intermediate, and poor collaterals.

feature names are distinguished starting with 'd' for the deviation map and 'c' for the contrast map. These are then fed into the One-Vs-Rest SVM classifier with the radial basis function (RBF) kernel using balanced class weight (to penalize the majority class) to classify collaterals into good, intermediate, or poor cases.

5.3 Results

We applied k-fold (k=10) cross-validation and the performance of the developed method was evaluated using receiver operating characteristic (ROC). We obtained an overall area under the curve (AUC) of 0.86 for best sensitivity and specificity with AUC of 0.85 for poor, 0.90 for intermediate, and 0.83 for good collaterals (Fig. 5.4). An average sensitivity of 80.33% and specificity of 79.33% were obtained by taking the mean true positive and false positive rates of the three classes. The feature ranking and classification were performed using scikit-learn [106].

Further, we compared SVM performance with the Random Forest (RF). However, due to our small sample size and imbalanced dataset, RF performed extensive model selection which led to over-fitting. Thus although we had a high training accuracy (98%), the testing accuracy was 55% in the case of RF.

All the experiments were performed on a Windows 7 machine equipped with an Intel(R) Core(TM) i7-4770 CPU @ 3.40GHz and 28 GB of RAM. The robust and automatic approach grades collaterals quickly (i.e. in approximately 10 minutes with 3 minutes for the registration step, less than 6 minutes for feature extraction and 1 minute to classify a single patient).

5.4 Discussion

In the current study, we developed an automatic approach for collateral grading based on ASPECTS terminology using NCCT. This is a novel modality for assessing collaterals which are an independent predictor of good clinical outcomes. Most of the studies use dynamic CTA to assess the collateral grading whereas we proposed classifying collaterals based on ischemic damage from NCCT which requires less time, no bolus, and less radiation. To analyze the relationship between ischemic damage and collateral status, the agreement between e-CTA, which assesses collaterals with machine learning using CTA-CS [31] and e-ASPECTS from e-STROKE SUITE (Brainomix Ltd.,) is shown in the study of Grunwald *et al.* [25]. Our method uses 4D CTA as ground truth to obtain the collateral status from multiple phases and assesses it automatically from single-phase NCCT which resembles radiologists' scores and methodology.

The proposed method automatically scores collaterals using SVM based classifier, which is a popular classification method used in many other ischemic stroke analysis studies and more insights can be found in the study of Kamal *et al.* [107]. Since each subject's bilateral sides resemble each other, the collateral extent can be identified from the non-symmetry of both sides. Following this idea, the difference of radiomic features between each side helps to identify ischemic damages automatically without manual intervention, indicating insufficient collateral regions in the MCA territory. The atlas-based ASPECTS regions mapping to individual patients helps to improve the performance of the classifier and validate our method based on the popular ASPECTS scoring terminology.

A limitation of this study is the small dataset, which is the main challenge of training a classifier. To have a reasonable ratio of train test data, we used 10-fold cross-validation. Since we have only 8 poor cases, this splitting does not confirm a poor case to test in each fold. In future work, we will validate our method's performance with more data before applying it to clinical trials.

5.5 Conclusion

In conclusion, we have implemented a machine learning-based collateral grading method using NCCT that may replace high-contrast and radiation-based CTA. By using 4D CTA-based collateral scoring as ground truth, this novel approach can evaluate collaterals from the tissue degeneration extracted by radiomic features in NCCT. Our results show the feasibility of using NCCT to help physicians identify suitable patients for revascularization in AIS.

Chapter 6

SCANED: Siamese Collateral Assessment Network for Evaluation of Collaterals from Ischemic Damage

Preface

This chapter is based on our submitted paper [108].

This chapter presents the use of Siamese networks in automatic collateral evaluation for ischemic damage cases, particularly when dealing with small and imbalanced datasets. Our method, Siamese Collateral Assessment Network (SCANED), is engineered to automatically assess collaterals in non-contrast computed tomography (NCCT) images, frequently used in emergencies. What sets SCANED apart is its utilization of Siamese networks, which excel in computing similarities and are exceptionally effective at managing imbalanced data. By harnessing the Siamese network's inherent ability to calculate similarity, SCANED extracts distinctive features from affected and healthy regions of the NCCT brain scans using the 3D ResNet network. It then employs Euclidean distance as the similarity metric to pinpoint areas of ischemic damage. Rather than relying on the conventional 0.5 thresholds often associated with Sigmoid-based methods, SCANED determines the optimal threshold through ROC analysis. Collaterals are subsequently categorized as either good/intermediate or poor based on the dissimilarity observed between the left and right hemispheres, reflecting the impact of ischemic damage on the affected side. In contrast to our previous study 5 collaterals from ischemic damage are identified in NCCT using a small imbalanced dataset without hand-engineering of radiomic features. Instead, we employ a deep learning-based Siamese network approach. This work is under review on Computerized Medical Imaging and Graphics.

Abstract

This study performs collateral evaluation automatically from ischemic damage using the deep learning-based Siamese network, mitigating the issues of a small imbalanced dataset. The collateral network provides an alternative pathway for oxygen and nutrient supply in ischemic stroke cases, influencing treatment decisions. The ongoing vibrant area of research focuses on automated collateral assessment using deep learning (DL) methods to expedite decision-making processes and enhance accuracy. Our study employed a 3D ResNet-based Siamese network, referred to as SCANED, to classify collaterals as good/intermediate or poor. Utilizing non-contrast computed tomography (NCCT), the network automates collateral identification and assessment by analyzing tissue degeneration around the ischemic site. Relevant features from the left/right hemispheres were extracted and Euclidean Distance (ED) was employed for similarity measurement. Finally, dichotomized classification of good/intermediate or poor collateral is performed by SCANED using an optimal threshold derived from ROC analysis. SCANED provides a sensitivity of 0.88, a specificity of 0.63, and a weighted F1 score of 0.86 in the dichotomized classification. The GitHub code will be made available.

6.1 Introduction

Collateral circulation networks are alternative paths for blood flow when main arteries become blocked due to ischemic stroke (caused by a clot). The network comprises small blood vessels that become active when the major vessels constrict and blood flow is reduced. Collateral circulation plays a crucial role in predicting the clinical outcome of ischemic stroke since good collaterals are a positive factor in providing endovascular thrombectomy treatment. This is one of the best treatment strategies for ischemic stroke patients, which cannot apply to all patients, particularly those with poor collaterals. Poor collaterals can lead to hemorrhagic stroke caused by excess bleeding to the brain during endovascular treatment. Therefore, determining the extent of collaterals is an obvious step to guide radiologists in the further treatment of ischemic stroke patients.

Collateral assessment by radiologists, through visual inspection, suffers from inter and intra-rater variability due to a myriad of factors, including the intricate nature of manually assessing 3D images [60, 25]. Thus, as well as being time-consuming, manual evaluation can lead to wrong treatment decisions. On the other hand, automatic evaluation of collaterals by computer-aided techniques can give a quick decision and be more accessible in acute clinical stroke. Therefore, developing robust automated scoring methods of collaterals is active research.

Deep learning (DL) is increasingly being studied in the medical image domain due to its capabilities of handling complex data, automatic and sophisticated feature extraction steps, and more accurate and consistent measurement criteria. DL approaches have been proposed for evaluating collaterals [79, 49, 50, 51], identifying ischemic damage [55, 56], and detecting stroke [109, 110]. The issues that arise with deep learning are when available training data are scarce. There are very limited open and large labeled datasets in the case of stroke data.

The Siamese network, introduced by Bromley *et al.* [59], is a potential alternative as it needs a minimal training set. This network has shown the outstanding performance of an AUC of 0.91 in detecting stroke in the study by Barman *et al.* [110] using the inception

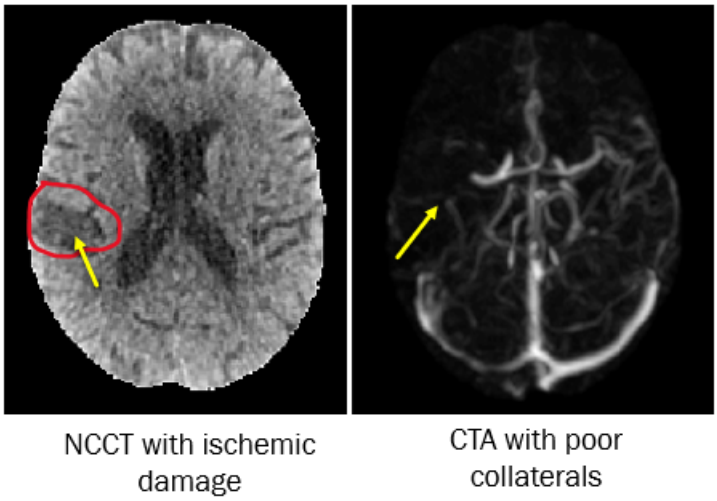

Figure 6.1: An example brain with ischemic damage shown in NCCT (left) and the poor collaterals caused by tissue degeneration (right)

module as the twin network to distinguish between left/right hemispheres. Stroke detection is also performed with a Siamese network considering the difference of feature vectors of the left/right hemispheres in the study by Vieira *et al.* [109] which achieved an F1 score of 0.72. Furthermore, Siamese networks have shown robustness in handling imbalance data issues by computing mRS scores (An accuracy of 0.67 for minority class and 0.61 for majority class) in stroke by Osama *et al.* [90] with a 6% higher AUC for the minority class over the majority class. Based on the concept of the Siamese network calculating the similarity with good performance and the robust power of handling imbalanced issues, we incorporated the network into our proposed method for the automatic evaluation of collaterals. To the best of our knowledge, this is the first use of Siamese networks for collateral circulation evaluation.

The developed method, Siamese Collateral Assessment Network (SCANED), aims to score collaterals automatically from non-contrast computed tomography (NCCT) images, which are most commonly used in an emergency setting (Figure 6.1). The distinguishing features from the affected and good sides of the NCCT brain are extracted using the 3D ResNet network. The main goal is to identify changes from NCCT due to ischemic damage, resulting in tissue degeneration and further lack of collaterals. In summary, the method performs the following steps: (1) extracts significant features with twin 3D Resnet Siamese networks, (2) uses Euclidean distance as the similarity metric to identify ischemic damage, (3) determines the optimal threshold by ROC analysis rather than 0.5 as a traditional threshold based on Sigmoid, and (4) classify collaterals into good/intermediate and poor categories based on the difference between left/right hemisphere caused by the ischemic damage in the affected side. The robustness of the proposed approach over our previous study [92] is to score collaterals from ischemic damage identified in NCCT using a small imbalanced dataset without hand-engineering of radiomic features. Instead, it employs a deep learning-based Siamese network approach.

6.1.1 Related Work

Collateral assessment is essential in determining the optimal treatment approach for ischemic stroke. As manual evaluation suffers from inter and intra-rater variabilities [60, 25], current studies focus on computer-aided systems to assess the extent of collaterals. In Table 6.1, we summarize the most related works to collateral circulation scoring, which we describe in more detail below.

Table 6.1: Recent automatic collateral scoring techniques

Ref	Data	Prediction criteria	Cases	Method	Results
[48]	3D MR Perfusion	Direct and cascaded 3-class collateral scoring done by trained neuroradiologists	183	ROI extraction followed by feature extraction with denoising, autoencoder and local image descriptor and classification with CNN, SVM, RF and KNN	Best performance with CNN+MLP with an overall accuracy of 0.72
[51]	Cone beam CT (CBCT)	Quantification of collateral scoring by radiologists into 3-classes	30 (4368 CBCT images)	2D ResNet-50 as a classifier to evaluate good and poor collaterals	Average sensitivity and specificity of 0.79 and 0.96 obtained are obtained
[50]	Peak arterial CT perfusion volumes	ESCAPE trial and Tan score [31]	200	A CNN model to evaluate collaterals automatically using axial and anteroposterior views into dichotomized as well as multiclass	Average sensitivity of 0.88 and 0.80 for dichotomized and multi-class collaterals respectively
[52]	Multiphase CTA	Positive (good + intermediate) and negative (poor) collaterals	82	CNN model	Accuracy of 0.75 in the validation group with an AUC of 0.70
[25]	CTA	Tan System [31]	98	Automated e-CTA score	90% agreement with radiologist, intra-class correlation coefficient of 0.93

[47]	CTA	4-grade Tan [31]	269	Deep-learning-based vessel extraction followed by automatic collateral scoring with best results obtained by the random forest classifier	Overall accuracy 0.80 & dichotomized collateral accuracy 0.90
[60]	4D CTA	ASPECTS	54	Quantification of filled versus unfilled vessels	An average AUC of 85.39%
[92]	NCCT	ASPECTS	64	SVM with radiomic features of left/right to identify ischemic damage	Overall AUC of 0.86 with an average sensitivity of 80.3% and specificity of 79.3%.
[79]	4D CTA	ASPECTS	83	Pre-trained EfficientNet B0 network/transfer learning	Mean sensitivity of 0.71 in multi-class and 0.89 in dichotomized classification
[49]	Multiphase CTA	Tan score [31]	173	Feature fusion from four phases with hybrid attention mechanism	Accuracy of 90.43%

The study by Frolich *et al.* [45] proposed a semi-quantitative approach for collateral assessment with 4D CTA proving tMIPs as the best indicator of regional leptomeningeal collateral rLMC scoring [32]. The method obtained an inter-rater agreement with an intra-class correlation coefficient of 0.78, limited to only good collaterals. Using 4D CTA and considering the difference of intensities between left and right hemispheres, Kersten *et al.* [42] found a correlation ratio of $r^2 = 0.71$ to radiologists' scoring. A low-rank and sparse decomposition method that quantified the ratio of filled vs. unfilled collaterals from left and right hemispheres resulted in an average AUC of 85.39% for three classes (good, intermediate, poor) [60].

