



Machine Learning Characterization of Vascular Functions in Stroke Perfusion Imaging

Ezequiel de la Rosa

Vollständiger Abdruck der von der TUM School of Computation, Information and Technology der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz:

Prof. Dr. Felix Kraemer

Prüfer der Dissertation:

1. Prof. Dr. Bjoern H. Menze
2. Prof. Dr. Jef Vandemeulebroucke

Die Dissertation wurde am 28.06.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Computation, Information and Technology am 09.11.2023 angenommen.

To my family.
Dedicado a mi familia.



Abstract

Perfusion CT (CTP) imaging is crucial in acute ischemic stroke for assessing brain tissue viability and guiding reperfusion treatments. Deconvolution algorithms enable the quantification of brain perfusion status and the identification of the salvageable *penumbra* and the irreversibly damaged *core*. However, the accuracy of deconvolution algorithms heavily relies on the selection of two input parameters: the arterial input function (AIF) and the venous output function. Therefore, their precise selection is crucial to ensure the accurate quantification of perfusion metrics and reliable treatment decision-making. This thesis investigates the characterization of vascular functions in acute ischemic stroke using data-driven machine-learning models. Three research questions ($Q1$, $Q2$, $Q3$) are addressed. Firstly, $Q1$ challenges the concept of the “optimal” AIF in perfusion imaging and leverages machine-learning methods to identify it. A differentiable deconvolution algorithm is integrated within a deep-learning framework, enabling the estimation of perfusion biomarkers during training stages and, therefore, enhancing the learning process. The proposed model shows superior performance compared to manually or clinically selected AIFs in core tissue segmentation. Secondly, $Q2$ targets the development of machine-learning models that can select vascular functions from clinical and real-world CTP scans. A novel approach called *AIFNet* is introduced, utilizing convolutional neural networks to accurately identify vascular functions from clinical CTP scans. The method is end-to-end trainable and can be easily fit using sparse annotations. Furthermore, *AIFNet* achieves comparable performance to experts in selecting vascular functions across various CTP data and quality scenarios. Lastly, $Q3$ aims to devise machine-learning models that use information from vascular functions to detect unreliable CTP scans affected by truncation artifacts. The method outperforms the suggested baseline, which solely relies on the scan duration, improving the interpretability of CTP analysis and enabling neurointerventionalists to make safer and more reliable treatment decisions. In summary, this dissertation explores the selection, estimation, and applicability of vascular functions in acute ischemic stroke, contributing to the development of more robust and reliable CTP software packages.



Zusammenfassung

Die CT-Perfusionsbildgebung (CTP) ist bei einem ischämischen Schlaganfall entscheidend, um die Viabilität des Hirngewebes zu beurteilen und Reperfusionstherapien anzuleiten. Dekonvolutionsalgorithmen ermöglichen die Quantifizierung des Hirnperfusionsstatus und die Identifizierung der Penumbra und des irreversibel beschädigten Kerns. Die Genauigkeit der Dekonvolutionsalgorithmen ist jedoch erheblich von der Auswahl zweier Eingangsparameter abhängig: der arteriellen Eingangsfunktion (AIF) und der venösen Ausgangsfunktion. Die präzise Auswahl dieser Parameter ist somit entscheidend, um die genaue Quantifizierung der Perfusionsmetrik und eine zuverlässige Entscheidungsfindung für die Behandlung sicherzustellen. Diese These untersucht die Charakterisierung vaskulärer Funktionen bei einem akuten ischämischen Schlaganfall unter Verwendung datengestützter Maschinenlernmodelle. Es werden drei Fragen (F1, F2, F3) adressiert. F1 hinterfragt zunächst das Konzept der „optimalen“ AIF in der Perfusionsbildgebung, und ermittelt dies mithilfe von Methoden aus dem Bereich des maschinellen Lernens. Ein differenzierbarer Dekonvolutionsalgorithmus ist in ein Deep-Learning-Framework integriert, um eine Echtzeit-Einschätzung von Perfusions-Biomarkern zu ermöglichen, welche den Trainingsprozess verbessern. Die vorgeschlagenen Modelle zeigen eine überlegene Leistung im Vergleich zu manuell oder klinisch ausgewählten AIFs bei der Segmentierung von Kerngewebe. F2 visiert daraufhin die Entwicklung von Maschinenlernmodellen an, die vaskuläre Funktionen von klinischen und realen CTP-Scans selektieren können. Es wird ein neuartiger Ansatz mit der Bezeichnung AIFNet vorgestellt, der gefaltete neuronale Netzwerke für die genaue Identifizierung vaskulärer Funktionen von klinischen CTP-Scans nutzt. Die Methode kann „Ende-zu-Ende“ trainiert werden und lässt sich mit wenigen Ergänzungen ohne Weiteres anpassen. AIFNet erzielt überdies bei der Selektion vaskulärer Funktionen im Rahmen verschiedener CTP-Daten und -Qualitätsszenarien eine Leistung, die mit der von Sachverständigen vergleichbar ist. F3 widmet sich schließlich der Gestaltung von Maschinenlernmodellen, die Informationen von vaskulären Funktionen nutzen, um unzuverlässige CTP-Scans zu detektieren, die von Trunktionsartefakten beeinträchtigt sind. Die Methode übertrifft die Leistung der vorgeschlagenen Baseline, die sich allein auf die Scandauer stützt, sodass die Interpretierbarkeit der CTP-Analyse verbessert wird und Fachkräfte im Bereich der

Neurointervention sicherere und zuverlässigere Behandlungsentscheidungen treffen können. Zusammenfassend erforscht diese Dissertation die Auswahl, Einschätzung und Anwendbarkeit vaskulärer Funktionen bei einem akuten ischämischen Schlaganfall, um zur Entwicklung robusterer und zuverlässigerer CTP-Softwarepakete beizutragen.



Acknowledgements

My passion for science dates back to my early childhood, where I was extremely curious about the world around us. Any object that could be explored attracted my attention, although I was more adept at the “opening” phase than the “rebuilding” one. The pursuit of doctoral studies completes, in a way, a long-awaited cycle that began in my childhood, leading me towards what can be called “being a researcher”. The journey towards my doctoral studies was rather convoluted than straightforward, and at every step, there were always supportive and helpful individuals who helped me achieve my dream.

I want to begin by expressing my gratitude to my supervisors at icometrix, who provided step-by-step support in every project undertaken during this thesis. We shared numerous ideas, discussions, and collaborations that helped me grow as a curious mind and develop a critical perspective on research. Diana, I am grateful to have had your close support throughout these doctoral years, from the very first day I stepped into icometrix. I have learned so much from you, both academically and in terms of research, as well as personal soft skills. Thank you for always being there when I needed it. David, our paths crossed in icometrix when I was feeling lost and confused about my doctoral career, and I am immensely grateful to you for helping me find a path in my PhD. This thesis bears your imprint, your ideas, and our discussions, as you introduced me to the fascinating world of perfusion imaging. I am grateful and proud to have had you by my side during these past years.

I would like to continue by thanking my supervisors at the Technical University of Munich, Bjoern and Jan. You have taught me that discussing science means discussing freedom, and for that, I am truly grateful. I always felt that my research ideas were challenged and nurtured, rather than discouraged or dismissed. Thank you for all these years of learning, discussions, collaborations, and opportunities. It has been incredible to share this journey with both of you.

Next, I want to express my appreciation to my colleagues and friends from Leuven. Roberto and Vince, we shared countless moments within and outside icometrix, enjoying dinners, cycling, and indulging in pintjes and fritjes. I am grateful to both of you for being there for me during good and bad times, listening to my complaints,

and celebrating my happy moments. I will never forget our time together in Leuven. Vince, thank you for supporting my craziness during our days as flatmates. It was a wonderful and fun time! Jaime, Ines, and Patri, my time in Leuven would not have been the same without you all. Thank you for being there when I needed it and for sharing so many memorable moments with me. Additionally, I want to thank all my icometrix colleagues who helped me grow, learn, and have fun throughout my doctoral studies. A special mention goes to Thijs: Thank you for being so open and welcoming during my first days at icometrix. You helped me integrate into the work environment, and I will never forget that.

I would also like to thank my dearest Argentinean friends. Herni, thank you for being there for me all these years while I am in Belgium (and even before that). These “European times” have brought us closer, sharing special moments while far from our families during festivities and Christmas, and providing support to each other when life wasn’t so cheerful. Thank you for the fun times we’ve had together in Poland, Germany, Belgium, and Spain. Next, I want to thank Gato, Maki, and Lucerito. I am fortunate to have you in my life, sharing important life moments both remotely and when I am back in Argentina. You are special, lovely people, and I’m grateful for that.

These doctoral years also brought many interesting and nice people, colleagues, and friends into my life. Among them, I would like to thank my TRABIT colleagues with whom we shared several trainings, seminars, and after-office hours. Thanks to Supro, Ivan, Ines, Luca, France, Stefano, Carmen, Sveinn, Daniel, Andrey, Amnah, Athena, Lucas, Thomas, and Florian. A special mention goes to Luca. I was fortunate to share not only our friendship but also our doctoral path in TRABIT. We had enough time for serious professional meetings and even more time to travel and have fun. Thanks for being there for me.

I would like to express deep gratitude to the person who accompanies me in this amazing journey called life. Thanks, Angie, for supporting me 24/7, for making me smile every day, and for helping me achieve my career goals. You are amazing. Thanks for filling in with colors this important part of my life. I love you.

Por último, mis mayores agradecimientos van para mi familia. Esta tesis está dedicada a ustedes. Gracias por el apoyo incondicional a más de 10 mil km de distancia, por ser mi sostén cuando las cosas no salían como debían. Gracias mamá y papá por haberme brindado educación y por haberme permitido estudiar; por haberme inculcado valores y por haberme dejado ser libre en mis pensamientos y sueños. A mis hermanos, Mati y Juan, gracias por haber estado siempre presentes y cercanos cuando los necesité; mis primeros pasos en el mundo de la “ciencia” los compartí

con ustedes, desarmado cosas, volviendo a armarlas (o al menos intentándolo), o simplemente explorando el mundo. Son mi motivación. A mis sobrinos, Martín y Mateo, que con solo verlos crecer son parte de este aliento a la distancia. Juani, esta tesis también está dedicada a vos. Desde donde estés, sé que estás inclinando la balanza de la vida a mi favor. Gracias!



Contents

| | |
|--|-------------|
| Abstract | iii |
| Zusammenfassung | v |
| Acknowledgements | vii |
| Contents | xi |
| List of Figures | xv |
| Publication List | xvii |
| | |
| I INTRODUCTION | 1 |
| | |
| 1 Foreword | 3 |
| | |
| 2 Background | 7 |
| 2.1 Stroke | 8 |
| 2.1.1 Reperfusion Therapies | 8 |
| 2.1.2 Time is Brain | 9 |
| 2.2 Imaging modalities in acute stroke | 12 |
| 2.2.1 Perfusion imaging | 13 |
| 2.2.2 Quantification of brain tissue status | 15 |
| 2.2.3 The arterial input and the venous output functions | 17 |
| 2.2.4 Quantification of stroke lesions | 19 |
| | |
| 3 Methodology | 21 |
| 3.1 Neural networks back in history | 22 |
| 3.2 Learning paradigms | 23 |
| 3.3 Building deep learning algorithms | 24 |

| | | |
|-------------------------------|--|------------|
| 3.3.1 | Activation functions | 25 |
| 3.3.2 | Model optimization | 26 |
| 3.3.3 | Backpropagation | 27 |
| 3.3.4 | Supervised learning: Regression | 28 |
| 3.3.5 | Supervised learning: Classification | 29 |
| 3.3.6 | Supervised learning: Segmentation | 30 |
| 3.3.7 | Convolutional Neural Networks | 30 |
| 3.4 | Learning from perfusion imaging data | 31 |
| 3.4.1 | Data challenges | 31 |
| 3.4.2 | Selecting vascular functions with machine-learning | 34 |
| II Contributions | | 37 |
| 4 | Summary of Contributions | 39 |
| 5 | Publications | 43 |
| 5.1 | Differentiable Deconvolution for Improved Stroke Perfusion Analysis | 44 |
| 5.2 | AIFNet: Automatic Vascular Function Estimation for Perfusion Analysis Using Deep Learning | 56 |
| 5.3 | Detecting CTP Truncation Artifacts in Acute Stroke Imaging from the Arterial Input and the Vascular Output Functions | 72 |
| III CONCLUDING REMARKS | | 91 |
| 6 | Discussion | 93 |
| 7 | Outlook | 99 |
| Bibliography | | 101 |
| IV APPENDICES | | 111 |
| A | Supplementary Material: Differentiable Deconvolution for Improved Stroke Perfusion Analysis | 113 |

B Supplementary Material: AIFNet: Automatic vascular function estimation for perfusion analysis using deep learning 115

C Supplementary Material: Detecting CTP truncation artifacts in acute stroke imaging from the arterial input and the vascular output functions 119

D A Radiomics Approach to Traumatic Brain Injury Prediction in CT Scans 121

E Publisher’s approval and licenses 127



List of Figures

| | | |
|-----|---|----|
| 2.1 | Time-dependent progression of the lesion. Green: hypoperfused (salvageable) tissue. Red: infarcted (irreversibly damaged) tissue. This image was generated using a scan from ISLES'18 [9, 23]. | 10 |
| 2.2 | A commonly used acute stroke CT imaging workflow. | 13 |
| 2.3 | Temporal perfusion profiles within an artery (AIF), a vein (VOF), a healthy brain area, and a necrotic (core) tissue area. | 14 |
| 2.4 | Perfusion parameter maps. The top row shows a healthy subject, while the bottom row shows a subject with acute ischemic stroke. The red arrow indicates abnormal perfusion areas in the pathological brain. The scans used to generate these images belong to the ISLES'18 dataset [9, 23]. . . | 16 |
| 2.5 | Example of a perfusion CT report using the software package icobrain cva . This image was generated using a scan from ISLES'18 [9, 23]. . . . | 19 |
| 3.1 | Types of machine-learning paradigms used in vascular function analysis. | 24 |
| 3.2 | Relevant CNN architectures. | 31 |



Publication List

The following publications constitute the foundations of my *cumulative doctoral thesis*.

- [1] Ezequiel de la Rosa, David Robben, Diana M Sima, Jan S Kirschke, and Bjoern Menze. “Differentiable deconvolution for improved stroke perfusion analysis.” In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VII 23*. Springer. 2020, pp. 593–602.
- [2] Ezequiel de la Rosa, Diana M Sima, Jan S Kirschke, Bjoern Menze, and David Robben. “Detecting CTP truncation artifacts in acute stroke imaging from the arterial input and the vascular output functions.” In: *PloS one* 18.3 (2023), e0283610.
- [3] Ezequiel de la Rosa, Diana M Sima, Bjoern Menze, Jan S Kirschke, and David Robben. “AIFNet: Automatic vascular function estimation for perfusion analysis using deep learning.” In: *Medical Image Analysis* 74 (2021), p. 102211.

The below listed additional works were also co-authored *during my doctoral studies*.

Journal Articles

- [1] Moritz R Hernandez Petzsche, Ezequiel de la Rosa, Uta Hanning, Roland Wiest, Waldo Valenzuela, Mauricio Reyes, Maria Meyer, Sook-Lei Liew, Florian Kofler, Ivan Ezhov, et al. “ISLES 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset.” In: *Scientific data* 9.1 (2022), p. 762.
- [2] Sook-Lei Liew, Bethany P Lo, Miranda R Donnelly, Artemis Zavaliangos-Petropulu, Jessica N Jeong, Giuseppe Barisano, Alexandre Hutton, Julia P Simon, Julia M Juliano, Anisha Suri, (...), Ezequiel de la Rosa, et al. “A large, curated, open-source stroke neuroimaging dataset to improve lesion segmentation algorithms.” In: *Scientific data* 9.1 (2022), p. 320.
- [3] Maria Ines Meyer, Ezequiel de la Rosa, Nuno Pedrosa de Barros, Roberto Paoletta, Koen Van Leemput, and Diana M Sima. “A contrast augmentation approach to improve multi-scanner generalization in MRI.” In: *Frontiers in Neuroscience* 15 (2021), p. 708196.
- [4] Mladen Rakić, Sophie Vercruyssen, Simon Van Eyndhoven, Ezequiel de la Rosa, Saurabh Jain, Sabine Van Huffel, Frederik Maes, Dirk Smeets, and Diana M Sima. “icobrain ms 5.1: Combining unsupervised and supervised approaches for improving the detection of multiple sclerosis lesions.” In: *NeuroImage: Clinical* 31 (2021), p. 102707.
- [5] Thijs Vande Vyvere, Ezequiel de la Rosa, Guido Wilms, Daan Nieboer, Ewout Steyerberg, Andrew IR Maas, Jan Verheyden, Luc Van den Hauwe, Paul M Parizel, CENTER-TBI Participants, and Investigators. “Prognostic validation of the NINDS common data elements for the radiologic reporting of acute traumatic brain injuries: a CENTER-TBI study.” In: *Journal of Neurotrauma* 37.11 (2020), pp. 1269–1282.
- [6] Mandy Melissa Jane Wittens, Diana Maria Sima, Ruben Houbrechts, Annemie Ribbens, Ellis Niemantsverdriet, Erik Fransen, Christine Bastin, Florence Benoit, Bruno Bergmans, Jean-Christophe Bier, (...), Ezequiel de la Rosa, et al. “Diagnostic performance of automated MRI volumetry by icobrain dm for Alzheimer’s disease in a clinical setting: a REMEMBER study.” In: *Journal of Alzheimer’s Disease* 83.2 (2021), pp. 623–639.

Conference Articles

- [1] Ezequiel de la Rosa, Diana M Sima, Thijs Vande Vyvere, Jan S Kirschke, and Bjoern Menze. “A radiomics approach to traumatic brain injury prediction in CT scans.” In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE. 2019, pp. 732–735.
- [2] Maria Ines Meyer, Ezequiel de la Rosa, Nuno Barros, Roberto Paoella, Koen Van Leemput, and Diana M Sima. “An augmentation strategy to mimic multi-scanner variability in MRI.” In: *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2021, pp. 1196–1200.
- [3] Maria Ines Meyer, Ezequiel de la Rosa, Koen Van Leemput, and Diana M Sima. “Relevance vector machines for harmonization of MRI brain volumes using image descriptors.” In: *OR 2.0 Context-Aware Operating Theaters and Machine Learning in Clinical Neuroimaging: Second International Workshop, OR 2.0 2019, and Second International Workshop, MLCN 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13 and 17, 2019, Proceedings*. Springer. 2019, pp. 77–85.
- [4] Jaime Simarro Viana, Ezequiel de la Rosa, Thijs Vande Vyvere, David Robben, Diana M Sima, and CENTER-TBI Participants and Investigators. “Unsupervised 3D brain anomaly detection.” In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part I*. Springer. 2021, pp. 133–142.
- [5] Jana Vujadinovic, Jaime Simarro Viana, Ezequiel de la Rosa, Els Ortibus, and Diana M Sima. “Knowledge-Guided Segmentation of Isointense Infant Brain.” In: *Perinatal, Preterm and Paediatric Image Analysis: 7th International Workshop, PIPPI 2022, Held in Conjunction with MICCAI 2022, Singapore, September 18, 2022, Proceedings*. Springer. 2022, pp. 105–114.

Conference Abstracts

- [1] Ezequiel de la Rosa, Diana M Sima, Melissa Verheijen, Dirk Smeets, and David Robben. “Detecting Truncation Artifacts in CT Perfusion Affecting Hypoperfused and Core Tissue Volumes in Acute Stroke.” 7th European Stroke Organization conference, September 1-3. 2021.
- [2] Maria Ines Meyer, Ezequiel de la Rosa, Greg Dudra, Eline Van Vlierberghe, Koen Van Leemput, and Diana M Sima. “Harmonization of Brain Volumes from MRI Images of MS Patients Using Image Descriptors.” In: *2019 Annual Meeting of the Consortium of Multiple Sclerosis Centers*. CMSC. 2019.
- [3] Roberto Paoella, Ezequiel de la Rosa, Vincenzo Anania, Stijn Denissen, Jeroen Van Schependom, Guy Nagels, Dominique Dive, Françoise Durand-Dubief, Dominique Sappey-Marinier, Dirk Smeets, et al. “Decoding the EDSS scores of multiple sclerosis patients from MRI biomarkers.” In: *Multiple Sclerosis Journal* 26.S3 (2020), pp. 119–120.
- [4] Mladen Rakic, Sophie Vercruyssen, Simon Van Eyndhoven, Ezequiel de la Rosa, Dirk Smeets, and Diana M Sima. “icobrain ms 5.1: A combined approach for improving the detection of multiple sclerosis lesions.” European Society of Neuroradiology (ESNR) Annual Meeting. 2021.
- [5] Mandy MJ Wittens, Ezequiel de la Rosa, Diana M Sima, Wim Van Hecke, Dirk Smeets, and Sebastiaan Engelborghs. “Reliable MRI volumetry for Alzheimer’s disease: diagnostic performance of icobrain dm on real-world data.” Computational approaches for ageing and age-related diseases (CompAge). 2020.

PART I

INTRODUCTION



Foreword

In recent years, the medical imaging field has witnessed remarkable advancements through the integration of machine-learning techniques. These cutting-edge methodologies have revolutionized the analysis and interpretation of medical scans, leading to improved diagnosis, treatment, and patient outcomes. Applications of these techniques are diverse, targeting a wide range of anatomical regions, organs, and pathologies. Machine-learning developments have significantly impacted most medical imaging modalities used in healthcare centers.

Stroke remains a leading cause of mortality and morbidity worldwide, necessitating urgent and accurate diagnosis for optimal treatment. Traditional methods of interpreting stroke images heavily rely on the radiologist's manual assessment, interaction, and quality control, which are time-consuming, subjective, and prone to errors. Machine-learning algorithms have introduced a new paradigm in stroke imaging analysis, enabling automated, objective, and quantitative evaluation with unprecedented precision.

Machine-learning techniques have shown exceptional capabilities for extracting complex patterns and features from data, even under scenarios with high data variability. The characterization of data through these techniques is informative and rich, leading to successful solutions for a wide range of tasks, including data classification, regression, and segmentation. In the field of medical imaging, convolutional neural networks (CNNs) have emerged as notable techniques, leveraging their hierarchical feature learning and pattern recognition abilities. By training CNNs on large amounts of medical data, researchers have achieved remarkable results on different tasks relevant to the clinical stroke workflow. For example, stroke lesion segmentation has been investigated in different imaging modalities such as MRI [34], non-contrast CT [45], and perfusion CT [23], considering various phases of the pathology, including acute [23, 45], sub-acute [34], and chronic [32]. Additionally, machine learning-based prediction of clinical outcomes in acute ischemic stroke has been explored, aiding in treatment decision-making and patient stratification [66].

During the acute stage of stroke, perfusion imaging plays a vital role in assessing

the perfusion status of brain tissue. This imaging technique enables the quantification of perfusion biomarkers, which help estimate brain abnormalities. The primary objective for neurointerventionalists is to determine the presence of salvageable brain tissue and identify the appropriate treatment options for the patient.

In the literature, several computational methods, commonly known as “deconvolution” models, are available to quantify the brain perfusion status. These techniques rely on perfusion data parameters that serve as reference information to solve a challenging and mathematically ill-posed problem. The two parameters of interest are the arterial input function (AIF) and the vascular output function (VOF).

In this doctoral dissertation, I develop data-driven methodologies for studying, selecting, and characterizing the AIF and VOF parameters. One might argue that, given the widespread use of machine-learning in various tasks, applying these techniques to the problem at hand should be a straightforward and quickly solvable task. Unfortunately, this ideal scenario is far from reality for two main reasons. The first reason is the inherent complexity and intricacy of the perfusion imaging modality and the post-processing algorithms required to make the images interpretable. This complexity arises from several factors, such as the 4D nature of the modality, the significant image variability among scanners and protocols used in different healthcare centers, and the diverse anatomical and physiological characteristics of patients’ brains, among others. The second reason that adds to the complexity of this problem is the high sensitivity of deconvolution algorithms to variations in vascular functions.

Thesis organization

The thesis is organized into three sections, each divided into chapters. These sections are:

- Section I: *Introduction*
- Section II: *Contributions*
- Section III: *Concluding Remarks*

Section I - Chapter 2 provides an overview of the clinical background of the field under study, focusing on *stroke* as a neurovascular disorder and its associated imaging techniques. In Section I - Chapter 3, I describe the machine-learning and deep-learning techniques that serve as the foundations for the research experiments conducted within this thesis.

Section II: *Contributions* consists of two chapters. In Chapter 4, I present the goals pursued within this thesis, including the research questions being addressed. The most significant contributions of my work are discussed in this chapter. Chapter 5 includes the publications produced during this thesis, which are self-contained.

Section III: *Concluding Remarks* is divided into two chapters. In Chapter 6, I discuss the results and outputs of this thesis, as well as their future perspectives and ongoing challenges. In Chapter 7, I present the conclusions of this thesis.

Additionally, there is an Appendix (Part IV) that includes supplementary material supporting the research publications of Chapter 5. The appendix also contains a peer-reviewed conference publication that, while not directly related to the goals of this thesis, contains findings of interest for the neuroimaging community.



Background

Over thousands of years of evolution, our brain has developed a remarkable vasculature that is finely tuned to maintain its health and functionality, allowing for optimal functioning of the organ. The brain's circulatory system is a complex structure consisting of a vast network of arteries, veins, and capillaries. This network serves to transport essential nutrients, such as glucose and oxygen to every corner of the organ. One of the most interesting capabilities of this intricate system is its fail-safe mechanism, which guarantees nutrient delivery even in scenarios where blood flow may be compromised. In other words, nature has equipped our brains with a backup plan in the form of *collateral* circulation. Collateral arteries act as alternative routes, redirecting blood to affected brain areas and ensuring their continued nourishment. Thanks to these biological, physiological, and anatomical adaptations, our brain has become remarkably resilient and adaptive to a wide range of circumstances.

In this chapter, I will present a pathology that is highly skilled at sabotaging the brain's network vasculature goals by hindering the delivery of nutrients and oxygen to the entire organ: *ischemic stroke*. The purpose of this chapter goes beyond understanding the pathology itself and aims to describe how we can utilize medical imaging technologies for optimal treatment decision-making. I will explore brain perfusion imaging, the algorithms involved in handling these images, and finally, I will demonstrate how we can estimate ischemic stroke lesions from these images.

2.1 Stroke

Stroke is a cerebrovascular event characterized by the sudden interruption or significant reduction of blood flow to a specific brain region, resulting in insufficient supply of oxygen and nutrients to the affected tissue. There are two types of stroke: hemorrhagic and ischemic. Among these, ischemic stroke is the most prevalent, accounting for 87% of all stroke cases [63]. Ischemic strokes are generally caused by the blockage or occlusion of a blood vessel within the brain, commonly caused by a thrombus (i.e., blood clot) or an embolus (dislodged clot or other material). The restricted blood flow caused by the occlusion triggers a cascade of molecular and cellular events, including major pathogenic mechanisms such as excitotoxicity, peri-infarct depolarizations, inflammation, and programmed cell death [14]. As a result of these pathophysiological conditions, brain tissue becomes compromised, ultimately leading to tissue death (i.e., infarction).

At the cellular level, the ischemic cascade begins with a reduction in oxygen and glucose supply, leading to a decrease in adenosine triphosphate production and subsequent failure of energy-dependent cellular processes. The lack of energy compromises neuronal function and triggers excitotoxicity, a process characterized by the over-activation of glutamate receptors that leads to dendrite degeneration and later cell death [36]. Excitotoxicity results in an excessive influx of calcium ions into neurons, activating various enzymes that promote cell damage, including proteases and lipases. These enzymes contribute to the breakdown of cellular components and membrane damage, exacerbating oxidative stress and further cellular injury [14].

The underlying risk factors for acute ischemic stroke include high blood pressure, diabetes, smoking, obesity, and high cholesterol [65]. These risk factors can contribute to the development of vascular pathologies, such as atherosclerotic plaques or the formation of blood clots, which can eventually obstruct cerebral arteries and trigger an ischemic stroke event. Acute ischemic stroke affects approximately 700,000 people per year and is among the leading causes of death and disability in the United States [65]. According to [63], a person in the United States suffers a stroke attack, on average, every 40 seconds.

2.1.1 Reperfusion Therapies

Thrombolysis and *mechanical thrombectomy* are medical interventions aimed at restoring blood flow to the brain by removing or dissolving a blood clot that is obstructing a brain vessel.

Thrombolysis, also known as intravenous thrombolysis or systemic thrombolysis, involves the administration of a medication as a thrombolytic agent, such as tissue plasminogen activator (tPA or *alteplase*). The medication is typically given intravenously within a few hours of the onset of stroke symptoms. Thrombolytic agents work by breaking down the blood clot, promoting the restoration of blood flow in the affected blood vessel. Besides alteplase, another prominent thrombolytic drug is *tenekteplase*. According to the European Stroke Organization guidelines, tenekteplase is preferred over alteplase when considering times from stroke onset less than 4.5 hours and when the patient is eligible for mechanical thrombectomy [6]. It is important to note that thrombolysis carries some risks, including an increased risk of bleeding, and it may not be suitable for all patients, depending on several factors such as the time from symptom onset and the patient's age.

Mechanical thrombectomy, on the other hand, consists of an invasive procedure in which a specialized catheter or stent retriever is used to physically remove the blood clot from the occluded vessel. The catheter is typically introduced through the patient's groin or arm and placed in the occluded artery using X-ray imaging. Typically, thrombectomy is reserved for patients with large vessel occlusions and can be effective within a certain time window, usually up to 6 to 24 hours from symptom onset under certain clinical conditions. Mechanical thrombectomy has been shown to be a highly effective treatment and has significantly improved outcomes for eligible patient populations [41, 1].

In previous attempts, a third method for acute ischemic stroke therapy involved trying to hinder the progression of the ischemic cascade, referred to in the literature as *neuroprotection*. Neuroprotection aims to protect against neuronal damage under different conditions. Although showing efficacy in experimental studies, neuroprotection has proven to be unsuccessful in several clinical trials [71, 52].

2.1.2 Time is Brain

The concept of “time is brain” was introduced by Dr. Camilo Gomez in 1993 and emphasizes the urgency of timely intervention in acute stroke. Early recognition of stroke symptoms, rapid activation of emergency medical services, and prompt access to specialized stroke care are essential to minimize brain damage, preserve neuronal function, and improve patient prognosis.

In acute ischemic stroke, the rapid restoration of blood flow to the affected brain region is crucial to limit the extent of ischemic damage and preserve neurological function. Time plays a critical role in determining the eligibility and effectiveness

2. BACKGROUND

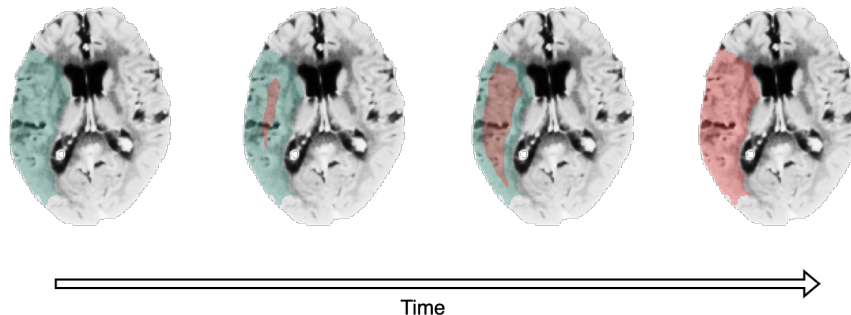


Figure 2.1: Time-dependent progression of the lesion. Green: hypoperfused (salvageable) tissue. Red: infarcted (irreversibly damaged) tissue. This image was generated using a scan from ISLES'18 [9, 23].

of various treatment options, such as thrombolysis and mechanical thrombectomy. The longer the wait for treating the patient, the larger the portion of irreversibly damaged brain tissue. In other words, timely intervention in acute ischemic stroke is the cornerstone for limiting neuronal damage in the hypoperfused ischemic penumbra [21]. A pictorial visualization of time-dependent lesion progression is shown in Figure 2.1. It is worth mentioning that the rate of infarct progression is influenced by several factors, such as the extent of collateral arterial circulation, duration of the insult, and the functional and metabolic state of the cells [5]. All these factors collectively impact the effectiveness of therapeutic interventions commonly used for acute ischemic stroke [40].

Intravenous thrombolysis with tPA is a time-sensitive therapy used in the treatment of acute ischemic stroke. Several clinical trials have demonstrated the benefits of intravenous tPA when administered within a specific time window after symptom onset and for certain patients [39, 60]. The European Stroke Organization (ESO) guidelines consider the administration of alteplase under different scenarios, including known or unknown time from stroke onset, as well as different brain imaging criteria. A summary of these recommendations is shown in Table 2.1. It can be observed that the success of thrombolysis treatment is time-dependent, and the benefit of this reperfusion therapy diminishes over time [15]. Therefore, timely recognition of stroke symptoms, prompt activation of the emergency response, and treatment of the patient are crucial for increasing the chances of a better patient outcome and reducing the potential disability resulting from the ischemic attack.

| Time from onset | Imaging criteria | Recommendation |
|--------------------------------|--------------------------------------|--|
| Intravenous Thrombolysis (IVT) | | |
| 0-4.5 hs | plain CT | IVT with alteplase |
| 4.5-9 hs | plain CT | No IVT |
| 4.5-9 hs | perfusion mismatch (CT/MR) | IVT with alteplase when thrombectomy is not indicated or planned |
| Wake-up stroke / unknown onset | DWI-FLAIR mismatch (MR) | IVT with alteplase when thrombectomy is not indicated or planned |
| Mechanical thrombectomy (MT) | | |
| 0-6 hs | * | MT + BMM with anterior circulation LVO |
| 6-24 hs from last known well | Fulfilling DAWN [41] or DEFUSE-3 [1] | MT + BMM with anterior circulation LVO |

Table 2.1: ESO recommendations guidelines for thrombolysis and mechanical thrombectomy treatment [6, 61]. BMM: Best medical management. LVO: Large vessel occlusion. *Although not explicitly mentioned in the ESO guidelines [61], ASPECTS (Alberta Stroke Program Early CT Score) scoring using plain CT has been utilized in clinical trials [22].

Like thrombolysis, mechanical thrombectomy is also a time-sensitive intervention used for large vessel occlusions in acute ischemic stroke. The DEFUSE-3 clinical trial has demonstrated that the combination of endovascular thrombectomy and standard medical therapy provided superior functional outcomes compared to standard medical therapy alone when performed within 6 to 16 hours after the patient was last known to be well [1]. The DAWN clinical trial extended the time window for thrombectomy to 24 hours in patients who exhibited a mismatch between their clinical deficit and infarct. It demonstrated that thrombectomy, when combined with standard care, led to less disability at 90-day outcomes in comparison to standard care alone [41]. A

summary of the ESO recommended time windows for the treatment of patients with mechanical thrombectomy is shown in Table 2.1.

It is worth mentioning that the concept of “time is brain” extends beyond reperfusion therapies. Early assessment, diagnosis, and management of stroke, including supportive care and secondary prevention measures, are also critical for optimal outcomes. Rapid identification of stroke symptoms, initiation of stroke protocols, and timely access to stroke expertise through telemedicine or regional stroke centers contribute to reducing treatment delays and improving patient outcomes.

2.2 Imaging modalities in acute stroke

Imaging of the brain status in acute stroke is performed with computed tomography (CT) and/or magnetic resonance imaging (MRI). Choosing one modality over the other depends mostly on scanner availability and also on its affordability by the country’s healthcare system. Considering this *availability-affordability* criterion, CT is a more widespread modality than MRI in the clinical setting, and therefore, most acute stroke research and clinical trials have preferred it over MRI. For the aims of this thesis, imaging of acute stroke will be described in the context of CT. Nonetheless, most of the concepts considered in this chapter completely (or at least partially) translate to MRI. Please note that perfusion CT also possesses technical advantages over perfusion MRI. Firstly, it is less susceptible to motion artifacts. Secondly, patient eligibility is simplified as there is no need to check for metallic implants or gadolinium allergies. Thirdly, perfusion CT benefits from a linear relationship between contrast agent concentration and brain tissue attenuation, making perfusion measurements easier to obtain. Perfusion MRI, instead, relies on a *non-linear* relationship between the contrast agent concentration and the induced T2* effect. Consequently, perfusion MRI requires heavier and more complex data preprocessing in order to convert the measured signal intensities into contrast agent concentrations.

Imaging in acute ischemic stroke aims to answer, at least, the following four major questions [27]:

1. Is there hemorrhage?
2. Is there a thrombus that can be targeted through reperfusion therapies?
3. Is there salvageable tissue (a.k.a. *penumbra*)?
4. Is there irreversibly damaged, infarcted tissue (a.k.a. *core*)?

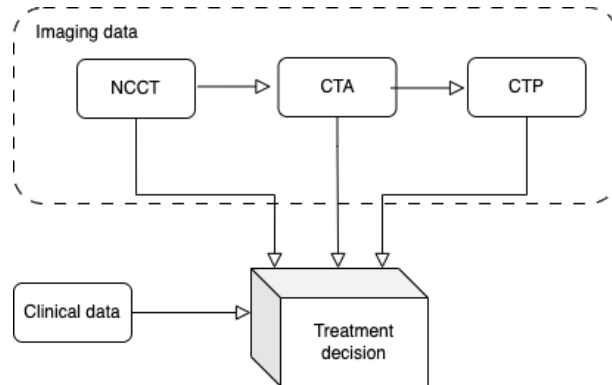


Figure 2.2: A commonly used acute stroke CT imaging workflow.

While the first question can be addressed through non-contrast-enhanced CT (NCCT), insights about the second question can be obtained through CT angiography (CTA). The last two questions, instead, can be answered through a single imaging modality which is perfusion CT (CTP). Please note that the *order* in which the scans are performed matters: under the presence of a hemorrhagic stroke, there is no longer a need to answer questions 2-4, and as such, CTA and CTP could be avoided. Figure 2.2 shows the acute stroke imaging workflow in the clinical setting.

2.2.1 Perfusion imaging

The beginnings of perfusion imaging date back to the early 80s with experiments using CT scanners by Leon Axel [4, 42]. At that time, CT scanners were technically much more limited than current machines, particularly in terms of speed and anatomical coverage. Despite these constraints, the experiments conducted by Axel and his team were groundbreaking as they provided the first insights into hemodynamics at the capillary level, significantly expanding the usage of CT machines. Within a short period of time, these perfusion concepts were translated to MRI technology.

Perfusion imaging in acute stroke aims to quantify blood flow at the brain tissue (capillary) level. Image acquisition is achieved through serial brain CT scanning after intravenous administration of an iodinated contrast bolus. This contrast bolus is intravenously injected into the patient over a period of 7-10 seconds, and image acquisition is later continued for approximately 60-70 seconds. As a result, a temporal sequence of 3D brain volumes is obtained, commonly referred to in the literature as a (spatio-temporal) 4D imaging modality. Figure 2.3 shows the temporal perfusion

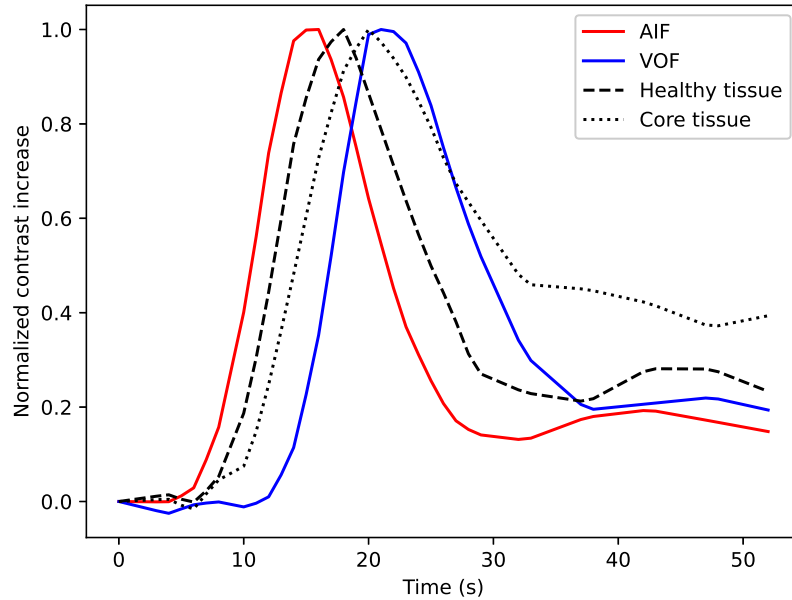


Figure 2.3: Temporal perfusion profiles within an artery (AIF), a vein (VOF), a healthy brain area, and a necrotic (core) tissue area.

profiles over different brain tissue areas, including an artery, a vein, a healthy brain area, and a necrotic brain area.

Current guidelines for CTP acquisition in the context of acute ischemic stroke recommend [10]:

- i) Scan order: CTP after CTA (with a 60-second delay in between)
- ii) Scan duration: 60-70 seconds, with 5-10 seconds of pre-contrast acquisition
- iii) Contrast injection: 40 mL of contrast at 4-6 mL/s, followed by 40 mL of saline at 4-6 mL/s
- iv) Radiation dose: ≥ 300 mGy
- v) Kilovoltage peak: 70-80 kV
- vi) Brain coverage: at least 8 cm of z-axis coverage

vii) Frame rate: 2 seconds or faster

2.2.2 Quantification of brain tissue status

Understanding whether a patient may benefit from recanalization techniques relies on the imaging measurements of the *core-penumbra* volumetric mismatch. Therefore, it is important to understand how the acquired CTP images help us answer this question. It should be noted that the acquired CTP images can be extremely complex and challenging to interpret radiologically due to the 4D nature of the imaging modality. However, the main difficulty in analyzing these 4D images lies in the fact that they are greatly influenced by patient-specific biological and physiological variables, as well as by differences in the contrast injection protocols used in clinical routine. Well-known variables that affect the measured time-attenuation curves include patient size and cardiovascular system dynamics (such as cardiac output), which alter the contrast delivery through the brain tissue [12]. Similarly, specific pathophysiological conditions (e.g., patients with multiple intracranial emboli or intravascular narrowing [35]) may prolong the delivery of the contrast agent in the brain. Therefore, it seems natural to assume that removing the effect of these confounders would make the subsequent image interpretation and quantification much easier.

To address the aforementioned issue, we can extract information about the blood flow in the brain from the CTP data. To this end, we introduce the indicator dilution equation, given by:

$$c_{\text{tissue}}(t) = c_{\text{art}}(t) \otimes h(t) \quad (2.1)$$

where $c_{\text{tissue}}(t)$, $c_{\text{art}}(t)$, and $h(t)$ represent the contrast increase in the tissue voxel under consideration, the contrast increase measured in a large feeding artery (referred to as the *arterial input function* or AIF), and the flow-scaled residue function, respectively. The symbol \otimes denotes the convolutional operator. The goal is, therefore, to estimate perfusion markers that are unaffected or less influenced by confounding factors. In this direction, a solution of the indicator dilution equation for $h(t)$ can be found through deconvolution methods. For the sake of simplicity, this thesis focuses on the most investigated and widely implemented approach in clinical software packages [62]: the *delay-invariant Singular Value Decomposition (SVD) deconvolution*. The SVD-based deconvolution has been validated using experimental [38] and clinical data [1, 41], and it can be considered, according to Fieselmann et al. [17], as the current standard approach for perfusion image analysis. It is important for the reader to note that the conclusions drawn in this thesis are applicable to other deconvolution

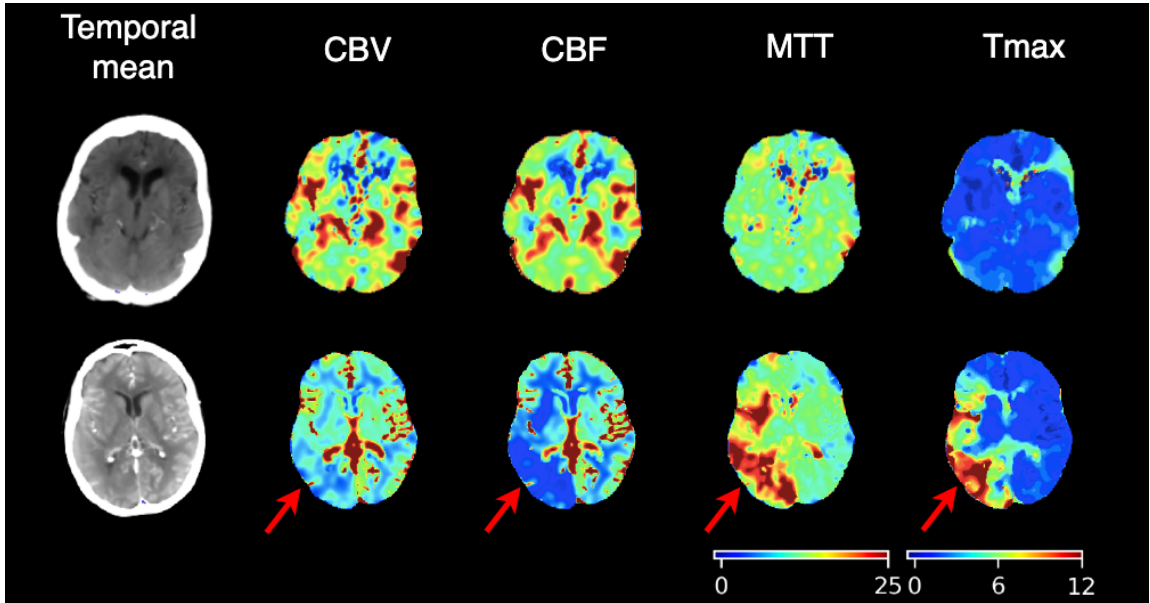


Figure 2.4: Perfusion parameter maps. The top row shows a healthy subject, while the bottom row shows a subject with acute ischemic stroke. The red arrow indicates abnormal perfusion areas in the pathological brain. The scans used to generate these images belong to the ISLES’18 dataset [9, 23].

methods as well, albeit to some extent. Other deconvolution algorithms encompass models assuming specific data distributions (e.g., exponential [30]), models based on the fast Fourier transform [43, 20], or using maximum likelihood expectation maximization [64], among others.

After image deconvolution, it is possible to quantify brain perfusion parameters using a physical model. These parameters, commonly referred to as *perfusion maps* or *parameter maps*, include cerebral blood flow (CBF), cerebral blood volume, mean transit time (MTT), and time-to-maximum to the flow-scaled residue function (Tmax). For example, CBF and Tmax can respectively be calculated as $CBF = \frac{\max(h(t))}{\rho}$ and $Tmax = \operatorname{argmax}(h(t))$, where ρ represents the mean density of the volume under consideration [g/ml]. The mathematical derivation of these perfusion maps is beyond the scope of this thesis; for more details, interested readers can refer to [17]. Other parameter maps, such as time-to-peak (TTP), are directly measured from the time-attenuation curves. Figure 2.4 shows an example of perfusion maps for a healthy brain and a brain affected by acute ischemic stroke due to middle cerebral artery

occlusion.

The implementation of the SVD deconvolution algorithm can be obtained by first discretizing Eq. 2.1. In the experiments conducted in this thesis, the Volterra discretization approach has been used, although different techniques can be used to tackle the problem [57]. After discretization, deconvolution is performed using singular value decomposition. It is important to mention that the problem at hand is ill-posed, meaning that small variations in the input can lead to significant variations in the deconvolution solution of the equation. This sensitivity to input perturbations can be particularly influenced by image quality factors, such as the signal-to-noise ratio. To address the challenges associated with ill-posedness, regularization techniques are employed. These techniques play a crucial role in obtaining physiologically plausible solutions by mitigating issues related to instability and noise amplification [17]. The goal of regularization is to strike a balance between mathematical and physical solutions, ultimately enhancing the robustness of perfusion estimation measurements and ensuring that the results align with expected physiological behaviors. Examples of regularization techniques include the Tikhonov approach and the generalized cross-validation approach [17, 58].

2.2.3 The arterial input and the venous output functions

In this thesis, the term “vascular functions” refers to the arterial input and venous output functions. These functions, particularly the arterial input function (AIF), play a crucial role in deconvolution-based perfusion imaging. The positioning of the AIF within the brain, along with its profile, explicitly shapes the flow-scaled residue function described in Equation 2.1, thereby impacting the quantified perfusion maps and, ultimately, the perfusion lesion volumes. Vascular functions are fundamental to this doctoral dissertation, and I will now proceed to introduce them and describe what they are, why they are important, and how they are used. The AIF is the time-attenuation or contrast concentration curve within an arterial vessel that supplies blood to the brain tissue. It is typically measured in large brain arteries such as the internal carotid or middle cerebral arteries, as they are less prone to partial volume effects. Given its significance in dilution theory and as one of the most crucial input parameters for deconvolution algorithms, AIFs have been extensively studied. Questions have been raised about from which location it should be estimated, whether in the ipsilesional or contralesional brain hemisphere, and whether a single global AIF or multiple local AIFs should be used [7]. To understand the need for using the AIF in deconvolution models, it is important to explain why direct analysis

of the time-attenuation curves is not feasible. The main reason is the presence of confounding factors. As mentioned earlier in Section 2.2.2, the time-attenuation curves measured by the scanner result from complex interactions of various variables, including biological, physiological, and anatomical patient-specific conditions, as well as the contrast injection protocols. In other words, the time-attenuation curves do not reflect the true microvasculature information specific to the tissue. However, measuring the contrast bolus entering the tissue under study (i.e., the AIF) allows us to isolate the true and desired tissue microvasculature status through deconvolution.

Similar to the AIF, the venous output function (VOF) represents the time-attenuation or bolus concentration curve within a large feeding vessel, which in this case is a vein. Thus, unlike the AIF, the VOF represents the bolus concentration leaving the brain tissue, its “output”. Figure 2.3 shows the arterial, venous, healthy tissue, and necrotic tissue time-concentration curves. Note the ordered, sequential timing of the events: the AIF (input to the tissue) reaches its maximum value first, followed by the contrast arrival in the healthy tissue, the delayed contrast arrival in the damaged tissue, and finally, the contrast reaching the vein, which functions as a collector of the tissue’s blood. The VOF, like the AIF, is measured by selecting a large vessel structure such as the superior sagittal sinus.

Up to this point, we have established similarities between the AIF and VOF in terms of their shape, what they represent, and the appropriate vessel to measure them. However, the most important difference between them is their relevance. The VOF is a much less critical parameter than its arterial counterpart in the deconvolution phase. The VOF is primarily used to rescale the perfusion metrics obtained through deconvolution. The idea behind this is that large venous structures, like the sagittal sinus, are usually wider (i.e. have larger diameters) than large arteries. As a result, the X-ray attenuation measured in those voxels is less likely to include non-venous structures, thereby reducing partial volume effects. In short, the VOF is used to rescale the absolute values of cerebral blood flow (CBF). It is important to note that in software packages where hypoperfused and core volumes are measured based on T_{max} and rCBF, the VOF has no influence on the measured volumes. Only in software packages using absolute CBF or CBV values, the selection of the VOF may lead to misleading estimates of perfusion lesions.

2.2. Imaging modalities in acute stroke

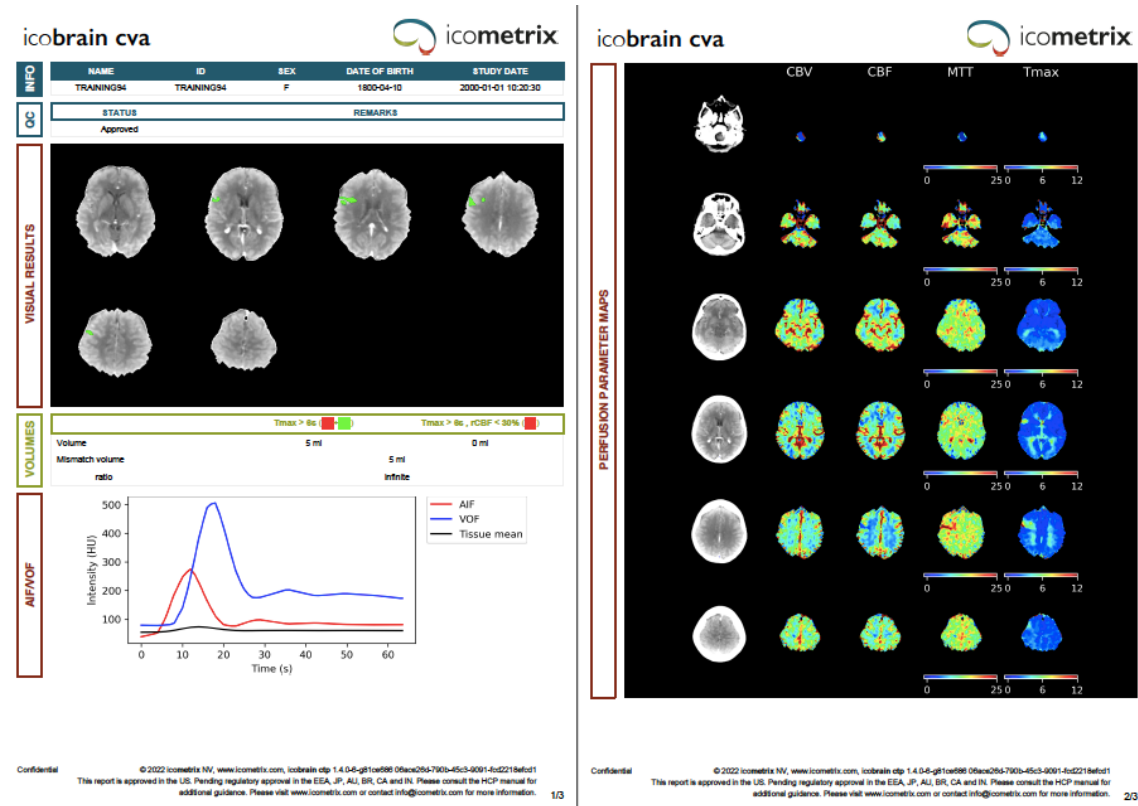


Figure 2.5: Example of a perfusion CT report using the software package *icobrain cva*. This image was generated using a scan from ISLES'18 [9, 23].

2.2.4 Quantification of stroke lesions

Once perfusion parameter maps have been obtained, we can address the key questions of interest: *How much of the brain is already infarcted, and how much can be saved through reperfusion therapies?* To answer these questions, this thesis follows the *conventional* approach used in clinical practice: thresholding of the perfusion maps.

To determine the *core* and *penumbra* volumes, I will introduce a related but slightly different volume of interest: the *hypoperfused* volume. This term refers to the total brain tissue experiencing an ischemic stroke attack, regardless of its tissue status (i.e., including both salvageable and irreversibly damaged tissue). In other words, the hypoperfused volume is the combination of the core and penumbra volumes. Typically, software packages used in clinical practice quantify the hypoperfused and core volumes,

2. BACKGROUND

with the penumbra representing the mismatch between the two. It is important to note that the *perfusion mismatch volume* (difference between the hypoperfused and core volumes) is a crucial volumetric marker for determining a patient’s eligibility for mechanical thrombectomy. Another marker used to identify patients who may benefit from recanalization is the perfusion mismatch ratio (hypoperfused/core), which can range from a unitary value (indicating that the entire hypoperfused volume is already necrotic) to infinity (meaning that the entire hypoperfused volume is potentially salvageable).

Estimating the hypoperfused and core volumes involves thresholding the parameter maps at specific cutoff values. Common parameter maps used in clinical practice for quantifying the *core* are cerebral blood flow (CBF) and cerebral blood volume (CBV). In the DAWN and DEFUSE-3 trials [41, 1], the core was determined using a relative CBF (rCBF) threshold of 30%. It is worth noting that rCBF is the CBF map normalized by healthy tissue perfusion values, which makes the perfusion map cutoffs more generalizable and robust to patient-specific perfusion differences. The hypoperfused volume, on the other hand, can be quantified using Tmax or TTP maps. Following the DEFUSE-3 study, a hypoperfused volume can be obtained by considering Tmax values greater than 6 seconds. Figure 2.5 shows an example of a clinical report generated from perfusion CT using the commercial software package **icobrain cva**. In this report, perfusion volumes are obtained by thresholding the Tmax and rCBF parameter maps at the aforementioned cutoff values. It illustrates various components relevant to radiological reading and quality control of perfusion imaging data, including the parameter maps, overlaid perfusion lesions, markers of interest (absolute perfusion volumes, perfusion mismatch ratio, and perfusion mismatch difference), and the vascular functions used as input for the software.

It is important to consider that thresholds are highly sensitive to the type and implementation of deconvolution algorithms and, therefore, depend on the specific parameter maps and software packages used. The lack of standardization in CTP deconvolution procedures has resulted in the use of various algorithms (not only for deconvolution but also for pre-processing and post-processing of images), different parameter maps, and different cutoff values for determining perfusion volumes. A limitation of this algorithmic variability and lack of standardization is the significant volumetric differences observed when using different software packages, particularly in quantifying the hypoperfused volumes [29, 47].



Methodology

In recent years, machine learning has made significant strides, driven by the abundance of data and improvements in computing power. Within the field of medical imaging, this technology has brought about a revolution in how images are understood and interpreted, greatly benefiting healthcare practitioners in their decision-making process. Machine learning applications have expanded to encompass various organs, pathologies, and imaging techniques, with some algorithms even matching or surpassing the performance of human experts in specific tasks.

This chapter delves into the fundamental concepts and algorithms of machine-learning that underpin the investigation of vascular functions in perfusion imaging. These techniques serve as the bedrock of this doctoral dissertation, as the contributions made in this work rely on these principles. The chapter provides a comprehensive overview of machine-learning algorithms, with a particular emphasis on neural networks and their optimization. Additionally, it sheds light on the challenges encountered when dealing with the intricacies and computational demands associated with 4D imaging data, such as those found in perfusion imaging.

3.1 Neural networks back in history

Machine-learning dates back to the 1940s and 1950s when researchers began exploring the concept of creating machines that could learn from data. During this time, Warren McCulloch and Walter Pitts introduced the concept of artificial neurons, which formed the foundation for neural networks. These artificial neurons aimed to mimic the behavior of biological neurons, enabling information processing and learning.

One significant development in the field of neural networks came in the late 1950s with the introduction of the perceptron by Frank Rosenblatt [51]. The perceptron was an early form of artificial neural network capable of learning from labeled data. It showed promise in pattern recognition tasks but had limitations due to its single-layer structure, restricting its ability to solve complex problems.

The field of artificial intelligence experienced a downturn in the 1970s, often referred to as the (first) “AI Winter”, which had an impact on neural network research. In the 1980s, the earlier introduced *backpropagation* algorithm gained attention. The algorithm allows efficiently training multi-layer neural networks by propagating errors backward through the network (see section 3.3.3), thus, enabling the training of deeper architectures, addressing the limitations of perceptrons, and sparking renewed interest in neural networks.

The 1990s witnessed the rise of neural networks, with applications emerging in various domains such as image and speech recognition. Researchers developed new types of neural networks, including radial basis function networks, recurrent neural networks, and convolutional neural networks (CNNs). These architectures expanded the capabilities of neural networks and paved the way for their successful application in different fields.

The 2010s marked a significant turning point for neural networks with the advent of deep-learning. Deep-learning leverages deep neural networks with many layers, enabling the extraction of hierarchical representations from large datasets. Breakthroughs in deep-learning, such as the ImageNet competition won by AlexNet in 2012, showcased the superior performance of deep neural networks in image classification tasks [13, 28]. This triggered a revolution in various domains, including computer vision, medical image analysis, natural language processing, etc.

Current goals in the field revolve around advancing the capabilities and understanding of deep neural networks. Researchers aim to improve model performance and generalization, reduce training time and resource requirements, and develop techniques for handling real-world challenges such as small datasets, model interpretability,

and adversarial attacks. One trend is the exploration of novel network architectures, including attention mechanisms, transformers, and graph neural networks, to tackle specific tasks and domains. Another trend is the integration of deep-learning with other fields such as reinforcement learning, meta-learning, and generative modeling. Additionally, there is a growing focus on ethical considerations, fairness, and transparency in deep-learning algorithms. Despite the remarkable progress, challenges persist, including the need for more robust training algorithms, addressing biases and limitations of large-scale datasets, and improving the interpretability and explainability of deep-learning models. Overall, machine-learning is evolving at an unprecedented speed, driven by the pursuit of advancing the state-of-the-art and addressing the practical challenges of deploying deep neural networks in real-world applications.

3.2 Learning paradigms

Machine- and deep-learning can be categorized into three main types: *supervised learning*, *unsupervised learning*, and *reinforcement learning*. In Figure 3.1, a summary of the learning paradigms used in vascular function selection is shown. *Supervised learning*, models are trained using labeled data examples, where both the input data and the desired output are provided. The model learns to map the input to the output by minimizing a predefined loss function. This type of learning is commonly used for tasks such as regression and classification, which are introduced in the section 3.3.2. Some traditional examples of supervised learning algorithms are random forests and support vector machines.

Unsupervised learning involves, instead, training models on unlabeled data without any predefined output to guide the learning process. The goal is to discover patterns, relationships, and structures within the data. *Unsupervised clustering*, where data points are grouped based on their similarity, and *dimensionality reduction*, which reduces the number of features while preserving important information, are common tasks in unsupervised learning. Typical algorithms for clustering are K-means and hierarchical clustering. Principal component analysis and autoencoders are examples of dimensionality reduction techniques.

Reinforcement learning focuses on training agents to make sequential decisions in an environment to maximize a reward signal. The agent interacts with the environment and learns through trial and error, receiving feedback in the form of rewards or penalties. The objective is to learn an optimal policy that leads to the

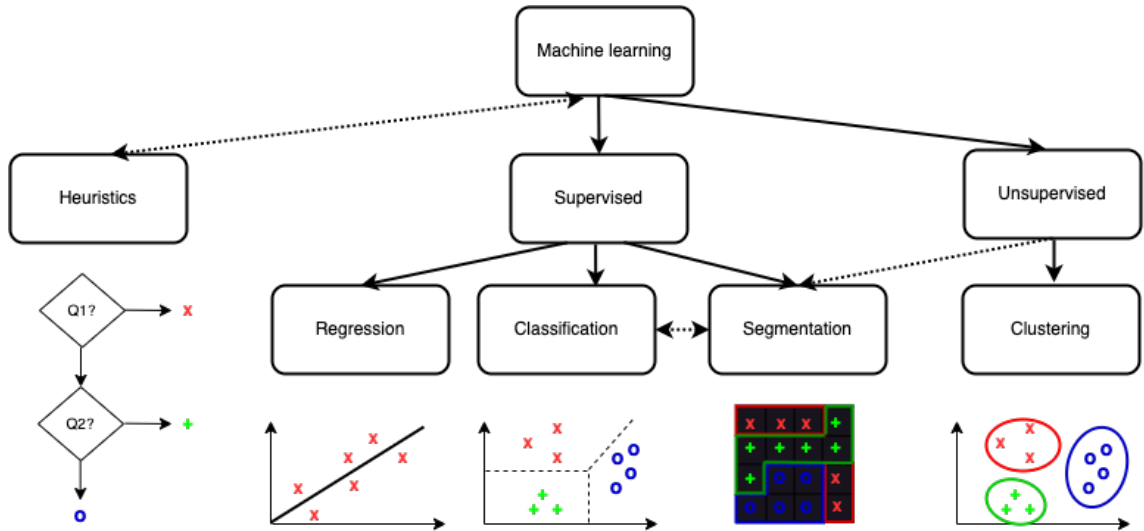


Figure 3.1: Types of machine-learning paradigms used in vascular function analysis.

highest cumulative reward. Reinforcement learning algorithms employ exploration and exploitation strategies to balance the trade-off between discovering new actions and exploiting known rewarding actions.

The various paradigms of machine-learning provide unique approaches for addressing diverse problem types. An interesting aspect to consider is that deep neural networks can be trained using any of the three learning paradigms mentioned earlier. This characteristic distinguishes neural networks from certain algorithms that were specifically designed to operate exclusively with labeled or unlabeled data (e.g., unsupervised K-means). Neural networks offer versatility by being adaptable to different learning scenarios, making them a powerful tool in machine-learning.

In this thesis I mostly exploit supervised learning techniques. The selection of algorithms and training strategies are determined based on the inherent nature of the problems and the suitability of the different algorithms in addressing them. Consequently, a combination of deep neural networks and traditional machine-learning algorithms IS utilized.

3.3 Building deep learning algorithms

Neural networks are a type of machine-learning model inspired by the structure and functioning of the human brain. They are composed of interconnected layers of

artificial neurons. These neurons (also known as nodes or units) work together to process input data and produce output predictions.

The basic building block of a neural network is a neuron, which takes in multiple inputs, applies weights to these inputs, sums them up, and passes the result through an activation function. The activation function introduces non-linearity to the network, allowing it to learn complex patterns in the data.

The output of a neuron can be represented as follows:

$$y = f \left(\sum_{i=1}^n w_i x_i + b \right) \quad (3.1)$$

where y is the output of the neuron, f is the activation function, w_i are the weights associated with the inputs x_i , b is the bias term. The weights and bias terms in the network are learned during the training process, where the network adjusts them to minimize the difference between its predictions and the true output values.

Neural networks consist of multiple layers, including an input layer, one or more hidden layers, and an output layer. The hidden layers receive the outputs from the previous layers as input, process them, and generate, as output, intermediate representations of the data. The output layer, instead, produces the final model predictions.

The computations in a neural network can be represented as follows:

$$\mathbf{y} = f(\mathbf{W} \cdot \mathbf{X} + \mathbf{b}) \quad (3.2)$$

where: \mathbf{y} is the vector of outputs, \mathbf{W} is the matrix of weights connecting the layers, \mathbf{X} is the vector of inputs, \mathbf{b} is the vector of bias terms.

Training a neural network involves an optimization process called *backpropagation*, where the network adjusts its weights and biases using optimization algorithms and the so-called chain rule of calculus. This allows the network to update its parameters in a way that reduces the difference between its predicted outputs and the true outputs.

3.3.1 Activation functions

Activation functions play a crucial role in deep learning by introducing non-linearity to the neural network architecture. They transform the input of a neuron into its output and allow modeling complex data relationships that overcome the limitations of linear transformations. Popular activation functions are summarized in Table 3.1,

3. METHODOLOGY

including the sigmoid function, the hyperbolic tangent function, the Rectified Linear Unit (ReLU) function [18] and variations of the latter one (e.g. Leaky ReLU [33] and the Parametric ReLU [25]).

| Function | Equation | Description |
|-----------------|--|---|
| Sigmoid | $\text{sigmoid}(x_i) = \frac{1}{1+e^{-x_i}}$ | Maps the input to the range $[0, 1]$. Common in binary classification. |
| Tanh | $\tanh(x_i) = \frac{e^{x_i} - e^{-x_i}}{e^{x_i} + e^{-x_i}}$ | Maps the input to the range $[-1, 1]$, symmetric and zero-centered output. |
| ReLU | $\text{ReLU}(x_i) = \max(0, x_i)$ | Devised to help mitigating the vanishing gradient problem. |
| Leaky ReLU | $\text{LReLU}(x_i) = \begin{cases} x_i & \text{if } x_i > 0 \\ 0.01x_i & \text{otherwise} \end{cases}$ | Introduces a small slope for negative inputs. |
| Parametric ReLU | $\text{PReLU}(x_i) = \begin{cases} x & \text{if } x_i > 0 \\ \beta_i x_i & \text{otherwise} \end{cases}$ | Flexible extension of LReLU introducing $\beta_i \in [0, 1]$. |

Table 3.1: Overview of common activation functions used in deep learning.

3.3.2 Model optimization

Machine-learning optimization aims to find the optimal values for the set of parameters of a model that minimize a given loss function. The process involves iteratively adjusting the set of those parameters to minimize the discrepancy between the predicted outputs of the model (\hat{y}) and the true values (y) in a training data subset. The optimal set of model parameters θ^* that minimizes the loss function $\mathcal{L}(\theta)$ can be written as:

$$\theta^* = \arg \min_{\theta} \mathcal{L}(\theta) \tag{3.3}$$

In order to find θ^* , optimization algorithms are used. A widely common used optimization algorithm is *gradient descent*. The algorithm iteratively updates the

parameters in the (opposite) direction of the loss function gradient with respect to the set of parameters. The update rule can be written as:

$$\theta^{(t+1)} = \theta^{(t)} - \alpha \nabla \mathcal{L}(\theta^{(t)}) \quad (3.4)$$

where $\theta^{(t)}$ represents the parameter values at iteration t , α is the learning rate determining the step size, and $\nabla \mathcal{L}(\theta^{(t)})$ is the gradient of the loss function with respect to the model parameters. The gradient provides the direction of the steepest descent, indicating how the parameters should be adjusted to reduce the loss function. This iterative process continues until convergence, typically defined by a stopping criterion (e.g., reaching a certain number of iterations or achieving a small change in the loss function). Upon convergence, the parameters θ^* represent the optimized values that best fit the training data and minimize the targeted loss. It is worth mentioning that optimization algorithms operate at a local level and cannot guarantee to reach the global optimum.

Stochastic gradient descent (SGD) is a variant of the standard gradient descent algorithm that randomly selects a subset of training samples at each iteration to estimate the gradient. This algorithm introduces randomness in the gradient estimation process, which can lead to faster convergence and better generalization performance, especially in large-scale datasets. The gradient $\nabla \mathcal{L}_i(\theta^{(t)})$ is computed in SGD on a randomly selected mini-batch i consisting of a random subset of training samples. Note that in SGD there is a faster computation and more frequent update of the parameters. The SGD algorithm also provides a solution when fitting machine-learning models over large and heavy datasets, which is sometimes challenging given computational processing constraints.

3.3.3 Backpropagation

The backpropagation algorithm adjusts the weights and biases of the network based on the error between the predicted outputs and the true outputs. It involves two main steps: forward propagation and backward propagation. During forward propagation, the input data is fed through the network, and the activations of each neuron are computed layer by layer until the output is obtained. Considering the input to the network as \mathbf{x} , the activation of a neuron in layer l can be computed as:

$$\mathbf{a}^{(l)} = f(\mathbf{W}^{(l)} \mathbf{a}^{(l-1)} + \mathbf{b}^{(l)}) \quad (3.5)$$

where $\mathbf{a}^{(l)}$ is the activation vector of layer l and $\mathbf{a}^0 = \mathbf{x}$, $\mathbf{W}^{(l)}$ is the weight matrix of layer l , $\mathbf{a}^{(l-1)}$ is the activation vector of the previous layer, $\mathbf{b}^{(l)}$ is the bias vector of layer l , and $f(\cdot)$ is the activation function.

After forward propagation, the predicted outputs of the network are compared with the true outputs to compute the error. The goal of backpropagation is to minimize this error by adjusting the weights and biases.

During backward propagation, the error is propagated back through the network to compute the gradients of the weights and biases. Let's denote the error vector as \mathbf{e} , which is the difference between the predicted outputs and the true outputs. The gradients of the weights and biases can be computed as:

$$\frac{\partial E}{\partial \mathbf{W}^{(l)}} = \frac{\partial \mathbf{e}}{\partial \mathbf{a}^{(l)}} \frac{\partial \mathbf{a}^{(l)}}{\partial \mathbf{W}^{(l)}} \quad (3.6)$$

$$\frac{\partial E}{\partial \mathbf{b}^{(l)}} = \frac{\partial \mathbf{e}}{\partial \mathbf{a}^{(l)}} \frac{\partial \mathbf{a}^{(l)}}{\partial \mathbf{b}^{(l)}} \quad (3.7)$$

where $\frac{\partial \mathbf{e}}{\partial \mathbf{a}^{(l)}}$ is the derivative of the error with respect to the activations of layer l , and $\frac{\partial \mathbf{a}^{(l)}}{\partial \mathbf{W}^{(l)}}$ and $\frac{\partial \mathbf{a}^{(l)}}{\partial \mathbf{b}^{(l)}}$ are the derivatives of the activations with respect to the weights and biases of layer l , respectively. These gradients are then used to update the weights and biases of the network using any optimization algorithm, such as the earlier introduced SGD.

3.3.4 Supervised learning: Regression

Machine-learning regression is a type of supervised learning algorithms that is used to predict continuous numeric values based on a set of input features. In regression tasks, the goal is to establish a relationship between the input variables (also known as independent variables or features) and the output variable (also known as the dependent variable or target). The idea is to find a mathematical function or model that can accurately represent the relationship between the input and output variables.

In mathematical terms, let's consider a domain set $\mathcal{X} \in \mathbb{R}^d$, with corresponding label outputs $\mathcal{Y} \in \mathbb{R}$. The problem is then constrained to the finding of a model or function $h: \mathcal{X} \rightarrow \mathcal{Y}$ ($\mathbb{R}^d \rightarrow \mathbb{R}$). The different existing algorithms will make use of different learning functions h to map the input feature vector to an output space. The choice of the regression algorithm depends on the available data and the problem tackled. Some common regression algorithms include linear regression, random forests, support vector machines, and neural networks.

Regression loss

A typical regression loss function based on the l_2 norm is the mean-squared-error loss, defined as:

$$\mathcal{L}_{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.8)$$

where y_i and \hat{y}_i represent the ground truth and predicted values for the sample i , respectively. Other loss functions are, for instance, the absolute mean error or the Huber loss.

3.3.5 Supervised learning: Classification

In classification tasks, the algorithm learns to predict the sample's output class from a set of input features. Thus, similar to the regression case, the goal is to find a mathematical function or model that can represent the mapping of the input variables into the output class space.

In the simplest classification scenario, *binary classification*, the data samples are allocated into one of two possible classes. In mathematical terms, let's consider a domain set $\mathcal{X} \in \mathbb{R}^d$, with corresponding (binary) label outputs $\mathcal{Y} \in \{0, 1\}$. The idea is to find a model or function $h: \mathcal{X} \rightarrow \mathcal{Y}$. As for the regression problem, different algorithms will make use of different learning functions h to map the input feature vector in the output class space. Typical classification algorithms are random forests, support vector machines, neural networks, etc.

Classification loss

Among the available classification loss functions, a commonly used one is categorical cross-entropy. In the simplest scenario having only two classes (i.e., binary classification), the categorical cross-entropy is defined as:

$$\mathcal{L}_{CE} = - \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (3.9)$$

where N represents the number of instances in the dataset, y_i is the ground truth label for instance i (0 for the negative class, 1 for the positive class), and p_i is the predicted probability of instance i to belong to the positive class.

3.3.6 Supervised learning: Segmentation

Segmentation refers to the process of partitioning input data (typically an image) into multiple distinct regions based on specific criteria or properties. It involves assigning a label (class) to each pixel or voxel in the input image, that indicates to which segment or region it belongs to. Image segmentation can be considered as an extension of the classification task introduced above, where multiple samples (pixels or voxels) from an image or volume are assigned to a specific region (class).

Segmentation loss

Loss functions based on overlap metrics are used. The categorical cross-entropy loss function of Equation 3.9 can be used for image segmentation. An alternative widely used segmentation loss function is the soft-Dice loss, defined as:

$$\mathcal{L}_{Dice} = 1 - \frac{2 * \sum_i y_i \hat{y}_i}{\sum_i y_i + \sum_i \hat{y}_i} \quad (3.10)$$

where in this case \hat{y}_i and y_i are the predicted output and the ground truth values for sample i , and \sum_i represents the summation over all pixels/voxels in the image.

3.3.7 Convolutional Neural Networks

Convolutional neural networks (CNNs) are a specific type of neural network used in computer vision tasks. They are designed to automatically learn hierarchical representations of visual data (i.e., images) through a series of convolutional operations and pooling layers. CNNs excel at capturing spatial relationships and local patterns in images, making them highly effective in tasks such as image classification, object detection, and segmentation.

The CNN architectures that have made crucial contributions in the field are LeNet-5 [31], AlexNet [28], VGGNet [55], GoogLeNet [59] (Inception), ResNet [24], and the U-Net [49]. Figure 3.2 shows a pictorial representation of some of these architectures. LeNet-5, proposed by LeCun et al. in 1998, laid the foundation for CNNs and introduced the concept of convolutional layers. AlexNet demonstrated the power of deep CNNs by winning the ImageNet Large Scale Visual Recognition Challenge [13] with a significant margin. VGGNet, proposed by Simonyan and Zisserman in 2014, introduced a deeper architecture with smaller filter sizes, achieving impressive performance on various vision tasks. GoogLeNet was presented by Szegedy et al. in 2014 and introduced the concept of inception modules, which allowed for

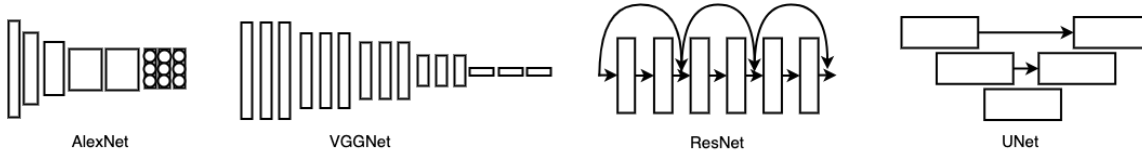


Figure 3.2: Relevant CNN architectures.

efficient and parallel processing at multiple scales. It showcased the benefits of network depth and computational efficiency. ResNet, proposed by He et al. in 2015, addressed the issue of vanishing gradients in deep networks by introducing skip connections, enabling the training of much deeper architectures.

The U-Net is a popular deep-learning architecture commonly used in medical imaging for tasks such as image segmentation. It was introduced by Ronneberger et al. in 2015 [49], specifically designed for 2D biomedical image analysis. The U-Net architecture consists of a contracting path (encoder) and an expansive path (decoder), with skip connections that help preserve spatial information. The encoder captures contextual information, while the decoder combines it with high-resolution features to generate detailed segmentation maps. Çiçek et al. [11] extended the U-Net architecture to its 3D version, allowing volumetric image analysis through 3D convolutional layers. More recently nnUnet was presented as a generalized end-to-end pipeline for training U-Net models in medical image datasets [26]. nnUnet showed remarkable performance in various image tasks, as shown in the Decathlon challenge [2].

3.4 Learning from perfusion imaging data

3.4.1 Data challenges

Perfusion imaging is a complex and heterogeneous modality that poses challenges for downstream post-processing tasks. Developing data-driven algorithms that can effectively handle this complexity is crucial for the success and robustness of machine learning methods. During my experimental work in this thesis, I have encountered these challenges and I have grouped them into three sources: *i*) real-world perfusion data is heterogeneous, *ii*) perfusion imaging involves the handling of large 4D images with limited computational resources and *iii*) perfusion imaging post-processing algorithms are susceptible to changes in the vascular functions.

Firstly, the heterogeneity in perfusion CT imaging data arises from various sources, including variations in scanner technologies, scanning protocols, and patient characteristics. Different scanner vendors and models incorporate distinct technological advancements that can influence the output images. For instance, the number of CT detectors and tube rotation speeds constrain the CTP acquisition, with the former determining the z-axis anatomical coverage and the latter affecting the temporal sampling resolution. While optimal acquisition entails whole-brain CTP coverage, this is not universally available. Modern clinical scans involve machines with z-axis coverage varying from 4 to 16 cm [10]. In scanners with limited z-axis coverage, the CTP can be obtained by moving the table during the acquisition or by scanning adjacent anatomical regions using separate injection protocols. In any case, the differences in the number of detectors (and, therefore, the anatomical coverage) and tube rotation speeds affect the scanned brain area, the number and thickness of the slices, and the frame rate of the CTP scan. Moreover, the scanning and contrast injection protocols employed in clinical practice are not standardized, leading to a wide range of image variability and quality. Parameters influencing image quality are typically listed in the recommended guidelines outlined in Chapter 2. For example, the CTP frame rate, which is recommended to be equal to or faster than 2 seconds per frame, may vary across centers, with some utilizing slower acquisitions (e.g., 3 seconds per frame) or employing faster frame rates during the contrast wash-in phase and lower frame rates during the contrast wash-out phase. Contrast injection protocols also impact the degree of tissue contrast enhancement, with some scans producing significantly higher contrast in vessels (a couple of hundred Hounsfield Units) while others exhibit poor vessel enhancement with the contrast agent. Furthermore, image variability also considers the inherent anatomical, physiological, and pathological differences among patients. Brain volume, shape, and structures naturally differ throughout an individual’s lifespan. Age also influences cardiac output, resulting in a slower distribution of the contrast agent within the brain and broader time-attenuation curves measured on CTP. Additionally, various brain conditions, disorders, and pathologies can significantly impact brain anatomy and physiology. Consequently, developing machine learning models that can effectively capture this extensive variability is extremely challenging. Such models require large amounts of data and are highly dependent on the characteristics and variability of the data, making them data-hungry and data-dependent.

Secondly, perfusion imaging data is computationally intensive due to its 4D nature, encompassing three-dimensional (3D) spatial information combined with temporal sampling. This involves capturing multiple 3D brain volumes at various time points

during the passage of the contrast agent. The recommended acquisition duration for CT perfusion is approximately 60-70 seconds, resulting in the generation of 60-70 3D brain volumes when the sampling frequency is one frame per second. In some centers, longer acquisitions are performed, leading to twice the number of temporal acquisitions. The z-axis coverage of brain images also exhibits consistent differences based on the CT scanner technology. For example, in the open ISLES'18 dataset, the number of slices in a scan ranges between 2 and 20. However, modern CT scanners can achieve a very high z-axis resolution, allowing some centers to acquire CTP images with hundreds of axial slices. Consequently, the voxel resolution, number of z-slices, temporal sampling rate, and acquisition length contribute to a substantial amount of data that needs to be processed. When working on machine learning tasks related to vascular functions, utilizing the entire dataset as input for the models may be necessary. This is because vascular functions are typically selected based on an “optimal” criterion. Thus, when considering a set of possible AIF/VOF candidates, only the best one (or a few) should be retained. In other words, the problem is inherently dependent on the entire available dataset information, as the *optimal* vascular function criterion demands a comparison among multiple (if not all) candidate options. Consequently, patch-based deep-learning approaches, commonly used to address memory constraints in tasks such as image segmentation, are not easily applicable to this problem. Therefore, alternative solutions are necessary to tackle these challenges effectively.

Thirdly, achieving voxel-level accuracy in perfusion imaging is crucial, particularly when analyzing vascular functions. Even a single suboptimal voxel can completely undermine the accuracy of downstream perfusion data analysis. This limitation applies not only to the use of a “global” vascular function but also to region-of-interest or subsets of voxels defining the AIF or VOF. An ideal AIF should exhibit the earliest bolus arrival time and the highest contrast increase among all candidate AIF voxels. Therefore, the presence of just one suboptimal AIF voxel can alter the resulting average curve, causing it to no longer represent the earliest or the most contrast-enhanced curve. It is worth noting that deconvolution algorithms, as described by the dilution theory, are extremely sensitive to the shape of the AIF curve (Equation 2.1). Consequently, even a slight deviation from the optimal AIF choice, such as a time delay, increased noise levels, or a wider profile due to partial volume effects, can significantly impact the quantification of perfusion lesion volumes. Interestingly, for many other machine learning tasks in medical imaging, achieving voxel-level accuracy may not be necessary or may not significantly affect the downstream radiological analysis. For example, a few false positives or negatives

in the segmentation of large structures or large lesions might be tolerable. However, when it comes to learning from vascular curves, machine-learning models need to be optimally tailored to handle the low error margins and deliver voxel-level accuracy performance.

In conclusion, the research problem under investigation presents challenges encompassing the highly heterogeneous clinical data, the computational constraints for handling complex 4D imaging modalities, and the necessity for precise and reliable machine learning models.

3.4.2 Selecting vascular functions with machine-learning

Initial software packages based on deconvolution performed manual vascular function selection. The process consisted of choosing a group of artery/vein candidate voxels based on their anatomical location (e.g., large vessels, contra-lateral to the ischemic region, etc.), signal-profile characteristics (like bolus arrival time or contrast increase) and signal quality (high SNR and as little contaminated with artifacts as possible) criteria. Alternatives to the voxel-by-voxel selection approach included drawing a region of interest within a specific vessel. Whatever the manual voxel selection approach was, the temporal signals of the chosen candidates were later averaged and used as vascular function inputs to the deconvolution algorithm. In the early times of perfusion imaging, these kinds of approaches seemed to be feasible. CT images were very limited in anatomical spatial coverage and in voxel resolution, which allowed the selection of *optimal voxel candidates* through exhaustive manual search. Alongside, computers were several orders of magnitude more limited in computational processing power than current machines. Furthermore, data-driven algorithms (such as machine-learning ones) were not as advanced as nowadays. Consequently, automatic and fast processing of large amounts of data was less feasible than now.

Improvements in CT scanner technologies (such as faster rotation speeds, larger head coverage, and higher image resolution) came with an increase in the amount of generated imaging data. For instance, initial CTP protocols for acute stroke imaging used to solely acquire a couple of slices around the middle cerebral artery. Nowadays, full brain coverage with a hundred axial slices is no longer rare. All these fantastic technological improvements came together with increasing difficulties to continue doing exhaustive manual search for the vascular functions, especially after research showed the importance of having time-efficient clinical practice protocols (as explained in section 2.1.2, reminder: *Time is brain!*). Larger computation capabilities and the fast development and deployment of data-centric algorithmic solutions opened the

possibility of automating the vascular function selection process in different ways. In Figure 3.1, a schematic representation of different families of algorithms used for the vascular function selection problem are shown. Please note that this section does not aim to detail algorithmic strategies, though a brief explanation is needed to understand their usage in the problem under study.

The first algorithms to be used for tackling the problem were *heuristics*. With *heuristics* I refer in this thesis to algorithmic rules based on expert’s knowledge and practical experience that allow a fast and acceptable estimation of the entity being searched - in this case, the vascular functions. Most of these methods define different rules based on the Hounsfield Units contrast increase of the temporal signals, the bolus arrival time, the noise measured in it, etc., to select good arterial/vein voxel candidates. The works of Rempp et al. [48] and Carroll et al. [8] are examples of such approaches.

More sophisticated methods to tackle this problem are based on different machine learning paradigms. We show in Figure 3.1 a classification of learning paradigms used for AIF/VOF selection. Unsupervised clustering methods have been widely explored. In a nutshell, unsupervised clustering algorithms aim to split data into a (known or unknown) number of groups (i.e., clusters) based on extracted features. Later on, vascular functions can be selected by applying heuristic rules over the cluster of interest, and finally, choose the voxels most likely to belong to an artery or vein. Examples of clustering methods for AIF selection include affinity propagation [54], normalized cuts clustering [72], and fuzzy clustering [37].

Among the supervised learning approaches, *classification* algorithms have also been used. For the training of classification algorithms that can choose vascular functions from the CT data, the image needs to be split into classes. Additionally, it is necessary to determine the type of samples (such as voxels, regions, clusters, etc.) that will be provided to the algorithm. A simple strategy for doing so would be, for instance, to label the AIF voxels as belonging to one class and to label ‘all-the-rest’ voxels as belonging to another class. Then, the algorithm’s aim is to find the ‘AIF’ class. Examples of such methods are the works of [67, 16].

Lastly, *segmentation* methods -a specific classification algorithm subtype- have also been used. In [16], a 3D-Unet is used to segment arterial voxels and later estimate the AIF. Considering the image processing definition of *segmentation* algorithms, we can see that labels might not always be necessary to split an image into regions. This is commonly known as *unsupervised segmentation*, a family of algorithms with common points with *unsupervised clustering* methods but exclusively working with imaging data. These types of algorithms are not discussed nor explored within this

3. METHODOLOGY

thesis. Therefore, I leave details about this family of methods as optional and external homework for curious minds.