Due to the advantages of machine learning over traditional methods, several studies focused on the automatic evaluation of collaterals with machine learning methods. Xiao *et al.* [5] developed an automatic method for scoring collaterals in ischemic stroke by applying support vector machines (SVM) to features extracted from 4D CTA images by PCA. The study was performed with 37 good, intermediate, and poor collateral patients' data with an accuracy of 82.2%. The e-CTA software tool is based on feature-based machine learning, validated by the study of Grunwald *et al.* [25]. Su *et al.* [47] evaluated collaterals using 4-grade Tan score [31] with an overall accuracy of 0.80 and dichotomized collateral accuracy

of 0.90. The proposed collateral assessment methods helped achieve generalizability to new data and good performance. However, one potential drawback of machine learning is the extensive hand engineering required, which can be a time-consuming and challenging process requiring significant domain expertise.

DL has become a popular and powerful tool for solving complex problems in medical imaging. Unlike traditional machine learning methods, DL models can automatically learn and extract relevant features from data, reducing the need for manual feature engineering. Therefore, current studies (e.g., [79, 49, 50, 51, 48, 52]) focus on DL-based methods to evaluate collaterals automatically. The proposed approach by Tetteh *et al.* [48] used a direct and cascaded 3-class collateral scoring method on 3D MR perfusion images. The study performed ROI extraction, feature extraction, and classification with traditional (SVM, RF, and KNN) and DL methods (CNN+MLP), showing a better overall accuracy of 0.72 with CNN+MLP. A CNN model is implemented in the study of Huang *et al.* [52] that scores good+intermediate and poor collaterals with an accuracy of 0.75. A computer-aided diagnostic technique using CNN named CCA4CTA based on a feature fusion network with a hybrid attention mechanism was proposed in the study by Tan *et al.* [49] for the automatic scoring of collateral circulation via multiphase cranial CTA. This study achieved 90.43% accuracy with a sample set of 173 subjects. A 2D CNN model and a 2D ResNet-50 are used by Rava *et al.* [50] and Ali *et al.* [51] respectively to evaluate collaterals automatically, which achieves promising results while using a large number of train dataset.

In the greater context of stroke, DL models have been used in automating ASPECTS score based on ischemic damage on NCCT [53, 54]. Chiang *et al.* [55] developed an architecture consisting of a 2D slice encoder, a slice feature aggregator, and a fully connected classifier for ASPECTS scoring with an accuracy of 80%. Cao *et al.* [56] performed ASPECTS scoring by developing a deep asymmetry network that compares the left/right hemispheres to identify ischemic status in NCCT. NCCT is more commonly used in these studies, being widely used in emergent settings and, thus, more accessible. Therefore, in our previous study [92], we developed a novel method of collateral evaluation from ischemic damage identified from NCCT, which causes tissue degeneration and, as a result, lack of collaterals in the affected side. The study obtained an overall AUC of 0.86 in classifying good, intermediate, and poor collaterals. Recently, we performed a DL-based collateral evaluation [79] in 4D CTA using the EfficientNet B0 network for transferring knowledge from the ImageNet dataset to our small dataset of 83 patients. This method used focal loss to overcome the issue of imbalanced datasets and achieved an overall sensitivity of 0.89 in the dichotomized classification of collaterals. As the Siamese network is designed to handle imbalanced data by identifying similar samples from the majority (good/intermediate) and minority (poor) classes, we evaluated a Siamese network architecture for collateral circulation scoring in this work.

6.2 Methods

The overview of the SCANED method, which was developed for automatic collateral scoring in the context of an absence of large open ischemic stroke datasets, is shown in Figure 6.2.

6.2.1 Scanning Protocol

We have used 4D CTA data as our ground truth to evaluate the collateral circulation. The imaging was performed at the Montreal Neurological Institute and Hospital with a Toshiba

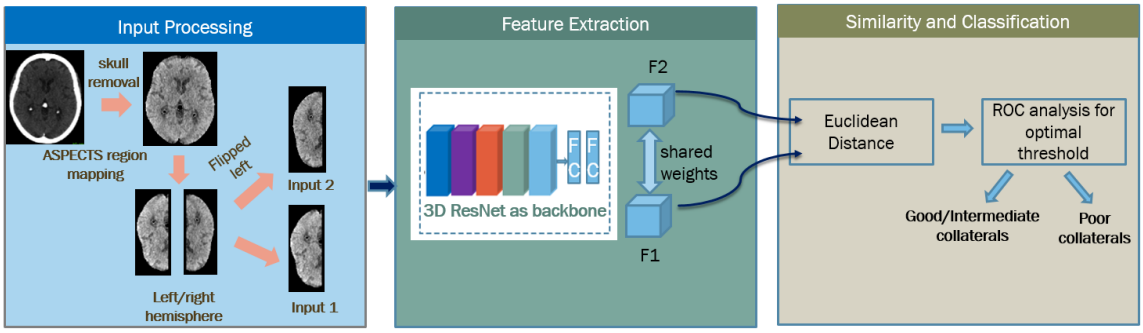

Figure 6.2: An overview of SCANED.

Aquilion ONE 320-row detector 640-slice cone beam CT (Toshiba medical systems, Tokyo, Japan). The 4D CTA images captured the dynamic vascular flow over 19 volumes, where the first volume is before applying contrast. Further, three-volume groups were captured to have early and peak arterial and venous phases. In each volume, axial CT scan images were acquired with imaging parameters of 0.5mm section thickness, 512x512 matrix, 240 mm axial scan field of view, and 160 mm scan length. Standard 80-kVp tube voltage was applied to all volumes. The first volume is an NCCT which is more widely available and commonly used in emergency rooms. Therefore, our study aims to evaluate the collaterals from ischemic damage identified in NCCT from the 4D CTA ground truth images.

6.2.2 Dataset Description

Our dataset comprises 64 subjects with NCCT (39 good, 17 intermediate, and 8 poor cases). Two radiologists scored them into three categories: good (100% collaterals), intermediate (above 50% to less than 100%), and poor (below 50%). In different studies [79, 51], only two groups are considered: good/intermediate and poor since the treatment benefits are either minimal or nonexistent for the latter group, as demonstrated in the MR CLEAN Trial [16]. This is because good collaterals can be a favorable factor in endovascular treatment, working as a robust blood flow network during blockage. In contrast, poor collaterals can adversely affect the treatment leading to excess bleeding. Therefore, in this study, we focused on the dichotomized classification of collaterals: good/intermediate-class 0 and poor-class 1 rather than multi-class classification similar to our previous study [79] as well as the study by Su *et al.* [47].

6.2.3 Pre-processing Steps

Before feeding the left/right hemispheres to the Siamese network, the NCCT brain went through a number of pre-processing steps: (1) skull and any calcification were removed from each brain using the FSL tool following the study of Muschelli *et al.* [103], (2) 10-ASPECTS regions were mapped to each brain which allows better assessment of ischemic damage in the ASPECTS regions as in [92], (3) left/right hemispheres were separated using the ANTs tool, and (4) the left side was flipped to overcome any left/right biases so that the model learns similar features from each side and accounts for any asymmetry between them.

6.2.4 Pairs Generation for Siamese Network

We must provide pairs of input to the twin network, implying the need for two input samples sharing identical weights and parameters. To follow this criterion, we made pairs of input as positive and negative samples using each brain's left and right hemispheres. Let us consider the left side, L, and the right side, R are denoted as $L_{affected}$, $R_{affected}$, L_{good}, and R_{good} for the affected stroke sides and the normal unaffected sides, respectively. In general, the label can be considered zero if both the inputs are from the $L_{affected}$ and $R_{affected}$ or L_{good}, and R_{good}, on the other hand, the label should be one if the input combinations are either $L_{affected}$, and R_{good} or $R_{affected}$ and L_{good}. Since the brains are categorized into three classes and the affected side is unknown from the ground truth, it is not feasible to create arbitrary pairs using the left/right side terminology described above. Therefore, pairs are considered in the case of dichotomized classification as either positive or negative pairs. Specifically, **similar/positive pairs** are those where the left and right hemispheres of the brains have more than 50% of collaterals (good and intermediate). These are denoted with zero and named class 0 since these brains should have a higher similarity (with more collaterals) between both sides. In contrast, **dissimilar/negative pairs** are those where the collaterals (poor) are below 50% denoted with one and named class 1 with higher dissimilarity between good and affected sides.

6.2.5 Feature Extraction

The input pair was further fed into the twin network, where 3D ResNet was used as a backbone for feature extraction. Rather than starting with random weights, the proposed approach used the pre-trained weights [87] of ResNet-50 from 23 medical domain datasets following the study of Chen *et al.* [111]. Since the pre-trained model [111] is for segmentation purposes, we replaced the last decoder layer with 2-fully connected layers in our study. The first layer comprises 512 filters, followed by a batch norm and a ReLU; the last layer has 64 filters. Two feature vectors, F1 and F2, are obtained from the network from input1 and input2, respectively, which were further used for similarity and classification purposes.

6.2.6 Similarity and Collateral Scoring

The Euclidean distance (ED) was used as a similarity metric between F1 and F2 to output a distance value. The higher the distance, the less similar the two sides of the brain, suggesting poor collaterals in dichotomized classification and vice versa. This distance value played a role in the training loss function to shape the optimal model. During testing, each brain consists of multiple overlapping patches with a sliding window of stride size half to the patch size to cover the whole brain. Since different regions between the left/right sides, especially the areas surrounding the boundaries of the brains could be similar even in the case of poor collaterals, the distance values from all patches are not significant. Therefore, we chose the top three distance values, computed their median, and adopted it as the definitive distance measure. Further, median distance values from the validation set were employed to establish a threshold using Youden's index. This index is based on the Receiver Operating Characteristics (ROC) curve, which aims to identify the optimal cutoff point by balancing the best sensitivity and specificity.

6.2.7 Loss Function

A contrastive loss function [112] was used in this study rather than the traditional binary cross-entropy to classify two classes. The Siamese network is a metric learning problem mainly focused on identifying similarities between pairs of inputs rather than classifying them. The contrastive loss function is designed as Equation (10) [112] in such a way that it encourages the model to learn to maximize the distance between dissimilar samples and vice versa. Also, it is more robust to the imbalances between majority and minority classes, where the binary cross-entropy can be sensitive. The contrastive loss $\mathcal{L}$ is defined as:

$$\mathcal{L} = (1 - label) \cdot \frac{1}{2} D^2 + label \cdot \frac{1}{2} \max\{0, margin - D\}^2 \tag{9}$$

where *label* stands for either 0 (similar) or 1 (dissimilar) input images. D is the distance metric that we named ED before and $margin = 1.0$ which is the maximum margin considered in our case. The equation demonstrates that if the collaterals are from class 0 (good/intermediate), the D between them will be minimized, whereas if the left/right inputs are dissimilar, meaning poor collaterals, the right side of the loss function will be penalized trying to keep the maximum distance within the margin. If the D is more than the margin, then the loss would be zero, and it is not necessary to try to move dissimilar points farther than the maximum margin.

6.2.8 Training

The NCCT data were downsampled to 221x221x160 voxels, and Gaussian filtering was applied to retain the most significant information. Subsequently, the left/right hemispheres were separated, and zeros were eliminated from each side, resulting in images with dimensions of 70x170x140 voxels. We utilized 3D patches of size 64x64x64 voxels, following a similar approach to the study by Roy *et al.* [113], where patch extraction improved results despite using a small dataset. Further, patches help to increase the models' generalization capability by learning more robust and variable features. Since the 3D model is computationally expensive, extracting small patches rather than the whole image leads to faster training and lower memory requirements. Further, to boost our imbalanced dataset, only augmentation was used in poor collaterals. Random rotation, flipping, zooming, and Gaussian noise were applied during training to the poor collateral brains to increase the model's generalizability. The patch extraction and data augmentation was performed using the MONAI open-source framework for deep learning [114]. In every epoch, one patch is considered for training.

Our evaluation employed 5-fold cross-validation, maintaining an identical data distribution for each class, consisting of a 70% training set, 10% validation set, and 20% test set. A testing set was initially isolated within each fold, followed by a train-validation split. The best model obtained from the validation process was then utilized to assess the test set. The final results were computed as an average over the 5-fold process. For the creation of the F1-64 and F2-64 feature vectors from 3D ResNet through fine-tuning, a selection of 64 filters was made. We employed the Adam optimizer with an exponential learning rate scheduler initialized at 0.001. All experiments commenced with an initial epoch count of 300 while continuously monitoring the best validation loss to obtain the optimal model from each fold.

6.3 Experimental Analysis and Results

To assess collaterals using SCANED, we established various experimental setups to evaluate the performance of dichotomized classification. As the patch size is a hyperparameter, we trained the network using patch sizes 16, 32, 45, and 64 to observe the results of 5-fold cross-validation. Our findings indicate that the model achieved the highest performance when trained with a patch size of 64 (refer to Table 6.2). The table shows that increasing the patch size improves performance for classes 1 (poor collaterals) and 0 (good collaterals). Notably, although the overall accuracy for patch size 45 is lower compared to sizes 16 and 32, it demonstrates a higher ability to classify poor collaterals. Given the significant impact of poor collaterals on treatment outcomes, our method prioritizes their accurate identification. Consequently, a patch size that can effectively differentiate and classify both classes is significant for our research objectives. The hypothesis from these results suggests that smaller patches with similarities can potentially confuse the loss function. This is because when ischemic damage occurs, the resulting degenerated tissue may be localized to a small region within the brain. However, since we have knowledge of the ground truth label for the entire 3D brain, other areas that exhibit similarities may still be present, even in cases with poor collateral circulation. Further, with the best-performing models using a

Table 6.2: Analyzing patch size in model training

Patch-size	**Accuracy**	**Sensitivity of Class 0**	**Sensitivity of Class 1**
16	0.78	0.77	0.13
32	0.64	0.70	0.25
45	0.61	0.63	0.50
64	0.84	0.88	0.63

patch size 64, we tested each brain by dividing it into 12 overlapping patches, encompassing the entire area. The final distance value for each brain obtained as described in Section 6.2.6 is compared to the threshold to decide the final class of the test brain.