PART II

Contributions



Summary of Contributions

In the previous chapters I have introduced the role of the arterial and venous functions for acute stroke perfusion analysis. It has been exhibited that their suboptimal selection and estimation may critically impact the quantification of brain perfusion biomarkers and, therefore, the treatment decision-making. Furthermore, machine-learning strategies that can help characterize these vascular functions have been introduced and stratified according to different learning paradigms (*supervised* and *unsupervised*). **The goal of this thesis is to characterize the perfusion vascular functions derived from real-world CTP scans by using data-driven machine-learning models.** In this chapter, I introduce the specific goals pursued within my thesis and the contributions made in the field under study. I would like to inform the reader that the research questions were strategically chosen to be as close as possible to the clinical routine and, hence, having high applicability and transferability potential.

The first research question targeted in this thesis challenges the *optimal* AIF concept and approaches it from a different perspective:

Q1: What is the best AIF that maximizes the deconvolution algorithm performance? And can I automatically estimate it with machine-learning models?

Talking about the *best* AIF can lead to a long debate about what is considered *best*. For this work, the optimal AIF has been assumed as the temporal function that, through SVD deconvolution, led to the best estimate of the necrotic *core*. Note that in CTP imaging, the penumbra and core regions are estimated through deconvolution analysis, meaning there is no way (not even manually) to delineate the lesions' ground truth. Consequently, to have a better insight into this research question, I used a very useful but rarely available clinical dataset where concomitant CTP acquisitions and DWI scans with manually annotated core ground truth were available. In Chapter 5.1 the proposed methodology is detailed through a peer-reviewed conference

article. In a nutshell, I proposed a differentiable implementation of the time-invariant SVD deconvolution integrated within a deep-learning scheme. The method allows fitting an AIF-generating CNN by turning the SVD deconvolution algorithm into an optimization (performance maximization) problem. Through this approach, I show that it is possible to find the *best* AIF (following the above-mentioned ‘best’ criteria) for each CTP scan. These AIFs produced better estimates of the core lesions than the results obtained when using AIFs chosen by a manual rater or by a clinical CTP software package, thus reducing the core delineation performance gap between CTP and DWI modalities. Furthermore, the presented work is also novel from a machine-learning perspective since it tackles a regression task by optimizing a segmentation loss function instead.

The second topic addressed in this thesis was the selection of vascular functions from clinical and real-world CTP scans. The question that I investigated was:

*Q2: Can I automatically select vascular functions from clinical CTP scans?
In that case, how close to a manual expert performance can I get?*

As shown in Chapter 2 this problem has been widely explored in previous works. The *key* question that I asked myself was: *Has this problem been solved?* The short answer that I found was negative. Failures in the AIF selection are frequent in clinical software packages [70, 56] and, therefore, may lead to completely unusable and uninterpretable deconvolution results. Under these circumstances the CTP study might be discarded or, in some cases, would require an extra manual AIF correction step. In any case, a wrong AIF selection delays the acute stroke workflow and may need repeating the acquisition, thus increasing the radiation dose and contrast agent delivered to the patient. On one hand, algorithms used in commercial software packages are rarely described in the literature, and limited descriptions can only be accessed through their patent description. On the other hand, open approaches available in the literature are limited on several fronts. For instance, most of them have been devised for perfusion MRI but have not been validated in CTP. Moreover, most of these methods have been tested over synthetic data or over small, homogenous, and/or quality-selected scans. Therefore, the challenge here was to devise an automatic AIF-VOF selection method that can work over real-world CTP data and under different image quality scenarios. Chapter 5.2 describes the proposed methodology *AIFNet* through its peer-reviewed manuscript. In this work, I show that vascular function selection can be accurately achieved (almost at expert level) through convolutional neural networks. The method is also very different from previous works

as it is not approached as a clustering or classification problem. Moreover, *AIFNet* is trained using sparse annotations, thus allowing faster data labeling and faster deployment. The reader may note that the work’s novelty goes beyond its clinical application. The method combines different learning strategies as a segmentation task is optimized through a regression loss function (*Regression loss* \Rightarrow *Segmentation task*). Hence, the proposed *mixed* learning scheme inverts the strategy used to tackle question *Q1* (*Segmentation loss* \Rightarrow *Regression task*). *AIFNet* has been implemented in the CTP software package (FDA-cleared and CE-marked) **icobrain cva** (icometrix, Leuven, Belgium) and is currently clinically used in several countries. Furthermore, the method’s novelty led to a patent [50].

Thirdly, I extended the applicability of the vascular functions to a different and common problem in stroke clinical setting: the *truncation* (i.e., early interruption of the scanning) of the CTP series. The question that I raised in this matter was:

Q3: Can I exploit the deconvolution-available vascular functions in order to detect truncated CTP scans automatically?

As discussed in chapter 2, the accurate and reliable estimation of acute stroke perfusion lesions is crucial for reperfusion treatment decision-making. However, short CTP acquisitions not satisfying the guidelines’ recommended duration are still common in clinical practice and may, sometimes, lead to inaccurate quantification of the perfusion lesions. In this last chapter of my thesis, I study and quantify CTP truncation artifacts by synthetically shortening the duration of the scans. First, the study allowed reporting statistical metrics describing those truncation artifacts, shedding some light on questions like ‘*How long should a scan be in order to quantify perfusion volumes reliably?*’ or ‘*How many seconds from the AIF’s peak are necessary to avoid truncation errors?*’. Second, after measuring the impact of those errors, I moved the research towards detecting scans impacted by truncation. Therefore, I propose machine-learning models that allow the detection of unreliably short CTP scans when fed with features extracted from the AIF and the VOF. The results show that such an approach based on the vascular functions is more accurate in detecting truncated scans than solely considering the scan duration. Through the proposed methodology, radiologists and neurointerventionalists could be warned of misleading perfusion volumes due to short CTP acquisitions. Details about this research are available in the peer-reviewed manuscript of Chapter 5.3.



Publications

The publications product of this thesis are included in the next chapters.

5.1 Differentiable Deconvolution for Improved Stroke Perfusion Analysis

Ezequiel de la Rosa, David Robben, Diana M. Sima, Jan S. Kirschke & Bjoern Menze

Conference: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020

Synopsis: Perfusion imaging is the current gold standard for acute ischemic stroke analysis. It allows quantification of the salvageable and non-salvageable tissue regions (*penumbra* and *core* areas respectively). In clinical settings, the singular value decomposition (SVD) deconvolution is one of the most accepted and used approaches for generating interpretable and physically meaningful maps. Though this method has been widely validated in experimental and clinical settings, it might produce suboptimal results because the chosen inputs to the model cannot guarantee optimal performance. For the most critical input, the arterial input function (AIF), it is still controversial how and where it should be chosen even though the method is very sensitive to this input. In this work we propose an AIF selection approach that is optimized for best performance of *core* lesion segmentation. The AIF is regressed by a neural network optimized through a differentiable SVD deconvolution, aiming to maximize the agreement of the core lesion segmentation with ground truth data. To our knowledge, this is the first work exploiting a differentiable deconvolution model with neural networks. We show that our approach can generate AIFs without any manual annotation, thus avoiding manual raters' intervention, which may lead to inter-rater variability. The method achieves manual expert performance in the ISLES18 dataset. We conclude that the methodology opens new possibilities for improving perfusion imaging quantification with deep neural networks.

Contributions of thesis author: Conceptualization of the project. Preprocessing of the data using custom image processing software. Implementation of a differentiable SVD deconvolution functions using Tensorflow. Integration of these functions within a deep learning pipeline. Experiment design, planning and conduction. Evaluation of results: performance analysis and statistical analysis. Comparison against other approaches. Manuscript writing.

Copyright: © 2020 Springer Nature Switzerland AG.



Differentiable Deconvolution for Improved Stroke Perfusion Analysis

Ezequiel de la Rosa^{1,2(✉)}, David Robben^{1,3,4}, Diana M. Sima¹,
Jan S. Kirschke⁵, and Bjoern Menze²

¹ icometrix, Leuven, Belgium

{ezequiel.delarosa,david.robben,diana.sima}@icometrix.com

² Department of Computer Science,

Technical University of Munich, Munich, Germany

bjoern.menze@tum.de

³ Medical Imaging Research Center (MIRC), KU Leuven, Leuven, Belgium

⁴ Department of Electrical Engineering, ESAT/PSI, KU Leuven, Leuven, Belgium

⁵ Neuroradiology, School of Medicine,

Technical University of Munich, Munich, Germany

jan.kirschke@tum.de

Abstract. Perfusion imaging is the current gold standard for acute ischemic stroke analysis. It allows quantification of the salvageable and non-salvageable tissue regions (penumbra and core areas respectively). In clinical settings, the singular value decomposition (SVD) deconvolution is one of the most accepted and used approaches for generating interpretable and physically meaningful maps. Though this method has been widely validated in experimental and clinical settings, it might produce suboptimal results because the chosen inputs to the model cannot guarantee optimal performance. For the most critical input, the arterial input function (AIF), it is still controversial how and where it should be chosen even though the method is very sensitive to this input. In this work we propose an AIF selection approach that is optimized for maximal core lesion segmentation performance. The AIF is regressed by a neural network optimized through a differentiable SVD deconvolution, aiming to maximize core lesion segmentation agreement with ground truth data. To our knowledge, this is the first work exploiting a differentiable deconvolution model with neural networks. We show that our approach is able to generate AIFs without any manual annotation, and hence avoiding manual rater's influences. The method achieves manual expert performance in the ISLES18 dataset. We conclude that the methodology opens new possibilities for improving perfusion imaging quantification with deep neural networks.

Keywords: Perfusion imaging · SVD deconvolution · Deep learning

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-59728-3_58) contains supplementary material, which is available to authorized users.

1 Introduction

Perfusion imaging techniques are the clinical standard for acute ischemic stroke lesion assessment. They acquire images of the passage of a contrast agent bolus through the brain tissue. Since the perfusion series are not directly clinically interpretable, they require the computation of physically meaningful parameter maps. Although different approaches may be used for their computation (e.g. compartmental models, which are mainly used over long acquisition time perfusion MRI), the preferred technique in perfusion CT analysis is the singular value decomposition (SVD) deconvolution [11,25]. The technique has been well validated in experimental [16] and clinical [1] settings and is widely implemented in perfusion CT software [7,25]. Cerebral blood flow (CBF) and time to the maximum residue function (Tmax) are typically used maps, though cerebral blood volume and time-to-peak maps are often considered as well. Parameter maps are critical for treatment decision making. They allow assessing the salvageable *penumbra* and irreversible *core* necrotic lesions, and hence determining if reperfusion techniques may reduce the disease damage severity.

The SVD deconvolution method requires as input an arterial input function (AIF), defined as the concentration time-curve inside an artery feeding the tissue under study. In practice, the AIF is mostly selected by a physician, a demanding, highly variable and poorly reproducible process. Its correct selection is the cornerstone for generating accurate maps, as has been shown that minimal changes in its location and/or shape may strongly impact the deconvolution process [13]. Although several works studied how and where the AIF should be chosen [4], the subject is still very controversial. Thus, AIF selection is suboptimal, since we do not know which function will maximize the deconvolution performance. Besides, the AIF's concept is defined based on the SVD-deconvolution theoretical model, which relies on several assumptions violated in clinical practice. For instance, limited voxel resolution, partial volume effect, time-curve delays, noise and other confounders are typically limiting the model's performance. Consequently, it is not straightforward to define which AIFs are the best to use in practice.

In this work we use neural networks to generate the AIF, aiming to find the *best* AIFs in core lesion segmentation terms. Through experiments on the ISLES18 database we show that the method is able to learn from scratch to generate AIFs that maximize the segmentation agreement with manually delineated ground truth. Thus, the AIFs are learned without any expert annotations of the AIFs themselves, hence avoiding potential rater's bias. We show, as well, that the approach is able to yield manual expert performance in the ISLES18 database.

2 Method

2.1 Differentiable Deconvolution

We propose the optimizable framework of Fig. 1 which generates the *best* AIF for SVD deconvolution. We define the *best* AIF as the one yielding the highest

agreement between the estimated core lesion and the ground truth core lesion. The input to the framework is 4D perfusion data and the output is a lesion map. It consists of a CNN that generates the AIF $c_{art}(t)$. The generated AIF together with the 4D input perfusion series pass through a differentiable SVD deconvolution block, which outputs relative CBF (rCBF) maps after image deconvolution. Finally, rCBF is transformed into the lesion probability map y_{pred} . The framework is end-to-end trainable, which means that gradients are backpropagated through all blocks including the SVD deconvolution, thus allowing to generate the best AIF candidate that maximizes the segmentation performance.

AIF Generating CNN. Unlike most previous works that use unsupervised clustering [14, 15, 18, 19, 21] or supervised segmentation [6] approaches, we propose a regression CNN for obtaining an AIF $c_{art}(t)$ from the 4D perfusion series. The architecture is fairly straightforward: the only particularity is that the input size varies in the z-axis from 2 to 8 slices. Therefore, it consists of 3D convolutional layers followed by an average pooling over the z-axis. Subsequently, there are 2D convolutional layers and finally a fully connected layer with the same number of output neurons as the number of time-points in the perfusion data. This final 1D vector represents $c_{art}(t)$. After each convolution, average pooling and dropout [23] are used. ReLU [10] activations are applied in all layers except for the final AIF output layer, where a linear activation is employed. For fitting 4D data into the CNN, we encode volume time-points as channels in the network.

SVD Deconvolution. For generating interpretable and physically-meaningful maps, perfusion series are deconvolved using the well-validated delay-invariant SVD deconvolution. Our GPU implementation of the algorithm uses Volterra discretization [22]. Differentiability of the SVD algorithm was studied in [17, 24]. Through analytical methods it was shown that the Jacobian of the SVD is computationally feasible [17]. Even more, in [8] a differentiable SVD was used in neural network layers for a variety of tasks including image segmentation. In this work, we take advantage of current deep learning libraries that support auto-differentiation and allow backpropagating SVD gradients. Given the fact that deconvolution is an ill-conditioned problem [7, 20], we use a Tikhonov regularization scheme. The output of the deconvolution block of Fig. 1 is rCBF.

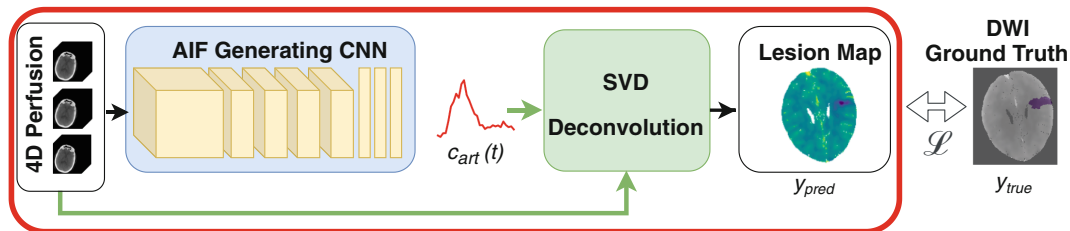


Fig. 1. Differentiable SVD deconvolution pipeline.

Although in clinical practice the Tmax and rCBF maps are typically used for identifying core lesions, we only consider rCBF since the differentiability of Tmax is not fully clear.

In a nutshell, the analysis of perfusion data can be described by the following convolution product:

$$c_{voi}(t_j) = \int_0^t c_{art}(\tau)k(t_j - \tau)d\tau \quad (1)$$

where c_{voi} is the agent concentration in the voxel under consideration, c_{art} is the concentration curve measured in an artery feeding the volume (i.e., the AIF), and k is the impulse response function that characterizes the tissue of interest [7, 22]. Note that here $t = 0$ is taken prior to the arrival of the contrast agent (such that $k(t) = 0$ for $t < 0$). For solving Eq. 1 (i.e., finding k) we rely on discretization methods, since in practice the measured AIF c_{art} and the agent concentration in the tissue volume c_{voi} are discretized at specific time points. Considering these time points as $t_j = (j - 1)\Delta t$ (for $j = 1, \dots, N$), and assuming that $c_{art}(t)$ is negligible for $t > N\Delta t$, the discretization of Eq. 1 can be approximated as:

$$c_{voi}(t_j) = \int_0^t c_{art}(\tau)k(t_j - \tau)d\tau \approx \Delta t \sum_{i=1}^N c_{art}(t_i)k(t_{j-i+1}) \quad (2)$$

which can be rewritten as a linear system:

$$\mathbf{c} = \mathbf{A}\mathbf{k} \quad (3)$$

where \mathbf{A} is the Volterra matrix with the following A_{ij} elements [22]:

$$\begin{cases} A_{i0} = (2c_{art}(t_i) + c_{art}(t_{i-1}))/6 & (0 < i \leq N - 1) \\ A_{ii} = (2c_{art}(t_i) + c_{art}(t_{i+1}))/6 & (0 < i \leq N - 1) \\ A_{ij} = \frac{2}{3}c_{art}(t_i) + \frac{c_{art}(t_{i-1})}{6} + \frac{c_{art}(t_{i+1})}{6} & (1 < i \leq N - 1, 0 < j < i) \\ A_{ij} = 0 & \text{elsewhere} \end{cases} \quad (4)$$

We assume, without loss of generality, that $\Delta t = 1s$, which is usually the case in clinical practice. For the derivation of \mathbf{A} and for a more in-depth understanding of discretization methods in perfusion imaging, the reader is referred to [22].

A classical way of solving Eq. 3 is by means of SVD as:

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_{i=1}^r \mathbf{u}_i \sigma_i \mathbf{v}_i^T \quad (5)$$

where $r = rank(\mathbf{A})$, $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_r]$ and $\mathbf{v} = [\mathbf{v}_1, \dots, \mathbf{v}_r]$ are the left and right singular vectors, respectively, and $\mathbf{\Sigma} = diag(\sigma_1, \dots, \sigma_r)$ is the diagonal matrix containing singular values in decreasing order. The least squares solution of Eq. 3 for \mathbf{k} is:

$$\mathbf{k} = \sum_{i=1}^r \frac{\mathbf{u}_i^T \mathbf{c}}{\sigma_i} \mathbf{v}_i \quad (6)$$

Nonetheless, in cases where \mathbf{A} is ill-conditioned, Eq. 6 is not a suitable solution of the linear system since a small variability in \mathbf{c} may generate very large variability in \mathbf{k} [7]. Thus, regularization is required for having a stable result as:

$$\mathbf{k}_\lambda = \sum_{i=1}^r \left(f_{\lambda,i} \frac{\mathbf{u}_i^T \mathbf{c}}{\sigma_i} \right) \mathbf{v}_i \quad (7)$$

where $f_{\lambda,i} = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2}$ are Tikhonov regularization parameters with $\lambda = \lambda_{rel} \sigma_i$. The parameter λ_{rel} should be chosen in the interval $(0, 1)$. In our implementation we empirically set $\lambda_{rel} = 0.3$. Finally, it can be proven that the cerebral blood flow can be obtained as:

$$\text{CBF} = \frac{1}{\rho_{voi}} \max(k(t_j)) \quad (8)$$

where $\rho_{voi} [\frac{g}{ml}]$ stands for the mean tissue density in the voxel. For a mathematical demonstration of this statement the reader is referred to [7]. Finally, following current clinical practice, CBF is normalized with mean healthy CBF values for obtaining a map in a subject-independent scale.

Ischemic Lesion Map. Generated rCBF maps require some sort of transformation to obtain lesion probability maps y_{pred} that can be compared with the binary ground truth masks. With this aim, we use sigmoid activations centered at rCBF = 0.38 for mapping the y_{pred} probability values. This cutoff previously was found to be optimal for this dataset [5]. It is worth to mention that it is possible to allow the network to choose the best cutoff, but we preferred to keep a fixed threshold as mostly used in clinical practice. In such a way, our proposed method is directly comparable with the results of a manual AIF selection using the same cutoff value, assuring that differences in results are only driven by the choice of AIF.

2.2 Implementation and Optimization

The framework is implemented using TensorFlow and Keras, where we ensure effective gradient propagation by only using differentiable operations. Given the class imbalance between *healthy* and *necrotic* brain tissue, a soft-Dice loss function is used as follows:

$$\mathcal{L} = 1 - \frac{2 \sum y_{true} y_{pred}}{\sum y_{true} + \sum y_{pred}} \quad (9)$$

where y_{pred} is the framework's output lesion map and y_{true} the ground truth manually delineated lesion mask.

Optimization is conducted using stochastic gradient descent with momentum, with a unitary batch size. In order to improve the network's learning stage and to overcome data limitations, two types of data augmentation are conducted. First, perfusion specific data augmentation [20] is implemented at an image level, which

allows mimicking AIF bolus delay arrivals and AIF peak concentration scaling. Second, traditional segmentation data augmentation is used, including image rotation, translation, flipping and random Gaussian noise addition.

2.3 Data and Experiments

The free and open ISLES18 database is used [9,12]. It consists of 4D CT perfusion series with ground truth core lesion delineations obtained from diffusion weighted imaging (DWI). From the total amount of scans provided in the challenge ($n = 156$), only the training set ($n = 94$) includes ground truth data and hence was used for our experiments. The dataset is multi-scanner and multi-center, obtained from different institutions from the U.S. and Australia. All provided images are already motion-corrected, co-registered for matching CTP with DWI modalities, and spatio-temporally resampled (with 256×256 images and 1 volume/second).

To compare our proposed method against the current clinical approach, an expert provided manual AIF annotations for the entire ISLES18 training set. A single global AIF was selected per case, following recommendations found in the literature [4]. Moreover, our results are compared with the automatic AIF selection approach included in **icobrain cva** (**icometrix**, Leuven, Belgium), an FDA cleared CTP analysis software package. In both cases, the CTP images are deconvolved with the chosen AIF. The same data preprocessing and deconvolution algorithm of Sect. 2.1 are used. As such, any difference in results is only caused by the AIFs themselves. rCBF maps are generated and the core lesions are quantified. For all our experiments and methods a fixed threshold rCBF = 0.38 is used for defining the core lesions [5].

Since the dataset is already preprocessed, the preprocessing on our side is limited to a spatio-temporal smoothing before the CTP data is deconvolved. The method's performance is assessed through 5-fold cross-validation. Results are evaluated at parameter map and lesion segmentation levels. The discriminant power of the rCBF maps for differentiating healthy and necrotic tissue are assessed through the area under the ROC curve (AUC). The lesion segmentation is assessed by means of Dice and Jaccard indexes, 95% Hausdorff distance and volumetric Bland-Altman analysis [3].

3 Results

Training of our model takes ~ 2.5 h on an Nvidia K80 GPU with 12 GB dedicated memory. During testing, the entire process of AIF selection, SVD deconvolution and lesion quantification takes ~ 1.15 s per case. On the other hand, the expert annotation of an AIF takes around one minute per case.

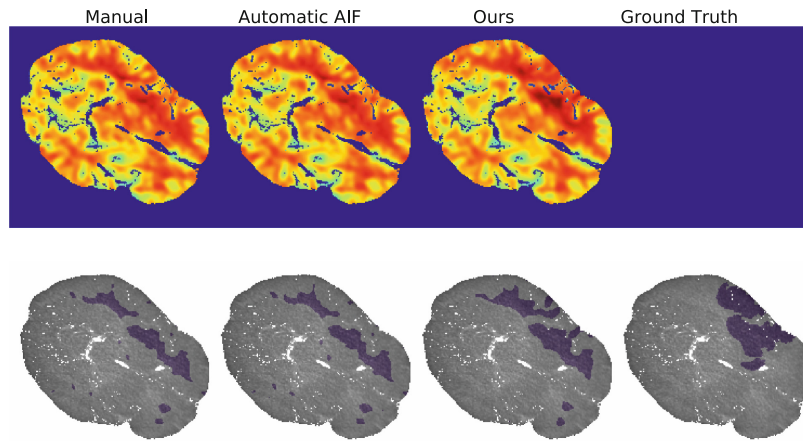


Fig. 2. Qualitative results. Top: rCBF maps; Bottom: Core-lesion segmentations.

In Fig. 2, the resulting rCBF maps generated with the different methods are shown for an example case. The corresponding AIFs selected or generated by these methods are included in the supplementary material. The automatic algorithm yields results that are visually similar to the ones generated with the expert AIF selection. However, our generated map better matches the DWI ground truth lesions than both other approaches. When assessing the ROC AUC values of rCBF for discerning healthy and necrotic tissue, our proposed method shows better performance than the other methods (Table 1). In all segmentation metrics the methods achieve comparable performance, with our new method outperforming the others, but the differences are not statistically significant (paired t-test).

Figure 3 shows the lesion volume quantification performance in Bland-Altman plots. Volumetric overestimation is found for all methods when comparing with results reported by [5], which shows better agreement with ground truth. This can be explained by their use of an extra Tmax criterion and due to their use of a modified ground truth, that excluded tissue with a low Tmax. The manual rater and the automatic AIF software yield comparable results, with the expert annotations having the best agreement with the ground truth (p-value non-statistically significant between these methods,

Table 1. Mean (standard deviation) of various segmentation metrics for the different methods. rCBF: relative cerebral blood flow; AUC: area under the ROC curve; HD: Hausdorff distance.

| Method | rCBF AUC | Dice | Jaccard | 95% HD (mm) |
|---------------|--------------|----------------------|----------------------|------------------------|
| Expert | 0.856 | 0.353 (0.201) | 0.233 (0.157) | 54.136 (19.449) |
| Automatic AIF | 0.856 | 0.351 (0.202) | 0.232 (0.158) | 55.015 (18.790) |
| Ours | 0.868 | 0.359 (0.201) | 0.238 (0.157) | 53.747 (18.875) |

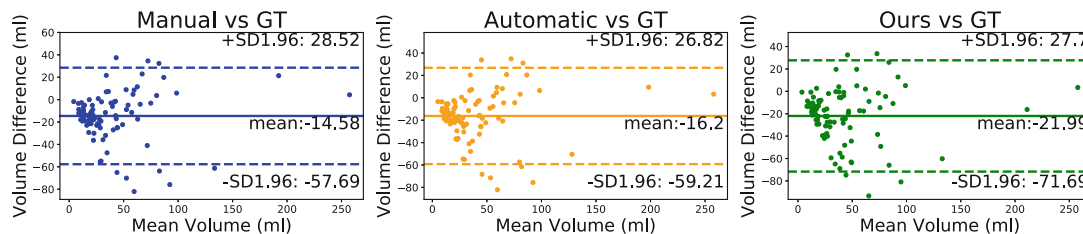


Fig. 3. Volumetric Bland-Altman plots. GT: Ground truth.

Mann-Whitney U test). Our approach shows a larger volumetric bias than these methods, that was only statistically significant when compared with manual results (p-value = 0.04 and p-value 0.09 when compared with manual and automatic methods respectively, Mann-Whitney U test). The reason for our method’s mismatch in segmentation and volumetric performance may be driven by the optimized loss function. As explained in [2], soft Dice loss can lead to volumetric bias.

4 Discussion and Conclusion

We present a neural network that regresses the AIF for CT perfusion analysis. Unlike previous methods that aim to imitate the AIF selection of a human rater, the training of our network requires no manual annotations. To this end, we implement a differentiable SVD deconvolution, allowing the AIF generating network to be optimized for generating the most discriminative rCBF maps with reference to DWI images.

There are no previous studies that use CNNs to regress the AIF. Moreover, we are first in applying SVD deconvolution differentiability for perfusion applications. Unlike previous works, our approach does not require manual annotations. This is a crucial finding for devising automatic methods free from manual rater’s influence. From a scientific point, it is interesting that our approach generates the ‘best’ AIF: current guidelines for the manual selection of AIFs do not have that guarantee.

Our experiments with the ISLES18 data yielded results matching expert segmentation performance. The rCBF maps that we obtained are slightly more informative for finding core-lesions than the ones an expert generates, as shown in the ROC analysis. In all segmentation metrics considered our method is comparable to an expert or an FDA-cleared software.

In future work, we aim to work with extra datasets and incorporate additional perfusion parameters. We currently only have 94 subjects and since each subject corresponds to a single regression, we effectively have only 94 samples. We expect that a larger dataset will further improve results. Similarly, additional data is needed to increase confidence in the method’s performance.

This work only optimizes the rCBF map. In future work, we also intend to optimize the Tmax parameter map (which is defined as $argmax_t k(t)$).

Tmax estimation is crucial for the applicability of the method and for improving the lesion quantification since an increased Tmax is indicative for tissue at risk. This will require finding a differentiable substitute for the *argmax* and a ground truth for tissue at risk (e.g. the final infarct in patients that did not have reperfusion).

Acknowledgements. This project received funding from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement TRABIT No 765148. EDLR, DR and DMS are employees of icometrix. DR is supported by an innovation mandate of Flanders Innovation & Entrepreneurship (VLAIO).

References

1. Albers, G.W., et al.: Ischemic core and hypoperfusion volumes predict infarct size in SWIFT PRIME. *Ann. Neurol.* **79**(1), 76–89 (2016)
2. Bertels, J., Robben, D., Vandermeulen, D., Suetens, P.: Optimization with soft dice can lead to a volumetric bias. arXiv preprint [arXiv:1911.02278](https://arxiv.org/abs/1911.02278) (2019)
3. Bland, J.M., Altman, D.: Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **327**(8476), 307–310 (1986)
4. Calamante, F.: Arterial input function in perfusion MRI: a comprehensive review. *Progr. Nucl. Magn. Reson. Spectrosc.* **74**, 1–32 (2013)
5. Cereda, C.W., et al.: A benchmarking tool to evaluate computer tomography perfusion infarct core predictions against a DWI standard. *J. Cereb. Blood Flow Metab.* **36**(10), 1780–1789 (2016)
6. Fan, S., et al.: An automatic estimation of arterial input function based on multi-stream 3d CNN. *Front. Neuroinform.* **13**, 49 (2019)
7. Fieselmann, A., Kowarschik, M., Ganguly, A., Hornegger, J., Fahrig, R.: Deconvolution-based CT and MR brain perfusion measurement: theoretical model revisited and practical implementation details. *J. Biomed. Imaging* **2011**, 14 (2011)
8. Ionescu, C., Vantzou, O., Sminchisescu, C.: Training deep networks with structured layers by matrix backpropagation. arXiv preprint [arXiv:1509.07838](https://arxiv.org/abs/1509.07838) (2015)
9. Kistler, M., Bonaretti, S., Pfahrer, M., Niklaus, R., Büchler, P.: The virtual skeleton database: an open access repository for biomedical research and collaboration. *J. Med. Internet Res.* **15**(11), e245 (2013)
10. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, pp. 1097–1105 (2012)
11. Lin, L., Bivard, A., Krishnamurthy, V., Levi, C.R., Parsons, M.W.: Whole-brain CT perfusion to quantify acute ischemic penumbra and core. *Radiology* **279**(3), 876–887 (2016)
12. Maier, O., et al.: ISLES 2015-a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. *Med. Image Anal.* **35**, 250–269 (2017)
13. Mlynash, M., Eynhorn, I., Bammer, R., Moseley, M., Tong, D.C.: Automated method for generating the arterial input function on perfusion-weighted MR imaging: validation in patients with stroke. *Am. J. Neuroradiol.* **26**(6), 1479–1486 (2005)
14. Mouridsen, K., Christensen, S., Gyldensted, L., Østergaard, L.: Automatic selection of arterial input function using cluster analysis. *Magn. Reson. Med.: Off. J. Int. Soc. Magn. Reson. Med.* **55**(3), 524–531 (2006)

15. Murase, K., Kikuchi, K., Miki, H., Shimizu, T., Ikezoe, J.: Determination of arterial input function using fuzzy clustering for quantification of cerebral blood flow with dynamic susceptibility contrast-enhanced mr imaging. *J. Magn. Reson. Imaging: Off. J. Int. Soc. Magn. Reson. Med.* **13**(5), 797–806 (2001)
16. Murphy, B., Chen, X., Lee, T.Y.: Serial changes in CT cerebral blood volume and flow after 4 hours of middle cerebral occlusion in an animal model of embolic cerebral ischemia. *Am. J. Neuroradiol.* **28**(4), 743–749 (2007)
17. Papadopoulo, T., Lourakis, M.I.A.: Estimating the Jacobian of the singular value decomposition: theory and applications. In: Vernon, D. (ed.) *ECCV 2000*. LNCS, vol. 1842, pp. 554–570. Springer, Heidelberg (2000). https://doi.org/10.1007/3-540-45054-8_36
18. Peruzzo, D., Bertoldo, A., Zanderigo, F., Cobelli, C.: Automatic selection of arterial input function on dynamic contrast-enhanced MR images. *Comput. Methods Programs Biomed.* **104**(3), e148–e157 (2011)
19. Rausch, M., Scheffler, K., Rudin, M., Radü, E.: Analysis of input functions from different arterial branches with gamma variate functions and cluster analysis for quantitative blood volume measurements. *Magn. Reson. Imaging* **18**(10), 1235–1243 (2000)
20. Robben, D., Suetens, P.: Perfusion parameter estimation using neural networks and data augmentation. In: Crimi, A., Bakas, S., Kuijff, H., Keyvan, F., Reyes, M., van Walsum, T. (eds.) *BrainLes 2018*. LNCS, vol. 11383, pp. 439–446. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11723-8_44
21. Shi, L., et al.: Automatic detection of arterial input function in dynamic contrast enhanced MRI based on affinity propagation clustering. *J. Magn. Reson. Imaging* **39**(5), 1327–1337 (2014)
22. Sourbron, S., Luypaert, R., Morhard, D., Seelos, K., Reiser, M., Peller, M.: Deconvolution of bolus-tracking data: a comparison of discretization methods. *Phys. Med. Biol.* **52**(22), 6761 (2007)
23. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**(1), 1929–1958 (2014)
24. Townsend, J.: Differentiating the singular value decomposition. Technical Report 2016 (2016). <https://j-towns.github.io/papers/svd-derivative>
25. Vagal, A., et al.: Automated CT perfusion imaging for acute ischemic stroke: pearls and pitfalls for real-world use. *Neurology* **93**(20), 888–898 (2019)

5.2 AIFNet: Automatic Vascular Function Estimation for Perfusion Analysis Using Deep Learning

Ezequiel de la Rosa, Diana M. Sima, Bjoern Menze, Jan S. Kirschke & David Robben

Journal: Medical Image Analysis

Synopsis: Perfusion imaging is crucial in acute ischemic stroke for quantifying the salvageable *penumbra* and irreversibly damaged *core* lesions. As such, it helps clinicians to decide on the optimal reperfusion treatment. In perfusion CT imaging, deconvolution methods are used to obtain clinically interpretable perfusion parameters that allow identifying brain tissue abnormalities. Deconvolution methods require the selection of two reference vascular functions as inputs to the model: the arterial input function (AIF) and the venous output function, with the AIF as the most critical model input. When manually performed, the vascular function selection is time demanding, suffers from poor reproducibility and is subject to the professionals' experience. This leads to potentially unreliable quantification of the penumbra and core lesions and, hence, might have negative impact on the treatment decision process. In this work we automatize the perfusion analysis with *AIFNet*, a fully automatic and end-to-end trainable deep learning approach for estimating the vascular functions. Unlike previous methods using clustering or segmentation techniques to select vascular voxels, *AIFNet* is directly optimized for the estimation of the vascular function, which allows to better recognise the time-curve profiles. Validation on the public ISLES18 stroke database shows that *AIFNet* approaches inter-rater performance for the vascular function estimation and, subsequently, for the parameter maps and core lesion quantification obtained through deconvolution. We conclude that *AIFNet* has potential for clinical transfer and could be incorporated in perfusion deconvolution software.

Contributions of thesis author: Manual annotation of the entire dataset. Preprocessing of the dataset using in-house developed software. Exploration of different deep learning models (architectures and optimization functions) targeting the problem. Implementation of the proposed AIFNet deep learning architecture. Integration of

5.2. AIFNet: Automatic Vascular Function Estimation for Perfusion Analysis Using Deep Learning

AIFNet in a deep learning pipeline. Experiment design, planning and conduction. Evaluation of results: performance analysis and statistical analysis. Comparison against other approaches. Manuscript writing.

Copyright: © 2021 The Author(s). Published by Elsevier B.V.



Challenge report

AIFNet: Automatic vascular function estimation for perfusion analysis using deep learning

Ezequiel de la Rosa^{a,b,*}, Diana M. Sima^a, Bjoern Menze^b, Jan S. Kirschke^c, David Robben^{a,d,e}^a *icomatrix*, Leuven, Belgium^b *Department of Computer Science, Technical University of Munich, Munich, Germany*^c *Neuroradiology, School of Medicine, Technical University of Munich, Munich, Germany*^d *Medical Imaging Research Center (MIRC), KU Leuven, Leuven, Belgium*^e *Medical Image Computing (MIC), ESAT-PSI, Department of Electrical Engineering, KU Leuven, Leuven, Belgium*

ARTICLE INFO

Article history:

Received 23 September 2020

Revised 25 June 2021

Accepted 4 August 2021

Available online 6 August 2021

Keywords:

Ischemic stroke

Perfusion imaging

Arterial input function

Deep learning

ABSTRACT

Perfusion imaging is crucial in acute ischemic stroke for quantifying the salvageable *penumbra* and irreversibly damaged *core* lesions. As such, it helps clinicians to decide on the optimal reperfusion treatment. In perfusion CT imaging, deconvolution methods are used to obtain clinically interpretable perfusion parameters that allow identifying brain tissue abnormalities. Deconvolution methods require the selection of two reference vascular functions as inputs to the model: the arterial input function (AIF) and the venous output function, with the AIF as the most critical model input. When manually performed, the vascular function selection is time demanding, suffers from poor reproducibility and is subject to the professionals' experience. This leads to potentially unreliable quantification of the penumbra and core lesions and, hence, might harm the treatment decision process. In this work we automatize the perfusion analysis with AIFNet, a fully automatic and end-to-end trainable deep learning approach for estimating the vascular functions. Unlike previous methods using clustering or segmentation techniques to select vascular voxels, AIFNet is directly optimized at the vascular function estimation, which allows to better recognise the time-curve profiles. Validation on the public ISLES18 stroke database shows that AIFNet almost reaches inter-rater performance for the vascular function estimation and, subsequently, for the parameter maps and core lesion quantification obtained through deconvolution. We conclude that AIFNet has potential for clinical transfer and could be incorporated in perfusion deconvolution software.

© 2021 The Author(s). Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

1. Introduction

Stroke is currently the second leading cause of mortality and the third leading cause of disability worldwide ([Stroke Unit Trialists Collaboration, 2013](#)). In physio-pathological terms, it is defined as an 'acute neurologic dysfunction of vascular origin with sudden (within seconds) or at least rapid (within hours) occurrence of symptoms and signs corresponding to involvement of focal areas in the brain' ([Force, 1989](#)). Two main types of the disease can be recognised: ischemic and hemorrhagic, representing 85% and 15% of total cases respectively ([Hinkle and Guanci, 2007](#)). We focus on the ischemic case, where there is a shortage in the blood supply to the brain tissue, cutting the provision of oxygen and glucose. During the ischemic event, brain tissue might become necrotic (i.e., cells

are dead and the tissue is irreversibly damaged, known as *core*) or in a hypo-perfused but salvageable state (i.e., tissue is at risk but could return to a healthy condition, known as *penumbra*).

1.1. Perfusion CT in acute ischemic stroke

Acute ischemic stroke therapies rely on reperfusion techniques, where the main goal is to reestablish the blood flow supply in the affected territories by thrombolysis or thrombectomy. Identifying which patients might benefit from these treatments is critical for clinical decision making ([Campbell and Parsons, 2018](#); [Albers et al., 2016](#)). To this end, assessment and quantification of the core and penumbra tissues are required. In the acute scenario, computer tomography (CT) is the most widely used imaging technique, where perfusion CT (CTP) enables the determination of the core and penumbra areas. An iodinated contrast agent is intra-venously injected in the patient for 7–10 s, and continuous CT acquisition

* Corresponding author.

E-mail address: ezequiel.delarosa@icomatrix.com (E. de la Rosa).

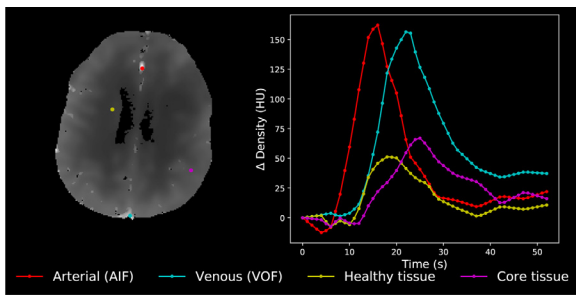


Fig. 1. Contrast enhancement curves for different brain tissues. Left: a perfusion CT example from ISLES18. Right: Corresponding time curves at the indicated locations. Healthy and diseased brain areas have been identified through diffusion weighted imaging. The healthy and core time-curves are scaled by a factor of six for visualization. HU: Hounsfield units.

is followed for around 50 s (Fieselmann et al., 2011). As such, 4D data is generated, resulting in a brain volume imaged during the agent passage through the brain vessels and parenchyma. The process for evaluating brain tissue status is performed by firstly obtaining parameter maps from the CTP time series and by later applying a tissue discrimination rule (mainly, thresholding). Typical maps include cerebral blood flow (CBF), cerebral blood volume (CBV), time to peak (TTP) and time to the maximum of the residue function (Tmax). It is worth saying that there is no *gold* standard for quantifying perfusion metrics (Lorenz et al., 2006), and all methods found in literature provide merely non-exact solutions. Experimental studies have shown that CBV and CBF discriminate ischemic and oligemic tissue with 90.6% and 93.3% sensitivity and specificity, respectively, when using histological measurements as ground truth (Murphy et al., 2007). The most widely used methods for CTP parameter map estimation are based on deconvolution (Konstas et al., 2009), which provides a solution to the indicator dilution theory described by:

$$c_{\text{tissue}}(t) = c_{\text{art}}(t) \circledast h(t) \quad (1)$$

where $c_{\text{tissue}}(t)$ represents the CTP contrast enhancement in a voxel of tissue, $c_{\text{art}}(t)$ is the contrast enhancement in the arteries (known as arterial input function, from now on ‘AIF’), $h(t)$ is the flow-scaled residue function and \circledast represents the convolution operator. The delay-invariant singular value decomposition deconvolution is the preferred technique for algebraically solving Eq. (1) and it is widely implemented in software packages (Fieselmann et al., 2011; Konstas et al., 2009; Kosior and Frayne, 2007; Kudo et al., 2010; Vagal et al., 2019). The method has been extensively validated in clinical practice, showing better performance compared to similar techniques (Konstas et al., 2009; Fieselmann et al., 2011) like the maximum slope approach (Konstas et al., 2009; Klotz and König, 1999), non-delay invariant deconvolution (Østergaard et al., 1996a,b), etc. Deconvolution methods require as input to the algorithm the CTP series and two vascular functions: the AIF and the venous output function (VOF). These vascular functions are reference time-curves representing the contrast concentration inlet and outlet to the tissue under consideration $c_{\text{tissue}}(t)$. Fig. 1 shows an example case of vascular functions (i.e. AIF and VOF) and contrast enhancement curves for healthy and core tissue areas. In clinical practice the AIF and VOF are generally selected by a radiologist, a time demanding and highly variable process that implies selecting in the CTP series the optimal candidate voxels. Frequently, a single voxel per vascular function is selected, which leads to low SNR curves. Voxel selection is, moreover, subject to the professionals’ training and experience, which not only introduces human bias (Lorenz et al., 2006) but it may also affect CBF maps depending which side of the brain the AIF is chosen from (Wu et al., 2003; Thijs et al., 2004). The AIF is so

critical for generating accurate maps that very small changes in its shape and/or location may produce a profound effect over the generated maps (Mlynash et al., 2005; Mouridsen et al., 2006). Besides, given the acute context of the disease, a fast voxel selection has to be performed. It has been shown that for every 30-minute delay in reperfusion, the probability of good outcome decreases by 20% (Khatri et al., 2014). Given these limitations, automatic, fast and reproducible core and penumbra quantification are highly desired.