Table 6.3 shows the confusion matrix for dichotomized classification of collaterals based on the average of five-fold cross-validation. From the matrix, the overall accuracy is 0.84. But for the imbalanced dataset, accuracy is not the best metric to consider since it can mislead the interpretation of the result. Therefore, the precision, recall (sensitivity), specificity, and weighted F1 score are reported considering the good/intermediate collaterals as the true positive class of 0.94, 0.88, 0.63, and 0.86 respectively.

Furthermore, to compare the performance, we conducted a patch-based evaluation by directly measuring the distance between two feature vectors, F1 and F2, instead of using ED. This approach involved using a fully connected layer to determine the output class rather than employing ROC analysis. However, our proposed method yielded higher performance compared to this alternative approach.

To assess the performance of the backbone network, we conducted experiments using ResNet-101 to analyze its impact. While ResNet-101 demonstrated improved performance in classifying good/intermediate collaterals (sensitivity: 0.89) due to its ability to learn from more layers, it struggled with accurately classifying poor collaterals (specificity: 0.38). This difficulty arose from the need for more samples for poor collaterals, resulting in overfitting.

Due to the scarcity of literature on collateral evaluation using NCCT, direct comparisons

Table 6.3: Confusion matrix for dichotomized collateral scoring

True Class	Predicted Class		Total
	Class 0	Class 1	
Class 0	49	7	56
Class 1	3	5	8
Total	52	12	64

with related methods are beyond our current capabilities. Therefore, we compared our method with a recent approach proposed by Ali *et al.* [51]. Their method involved classifying collaterals into good and poor categories using 2D cone beam CT (CBCT) images and ResNet-50. They conducted their study on 4368 CBCT images from 30 ischemic stroke patients, achieving an overall accuracy of 76.79%, a sensitivity of 79.33%, and a specificity of 96.02%.

To apply the 2D ResNet-50 following their methodology, we considered 2D slices from each brain as separate images, excluding the first and last ten slices that contained no relevant information. By employing the same data augmentation and cross-validation setting, we obtained an overall accuracy of 0.76 and an average sensitivity of 0.58 for our experiment, with mostly good/intermediate collaterals correctly classified. This discrepancy can be attributed to the imbalanced dataset, with insufficient samples representing poor collaterals compared to the good/intermediate ones. Based on the findings from our experiments with both 3D ResNet and 2D ResNet, as presented in the study by Ali *et al.* [51], it can be concluded that while ResNet performs well as a classifier for balanced datasets, for imbalanced data, a Siamese network with ResNet as a feature extractor proves to be more effective. The addition of sequential features by the use of 3D rather than 2D further enhances the classification process. The results for different experiments are shown in Table 6.4.

Table 6.4: Performance comparison of different experimental settings

Methods	**Sensitivity**	**Specificity**	**Weighted F1 score**
SCANED (backbone: ResNet-50)	0.88	**0.63**	**0.86**
SCANED (backbone: ResNet-101)	**0.89**	0.38	0.82
SCANED (direct distance)	0.77	0.5	0.77
2D ResNet-50 proposed by Ali *et al.* [51]	0.58	0.25	0.71

Table 6.4 shows that SCANED achieves the best weighted F1 score. This demonstrates the effectiveness of our approach, particularly in addressing the challenges posed by

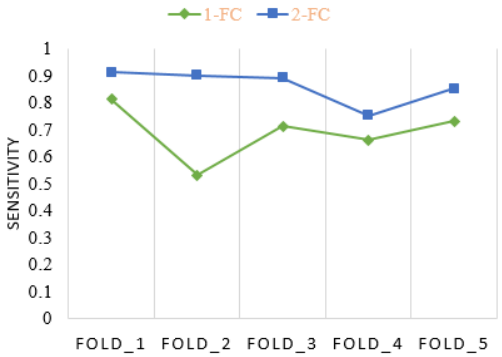

Figure 6.3: 1-fully connected layer(1-FC) vs. 2-fully connected layers(2-FC)

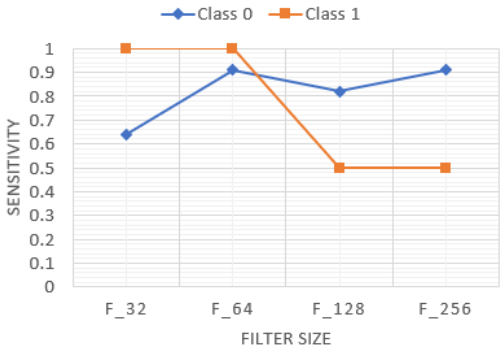

Figure 6.4: Determination of best filter size

imbalanced and small datasets. Additionally, our method considers the importance of both good/intermediate and poor collateral evaluations, which is important for making treatment decisions related to ischemic stroke.

Moreover, by considering 3D brain analysis instead of 2D analysis, our method leverages the advantages of capturing sequential information between slices. In contrast, 2D analysis treats each slice as a separate image, disregarding the valuable contextual information present in the sequential nature of brain scans.

6.3.1 Ablation Study

Before the final evaluation step, an ablation study defines the optimal model for further validation. Two fully connected layers are considered in the final model, which is decided based on an experiment on a subset of the dataset. After the backbone output of 2048 filters, the first fully connected layer is with 512 filters, followed by BatchNorm and ReLU layers. The second and final top layer consists of 64 filters. Through two separate experiments with only 1-fully connected layer and 2-fully connected layers, respectively, an improvement of 15% from the average sensitivity of 0.71 to 0.86 is obtained in the case of good collaterals with more true positive results (Fig. 6.3).

We found that a 2-fully connected layer added to the backbone provides the optimal overall performance, although the performance is the same for the poor collaterals. The hypothesis here is that the 2-fully connected layers with BatchNorm and ReLU layers analyzed features in a more sophisticated way and helped to classify more good collateral cases while poor cases' performance was unaffected.

To fine-tune the model, the top layer of the network was tweaked in a subset of the dataset with different filter sizes: 32, 64, 128, and 256 (Fig. 6.4), which demonstrates that with higher values of filter, the performance is increased for the majority Class 0 but deteriorated for the minority Class 1 whereas a reasonable performance for both classes is obtained with the filter size of 64. Therefore, to balance the performance for both classes and consider the significance of distinguishing the poor cases in the treatment of ischemic stroke, a filter size of 64 is chosen. The hypothesis is that for the minority class samples, a larger filter size can lead to overfitting due to a lack of samples to learn, which matches our previous study [79].

6.4 Discussion

We propose SCANED for collateral evaluation using NCCT. The approach shows robustness in handling the scarcity of stroke data and imbalanced class issues. Using 3D patch-based classification helps to extract significant sequential features considering different regions of the brain, which can be missed using 2D patches.

This study builds upon our previous work [92] by incorporating DL-based automatic extraction instead of traditional machine learning-based radiomic feature extraction tasks. DL-based methods need many training samples, which are rare in the medical domain, especially in the case of stroke data. The Siamese network can handle limited and imbalanced data since it focuses on the distance between two inputs. Therefore, the network can identify similarities using a few samples from each class. Since our collateral classification is based on the left/right hemispheres' similarity, the Siamese network is a natural fit. Further, it has a higher generalization capability and is less sensitive to variations of input samples, and as such, it helped mitigate the imbalanced issue that can arise easily from the traditional DL methods. Lastly, Siamese networks are more computationally efficient than traditional DL methods as they share the same weights and parameters for both inputs.

Compared to our previous study [92] for collateral evaluation from ischemic damage and the same dataset, the current study has shown efficacy over it. While our previous study [92] focused on radiomic features GLCM, GLSZM, and global features and classification with SVM, our recent study expands upon this by considering ResNet-50 as a backbone for automatic feature extraction rather than hand engineering of features. Furthermore, the idea of a Siamese network gives more robustness than the previous study [92] since it solves the imbalance issue. Although the same number of data, 64 and poor: 8, are used in this study, the previous study used a 10-fold cross-validation, which did not confirm a poor case to test in each fold. On the other hand, the recent study considered 5-fold cross-validation, which ensures at least one test case in each fold. Also, the dataset is split into training, validation, and test sets in the current work with only four poor cases, even in the case of the training set in a fold. This proves the Siamese network could distinguish between the majority and minority class samples based on similarity metrics with very few samples. These confirm that the current work is more robust than the previous study [92].

In general, if we consider the performance metric for different recent studies (for example, the study by Tetteh et al. [48], Huang *et al.* [52] and Tan *et al.* [49]) for automatic evaluation of collaterals, our method has a better performance. Although the study by Tetteh *et al.* has obtained a reasonable accuracy of 0.81 in cascaded two-class classification, they used 183 patients for training, whereas our study used 64 patients only and also obtained a higher sensitivity. If we consider the study by Huang *et al.* [52] in the CNN model's positive and negative collateral evaluation, their accuracy is 0.75, which needs further experiments and validation to obtain higher performance. Our previous study [79] for collateral evaluation with 4D CTA as ground truth obtained a promising performance of sensitivity=0.89 in dichotomized classification using transfer learning. The study addressed the imbalanced issue by employing focal loss. However, it should be noted that this method may exhibit sensitivity in different domains due to the utilization of transfer learning with pre-trained weights from non-matching domains. In contrast to the previous study [79], our current research leverages pre-trained weights specifically from the medical domain. Additionally, it incorporates a Siamese network, demonstrating superior generalizability compared to traditional transfer learning methods. Moreover, the use of focused 3D patches facilitates

the extraction of finer details and sequential 3D features from each brain. While our study achieved a performance 6% lower than the collateral evaluation conducted by Tan *et al.* [49], wherein they presented a promising method involving feature fusion and achieved an accuracy of 90.43%, our approach demonstrated a closer performance even with a smaller dataset of only 64 patients, compared to their usage of 173 patients' data.

In our current work, we acknowledge certain limitations. While testing new cases, we utilized multiple crops (overlapping patches covering the entire brain). However, during training, we relied on a single crop, using one patch with a size of 64 in each epoch. Incorporating multi-crops directly into training poses challenges due to tissue degeneration from ischemic stroke, which may only appear in specifically larger regions of the brain rather than affecting the entire brain. This situation can lead to confusion in the loss function when multiple overlapping/non-overlapping patches are assigned the same label as the whole brain. To address this issue in future research, we propose following the approach presented by Gessert *et al.* [115]. Specifically, we suggest integrating a patch-based attention mechanism to capture the global context between patches. Additionally, applying patch regularization, as seen in that study [115] on skin lesion classification, could prove beneficial.

SCANED can be further validated to another dataset, if available. Another interesting research could be considering each ASPECTS region to obtain similarity between left/right hemispheres and calculate the final score based on all regions as it is done for ASPECTS scoring.

6.5 Conclusion

In summary, SCANED is a promising approach for the assessment of collaterals, which is less prone to human errors. Siamese networks with a deep model as the backbone to perform feature extraction help the approach to be more robust and faster and have the potential for ischemic stroke diagnosis while handling the scarcity of data and acquisition cost. SCANED can evaluate collaterals from ischemic damage identified from NCCT imaging, which is lower cost and more commonly available than CTAs. With further validation of patients, the study can significantly impact the treatment outcomes of ischemic stroke patients.

Chapter 7

VesselShot: Few-shot learning for cerebral blood vessel segmentation

Preface

This chapter is based on our accepted conference paper [116].

While the previous chapters (3, 4, 5, and 6) introduced methods for automatic collateral scoring, this chapter presents a novel approach for segmenting 3D cerebral blood vessels, even in the absence of large annotated training data.

Cerebral blood vessel segmentation serves as a crucial prerequisite for the assessment of collateral circulation in ischemic stroke, as discussed in 3. To effectively extract blood vessels with contrast flow and distinguish them from other brain components, such as white and grey matter and calcifications, it is important to have reliable segmentation methods. However, the performance of this segmentation can be significantly hindered in the absence of ground truth data. Creating manually annotated 3D brain imaging data is labor-intensive, requiring expert annotation and substantial time investment. In light of these challenges, this chapter endeavors to tackle the blood vessel segmentation task using a limited set of annotated data to enhance the performance in computer-aided collateral evaluation. The proposed method, named "VesselShot" utilizes a publicly available dataset with a pre-existing ground truth of segmented cerebrovasculature. This work was presented in Machine Learning in Clinical Neuroimaging (MLCN) workshop in MICCAI, 2023.

Abstract

Angiography is widely used to detect, diagnose, and treat cerebrovascular diseases. While numerous techniques have been proposed to segment the vascular network from different imaging modalities, deep learning (DL) has emerged as a promising approach. However, existing DL methods often depend on proprietary datasets and extensive manual annotation. Moreover, the availability of pre-trained networks specifically for medical domains and 3D volumes is limited. To overcome these challenges, we propose a few-shot learning approach called VesselShot for cerebrovascular segmentation. VesselShot leverages knowledge from a few annotated support images and mitigates the scarcity of labeled data and the need for extensive annotation in cerebral blood vessel segmentation. We evaluated the performance of VesselShot using the publicly available TubeTK dataset for the segmentation task, achieving a mean Dice coefficient (DC) of 0.62±0.03.