1.2. Automatic core and penumbra segmentation

Automatic machine and deep learning approaches for core and penumbra quantification have been explored in two ways: 1) by direct parameter maps estimation and 2) by direct lesions segmentation. On one hand, automatic parameters maps estimation (i.e., bypassing deconvolution) was explored in (McKinley et al., 2018; Meier et al., 2019; Robben and Suetens, 2018; Ulas et al., 2018a,b). However, the main drawback of these methods is the fact that *silver* standard maps obtained through deconvolution or other methods (e.g. compartmental models in the case of perfusion MRI) are approximated. Note that in these approaches there is also an AIF assumption behind the parameter maps ground truth. As such, these methods do not improve the perfusion gold standard, but aim to reproduce it with a different model. On the other hand, direct lesion segmentation approaches use native CTP data with or without perfusion maps as model inputs. Thus, the neural networks are used for finding a non-linear transformation from CTP and/or CBF, CBV, MTT and Tmax that estimates brain lesions. For instance, in (Bertels et al., 2018) and (Robben et al., 2020) direct lesion segmentation is conducted by only using CTP images. While the former work exploits contralateral brain information into a U-Net based architecture, the latter work includes metadata and vascular functions into a multiresolution DeepMedic-based (Kamnitsas et al., 2017) architecture. Other works include parameter maps obtained through deconvolution as inputs to the model (Clèrigues et al., 2019; Abulnaga and Rubin, 2018; Song and Huang, 2018; Wang et al., 2020). Similarly as in Bertels et al. (2018), Clèrigues et al. (2019) exploit brain symmetry information with U-nets. Song and Huang (2018) and Wang et al. (2020) propose, instead, to synthesize pseudo diffusion weighted imaging (DWI) data to improve core lesion segmentation. While deep learning based approaches showed good overall performance, their main limitation is the poor model’s explainability and lack of quality control. Since these fully ‘black-box’ methods do not allow AIF or perfusion maps inspection, they preclude physicians to recompute the parameter maps with a manually corrected AIF in clinically or technically challenging cases. As such, the clinical transferability potential of these models is limited. In this work we aim to automatize, instead, the well validated deconvolution process by the automatic selection of vascular functions. In this way, we avoid approximating parameters that can be directly estimated through a physical model while also preserving explainability and quality control in clinical settings.

1.3. Automatic vascular function selection

Automatic vascular function selection has been explored for perfusion MRI in (Murase et al., 2001; Mouridsen et al., 2006; Peruzzo et al., 2011; Shi et al., 2014; Shi and Malik, 2000; Yin et al., 2015; Fan et al., 2019; Winder et al., 2020). These methods mainly rely on clustering techniques, where fuzzy c-means (Murase et al., 2001), K-means (Mouridsen et al., 2006), hierarchical clustering (Peruzzo et al., 2011), gamma-variate based clustering (Rausch et al., 2000) and affine propagation clustering (Shi et al., 2014) were explored. Heuristic approaches have also

been traditionally used, where some rules are defined for finding the best-matching curve, such as in (Mlynash et al., 2005; Rempp et al., 1994). Other techniques use normalized cuts (Shi and Malik, 2000; Yin et al., 2015) and independent component analysis (Calamante et al., 2004). Moreover, vascular function estimation using deep neural networks can be conducted through segmentation approaches aiming to detect arterial voxels candidates. Fan et al. (2019) proposed a deep learning segmentation approach for delineating AIF candidates in perfusion MRI. The method uses two independently optimized 3D CNNs for conducting arterial tissue segmentation: one extracting spatial information in the $x - y - z$ axis, and another one extracting temporal-information in the $x - y - t$ axis (with t representing the temporal domain). Afterwards, the networks' results are merged using a late-fusion support vector machine. More recently, Winder et al. (2020) proposed a binary output CNN for classifying arterial vs non-arterial voxels in CTP and perfusion MRI. The AIF is then estimated by geometrically averaging the most probable arterial voxels. Though segmentation or classification methods can identify potential good curves, they have some limitations: i) They require complete manual annotation of all "good-looking" voxel curves, which is very time demanding and ii) They could not always guarantee optimal AIF curve selection (for a possible definition of optimal AIF selection, see Methods 3.1.2) since the algorithms are mainly optimized to perform selection based on spatial information rather than on time profiles. Segmentation methods may lead, for instance, to the undesired selection of noise-corrupted, low contrast enhanced or time delayed AIFs, which introduce errors in the deconvolution algorithms. In CTP imaging, however, vascular function selection is under-explored. Excepting the work of Winder et al. (2020) the few existing methods are mostly private and patented. Besides, most of the methods developed for perfusion MRI have not been validated for CTP. Despite perfusion CT and perfusion MRI having common working points, there are still technical differences that may affect the automatic selection of CTP vascular functions (such as lower tissue-density contrasts and lower SNR of CT compared to MRI). Moreover, additional challenges in CTP include overlapping density distribution of bone, artifacts and calcifications with the iodine contrast.

In this work we propose AIFNet, an end-to-end supervised convolutional neural network devised for estimating vascular functions (i.e. AIF and VOF) in perfusion imaging. The model is easy to train and deploy given the minimal data annotation required, which can be as little as a single voxel per vascular function. AIFNet receives 4D CTP series as input and generates as output i) the estimated AIF and VOF curves and ii) a voxel-wise, interpretable probability map representing the voxelwise contribution to the estimated vascular signal. Unlike other approaches, AIFNet is optimized at a vascular function level, which helps the network to better learn the time-curve profiles. The method preserves clinical interpretability and also enables quality control of the selected AIF/VOF brain vasculature, thus enhancing its clinical transferability potential. Through an extensive analysis at signal, parameter maps and lesion quantification levels, we show that our method performs almost as good as manual raters on the open ISLES18 acute stroke database.

2. Methods

2.1. Function estimation with deep learning

AIFNet is a fully end-to-end deep learning approach for vascular function estimation. It works by estimating a 3D probabilistic volume that represents the voxelwise contribution to the vascular signal. The advantage of having an averaged curve using multiple voxels lies on the higher function's SNR as well as on the method

robustness. The network receives as input the 4D perfusion series $x(t)$ and outputs the predicted arterial and venous functions as $\hat{y}(t) = AIFNet(x(t))$, being $x(t) = \{x_t; t = 1, 2, \dots, T\}$ with x_t representing the sampled time point volumes of dimension $M \times N \times Q$. We want to find for the considered volume, its corresponding vascular functions (AIF and VOF, for simplicity not differentiated in the notation) represented by $\hat{y}(t) = \{\hat{y}_t; t = 1, 2, \dots, T\}$, where \hat{y}_t is the estimated signal at time t (in Hounsfield units). For finding $\hat{y}(t)$, we represent each time point \hat{y}_t as a weighted average of all voxels of the volume x_t at that t time point as:

$$\hat{y}_t = \sum_{q=1}^Q \sum_{n=1}^N \sum_{m=1}^M x_t(m, n, q) * P_{vol}(m, n, q) \quad (2)$$

where P_{vol} is the 3D probabilistic volume containing the voxelwise contribution to the vascular function and fulfilling:

$$\sum_{q=1}^Q \sum_{n=1}^N \sum_{m=1}^M P_{vol}(m, n, q) = 1 \quad (3)$$

Our problem is hence confined to finding P_{vol} . With this aim, AIFNet receives as input native CTP series, and generates as outputs P_{vol} and its associated vascular function. To find \hat{y}_t it is important to optimize the similarity of the shape rather than the amplitude. This is due to two facts: 1) the absolute contrast values of the AIF can be disregarded, since given the high partial volume effect in the arteries, the AIF is later recalibrated with the VOF (Fieselmann et al., 2011) (also see Section 2.4.1) and 2) a suboptimal deconvolution might occur by selecting delayed input functions. The penalty in the time domain is introduced by using Pearson's correlation as loss function as follows:

$$\mathcal{L}(y(t), \hat{y}(t)) = - \frac{\sum_{t=1}^T (y_t - \bar{y})(\hat{y}_t - \bar{\hat{y}})}{\sqrt{\sum_{t=1}^T (y_t - \bar{y})^2} \sqrt{\sum_{t=1}^T (\hat{y}_t - \bar{\hat{y}})^2}} \quad (4)$$

where $y(t)$ and $\hat{y}(t)$ are the ground truth and predicted vascular functions with respective mean values $\bar{y} = \frac{1}{T} \sum_{t=1}^T y_t$ and $\bar{\hat{y}} = \frac{1}{T} \sum_{t=1}^T \hat{y}_t$.

2.2. Architecture

AIFNet architecture is shown in Fig. 2. It uses 3D convolutional layers for volumetric feature extraction, which are finally translated into a probabilistic volume through a 3D softmax operation. After finding P_{vol} , a voxelwise multiplication and 3D average pooling blocks are used for obtaining $\hat{y}(t)$, by means of Eq. (2). Each convolutional layer $L_k = \{k = 1, 2, \dots, K\}$ has 2^{3+k} filters with a $3 \times 3 \times 3$ kernel with exception of L_1 , which uses a $3 \times 3 \times 1$ one with the aim of compensating the lower image resolution along the z -axis. The CTP time points are incorporated as channel information into the network. A fixed number of T time points are used for all scans. In our experiments we use a T equal to the smallest number of time points found among all scans. Rectified linear units are used as activation functions (Krizhevsky et al., 2012). For mapping the convolutional layers to a single probabilistic volume, we add an extra convolution block (L_{out}) with only one filter in between L_K and the softmax operator.

2.3. Training phase

The network is optimized using stochastic gradient descent with momentum. A batch size of one sample is used. Regularization of the model is reached using a perfusion-specific data augmentation approach.

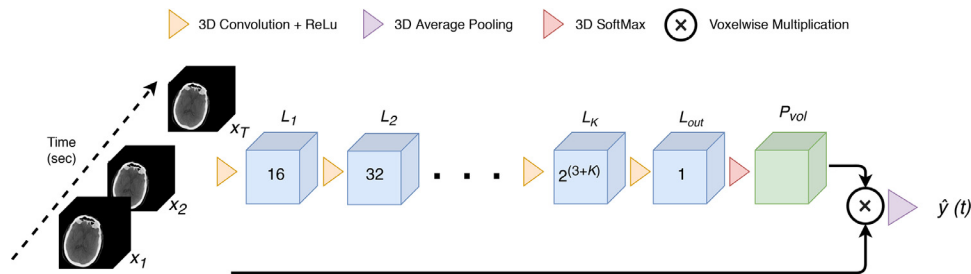


Fig. 2. AIFNet architecture. The CTP time points x_t ($t = 1, 2, \dots, T$) are incorporated as channels in the network. All convolutional layers use $3 \times 3 \times 3$ kernels except L_1 , which uses $3 \times 3 \times 1$. L_k is the k -th convolutional layer (with $k = 1, 2, \dots, K$). Inside each feature block the number of channels used is indicated. P_{vol} is the probabilistic volume. The 3D average pooling block averages the volumetric information along the $x - y - z$ axes, such that the predicted vascular function $\hat{y}(t)$ is a 1D vector of length T .

2.3.1. Perfusion specific data augmentation

We adapt the data augmentation strategy proposed in (Robben and Suetens, 2018) for working at an image level. Two perfusion specific phenomena are modelled: *i*) the variability of the contrast bolus arrival, which depends on the injection protocol and the patient's cardiovascular system and *ii*) the variability of the curve's peak-to-baseline (PTB) values, which depends on the iodine concentration in the contrast agent. Bolus arrival changes are simulated by randomly shifting the time attenuation curves, for which the first or last CT volumes are replicated (late or early simulated arrivals respectively). On the other hand, curve PTB changes are simulated in a three-step approach. Firstly, the pre-contrast averaged volume is subtracted from the perfusion series. Secondly, a random scaling is applied. Thirdly, the pre-contrast volume is re-added to the perfusion series. Uniform distributions are used for simulating the random time shifts and the random PTB scaling.

2.4. Testing phase

In the testing scenario, vascular function predictions are obtained by feeding the parametrized AIFNet model with the unseen CTP scans. The voxelwise multiplication and 3D average pooling blocks of AIFNet are performed over the full-length CTP perfusion series, with the aim of obtaining vascular function predictions that preserve the same number of time points as the native CTP scan. For VOF a signal recalibration step is also applied, as detailed below.

2.4.1. VOF signal recalibration

Our multiple signal averaging approach has the disadvantage of underestimating the VOF peaks. Since the VOF's role in deconvolution-based perfusion analysis is to compensate for partial volume effect in the AIF by its recalibration, it is important that its PTB matches the same amplitude as single CTP candidate voxels. Ideally, a suitable VOF curve has the highest PTB value among all venous voxel candidates. Therefore, we use a probabilistic volume that encodes voxelwise contribution to the function estimation. Firstly, we generate a 3D volume encoding the voxelwise PTB values. Secondly, we scale this volume with P_{vol} in order to obtain probabilistic-weighted PTB values. The VOF is finally recalibrated with the maximal value found in the weighted PTB distribution. We prefer using weighted PTB instead of only considering P_{vol} , since the highest probability voxel of P_{vol} might have a low PTB, thus leading to an underestimation of the VOF PTB value.

3. Experiments

3.1. Data

3.1.1. ISLES18

The large public multi-center and multi-scanner ISLES18 dataset is used for our experiments (Maier et al., 2017; Kistler et al., 2013; Cereda et al., 2016). It consists of 156 CTP acquisitions acquired from 103 acute stroke patients from three US centers and one Australian center. In the ISLES challenge, data is split into a train (94 CTP volumes scanned from 63 patients) and a test (62 CTP volumes scanned from 40 patients) sets. The mismatch between patients and scans is due to the limited field of view of some scanners, which leads to two independent CTP acquisitions from different brain regions in some cases. We have directly accessed the clean and preprocessed data through the ISLES challenge site (<http://www.isles-challenge.org/>). For each acquisition, CTP and DWI data were performed within 3 hours of each other. The open database provides CTP scans for the whole dataset and infarct core lesion masks (delineated in DWI images) for the training set only. Subjects having more than 50% of the DWI lesion with normal perfusion at the moment of the CTP acquisition were excluded, as well as those subjects with bad quality of the baseline CTP data and/or with inappropriate image coregistration due to distortions (Cereda et al., 2016). CTP volumes have been motion corrected and coregistered for matching the DWI lesion masks. Finally, scans have been spatio-temporally resampled (with a 256×256 dimension matrix and with a temporal resolution of one volume per second). For a more detailed description of this database the reader is referred to (Cereda et al., 2016).

3.1.2. Vascular function annotation

All training and testing scans are in-house annotated by two independent raters (DR & EdIR). A single *global* AIF and VOF per scan is selected (i.e., functions are measured from a major artery/vein and used as global inputs for the tissue in the whole brain (Calamante, 2013)), where the following AIF time attenuation curves are preferred: *i*) contralateral voxels to the affected area (rather than ipsilateral ones) (Kealey et al., 2004; Calamante, 2013), *ii*) Early bolus arrival AIF curves with a large and narrow peak enhancement (Calamante, 2013) *iii*) Curves with high contrast-to-noise ratio and, ideally, less affected by partial volume effect (qualitatively assessed) (Calamante, 2013). The *best* voxel candidate (following the just mentioned criterion) among the anterior cerebral arteries, middle cerebral arteries, internal carotid arteries or the basilar artery are chosen as AIF. On the other hand, VOF curves are located in the superior sagittal, transverse or sigmoid sinuses, which are large vessels less affected by partial volume effect than other vessels. All vascular function annotations are provided as supplementary material.

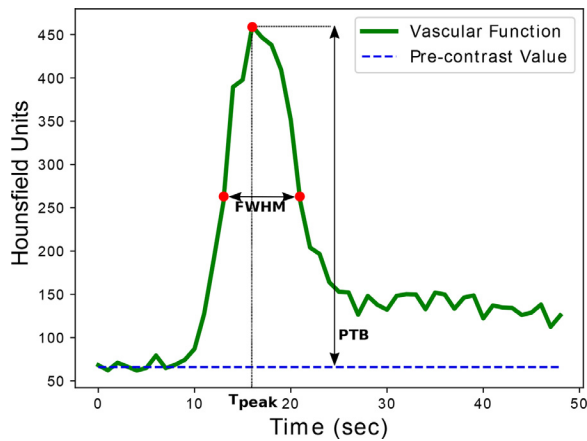


Fig. 3. Vascular function metrics. FWHM: Full width at half maximum; PTB: Peak to baseline; T_{peak} : Time at which the curve peak occurs.

3.2. Performance assessment

In order to evaluate the performance of AIFNet, we conduct a 5-fold (train 70%, validation 10%, test 20%) cross-validation experiment using the annotations of rater #1. All training and testing cases of the ISLES18 database are used in this experiment. In each fold, the train set is used to parametrize the network, the validation set to apply an early-stopping criterion (with the aim of avoiding overfitting) and the test set is independently used for predicting unseen cases. For an in-depth evaluation of the proposed method, results are assessed at a signal, parametric map and lesion quantification level.

3.2.1. Vascular function

Since there is no ground truth for the vascular functions we compare the predictions $\hat{y}(t)$ against the manual annotations $y(t)$ of both the raters (from now on, we specifically refer to our method as $\hat{y}_{\text{AIFNet}}(t)$, to rater #1 as $y_{r1}(t)$ and to rater #2 as $y_{r2}(t)$). The agreement between $y(t)$ and $\hat{y}_{\text{AIFNet}}(t)$ is only computed over the time domain since the AIF absolute contrast values rely on the VOF rescaling (see sections 2.3 and 2.4.1). To this end, we measure the time at which the curve peak occurs (namely T_{peak}), which should indicate potential time shifts of the predictions with respect to ground truth. Moreover, as a measure of the function's width, we quantify the full-width at half-maximum (FWHM) interval. FWHM points are preferred to the curve's onset/offset since these are more difficult to measure and in many cases the offset point is missing. Besides, for evaluating the VOF recalibration strategy, we assess the whole signal agreement using mean squared error and the measured PTB values. All signal metrics are illustrated in Fig. 3.

3.2.2. Parameter maps and lesion quantification

Parameter maps (CBF, CBV, MTT, Tmax) are computed using the well known time-delay invariant singular value decomposition deconvolution. The method is the most commonly found in (clinical) software. Since deconvolution is a mathematically ill-conditioned problem, regularization techniques are necessary. We use Tikhonov regularization over the singular values under a Volterra discretization scheme (Sourbron et al., 2007). Absolute and relative parameter maps are computed, where the relative ones are calculated by the voxelwise normalization of the absolute ones with the mean control tissue region value. Control tissue is defined as the region with normal perfusion (i.e., $T_{\text{max}} < 6$ s (Lin et al., 2016)).

To understand the impact of the vascular functions on the perfusion metrics we compare the parameter maps (obtained through

deconvolution) between the automatically and manually annotated vascular functions. The same deconvolution strategy is adopted in both cases. In this way we are sure that variations are only due to the vascular functions. The assessment of the method in terms of lesion quantification is conducted by comparing i) the hypoperfused and core masks obtained from CTP with automatically annotated vascular functions against the ones obtained using the curves labeled by the two experts and ii) by comparing the obtained core masks against the ISLES18 DWI masks. Note that this latter comparison is only done for the ISLES18 training set, since DWI lesion masks are not available for the test set. Hypoperfused tissue was defined as brain tissue with $T_{\text{max}} > 6$ s (Lin et al., 2016) and core tissue was defined within the hypoperfused area as $\text{rCBF} < 38\%$ (a cutoff previously found to be optimal by Cereda et al. (2016)).

3.2.3. Comparison with other methods

Besides the comparison with the two manual raters, our method is compared with an unsupervised clustering AIF selection approach (Mouridsen et al., 2006) and with two similar CNNs: a regression and a segmentation network, both of them modified versions of AIFNet. For having comparable deep learning approaches, we keep the network's architecture and configuration as close as possible to AIFNet. The same perfusion-specific data augmentation of Section 2.3.1 is used for both the segmentation and regression networks and a unitary batch size is used. The networks are tested following the same 5-fold cross-validation experiments used for AIFNet by assuring that for all the models the same train-validation-test splits are preserved.

Unsupervised clustering AIF selection We use an in-house reimplementation of the unsupervised clustering approach of (Mouridsen et al., 2006). Firstly, non-arterial voxels are discarded using the area under the curve (with a threshold set at the 90th percentile of candidates) and the roughness (defined as $\int_0^T [C''(t)]^2 dt$, with a threshold set at the 25% most irregular candidates) of each voxels' time curve $C(t)$. Secondly, a two-steps K-means (with 5 clusters) is performed for separating other tissues and venous voxels from the arterial cluster. In each K-means iteration, the arterial cluster is selected as the one having lowest first moment of its mean curve. The final AIF is the mean curve with lowest first moment obtained after applying K-means twice.

Regression AIFNet The regression CNN has been introduced by our group in (de la Rosa et al., 2020). It consists of a 3D + 2D neural network equipped with six convolutional layers with average pooling and with two fully connected layers at the end. The last fully connected layer is a 1D vector with same number of neurons as time points in the perfusion CTP and represents the vascular function prediction $\hat{y}_{\text{Reg}}(t)$. The 3D to 2D data transformation in the network is conducted by squeezing the z-axis information by means of average pooling. Homogenizing the z-axis dimension is required for dealing with the variable CTP coverage (varying between 2 and 16 slices per scan). The 3D convolution kernels have dimension $3 \times 3 \times 3$ and the 2D convolution ones have dimension 3×3 . Unlike the original work where optimization was carried out with a segmentation loss for the core tissue, in this work the Pearson's correlation coefficient is preserved as loss function. The network is optimized with RMSprop gradient descent (Hinton et al., 2012).

Segmentation AIFNet This network is fed with AIF binary masks as ground truth. It is similar to AIFNet by preserving the whole architecture except the voxelwise multiplication and 3D average pooling blocks (see Fig. 2). Besides, the last convolutional block (L_{out}) has two kernels followed by a softmax operation for conducting background and foreground segmentation. For compensating the large class imbalance, this network is trained using weighted categorical cross-entropy as loss function and is optimized using stochastic gradient descent with momentum. The AIF is then esti-

Table 1

AIF agreement among methods and rater #1 for the AIF signals as a whole, as well as for their T_{peak} and FWHM parameters. Pearson's correlation coefficients are computed between pairs of AIF signals; the mean (standard deviation) and 5th - 95th percentile interval are provided. For the T_{peak} and FWHM parameters, Pearson's correlation coefficients and errors (in seconds) are reported across all scans as mean (standard deviation). $y_{r1}(t)$, $y_{r2}(t)$: AIF annotated by raters #1 and #2; $\hat{y}_{Kmeans}(t)$, $\hat{y}_{Seg}(t)$, $\hat{y}_{Reg}(t)$, $\hat{y}_{AIFNet}(t)$: AIF estimated with the K-means approach (Mouridsen et al., 2006), with the regression CNN (de la Rosa et al., 2020), with the segmentation CNN and with AIFNet, respectively; r : Pearson's correlation coefficient; T_{peak} : time at which the peak of the curve occurs; FWHM: full-width at half-maximum. P5th: 5th percentile; P95th: 95th percentile. Paired significance tests are performed between AIFNet and the other approaches. *: p-value < 0.05; †: p-value < 0.01. The values in bold indicate the outperforming method for the metric under consideration.

| | | Signal | | T_{peak} | | FWHM | |
|----------------|---|---------------------|---|-------------------|---------------------|--------------|---------------------|
| | | r | r (P5 th , P95 th) | r | Error [s] | r | Error [s] |
| Inter-rater | | 0.971 (0.075)† | (0.883, 1) | 0.964 | -0.14 (1.29)† | 0.902 | -0.08 (1.74)† |
| $y_{r1}(t)$ vs | $\hat{y}_{Kmeans}(t)$ | 0.610 (0.315)† | (-0.101, 0.955) | 0.678 | -5.37 (5.55)† | 0.369 | -1.95 (6.27)† |
| | $\hat{y}_{Seg}(t)$ | 0.677 (0.15)† | (0.393, 0.897) | 0.851 | -3.95 (3.43)† | 0.587 | -4.43 (5.03)† |
| | $\hat{y}_{Reg}(t)$ | 0.837 (0.260)† | (0.563, 0.986) | 0.740 | -0.30 (3.30) | 0.419 | -2.86 (3.65)† |
| | $\hat{y}_{AIFNet}(t)$ | 0.965 (0.05) | (0.838, 0.997) | 0.940 | -0.55 (1.75) | 0.854 | -0.89 (2.14) |

mated as the average function among the top ranked voxels, such that the AIF Pearson's correlation is maximized.

3.2.4. Statistical analysis

For evaluating the vascular signals, Pearson's correlation coefficients are computed between pairs of signals and across all scans for the considered metrics (i.e. T_{peak} , FWHM and PTB). Mean, standard deviation and (5th, 95th) percentiles are provided. Additionally, to assess a potential bias of the different metrics, we compute the mean and standard deviation of the errors. The assessment of the parameter maps is performed using Pearson's correlation coefficients computed per scan within the brain masks (excluding background, skull, ventricles and vessels). Hypoperfused and core tissue segmentations are evaluated by comparing the CTP masks obtained by the different methods with the CTP masks obtained by the experts. Additionally, for the core tissue we compare the experts and the different methods CTP masks with the ground truth DWI masks from ISLES18. The mean volume error and the mean absolute volume error are used for evaluating lesion volumetric agreement with the ground truth. The Dice coefficient is used as a general segmentation performance metric. In all cases, paired t-tests are performed after visual inspection of the data distributions. Under the presence of non-normal distributions, outliers, or heteroskedasticity, a paired Wilcoxon-test is preferred. The significance level is set in all cases to $\alpha = 0.05$.

4. Results and discussion

All models are trained on a machine with a Tesla K80 Nvidia GPU (12 Gb dedicated), with 64 gb RAM and an Intel Xeon E5-2686 v4 multiprocessor. The training stage takes ~11 hours for an AIF/VOF model. Manual annotations take between 2 and 4 minutes for both functions per scan, depending on the number of slices of the volume. On the other hand, predictions take ~6 seconds per each vascular function per scan.

4.1. Signal agreement

4.1.1. AIF

Table 1 shows a summary of the different methods' performance compared to rater #1. Likewise, the comparison with rater #2 is shown in Table S1 (supplementary materials). The automatic predictions of AIFNet obtain high agreement with both raters in all the metrics considered. There is an overall better agreement with $y_{r2}(t)$, even when the network is trained using the $y_{r1}(t)$ annotations, suggesting good generalization at inter-rater level. When the entire vascular signal is evaluated, the method obtains Pearson's r values reaching the raters range. A slightly lower 5th percentile is observed in the agreement between $\hat{y}_{AIFNet}(t)$ and $y_{r1}(t)$

when compared with the inter-rater agreement. This discordance is, however, not found when comparing $\hat{y}_{AIFNet}(t)$ with $y_{r2}(t)$, which obtains fully overlapping ranges with the inter-rater performance. The 95th percentile obtained between AIFNet and the raters is, as expected, close to $r = 1$ but never reaching perfect agreement, due to the weighted multivoxel selection strategy proposed.

When the method performance is assessed in terms of T_{peak} , a high correlation with the manual annotations is found. It can be observed from the inter-rater comparison that the T_{peak} annotations of $y_{r2}(t)$ are slightly delayed when compared with the ones of $y_{r1}(t)$. The AIF functions that AIFNet selects are on average ~ 0.5 seconds delayed when compared with the raters. This temporal trend toward delayed events explains the slight overall lower agreement between AIFNet and both raters. Similarly, the agreement that is obtained for the FWHM between AIFNet and the raters is slightly lower than the inter-raters level. The predicted FWHM windows are on average ~ 1 second longer than the manual ones. These time differences found in T_{peak} and FWHM with our method are below the temporal CTP resolution (one frame, the minimal possible). The main reason behind these differences is the flip side of the coin of the multivoxel selection strategy. Thus, vascular function estimation based on multiple voxels could not always provide the *earliest* bolus arrival with the *highest* and *narrowest* curves, but averaged values over the activated voxels. Selecting vascular functions with these characteristics is, hence, not always fully possible with our strategy, since generally a single or just a few voxels fulfill these requirements for AIF.

The comparison of the different methods shows that AIFNet has a much better agreement with the raters than the other approaches. While the segmentation CNN slightly outperforms the K-means method, the regression CNN outperforms both the segmentation CNN and the K-means approach (Table 1). The regression CNN not only correlates better with the raters at signal level but also localized with less delay T_{peak} than these two other methods. There are no statistically significant differences in the T_{peak} errors of the regression CNN and the ones of AIFNet. An explanation to this observation could be in the optimized loss function: the regression network, likewise AIFNet, is optimized at the predicted time-curve level instead of at the image spatial level (which is the case for the segmentation CNN). Another observation is that the segmentation CNN (which comprises almost the same architecture as AIFNet but optimized with a segmentation loss) obtained a much worse performance than our proposal. The segmentation CNN fails in localizing properly the AIF peaks, and provides delayed and much wider curves. These results suggest that the used Pearson loss function enhances the task performance by *i*) mostly activating arterial voxels with good AIF curves and by *ii*) selectively discarding suboptimal arterial voxels whose AIFs are delayed, highly noise-corrupted or with poor contrast enhancement.

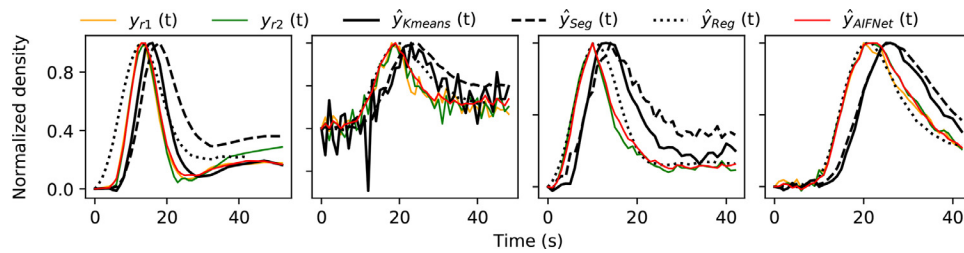


Fig. 4. Predicted arterial input functions (AIF) on diverse quality ISLES18 scans. First image: a noise-free scan. Second image: a head-motion corrupted scan. Third image: an early bolus arrival scan (the pre-contrast increase signal is missing). Fourth image: a truncated perfusion scan (the curve tail is missing). $y_{r1}(t)$, $y_{r2}(t)$: AIF annotated by raters 1 and 2; $\hat{y}_{Kmeans}(t)$, $\hat{y}_{Seg}(t)$, $\hat{y}_{Reg}(t)$, $\hat{y}_{AIFNet}(t)$: AIF estimated with K-means (Mouridsen et al., 2006), with the regression CNN (de la Rosa et al., 2020), with the segmentation CNN and with AIFNet, respectively.

Table 2

VOF agreement between AIFNet and both the raters for the VOF signals as a whole, as well as for their T_{peak} , FWHM and PTB parameters. Pearson's correlation coefficients are computed between pairs of VOF signals; the mean (standard deviation) and 5th - 95th percentile interval are provided. For the T_{peak} , FWHM and PTB parameters, Pearson's correlation coefficients and errors (in seconds for T_{peak} and FWHM and in Hounsfield units for PTB) are reported across all scans as mean (standard deviation). $y_{r1}(t) - y_{r2}(t)$: annotated VOF signals by raters #1 and #2; $\hat{y}_{AIFNet}(t)$: predicted VOF signals with AIFNet; r : Pearson's correlation coefficient; MSE: Mean squared error; HU: Hounsfield units; T_{peak} : time at which the peak of the curve occurs; FWHM: full-width at half-maximum; PTB: Peak-to-baseline; P5th: 5th percentile; P95th: 95th percentile. A paired significance test is performed between the inter-rater results and the AIFNet-raters ones. *: p-value < 0.05; †: p-value < 0.01.

| | | Inter-rater | $y_{r1}(t)$ vs $\hat{y}_{AIFNet}(t)$ | $y_{r2}(t)$ vs $\hat{y}_{AIFNet}(t)$ |
|------------------------------|---|---------------|--------------------------------------|--------------------------------------|
| Signal | r | 0.985 (0.047) | 0.981 (0.069) | 0.983 (0.051) |
| | r (P5 th , P95 th) | (0.944, 1) | (0.914, 0.999) | (0.925, 0.999) |
| | MSE [HU] | 1424 (3622) | 1235 (2623) | 1558 (3740) |
| | MSE (P5 th , P95 th) | (0, 7148) | (17, 7024) | (25, 8213) |
| T_{peak} | r | 0.980 | 0.955 | 0.963 |
| | Error [s] | 0.27 (1.14) | -0.07 (1.69)† | -0.33 (1.51)† |
| FWHM | r | 0.829 | 0.827 | 0.911 |
| | Error [s] | 0.12 (2.28) | -0.04 (2.42) | -0.15 (1.74) |
| PTB | r | 0.921 | 0.953 | 0.919 |
| | Error [HU] | 11 (55) | 9 (44) | -2 (58) |

In Fig. 4, AIF predictions with the different methods over different quality scans are shown. As it can be seen, AIFNet closely follows the manual rater annotations even under challenging scenarios.

4.1.2. VOF

In Table 2 a summary of the performance of our method for VOF estimation is shown. A high agreement with the manual annotations is obtained, which is better than the performance obtained for the AIF estimation. These results can be expected since VOF compared to AIF is less affected by partial volume effect, has higher SNR and hence provides lower inter-rater variability (Table 2).

For all the considered metrics excepting T_{peak} there are no statistically significant differences between the inter-rater agreement and the AIFNet vs raters agreement. When the entire VOF signals are considered, a high correlation with the manual annotations is achieved, reaching inter-rater variability ranges. In terms of T_{peak} , a good performance is obtained though the same delaying effect previously described for AIF is found. In this case, however, the delays are within the inter-rater range. For FWHM, the agreement between our method and rater #2 is much higher than among raters. Unlike the AIF analysis, it is worth noticing that there is no flattening or widening of the VOF curves predicted with AIFNet. The evaluation of the recalibration strategy using the curve mean squared error and the PTB metric shows a high agreement and high correlation between AIFNet and the manual annotations reaching inter-rater ranges. In the assessment of the mean squared error, the inter-rater's 5th percentile is zero, which implies that the raters have sometimes selected the exact same voxel. The evaluation of the PTB times shows no clear trend of our method towards

under/over-estimation of the VOF signals, suggesting a good overall performance of the recalibration strategy.

4.1.3. Arterial localization

The anatomical localization that AIFNet conducts can be assessed from the voxelwise activation encoded in P_{vol} . Unlike most AIF selection approaches selecting only few candidates, AIFNet allows multiple voxel contribution for building the vascular functions.

In Fig. 5 the best and worst AIF (in correlations terms) among all predictions are shown. While the prediction with higher agreement achieves a Pearson's $r = 0.999$ (left-side of the figure), the case with poorest agreement achieves an $r = 0.674$ (right-side of the figure). Both raters have chosen the same AIF voxel in the best performance scenario. In the top-left part of Fig. 5 it can be seen that just a few voxels are activated in the displayed CT slice, having high activation values. The AIF voxel selected by the raters ($y_{r1,r2}(t)$) is also being activated by AIFNet, being the second highest value of P_{vol} . Mainly voxels belonging to the anterior cerebral artery are chosen. Besides, the AIF that our method predicts follows closely the raters' function, with no observable delays and with almost no differences in the curves' shape. On the other hand, localization results from the worst Pearson's correlation case shows a different behaviour. Several voxels belonging to different arteries are enhanced by the network with a homogeneous activation distribution. The anterior cerebral artery and middle cerebral arteries are mainly selected. When assessing $y_{r1}(t)$ and $\hat{y}_{AIFNet}(t)$, it is noticeable that the low Pearson's r is driven by the time shift between the functions (which is 4 seconds measured at the curve peaks). In this case, AIFNet outperforms rater #1 by estimating a vascular function with high agreement in morphology, which oc-

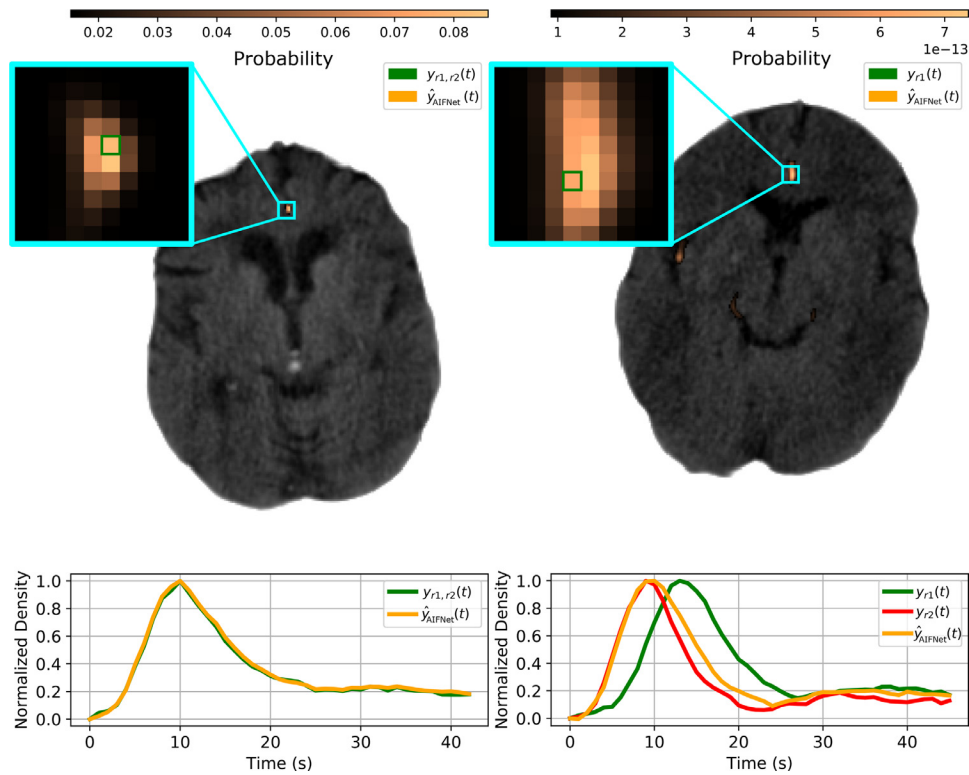


Fig. 5. Best (left) and worst (right) prediction performance in terms of Pearson's correlation between AIF functions. Above, the voxels selected by rater #1 ($y_{r1}(t)$), rater #2 ($y_{r2}(t)$) and AIFNet ($\hat{y}_{AIFNet}(t)$) as arterial input functions. Below, their corresponding vascular function. Note that in the best performance case, both raters have chosen the exact same voxel as AIF. In the worst performance case, the selected voxel location for $y_{r2}(t)$ is not shown since was annotated in a different volume slice.

Table 3

Parameter maps agreement among methods in terms of Pearson's correlation. Mean (standard deviation) values are provided. Correlation has been computed per scan using all the voxels within the brain tissue (excluding background, skull, ventricles and vessels). y_{r1} : Parameter maps obtained after deconvolving the images with the AIF of rater 1; \hat{y}_{Kmeans} , \hat{y}_{Seg} , \hat{y}_{Reg} , \hat{y}_{AIFNet} : Parameter maps obtained after deconvolving the images with the AIF predicted with K-means (Mouridsen et al., 2006), with the regression CNN (de la Rosa et al., 2020), with the segmentation CNN and with AIFNet, respectively. CBF: cerebral blood flow; CBV: cerebral blood volume; MTT: mean transit time; T_{max} : time to the maximum of the residue function. The values in bold indicate the outperforming method (in terms of Pearson's r) for each parameter map.

| | | Pearson's r coefficient | | | |
|-------------|--------------------|-------------------------|----------------------|----------------------|----------------------|
| | | CBF | CBV | T_{max} | MTT |
| Inter-rater | | 0.998 (0.016) | 0.987 (0.157) | 0.944 (0.086) | 0.927 (0.204) |
| y_{r1} vs | \hat{y}_{Kmeans} | 0.967 (0.098) | 0.960 (0.274) | 0.786 (0.217) | 0.772 (0.291) |
| | \hat{y}_{Seg} | 0.974 (0.118) | 0.947 (0.315) | 0.781 (0.237) | 0.758 (0.310) |
| | \hat{y}_{Reg} | 0.990 (0.027) | 0.972 (0.160) | 0.809 (0.179) | 0.791 (0.253) |
| | \hat{y}_{AIFNet} | 0.998 (0.007) | 1.000 (0.003) | 0.921 (0.094) | 0.908 (0.188) |

curs much earlier than the manually selected one. We consider the annotation of rater #1 suboptimal, probably because the voxel was chosen from an artery branch already affected by the occlusion. However, our prediction follows more closely $y_{r2}(t)$ (Pearson's $r = 0.980$). There is no observable function delay between $y_{r2}(t)$ and $\hat{y}_{AIFNet}(t)$, though a slightly wider FWHM can be appreciated for $\hat{y}_{AIFNet}(t)$.

4.2. Parameter maps and lesion quantification

The parameter maps correlation between rater #1 and the different approaches is shown in Table 3. Similar results are obtained when comparing the parameter maps with rater #2 (Table S2). Correlation values are computed for each scan within the brain tissue (excluding background, skull, ventricles and vessels). Among all the compared methods AIFNet obtains the best agreement with the rater for each of the parameter maps, showing consistency with the experts' results. An outstanding agreement is observed between the raters and AIFNet for CBF and CBV, reaching inter-

rater performance. For T_{max} and MTT, however, the agreement is still high but marginally under the inter-rater performance. We hypothesize that the found lower consistency in these parameter maps could be driven by the ~ 0.5 seconds delay in T_{peak} and by the ~ 1 second wider FWHM of AIFNet predictions. It is worth to notice that the other automatic AIF selection methods also obtained better performance for CBF and CBV than for T_{max} and MTT. Parameter maps and lesion masks obtained with all methods are shown in Fig. 6 for the scan with median AIF Pearson correlation (y_{r1} vs \hat{y}_{AIFNet} comparison). There is a high qualitative correspondence between raters and AIFNet at all levels. It is also seen here that there is a better correspondence between rater #1 and all automatic AIF selection methods for estimating rCBF than estimating T_{max} .

In Fig. 7 the hypoperfused and core volumes between rater #1 and all the methods are shown. Likewise, the methods' agreement with rater #2 follows a similar pattern (please see the Supplementary material, Fig. S1). In Table 4 (Table S3), the lesion volumes quantification performance is shown for all the methods when

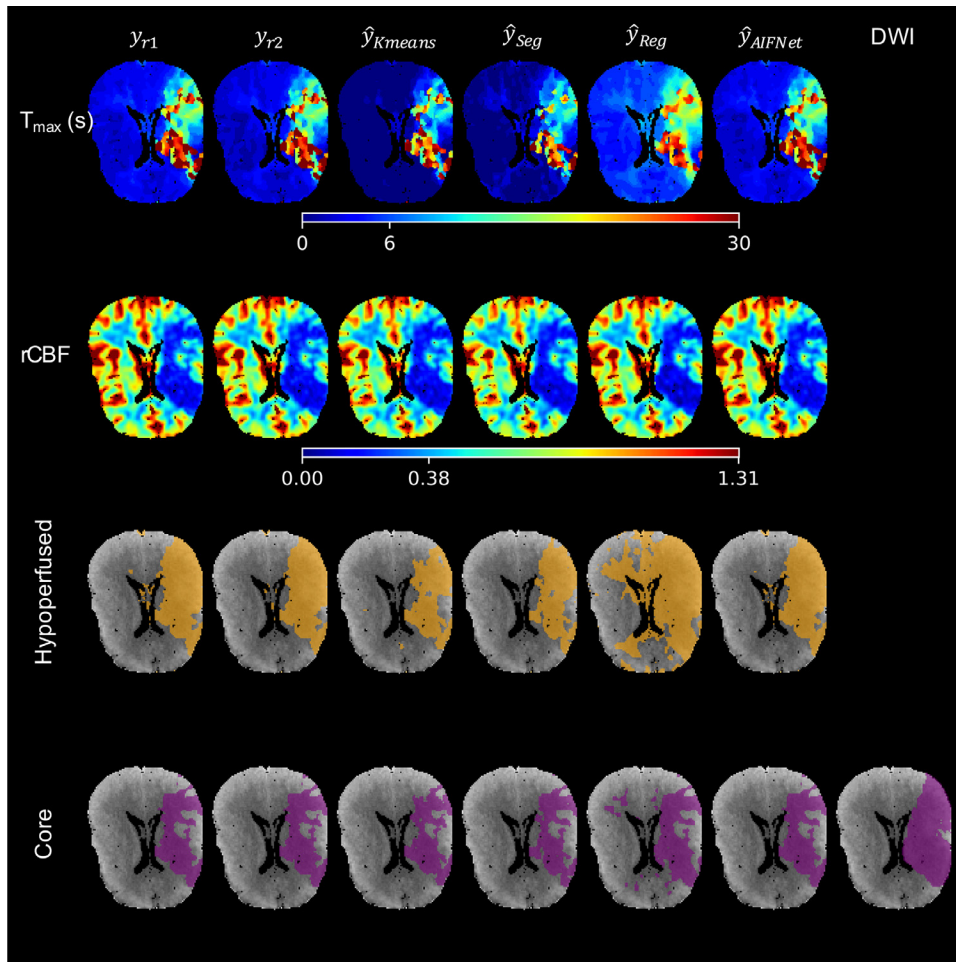


Fig. 6. Estimated parameter maps and brain lesions for all the methods, obtained after deconvolving the CTP images with the annotated or predicted vascular functions. The shown example is the scan with median AIF Pearson correlation (y_{r1} vs \hat{y}_{AIFNet} comparison). Hypoperfused tissue is obtained after thresholding $T_{max} < 6$ s. Core tissue is obtained after thresholding the rCBF map at 38% over the entire hypoperfused region. y_{r1} , y_{r2} : Results obtained after deconvolving the images with the AIF of rater 1 and rater 2; \hat{y}_{Kmeans} , \hat{y}_{Seg} , \hat{y}_{Reg} , \hat{y}_{AIFNet} : Results obtained after deconvolving the images with the AIF predicted with K-means (Mouridsen et al., 2006), with the regression CNN (de la Rosa et al., 2020), with the segmentation CNN and with AIFNet, respectively. rCBF: relative cerebral blood flow map; T_{max} : time to the maximum of the residue function map. DWI: Ground truth delineated over DWI in ISLES18.

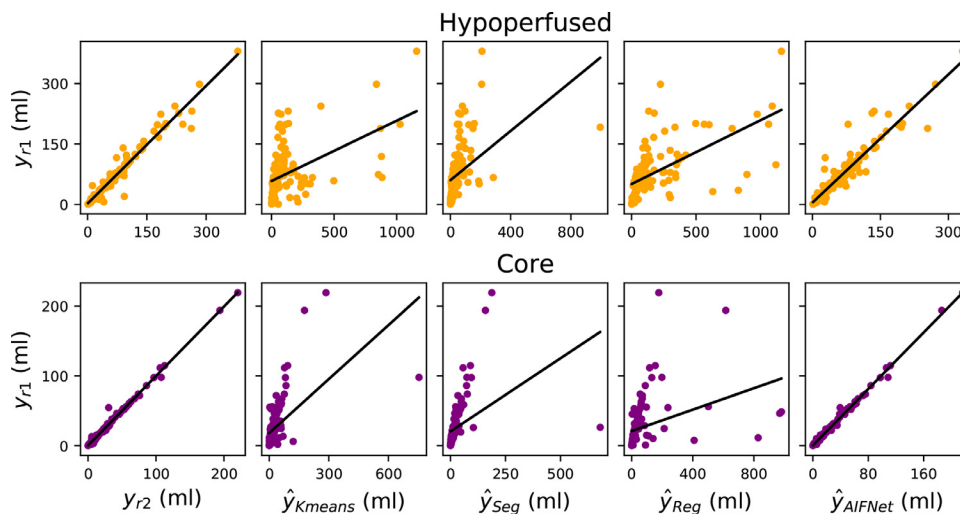


Fig. 7. Lesion volume agreement among methods. y_{r1} , y_{r2} : Lesion volumes obtained after deconvolving the images with the AIF of rater 1 and rater 2; \hat{y}_{Kmeans} , \hat{y}_{Seg} , \hat{y}_{Reg} , \hat{y}_{AIFNet} : Lesion volumes obtained after deconvolving the images with the AIF predicted with K-means (Mouridsen et al., 2006), with the regression CNN (de la Rosa et al., 2020), with the segmentation CNN and with AIFNet, respectively.

Table 4

Brain lesion quantification performance for all the methods, obtained after deconvolving the CTP images with the manual and automatic vascular functions. Mean (standard deviation) values are provided. Hypoperfused tissue is obtained after thresholding $T_{\max} < 6$ s. Core tissue is obtained after thresholding the rCBF map at 38% over the entire hypoperfused region. y_{r1} : Lesion volumes obtained after deconvolving the images with the AIF of rater 1; \hat{y}_{Kmeans} , \hat{y}_{Seg} , \hat{y}_{Reg} , \hat{y}_{AIFNet} : Lesion volumes obtained after deconvolving the images with the AIF predicted with K-means (Mouridsen et al., 2006), with the regression CNN (de la Rosa et al., 2020), with the segmentation CNN and with AIFNet, respectively. DWI: Agreement obtained when comparing the different CTP approaches with the diffusion weighted imaging ground truth provided in ISLES18. VE: Volume error; AVE: Absolute volume error. A paired significance test is performed between AIFNet and the other approaches.*: p-value < 0.05; †: p-value < 0.01. The values in bold indicate the outperforming method for the metric under consideration.

| | | Hypoperfused | | | Core | | |
|-------------|--------------------|--------------------|-------------------|--------------------|-----------------------|---------------------|--------------------|
| | | Dice [%] | VE [ml] | AVE [ml] | Dice [%] | VE [ml] | AVE [ml] |
| Inter-rater | | 91.7 (13.8)† | 0.5 (13.5)† | 6.3 (12.0)† | 91.8 (14.1)† | 0.1 (2.5)† | 0.9 (2.3)† |
| y_{r1} vs | \hat{y}_{Kmeans} | 48.3 (28.2)† | -36.9 (174.3) | 82.7 (157.8)† | 58.3 (27.0)† | -2.6 (55.2)† | 13.3 (53.6)† |
| | \hat{y}_{Seg} | 51.3 (19.4)† | 26.5 (80.8)† | 45.5 (71.8)† | 61.9 (20.7)† | 1.7 (53.9)† | 12.0 (52.6)† |
| | \hat{y}_{Reg} | 70.1 (23.4)† | -78.5 (198.2)† | 94.2 (191.3)† | 72.3 (23.3)† | -32.0 (136.4) | 35.7 (135.5)† |
| | \hat{y}_{AIFNet} | 87.3 (13.0) | 8.8 (20.3) | 12.7 (18.2) | 88.3 (13.6) | 0.5 (2.3) | 1.3 (1.9) |
| y_{r1} | | | | | 38.3 (19.4) | 6.8 (20.0) | 14.3 (15.6) |
| y_{r2} | | | | | 38.3 (19.5) | 6.6 (20.2)† | 14.2 (15.7) |
| DWI vs | \hat{y}_{Kmeans} | | | | 31.5 (19.9)† | 1.2 (72.9)† | 25.1 (68.4)† |
| | \hat{y}_{Seg} | | | | 32.7 (19.0)† | 12.7 (22.2)† | 16.4 (19.6)* |
| | \hat{y}_{Reg} | | | | 35.1 (19.8)* | -34.2 (154.3)† | 52.6 (149.0)† |
| | \hat{y}_{AIFNet} | | | | 38.1 (19.5) | 7.2 (20.5) | 14.5 (16.2) |

compared with rater #1 (#2). AIFNet consistently outperforms the other approaches (i.e., with statistically significant comparisons for almost all the core and hypoperfused metrics) and reaches a high agreement with the experts slightly below the inter-rater performance (there are statistically significant differences between raters and AIFNet as well). While the agreement with the raters for the core volumes is very high, a slight bias in the hypoperfused volumes can be appreciated, suggesting that AIFNet tends to underestimate these tissue areas. This bias is a consequence of the AIF differences obtained with AIFNet (described in Section 4.1) that impact over T_{\max} .

Furthermore, Table 4 compares the predicted core volumes with the ground truth DWI core masks. In this comparison, AIFNet closely follows the manual raters performance: except the volume error comparison between AIFNet and rater #2, there are no statistically significant differences with the raters in the quantified metrics. Our results suggests that AIFNet estimates the acute brain infarcts as good as manual raters do. The core volume correlation between the different CTP methods and the DWI ground truth is depicted in Figure S2. When comparing the core CTP volumes of all the investigated approaches (raters and automatic algorithms) with the DWI ones, there are statistically significant differences for all the methods (p-values < 0.01, Wilcoxon test). Nonetheless, this cross-modality (i.e. CTP-DWI) performance evaluation should be carefully judged, since external source of errors are being introduced (such as brain perfusion changes due to time acquisition differences, modality co-registration errors, etc). These limitations in cross-modality correspondences explain the low overall Dice coefficients even found for the raters in the CTP-DWI comparison.

Overall, the parameter maps and lesion quantification analysis show that the small AIF differences found with AIFNet (mainly in T_{peak} and FWHM) do not produce a large impact over the deconvolution process. The main observation related to these AIF differences is the bias found in the hypoperfused volumes, showing a trend in AIFNet to slightly underestimate this tissue. There are no important differences in the core predictions of AIFNet in comparison to the ones of the manual raters. There are neither differences in the CTP-DWI core agreement: the raters and AIFNet are both equally consistent compared to the ISLES18 DWI ground truth. Overall, our analysis of the vascular functions, of the parameters maps and lesion volumes obtained after CTP deconvolution show that AIFNet obtains state-of-the-art performance in automatic AIF selection. The method almost behaves as a manual expert in all

the considered deconvolution stages and its performance is close to the inter-rater one.

4.3. Comparison with other methods

The comparison of different AIF selection methods shows that our innovative CNN is the most suitable approach among the investigated ones for performing the task. Unlike AIFNet, the other automatic algorithms predict AIFs that consistently compromise the parameter maps quality and, hence, the estimation of the brain lesions. We also observe that our method can robustly work under challenging perfusion cases, as shown for different quality scans in Fig. 4.

The K-means method (Mouridsen et al., 2006) shows several limitations and provides a large performance variability among the computed AIFs (large standard deviations of the AIF metrics). These results could be expected since the method has been originally devised for perfusion MRI, which accounts with higher signal-to-noise ratio and smoother perfusion curves than CTP. As such, the AIF unsupervised clustering might better perform in perfusion MRI than in CTP. The results from the segmentation CNN show similar performance to the K-means method and, as such, worse than the AIFNet performance. Our approach not only shows to be quantitatively better than the segmentation CNN, but also provides the advantage of requiring minimal data labels: while classification or segmentation approaches would ideally annotate most of the target class samples for their training, AIFNet can be trained with a single voxel annotation. It is worth reminding the reader that the only difference between the segmentation CNN and AIFNet is the optimized loss function. The segmentation network works at a spatial level by recognising similar anatomical/shape areas to the target class, but it fails in selecting good AIF candidate voxels. We show that the AIF selection task is consistently improved when the network learns from the predicted AIF (as happens with the proposed Pearson correlation loss) rather than from the selected voxels only, as happens when using segmentation loss functions. While arterial voxels can be extremely similar in terms of density values, localization and/or anatomical context, their perfusion curves can differ significantly. As such, segmentation-based methods can misleadingly activate arterial voxels whose perfusion curves are suboptimal. These findings are also supported by the better results obtained with the regression CNN (whose training loss is also the Pearson correlation, as for AIFNet) compared to

the segmentation CNN. Nonetheless, the regression CNN results are much worse than the ones of our method. Moreover, fully AIF regression CNNs as earlier proposed in (de la Rosa et al., 2020) have some drawbacks. First, from a qualitative point of view, the model does not provide voxelwise arterial localization, making it less interpretable than other approaches and less transferable to clinical settings. Second, the model requires that the scans' duration is homogeneous. As such, truncation artifacts might be introduced during this process (Kasasbeh et al., 2016). It is interesting to observe that AIFNet overcomes the segmentation and regression CNN limitations by working as a hybrid segmentation-regression network. Its architecture allows voxelwise activation and selection like most segmentation approaches, but its optimization is performed like a regression CNN, which shows to enhance the performance. In other words, AIFNet explores the advantages of segmentation and regression models for improving the AIF estimation.

Another successful deep learning approach for AIF selection has been presented by Winder et al. (2020). Similar to AIFNet, the approach achieves close agreement with the manual raters in the selection of AIF and in the estimated perfusion lesions. The approach differs from AIFNet in several ways. First, the CNN is a 1D model receiving as input single voxel perfusion curves (i.e., there is no spatial/contextual information considered but only temporal information). Second, the method is a binary classifier CNN (AIF vs all-the-rest) and hence it is optimized with a classification loss function. Third, the method estimates the AIF by means of a sophisticated geometric averaging approach. It is worth noting that both the methods (AIFNet and (Winder et al., 2020)) avoid delayed AIF curves in different ways. On one hand, (Winder et al., 2020) isolates potential AIF candidates with the classification CNN and then corrects the time curves with the geometric averaging technique. On the other hand, our proposal discards delayed AIF candidates by restricting the CNN learning through a suitable loss function. Some advantages on the usage of AIFNet over the work of Winder et al. (2020) are *i*) the number of manual annotations required for the model training (around < 20 times labeled data), *ii*) almost no pre-processing required, *iii*) the automatic selection of the number of voxels to average (which is a parameter to set in (Winder et al., 2020)) and *iv*) its full end-to-end framework, which makes our approach easier and faster to deploy.

4.4. Ablation analysis

We ablate our network for finding the optimal architecture and training strategy for computing $\hat{y}_{AIFNet}(t)$. The ablation is conducted for the AIF since it is the most critical input to the deconvolution model and it is much more difficult to estimate than the VOF. These experiments are performed using the original train-test data split of the ISLES18 challenge ($n_{train} = 94$, $n_{test} = 62$). For training purposes we randomly exclude 10% of the training data and use it as validation set, assuring that in all the ablation experiments the same train-validation-test sets are used.

Results for the ablation analysis are shown in Table 5. Our experiments show that $K = 5$ convolutional layers are optimal for AIF prediction. The usage of less convolutional blocks leads not only to lower mean performance but also to higher variability. Besides, results do not improve when considering more than $K = 5$ convolutional layers. It is worth to point out the considerable improvement in robustness when problem-specific data augmentation is considered for training the models. Overall, a much higher 5th percentile is obtained with rather than without data augmentation, showing better generalization over challenging cases. For VOF prediction, $K = 2$ convolutional layers are enough to estimate the function at inter-rater performance. Thus, less features are required for finding good VOF voxel candidates. These results are expected given the higher task difficulty for selecting AIF over VOF, as shown in

Table 5

AIFNet ablation performance. K : Number of convolutional layers in the CNN. Given GPU memory constrains, the AIFNet experiment with $K = 6$ is conducted with 2^{2+k} kernels per layer instead of 2^{3+k} (such as the first layer has 8 kernels and the sixth one 256). DA: Data augmentation; std: Standard deviation; Perc: Percentile. The values in bold indicate the outperforming approach for the metric under consideration.

| AIFNet #Layers (K) | | | | | Pearson's r | |
|--------------------|---|---|---|----|----------------------|-----------------------|
| 3 | 4 | 5 | 6 | DA | Mean (std) | (5th, 95th Perc) |
| x | | | | | 0.943 (0.133) | (0.661, 0.999) |
| | x | | | | 0.947 (0.107) | (0.669, 0.999) |
| | | x | | | 0.950 (0.088) | (0.694, 0.999) |
| | | | x | | 0.946 (0.094) | (0.682, 0.999) |
| | | | | x | 0.957 (0.057) | (0.870, 0.999) |

Tables 1 and 2, where a better agreement between raters is shown for VOF than for AIF.

4.5. Limitations and future perspectives

A limitation of this work is the lesion ground truth used, since currently there is no gold standard for the penumbra and for the ischemic core. We use, as provided in ISLES18, the core masks delineated in DWI. However, the acquisition delay between CTP and DWI imaging may introduce ischemic core modifications. Another source of mismatch between the imaging modalities might be introduced by the reperfusion therapy, since reversal of the DWI lesion may happen after reperfusion (Campbell et al., 2012). Even more, mismatch errors could also appear during the cross-modality image registration. Consequently, a full correspondence between CTP and DWI core lesions is unlikely to happen. In our experiments, this mismatch could explain the statistically significant differences found when comparing all CTP volumetric core predictions (from raters and AIFNet) against the delineated DWI ground truth. For a better understanding of the different methods' performance, we include a CTP-CTP analysis by considering as hypoperfused and core ground truth the volumes obtained by using the manual rater annotations in the CTP deconvolution. In such a way, the aforementioned cross-modality limitations are no longer affecting the analysis.

As future perspectives for this work we could consider the validation of AIFNet over a larger database, as well as over other imaging modalities, such as perfusion MRI and PET images. Given the challenging task behind vascular estimation over CTP, we expect the method to be easy to adapt to images of better quality (such as MRIs). Exploring whether the technique is generalizable to other organs and pathologies where perfusion analysis is used (such as in brain tumors, myocardial infarction, etc.) also constitutes potential research lines.

5. Conclusions

We have presented AIFNet, a new automatic method for vascular function estimation in brain perfusion imaging. It is developed and validated over the public ISLES18 database, which consists of stroke perfusion CT cases. To our knowledge, this is one of the first automatic methods described in literature fully developed and validated over perfusion CT data. Most of the approaches previously described have been devised and tested over perfusion MRI instead. For tackling the problem, we make use of a fully end-to-end trainable CNN, that is optimized for the prediction of vascular functions. We exploit prior knowledge by performing modality-specific data augmentation during the training stage. Our approach consistently differs from the previous ones,

which mainly rely on clustering or statistical techniques. Additionally, most of these techniques require the definition of a decision rule (mainly a cutoff) for selecting the optimal voxels, a strategy that might be dataset-dependent and, hence, requires parameters tuning. Unlike these methods, we present a non-heuristic function estimation strategy that combines information from multiple voxels by means of a 3D probabilistic volume. AIFNet allows arterial localization and, hence, clinical interpretability. The method is easier to train and deploy compared to other approaches due to its architecture and due to the minimal voxel annotations required as ground truth (one single voxel per vascular function and per scan is enough to parametrize the network). As a consequence, the database labeling process is very fast. This is a clear advantage of AIFNet when compared against segmentation approaches, since the latter are more time consuming by requiring a vessel region annotation and multiple vascular functions checks. We show, as well, that using a suitable loss function enhances the task performance. After validating AIFNet in the ISLES18 dataset, the method outperformed existing methods and achieved results close to the inter-rater agreement, being able to make predictions of vascular functions, parameter maps and perfusion lesions with similar performance as human experts. Besides, the approach shows to be robust under poor quality-scan scenarios. Our results suggests that AIFNet could be implemented in clinical scenarios and, hence, could potentially be included in future brain perfusion deconvolution software. For better reproducibility and direct comparison against future methods, we provide both raters' vascular annotations as supplementary material.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

EdIR and DR are co-inventors in technology related to this research; a patent application has been submitted and is pending. EdIR, DMS and DR are employees of **icometrix**.

CRedit authorship contribution statement

Ezequiel de la Rosa: Data curation, Investigation, Methodology, Software, Writing – original draft. **Diana M. Sima:** Supervision, Writing – review & editing, Funding acquisition. **Bjoern Menze:** Conceptualization, Methodology, Writing – review & editing. **Jan S. Kirschke:** Conceptualization, Writing – review & editing. **David Robben:** Data curation, Formal analysis, Supervision, Conceptualization, Writing – review & editing.

Acknowledgement

This project received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement TRABIT No 765148. DR is supported by an innovation mandate of Flanders Innovation & Entrepreneurship (VLAIO). BM received funding from the project "Stroke treatment goes personalized: Gaining added diagnostic yield by computer-assisted treatment selection" (Deutsche Forschungsgemeinschaft (DFG) - Projekt number 326824585).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at [10.1016/j.media.2021.102211](https://doi.org/10.1016/j.media.2021.102211)

References

- Abulnaga, S.M., Rubin, J., 2018. Ischemic stroke lesion segmentation in CT perfusion scans using pyramid pooling and focal loss. In: International MICCAI Brainlesion Workshop. Springer, pp. 352–363.
- Albers, G.W., Goyal, M., Jahan, R., Bonafe, A., Diener, H.-C., Levy, E.I., Pereira, V.M., Cognard, C., Cohen, D.J., Hacke, W., et al., 2016. Ischemic core and hypoperfusion volumes predict infarct size in SWIFT PRIME. *Ann. Neurol.* 79 (1), 76–89.
- Bertels, J., Robben, D., Vandermeulen, D., Suetens, P., 2018. Contra-lateral information CNN for core lesion segmentation based on native CTP in acute stroke. In: International MICCAI Brainlesion Workshop. Springer, pp. 263–270.
- Calamante, F., 2013. Arterial input function in perfusion MRI: a comprehensive review. *Prog Nucl Magn Reson Spectrosc* 74, 1–32.
- Calamante, F., Mørup, M., Hansen, L.K., 2004. Defining a local arterial input function for perfusion MRI using independent component analysis. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 52 (4), 789–797.
- Campbell, B.C.V., Parsons, M.W., 2018. Imaging selection for acute stroke intervention. *International Journal of Stroke* 13 (6), 554–567.
- Campbell, B.C.V., Purushotham, A., Christensen, S., Desmond, P.M., Nagakane, Y., Parsons, M.W., Lansberg, M.G., Mlynash, M., Straka, M., De Silva, D.A., et al., 2012. The infarct core is well represented by the acute diffusion lesion: sustained reversal is infrequent. *Journal of Cerebral Blood Flow & Metabolism* 32 (1), 50–56.
- Cereda, C.W., Christensen, S., Campbell, B.C.V., Mishra, N.K., Mlynash, M., Levi, C., Straka, M., Wintermark, M., Bammer, R., Albers, G.W., et al., 2016. A benchmarking tool to evaluate computer tomography perfusion infarct core predictions against a DWI standard. *Journal of Cerebral Blood Flow & Metabolism* 36 (10), 1780–1789.
- Clèrigues, A., Valverde, S., Bernal, J., Freixenet, J., Oliver, A., Lladó, X., 2019. Acute ischemic stroke lesion core segmentation in CT perfusion images using fully convolutional neural networks. *Comput. Biol. Med.* 115, 103487.
- Fan, S., Bian, Y., Wang, E., Wang, D.J.J., Yang, Q., Ji, X., Kang, Y., 2019. An automatic estimation of arterial input function based on multi-stream 3D CNN. *Front Neuroinform* 13, 49.
- Fieselmann, A., Kowarschik, M., Ganguly, A., Hornegger, J., Fahrigr, R., 2011. Deconvolution-based CT and MR brain perfusion measurement: theoretical model revisited and practical implementation details. *Journal of Biomedical Imaging* 2011, 14.
- Force, W.T., 1989. Stroke-1989. Recommendations on stroke prevention, diagnosis, and therapy. Report of the WHO Task Force on Stroke and other Cerebrovascular Disorders. *Stroke* 20 (10), 1407–1431.
- Hinkle, J.L., Guanci, M.M., 2007. Acute ischemic stroke review. *Journal of neuroscience nursing* 39 (5), 285–293.
- Hinton, G., Srivastava, N., Swersky, K., 2012. Neural networks for machine learning—Lecture 6a: Overview of mini-batch gradient descent. https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf/, Accessed: May 5, 2021 [Online].
- Kamnitsas, K., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B., 2017. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal* 36, 61–78.
- Kasasbeh, A.S., Christensen, S., Straka, M., Mishra, N., Mlynash, M., Bammer, R., Albers, G.W., Lansberg, M.G., 2016. Optimal computed tomographic perfusion scan duration for assessment of acute stroke lesion volumes. *Stroke* 47 (12), 2966–2971.
- Kealey, S.M., Loving, V.A., DeLong, D.M., Eastwood, J.D., 2004. User-defined vascular input function curves: influence on mean perfusion parameter values and signal-to-noise ratio. *Radiology* 231 (2), 587–593.
- Khatri, P., Yeatts, S.D., Mazighi, M., Broderick, J.P., Liebeskind, D.S., Demchuk, A.M., Amarenco, P., Carrozella, J., Spilker, J., Foster, L.D., et al., 2014. Time to angiographic reperfusion and clinical outcome after acute ischaemic stroke: an analysis of data from the Interventional Management of Stroke (IMS III) phase 3 trial. *The Lancet Neurology* 13 (6), 567–574.
- Kistler, M., Bonaretti, S., Pfahrer, M., Niklaus, R., Büchler, P., 2013. The virtual skeleton database: an open access repository for biomedical research and collaboration. *J. Med. Internet Res.* 15 (11), e245.
- Klotz, E., König, M., 1999. Perfusion measurements of the brain: using dynamic CT for the quantitative assessment of cerebral ischemia in acute stroke. *Eur J Radiol* 30 (3), 170–184.
- Konstas, A.A., Goldmakher, G.V., Lee, T.-Y., Lev, M.H., 2009. Theoretic basis and technical implementations of CT perfusion in acute ischemic stroke, part 1: theoretic basis. *American Journal of Neuroradiology* 30 (4), 662–668.
- Kosior, J.C., Frayne, R., 2007. PerfTool: a software platform for investigating bolus-tracking perfusion imaging quantification strategies. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 25 (3), 653–659.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, pp. 1097–1105.
- Kudo, K., Sasaki, M., Yamada, K., Momoshima, S., Utsunomiya, H., Shirato, H., Ogasawara, K., 2010. Differences in CT perfusion maps generated by different commercial software: quantitative analysis by using identical source data of acute stroke patients. *Radiology* 254 (1), 200–209.
- Lin, L., Bivard, A., Krishnamurthy, V., Levi, C.R., Parsons, M.W., 2016. Whole-brain CT perfusion to quantify acute ischemic penumbra and core. *Radiology* 279 (3), 876–887.
- Lorenz, C., Benner, T., Chen, P.J., Lopez, C.J., Ay, H., Zhu, M.W., Menezes, N.M., Aro-

- nen, H., Karonen, J., Liu, Y., et al., 2006. Automated perfusion-weighted MRI using localized arterial input functions. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 24 (5), 1133–1139.
- Maier, O., Menze, B.H., von der Gablentz, J., Häni, L., Heinrich, M.P., Liebrand, M., Winzeck, S., Basit, A., Bentley, P., Chen, L., et al., 2017. ISLES 2015-A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. *Med Image Anal* 35, 250–269.
- McKinley, R., Hung, F., Wiest, R., Liebeskind, D.S., Scalzo, F., 2018. A machine learning approach to perfusion imaging with dynamic susceptibility contrast MR. *Front Neurol* 9, 717.
- Meier, R., Lux, P., Jung, S., Fischer, U., Gralla, J., Reyes, M., Wiest, R., McKinley, R., Kaesmacher, J., 2019. Neural network-derived perfusion maps for the assessment of lesions in patients with acute ischemic stroke. *Radiology: artificial intelligence* 1 (5), e190019.
- Mlynash, M., Eyngorn, I., Bammer, R., Moseley, M., Tong, D.C., 2005. Automated method for generating the arterial input function on perfusion-weighted MR imaging: validation in patients with stroke. *American Journal of Neuroradiology* 26 (6), 1479–1486.
- Mouridsen, K., Christensen, S., Gyldensted, L., Østergaard, L., 2006. Automatic selection of arterial input function using cluster analysis. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 55 (3), 524–531.
- Murase, K., Kikuchi, K., Miki, H., Shimizu, T., Ikezoe, J., 2001. Determination of arterial input function using fuzzy clustering for quantification of cerebral blood flow with dynamic susceptibility contrast-enhanced MR imaging. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 13 (5), 797–806.
- Murphy, B.D., Chen, X., Lee, T.-Y., 2007. Serial changes in CT cerebral blood volume and flow after 4 hours of middle cerebral occlusion in an animal model of embolic cerebral ischemia. *American Journal of Neuroradiology* 28 (4), 743–749.
- Østergaard, L., Sorensen, A.G., Kwong, K.K., Weisskoff, R.M., Gyldensted, C., Rosen, B.R., 1996. High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. Part II: Experimental comparison and preliminary results. *Magn Reson Med* 36 (5), 726–736.
- Østergaard, L., Weisskoff, R.M., Chesler, D.A., Gyldensted, C., Rosen, B.R., 1996. High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. Part I: Mathematical approach and statistical analysis. *Magn Reson Med* 36 (5), 715–725.
- Peruzzo, D., Bertoldo, A., Zanderigo, F., Cobelli, C., 2011. Automatic selection of arterial input function on dynamic contrast-enhanced MR images. *Comput Methods Programs Biomed* 104 (3), e148–e157.
- Rausch, M., Scheffler, K., Rudin, M., Radü, E.W., 2000. Analysis of input functions from different arterial branches with gamma variate functions and cluster analysis for quantitative blood volume measurements. *Magn Reson Imaging* 18 (10), 1235–1243.
- Rempp, K.A., Brix, G., Wenz, F., Becker, C.R., Gückel, F., Lorenz, W.J., 1994. Quantification of regional cerebral blood flow and volume with dynamic susceptibility contrast-enhanced MR imaging. *Radiology* 193 (3), 637–641.
- Robben, D., Boers, A.M.M., Marquering, H.A., Langezaal, L.L., Roos, Y.B., van Oostenbrugge, R.J., van Zwam, W.H., Dippel, D.W.J., Majoie, C.B., van der Lugt, A., et al., 2020. Prediction of final infarct volume from native CT perfusion and treatment parameters using deep learning. *Med Image Anal* 59, 101589.
- Robben, D., Suetens, P., 2018. Perfusion parameter estimation using neural networks and data augmentation. In: *International MICCAI Brainlesion Workshop*. Springer, pp. 439–446.
- de la Rosa, E., Robben, D., Sima, D.M., Kirschke, J.S., Menze, B., 2020. Differentiable deconvolution for improved stroke perfusion analysis. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 593–602.
- Shi, J., Malik, J., 2000. Normalized cuts and image segmentation. *Departmental Papers (CIS)* 107.
- Shi, L., Wang, D., Liu, W., Fang, K., Wang, Y.-X.J., Huang, W., King, A.D., Heng, P.A., Ahuja, A.T., 2014. Automatic detection of arterial input function in dynamic contrast enhanced MRI based on affinity propagation clustering. *J. Magn. Reson. Imaging* 39 (5), 1327–1337.
- Song, T., Huang, N., 2018. Integrated extractor, generator and segmentor for ischemic stroke lesion segmentation. In: *International MICCAI Brainlesion Workshop*. Springer, pp. 310–318.
- Sourbron, S., Luypaert, R., Morhard, D., Seelos, K., Reiser, M., Peller, M., 2007. Deconvolution of bolus-tracking data: a comparison of discretization methods. *Physics in Medicine & Biology* 52 (22), 6761.
- Stroke Unit Trialists Collaboration, 2013. Organised inpatient (stroke unit) care for stroke. *Cochrane Database Syst Rev* 9 (9).
- Thijs, V.N., Somford, D.M., Bammer, R., Robberecht, W., Moseley, M.E., Albers, G.W., 2004. Influence of arterial input function on hypoperfusion volumes measured with perfusion-weighted imaging. *Stroke* 35 (1), 94–98.
- Ulas, C., Das, D., Thrippleton, M.J., del C. Valdés, H.M., Armitage, P.A., Makin, S.D., Wardlaw, J.M., Menze, B.H., 2018. Convolutional neural networks for direct inference of pharmacokinetic parameters: Application to stroke dynamic contrast-enhanced MRI. *Front Neurol* 9.
- Ulas, C., Tetteh, G., Thrippleton, M.J., Armitage, P.A., Makin, S.D., Wardlaw, J.M., Davies, M.E., Menze, B.H., 2018. Direct estimation of pharmacokinetic parameters from DCE-MRI using deep CNN with forward physical model loss. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 39–47.
- Vagal, A., Wintermark, M., Nael, K., Bivard, A., Parsons, M., Grossman, A.W., Khatri, P., 2019. Automated CT perfusion imaging for acute ischemic stroke: pearls and pitfalls for real-world use. *Neurology* 93 (20), 888–898.
- Wang, G., Song, T., Dong, Q., Cui, M., Huang, N., Zhang, S., 2020. Automatic ischemic stroke lesion segmentation from computed tomography perfusion images by image synthesis and attention-based deep neural networks. *Med Image Anal* 101787.
- Winder, A., d'Esterre, C.D., Menon, B.K., Fiehler, J., Forkert, N.D., 2020. Automatic arterial input function selection in CT and MR perfusion datasets using deep convolutional neural networks. *Med Phys*.
- Wu, O., Østergaard, L., Weisskoff, R.M., Benner, T., Rosen, B.R., Sorensen, A.G., 2003. Tracer arrival timing-insensitive technique for estimating flow in MR perfusion-weighted imaging using singular value decomposition with a block-circulant deconvolution matrix. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 50 (1), 164–174.
- Yin, J., Sun, H., Yang, J., Guo, Q., 2015. Automated detection of the arterial input function using normalized cut clustering to determine cerebral perfusion by dynamic susceptibility contrast-magnetic resonance imaging. *J. Magn. Reson. Imaging* 41 (4), 1071–1078.

5.3 Detecting CTP Truncation Artifacts in Acute Stroke Imaging from the Arterial Input and the Vascular Output Functions

Ezequiel de la Rosa, Diana M. Sima, Jan S. Kirschke, Bjoern Menze & David Robben

Journal: PLOS One

Synopsis: *Background:* Current guidelines for CT perfusion (CTP) in acute stroke suggest acquiring scans with a minimal duration of 60-70 s. But even then, CTP analysis can be affected by truncation artifacts. Conversely, shorter acquisitions are still widely used in clinical practice and may, sometimes, be sufficient to reliably estimate lesion volumes. We aim to devise an automatic method that detects scans affected by truncation artifacts. *Methods:* Shorter scan durations are simulated from the ISLES'18 dataset by consecutively removing the last CTP time-point until reaching a 10 s duration. For each truncated series, perfusion lesion volumes are quantified and used to label the series as unreliable if the lesion volumes considerably deviate from the original untruncated ones. Afterwards, several features from the arterial input function (AIF) and the vascular output function (VOF) are derived and used to fit machine-learning models with the goal of detecting unreliably truncated scans. Methods are compared against a baseline classifier solely based on the scan duration, which is the current clinical standard. The ROC-AUC, precision-recall AUC and the F1-score are measured in a 5-fold cross-validation setting. *Results:* The best performing classifier obtained an ROC-AUC of 0.982, precision-recall AUC of 0.985 and F1-score of 0.938. The most important feature was the AIFcoverage, measured as the time difference between the scan duration and the AIF peak. When using the AIFcoverage to build a single feature classifier, an ROC-AUC of 0.981, precision-recall AUC of 0.984 and F1-score of 0.932 were obtained. In comparison, the baseline classifier obtained an ROC-AUC of 0.954, precision-recall AUC of 0.958 and F1-Score of 0.875. *Conclusions:* Machine learning models fed with AIF and VOF features accurately detected unreliable stroke lesion measurements due to truncated CTP acquisitions. The AIFcoverage was the most predictive feature of truncation and identified unreliable short scans almost as good as machine learning. We conclude that AIF/VOF based classifiers are more accurate than the scans' duration for detecting

5.3. Detecting CTP Truncation Artifacts in Acute Stroke Imaging from the Arterial Input and the Vascular Output Functions



truncation. These methods could be transferred to perfusion analysis software in order to increase the interpretability of CTP outputs.

Contributions of thesis author: Preprocessing of the data using in-house perfusion imaging software. Generation of synthetic samples from the original raw data. Quantification of the data-features used in the proposed algorithm. Experiment design, planning and conduction. Evaluation of results: performance analysis and statistical analysis. Comparison against different machine learning approaches. Manuscript writing.

Copyright: © 2023 de la Rosa et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

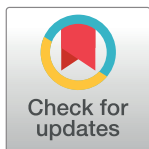
RESEARCH ARTICLE

Detecting CTP truncation artifacts in acute stroke imaging from the arterial input and the vascular output functions

Ezequiel de la Rosa ^{1,2,*}, Diana M. Sima^{1,3}, Jan S. Kirschke ⁴, Bjoern Menze^{2,5}, David Robben ^{1,6,7}

1 icometrix, Leuven, Belgium, **2** Department of Informatics, Technical University of Munich, Munich, Germany, **3** AIMS Laboratory, Center for Neurosciences, Vrije Universiteit Brussel, Brussels, Belgium, **4** Neuroradiology, School of Medicine, Technical University of Munich, Munich, Germany, **5** Department of Quantitative Biomedicine, University of Zurich, Zurich, Switzerland, **6** Medical Imaging Research Center (MIRC), KU Leuven, Leuven, Belgium, **7** Department of Electrical Engineering, Medical Image Computing (MIC), ESAT-PSI, KU Leuven, Leuven, Belgium

* ezequiel.delarosa@icometrix.com



OPEN ACCESS

Citation: de la Rosa E, Sima DM, Kirschke JS, Menze B, Robben D (2023) Detecting CTP truncation artifacts in acute stroke imaging from the arterial input and the vascular output functions. PLoS ONE 18(3): e0283610. <https://doi.org/10.1371/journal.pone.0283610>

Editor: Stephan Meckel, RKH Klinikum Ludwigsburg, GERMANY

Received: May 16, 2022

Accepted: March 13, 2023

Published: March 30, 2023

Copyright: © 2023 de la Rosa et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All CTP images used in this study are public and free, and have been released for the Isles'18 challenge (<http://www.isles-challenge.org/>) under the following repository: <https://www.smir.ch/>. Besides, manual vascular function annotations used in this work are available in the supporting material of the following manuscript: de la Rosa, Ezequiel, et al. "AIFNet: Automatic vascular function estimation for perfusion analysis using deep learning." Medical Image Analysis 74 (2021): 102211.

Abstract

Background

Current guidelines for CT perfusion (CTP) in acute stroke suggest acquiring scans with a minimal duration of 60-70 s. But even then, CTP analysis can be affected by truncation artifacts. Conversely, shorter acquisitions are still widely used in clinical practice and may, sometimes, be sufficient to reliably estimate lesion volumes. We aim to devise an automatic method that detects scans affected by truncation artifacts.

Methods

Shorter scan durations are simulated from the ISLES'18 dataset by consecutively removing the last CTP time-point until reaching a 10 s duration. For each truncated series, perfusion lesion volumes are quantified and used to label the series as *unreliable* if the lesion volumes considerably deviate from the original untruncated ones. Afterwards, nine features from the arterial input function (AIF) and the vascular output function (VOF) are derived and used to fit machine-learning models with the goal of detecting unreliably truncated scans. Methods are compared against a baseline classifier solely based on the scan duration, which is the current clinical standard. The ROC-AUC, precision-recall AUC and the F1-score are measured in a 5-fold cross-validation setting.

Results

The best performing classifier obtained an ROC-AUC of 0.982, precision-recall AUC of 0.985 and F1-score of 0.938. The most important feature was the $AIF_{coverage}$, measured as the time difference between the scan duration and the AIF peak. When using the $AIF_{coverage}$ to build a single feature classifier, an ROC-AUC of 0.981, precision-recall AUC of 0.984 and F1-score of 0.932 were obtained. In comparison, the baseline classifier obtained an ROC-AUC of 0.954, precision-recall AUC of 0.958 and F1-Score of 0.875.

Funding: This project received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement TRABIT No 765148. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: EdIR, DMS and DR are employees of icometrix.

Conclusions

Machine learning models fed with AIF and VOF features accurately detected unreliable stroke lesion measurements due to insufficient acquisition duration. The AIF_{coverage} was the most predictive feature of truncation and identified unreliable short scans almost as good as machine learning. We conclude that AIF/VOF based classifiers are more accurate than the scans' duration for detecting truncation. These methods could be transferred to perfusion analysis software in order to increase the interpretability of CTP outputs.

Introduction

Treatment decision making in acute ischemic stroke is mostly guided by computed tomography (CT) imaging, as the technique allows to answer (at least) four crucial questions regarding the patient brain's condition: 1) Is there hemorrhage? 2) Is there any thrombus that could be targeted? 3) Is there already irreversibly damaged tissue (a.k.a. *core*)? 4) Is there salvageable tissue (a.k.a. *penumbra*, tissue at risk but potentially recoverable)? [1]. While the first two questions can be answered with non-contrast CT and CT angiography, respectively, the last two questions are typically addressed through CT perfusion (CTP). CTP is of major importance for neuroradiologists as it allows the identification of patients that could benefit from recanalization therapies [2]. In this context, distinguishing potentially salvageable brain tissue from already necrosed areas drive the therapeutical decision making.

In clinical routine, CTP post-processing software is used to estimate perfusion maps and to quantify perfusion lesion volumes. The perfusion maps used in acute ischemic stroke are derived from the CTP contrast attenuation curves and are cerebral blood volume, cerebral blood flow (CBF), mean transit time and time to the maximum of the residue function (Tmax). There exist several different techniques implemented in clinical and/or research software packages to estimate these perfusion metrics. Among the most widely used are the Fourier transform and the delay-invariant singular value decomposition deconvolution techniques using time-shift [3] or block-circulant approaches [4, 5]. Independently of their functioning, the end goal of CTP software packages is the accurate quantification of perfusion maps and, consequently, the reliable volumetric quantification of the brain lesions. Despite the vast adoption of CT perfusion software in clinical routine, there are well known and persistent pitfalls of these techniques that hamper the brain lesion quantification and hence their interpretation, as described in [6–9]. This work focuses on the so called *truncation* of the time attenuation curves, which could be defined as the early ending of the CTP acquisition that precludes the entire capture of the tissue perfusion phases [8].

CTP truncation artifacts have extensively been observed in previous works [10–17] and are related to several sources: *i*) the type of deconvolution used to process the CTP images, *ii*) the biological and physiological variability of the patients (e.g. patient size and the cardiac output alter the contrast delivery through the brain [16]), *iii*) physiopathological conditions that prolong the contrast-agent passage through the affected tissue, which happens in the hypoperfused tissue due to the ischemic occlusion [10, 13] or in patients with severe intracranial vascular narrowing or multiple intracranial emboli [6], and *iv*) the contrast injection and CTP acquisition protocols (e.g. contrast injection rate, the pre-contrast scanning duration, synchronization between contrast injection and acquisition, etc.).

As described in practical acute stroke imaging recommendations, the CTP analysis should include a quality control step that checks for complete acquisition of the perfusion curves including both the contrast agent wash-in and wash-out phases [8, 9, 18]. Visual identification of truncated AIF/VOF and/or time attenuation curves has been conducted in previous studies [14, 15]. Despite the fact that visual quality control could easily detect truncated perfusion curves, it is not straightforward to understand the implications of such curves truncations over the quantified lesion volumes. Thus, finding whether the truncation effects are strong enough to considerably perturb the quantified perfusion volumes could only be assessed through quantitative analyses. A major step in understanding the quantitative impact of truncation artifacts over the perfusion maps was done in [16]. The work showed that truncation artifacts depend on the truncation degree and affect the perfusion metrics differently depending on the used deconvolution algorithm. Moreover, the CTP truncation effects over the brain lesion volumes were studied in [17]. The authors found that a 60 second scan duration is enough to avoid volumetric errors in 95% of their analyzed scans. These results have later been adopted as a practical recommendation for the implementation of CTP in acute stroke [18]. In clinical routine, however, different centers or scanner operators make use of post-processing software from different vendors (and with diverse deconvolution algorithms), as well as different contrast injection and CTP acquisition protocols. Shorter acquisitions are frequently adopted by centers in order to reduce the exposure of the patient to ionizing radiation under the ALARA (i.e. as low as reasonably achievable) principle. Based on these considerations, it is possible that scans with shorter than 60 second scan duration could reliably estimate lesion volumes while scans with different characteristics could suffer from truncation errors even while having a 60–70 second acquisition duration.