7.1 Introduction

Cerebral blood vessel segmentation plays a vital role in various applications, such as diagnosing cerebrovascular diseases (e.g., stroke), planning surgical interventions for conditions such as aneurysms and arteriovenous malformations, studying brain functions, and assessing the impact of new treatments for cerebrovascular disorders. However, publicly labeled data in this domain is limited, hindering the progress of deep learning-based research on cerebral blood vessel segmentation. Since 2017, numerous deep learning-based methods have been proposed for cerebral blood vessel segmentation. However, most of the previous work has been performed on data from private sources as there is very little publicly available labeled data [117]. Traditionally, deep learning (DL) models for semantic segmentation require a large amount of training data with manual annotation, which is time-consuming and labor-intensive, particularly in 3D, where many slices must be inspected. The need for more annotated data negatively impacts DL models' training and generalization capabilities.

To address this challenge, few-shot learning emerges as a promising alternative that reduces the need for extensive manual annotation. For semantic image segmentation, few-shot learning aims to enable DL models to learn underlying visual patterns and semantics from a limited set of labeled examples. This also allows them to generalize effectively to unseen object categories during the segmentation process. To date, few-shot segmentation has been explored in several medical imaging contexts. In the study of Roy *et al.* [118], the authors proposed a two-armed few-shot architecture to extract support and query images with squeeze-and-excitation modules for the segmentation of abdominal organs (each organ type is considered as a separate class) using 3D volumetric scans, obtaining an average Dice coefficient (DC) of 48.5% [118]. Similarly, Tang *et al.* [119] used a few-shot framework to refine the segmentation masks with a recurrent module and achieved a mean DC of 81.91%. To eliminate expert annotation for training medical image segmentation algorithms, Ouyang *et al.* [120] employed a super-pixel-based self-supervised segmentation approach with few-shot learning. Preserving the local information to alleviate the foreground vs. background imbalance issue with an adaptive local prototype pooling, their study achieved a maximum DC of 78.84%. Semi-supervised segmentation was also incorporated in a few-shot paradigm by considering a generative adversarial network (GAN) [121]. This method had comparable performance to fully supervised approaches in multi-modal 3D medical image segmentation. Few-shot learning techniques have also

shown efficacy in cardiac image sequence segmentation tasks following a multi-level semantic adaptation, with a DC of 92.43% [122]. Lastly, Xu *et al.* [123] proposed a few-shot learning method with a multi-scale class prototype and attention module for 2D retinal blood vessel segmentation.

In this chapter, we aimed to develop a few-shot learning approach called "VesselShot" for segmenting cerebral blood vessels. Building upon the PANet few-shot segmentation method introduced by Wang *et al.* [3] for natural image segmentation based on metric learning, VesselShot leverages DL models' ability to learn a consistent embedding space that minimizes the distance between support and query prototypes (see Section 7.2.2). To the best of our knowledge, our method is the first attempt to employ few-shot learning for 3D segmentation of brain vascular images. The proposed VesselShot technique aims to overcome the limitations of the existing deep-learning models for cerebral blood vessel segmentation and explore the potential of few-shot learning in this domain.

7.2 Methodology

7.2.1 Dataset and Pre-processing

We used the publicly available TubeTK dataset [1]. The dataset contains T1 and T2 acquired at 1x1x1 mm3, Magnetic Resonance Angiography (MRA) acquired at 0.5 x 0.5 x 0.8 mm3, and Diffusion Tensor Imaging (DTI) using 6 directions and a voxel size of 2x2x2 mm3 from where we used the MRA images. Among the 100 MRA of healthy subjects, a subset of 42 have manual segmentation of the intracranial vasculature. The original dimension of the images is 448x448x128 voxels at a resolution of $0.5\times0.5\times0.8$ mm^3. The images were pre-processed as follows. First, all images were down-sampled to a resolution of 1x1x1 mm^3, resulting in a dimension of 230x230x102 voxels. To allow spatial consistency, all the brains were registered to one subject's image as a template with affine transformations. Fifteen patches that contain blood vessels were randomly extracted from each brain using the technique introduced in the study of Wang *et al.* [124], with a size of 64x64x16 voxels to fit the GPU memory.

7.2.2 Problem Definition

To segment cerebral blood vessels with a small amount of annotated training data, we built upon the few-shot segmentation method proposed by Wang *et al.* [3]. In general, few-shot learning involves training and testing episodes with support and query sets, following a "C-way K-shot" paradigm. The support set comprises labeled examples that a DL model can use to learn about target classes, while the query set contains unseen test cases to be classified during inference. In C-way K-shot segmentation, we obtain K {image, mask} pairs per semantic class in the support set, with a total of C classes. The training episodes consist of $S_{i,k}$, $M_{i,k}$ and $Q_{i,k}$, denoting support, mask, and query sets, respectively with $i = 1, 2, ...c$ for c classes and $k = 1, 2, ...s$ for s samples/shots. Both the support and query sets share knowledge extracted by the DL model to perform the final segmentation. In our experiments, following the problem framing of Roy *et al.* [118], who categorized classes with the designated segmentation tasks, we primarily focused on building our algorithm based on one class (i.e., blood vessel segmentation) or a 1-way K-shot approach. Furthermore, to account for individual vascular differences between subjects, we also considered the problem

[1] https://public.kitware.com/Wiki/TubeTK/Data

framing from the work of Xu *et al.* [123], who treated each subject as a separate class with its image patches as members of the class for retinal vessel segmentation. In this case, the segmentation was extended to a C-way K-shot setting.

7.2.3 Model Design

To perform support-to-query segmentation, we built robust prototypes from the target class of the support set. We used the nn-UNet [2] architecture as a backbone network to extract deep features from support and query images. Upsampling along with masked average pooling [3] was performed to obtain the final segmentation. Figure 7.1 shows an example of a 1-way 3-shot learning paradigm for query mask generation.

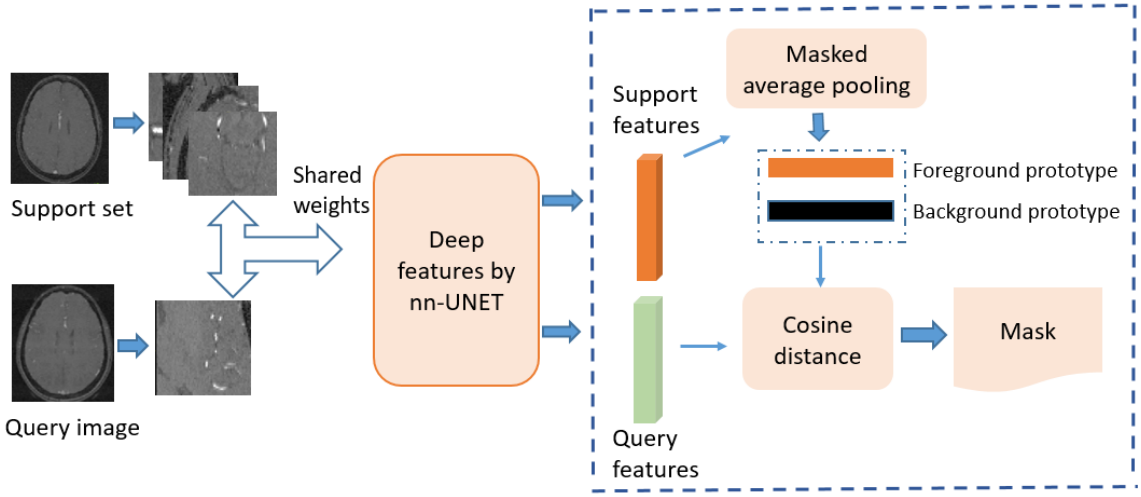

Figure 7.1: VesselShot 1-way 3-shot learning: 1 brain with 3 sample patches is considered for the support set. Knowledge is shared between support and query set by extracting deep features using nn-UNet [2] which are further embedded into foreground and background prototypes using masked average pooling [3]. Cosine similarity is used between support and query prototypes to obtain the segmented query mask.

To generate a prototype from a class, the feature map, F_I of an image, I was extracted from the support set, S and by masked average pooling (Equation 10), the obtained feature maps were compared across different class indices of I, similar to the prototype extraction approach used in PANet [3].

$$prototype_{\text{class}} = \frac{1}{N_s} \sum_{i=1}^{N_s} \frac{\sum_{x,y,z} F_{I,\text{class}}^{(x,y,z)} 1[Mask_{I,\text{class}}^{(x,y,z)} == \text{class}]}{\sum_{x,y,z} 1[Mask_{I,\text{class}}^{(x,y,z)} == \text{class}]} \tag{10}$$

where N_s is the total number of support images, and $1[\cdot]$ takes a value of 1 when the condition $[Mask_{I,class}^{(x,y,z)} == class]$ is true and 0 otherwise. The background prototype was built following the same equation with the constraint of $[Mask_{I,class} \neq class]$, which means the feature map values do not belong to the corresponding class index. For the evaluation, all brain class indices were assigned a value of 1 for foreground blood vessels (our main target) and 0 for the background. After prototype extraction, the feature map of the query image was compared with the support prototypes using cosine similarity. Each voxel at the spatial location (x, y, z) was classified based on the maximum similarity between the

query feature map and the support prototypes. Finally, a segmentation mask was predicted for the query image based on the maximum probability values obtained by Softmax that was applied to the distance map. Note that our 3D segmentation was performed similarly to the work of Roy *et al.* [118], but we experimented with the scenarios of 1-way K-shot and C-way K-shot as mentioned in Section 7.2.2. During inference time, we extracted 54 non-overlapping patches from a test MRA, where each was a query image at a given time and was paired with the support set to obtain the final segmentation. For our experiments, the support sets were created from the training data.

For training, a combination of both cross-entropy loss, CE_{loss} and Dice loss, D_{loss} was used. Our approach emphasized the Dice loss since blood vessels occupy a minimal area considering the brain space, which can significantly affect learning with a high-class imbalance. The Dice loss only focuses on the agreement of an image's predicted segmentation and ground truth label. However, it is not ideal to overlook the background entirely, as this can affect the robustness of significant features of the network [125]. Therefore, the following hybrid loss function was used to handle both class imbalance and increase the strength of features: $Loss = 0.6 * CE_{loss} + 0.7 * D_{loss}$. Note that the weights were determined empirically.

7.3 Experimental Setup

We employed the nn-UNet [2] model to extract deep features from the support and query images. Since some fine-grained image features are lost while downsampling in feature extraction [123], upsampling is an important step and pre-requisite for further background and foreground prototype extraction. Unlike Wang *et al.* [3], our preliminary testing showed that nn-UNet [2] was more effective as a decoder than trilinear upsampling for our 3D data (PANet used bilinear upsampling for 2D images [3]).

To perform training in our few-shot blood vessel segmentation, a maximum iteration of 20,000 was used while monitoring the best DC value for early stopping. A learning rate (lr) scheduler with an SGD optimizer was used at an initial lr=0.01 and momentum of=0.99. For data augmentation, random flipping, random Gaussian blurring, noise addition, and contrast changes were applied during training. To evaluate the performance of our method, different metrics, including DC, precision, sensitivity, and Intersection over Union (IoU) were computed. We performed multiple experiments to prove the efficacy of the proposed few-shot segmentation method. These experiments encompassed various settings, namely 1-way 1-shot, 1-way 4-shot, 1-way 5-shot, and 3-way 5-shot learning. In addition, we also used a fully supervised UNet with four layers of hierarchies as a baseline to assess the performance of the proposed method using the same patch size and data augmentation techniques. It was trained using the Dice loss with an Adam optimizer ($lr = 0.0001$) and a CosineAnnealingLR scheduler ($T_{max} = 5, eta_{min} = 0.000001$). To compare the different few-shot learning settings, as well as the UNet baseline, we divided the data into three subsets: a training set (78%, 33 cases), a validation set (7%, 3 cases), and a test set (15%, 6 cases). The best setting was determined from their performance based on the test set. Subsequently, we used the best setting to conduct a full 4-fold cross-validation. This way, we could obtain segmentation results for all the subjects to offer a more comprehensive evaluation.

7.4 Results

Table 7.1 presents the performance for various few-shot segmentation settings. It is important to note that the reported performance in the table was obtained from patch-based evaluations, where the averages of all classes in the test set were considered. The highest average DC of 0.67 was obtained with 1-way 1-shot learning. In terms of the mean values, 1-way 4-shot and 1-way 5-shot give similar results in all the evaluation criteria. However, the performance of the 3-way 5-shot method was notably inferior. This discrepancy may be attributed to the inclusion of three classes represented by distinct brains characterized by significant similarities. Excess prototype generation in this approach likely contributed to overfitting, resulting in the observed decline in performance. The results of the 1-way 1-shot proved that the few-shot paradigm could offer sufficient segmentation performance with even a single sample from a single class, which makes faster convergence and mitigates the issue of a small annotated dataset.

It is essential to note that while the single-split result indicated a slight advantage for the 1-way 1-shot model in the case of DC, the more comprehensive evaluation through cross-validation provided a more precise and more reliable picture of the model's performance. Therefore, considering all performance metrics, the 1-way 5-shot setting emerged as the top-performing setting among the options tested in this study. Based on the 1-way 5-shot setting, a full 4-fold cross-validation was performed. The metrics of DC, sensitivity, precision, and IoU were obtained as 0.62±0.03, 0.53±0.02, 0.72±0.02, and 0.43±0.02, respectively. For qualitative evaluation, segmentation maps of four random patches are shown in Fig. 7.2 for the 1-way 5-shot setting. Furthermore, by recombining segmented image patches from the same brain spatially, we also demonstrate a case in Fig. 7.3 for the same setting. Finally, the fully supervised UNet performed poorly with the limited annotated data and achieved a DC of 0.27±0.27. UNet typically requires a larger, well-labeled dataset to achieve reasonable performance, as demonstrated by the study of Livne *et al.* [126].