In this work we propose a tool for the automatic identification of unreliable perfusion volumes due to insufficient scan duration. Our proposal makes use of simple and easy to extract features derived from the vascular perfusion curves (i.e. the arterial input function, AIF, and the vascular output function, VOF). Experiments on the public ISLES'18 dataset show that truncation artifacts impact the perfusion-derived features, hence allowing their identification with machine learning models. The proposed approach increases the interpretability of acute ischemic stroke outputs obtained in clinical practice with CTP post-processing software.

Materials and methods

Data

The ISLES'18 dataset is used for our experiments [19, 20]. The database is multi-center and multi-scanner and includes 156 CTP scans obtained from 103 acute stroke patients. For our experiments, we have used the preprocessed scans from the ISLES 2018 challenge (<http://www.isles-challenge.org/>). The CTP volumes have been motion corrected, coregistered and spatio-temporally resampled (256×256 matrix, 1 volume per second). A full dataset description can be found in [19].

Simulating shorter CTP scans

We simulate shorter CTP scan durations by repeatedly discarding a 1 second timepoint from the end of the series until reaching the 10 first seconds of it. Note that the number of truncated simulated series varies from scan to scan, depending on its original total duration.

CTP post-processing

Each truncated CTP series is analyzed using a research version of **icobrain cva** 1.4.1 (icometrix, Leuven, Belgium), an FDA-cleared and CE-marked software for acute stroke CTP post-processing. Each truncated series is processed using experts' manually annotated vascular functions available in [21]. Please note that the manual AIF/VOF does not change location for all shorter versions of a same scan. The vascular functions from each truncated scan are retained for the subsequent experiments.

Perfusion maps (Tmax, CBF, cerebral blood volume and mean transit time) are obtained through delay-invariant singular value decomposition deconvolution. Absolute and relative CBF maps are computed, where the relative rCBF map is obtained after normalization of the absolute one using mean control tissue values. Control tissue is defined by the software as $T_{max} < 6s$ [22]. Quantification of the hypoperfused and core lesion volumes is automatically obtained by the software using $T_{max} > 6s$ [22] and $rCBF < 0.38$ (within the hypoperfused tissue area), respectively. The used rCBF cutoff (which is set in the software just for the purpose of these experiments) has been identified as optimal for the ISLES'18 dataset [19].

Defining truncation artifacts

In order to label each shorter scan version as *reliable* or *unreliable* (i.e., considerably suffering from truncation artifacts), we first check that the original unshortened scan does not already suffer from truncation artifacts. As such, scans are labeled to be *stable* if truncation of the final 6 frames or less did not impact the computed volumes by more than 2.5 ml [17]; otherwise, scans are labelled as *unstable* ones. For our experiments, all unstable scans have been discarded from further analyses.

The truncated series from all stable scans are labelled as *reliable* if the corresponding hypoperfused and core volumes deviated $< 10\%$ or < 5 ml from the untruncated volume estimates. Otherwise, the truncated scan (and all its shorter versions) are labelled as *unreliable*. Scans with a stable hypoperfused lesion smaller than 5 ml are excluded from the analysis as their reliability can not be trusted. Besides, for each CTP series, the optimal scan duration (OSD) is defined as the shortest scan duration providing reliable volume estimates. Fig 1 shows a stable CTP scan example with its corresponding reliability truncation labels.

Machine learning for CTP truncation detection

Machine learning algorithms have been widely used to assess the quality of medical images [23–25]. We explore different machine learning models that could detect unreliably truncated scans by solely using information extracted from the vascular functions. The benefits of using the AIF and VOF to detect truncation artifacts are two-fold. First, the perfusion curves are always available in this imaging modality. Second, as they cover the entire perfusion event (note that these curves represent the contrast concentration inlet and outlet to the brain), they contain rich information for the problem under study. Consequently, it is needed to extract meaningful perfusion features that are impacted by an insufficient scan duration and that are, also, predictive of the truncation artifacts. Those features should capture the perfusion phases of the contrast-agent wash-in and wash-out and, ideally, they should be unaltered by the different CTP protocols used in clinical routine.

Feature extraction. All the explored machine learning algorithms are fed with the following 9 AIF/VOF derived features:

- Scan duration
- AIF/VOF time to the peak of the function ($\text{argmax}\{\text{AIF}\}$, $\text{argmax}\{\text{VOF}\}$)

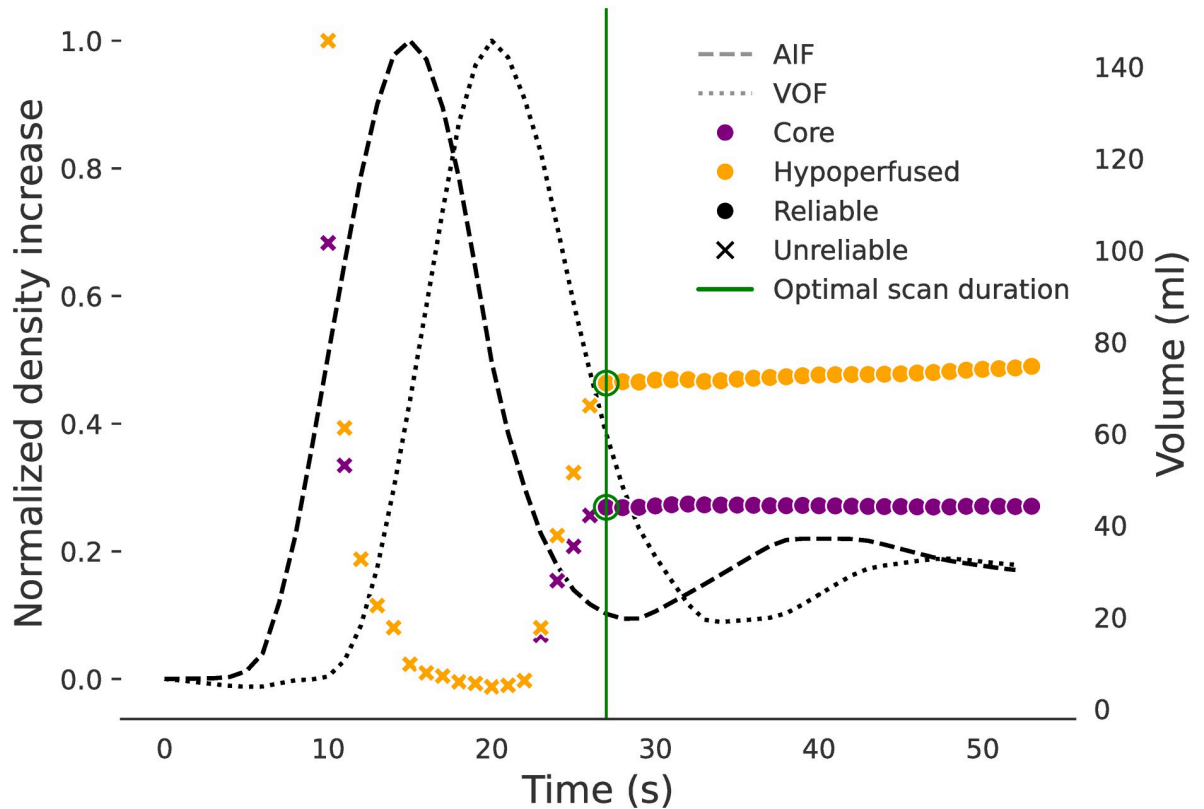


Fig 1. Reliable/unreliable lesion volumes computed at various scan durations. The arterial input function (AIF) and the vascular output function (VOF) are displayed as reference.

<https://doi.org/10.1371/journal.pone.0283610.g001>

- The *AIF/VOF coverage*, defined as the time difference between the peak of a signal and the scan duration:

$$* AIF_{coverage} = scan\ duration - argmax\{AIF\}$$

$$* VOF_{coverage} = scan\ duration - argmax\{VOF\}$$

- *AIF/VOF upward and downward contrast increase*

$$* AIF_{UCI} = AIF_{t=argmax\{AIF\}} - AIF_{t=0}$$

$$* AIF_{DCI} = AIF_{t=argmax\{AIF\}} - AIF_{t=scan\ duration}$$

$$* VOF_{UCI} = VOF_{t=argmax\{VOF\}} - VOF_{t=0}$$

$$* VOF_{DCI} = VOF_{t=argmax\{VOF\}} - VOF_{t=scan\ duration}$$

All features are visually represented in Fig 2.

Classifiers & model fitting. We train six statistical/machine learning classifiers with the aim of detecting *reliable* and *unreliable* truncated scans. The trained models make use of linear or non-linear decision functions and are: *i*) random forests, *ii*) multivariate logistic-regression, *iii*) support vector machines with linear kernel, *iv*) support vector machines with radial basis kernel, *v*) Adaptive boosting (aka, Adaboost [26]) and *vi*) Gradient boosting [27]. In order to find the optimal set of parameters to fit a model, a Bayesian search is conducted by sampling over the parameter-space of each model and by evaluating its performance in a train set, 3-fold

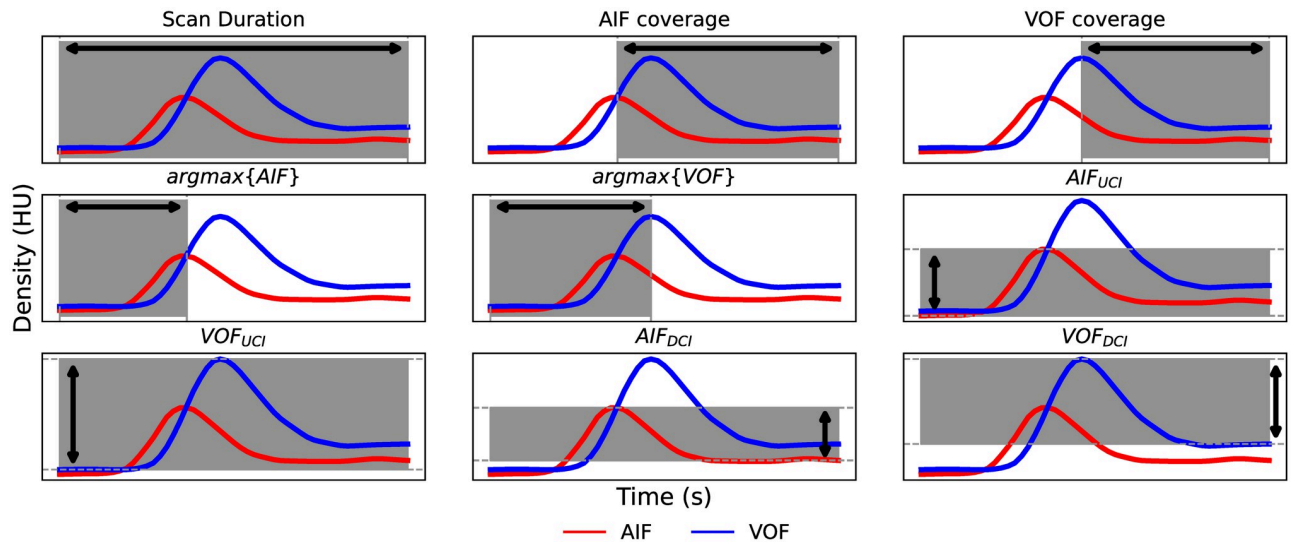


Fig 2. AIF and VOF derived features used to feed the machine learning algorithms. AIF: arterial input function; VOF: venous output function; HU: Hounsfield units; UCI: upward contrast increase; DCI: downward contrast increase.

<https://doi.org/10.1371/journal.pone.0283610.g002>

cross-validation setting. The final set of parameters chosen to parametrize the model is the one maximizing the area under the precision-recall curve (PR-AUC). All models are trained and optimized using the scikit-learn Python library [28].

Data augmentation. We augment the training sets by generating synthetic samples in order to model different perfusion scenarios, such as the variable pre-contrast agent duration and the variable contrast increases of the perfusion curves. Note that a different timing in the contrast bolus arrival alters the CTP scan duration but does not alter the presence of truncation artifacts. Likewise, the AIF and VOF contrast increase depends on the contrast agent iodine concentration. However, as the deconvolution algorithm is independent from the AIF/VOF absolute amplitudes, a variable vascular contrast increase does not alter the presence of truncation artifacts.

Simulation of contrast injection protocol variations is conducted by perfusion-specific data augmentation as similarly done in [21, 29]. Uniform distributions are used to randomly modify the pre-contrast agent duration and vascular contrast increases. When simulating variable pre-contrast duration, pre-contrast timing dependent features are increased or decreased by the same random factor ($\text{argmax}\{\text{AIF}\}$, $\text{argmax}\{\text{VOF}\}$ and *scan duration*). For modelling variable contrast increases, the features AIF_{UCI} , AIF_{DCI} , VOF_{UCI} and VOF_{DCI} are scaled by a random factor.

Experiments. We perform a 5-fold cross-validation experiment using an 80–20% train-test data split. The data splitting is conducted at the *scans* level, assuring that *i*) all untruncated and truncated versions of a same scan belong to the same fold and *ii*) the same data-splits are used to fit all the considered models. Only the training data is used to parametrise the models and to select the classifiers' operating point. Truncation predictions are later inferred over the unseen test data.

Besides, we compare the machine learning models against a baseline classifier which solely uses the *scan duration* as discriminant-rule. The classifier g operates as follows:

$$g(\text{scan duration}, \theta) = \begin{cases} \text{reliable} & \text{if scan duration} \geq \theta \\ \text{unreliable} & \text{if scan duration} < \theta \end{cases} \quad (1)$$

with θ a scan duration cutoff. This baseline is motivated by the CTP guidelines, which only consider the duration of a scan to avoid truncation artifacts during CTP acquisition [18]. Specifically, these guidelines suggest a cutoff of $\theta = 60$ seconds in Eq 1.

In order to understand the relevance of the AIF-VOF extracted features to discriminate *reliably* and *unreliably* truncated scans, we conduct a bootstrapping experiment by resampling 100 times the original database. In each iteration, a sample was drawn with replacement and was used to fit a classifier as described in Section *Classifiers & Model Fitting*. The relative feature importance is measured as defined in [27] for decision tree ensembles. Briefly, the feature importance is calculated at the classifiers' tree level as the impurity decay across all the nodes where that feature was used to create a split [30]. The final feature importance is computed as the average feature importance over all the considered trees. The mean and standard deviation feature importance for all features are reported. The chosen classifier for this experiment is the best performing one in terms of precision-recall AUC.

Performance evaluation. The mean, standard deviation, 5th-95th percentiles and minimum and maximum of the scan duration and the optimal scan duration are reported for the entire dataset. The different algorithms' performance are evaluated by conducting ROC and precision-recall (PR) analysis. The area under the ROC and precision-recall curves are used as general classifier performance metrics. Besides, we measure the binary classification performance at the operating point closest to an ideal classifier with $precision = recall = 1$. The operating point is chosen from the fitted classifier as $\operatorname{argmin}\{\sqrt{(precision_t - 1)^2 + (recall_t - 1)^2}\}$, with t different classifier thresholds. Performance is measured in terms of precision ($\frac{TP}{TP+FP}$), recall ($\frac{TP}{TP+FN}$) and F1-score ($\frac{2*TP}{2*TP+FP+FN}$), where acronyms represent TP: true positives, TN: true negatives, FP: false positives and FN: false negatives. The same binary classification metrics are reported for our baseline scan duration classifier, by making use of cutoffs $\theta = [30, 40, 50, 60]$ s. For these defined metrics, an *unreliable* truncation sample is considered as positive and a *reliable* truncation sample as negative.

Results & discussion

From the 156 analyzed scans, 123 scans (78.8%) are retained for further analysis. The remaining scans are discarded since 18 (11.5%) are unstable, 14 (8.9%) have hypo-perfused volumes < 5 ml or are free from CTP lesions, and 1 scan (0.6%) is corrupted by motion artifacts. A total of 4621 synthetically truncated scans are obtained from the retained stable cases, from which 2353 (50.9%) are labelled as *reliable* and 2268 (49.1%) as *unreliable*.

Descriptive statistics about the optimal scan duration are summarized in Table 1. It can be appreciated that ~ 40 s scan duration suffices to get accurate perfusion volumes in 95% of the dataset and 43 s scan duration avoids truncation in the entire ISLES'18 dataset. At first sight these OSD values might seem much shorter than the 60-second recommended duration in the CTP guidelines [18]. However, the used ISLES'18 dataset has very short pre-contrast acquisitions that does not always reach the guidelines' recommended 5–10 s. Note that the median

Table 1. Descriptive statistics of the stable CTP scans. SD: Scan duration; OSD: Optimal scan duration. Std: standard deviation; P: percentile. AIF: arterial input function. All metrics are reported in seconds.

| | SD | OSD | (OSD— $\operatorname{argmax}\{\text{AIF}\}$) |
|---|--------------|--------------|---|
| Mean (std) | 46.6 (5.3) | 28.4 (5.7) | 12.9 (3.7) |
| (Min, Max) | (31.0, 64.0) | (19.0, 43.0) | (8.0, 30.0) |
| (P5 th , P95 th) | (43.0, 60.2) | (20.0, 39.8) | (10.0, 19.9) |

<https://doi.org/10.1371/journal.pone.0283610.t001>

AIF peak (i.e. $\text{argmax}\{\text{AIF}\}$) across all scans is 15.6 ± 4.5 s, with a minimum AIF peak at 5 s. Thus, some ISLES'18 scans have no pre-contrast acquisition at all, as seen in [S1 Fig](#). Therefore, compensating for the short pre-contrast duration would increase the 95th percentile OSD value of [Table 1](#) to a $\sim 45\text{--}50$ s scan acquisition. This result is in line with earlier research: it was shown that in scans with 10 s of pre-contrast duration a ~ 53 s acquisition is needed to get reliable perfusion volumes in 90% of the scans [17].

The optimal scan duration reported in [Table 1](#) depends on the pre-contrast scan duration, which is not standardized in clinical practice. In order to have a more informative metric less biased by the different CTP protocols, we compute the time difference between the OSD and the AIF peak ($\text{argmax}\{\text{AIF}\}$). In [Table 1](#) this metric is reported for our entire database. Results show that on average our scans require ~ 13 s following the AIF peak to obtain reliable perfusion volumes. In [17] the median OSD for scans with a 10 s baseline is ~ 33 s. Let's assume an average AIF peak of $\sim 15\text{--}20$ s in a standard CTP acquisition protocol. Then $\sim 13\text{--}18$ s following the AIF peak are needed in [17] to get reliable volumes in 50% of their scans. Thus, these results are comparable to our finding of ~ 13 s on average following the AIF peak to get reliable volumes.

These OSD analyses of ISLES'18 show that an acquisition protocol using 60–70 s scan duration, with 5–10 s of pre-contrast acquisition, avoids truncation errors in the whole analyzed dataset. Thus, supporting the recommended scan duration of the CTP guidelines. Depending on the patients' physiology, the contrast injection and/or the CTP acquisition protocols, even shorter acquisitions may, sometimes, reliably quantify CTP lesion volumes. However, from a risk-benefit perspective it is strongly inadvisable to shorten CTP scan durations since *i*) it is not possible to know *a-priori* the OSD needed for a particular patient/scan, and *ii*) the risk of inaccurately estimating the lesion volumes due to a short acquisition is significantly larger than exposing the patient to an additional radiation exposure. Thus, an unreliable estimation of the core-penumbra mismatch may lead to a change in the treatment decision, which may have a drastic impact on the patient's outcome. In a different scenario, an insufficient CTP acquisition may lead to the full re-scanning of the patient, thus significantly increasing the exposure to ionizing radiation and to the iodine contrast and, ultimately, delaying the treatment of the patient.

Effect of the AIF choice

In order to understand the impact of different AIFs over the computed optimal scan duration, an inter-rater analysis is performed. In this experiment, we simulate truncation artifacts as described in section *Simulating Shorter CTP Scans* but using vascular functions selected by a different expert. The used annotations are the ones available in [21] and labelled in the work as *Rater #2*.

In [Fig 3](#) a histogram and a Bland-Altman plot of the inter-rater OSD values are shown (left and right figures, respectively). The majority of the scans ($n = 78$, $\sim 64\%$) show no time differences in the OSD obtained with vascular functions selected by the two raters. The 5th and 95th percentiles of the absolute OSD differences are respectively 0 s and 2.30 s. The maximum OSD difference between the raters is 8 s.

Truncation artifacts detection

[Fig 4](#) shows the ROC and precision-recall curves obtained with the different classifiers when differentiating *reliable* from *unreliable* truncated acquisitions. Overall it can be seen that classifiers yielded a similar high performance for both the considered metrics. The gradient boosting algorithm outperformed the remaining classifiers yielding an ROC-AUC of 0.982 and a

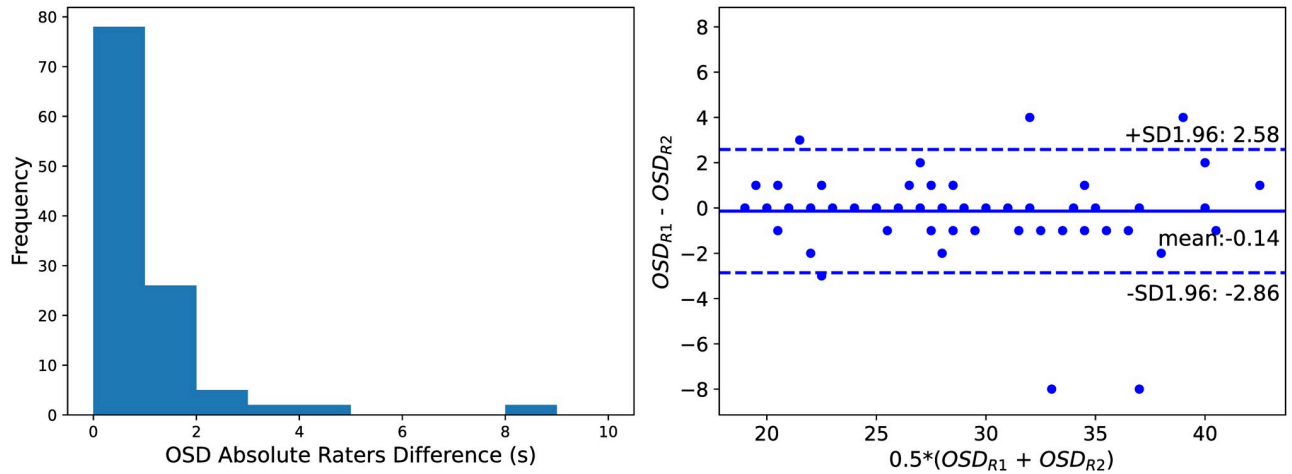


Fig 3. Optimal scan duration (OSD) inter-rater analysis. Left: Histogram of the OSD absolute differences between raters. Right: Bland-Altman plot of the OSD values for the raters. R1: Rater #1. R2: Rater #2.

<https://doi.org/10.1371/journal.pone.0283610.g003>

precision-recall AUC of 0.985. When assessing the classifiers capability for detecting truncation artifacts at the chosen operating point, the results of Table 2 are obtained. The Gradient boosting method obtained the highest performance for detecting unreliable perfusion volumes (F1-Score = 0.938). All machine learning models have considerably outperformed the baseline classifier *g* in terms of ROC-AUC, PR-AUC and F1-Score (Table 2).

For the baseline classifier *g*, the highest detection performance is obtained at the *optimal operating point* $\theta = 27$ s (F1-score = 0.875). When using the standard cutoff $\theta = 60$ s, the baseline classifier showed the maximal recall of 1.0 with a low precision of 0.493 and low F1-score of 0.660. These results are expected as all the analyzed scans have OSD values much lower than 60 s (Table 1). The ROC and precision-recall operating points at $\theta = 60$ s are shown in Fig 4. It

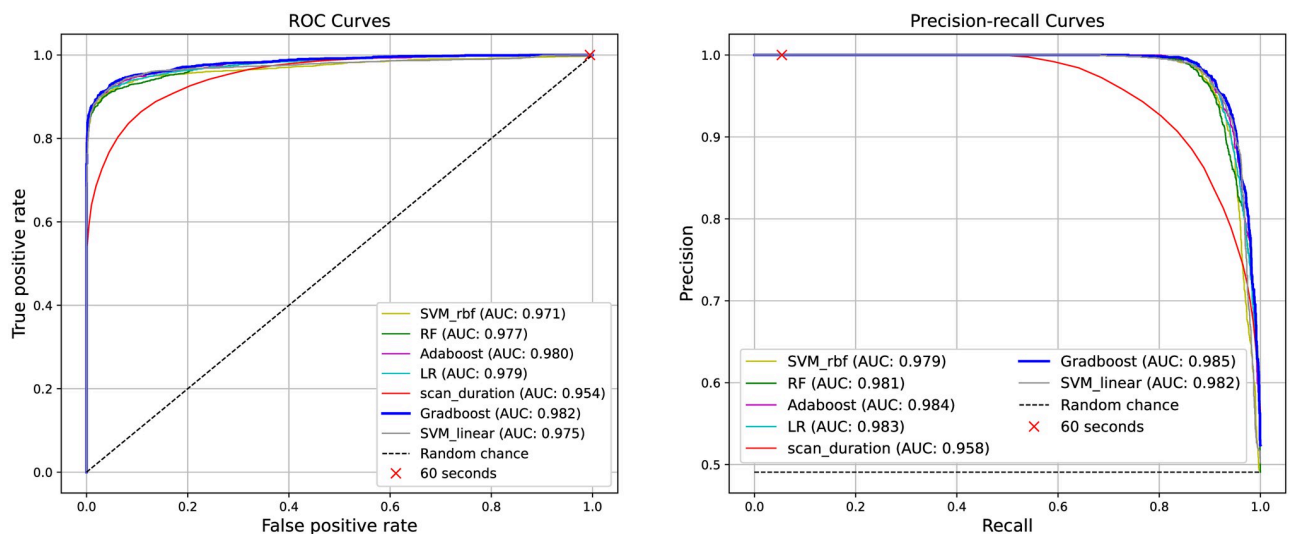


Fig 4. Receiver operating characteristic (left) and precision-recall (right) curves. AUC: area under the curve; SVM_linear: support-vector machine with linear kernel; SVM_rbf: support-vector machine with radial basis function kernel; RF: Random forests; LR: Logistic-regression; Adaboost: Adaptive boosting; Gradboost: Gradient boosting.

<https://doi.org/10.1371/journal.pone.0283610.g004>

Table 2. Classifiers' performance for detecting truncation artifacts. The used operating points are $\theta = [27, 30, 40, 50, 60]$ s for the baseline classifier g . Note that 27 s is the *optimal operating point* for g , defined as the closest point to the ideal classifier with $precision = recall = 1$. Reported metrics for the machine learning approaches are obtained at the optimal operating points. Outperforming values for each metric are shown in bold. SVM_{Linear}: support-vector machine with linear kernel; SVM_{RBF}: support-vector machine with radial basis function kernel; RF: Random forests; LR: Logistic regression; Gradboost: Gradient boosting classifier. Baseline classifier: threshold on scan duration.

| Classifier | ROC-AUC | PR-AUC | Recall | Precision | F1-Score |
|-----------------------|--------------|--------------|----------------|--------------|---|
| SVM _{Linear} | 0,975 | 0,982 | 0,929 | 0,942 | 0,9305 |
| SVM _{RBF} | 0,971 | 0,979 | 0,918 | 0,942 | 0,930 |
| RF | 0,977 | 0,981 | 0,898 | 0,964 | 0,929 |
| Adaboost | 0,980 | 0,984 | 0,922 | 0,945 | 0,934 |
| LR | 0,979 | 0,983 | 0,926 | 0,936 | 0,931 |
| Gradboost | 0,982 | 0,985 | 0,930 | 0,946 | 0,938 |
| Baseline classifier | 0,954 | 0,958 | [0,885] | 0,864 | 0,875] _{$\theta=27s$} |
| | | | [0,926] | 0,813 | 0,866] _{$\theta=30s$} |
| | | | [0,997] | 0,597 | 0,747] _{$\theta=40s$} |
| | | | [1,000] | 0,503 | 0,670] _{$\theta=50s$} |
| | | | [1,000] | 0,493 | 0,660] _{$\theta=60s$} |

<https://doi.org/10.1371/journal.pone.0283610.t002>

can be appreciated that this operating point falls on the boundaries of the classifiers' ROC and precision-recall curves. These results suggest that using a 60 s scan duration is a safe and needed recommendation for acquiring CTP images, but it is a poor criterion in order to identify truncation errors at CTP post-processing stages.

Effect of the data augmentation. We evaluated the impact of the perfusion specific data augmentation strategy over the best performing machine learning model. For this experiment, a Gradient boosting classifier is trained using the original, un-augmented dataset as described in the section *Classifiers & Model Fitting*. While an ROC-AUC = 0.982, PR-AUC = 0.985 and F1-Score = 0.938 are obtained when using data augmentation, an ROC-AUC = 0.980, PR-AUC = 0.983 and F1-Score = 0.933 are obtained when training the model without augmenting the dataset. Our results show that simulating different perfusion scenarios (namely, variable contrast-increases and variable bolus arrival times) improves the model's performance.

Importance of the AIF and VOF features

Fig 5 summarizes the different features' relevance obtained when fitting 100 Gradient boosting classifiers in a resampling with replacement bootstrapping fashion. The AIF_{coverage} shows to be the most crucial feature for detecting *unreliable* perfusion volumes due to truncated acquisitions. Besides, the VOF_{DCI} and the VOF_{coverage} also result to be important features for the machine learning model. The large predictive value of the AIF_{coverage} and the VOF_{coverage} features can be related to their robustness to variable pre-contrast agent duration. The *scan duration* feature, instead, is affected by the CTP acquisition protocols and as such, shows slightly less relevance for the fitted models.

We also explore for each single AIF/VOF feature its discriminant power to detect truncation artifacts. To this end, we generate new classifiers g' that operate as described in Eq 1 by using the considered AIF/VOF feature instead of the *scan duration* feature. The selection of the operating points and the validation of the classifiers are performed with the same criteria described in section *Performance evaluation* for the machine-learning models. In Table 3 the detection performance metrics achieved with the different feature classifiers g' are summarized. It is worth noting that the top-ranked features in the bootstrapping experiment (namely

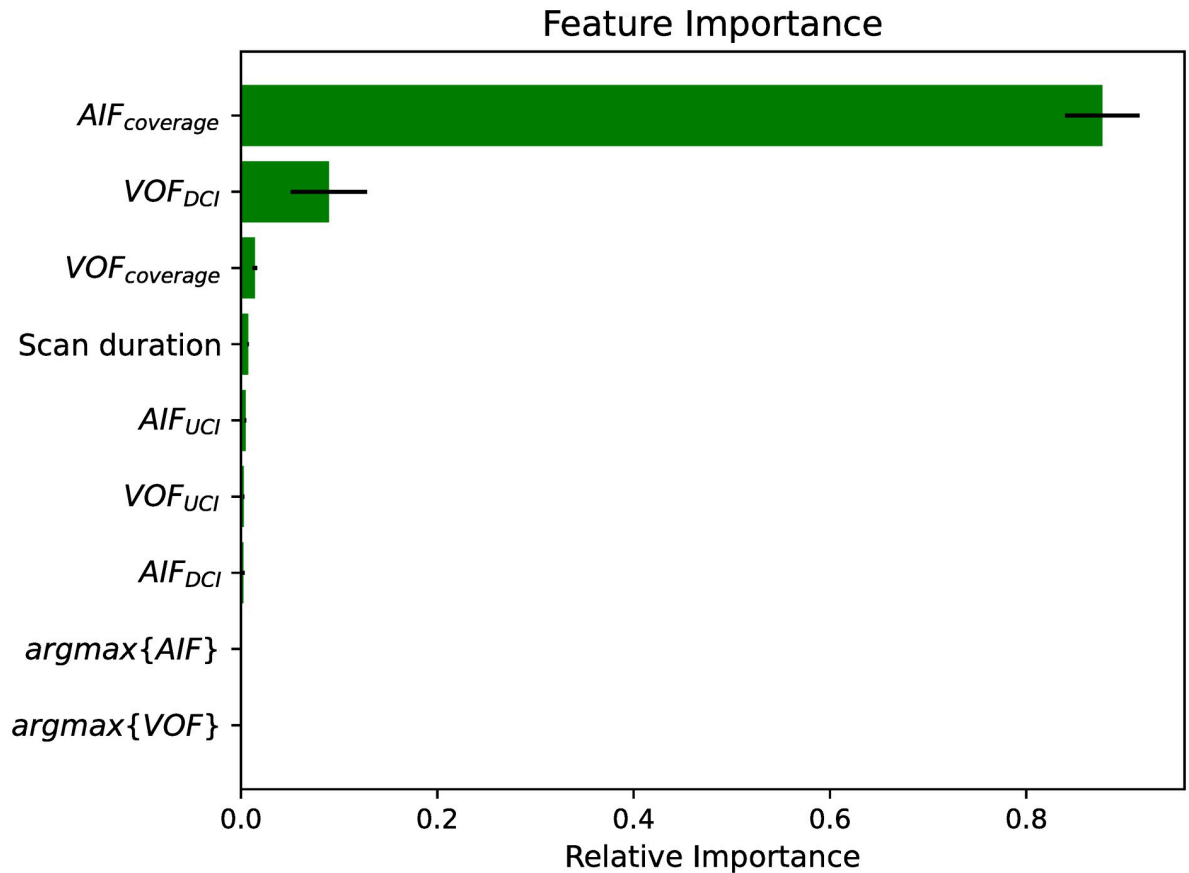


Fig 5. Relative feature importance for 100 bootstraps with a Gradient boosting classifier. Bars (error-bars) represent mean (standard deviation). AIF: arterial input function; VOF: venous output function; UCI: upward contrast increase; DCI: downward contrast increase.

<https://doi.org/10.1371/journal.pone.0283610.g005>

$AIF_{coverage}$, VOF_{DCI} and, $VOF_{coverage}$, see Fig 6) are the ones achieving the highest discriminant performance for detecting truncation artifacts. The best classifier $g'(AIF_{coverage})$ yielded a much better performance than the baseline classifier $g(\text{scan duration})$ (Tables 2 and 3). These results evidence that $AIF_{coverage}$ is a strong discriminant feature for detecting truncation artifacts. Similar to our results [17], has observed that the $VOF_{coverage}$ is an important feature

Table 3. Single features' classification performance. The used cutoff is always the *optimal operating point*, defined as the closest point to the ideal classifier with *precision = recall = 1*. AIF: arterial input function; VOF: venous output function; UCI: upward contrast increase; DCI: downward contrast increase; ROC: receiver operating characteristic curve; PR: precision-recall curve; AUC: area under the curve.

| Feature | ROC-AUC | PR-AUC | Precision | Recall | F1-score |
|------------------------------------|--------------|--------------|--------------|--------------|--------------|
| AIF_{DCI} | 0.842 | 0.879 | 0.815 | 0.724 | 0.767 |
| AIF_{UCI} | 0.597 | 0.631 | 0.571 | 0.465 | 0.513 |
| $argmax\{AIF\}$ | 0.509 | 0.525 | 0.502 | 0.484 | 0.493 |
| $argmax\{VOF\}$ | 0.693 | 0.752 | 0.649 | 0.592 | 0.619 |
| $AIF_{coverage}$ | 0.981 | 0.984 | 0.936 | 0.928 | 0.932 |
| $VOF_{coverage}$ | 0.956 | 0.962 | 0.874 | 0.880 | 0.877 |
| VOF_{DCI} | 0.958 | 0.969 | 0.929 | 0.901 | 0.915 |
| VOF_{UCI} | 0.765 | 0.816 | 0.765 | 0.597 | 0.671 |

<https://doi.org/10.1371/journal.pone.0283610.t003>

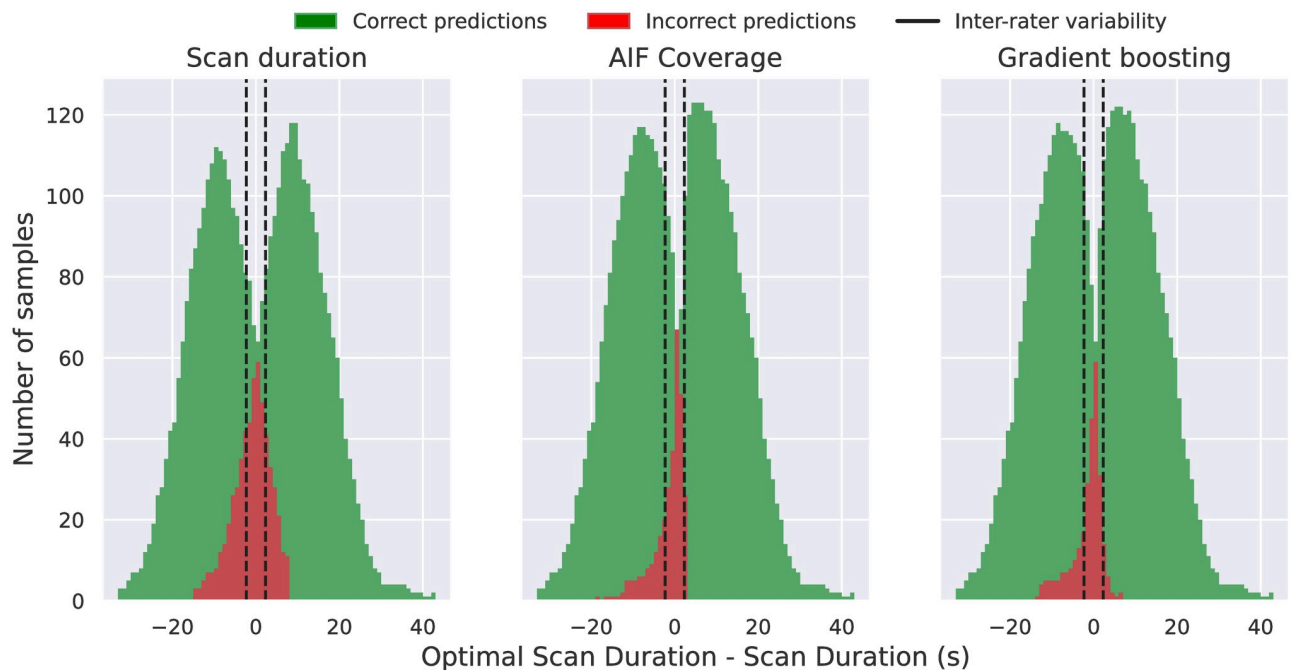


Fig 6. Histograms showing the difference between optimal scan duration and the scan duration for each classifiers' predicted samples. *Correct predictions* comprises scans properly labelled as reliable or unreliable. *Incorrect predictions* comprises samples wrongly labelled as reliable or unreliable. Inter-rater variability lines are drawn at the 95% inter-rater values (± 2.30 s).

<https://doi.org/10.1371/journal.pone.0283610.g006>

impacted by truncation artifacts. Although $AIF_{coverage}$ and $VOF_{coverage}$ are strongly correlated features, our experiments show that $AIF_{coverage}$ carries more truncation predictability than $VOF_{coverage}$ (Table 3). The reason for this finding is that in severely truncated CTP series where the scan acquisition does not reach the VOF peak but it does reach the AIF peak, only the $AIF_{coverage}$ can detect a truncation, as the $VOF_{coverage}$ feature can not be reliably estimated. However, in cases where the CTP acquisition reaches the VOF signal peak, both the $AIF_{coverage}$ and $VOF_{coverage}$ carry similar truncation predictability. We have confirmed this statement after discarding very short CTP scans (e.g. with scan durations lesser than 15 and 20 s) from the database and after re-measuring the single features detection performance. The obtained ROC-AUCs for the AIF and VOF coverages were, respectively, 0.972 and 0.962 when discarding scans with durations shorter than 15 s, and 0.955 and 0.952 when discarding scans shorter than 20 s. Our results showed that the performance agreement between $AIF_{coverage}$ and $VOF_{coverage}$ increased while discarding short CTP series that did not cover the VOF signal peak. In clinical routine, a short duration scan not covering the AIF peak would lead to completely unreliable perfusion volumes and, as such, would hardly be used for treatment decision making. However, in practice it is still common to find CTP scans where the VOF peak has not been reached. Hence, we can conclude that both the $AIF_{coverage}$ and $VOF_{coverage}$ are features with good overall truncation predictability, though the $AIF_{coverage}$ is a more robust feature as it works in a wider range of CTP truncation scenarios.

Scan duration, AIF coverage or machine-learning?

We compare the three classifiers outlined in this work: the baseline classifier g using the *scan duration*, the classifier g' based on the *AIF coverage* and the Gradient boosting classifier which uses multiple AIF and VOF features. It is evident that both the AIF coverage g' and the

Gradient-boosting classifier considerably outperform the baseline approach g based on the scan duration. However, when comparing the Gradient boosting method with g' , a marginal improvement in performance is found (as seen from Tables 2 and 3). Moreover, the study of the feature's importance has shown that the Gradient boosting method predicts outputs mostly using the AIF_{coverage} feature, together with some other features as VOF_{DCI} and VOF_{coverage} (see Fig 2). Thus, it is valid asking whether there is any benefit of using machine learning over a (simpler) classifier as g' .

In order to address this question we evaluate, for the three considered models, the distribution of the predicted samples in terms of their scan duration to optimal scan duration difference, as shown in Fig 6. While g misclassifies samples in a [-15, 8] s vicinity of OSD, g' and the Gradient boosting classifier do it in a vicinity of [-19, 3] s and [-14, 8] s, respectively. Thus, within these temporal windows the classifiers struggled the most to correctly detect reliable/unreliable volumes and, outside this temporal window, the classifiers correctly predicted all samples. At their optimal operating points, all the three models are more accurate predicting truncation over very short scans (i.e. $OSD \gg \text{scan duration}$) rather than over long duration ones ($OSD \ll \text{scan duration}$). Besides, the closer a scan duration is to its OSD, the harder for the models is to correctly classify a sample. It is worth to point out that g' and the Gradient boosting classifier generate most of the mis-classifications within the inter-rater OSD variability, a 'gray zone' interval where the *reliable* and *unreliable* labels suffer from larger uncertainty. With the aim of measuring the method's performance over AIF-choice unbiased samples, we have quantified the classifiers' error rates ($100 \times \frac{\# \text{ incorrect samples}}{\# \text{ all samples}}$) outside the inter-rater variability range ([-2.30, 2.30] s). While g yielded an error rate outside the inter-rater interval of 7.8% ($n = 314$ misclassified samples), g' and the Gradient boosting method yielded, respectively, a 2.5% ($n = 102$ misclassifications) and 2.4% ($n = 98$ misclassifications) error rates.

Our results show that both the Gradient boosting and the AIF_{coverage} based classifiers provide a good overall truncation predictability and consistently outperform the scan duration based approach. These models could increase the interpretability of CTP post-processing software by warning clinicians about potentially misleading and/or unreliable perfusion lesion volumes (i.e., volumetric errors greater than 10% or greater than 5 ml), which should be taken into account during treatment decision making. Additionally, such methods help identify reliable perfusion lesion measurements in scans not reaching the recommended 60–70 s acquisition. Considering their implementation in clinical routine CTP deconvolution packages, the AIF_{coverage} classifier might be preferred as *i*) it is more robust to the variable scan quality than the multi-feature machine learning (i.e., it requires measuring an always available and easy to extract feature, the AIF peak) *ii*) its implementation is straightforward (the model is free from hyper-parameters fitting) and *iii*) it has the advantage of being interpretable by radiologists and neurointerventionalists.

Limitations and future directions

There are some considerations about this research that should be cautiously taken. Firstly, there is no consistent definition for a scan to be *reliable* and, as such, its definition is somehow arbitrary. In this work, in order to define *reliability* we have adopted a quantitative criterion based on the perfusion volumes as similarly done in [17]. Alternative definitions of scan reliability could be based on changes on the treatment eligibility criteria used in the DAWN or DEFUSE-3 trials [31, 32]. An advantage of using a volumetry based criterion over a treatment eligibility one is the capability to identify subtle lesion changes due to inaccurate perfusion measurements that may not impact the treatment decision (e.g. in patients with a large perfusion lesion mismatch).

Secondly, it is worth to mention that our conclusions only hold for CTP analysis using time-invariant singular value decomposition deconvolution. Other techniques used for perfusion analysis might behave differently under truncation scenarios. Still, the delay-invariant singular value decomposition deconvolution is the most widespread and used algorithm in software packages [8, 33, 34]. Readers interested in the effect of CTP truncation over different parameter map estimation methods are referred to the work of [16], as such inter-algorithm comparisons are out of the scope of this research. It is worth saying that while the devised models only hold for the ISLES'18 database characteristics and for the deconvolution algorithm used in this study, the extracted features are generalizable and allow the adaptation of these models to other deconvolution algorithms or imaging modalities (as perfusion MRI).

Thirdly, the deployment of a truncation artifacts detection method in automatic CTP evaluation software is limited to the AIF/VOF selection performance. In this work, all the experiments have been conducted using manually annotated vascular functions. As such, failures in the CTP curves selection could produce a misleading truncation analysis using our proposed methodology. Nonetheless, recent approaches using dedicated artificial intelligence methods show efficacy and robustness to select vascular functions even under low quality CTP scenarios [21, 35].

Finally, future directions for this work might involve the machine-learning prediction of missing CTP time-points at the end of the series. As such, reconstructing the ending perfusion phase of the vascular functions could help improve the detection of truncation artifacts.

Conclusion

We observe that acquiring 60 s CTP scans is sufficient to avoid truncation artifacts in the entire multi-center/scanner ISLES'18 dataset. However, at CTP post-processing stages, using the scans' duration to detect truncation errors is sub-optimal. Depending on the patients' physiology, the contrast injection and/or the CTP acquisition protocols, even shorter acquisitions may sometimes provide reliable lesion volumes. In order to identify unreliable short scans we extract AIF and VOF features that are impacted by truncation errors. These features are simple, robust to extract even in low quality acquisitions and are independent from the contrast injection and CTP acquisition protocols. The AIF_{coverage} proves to be the most predictive feature of truncation. Furthermore, when training classifiers with AIF/VOF derived features a high truncation detection performance is obtained. We conclude that these methods could be transferred to perfusion analysis software and may increase the interpretability of CTP outputs.

Supporting information

S1 Fig. Example case of an ISLES'18 scan with no pre-contrast acquisition. AIF: Arterial input function; VOF: Venous output function.
(TIF)

Acknowledgments

The authors of this work would like to acknowledge the very insightful points raised by the reviewers, which significantly improve the quality of the manuscript and the message delivered through it.

Disclosure: Preliminary analysis of this work has been presented as an abstract at the 7th European Stroke Conference (ESOC 2021).

Author Contributions

Conceptualization: Diana M. Sima, Jan S. Kirschke, Bjoern Menze, David Robben.

Data curation: Ezequiel de la Rosa.

Formal analysis: Ezequiel de la Rosa, David Robben.

Methodology: Ezequiel de la Rosa, Diana M. Sima, Jan S. Kirschke, Bjoern Menze, David Robben.

Resources: Diana M. Sima, David Robben.

Software: Ezequiel de la Rosa, David Robben.

Supervision: Diana M. Sima, David Robben.

Validation: Ezequiel de la Rosa, David Robben.

Visualization: Ezequiel de la Rosa.

Writing – original draft: Ezequiel de la Rosa.

Writing – review & editing: Diana M. Sima, Jan S. Kirschke, Bjoern Menze, David Robben.

References

1. Konstas A, Goldmakher G, Lee TY, Lev M. Theoretic basis and technical implementations of CT perfusion in acute ischemic stroke, part 1: theoretic basis. *American Journal of Neuroradiology*. 2009; 30(4):662–668. <https://doi.org/10.3174/ajnr.A1487> PMID: 19270105
2. Albers GW, Goyal M, Jahan R, Bonafe A, Diener HC, Levy EI, et al. Ischemic core and hypoperfusion volumes predict infarct size in SWIFT PRIME. *Annals of neurology*. 2016; 79(1):76–89. <https://doi.org/10.1002/ana.24543> PMID: 26476022
3. Smith M, Lu H, Trochet S, Frayne R. Removing the effect of SVD algorithmic artifacts present in quantitative MR perfusion studies. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*. 2004; 51(3):631–634. <https://doi.org/10.1002/mrm.20006> PMID: 15004809
4. Wu O, Østergaard L, Koroshetz WJ, Schwamm LH, O'Donnell J, Schaefer PW, et al. Effects of tracer arrival time on flow estimates in MR perfusion-weighted imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*. 2003; 50(4):856–864. <https://doi.org/10.1002/mrm.10610> PMID: 14523973
5. Wittsack HJ, Wohlschläger AM, Ritzl EK, Kleiser R, Cohnen M, Seitz RJ, et al. CT-perfusion imaging of the human brain: advanced deconvolution analysis using circulant singular value decomposition. *Computerized Medical Imaging and Graphics*. 2008; 32(1):67–77. <https://doi.org/10.1016/j.compmedimag.2007.09.004> PMID: 18029143
6. Mangla R, Ekholm S, Jahromi BS, Almast J, Mangla M, Westesson PL. CT perfusion in acute stroke: know the mimics, potential pitfalls, artifacts, and technical errors. *Emergency radiology*. 2014; 21(1):49–65. <https://doi.org/10.1007/s10140-013-1125-9> PMID: 23771605
7. Potter CA, Vagal AS, Goyal M, Nunez DB, Leslie-Mazwi TM, Lev MH. CT for treatment selection in acute ischemic stroke: a code stroke primer. *Radiographics*. 2019; 39(6):1717–1738. <https://doi.org/10.1148/rg.2019190142> PMID: 31589578
8. Vagal A, Wintermark M, Nael K, Bivard A, Parsons M, Grossman AW, et al. Automated CT perfusion imaging for acute ischemic stroke: pearls and pitfalls for real-world use. *Neurology*. 2019; 93(20):888–898. <https://doi.org/10.1212/WNL.0000000000008481> PMID: 31636160
9. Chung CY, Hu R, Peterson RB, Allen JW. Automated Processing of Head CT Perfusion Imaging for Ischemic Stroke Triage: A Practical Guide to Quality Assurance and Interpretation. *American Journal of Roentgenology*. 2021; 217(6):1401–1416. <https://doi.org/10.2214/AJR.21.26139> PMID: 34259036
10. Campbell BC, Christensen S, Levi CR, Desmond PM, Donnan GA, Davis SM, et al. Cerebral blood flow is the optimal CT perfusion parameter for assessing infarct core. *Stroke*. 2011; 42(12):3435–3440. <https://doi.org/10.1161/STROKEAHA.111.618355> PMID: 21980202
11. Kamalian S, Kamalian S, Konstas A, Maas M, Payabvash S, Pomerantz S, et al. CT perfusion mean transit time maps optimally distinguish benign oligemia from true “at-risk” ischemic penumbra, but

- thresholds vary by postprocessing technique. *American journal of neuroradiology*. 2012; 33(3):545–549. <https://doi.org/10.3174/ajnr.A2809> PMID: 22194372
12. d'Esterre CD, Roversi G, Padroni M, Bernardoni A, Tamborino C, De Vito A, et al. CT perfusion cerebral blood volume does not always predict infarct core in acute ischemic stroke. *Neurological Sciences*. 2015; 36(10):1777–1783. <https://doi.org/10.1007/s10072-015-2244-8> PMID: 25981225
 13. Mikkelsen IK, Jones PS, Ribe LR, Alawneh J, Puig J, Bekke SL, et al. Biased visualization of hypoperfused tissue by computed tomography due to short imaging duration: improved classification by image down-sampling and vascular models. *European Radiology*. 2015; 25(7):2080–2088. <https://doi.org/10.1007/s00330-015-3602-x> PMID: 25894005
 14. Geuskens RR, Borst J, Lucas M, Boers AM, Berkhemer OA, Roos YB, et al. Characteristics of misclassified CT perfusion ischemic core in patients with acute ischemic stroke. *PLoS One*. 2015; 10(11): e0141571. <https://doi.org/10.1371/journal.pone.0141571> PMID: 26536226
 15. Borst J, Marquering HA, Beenen LF, Berkhemer OA, Dankbaar JW, Riordan AJ, et al. Effect of extended CT perfusion acquisition time on ischemic core and penumbra volume estimation in patients with acute ischemic stroke due to a large vessel occlusion. *PLoS One*. 2015; 10(3):e0119409. <https://doi.org/10.1371/journal.pone.0119409> PMID: 25789631
 16. Copen W, Deipolyi A, Schaefer P, Schwamm L, González R, Wu O. Exposing hidden truncation-related errors in acute stroke perfusion imaging. *American Journal of Neuroradiology*. 2015; 36(4):638–645. <https://doi.org/10.3174/ajnr.A4186> PMID: 25500309
 17. Kasasbeh AS, Christensen S, Straka M, Mishra N, Mlynash M, Bammer R, et al. Optimal computed tomographic perfusion scan duration for assessment of acute stroke lesion volumes. *Stroke*. 2016; 47(12):2966–2971. <https://doi.org/10.1161/STROKEAHA.116.014177> PMID: 27895299
 18. Christensen S, Lansberg MG. CT perfusion in acute stroke: practical guidance for implementation in clinical practice. *Journal of Cerebral Blood Flow & Metabolism*. 2019; 39(9):1664–1668. <https://doi.org/10.1177/0271678X18805590> PMID: 30346227
 19. Cereda CW, Christensen S, Campbell BC, Mishra NK, Mlynash M, Levi C, et al. A benchmarking tool to evaluate computer tomography perfusion infarct core predictions against a DWI standard. *Journal of Cerebral Blood Flow & Metabolism*. 2016; 36(10):1780–1789. <https://doi.org/10.1177/0271678X15610586> PMID: 26661203
 20. Hakim A, Christensen S, Winzeck S, Lansberg MG, Parsons MW, Lucas C, et al. Predicting infarct core from computed tomography perfusion in acute ischemia with machine learning: Lessons from the ISLES Challenge. *Stroke*. 2021; 52(7):2328–2337. <https://doi.org/10.1161/STROKEAHA.120.030696> PMID: 33957774
 21. de la Rosa E, Sima DM, Menze B, Kirschke JS, Robben D. AIFNet: Automatic vascular function estimation for perfusion analysis using deep learning. *Medical Image Analysis*. 2021; 74:102211. <https://doi.org/10.1016/j.media.2021.102211> PMID: 34425318
 22. Lin L, Bivard A, Krishnamurthy V, Levi CR, Parsons MW. Whole-brain CT perfusion to quantify acute ischemic penumbra and core. *Radiology*. 2016; 279(3):876–887. <https://doi.org/10.1148/radiol.2015150319> PMID: 26785041
 23. Menze BH, Kelm BM, Weber MA, Bachert P, Hamprecht FA. Mimicking the human expert: pattern recognition for an automated assessment of data quality in MR spectroscopic images. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*. 2008; 59(6):1457–1466. <https://doi.org/10.1002/mrm.21519> PMID: 18421692
 24. Kyathanahally SP, Mocioiu V, Pedrosa de Barros N, Slotboom J, Wright AJ, Julià-Sapé M, et al. Quality of clinical brain tumor MR spectra judged by humans and machine learning tools. *Magnetic resonance in medicine*. 2018; 79(5):2500–2510. <https://doi.org/10.1002/mrm.26948> PMID: 28994492
 25. Wei L, Rosen B, Vallières M, Chotchutipan T, Mierzwa M, Eisbruch A, et al. Automatic recognition and analysis of metal streak artifacts in head and neck computed tomography for radiomics modeling. *Physics and imaging in radiation oncology*. 2019; 10:49–54. <https://doi.org/10.1016/j.phro.2019.05.001> PMID: 33458268
 26. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*. 1997; 55(1):119–139. <https://doi.org/10.1006/jcss.1997.1504>
 27. Friedman JH. Greedy function approximation: a gradient boosting machine. *Annals of statistics*. 2001; p. 1189–1232.
 28. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*. 2011; 12:2825–2830.
 29. Robben D, Suetens P. Perfusion parameter estimation using neural networks and data augmentation. In: *International MICCAI Brainlesion Workshop*. Springer; 2018. p. 439–446.

30. Kazemitabar J, Amini A, Bloniarz A, Talwalkar AS. Variable importance using decision trees. *Advances in neural information processing systems*. 2017;30.
31. Nogueira RG, Jadhav AP, Haussen DC, Bonafe A, Budzik RF, Bhuva P, et al. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. *New England Journal of Medicine*. 2018; 378(1):11–21. <https://doi.org/10.1056/NEJMoa1706442> PMID: 29129157
32. Albers GW, Marks MP, Kemp S, Christensen S, Tsai JP, Ortega-Gutierrez S, et al. Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging. *New England Journal of Medicine*. 2018; 378(8):708–718. <https://doi.org/10.1056/NEJMoa1713973> PMID: 29364767
33. Fieselmann A, Kowarschik M, Ganguly A, Hornegger J, Fahrig R. Deconvolution-based CT and MR brain perfusion measurement: theoretical model revisited and practical implementation details. *Journal of Biomedical Imaging*. 2011; 2011:14.
34. Kudo K, Sasaki M, Yamada K, Momoshima S, Utsunomiya H, Shirato H, et al. Differences in CT perfusion maps generated by different commercial software: quantitative analysis by using identical source data of acute stroke patients. *Radiology*. 2010; 254(1):200–209. <https://doi.org/10.1148/radiol.254082000> PMID: 20032153
35. Winder A, d'Esterre CD, Menon BK, Fiehler J, Forkert ND. Automatic arterial input function selection in CT and MR perfusion datasets using deep convolutional neural networks. *Medical Physics*. 2020. <https://doi.org/10.1002/mp.14351> PMID: 32583617

PART III

CONCLUDING REMARKS



Discussion

In the previous chapters, I have described the research contributions of this doctoral dissertation. In Chapters 5.1-5.3, the peer-reviewed conference and journal manuscripts produced within this thesis were presented. These works are self-contained and provide a technical, detailed discussion of the targeted research questions, including their limitations and possible improvements. In this chapter, I will provide a big-picture analysis of these contributions, discussing their impact and what comes next.

Chapter 5.1 - Differentiable Deconvolution for Improved Stroke Perfusion Analysis

In Chapter 5.1, I provided insights about the question, “What is the best arterial input function (AIF) that maximizes the deconvolution algorithm’s performance? Can it be automatically estimated with machine-learning models?” The proposed model generated temporal functions that led to more discriminative CBF perfusion maps and, consequently, improved the delineation of core tissue. These results suggest that although SVD deconvolution has been widely investigated for over two decades, there is still room for improving its optimization. Beyond searching for the optimal AIF for SVD deconvolution, the proposed methodology could be adapted to improve the fitting of other parameters of the model. For instance, in Chapter 2, I mentioned that the regularization of the algorithm is impacted by the quality of the analyzed images. Under this consideration, regularization could be performed on a scan-custom basis rather than using a fixed, global value. Through machine-learning models, one could select the optimal regularization term that maximizes perfusion lesion segmentation based on the input image, thus leading to a personalized CTP analysis for each patient and scan.

The type of AIF-generating approach used in this work has advantages for research purposes but has clinical applicability drawbacks. On one hand, since the proposed AIF is formulated as an unconstrained regression problem, it does not impose any prior knowledge on the type and profile of the temporal curve that maximizes core segmentation performance. This allows, for instance, pushing the boundaries of

SVD deconvolution performance and hitting its sweetest optimization spot. On the other hand, the AIF-generating CNN proposes temporal curves that might be unrealistic from a physiological point of view, or ‘invent’ information that is not actually present in the CTP scan. In other words, they are mathematical solutions to the maximization problem that may not be physically plausible. Additionally, the CNN used in this study is non-interpretable (as the curve does not come from the temporal average of specific voxels), and its fixed-length output is a recipe for truncation errors. A potential solution to overcome these drawbacks is to replace the AIF-generating CNN with AIFNet (the method proposed in Chapter 5.2). In such a case, the SVD deconvolution maximization performance could be obtained using a realistic AIF selection method.

Moreover, the methodology proposed in this chapter proved that the time-invariant SVD deconvolution is fully differentiable and, therefore, it can be end-to-end integrated and used within deep-learning frameworks. The possibilities that the differentiable SVD deconvolution opens are endless, as it allows neural networks to be aware of perfusion biomarkers and, as such, allows models to ensure certain brain hemodynamics behaviors. In other words, quantifying perfusion biomarkers (e.g., T_{max}, CBF, etc.) and perfusion lesions (core and penumbra) during the training phase of deep-learning models can be advantageous. Relevant (non-exclusive) applications that could benefit from this work are:

- Low-dose CT perfusion imaging. Reducing the radiation dose delivered to the patient is of high interest to radiologists and neuroradiologists [73]. Reducing the radiation in CTP scans can be achieved in different ways. For instance, by diminishing the CTP acquisition sampling rate (recommended to be 2s or faster) or by reducing the CT tube current. In both scenarios, altering these parameters encompasses degraded image quality that could impact the perfusion volumes and the endovascular treatment decision-making. Hence, machine-learning models can be used in this context to improve or restore high-fidelity CTP scans from low-dose quality ones. Such approaches are not new and were studied, for instance, in [68, 69, 74]. The proposed methodology in Chapter 5.1 could be integrated within such approaches, aiming to devise perfusion-aware CTP restoration models.
- Truncated CTP scan restoration. As shown in Chapter 5.3, the short acquisition of CTP series significantly impacts the perfusion lesion volumes estimated through deconvolution. Restoring the truncated time-attenuation curves is a complex problem that may convert unusable and uninterpretable CTP results

into clinically useful ones. Such a problem could be tackled using supervised or unsupervised machine-learning models that propose a synthetic restoration of the missing CTP time points. The differentiable deconvolution algorithm may play a key role by constraining these algorithms to restore the image series by ensuring physiologically feasible results.

- **Synthetic CTP generation.** Generative artificial intelligence, such as generative adversarial networks and diffusion models, is revolutionizing medical imaging by yielding realistic-looking synthetic samples. Generating synthetic 4D CTP images will allow an increase in perfusion imaging databases. It is noteworthy that perfusion imaging is not easily accessible, as it is mainly restricted to high-complexity healthcare centers. As far as I know, there are only two open acute ischemic stroke CTP datasets, namely ISLES18 [9, 23] and UniToBrain [44, 19]. Consequently, creating synthetic CTP data may considerably impact perfusion imaging research. In this sense, the proposed differentiable deconvolution algorithm may serve to generate brain perfusion plausible datasets.

There are also challenging aspects that should be overcome to investigate some of the works mentioned above. On one hand, access to concomitant CTP and DWI is not easy. Therefore, similar data to ISLES'18 might not be available as it does not follow the recommended acute stroke clinical workflow (typically, either CTP or perfusion MRI with DWI is performed, but rarely both of them). An alternative solution could be to use follow-up imaging (DWI or NCCT) as ground truth. However, in such a case, more sources of variability influence the final infarct ground truth (such as the time to endovascular treatment or the grade of reperfusion achieved through thrombectomy). Furthermore, ground truth for the tissue at risk (i.e., penumbra) is hard to obtain as there is currently no gold standard for it. A solution could be to use follow-up imaging in non-recanalized patients, which provides ground truth for the perfusion lesion and, after subtraction of the core, gives the penumbra [46].

Chapter 5.2 - AIFNet: Automatic Vascular Function Estimation for Perfusion Analysis Using Deep Learning

In this chapter of my dissertation, I provided a novel vascular function selection method based on CNNs by devising a new approach to target the question *Q2 (Can I automatically select vascular functions from clinical CTP scans? In that case, how close to a manual expert performance can I get?)*. *AIFNet* showed outstanding performance for the considered task and produced results close to those achieved

by experts for a wide range of heterogeneous scenarios, including different scanners, vendors, centers, injection protocols, and image quality. This closes the academic-clinical transferability gap for working with real-world data.

An essential next direction for the vascular function selection problem is *quality control*. Checking the AIF-VOF profile and its anatomical placement is highly recommended in the CTP analysis workflow. Hence, the next steps in this matter should consider algorithms that provide *uncertainty* metrics about their performance. This way, quality control could be improved by warning the medical staff of potentially misleading perfusion outputs.

Considering the way that *AIFNet* operates, the method could be extended to other imaging modalities (such as brain perfusion MRI or PET imaging) and potentially to other organs and pathologies (such as myocardial blood-flow quantification from stress perfusion MRI [53]).

Chapter 5.3 - Detecting CTP Truncation Artifacts in Acute Stroke Imaging from the Arterial Input and the Vascular Output Functions

In Chapter 5.3, I recycled the available vascular functions used for deconvolution to provide insights about question *Q3* (*Can I exploit the deconvolution-available vascular functions to automatically detect truncated CTP scans?*). I studied the impact of short CTP acquisitions on the quantified perfusion volumes by using synthetically shortened CTP samples. In this study, I also fit machine-learning models that can identify potentially unreliable, insufficient CTP acquisitions. The proposed models exploited the available AIF and VOF information and provided better truncation detection performance than a baseline model solely considering the scan duration.

A next step for enriching question *Q3* and continuing to improve the performance of these models could be the inclusion of patient-specific information. Several factors, such as the patient's age and weight or certain pathophysiological conditions, impact the profile of the time-attenuation curves (see Chapter 2 for details). Therefore, devising strategies that consider this information may improve truncation detection performance. It is worth mentioning that in most CTP software packages, the vascular functions are automatically detected. As such, there is a clear risk of failing to detect truncation problems when the AIF/VOF curves are suboptimally chosen. Under these situations, the baseline model using the scan duration information could provide a backup strategy.

Another important topic for future research is the quantitative estimation of the truncation impact over the quantified lesion volumes: *How much do the unreliable*

perfusion volumes deviate from their real values? The current truncation detection algorithm has limitations, as it only alerts about unreliable perfusion volumes that deviate by more than 5 ml or more than 10% from the expected real lesion volumes. However, the impact of these errors on endovascular treatment decisions can vary. While they may change treatment decisions in certain cases, they may not impact treatment in others. By providing volumetric error ranges, radiologists and neurointerventionalists can better assess the clinical utility of short acquisitions, despite their unreliability.

Lastly, a natural and challenging next step for question *Q3* is restoring truncated CTP series using data-driven techniques, as discussed earlier in this chapter. In this sense, turning unusable scans into interpretable images with clinical utility can be groundbreaking. Various methods, such as supervised machine-learning or generative models, could be explored to tackle this task. The differentiable deconvolution method proposed in Chapter 5.1 could provide perfusion information to such models and constrain the restoration to plausible perfusion results.



Outlook

Overall, this thesis provides insights into the vascular function components used in stroke perfusion imaging pipelines. The work contributed to the study and characterization of vascular functions using data-driven learning techniques. Three main contributions with high clinical transferability potential are described.

The first contribution addresses the optimal arterial input function selection for deconvolution algorithms. A differentiable implementation of the SVD deconvolution algorithm integrated within a deep-learning scheme is proposed. It demonstrates the possibility of finding a scanner-custom “best” AIF, leading to improved core lesion estimates compared to manual or clinical software-selected AIFs. The second contribution focuses on the automatic selection of vascular functions from clinical CTP scans. A novel methodology called *AIFNet*, based on convolutional neural networks, achieves accurate vascular function selection comparable to expert performance. The method already entered the clinical routine workflow by being implemented in the CTP software package **icobrain cva** (**icometrix**, Leuven, Belgium). The third contribution deals with the detection of truncated CTP scans. Machine-learning models using vascular functions (AIF and VOF) information are used to detect unreliably short CTP scans. This approach proves more accurate than considering scan duration alone.

Nonetheless, there are significant challenges that need to be addressed to further enhance the utility of perfusion imaging in stroke and in neurovascular pathologies in general. Firstly, a notable limitation of perfusion imaging arises from the absence of clear gold standards for core and penumbra tissue. This lack of well-defined ground truth data increases the complexity of the problem and potential solutions. Consequently, end-to-end supervised learning methods may not always be suitable for this imaging modality, given the weak or uncertain nature of the available data. Secondly, current deep-learning models require a large amount of data for training. However, meeting this requirement is challenging since perfusion imaging is not universally available in healthcare centers and, as mentioned earlier, the limited available data is prone to uncertain labeling. Thirdly, perfusion imaging involves

4D imaging data, which is computationally intensive to process. Training current deep-learning models with such data may be constrained by the need for larger GPUs. It is important to note that a perfusion CT scan can be more than 60 times larger than a comparable 3D CT or MRI scan.

These challenges highlight the need for innovative machine-learning approaches in perfusion imaging that are less data-hungry and can effectively learn from uncertain data scenarios. Providing these models with better guidance for the specific task at hand can be immensely helpful. An example of such an approach is the incorporation of perfusion-awareness guidance in the models, as demonstrated in Chapter 5.1. By leveraging complementary feedback information, these models can be optimized even with limited and/or weakly labeled data. There are several other perspectives that can also contribute to addressing these challenges. For example, one approach is to incorporate 4D convolutional blocks, which can better capture the spatio-temporal characteristics of perfusion data. However, extending the 3D convolution operation to include a fourth dimension significantly increases the number of model parameters, which, as previously mentioned, is undesirable. Therefore, an alternative approach is the development of efficient and learnable layers, such as those based on matrix factorization [3], which can facilitate the creation of lighter models. These lighter models would require fewer data for training while still leveraging the inherent 4D nature of the modality. Other machine-learning paradigms, such as generative modeling with techniques like generative adversarial networks or diffusion models, can also play a crucial role. These approaches can help overcome data scarcity by generating synthetic data. Self-supervised learning techniques, instead, can leverage larger amounts of unlabeled and unused data to learn specific perfusion imaging tasks. In conclusion, by exploring these avenues, the above-mentioned technical limitations could be better managed, thus allowing a deeper understanding of acute ischemic stroke and its better management in the clinical workflow.



Bibliography

- [1] Gregory W Albers, Michael P Marks, Stephanie Kemp, Soren Christensen, Jenny P Tsai, Santiago Ortega-Gutierrez, Ryan A McTaggart, Michel T Torbey, May Kim-Tenser, Thabele Leslie-Mazwi, et al. “Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging.” In: *New England Journal of Medicine* 378.8 (2018), pp. 708–718.
- [2] Michela Antonelli, Annika Reinke, Spyridon Bakas, Keyvan Farahani, Annette Kopp-Schneider, Bennett A Landman, Geert Litjens, Bjoern Menze, Olaf Ronneberger, Ronald M Summers, et al. “The medical segmentation decathlon.” In: *Nature communications* 13.1 (2022), p. 4128.
- [3] Pooya Ashtari, Diana M Sima, Lieven De Lathauwer, Dominique Sappey-Marinier, Frederik Maes, and Sabine Van Huffel. “Factorizer: A scalable interpretable approach to context modeling for medical image segmentation.” In: *Medical Image Analysis* 84 (2023), p. 102706.
- [4] Leon Axel. “Cerebral blood flow determination by rapid-sequence computed tomography: theoretical analysis.” In: *Radiology* 137.3 (1980), pp. 679–686.
- [5] Elisabetta Bandera, Marco Botteri, Cosetta Minelli, Alex Sutton, Keith R Abrams, and Nicola Latronico. “Cerebral blood flow threshold of ischemic penumbra and infarct core in acute ischemic stroke: a systematic review.” In: *Stroke* 37.5 (2006), pp. 1334–1339.
- [6] Eivind Berge, William Whiteley, Heinrich Audebert, Gian Marco De Marchis, Ana Catarina Fonseca, Chiara Padiglioni, Natalia Pérez de la Ossa, Daniel Strbian, Georgios Tsivgoulis, and Guillaume Turc. “European Stroke Organisation (ESO) guidelines on intravenous thrombolysis for acute ischaemic stroke.” In: *European stroke journal* 6.1 (2021), pp. I–LXII.
- [7] Fernando Calamante. “Arterial input function in perfusion MRI: a comprehensive review.” In: *Progress in nuclear magnetic resonance spectroscopy* 74 (2013), pp. 1–32.

- [8] Timothy J Carroll, Howard A Rowley, and Victor M Haughton. “Automatic calculation of the arterial input function for cerebral perfusion imaging with MR imaging.” In: *Radiology* 227.2 (2003), pp. 593–600.
- [9] Carlo W Cereda, Søren Christensen, Bruce CV Campbell, Nishant K Mishra, Michael Mlynash, Christopher Levi, Matus Straka, Max Wintermark, Roland Bammer, Gregory W Albers, et al. “A benchmarking tool to evaluate computer tomography perfusion infarct core predictions against a DWI standard.” In: *Journal of Cerebral Blood Flow & Metabolism* 36.10 (2016), pp. 1780–1789.
- [10] Soren Christensen and Maarten G Lansberg. “CT perfusion in acute stroke: practical guidance for implementation in clinical practice.” In: *Journal of Cerebral Blood Flow & Metabolism* 39.9 (2019), pp. 1664–1668.
- [11] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. “3D U-Net: learning dense volumetric segmentation from sparse annotation.” In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*. Springer. 2016, pp. 424–432.
- [12] WA Copen, AR Deipolyi, PW Schaefer, LH Schwamm, RG González, and O Wu. “Exposing hidden truncation-related errors in acute stroke perfusion imaging.” In: *American Journal of Neuroradiology* 36.4 (2015), pp. 638–645.
- [13] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. “Imagenet: A large-scale hierarchical image database.” In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee. 2009, pp. 248–255.
- [14] Ulrich Dirnagl, Costantino Iadecola, and Michael A Moskowitz. “Pathobiology of ischaemic stroke: an integrated view.” In: *Trends in neurosciences* 22.9 (1999), pp. 391–397.
- [15] Jonathan Emberson, Kennedy R Lees, Patrick Lyden, Lisa Blackwell, Gregory Albers, Erich Bluhmki, Thomas Brott, Geoff Cohen, Stephen Davis, Geoffrey Donnan, et al. “Effect of treatment delay, age, and stroke severity on the effects of intravenous thrombolysis with alteplase for acute ischaemic stroke: a meta-analysis of individual patient data from randomised trials.” In: *The Lancet* 384.9958 (2014), pp. 1929–1935.
- [16] Shengyu Fan, Yueyan Bian, Erling Wang, Yan Kang, Danny JJ Wang, Qi Yang, and Xunming Ji. “An automatic estimation of arterial input function based on multi-stream 3D CNN.” In: *Frontiers in neuroinformatics* 13 (2019), p. 49.

-
- [17] Andreas Fieselmann, Markus Kowarschik, Arundhuti Ganguly, Joachim Hornegger, and Rebecca Fahrig. “Deconvolution-based CT and MR brain perfusion measurement: theoretical model revisited and practical implementation details.” In: *Journal of Biomedical Imaging* 2011 (2011), pp. 1–20.
- [18] Kunihiko Fukushima. “Cognitron: A self-organizing multilayered neural network.” In: *Biological cybernetics* 20.3-4 (1975), pp. 121–136.
- [19] Umberto Gava, Federico D’Agata, Bennink Edwin, Enzo Tartaglione, Daniele Perlo, Annamaria Vernone, Francesca Bertolino, Eleonora Ficiarà, Alessandro Cicerale, Fabrizio Pizzagalli, et al. “UniToBrain Dataset.” In: (2021).
- [20] GT Gobbel and JR Fike. “A deconvolution method for evaluating indicator-dilution curves.” In: *Physics in Medicine & Biology* 39.11 (1994), p. 1833.
- [21] Camilo R Gomez. *Time is brain!* 1993.
- [22] Mayank Goyal, Bijoy K Menon, Wim H Van Zwam, Diederik WJ Dippel, Peter J Mitchell, Andrew M Demchuk, Antoni Dávalos, Charles BLM Majoie, Aad van Der Lugt, Maria A De Miquel, et al. “Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials.” In: *The Lancet* 387.10029 (2016), pp. 1723–1731.
- [23] Arsany Hakim, Søren Christensen, Stefan Winzeck, Maarten G Lansberg, Mark W Parsons, Christian Lucas, David Robben, Roland Wiest, Mauricio Reyes, and Greg Zaharchuk. “Predicting infarct core from computed tomography perfusion in acute ischemia with machine learning: Lessons from the ISLES challenge.” In: *Stroke* 52.7 (2021), pp. 2328–2337.
- [24] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Deep residual learning for image recognition. arXiv 2015.” In: *arXiv preprint arXiv:1512.03385* 14 (2015).
- [25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification.” In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1026–1034.
- [26] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation.” In: *Nature methods* 18.2 (2021), pp. 203–211.

- [27] AA Konstas, GV Goldmakher, T-Y Lee, and MH Lev. “Theoretic basis and technical implementations of CT perfusion in acute ischemic stroke, part 1: theoretic basis.” In: *American Journal of Neuroradiology* 30.4 (2009), pp. 662–668.
- [28] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “Imagenet classification with deep convolutional neural networks.” In: *Advances in neural information processing systems* 25.1106-1114 (2012), p. 1.
- [29] Kohsuke Kudo, Makoto Sasaki, Kei Yamada, Suketaka Momoshima, Hidetsuna Utsunomiya, Hiroki Shirato, and Kuniaki Ogasawara. “Differences in CT perfusion maps generated by different commercial software: quantitative analysis by using identical source data of acute stroke patients.” In: *Radiology* 254.1 (2010), pp. 200–209.
- [30] Kenneth B Larson, William H Perman, Joel S Perlmutter, Mokhtar H Gado, John M Ollinger, and Kenneth Zierler. “Tracer-kinetic analysis for measuring regional cerebral blood flow by dynamic nuclear magnetic resonance imaging.” In: *Journal of theoretical biology* 170.1 (1994), pp. 1–14.
- [31] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. “Gradient-based learning applied to document recognition.” In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.
- [32] Sook-Lei Liew, Bethany P Lo, Miranda R Donnelly, Artemis Zavaliangos-Petropulu, Jessica N Jeong, Giuseppe Barisano, Alexandre Hutton, Julia P Simon, Julia M Juliano, Anisha Suri, (...), Ezequiel de la Rosa, et al. “A large, curated, open-source stroke neuroimaging dataset to improve lesion segmentation algorithms.” In: *Scientific data* 9.1 (2022), p. 320.
- [33] Andrew L Maas, Awni Y Hannun, Andrew Y Ng, et al. “Rectifier nonlinearities improve neural network acoustic models.” In: *Proc. icml*. Vol. 30. 1. Atlanta, Georgia, USA. 2013, p. 3.
- [34] Oskar Maier, Bjoern H Menze, Janina Von der Gablentz, Levin Häni, Mattias P Heinrich, Matthias Liebrand, Stefan Winzeck, Abdul Basit, Paul Bentley, Liang Chen, et al. “ISLES 2015-A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI.” In: *Medical image analysis* 35 (2017), pp. 250–269.

-
- [35] Rajiv Mangla, Sven Ekholm, Babak S Jahromi, Jeevak Almast, Manisha Mangla, and Per-Lennart Westesson. “CT perfusion in acute stroke: know the mimics, potential pitfalls, artifacts, and technical errors.” In: *Emergency radiology* 21 (2014), pp. 49–65.
- [36] Mark P Mattson. “Excitotoxicity.” In: *Neurodegeneration* (2017), pp. 37–45.
- [37] Kenya Murase, Keiichi Kikuchi, Hitoshi Miki, Teruhiko Shimizu, and Junpei Ikezoe. “Determination of arterial input function using fuzzy clustering for quantification of cerebral blood flow with dynamic susceptibility contrast-enhanced MR imaging.” In: *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 13.5 (2001), pp. 797–806.
- [38] BD Murphy, X Chen, and T-Y Lee. “Serial changes in CT cerebral blood volume and flow after 4 hours of middle cerebral occlusion in an animal model of embolic cerebral ischemia.” In: *American Journal of Neuroradiology* 28.4 (2007), pp. 743–749.
- [39] National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. “Tissue plasminogen activator for acute ischemic stroke.” In: *New England Journal of Medicine* 333.24 (1995), pp. 1581–1588.
- [40] Keqing Nian, Ian C Harding, Ira M Herman, and Eno E Ebong. “Blood-brain barrier damage in ischemic stroke and its regulation by endothelial mechanotransduction.” In: *Frontiers in Physiology* 11 (2020), p. 605398.
- [41] Raul G Nogueira, Ashutosh P Jadhav, Diogo C Haussen, Alain Bonafe, Ronald F Budzik, Parita Bhuvra, Dileep R Yavagal, Marc Ribo, Christophe Cognard, Ricardo A Hanel, et al. “Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct.” In: *New England Journal of Medicine* 378.1 (2018), pp. 11–21.
- [42] David Norman, Leon Axel, Walter H Berninger, Michael S Edwards, Christopher E Cann, Rowland W Redington, and Lauranne Cox. “Dynamic computed tomography of the brain: techniques, data analysis, and applications.” In: *American Journal of Neuroradiology* 2.1 (1981), pp. 1–12.
- [43] Leif Østergaard, Robert M Weisskoff, David A Chesler, Carsten Gyldensted, and Bruce R Rosen. “High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. Part I: Mathematical approach and statistical analysis.” In: *Magnetic resonance in medicine* 36.5 (1996), pp. 715–725.

- [44] Daniele Perlo, Enzo Tartaglione, Umberto Gava, Federico D’Agata, Edwin Beninck, and Mauro Bergui. “UniToBrain Dataset: A Brain Perfusion Dataset.” In: *Image Analysis and Processing. ICIAP 2022 Workshops: ICIAP International Workshops, Lecce, Italy, May 23–27, 2022, Revised Selected Papers, Part I*. Springer. 2022, pp. 498–509.
- [45] Wu Qiu, Hulin Kuang, Ericka Teleg, Johanna M Ospel, Sung Il Sohn, Mohammed Almekhlafi, Mayank Goyal, Michael D Hill, Andrew M Demchuk, and Bijoy K Menon. “Machine learning for detecting early infarction in acute stroke with non-contrast-enhanced CT.” In: *Radiology* 294.3 (2020), pp. 638–644.
- [46] RA Rava, KV Snyder, M Mokin, M Waqas, AB Allman, JL Senko, AR Podgorsak, MM Shiraz Bhurwani, Y Hoi, AH Siddiqui, et al. “Assessment of a Bayesian Vitrea CT perfusion analysis to predict final infarct and penumbra volumes in patients with acute ischemic stroke: a comparison with RAPID.” In: *American Journal of Neuroradiology* 41.2 (2020), pp. 206–212.
- [47] Ryan A Rava, Kenneth V Snyder, Maxim Mokin, Muhammad Waqas, Ariana B Allman, Jillian L Senko, Alexander R Podgorsak, Mohammad Mahdi Shiraz Bhurwani, Jason M Davies, Elad I Levy, et al. “Effect of computed tomography perfusion post-processing algorithms on optimal threshold selection for final infarct volume prediction.” In: *The neuroradiology journal* 33.4 (2020), pp. 273–285.
- [48] Katrin A Rempp, Gunnar Brix, Frederik Wenz, Christoph R Becker, Friedemann Gückel, and Walter J Lorenz. “Quantification of regional cerebral blood flow and volume with dynamic susceptibility contrast-enhanced MR imaging.” In: *Radiology* 193.3 (1994), pp. 637–641.
- [49] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation.” In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III* 18. Springer. 2015, pp. 234–241.
- [50] Ezequiel de la Rosa and David Robben. “Computer-implemented method, system and computer program product for determining a vascular function of a perfusion imaging sequence.” European pat. EP3910587A1. Nov. 17, 2021.
- [51] Frank Rosenblatt. *Principles of neurodynamics. perceptrons and the theory of brain mechanisms*. Tech. rep. Cornell Aeronautical Lab Inc Buffalo NY, 1961.

-
- [52] Sean I Savitz, Jean-Claude Baron, Midori A Yenari, Nerses Sanossian, and Marc Fisher. “Reconsidering neuroprotection in the reperfusion era.” In: *Stroke* 48.12 (2017), pp. 3413–3419.
- [53] Cian M Scannell, Ebrahim Alskaf, Noor Sharrack, Reza Razavi, Sebastien Ourselin, Alistair A Young, Sven Plein, and Amedeo Chiribiri. “AI-AIF: artificial intelligence-based arterial input function for quantitative stress perfusion cardiac magnetic resonance.” In: *European Heart Journal-Digital Health* 4.1 (2023), pp. 12–21.
- [54] Lin Shi, Defeng Wang, Wen Liu, Kui Fang, Yi-Xiang J Wang, Wenhua Huang, Ann D King, Pheng Ann Heng, and Anil T Ahuja. “Automatic detection of arterial input function in dynamic contrast enhanced MRI based on affinity propagation clustering.” In: *Journal of Magnetic Resonance Imaging* 39.5 (2014), pp. 1327–1337.
- [55] Karen Simonyan and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition.” In: *arXiv preprint arXiv:1409.1556* (2014).
- [56] Houman Sotoudeh, Asim K Bag, and Michael David Brooks. ““Code-stroke” CT perfusion; challenges and pitfalls.” In: *Academic Radiology* 26.11 (2019), pp. 1565–1579.
- [57] S Sourbron, Robert Luypaert, D Morhard, K Seelos, M Reiser, and M Peller. “Deconvolution of bolus-tracking data: a comparison of discretization methods.” In: *Physics in Medicine & Biology* 52.22 (2007), p. 6761.
- [58] Steven Sourbron, Robert Luypaert, Peter Van Schuerbeek, Martine Dujardin, and Tadeusz Stadnik. “Choice of the regularization parameter for perfusion quantification with MRI.” In: *Physics in Medicine & Biology* 49.14 (2004), p. 3307.
- [59] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. “Going deeper with convolutions (2014).” In: *arXiv preprint arXiv:1409.4842* 10 (2014).
- [60] Götz Thomalla, Claus Z Simonsen, Florent Boutitie, Grethe Andersen, Yves Berthezene, Bastian Cheng, Bharath Cheripelli, Tae-Hee Cho, Franz Fazekas, Jens Fiehler, et al. “MRI-guided thrombolysis for stroke with unknown time of onset.” In: *New England Journal of Medicine* 379.7 (2018), pp. 611–622.