Table 7.1: Performance metrics of VesselShot for different settings with the UNet as a baseline, including DC, Sensitivity, Precision, and IoU.

Methods	**DC (SD)**	**Sensitivity(SD)**	**Precision(SD)**	**IoU(SD)**
1-way 1-shot	**0.67(0.02)**	0.50 (0.03)	0.68 (0.02)	0.40 (0.02)
1-way 4-shot	0.66 (0.02)	0.54 (0.03)	**0.68 (0.02)**	0.41 (0.02)
1-way 5-shot	0.66 (0.02)	0.58 (0.02)	**0.71 (0.03)**	**0.45 (0.02)**
3-way 5-shot	0.52 (0.04)	0.39 (0.04)	0.57 (0.08)	0.23 (0.03)
UNet	0.27 (0.27)	0.47 (0.09)	0.35 (0.06)	0.15 (0.04)

7.5 Discussion

This chapter proposed a novel few-shot learning approach for 3D cerebral blood vessel segmentation. The method achieves the segmentation by building robust prototypes with masked average pooling based on embedded features that are extracted from an nn-UNet [2]. Inspired by the PANet [3], we adopted foreground and background prototypes and used them to compare the query feature map with the support set's prototypes for blood vessel segmentation. To further enhance generalizability, Wang et al. [3] performed prototype

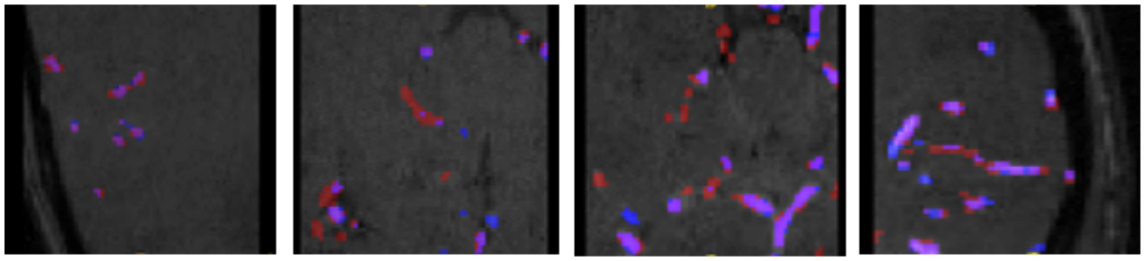

Figure 7.2: Segmentation maps of four samples, where red represents the original cerebral blood vessels, blue shows the prediction, and purple represents the overlap.

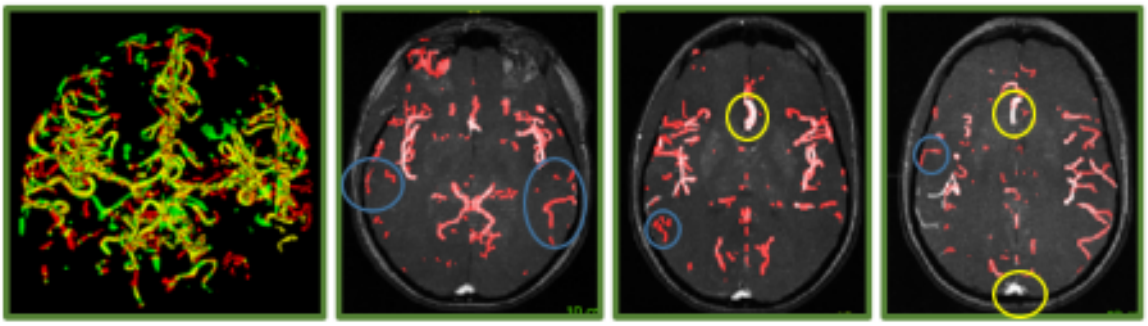

Figure 7.3: *From left to right:* 3D segmentation result with the overlap of GT and predicted labels (yellow = overlap, green = GT and red = prediction). 2D Maximum intensity projections (MIPs) of 10 slices in Slice 15-25, 25-35, and 35-45 from a total of 102 brain slices, with the overlay of the original MRA and segmentation (in red). The blue circles show the wrong prediction of large vessels and the yellow circles indicate missed blood vessels.

alignment regularization (PAR) of the predicted query mask with the support mask through additional information extraction. We have also experimented with this technique for our application, but unfortunately, it did not lead to performance gains during evaluation and resulted in slower convergence in model training.

In our approach, we considered two scenarios: (1) a single class (blood vessel segmentation) in a 1-way K-shot setting and (2) a C-way K-shot setting with each brain as a separate class. While the first case aligns with the approach of Roy *et al.* [118], the latter resembles the problem framing of Xu *et al.* [123]. In the 1-way K-shot setting, the best results came from the 1-way 5-shot setting, which was superior to treating different brains as their own classes. We hypothesize that this was due to the high structural similarity between MRAs after spatial normalization, which emphasizes the primary vasculature networks. This is in contrast to the results of Xu *et al.* [123]. Since the core task involves only two classes, blood vessels, and the background, the C-way K-shot paradigm may lead to overfitting and compromise performance. In the future, we will continue to explore different framings of the C-way K-shot setup for improved accuracy. For example, treating 3D image patches from consistent spatial locations in a stereotactic space as distinct classes to allow enhanced feature encoding.

We found that misclassifications predominantly occurred near the brain surface, where surface veins and the dura (both with bright signals) reside. This was partially due to

the fact that the manual ground truths of the MRA segmentation primarily focus on the main arteries rather than the surface vasculatures, which is of interest in neurosurgical planning [127, 128]. The misclassifications may also be caused by training the model with random patches that were mostly taken from the center of the brain. In the future, we will incorporate random patches that consider both vessel and non-vessel regions, along with an increased number of patches.

The recent work by Li *et al.* [129] introduced a global vascular context network (GVC-Net) with a hybrid loss to address over-segmentation issues caused by sparse labels and skull vessels. They also utilized the TubeTk dataset by training on 42 data points and testing on 10 data points, achieving a sensitivity of 61.24%, precision of 75.58%, DC of 67.66%, intersection over the union of 51.13%, and centerline Dice coefficient of 83.79%. Although their method has a 5% higher Dice coefficient, it is important to note that their reported result is based on a single fold while our method's performance was from a full 4-fold cross-validation. In a separate study, Tang *et al.* [119] compared their proposed RPNet approach to PANet [3] in 3D abdominal image segmentation. RPNet outperformed PANet, achieving approximately 33% higher performance. Given these promising results, it would be worthwhile to investigate RPNet's application for 3D cerebral blood vessel segmentation task.

One major benefit of few-shot learning is the capacity to allow high flexibility and adaptability for unseen classes, which can include new classification/segmentation tasks and image contrasts. Our proposed approach shows sufficient generalizability to new classes as only a few image patches in the support set allowed the segmentation of the whole brain volume. In contrast, Holroyd *et al.* [130] developed tUbe net, a model that achieved high performance in segmenting new blood vessels through transfer learning. However, tUbe net requires a large training dataset, which may not always be available. Our method is independent of pre-trained weights and can be potentially applicable to diverse applications.

Despite efforts to utilize limited annotated data, the current performance of VesselShot is not sufficiently accurate for clinical deployment. However, we will explore the strategies mentioned above to improve the accuracy and robustness of few-shot cerebral vascular segmentation. Despite its limitations, our proposed method represents a preliminary step in addressing limited annotated data in the challenging task of 3D cerebral blood vessel segmentation.

7.6 Conclusion

Our novel method utilizes few-shot learning to address the challenges of limited labeled datasets in 3D cerebral blood vessel segmentation. This approach shows promise in overcoming the bottleneck of limited manually annotated datasets and could aid in clinical tasks with further improvement in the future. While the present achievement may not yet find direct application in clinical environments, it signifies an advancement in this domain.

Chapter 8

Conclusions and Future Work

Determining the best treatment strategies for ischemic stroke patients is vital. IV-tPA medication can only help when patients reach the hospital within 3 to 4.5 hours of symptom onset; thus, EVT is considered one of the best ways of treating ischemic stroke patients that is still possible after 6 hours of symptom onset. However, not all patients are suitable candidates, and the decision depends on several factors. Collateral circulation is one of the most important independent predictors, as good collateral circulation is correlated to positive clinical outcomes. For these reasons, evaluating the collateral circulation in ischemic stroke patients before choosing the appropriate treatment strategy is necessary. Numerous grading approaches with different imaging modalities have been used to score collateral circulation, providing no gold standard for the best collateral grading approach. However, it has been shown that manual interpretation by visually evaluating collaterals can be time-consuming, inaccurate, and suffer from inter- and intra-rater variability.

For this reason, automatic computer-based collateral grading techniques have become more prevalent in recent years. The automated approaches with an analog to the radiologists' grading criteria can evaluate collaterals more robustly than human-rater scoring, and these are more easily understandable in the clinical environment. This thesis explored different methods to automatically assess collaterals, overcoming the issues of small, imbalanced datasets, a common problem in the case of ischemic stroke datasets.

8.1 Summary of Findings

The thesis focused on evaluating collaterals in ischemic stroke automatically. Five computer-based automatic methods were developed, focusing on machine learning and deep learning approaches. All the methods' significance and limitations are summarized in Table 8.1.

Our first method, ACCESS, evaluates collaterals by quantifying the percentage of unfilled versus filled vessels in a test brain. ACCESS leverages 4D CTA for collateral assessment, providing detailed and dynamic information, which is superior to single-phase CTA, potentially leading to more accurate results. Additionally, it is less dependent on feature selection and training data, and the fRMC approach used to obtain the sparsity and low-rank metrics of blood vessels is independent of tunable parameters. Despite these advantages, the method has some limitations. Since the final evaluation is performed using 2D in contrast to the 3D scoring done by radiologists, the direct scoring terminology was inappropriate for our method. Therefore, we used the ROC curve to determine the optimal thresholds to decide the final collateral classes. The experiment's sample size was also

Table 8.1: Highlights of the proposed research studies

Method	Significance	Limitations
ACCESS [60]: Low-rank and sparse decomposition	Visual indications (e.g. filled versus unfilled vessels) that radiologists use	Lack of validation due to shortage of data
Deep transfer learning[79]	Handles small & imbalanced datasets	Performance can be enhanced with more data and pre-trained weights from medical domain
SVM with radiomic features of both sides of the brain [92]	NCCT is quick, no bolus, less radiation, more frequently in ERs	Hand-engineering of features and limited performance assessment
SCANED: Siamese network with Resnet3D as a backbone [108]	Robust in handling imbalanced and small dataset	Performance limitation of poor class which has a significant impact
VesselShot [116]: Few-shot learning with nnUNet as a backbone	Novel method for 3D complex blood vessels segmentation, which requires few annotated slices to segment the whole 3D brain; can further improve collateral evaluation's performance.	Performance needs to be improved by advanced feature learning before applying to clinical settings.

quite small (46 patients), therefore, further validation is needed before considering clinical application.

The next developed method aimed to assess collaterals using deep learning, specifically focusing on the sequential contrast flow data. We utilized the same dataset with a larger sample size for automatic collateral scoring through deep neural networks with feature extraction. While we employed 2D MIPs in this method, they encapsulate a sequence of contrast flow information. To address the data-hungry nature of deep neural networks, we applied transfer learning, leveraging pre-trained weights. We enhanced the effectiveness of this automatic method by using focal loss with class weights, particularly benefiting minority class samples. Consequently, our approach demonstrates greater robustness than the previous one, efficiently evaluating collaterals while addressing imbalanced small dataset challenges. Unlike the previous method, we compared our work with state-of-the-art (SOTA) methods and found efficacy over them. However, although the dichotomized classification shows efficacy over the SOTA methods, the multi-class classification still requires better performance. The reason behind this could be due to missing features as the whole 3D MIP of a brain was not used in training.

Our target was to implement automatic evaluation systems for collaterals using both 4D CTA and NCCT. NCCT can help determine collateral scoring from identified ischemic damage and we utilized this imaging to evaluate collaterals using both classic machine learning and deep learning methods. The radiomic features significantly impact ischemic damage, and these features from the affected and non-affected sides of ischemic patients can play an essential role in identifying non-symmetry between both sides. This non-symmetry results in tissue degeneration, which represents the lack of collaterals. This

idea is implemented in Chapter 5, which evaluated collaterals based on ischemic damage, feeding the radiomic features to SVM. The method distinguishes itself from most studies that rely on dynamic CTA by classifying collaterals based on ischemic damage from NCCT, saving time, avoiding the need for bolus injection, and reducing radiation exposure. Additionally, the method employs popular machine learning techniques, such as SVM, and leverages bilateral symmetry in contrast to the above two methods and radiomic features to automatically identify insufficient collateral regions, aligning more with radiologists' scores and methodology. A notable limitation of this study is the relatively small dataset, which poses a challenge for training a classifier. To mitigate this limitation, we employed 10-fold cross-validation with only eight poor cases; therefore, future work should assess the method's performance more comprehensively.

Unlike the previous method described in Chapter 5, the method described in Chapter 6 considered deep learning with a Siamese network, which has several advantages over the previous method. It does not need any manual intervention/hand engineering of features because it uses deep learning rather than classic machine learning methods. Also, the previous method does not consider the imbalanced issue, which is the primary concern of this method. Using the Siamese network, we developed a more robust approach that gives reasonable performance and handles imbalanced and small dataset issues. Comparison to the SOTA method showed the efficacy of our developed method to evaluate collaterals from 3D NCCT brains with a small imbalanced dataset.