- [61] Guillaume Turc, Pervinder Bhogal, Urs Fischer, Pooja Khatri, Kyriakos Lobotesis, Mikaël Mazighi, Peter D Schellinger, Danilo Toni, Joost De Vries, Philip White, et al. “European stroke organisation (ESO)–European society for minimally invasive neurological therapy (ESMINT) guidelines on mechanical thrombectomy in acute ischaemic Stroke Endorsed by stroke alliance for Europe (SAFE).” In: *European stroke journal* 4.1 (2019), pp. 6–12.
- [62] Achala Vagal, Max Wintermark, Kambiz Nael, Andrew Bivard, Mark Parsons, Aaron W Grossman, and Pooja Khatri. “Automated CT perfusion imaging for acute ischemic stroke: pearls and pitfalls for real-world use.” In: *Neurology* 93.20 (2019), pp. 888–898.
- [63] Salim S Virani, Alvaro Alonso, Emelia J Benjamin, Marcio S Bittencourt, Clifton W Callaway, April P Carson, Alanna M Chamberlain, Alexander R Chang, Susan Cheng, Francesca N Delling, et al. “Heart disease and stroke statistics—2020 update: a report from the American Heart Association.” In: *Circulation* 141.9 (2020), e139–e596.
- [64] Evert-jan PA Vonken, Freek J Beekman, Chris JG Bakker, and Max A Viergever. “Maximum likelihood estimation of cerebral blood flow in dynamic susceptibility contrast MRI.” In: *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 41.2 (1999), pp. 343–350.
- [65] Kristin Walter. “What is acute ischemic stroke?” In: *JAMA* 327.9 (2022), pp. 885–885.
- [66] Wenjuan Wang, Martin Kiik, Niels Peek, Vasa Curcin, Iain J Marshall, Anthony G Rudd, Yanzhong Wang, Abdel Douiri, Charles D Wolfe, and Benjamin Bray. “A systematic review of machine learning models for predicting outcomes of stroke with structured data.” In: *PloS one* 15.6 (2020), e0234722.
- [67] Anthony Winder, Christopher D d’Esterre, Bijoy K Menon, Jens Fiehler, and Nils D Forkert. “Automatic arterial input function selection in CT and MR perfusion datasets using deep convolutional neural networks.” In: *Medical Physics* 47.9 (2020), pp. 4199–4211.
- [68] Yao Xiao, Ajay Gupta, Pina C Sanelli, and Ruogu Fang. “STAR: spatio-temporal architecture for super-resolution in low-dose CT perfusion.” In: *Machine Learning in Medical Imaging: 8th International Workshop, MLMI 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 10, 2017, Proceedings 8*. Springer. 2017, pp. 97–105.

-
- [69] Yao Xiao, Peng Liu, Yun Liang, Skylar Stolte, Pina Sanelli, Ajay Gupta, Jana Ivanidze, and Ruogu Fang. “STIR-net: deep spatial-temporal image restoration net for radiation reduction in CT perfusion.” In: *Frontiers in neurology* 10 (2019), p. 647.
- [70] Yunyun Xiong, Chiwen C Huang, Marc Fisher, David B Hackney, Rafeeqe A Bhadelia, and Magdy H Selim. “Comparison of automated CT perfusion softwares in evaluation of acute ischemic stroke.” In: *Journal of Stroke and Cerebrovascular Diseases* 28.12 (2019), p. 104392.
- [71] Yunyun Xiong, Ajay K Wakhloo, and Marc Fisher. “Advances in acute ischemic stroke therapy.” In: *Circulation Research* 130.8 (2022), pp. 1230–1251.
- [72] Jiandong Yin, Hongzan Sun, Jiawen Yang, and Qiyong Guo. “Automated detection of the arterial input function using normalized cut clustering to determine cerebral perfusion by dynamic susceptibility contrast-magnetic resonance imaging.” In: *Journal of Magnetic Resonance Imaging* 41.4 (2015), pp. 1071–1078.
- [73] Sebastian Zensen, Nika Guberina, Marcel Opitz, Martin Köhrmann, Cornelius Deuschl, Michael Forsting, Axel Wetter, and Denise Bos. “Radiation exposure of computed tomography imaging for the assessment of acute stroke.” In: *Neuroradiology* 63 (2021), pp. 511–518.
- [74] Haichen Zhu, Dan Tong, Lu Zhang, Shijie Wang, Weiwen Wu, Hui Tang, Yang Chen, Limin Luo, Jian Zhu, and Baosheng Li. “Temporally downsampled cerebral CT perfusion image restoration using deep residual learning.” In: *International journal of computer assisted radiology and surgery* 15 (2020), pp. 193–201.

PART IV
APPENDICES



**Supplementary Material:
Differentiable Deconvolution for
Improved Stroke Perfusion
Analysis**

– SUPPLEMENTARY MATERIAL –
Differentiable Deconvolution for Improved
Stroke Perfusion Analysis

Ezequiel de la Rosa^{1,2}, David Robben^{1,3}, Diana M. Sima¹, Jan S. Kirschke⁴,
and Bjoern Menze²

¹ icometrix, Leuven, Belgium

{ezequiel.delarosa, david.robben, diana.sima}@icometrix.com

² Department of Computer Science, Technical University of Munich, Munich,
Germany

bjoern.menze@tum.de

³ Medical Imaging Research Center (MIRC), KU Leuven, Leuven, Belgium

⁴ Neuroradiology, School of Medicine, Technical University of Munich, Munich,
Germany

jan.kirschke@tum.de

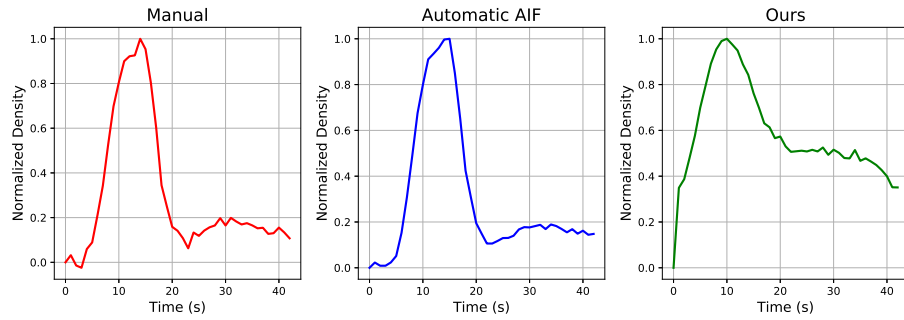


Fig. S1. Selected and generated AIFs for the example case displayed in the main text (i.e., Fig. 2). AIF: Arterial input function. Note that our generated AIF has consistently different shape and is shifted in time when compared against the manual and automatically chosen ones.



**Supplementary Material: AIFNet:
Automatic vascular function
estimation for perfusion analysis
using deep learning**

Supplementary Materials

AIFNet: Automatic Vascular Function Estimation for Perfusion Analysis Using Deep Learning

Ezequiel de la Rosa^{a,b,*}, Diana M. Sima^a, Bjoern Menze^b, Jan S. Kirschke^c, David Robben^{a,d,e}

^aicometrix, Leuven, Belgium

^bDepartment of Computer Science, Technical University of Munich, Munich, Germany

^cNeuroradiology, School of Medicine, Technical University of Munich, Munich, Germany

^dMedical Imaging Research Center (MIRC), KU Leuven, Leuven, Belgium

^eMedical Image Computing (MIC), ESAT-PSI, Department of Electrical Engineering, KU Leuven, Leuven, Belgium

*Corresponding author

Email address: ezequiel.delarosa@icometrix.com (Ezequiel de la Rosa)

| | Signal | | T_{peak} | | FWHM | | |
|--------------------|----------------------------|---------------------------------|-----------------------|---------------------------|---------------------------|---------------------------|---------------------------------|
| | r | r (P5th, P95th) | r | Error (s) | r | Error (s) | |
| Inter-rater | 0.971 (0.075) [†] | (0.883, 1) | 0.964 | -0.14 (1.29) | 0.902 | -0.08 (1.74) [†] | |
| \hat{Y}_{Kmeans} | 0.612 (0.329) [†] | (-0.108, 0.974) | 0.657 | -5.24 (5.69) [†] | 0.421 | -1.88 (6.05) [†] | |
| y_{r2} vs | \hat{Y}_{Seg} | 0.683 (0.155) [†] | (0.419, 0.879) | 0.846 | -3.81 (3.47) [†] | 0.564 | -4.35 (5.14) [†] |
| | \hat{Y}_{Reg} | 0.843 (0.240) [†] | (0.545, 0.981) | 0.747 | -0.15(3.25) [†] | 0.471 | -2.77 (3.51) [†] |
| | \hat{Y}_{AIFNet} | 0.969 (0.04)[†] | (0.884, 0.997) | 0.951 | -0.41(1.60) | 0.853 | -0.81 (2.14)[†] |

Table S1. AIF agreement among methods. Mean (standard deviation) provided. Note that the AIF agreement is measured over the time domain only, since a posterior signal recalibration using the VOF is required due to the high partial volume effect in the arteries. $y_{r2}(t)$: AIF annotated by raters 1 and 2; \hat{Y}_{Kmeans} , \hat{Y}_{Seg} , \hat{Y}_{Reg} , \hat{Y}_{AIFNet} : AIF estimated with the Kmeans approach (Mouridsen et al., 2006), with the regression CNN (de la Rosa et al., 2020), with the segmentation CNN and with AIFNet, respectively. r: Pearson’s correlation coefficient; T_{peak} : time at which the peak of the curve occurs; FWHM: full-width at half-maximum. P5th: 5th percentile; P95th: 95th percentile. Paired significance tests are performed between AIFNet and the other approaches. *: p-value < 0.05; [†]: p-value < 0.01. The values in bold indicate the outperforming method for the metric under consideration.

| | | Pearson's r coefficient | | | |
|--------------------|--------------------|-------------------------|----------------------|----------------------|----------------------|
| | | CBF | CBV | T _{max} | MTT |
| y _{r2} vs | \hat{y}_{Kmeans} | 0.968 (0.098) | 0.973 (0.224) | 0.785 (0.217) | 0.784 (0.266) |
| | \hat{y}_{Seg} | 0.976 (0.115) | 0.960 (0.273) | 0.784 (0.236) | 0.758 (0.294) |
| | \hat{y}_{Reg} | 0.991 (0.018) | 0.984 (0.025) | 0.809 (0.174) | 0.796 (0.213) |
| | \hat{y}_{AIFNet} | 0.998 (0.006) | 0.988 (0.148) | 0.917 (0.087) | 0.878 (0.248) |

Table S2. Parameter maps agreement among methods. Mean (standard deviation) values provided. Correlation has been computed per scan within the brain region (excluding background, ventricles and vessels). y_{r2}: Parameter maps obtained after deconvolving the images with the AIF of rater 2; \hat{y}_{Kmeans} , \hat{y}_{Seg} , \hat{y}_{Reg} , \hat{y}_{AIFNet} : Parameter maps obtained after deconvolving the images with the AIF predicted with Mouridsen et al. (2006), with the regression CNN (de la Rosa et al., 2020), with the segmentation CNN and with AIFNet, respectively. CBF: cerebral blood flow; CBV: cerebral blood volume; MTT: mean transit time; T_{max}: time to the maximum of the residue function. The values in bold indicate the outperforming method (in terms of Pearson's r) for each parameter map.

| | | Hypoperfused | | | Core | | |
|--------------------|--------------------|--------------------------|----------------------------|---------------------------|--------------------------|--------------------------|---------------------------|
| | | Dice (%) | VE (ml) | AVE (ml) | Dice (%) | VE (ml) | AVE (ml) |
| Inter-rater | | 91.7 (13.8) [†] | 0.5 (13.5) [†] | 6.3 (12.0) [†] | 91.8 (14.1) [†] | 0.1 (2.5) [†] | 0.9 (2.3) [†] |
| y _{r2} vs | \hat{y}_{Kmeans} | 48.8 (28.4) [†] | -37.4 (170.8) | 80.5 (155.2) [†] | 58.4 (27.2) [†] | -2.7 (54.3) [†] | 13.0 (52.8) [†] |
| | \hat{y}_{Seg} | 51.2 (19.6) [†] | 26.1 (81.2) [†] | 45.7 (72.0) [†] | 62.0 (20.9) [†] | 1.6 (53.9) [†] | 11.9 (52.6) [†] |
| | \hat{y}_{Reg} | 70.0 (23.5) [†] | -79.0 (198.4) [†] | 94.6 (191.4) [†] | 71.7 (23.8) [†] | -32.1 (136.8) | 35.9 (135.8) [†] |
| | \hat{y}_{AIFNet} | 87.5 (12.9) | 8.4 (19.7) | 11.5 (18.0) | 87.8 (14.1) | 0.4 (1.9) | 1.2 (1.5) |

Table S3. Brain lesion quantification performance for all the methods, obtained after deconvolving the CTP images with the annotated and predicted vascular functions. Mean (standard deviation) values are provided. Hypoperfused tissue is obtained after thresholding T_{max} < 6 s. Core tissue is obtained after thresholding the rCBF map at 38% over the entire hypoperfused region. y_{r2}: Lesion volumes obtained after deconvolving the images with the AIF of rater 2; \hat{y}_{Kmeans} , \hat{y}_{Seg} , \hat{y}_{Reg} , \hat{y}_{AIFNet} : Lesion volumes obtained after deconvolving the images with the AIF predicted with Mouridsen et al. (2006), with the regression CNN (de la Rosa et al., 2020), with the segmentation CNN and with AIFNet, respectively. VE: Volume error; AVE: Absolute volume error. Paired significance tests are performed between AIFNet and the other approaches. *: p-value < 0.05; †: p-value < 0.01. The values in bold indicate the outperforming method for the metric under consideration.

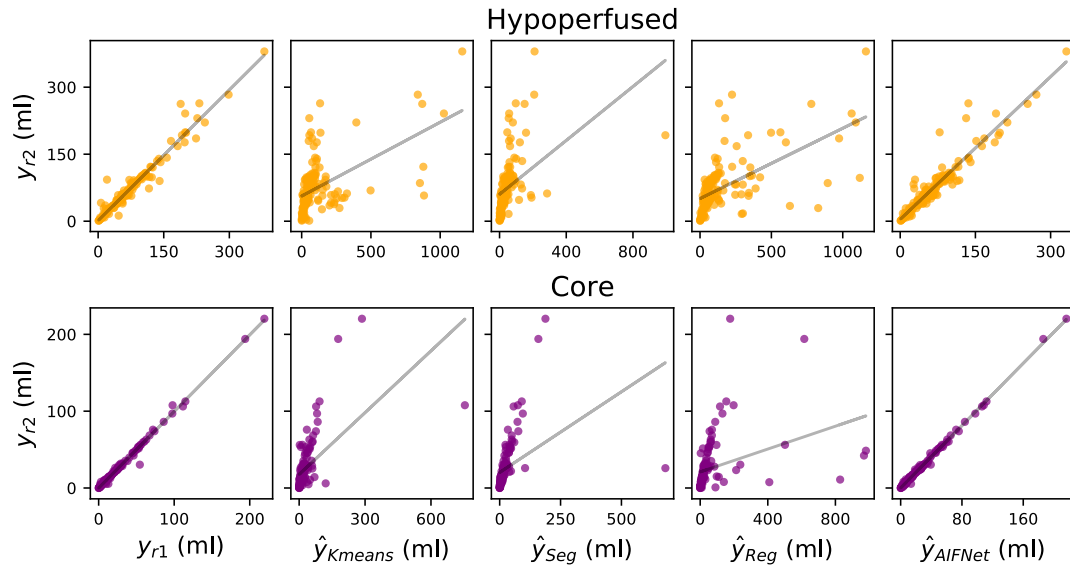


Figure S1. Lesion volume agreement among methods. y_{r1} - y_{r2} : Volumes obtained after deconvolving the images with the AIF of raters 1 and 2; \hat{y}_{Kmeans} , \hat{y}_{Seg} , \hat{y}_{Reg} , \hat{y}_{AIFNet} : Volumes obtained after deconvolving the images with the AIF predicted with Mouridsen et al. (2006), with the regression CNN (de la Rosa et al. (2020)), with the segmentation CNN and with AIFNet, respectively.

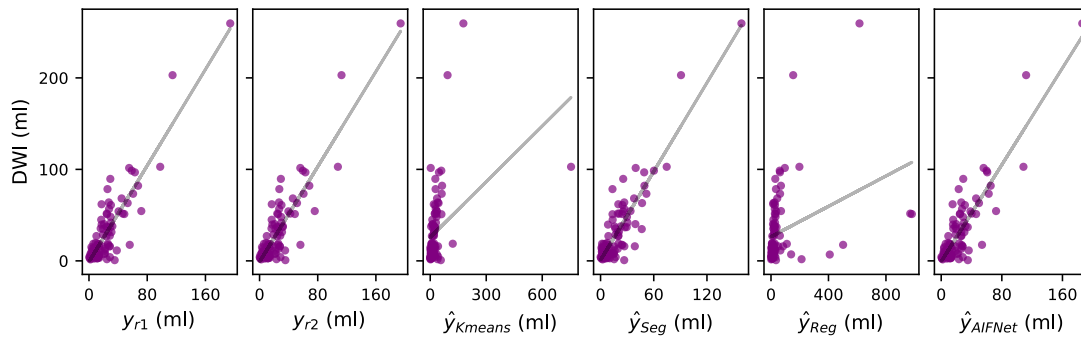


Figure S2. Lesion volume agreement among methods. y_{r1} - y_{r2} : Volumes obtained after deconvolving the images with the AIF of raters 1 and 2; \hat{y}_{Kmeans} , \hat{y}_{Seg} , \hat{y}_{Reg} , \hat{y}_{AIFNet} : Volumes obtained after deconvolving the images with the AIF predicted with Mouridsen et al. (2006), with the regression CNN (de la Rosa et al. (2020)), with the segmentation CNN and with AIFNet, respectively. DWI: Ground truth delineated in diffusion weighted imaging and provided in ISLES18.



**Supplementary Material: Detecting
CTP truncation artifacts in acute
stroke imaging from the arterial
input and the vascular output
functions**

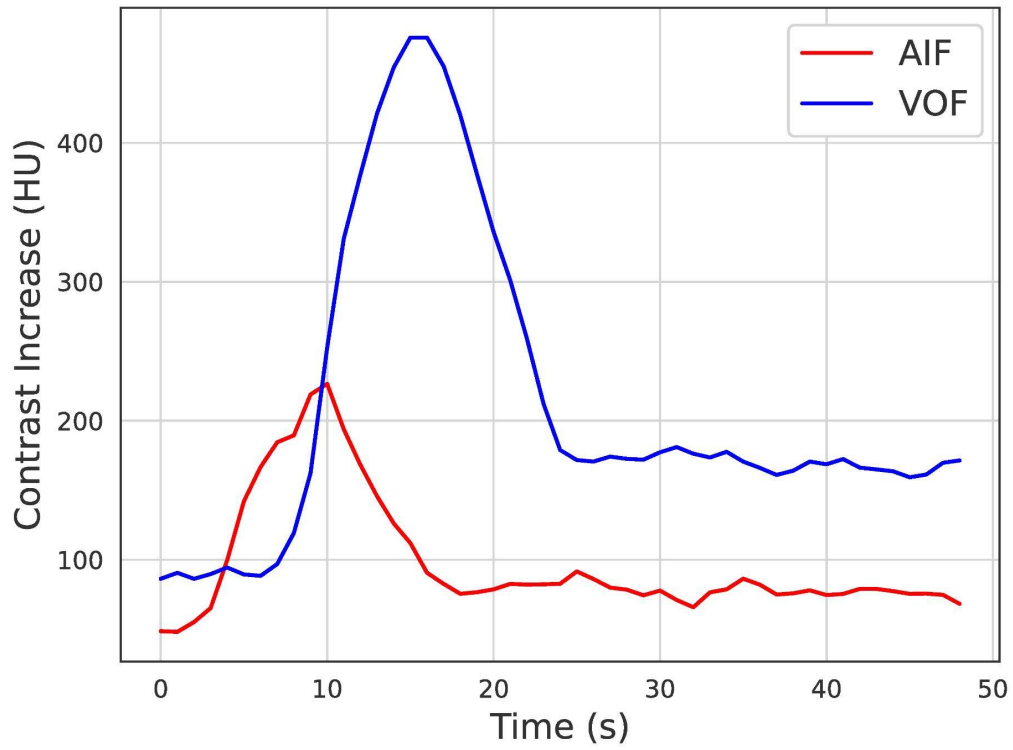


Fig. S1. Example case of an ISLES'18 scan with no pre-contrast acquisition.
AIF: Arterial input function; VOF: Venous output function.



A Radiomics Approach to Traumatic Brain Injury Prediction in CT Scans

A RADIOMICS APPROACH TO TRAUMATIC BRAIN INJURY PREDICTION IN CT SCANS

Ezequiel de la Rosa^{1,2}, Diana M. Sima¹, Thijs Vande Vyvere¹, Jan S. Kirschke², Bjoern Menze²

¹icometrix, Leuven, Belgium

²Department of Computer Science, Technical University of Munich, Germany

ABSTRACT

Computer Tomography (CT) is the gold standard technique for brain damage evaluation after acute Traumatic Brain Injury (TBI). It allows identification of most lesion types and determines the need of surgical or alternative therapeutic procedures. However, the traditional approach for lesion classification is restricted to visual image inspection. In this work, we characterize and predict TBI lesions by using CT-derived radiomics descriptors. Relevant *shape*, *intensity* and *texture* biomarkers characterizing the different lesions are isolated and a lesion predictive model is built by using Partial Least Squares. On a dataset containing 155 scans (105 train, 50 test) the methodology achieved 89.7% accuracy over the unseen data. When a model was built using only texture features, a 88.2% accuracy was obtained. Our results suggest that selected radiomics descriptors could play a key role in brain injury prediction. Besides, the proposed methodology is close to reproduce radiologists lesion labelling. These results open new possibilities for radiomics-inspired brain lesion detection, segmentation and prediction.

Index Terms— Traumatic Brain Injury, CT, Radiomics

1. INTRODUCTION

Traumatic Brain Injury (TBI) is a complex disease process that encompasses a whole spectrum of different pathologies. In the acute phase after injury, non-contrast Computed Tomography (CT) is the most commonly used imaging modality. It can detect most abnormalities, but is especially important for identifying the presence of large extra- or intra-axial space-occupying lesions (i.e. subdural hematomas, epidural hematomas or contusions). For instance, extra-axial hematomas that cause mass effect (i.e. basal cistern compression and midline shift) may need urgent neurosurgical evacuation. On the other hand, contusions may require a non-surgical treatment approach. In this regard, detection and classification of these lesions is of paramount importance in the medical decision-making process [1].

Radiomics, “the conversion of digital medical images into mineable high-dimensional data” [2] takes advantage of image analysis techniques for describing underlying physiopathology in medical scans. As a result, a descriptive

feature vector is obtained and subsequent interpretation by computer driven techniques is performed. For dealing with the high-dimensional space, dimensionality reduction methods turns crucial. In the last few years, the multivariate Partial Least Squares (PLS) method for analyzing radiomics-derived descriptors has been explored. The technique is extensively used in the *OMICS* field, since it is suitable for problems where the number of features is larger than the number of samples.

In this work, we aim to characterize and predict TBI lesions using CT-derived descriptors. Previous works detect them using basic density and shape features [3] or classify TBI scans using deep-learning [4]. However, the multivariate characterization and interpretation of these lesions remain unaddressed. The main contributions of this work are *i*) identification of distinctive radiomic biomarkers characterizing brain lesions, *ii*) fitting of PLS models that accurately predict lesion classes and *iii*) exploring whether *texture*-based models outperform *shape* and *intensity*-based ones for the desired task.

2. MATERIALS AND METHODS

2.1. Database

In this work, data from the CENTER-TBI study (www.center-tbi.com, NCT02210221) coming from more than 50 academic and non-academic centres in 20 countries was used. CT volumes were obtained from several scanners from all major manufacturers, including General Electric, Siemens, Philips and Toshiba. Images were acquired following several acquisition and reconstruction parameters, with variations on slice thickness, pixel spacing and scanner settings, among others. A sub-cohort of 3D volumes was randomly chosen, including three type of brain mass lesions: *i*) epidural hematoma (class 1), *ii*) acute subdural hematoma (class 2) and *iii*) contusion (class 3). 105 scans (train set) were used for data analysis, characterization, and model fitting. Besides, we included 50 extra scans (test set) for validating our models. The training (test) set contains 170 (69) annotated lesions, with 49 (20) in class #1, 54 (20) in class #2 and 67 (29) in class #3. Volume’s slices were delineated in 3D Slicer (<https://www.slicer.org/>) by three trained operators, supervised by an experienced neuroradiologist.

2.2. Radiomics Feature Extraction

With the aim of extracting robust biomarkers insensitive to scan acquisition and reconstruction parameters, all scans were resliced into an homogenous voxel dimension of [1x1x1 mm³]. As demonstrated in [5], voxel resampling is a strongly recommended step for obtaining reproducible radiomic descriptors. For each lesion present in the scan we extracted radiomic features using the ground truth segmentations. With this aim, the open-source *PyRadiomics* [6] library was used and 105 features from the original volumes were obtained grouped as follows: *i*) Shape (13), *ii*) First Order Statistics (FOS, 18), *iii*) Gray-level Co-occurrence Matrix (GLCM, 23), *iv*) Gray level Difference Matrix (GLDM, 14), *v*) Gray-level Run Length Matrix (GLRLM, 16), *vi*) Gray-level Size Zone matrix (GLSZM, 16) and *vii*) Neighborhood Gray-Tone Difference Matrix (NGDM, 5). Afterwards we built the so called feature \mathbf{X} ($N \times 105$) and responses \mathbf{Y} ($N \times 3$) matrices (N being the number of lesions in the training set).

2.3. Data Analysis: Partial-Least-Squares

With the aim of characterizing brain lesions using radiomics and generating a predictive TBI model, we used the multivariate PLS technique. PLS discriminant analysis (PLS-DA) is a flexible tool used for descriptive and predictive modelling, as well as for discriminant feature selection. The algorithm incorporates dimensionality reduction with discriminant analysis for high-dimensional data interpretation. In a nutshell, the algorithm involves two steps: i) PLS latent variables (LVs) construction and ii) predictive model building [7]. The LVs are computed as linear combinations of the independent \mathbf{X} variables, \mathbf{XW} , where the loading weights in \mathbf{W} are chosen in such a way that the corresponding LVs have maximal covariance with the responses in the \mathbf{Y} matrix. For multiclass problems, \mathbf{Y} is a dummy matrix encoding the class membership information. After building the model, the responses of unknown class data can be predicted from their independent variables [8].

2.3.1. Feature Selection

For finding informative and distinctive markers allowing classes discrimination, the Variables Importance on Prediction (VIP) criterion was used. VIP scores help in detecting and ranking those features contributing in the model fitting. For each variable, the VIP score is equal to the accumulated weights (\mathbf{w}) over all selected LV's. For feature selection it has been widely suggested to retain those features with VIP's greater than the unity [8].

2.3.2. Model Fitting

In this work, firstly a model using all the considered features was fitted. Secondly, by using the above explained VIP crite-

ron, the model was retrained with the retained features only. Model fitting was performed over the training set in a 10-fold cross-validation strategy by changing the number of LV's and by assessing the classification error rate. The model that minimized the classification error was preferred. Afterwards, PLS models were validated by predicting the unseen test samples.

2.4. Experiments

With the aim of assessing the feature-class effect for predicting the lesions, PLS models were assessed under different feature combinations, as indicated in Table 1. Selected experiments were chosen on the basis of comparing radiologist's *observable* descriptors (shape and FOS) and assessing their importance on lesion labelling. On the other hand, we explore as well the models behaviour under the inclusion of more complex computational descriptors (texture ones).

Table 1. Experiments.

| Feat\Exp | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 |
|----------|----|----|----|----|----|----|----|----|
| Shape | ✓ | ✓ | ✓ | | ✓ | | ✓ | |
| FOS | ✓ | ✓ | | ✓ | ✓ | | | ✓ |
| Texture | ✓ | ✓ | | | | ✓ | ✓ | ✓ |

Note: Exp: Experiment; Feat: Features used. Colored fields indicate that feature selection was applied.

2.5. Performance and statistics

The most relevant per-class descriptors in terms of VIP scores were assessed by means of the non-parametric Kruskal-Wallis test. When p-values exhibited statistical significance, the Dunn's test for multiple comparisons was performed. The Benjamini & Hochberg procedure for controlling the false discovery rate was applied. Two-tailed tests with a 0.05 significance level were used. For addressing mass lesion classifiers using PLS-DA, accuracy, sensitivity and specificity were used as performance metrics.

3. RESULTS

A summary of the predictive models fitness obtained for each experiment is shown in Table 2. As expected, the model where feature selection was conducted considering all descriptors (Exp #2) obtained the lowest error-rate. Besides, it outperformed the model where no feature selection was applied (Exp #1), suggesting that non-explanatory features have been discarded.

The worst performing model in terms of errors was the one fitted in Exp #4 (only FOS information considered), suggesting that these features are not sufficient for distinctively characterizing the lesion classes. However, when *FOS* and *shape* features were jointly considered, the error-rate considerably decreased. It is worth mentioning that these are

Table 2. Fitted models summary and number (%) of selected features.

| Exp | Model | | Features | | | | | | | | |
|-----|-------|----|------------|-----------|-----------|-----------|----------|----------|----------|----------|---------|
| | ER | LV | Considered | Selected | | | | | | | |
| | | | Total | Total | Shape | FOS | GLCM | GLDM | GLRLM | GLSZM | NGDM |
| #1 | 0.12 | 5 | 105 | 105 (100) | 13 (100) | 18 (100) | 23 (100) | 14 (100) | 16 (100) | 16 (100) | 5 (100) |
| #2 | 0.10 | 7 | 105 | 40 (38.1) | 12 (92.3) | 8 (44.4) | 8 (34.7) | 1 (7) | 7 (43.7) | 4 (25) | 0 (0) |
| #3 | 0.25 | 4 | 13 | 4 (30.7) | 4 (30.7) | | | | | | |
| #4 | 0.3 | 8 | 18 | 11 (61.1) | | 11 (61.1) | | | | | |
| #5 | 0.15 | 7 | 31 | 11 (35.5) | 6 (46.1) | 5 (27.7) | | | | | |
| #6 | 0.13 | 15 | 74 | 17 (22.9) | | | 6 (26.1) | 1 (7) | 6 (37.5) | 4 (25) | 0 (0) |
| #7 | 0.08 | 13 | 87 | 30 (34.4) | 11(84.6) | | 7 (30.4) | 1 (7) | 5 (31.2) | 6 (37.5) | 0 (0) |
| #8 | 0.09 | 18 | 92 | 30 (32.6) | | 10 (55.5) | 3 (13) | 4 (28.9) | 6 (37.5) | 1 (20) | |

Note: Exp: Experiment; ER: Error Rate.

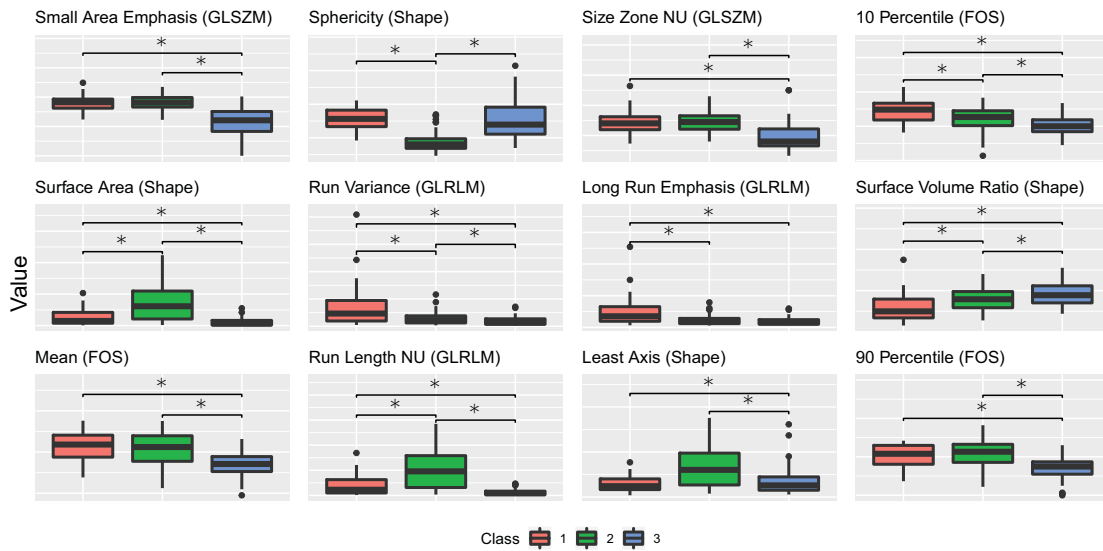


Fig. 1. Common selected features boxplots. Classes 1, 2 and 3 belong to *epidural hematoma*, *acute subdural hematoma* and *contusion*. NU: Non-uniformity. *: significant p-value.

observable features for radiologists, having a valuable meaning for lesion labelling.

When texture features were only used for fitting the model (Exp #6), a low error-rate was obtained. Among all feature classes, *GLSZM* and *GLCM* features were preferred. The *NGDM* features have never been selected, suggesting non informative descriptors for the considered task. The inclusion of *shape* (Exp #7) and *FOS* (Exp #8) to the *texture* model decreased the error-rate.

Boxplots of the shared retained features among the different models are shown in Fig. 1. It is possible to appreciate that all these features exhibited statistical significance among some or all groups. Considering radiological interpretable descriptors, *Sphericity* showed to be a distinctive one. The results are supported by lesion morphology configurations, namely that subdural ones are less spherical than epidural and

contusion ones [1]. On the other hand, *Surface Areas* were greater for subdural hematoma when compared against the other lesions.

Table 3 shows achieved classification performance on the unseen data. The best performance was obtained for model #2, which improved by ~ 7% on accuracy the model without feature selection. While models using independently *shape* or *FOS* information provided low performances, the model considering both features was able to considerably improve the classification by performing ~ 10% better on accuracy terms. For Exp #6, *texture* features were able to almost reproduce the best performing model by achieving 88.2% accuracy. Besides, adding *FOS* or *Shape* descriptors to the *texture* model (Exp #7 and Exp #8) has not improved the task performance. These results suggest that *texture* features encode relevant information for differentiating among

TBI lesions. The importance of these results relies on the fact that texture analysis can be performed patch-wise, suggesting potential for lesion detection, segmentation and classification.

Table 3. Classification performances.

| Exp | Acc | Se | | | Sp | | |
|-----|-------------|-----------|-------------|-------------|------------|-------------|-------------|
| | | C #1 | C #2 | C #3 | C #1 | C #2 | C #3 |
| #1 | 82.3 | 85 | 89.5 | 75.9 | 89.6 | 89.8 | 94.9 |
| #2 | 89.7 | 80 | 94.7 | 93.1 | 100 | 89.8 | 94.9 |
| #3 | 72.1 | 60 | 94.7 | 65.5 | 95.8 | 73.5 | 89.7 |
| #4 | 72.1 | 65 | 42.1 | 96.5 | 85.4 | 89.8 | 82.1 |
| #5 | 83.8 | 75 | 89.5 | 86.2 | 97.9 | 85.7 | 92.3 |
| #6 | 88.2 | 85 | 94.7 | 86.2 | 95.8 | 93.9 | 92.3 |
| #7 | 88.2 | 80 | 94.7 | 89.6 | 100 | 89.8 | 92.3 |
| #8 | 77.9 | 70 | 89.5 | 75.9 | 85.4 | 91.8 | 89.7 |

Note: Exp: Experiment; Acc: Accuracy (%); Se: Sensitivity (%); Sp: Specificity (%); C#: Classes.

4. DISCUSSION AND CONCLUSIONS

In this work several PLS models using radiomic CT features were compared. Three different brain lesion types were considered and characterized by means of different descriptors. We isolated several radiomarkers that may play a key role in discrimination of TBI lesions, all behaving differentially among classes (p-val significant). We devised, as well, an automatic method for classifying the lesions. When using selected descriptors chosen over all feature types, a 89.9% classification accuracy over the test set was achieved. Our results suggest that radiomic features are very close to reproduce radiologists lesion labelling. When only using *shape* descriptors, it was found that *Sphericity*, *Surface* and the *Surface to Volume Ratio* play an important role. However, neither *shape* nor *FOS* features were enough to accurately discriminate the lesion types. When considering both feature classes together, even by achieving a consistent improvement (~ 10% greater accuracy) these features were not able to reproduce the obtained results with the whole database. An important result of this work regards the potential and capability of texture descriptors for characterizing the lesions. We were able to build a model using only texture information which predicts the test set with 88.2% accuracy, behaving almost as the model that included all features. This result is relevant since it suggests texture CT information might help in lesion detection, multi-class segmentation and classification. Besides, it suggests the existence of non-observable data structures helping in discriminating TBI damage. Potentially, the developed models could be used for performing automatic quality control of segmented and predicted lesions. It is important to keep in mind that radiological inspired markers might be preferred for clinical applications when conducting these tasks. In future work other radiomics features and classification methods will be explored.

5. ACKNOWLEDGEMENTS

This work was funded by the H2020 Marie-Curie ITN TRABIT (765148) project. Data used for this work came from CENTER-TBI (FP7-COOPERATION-2013-602150). We thank Charlotte Timmermans, Nathan Vanalken and Thijs Vande Vyvere for conducting the manual segmentations.

6. REFERENCES

- [1] P.M. Parizel, J.W. Van Goethem, Ö. Özsarlak, M. Maes, and C.D. Phillips, “New developments in the neuroradiological diagnosis of craniocerebral trauma,” *European radiology*, vol. 15, no. 3, pp. 569–581, 2005.
- [2] R.J. Gillies, P.E. Kinahan, and H. Hricak, “Radiomics: images are more than pictures, they are data,” *Radiology*, vol. 278, no. 2, pp. 563–577, 2015.
- [3] Esther L Yuh, Alisa D Gean, Geoffrey T Manley, Andrew L Callen, and Max Wintermark, “Computer-aided assessment of head computed tomography (ct) studies in patients with suspected traumatic brain injury,” *Journal of Neurotrauma*, vol. 25, no. 10, pp. 1163–1172, 2008.
- [4] Sasank Chilamkurthy, Rohit Ghosh, Swetha Tanamala, Mustafa Biviji, Norbert G Campeau, Vasantha Kumar Venugopal, Vidur Mahajan, Pooja Rao, and Prashant Warier, “Deep learning algorithms for detection of critical findings in head ct scans: a retrospective study,” *The Lancet*, vol. 392, no. 10162, pp. 2388–2396, 2018.
- [5] M. Shafiq-ul Hassan, G.G. Zhang, K. Latifi, G. Ullah, D.C. Hunt, Y. Balagurunathan, M.A. Abdalah, M.B. Schabath, D.G. Goldgof, D. Mackin, et al., “Intrinsic dependencies of ct radiomic features on voxel size and number of gray levels,” *Medical physics*, vol. 44, no. 3, pp. 1050–1062, 2017.
- [6] J.J.M. van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R.G.H. Beets-Tan, J.C. Fillion-Robin, S. Pieper, and H.J.W.L. Aerts, “Computational radiomics system to decode the radiographic phenotype,” *Cancer research*, vol. 77, no. 21, pp. e104–e107, 2017.
- [7] L.C. Lee, C.Y. Liong, and A.A. Jemain, “Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps,” *Analyst*, 2018.
- [8] G. Aliakbarzadeh, H. Parastar, and H. Sereshti, “Classification of gas chromatographic fingerprints of saffron using partial least squares discriminant analysis together with different variable selection methods,” *Chemometrics and Intelligent Laboratory Systems*, vol. 158, pp. 165–173, 2016.



Publisher's approval and licenses

Publisher's approval and licences

Doctoral candidate: Ezequiel de la Rosa

1. de la Rosa, Ezequiel, et al. "Differentiable deconvolution for improved stroke perfusion analysis." *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VII* 23. Springer International Publishing, 2020.

<https://www.springernature.com/gp/partners/rights-permissions-third-party-distribution>

Permission FAQs

- > [Translation, audio, or book reprint rights](#)
- > [Commercial reprints](#)
- > [High-res image requests](#)
- > [TV licences](#)
- > [Third-Party distribution \(TPD\) requests](#)
- > [SpringerMaterials](#)
- > [Institutional Requests](#)
- > [Accessibility requests](#)
- > [Patients and Caregivers](#)
- > [Content published under a CC-BY licence](#)
- > [STM Permission Guidelines](#)

Springer Nature Author FAQs

- ∨ [Reuse in an Author's Dissertation or Thesis](#)

Springer Nature Book and Journal Authors have the right to reuse the Version of Record, in whole or in part, in their own thesis. Additionally, they may reproduce and make available their thesis, including Springer Nature content, as required by their awarding academic institution. Authors must properly cite the published work in their thesis according to current citation standards and include the following acknowledgement: *'Reproduced with permission from Springer Nature.'*

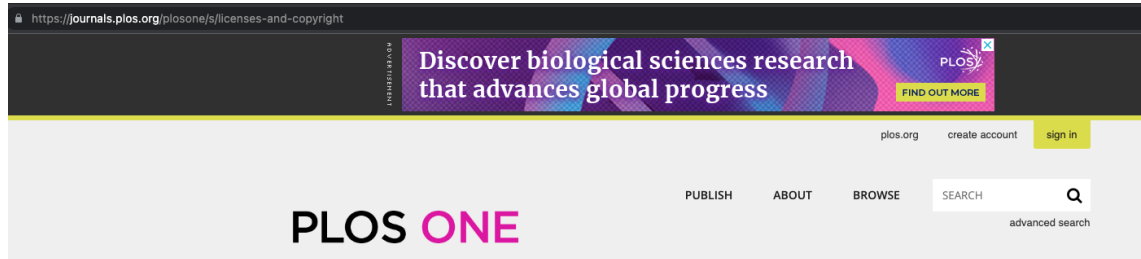
2. de la Rosa, Ezequiel, et al. "AIFNet: Automatic vascular function estimation for perfusion analysis using deep learning." *Medical Image Analysis* 74 (2021): 102211.

Author rights

The below table explains the rights that authors have when they publish with Elsevier, for authors who choose to publish either open access or subscription. These apply to the corresponding author and all co-authors.

| Author rights in Elsevier's proprietary journals | Published open access | Published subscription |
|--|-----------------------|--|
| Retain patent and trademark rights | √ | √ |
| Retain the rights to use their research data freely without any restriction | √ | √ |
| Receive proper attribution and credit for their published work | √ | √ |
| Re-use their own material in new works without permission or payment (with full acknowledgement of the original article): 1. Extend an article to book length 2. Include an article in a subsequent compilation of their own work 3. Re-use portions, excerpts, and their own figures or tables in other works. | √ | √ |
| Use and share their works for scholarly purposes (with full acknowledgement of the original article): 1. In their own classroom teaching. Electronic and physical distribution of copies is permitted 2. If an author is speaking at a conference, they can present the article and distribute copies to the attendees 3. Distribute the article, including by email, to their students and to research colleagues who they know for their personal use 4. Share and publicize the article via Share Links, which offers 50 days' free access for anyone, without signup or registration 5. Include in a thesis or dissertation (provided this is not published commercially) 6. Share copies of their article privately as part of an invitation-only work group on commercial sites with which the publisher has a hosting agreement | √ | √ |
| Publicly share the preprint on any website or repository at any time. | √ | √ |
| Publicly share the accepted manuscript on non-commercial sites | √ | √ using a CC BY-NC-ND license and usually only after an embargo period (see Sharing Policy for more information) |

3. de la Rosa, Ezequiel, et al. "Detecting CTP truncation artifacts in acute stroke imaging from the arterial input and the vascular output functions." *PloS one* 18.3 (2023): e0283610.



Reuse of PLOS Article
Content
Content Owned by Someone
Else
Using Article Content
Previously Published in
Another Journal
Maps
Removal of Content Used
Without Clear Rights
Trademarks and Symbols
Giving Proper Attribution for
Use of Content

Licenses and Copyright

The following policy applies to all PLOS journals, unless otherwise noted.

Reuse of PLOS Article Content

PLOS applies the [Creative Commons Attribution 4.0 International \(CC BY\) license](#), or other comparable licenses that allow free and unrestricted use, to articles and other works we publish. If you submit your paper for publication by PLOS, you agree to have the CC BY license applied to your work. If your institution or funder requires your work or materials to be published under a different license or dedicated to the public domain - for example [Creative Commons 1.0 Universal \(CC0\)](#) or [Open Governmental License](#) - this is permitted for those licenses where the terms are equivalent to or more permissive than CC BY. PLOS requires that you as the author agree that anyone can reuse your article content in whole or part for any purpose, for free, even for commercial purposes. These permitted uses include but are not limited to self-archiving by authors of submitted, accepted and published versions of their papers in institutional repositories. Anyone may copy, redistribute, reuse, or modify the content as long as the author and original source are properly cited. This facilitates freedom in reuse and also ensures that PLOS content can be mined without barriers for the needs of research.