Finally, our VesselShot method, presented in Chapter 7, is developed to segment cerebral vasculature from a small annotated labeled dataset. This novel method to extract blood vessels from the 3D brain requires a small number of annotated slices as ground truth. Using the few-shot approach with nn-UNet as a backbone, this novel method can segment 3D cerebral blood vessels with a few annotated slices. This method can help as a preliminary step in blood vessel extraction, which can be a prerequisite for collateral evaluation using 4D CTA. The method has limited performance and needs more validation before applying it to clinical settings.

8.2 Future Work

The limitations discussed above create opportunities for further research. In the future, the study can be extended, focusing on the limitations of the current work.

SCANED is a robust method to evaluate collaterals with deep learning while handling imbalanced and minor dataset issues. However, the performance is limited to assessing the poor collaterals, which play a significant role in the treatment decision. Therefore, some new approaches, for example, few-shot learning with prototypical networks, can be explored. A prototypical network [131] offers a more straightforward yet practical approach that creates a metric space, making classification as precise as measuring distances to prototype representations of each class. Few-shot settings with prototypical networks can help classify new, unseen classes with only a handful of examples. With this concept, poor collateral cases can be classified as an unseen class with few examples.

Although slice-based classification with 2D MIPs from 4D CTA helped us mitigate the problem of fewer training samples, 2D MIPs can still miss some critical sequential features from 3D. Thus, automatic approaches that directly use 3D volumes/3D MIPs may be more robust. A contralateral comparison can be performed by applying the 3D CNN separately to the left/right hemisphere in the MCA territory to investigate performance improvement.

For this purpose, a 3D CNN can be trained from scratch or using some existing renowned pre-trained networks; for example, EfficientNet, VGG, AlexNet, and ResNet with fine-tuning can be performed to get an optimal model for testing. Here, we can explore freezing some network layers given the limited training data [132, 133].

The pre-trained networks can be applied directly for classification. They can be considered feature extractors to feed the extracted features into classic machine learning algorithms, e.g., SVM, to see the performance. The ensemble of the pre-trained networks to mitigate issues by one single network following the sum rule introduced in the study by Nanni *et al.* [134] can also be explored. Following the findings of Su *et al.* [47], where the ratio of the number of collaterals between affected and non-affected sides as well as the vessel length are identified as significant features, these features can be incorporated further into pre-trained deep learning models or traditional machine learning classifiers to assess their performance.

Another research direction could be to explore the performance of the 3D collateral evaluation by extracting the vasculature through our proposed VesselShot [116] method presented in Chapter 7 rather than low-rank decomposition as used in our current works. Since VesselShot [116] needs few-annotated data to segment the cerebral blood vessels and ground truth is absent in our ischemic stroke dataset, this approach could be extended to self-supervised segmentation approach before using it as a pre-processing step in the collateral evaluation. In addition, test-time augmentation [135] can be used to both improve the results and provide uncertainty estimates.

Additionally, considering 3D volumes from 4D CTA directly provides temporal information, which can help distinguish between the collateral classes more significantly. The study by Li*et al.* [136] proposed a c3D-LSTM model that extracts spatial features from each 3D volume by combining a series of 3D CNN and further input those to a long short-term memory (LSTM) which can capture time-varying information. To consider temporal information in collateral evaluation, following the study by Li*et al.* [136], direct 3D volumes from the 4D CTA can be considered to utilize the contrast flow timing.

Another approach could be considering a Siamese network between the 3D volumes of a patient's 4D CTA to identify the differences between those with time. This way, the temporal flow of each case can be obtained, which could be a significant feature to distinguish it from other patients. Since multiple CNNs can be computationally expensive, four phases, arterial, arterio-venous, venous, and late venous, can be considered from the 19 volumes of 4D CTA. This phase extraction could be similar to the study by Huang *et al.* [52] or Tan *et al.* [49] who used the multiphase CTA to evaluate collaterals using deep learning.

In the thesis, a consensus of two neuroradiologists was used as the ground truth. Further, a better ground truth can also consider the clinical outcome, especially for intermediate collateral circulation cases where classification into poor and good circulation is challenging based on the imaging data alone.

Finally, validating our proposed methods in a larger dataset or comparing them with a small imbalanced dataset relevant to our problem-setting or obtained from another scanning protocol could be another research direction before applying those methods in clinical environments.

References

[1] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520, 2018.

[2] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature methods*, vol. 18, no. 2, pp. 203–211, 2021.

[3] K. Wang, J. H. Liew, Y. Zou, D. Zhou, and J. Feng, "PANet: Few-shot image semantic segmentation with prototype alignment," in *proceedings of the IEEE/CVF international conference on computer vision*, pp. 9197–9206, 2019.

[4] Y.-C. Kim, J.-W. Chung, O. Y. Bang, M. Hong, W.-K. Seo, G.-M. Kim, E. Yeop Kim, J. S. Lee, J. M. Hong, D. S. Liebeskind, and J. L. Saver, "A deep learning-based automatic collateral assessment in patients with acute ischemic stroke," *Translational Stroke Research*, pp. 1–7, 2022.

[5] Y. Xiao, A. Alamer, V. Fonov, B. W. Lo, D. Tampieri, D. L. Collins, H. Rivaz, and M. Kersten-Oertel, "Towards automatic collateral circulation score evaluation in ischemic stroke using image decompositions and support vector machines," in *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment*, pp. 158–167, Springer, 2017.

[6] V. K. Sharma, H. L. Teoh, L. Y. Wong, J. Su, B. K. Ong, and B. P. Chan, "Recanalization therapies in acute ischemic stroke: pharmacological agents, devices, and combinations," *Stroke research and treatment*, vol. 2010, 2010.

[7] N. I. of Neurological Disorders and S. rt PA Stroke Study Group, "Tissue plasminogen activator for acute ischemic stroke," *New England Journal of Medicine*, vol. 333, no. 24, pp. 1581–1588, 1995.

[8] W. Hacke, M. Kaste, E. Bluhmki, M. Brozman, A. Dávalos, D. Guidetti, V. Larrue, K. R. Lees, Z. Medeghri, T. Machnig, *et al.*, "Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke," *New England journal of medicine*, vol. 359, no. 13, pp. 1317–1329, 2008.

[9] G. J. Del Zoppo, J. L. Saver, E. C. Jauch, and H. P. Adams Jr, "Expansion of the time window for treatment of acute ischemic stroke with intravenous tissue plasminogen activator: a science advisory from the american heart association/american stroke association," *Stroke*, vol. 40, no. 8, pp. 2945–2948, 2009.

[10] R. G. Nogueira, A. P. Jadhav, D. C. Haussen, A. Bonafe, R. F. Budzik, P. Bhuva, D. R. Yavagal, M. Ribo, C. Cognard, R. A. Hanel, *et al.*, "Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct," *New England Journal of Medicine*, vol. 378, no. 1, pp. 11–21, 2018.

[11] S. A. Munich, K. Vakharia, and E. I. Levy, "Overview of mechanical thrombectomy techniques," *Neurosurgery*, vol. 85, no. suppl_1, pp. S60–S67, 2019.

[12] S. Jung, R. Wiest, J. Gralla, R. McKinley, H. Mattle, and D. Liebeskind, "Relevance of the cerebral collateral circulation in ischaemic stroke: time is brain, but collaterals set the pace.," *Swiss medical weekly*, vol. 147, no. w14538, p. w14538, 2017.

[13] S. Zhang, W. Chen, H. Tang, Q. Han, S. Yan, X. Zhang, Q. Chen, M. Parsons, S. Wang, and M. Lou, "The prognostic value of a four-dimensional ct angiography-based collateral grading scale for reperfusion therapy in acute ischemic stroke patients," *PLoS One*, vol. 11, no. 8, 2016.

[14] R. Cao, P. Qi, Y. Liu, X. Ma, Z. Shen, and J. Chen, "Improving prognostic evaluation by 4d cta for endovascular treatment in acute ischemic stroke patients: A preliminary study," *Journal of Stroke and Cerebrovascular Diseases*, vol. 28, no. 7, pp. 1971–1978, 2019.

[15] M. Goyal, A. M. Demchuk, B. K. Menon, M. Eesa, J. L. Rempel, J. Thornton, D. Roy, T. G. Jovin, R. A. Willinsky, B. L. Sapkota, *et al.*, "Randomized assessment of rapid endovascular treatment of ischemic stroke," *New England Journal of Medicine*, vol. 372, no. 11, pp. 1019–1030, 2015.

[16] O. A. Berkhemer, I. G. Jansen, D. Beumer, P. S. Fransen, L. A. Van Den Berg, A. J. Yoo, H. F. Lingsma, M. E. Sprengers, S. F. Jenniskens, G. J. Lycklama à Nijeholt, *et al.*, "Collateral status on baseline computed tomographic angiography and intra-arterial treatment effect in patients with proximal anterior circulation stroke," *Stroke*, vol. 47, no. 3, pp. 768–776, 2016.

[17] M. El Amki and S. Wegener, "Improving cerebral blood flow after arterial recanalization: a novel therapeutic strategy in stroke," *International journal of molecular sciences*, vol. 18, no. 12, p. 2669, 2017.

[18] J. J. Heit, G. Zaharchuk, and M. Wintermark, "Advanced neuroimaging of acute ischemic stroke: penumbra and collateral assessment," *Neuroimaging Clinics*, vol. 28, no. 4, pp. 585–597, 2018.

[19] L. Liu, J. Ding, X. Leng, Y. Pu, L.-A. Huang, A. Xu, K. S. L. Wong, X. Wang, and Y. Wang, "Guidelines for evaluation and management of cerebral collateral circulation in ischaemic stroke 2017," *Stroke and Vascular Neurology*, vol. 3, no. 3, pp. 117–130, 2018.

[20] H. Kortman, E. Smit, M. Oei, R. Manniesing, M. Prokop, and F. Meijer, "4d-cta in neurovascular disease: a review," *American Journal of Neuroradiology*, vol. 36, no. 6, pp. 1026–1033, 2015.

[21] C. Yuan, T. S. Hatsukami, and M. Mossa-Basha, *Vessel Based Imaging Techniques: Diagnosis, Treatment, and Prevention*. Springer Nature, 2019.

[22] J. C. Grotta, D. Chiu, M. Lu, S. Patel, S. R. Levine, B. C. Tilley, T. G. Brott, E. C. Haley Jr, P. D. Lyden, R. Kothari, *et al.*, "Agreement and variability in the interpretation of early ct changes in stroke patients qualifying for intravenous rtpa therapy," *Stroke*, vol. 30, no. 8, pp. 1528–1533, 1999.

[23] S. B. Coutts, M. D. Hill, A. M. Demchuk, P. A. Barber, J. Pexman, A. Buchan, H. Mak, K. Yau, and B. Chan, "Aspects reading requires training and experience," *Stroke*, vol. 34, no. 10, p. e179, 2003.

[24] R. von Kummer, R. Holle, U. Gizyska, E. Hofmann, O. Jansen, D. Petersen, M. Schumacher, and K. Sartor, "Interobserver agreement in assessing early ct signs of middle cerebral artery infarction.," *American journal of neuroradiology*, vol. 17, no. 9, pp. 1743–1748, 1996.

[25] I. Q. Grunwald, J. Kulikovski, W. Reith, S. Gerry, R. Namias, M. Politi, P. Papanagiotou, M. Essig, S. Mathur, O. Joly, *et al.*, "Collateral automation for triage in stroke: Evaluating automated scoring of collaterals in acute stroke on computed tomography scans," *Cerebrovascular Diseases*, vol. 47, no. 5-6, pp. 217–222, 2019.

[26] J. W. Pexman, P. A. Barber, M. D. Hill, R. J. Sevick, A. M. Demchuk, M. E. Hudon, W. Y. Hu, and A. M. Buchan, "Use of the alberta stroke program early ct score (aspects) for assessing ct scans in patients with acute stroke," *American Journal of Neuroradiology*, vol. 22, no. 8, pp. 1534–1542, 2001.

[27] R. T. Higashida and A. J. Furlan, "Trial design and reporting standards for intra-arterial cerebral thrombolysis for acute ischemic stroke," *Stroke*, vol. 34, no. 8, pp. e109–e137, 2003.

[28] G. A. Christoforidis, Y. Mohammad, D. Kehagias, B. Avutu, and A. P. Slivka, "Angiographic assessment of pial collaterals as a prognostic indicator following intra-arterial thrombolysis for acute ischemic stroke," *American Journal of Neuroradiology*, vol. 26, no. 7, pp. 1789–1797, 2005.

[29] F. Miteff, C. R. Levi, G. A. Bateman, N. Spratt, P. McElduff, and M. W. Parsons, "The independent predictive utility of computed tomography angiographic collateral status in acute ischaemic stroke," *Brain*, vol. 132, no. 8, pp. 2231–2238, 2009.

[30] M. B. Maas, M. H. Lev, H. Ay, A. B. Singhal, D. M. Greer, W. S. Smith, G. J. Harris, E. Halpern, A. Kemmling, W. J. Koroshetz, *et al.*, "Collateral vessels on ct angiography predict outcome in acute ischemic stroke," *Stroke*, vol. 40, no. 9, pp. 3001–3005, 2009.

[31] I. Tan, A. Demchuk, J. Hopyan, L. Zhang, D. Gladstone, K. Wong, M. Martin, S. Symons, A. Fox, and R. Aviv, "Ct angiography clot burden score and collateral score: correlation with clinical and radiologic outcomes in acute middle cerebral artery infarct," *American Journal of Neuroradiology*, vol. 30, no. 3, pp. 525–531, 2009.

[32] B. Menon, E. Smith, J. Modi, S. Patel, R. Bhatia, T. Watson, M. Hill, A. Demchuk, and M. Goyal, "Regional leptomeningeal score on ct angiography predicts clinical and imaging outcomes in patients with acute anterior circulation occlusions," *American journal of neuroradiology*, vol. 32, no. 9, pp. 1640–1645, 2011.

[33] B. K. Menon, C. D. d'Esterre, E. M. Qazi, M. Almekhlafi, L. Hahn, A. M. Demchuk, and M. Goyal, "Multiphase ct angiography: a new tool for the imaging triage of patients with acute ischemic stroke," *Radiology*, vol. 275, no. 2, pp. 510–520, 2015.

[34] J. J. Kim, N. J. Fischbein, Y. Lu, D. Pham, and W. P. Dillon, "Regional angiographic grading system for collateral flow: correlation with cerebral infarction in patients with middle cerebral artery occlusion," *Stroke*, vol. 35, no. 6, pp. 1340–1344, 2004.

[35] R. D. Henderson, M. Eliasziw, A. J. Fox, P. M. Rothwell, and H. J. Barnett, "Angiographically defined collateral circulation and risk of stroke in patients with severe carotid artery stenosis," *Stroke*, vol. 31, no. 1, pp. 128–132, 2000.

[36] H. C. Alves, K. M. Treurniet, B. G. Dutra, I. G. Jansen, A. M. Boers, E. M. Santos, O. A. Berkhemer, D. W. Dippel, A. van der Lugt, W. H. van Zwam, *et al.*, "Associations between collateral status and thrombus characteristics and their impact in anterior circulation stroke," *Stroke*, vol. 49, no. 2, pp. 391–396, 2018.

[37] A. Flores, M. Rubiera, M. Ribó, J. Pagola, D. Rodriguez-Luna, M. Muchada, S. Boned, L. Seró, E. Sanjuan, P. Meler, *et al.*, "Poor collateral circulation assessed by multiphase computed tomographic angiography predicts malignant middle cerebral artery evolution after reperfusion therapies," *Stroke*, vol. 46, no. 11, pp. 3149–3153, 2015.

[38] J. Saarinen, H. Rusanen, and N. Sillanpää, "Collateral score complements clot location in predicting the outcome of intravenous thrombolysis," *American Journal of Neuroradiology*, vol. 35, no. 10, pp. 1892–1896, 2014.

[39] B. K. Menon, B. O'brien, A. Bivard, N. J. Spratt, A. M. Demchuk, F. Miteff, X. Lu, C. Levi, and M. W. Parsons, "Assessment of leptomeningeal collaterals using dynamic ct angiography in patients with acute ischemic stroke," *Journal of Cerebral Blood Flow & Metabolism*, vol. 33, no. 3, pp. 365–371, 2013.

[40] B. K. Menon, E. Qazi, V. Nambiar, L. D. Foster, S. D. Yeatts, D. Liebeskind, T. G. Jovin, M. Goyal, M. D. Hill, T. A. Tomsick, *et al.*, "Differential effect of baseline computed tomographic angiography collaterals on clinical outcome in patients enrolled in the interventional management of stroke iii trial," *Stroke*, vol. 46, no. 5, pp. 1239–1244, 2015.

[41] F. Seker, A. Potreck, M. Möhlenbruch, M. Bendszus, and M. Pham, "Comparison of four different collateral scores in acute ischemic stroke by ct angiography," *Journal of neurointerventional surgery*, vol. 8, no. 11, pp. 1116–1118, 2016.

[42] M. Kersten-Oertel, A. Alamer, V. Fonov, B. Lo, D. Tampieri, and L. Collins, "Towards a computed collateral circulation score in ischemic stroke," *arXiv preprint arXiv:2001.07169*, 09 2016.

[43] A. Boers, R. S. Barros, I. Jansen, O. Berkhemer, L. Beenen, B. K. Menon, D. Dippel, A. van der Lugt, W. van Zwam, Y. Roos, and van Oostenbrugge RJ, "Value of quantitative collateral scoring on ct angiography in patients with acute ischemic stroke," *American Journal of Neuroradiology*, vol. 39, no. 6, pp. 1074–1082, 2018.

[44] Y. Shieh, C.-H. Chang, M. Shieh, T.-H. Lee, Y. J. Chang, H.-F. Wong, S. C. Chin, and S. Goodwin, "Computer-aided diagnosis of hyperacute stroke with thrombolysis decision support using a contralateral comparative method of ct image analysis," *Journal of digital imaging*, vol. 27, no. 3, pp. 392–406, 2014.

[45] A. M. Frolich, S. L. Wolff, M. N. Psychogios, E. Klotz, R. Schramm, K. Wasser, M. Knauth, and P. Schramm, "Time-resolved assessment of collateral flow using 4d ct angiography in large-vessel occlusion stroke," *European radiology*, vol. 24, no. 2, pp. 390–396, 2014.

[46] H. Kuang, M. Najm, D. Chakraborty, N. Maraj, S.-I. Sohn, M. Goyal, M. Hill, A. Demchuk, B. Menon, and W. Qiu, "Automated aspects on non-contrast ct scans in acute ischemic stroke patients using machine learning," *American Journal of Neuroradiology*, 12 2018.

[47] J. Su, L. Wolff, A. C. M. van Es, W. van Zwam, C. Majoie, D. W. Dippel, A. van der Lugt, W. J. Niessen, and T. Van Walsum, "Automatic collateral scoring from 3d cta images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 2190–2200, 2020.

[48] G. Tetteh, F. Navarro, R. Meier, J. Kaesmacher, J. C. Paetzold, J. S. Kirschke, C. Zimmer, R. Wiest, and B. H. Menze, "A deep learning approach to predict collateral flow in stroke patients using radiomic features from perfusion images," *Frontiers in neurology*, vol. 14, p. 1039693, 2023.

[49] D. Tan, J. Wang, R. Yao, J. Liu, J. Wu, S. Zhu, Y. Yang, S. Chen, and Y. Li, "Cca4cta: A hybrid attention mechanism based convolutional network for analysing collateral circulation via multi-phase cranial cta," in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1201–1206, IEEE, 2022.

[50] R. A. Rava, S. E. Seymour, K. V. Snyder, M. Waqas, J. M. Davies, E. I. Levy, A. H. Siddiqui, and C. N. Ionita, "Automated collateral flow assessment in patients with acute ischemic stroke using computed tomography with artificial intelligence algorithms," *World Neurosurgery*, vol. 155, pp. e748–e760, 2021.

[51] N. H. Ali, A. R. Abdullah, N. M. Saad, and A. S. Muda, "Brain cone beam computed tomography image analysis using resnet50 for collateral circulation classification," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 5, pp. 5843–5852, 2023.

[52] C.-C. Huang, H.-F. Chiang, C.-C. Hsieh, C.-L. Chou, Z.-Y. Jhou, T.-Y. Hou, and J.-S. Shaw, "Using deep-learning-based artificial intelligence technique to automatically evaluate the collateral status of multiphase cta in acute ischemic stroke," *Tomography*, vol. 9, no. 2, pp. 647–656, 2023.

[53] L.-N. Do, I. Park, H.-J. Yang, B.-H. Baek, Y. Nam, and W. Yoon, "Automatic assessment of dwi-aspects for assessment of acute ischemic stroke using 3d convolutional neural network," in *Proceedings of the The 6th International Conference on Big Data Applications and Services, Zhengzhou, China*, pp. 19–22, 2018.

[54] L.-N. Do, B. H. Baek, S. K. Kim, H.-J. Yang, I. Park, and W. Yoon, "Automatic assessment of aspects using diffusion-weighted imaging in acute ischemic stroke using

recurrent residual convolutional neural network," *Diagnostics*, vol. 10, no. 10, p. 803, 2020.

[55] P.-L. Chiang, S.-Y. Lin, M.-H. Chen, Y.-S. Chen, C.-K. Wang, M.-C. Wu, Y.-T. Huang, M.-Y. Lee, Y.-S. Chen, and W.-C. Lin, "Deep learning-based automatic detection of aspects in acute ischemic stroke: Improving stroke assessment on ct scans," *Journal of Clinical Medicine*, vol. 11, no. 17, p. 5159, 2022.

[56] Z. Cao, J. Xu, B. Song, L. Chen, T. Sun, Y. He, Y. Wei, G. Niu, Y. Zhang, Q. Feng, *et al.*, "Deep learning derived automated aspects on non-contrast ct scans of acute ischemic stroke patients," tech. rep., Wiley Online Library, 2022.

[57] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," *Advances in neural information processing systems*, vol. 27, 2014.

[58] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, pp. 6105–6114, PMLR, 2019.

[59] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a" siamese" time delay neural network," *Advances in neural information processing systems*, vol. 6, 1993.

[60] M. Aktar, D. Tampieri, H. Rivaz, M. Kersten-Oertel, and Y. Xiao, "Automatic collateral circulation scoring in ischemic stroke using 4d ct angiography with low-rank and sparse matrix decomposition," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, pp. 1501–1511, 2020.

[61] E. Cuccione, G. Padovano, A. Versace, C. Ferrarese, and S. Beretta, "Cerebral collateral circulation in experimental ischemic stroke," *Experimental & translational stroke medicine*, 1 2016.

[62] G. S. Piedade, C. M. Schirmer, O. Goren, H. Zhang, A. Aghajanian, J. E. Faber, and C. J. Griessenauer, "Cerebral collateral circulation: a review in the context of ischemic stroke and mechanical thrombectomy," *World neurosurgery*, vol. 122, pp. 33–42, 2019.

[63] B. Rezaei and S. Ostadabbas, "Background subtraction via fast robust matrix completion," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1871–1879, 2017.

[64] R. M. Freund, P. Grigas, and R. Mazumder, "An extended frank–wolfe method with "in-face" directions, and its application to low-rank matrix completion," *SIAM Journal on Optimization*, vol. 27, no. 1, pp. 319–346, 2017.

[65] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain," *Medical image analysis*, vol. 12, no. 1, pp. 26–41, 2008.

[66] V. Fonov, A. C. Evans, K. Botteron, C. R. Almli, R. C. McKinstry, D. L. Collins, and B. D. C. Group, "Unbiased average age-appropriate atlases for pediatric studies," *Neuroimage*, vol. 54, no. 1, pp. 313–327, 2011.

[67] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.

[68] M. Ashikuzzaman, C. Belasso, M. G. Kibria, A. Bergdahl, C. J. Gauthier, and H. Rivaz, "Low rank and sparse decomposition of ultrasound color flow images for suppressing clutter in real-time," *IEEE Transactions on Medical Imaging*, 2019.

[69] S. Moccia, E. De Momi, S. El Hadji, and L. S. Mattos, "Blood vessel segmentation algorithms—review of methods, datasets and evaluation metrics," *Computer methods and programs in biomedicine*, vol. 158, pp. 71–91, 2018.

[70] X. Yang, C. Liu, H. Le Minh, Z. Wang, A. Chien, and K.-T. T. Cheng, "An automated method for accurate vessel segmentation," *Physics in Medicine & Biology*, vol. 62, no. 9, p. 3757, 2017.

[71] M. Meijs, A. Patel, S. C. van de Leemput, M. Prokop, E. J. van Dijk, F.-E. de Leeuw, F. J. Meijer, B. van Ginneken, and R. Manniesing, "Robust segmentation of the full cerebral vasculature in 4d ct of suspected stroke patients," *Scientific reports*, vol. 7, no. 1, pp. 1–12, 2017.

[72] M. Jin, D. Hao, S. Ding, and B. Qin, "Low-rank and sparse decomposition with spatially adaptive filtering for sequential segmentation of 2d+ t vessels," *Physics in Medicine & Biology*, vol. 63, no. 17, p. 17LT01, 2018.

[73] T. Jerman, F. Pernuš, B. Likar, and Ž. Špiclin, "Enhancement of vascular structures in 3d and 2d angiographic images," *IEEE transactions on medical imaging*, vol. 35, no. 9, pp. 2107–2118, 2016.

[74] J. Huck, Y. Wanner, A. P. Fan, A.-T. Jäger, S. Grahl, U. Schneider, A. Villringer, C. J. Steele, C. L. Tardif, P.-L. Bazin, and C. J. Gauthier, "High resolution atlas of the venous brain vasculature from 7 t quantitative susceptibility maps," *Brain Structure and Function*, vol. 224, no. 7, pp. 2467–2485, 2019.

[75] T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[76] J. L. Fleiss, "Measuring nominal scale agreement among many raters.," *Psychological bulletin*, vol. 76, no. 5, p. 378, 1971.

[77] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *biometrics*, pp. 159–174, 1977.

[78] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and psychological measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[79] M. Aktar, J. Reyes, D. Tampieri, H. Rivaz, Y. Xiao, and M. Kersten-Oertel, "Deep learning for collateral evaluation in ischemic stroke with imbalanced data," *International Journal of Computer Assisted Radiology and Surgery*, vol. 18, no. 4, pp. 733–740, 2023.

[80] S.-M. Jung and T.-K. Whangbo, "A deep learning system for diagnosing ischemic stroke by applying adaptive transfer learning," *Journal of Internet Technology*, vol. 21, no. 7, pp. 1957–1968, 2020.

[81] A. Neethi, S. Niyas, S. K. Kannath, J. Mathew, A. M. Anzar, and J. Rajan, "Stroke classification from computed tomography scans using 3d convolutional neural network," *Biomedical Signal Processing and Control*, vol. 76, p. 103720, 2022.

[82] Z. N. K. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali, S. Ahmed, and J. Lu, "Brain tumor classification for mr images using transfer learning and fine-tuning," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 34–46, 2019.

[83] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.

[84] K. Ali, Z. A. Shaikh, A. A. Khan, and A. A. Laghari, "Multiclass skin cancer classification using efficientnets–a first step towards preventing skin cancer," *Neuroscience Informatics*, p. 100034, 2021.

[85] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017.

[86] S. Sarraf, D. D. DeSouza, J. Anderson, and G. Tofighi, "Deepad: Alzheimer's disease classification via deep convolutional neural networks using mri and fmri," *BioRxiv*, p. 070441, 2017.

[87] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[88] R. Mehra, "Breast cancer histology images classification: Training from scratch or transfer learning?," *ICT Express*, vol. 4, no. 4, pp. 247–254, 2018.

[89] T. Kaur and T. K. Gandhi, "Automated brain image classification based on vgg-16 and transfer learning," in *2019 International Conference on Information Technology (ICIT)*, pp. 94–98, IEEE, 2019.

[90] S. Osama, K. Zafar, and M. U. Sadiq, "Predicting clinical outcome in acute ischemic stroke using parallel multi-parametric feature embedded siamese network," *Diagnostics*, vol. 10, no. 11, p. 858, 2020.

[91] C.-F. Liu, S. Padhy, S. Ramachandran, V. X. Wang, A. Efimov, A. Bernal, L. Shi, M. Vaillant, J. T. Ratnanather, A. V. Faria, B. Caffo, M. Albert, and M. I. Miller, "Using deep siamese neural networks for detection of brain asymmetries associated with alzheimer's disease and mild cognitive impairment," *Magnetic resonance imaging*, vol. 64, pp. 190–199, 2019.

[92] M. Aktar, Y. Xiao, D. Tampieri, H. Rivaz, and M. Kersten-Oertel, "A radiomics-based machine learning approach to assess collateral circulation in ischemic stroke on non-contrast computed tomography," in *Multimodal Learning for Clinical Decision Support and Clinical Image-Based Procedures*, pp. 24–33, Springer, 2020.

[93] C. Maegerlein, J. Fischer, S. Mönch, M. Berndt, S. Wunderlich, C. L. Seifert, M. Lehm, T. Boeckh-Behrens, C. Zimmer, and B. Friedrich, "Automated calculation of the alberta stroke program early ct score: feasibility and reliability," *Radiology*, vol. 291, no. 1, pp. 141–148, 2019.

[94] E. Kellner, M. Reisert, V. Kiselev, C. Maurer, H. Urbach, and K. Egger, "Comparison of automated and visual dwi aspects in acute ischemic stroke," *Journal of Neuroradiology*, vol. 46, no. 5, pp. 288–293, 2019.

[95] F. Sallustio, C. Motta, S. Pizzuto, M. Diomedi, B. Rizzato, M. Panella, F. Alemseged, M. Stefanini, S. Fabiano, R. Gandini, *et al.*, "Ct angiography aspects predicts outcome much better than noncontrast ct in patients with stroke treated endovascularly," *American Journal of Neuroradiology*, vol. 38, no. 8, pp. 1569–1573, 2017.

[96] V. Sundaram, J. Goldstein, D. Wheelwright, A. Aggarwal, P. Pawha, A. Doshi, J. Fifi, R. De Leacy, J. Mocco, J. Puig, *et al.*, "Automated aspects in acute ischemic stroke: a comparative analysis with ct perfusion," *American Journal of Neuroradiology*, vol. 40, no. 12, pp. 2033–2038, 2019.

[97] H. Kuang, M. Najm, D. Chakraborty, N. Maraj, S. Sohn, M. Goyal, M. Hill, A. Demchuk, B. Menon, and W. Qiu, "Automated aspects on noncontrast ct scans in patients with acute ischemic stroke using machine learning," *American Journal of Neuroradiology*, vol. 40, no. 1, pp. 33–38, 2019.

[98] Y. Shieh, C.-H. Chang, M. Shieh, T.-H. Lee, Y. J. Chang, H.-F. Wong, S. C. Chin, and S. Goodwin, "Computer-aided diagnosis of hyperacute stroke with thrombolysis decision support using a contralateral comparative method of ct image analysis," *Journal of digital imaging*, vol. 27, no. 3, pp. 392–406, 2014.

[99] F. Fahmi, H. A. Marquering, C. B. Majoie, M. A. van Walderveen, G. J. Streekstra, *et al.*, "Image based automated aspect score for acute ischemic stroke patients," in *2017 5th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering*, pp. 1–5, IEEE, 2017.

[100] R. Bhatia, S. S. Bal, N. Shobha, B. K. Menon, S. Tymchuk, V. Puetz, I. Dzialowski, S. B. Coutts, M. Goyal, P. A. Barber, *et al.*, "Ct angiographic source images predict outcome and final infarct volume better than noncontrast ct in proximal vascular occlusions," *Stroke*, vol. 42, no. 6, pp. 1575–1580, 2011.

[101] J. Y. Choi, E. J. Kim, J. M. Hong, S. E. Lee, J. S. Lee, Y. C. Lim, and H. S. Kim, "Conventional enhancement ct: a valuable tool for evaluating pial collateral flow in acute ischemic stroke," *Cerebrovascular Diseases*, vol. 31, no. 4, pp. 346–352, 2011.

[102] Y. Mokli, J. Pfaff, D. P. dos Santos, C. Herweh, and S. Nagel, "Computer-aided imaging analysis in acute ischemic stroke–background and clinical applications," *Neurological Research and Practice*, vol. 1, no. 1, p. 23, 2019.

[103] J. Muschelli, N. L. Ullman, W. A. Mould, P. Vespa, D. F. Hanley, and C. M. Crainiceanu, "Validated automatic brain extraction of head ct images," *Neuroimage*, vol. 114, pp. 379–385, 2015.

[104] P. Burt and E. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on communications*, vol. 31, no. 4, pp. 532–540, 1983.

[105] J. J. Van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. Beets-Tan, J.-C. Fillion-Robin, S. Pieper, and H. J. Aerts, "Computational radiomics system to decode the radiographic phenotype," *Cancer research*, vol. 77, no. 21, pp. e104–e107, 2017.

[106] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[107] H. Kamal, V. Lopez, and S. A. Sheth, "Machine learning in acute ischemic stroke neuroimaging," *Frontiers in neurology*, vol. 9, p. 945, 2018.

[108] M. Aktar, Y. Xiao, A. K. Z. Tehrani, D. Tampieri, H. Rivaz, and M. Kersten-Oertel, "Scaned: Siamese collateral assessment network for evaluation of collaterals from ischemic damage," *submitted to Computerized medical imaging and graphics*, 2023.

[109] A. B. Vieira, A. C. Fonseca, J. Ferro, and A. L. Oliveira, "Using a siamese network to accurately detect ischemic stroke in computed tomography scans," in *Progress in Artificial Intelligence: 21st EPIA Conference on Artificial Intelligence, EPIA 2022, Lisbon, Portugal, August 31–September 2, 2022, Proceedings*, pp. 159–170, Springer, 2022.

[110] A. Barman, M. E. Inam, S. Lee, S. Savitz, S. Sheth, and L. Giancardo, "Determining ischemic stroke from ct-angiography imaging using symmetry-sensitive convolutional networks," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 1873–1877, IEEE, 2019.

[111] S. Chen, K. Ma, and Y. Zheng, "Med3d: Transfer learning for 3d medical image analysis," *arXiv preprint arXiv:1904.00625*, 2019.

[112] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2, pp. 1735–1742, IEEE, 2006.

[113] K. Roy, D. Banik, D. Bhattacharjee, and M. Nasipuri, "Patch-based system for classification of breast histology images using deep learning," *Computerized Medical Imaging and Graphics*, vol. 71, pp. 90–103, 2019.

[114] M. J. Cardoso, W. Li, R. Brown, N. Ma, E. Kerfoot, Y. Wang, B. Murrey, A. Myronenko, C. Zhao, D. Yang, *et al.*, "Monai: An open-source framework for deep learning in healthcare," *arXiv preprint arXiv:2211.02701*, 2022.

[115] N. Gessert, T. Sentker, F. Madesta, R. Schmitz, H. Kniep, I. Baltruschat, R. Werner, and A. Schlaefer, "Skin lesion classification using cnns with patch-based attention and diagnosis-guided loss weighting," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 2, pp. 495–503, 2019.

[116] M. Aktar, H. Rivaz, M. Kersten-Oertel, and Y. Xiao, "Vesselshot: Few-shot learning for cerebral blood vessel segmentation," *arXiv preprint arXiv:2308.14626*, 2023.

[117] M. R. Goni, N. I. R. Ruhaiyem, M. Mustapha, A. Achuthan, and C. M. N. C. M. Nassir, "Brain vessel segmentation using deep learning-a review," *IEEE Access*, 2022.

[118] A. G. Roy, S. Siddiqui, S. Pölsterl, N. Navab, and C. Wachinger, "'squeeze & excite'guided few-shot segmentation of volumetric images," *Medical image analysis*, vol. 59, p. 101587, 2020.

[119] H. Tang, X. Liu, S. Sun, X. Yan, and X. Xie, "Recurrent mask refinement for few-shot medical image segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 3918–3928, 2021.

[120] C. Ouyang, C. Biffi, C. Chen, T. Kart, H. Qiu, and D. Rueckert, "Self-supervision with superpixels: Training few-shot medical image segmentation without annotation," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX 16*, pp. 762–780, Springer, 2020.

[121] A. K. Mondal, J. Dolz, and C. Desrosiers, "Few-shot 3d multi-modal medical image segmentation using generative adversarial learning," *arXiv preprint arXiv:1810.12241*, 2018.

[122] S. Guo, L. Xu, C. Feng, H. Xiong, Z. Gao, and H. Zhang, "Multi-level semantic adaptation for few-shot segmentation on cardiac image sequences," *Medical Image Analysis*, vol. 73, p. 102170, 2021.

[123] J. Xu, J. Shen, C. Wan, Q. Jiang, Z. Yan, and W. Yang, "A few-shot learning-based retinal vessel segmentation method for assisting in the central serous chorioretinopathy laser surgery," *Frontiers in Medicine*, vol. 9, 2022.

[124] Y. Wang, G. Yan, H. Zhu, S. Buch, Y. Wang, E. M. Haacke, J. Hua, and Z. Zhong, "VC-Net: Deep volume-composition networks for segmentation and visualization of highly sparse and noisy image data," *IEEE transactions on visualization and computer graphics*, vol. 27, no. 2, pp. 1301–1311, 2020.

[125] J. Su, Z. Liu, J. Zhang, V. S. Sheng, Y. Song, Y. Zhu, and Y. Liu, "DV-Net: Accurate liver vessel segmentation via dense connection model with d-bce loss function," *Knowledge-Based Systems*, vol. 232, p. 107471, 2021.

[126] M. Livne, J. Rieger, O. U. Aydin, A. A. Taha, E. M. Akay, T. Kossen, J. Sobesky, J. D. Kelleher, K. Hildebrand, D. Frey, *et al.*, "A u-net deep learning framework for high performance vessel segmentation in patients with cerebrovascular disease," *Frontiers in neuroscience*, vol. 13, p. 97, 2019.

[127] O. Hellum, Y. Mu, M. Kersten-Oertel, and Y. Xiao, "A novel prototype for virtual-reality-based deep brain stimulation trajectory planning using voodoo doll annotation and eye-tracking," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 10, no. 4, pp. 418–424, 2022.

[128] S. Bériault, Y. Xiao, L. Bailey, D. L. Collins, F. Sadikot, Abbas, and G. B. Pike, "Towards computer-assisted deep brain stimulation targeting with multiple active

contacts," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012: 15th International Conference, Nice, France, October 1-5, 2012, Proceedings, Part I 15*, pp. 487–494, Springer, 2012.

[129] M. Li, S. Li, Y. Han, and T. Zhang, "GVC-Net: Global vascular context network for cerebrovascular segmentation using sparse labels," *IRBM*, vol. 43, no. 6, pp. 561–572, 2022.

[130] N. A. Holroyd, Z. Li, C. Walsh, E. E. Brown, R. J. Shipley, and S. Walker-Samuel, "tUbe net: a generalizable deep learning tool for 3d vessel segmentation," *bioRxiv*, pp. 2023–07, 2023.

[131] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," *Advances in neural information processing systems*, vol. 30, 2017.

[132] M. Amiri, R. Brooks, and H. Rivaz, "Fine tuning u-net for ultrasound image segmentation: Which layers?," in *MICCAI Workshop on Domain Adaptation and Representation Transfer*, pp. 235–242, Springer, 2019.

[133] M. Amiri, R. Brooks, and H. Rivaz, "Fine-tuning u-net for ultrasound image segmentation: different layers, different outcomes," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 12, pp. 2510–2518, 2020.

[134] L. Nanni, M. Interlenghi, S. Brahnam, C. Salvatore, S. Papa, R. Nemni, I. Castiglioni, A. D. N. Initiative, *et al.*, "Comparison of transfer learning and conventional machine learning applied to structural brain mri for the early diagnosis and prognosis of alzheimer's disease," *Frontiers in neurology*, vol. 11, 2020.

[135] M. Amiri, R. Brooks, B. Behboodi, and H. Rivaz, "Two-stage ultrasound image segmentation using u-net and test time augmentation," *International journal of computer assisted radiology and surgery*, vol. 15, pp. 981–988, 2020.

[136] W. Li, X. Lin, and X. Chen, "Detecting alzheimer's disease based on 4d fmri: An exploration under deep learning framework," *Neurocomputing*, vol. 388, pp. 280–287, 2020.