QUANTITATIVE NEUROIMAGING WITH HANDCRAFTED AND DEEP RADIOMICS IN NEUROLOGICAL DISEASES

Elizaveta Lavrova

Doctoral thesis

QUANTITATIVE NEUROIMAGING WITH HANDCRAFTED AND DEEP RADIOMICS IN NEUROLOGICAL DISEASES

Elizaveta Lavrova

2024

QUANTITATIVE NEUROIMAGING WITH HANDCRAFTED AND DEEP RADIOMICS IN NEUROLOGICAL DISEASES

Dissertation

To obtain the degree of Doctor at Maastricht University and
Université de Liège,
on the authority of the Rector Magnifici,
Prof. Dr. Anne-Sophie Nyssen and Prof. Dr. Pamela Habibović,
in accordance with the decision of the Board of Deans,
to be defended in public
on 31st of May 2024, at 10:00 hours

by

Elizaveta A. Lavrova
born on September 6, 1992
in Moscow, Russia

Promotors

Prof. Dr. Philippe Lambin, Maastricht University
Prof. Dr. Ir. Christophe Phillips, Liege University

Co-promotors

Prof. Dr. Eric Salmon, Liege University
Dr. Henry C. Woodruff, Maastricht University

Assessment Committee

Chair:
Prof. Dr. Ir. Walter H. Backes, Maastricht University
Members:
Prof. Dr. Ir. Gilles Louppe, Liege University
Prof. Dr. Fabienne Collette, Liege University
Prof. Dr. Adrien Depeursinge, University of Applied Sciences Western Switzerland
Dr. Johan van Soest, Maastricht University
Dr. Harini Veeraraghavan, Memorial Sloan-Kettering Cancer Center, New York, USA

This thesis was accomplished with financial support from the Liege-Maastricht Imaging Valley Grant.

Cover Dall-E 2, query: ”brain in Rembrandt style”, 2023
Production Publisher
ISBN xxx-xxxx-xxx-xx-x

To my parents

Contents

1
Introduction

1.1 Motivation

Neurological diseases have emerged as a significant global health challenge, representing both a leading cause of disability and a major contributor to mortality worldwide [1]. Among the most impactful neurological conditions are Alzheimer's and Parkinson's diseases, multiple sclerosis, and stroke [2, 3]. These conditions appear from diverse causes, including genetic disorders, congenital abnormalities, infections, lifestyle or environmental factors, and various injuries [3]. Most of them exhibit a prolonged development phase [4, 2, 5, 6]. Early diagnosis of neurological diseases remains a challenge as symptomatic changes become apparent only at later stages. At the moment, diagnosis primarily relies on the symptoms, clinical history, and the occurrence of recurrent clinical episodes [3, 7, 8].
Most neurological diseases currently lack a definitive cure [9]. However, timely and appropriate personalized treatment can effectively slow down or even halt the disease progression, leading to a substantial improvement in the quality of life for patients and their families [6]. The availability of disease-modifying and symptomatic medications, while promising, poses a significant financial burden on national healthcare systems [10] and creates risks for patients while not selected correctly. Therefore, gaining insights into diagnosis and treatment response beforehand becomes paramount in optimizing resource allocation and ensuring efficient healthcare management for patients. This leads us to a pressing unmet clinical need for reliable and objective biomarkers discovery to enable early detection and personalized treatment of neurological diseases [11].
In modern clinical practice, medical imaging has become an essential tool to investigate anatomical, morphological, and functional abnormalities. In neurology, several imaging modalities are commonly employed [12]: magnetic resonance imaging (MRI) is known for its ability to provide optimal soft tissue contrast without ionizing radiation and facilitate both anatomical and functional studies [13, 14, 15], positron emission tomography (PET) imaging is an instrument to study metabolism [16, 17], while X-ray

computed tomography (CT) scans offer structural visualization with rapid observation capabilities [18]. Additionally, CT angiography (CTA) enables visualization of the blood vessels after injecting an intravascular tracer and therefore CTA is mostly used for stroke evaluation [19]. A more advanced technique, CT perfusion (CTP) allows for blood flow quantification [20]. Blood vessel visualization for stroke patient management is possible with MR angiography and perfusion as well [21]. Although integrated into clinical protocols, the existing imaging criteria for diagnostics remain subjective. Advancements in medical imaging technology have led to improved image quality, accompanied by massive data aggregation [22]. As a consequence, neuroimaging has emerged as a promising field for data-driven approaches.

According to ISO/IEC TR 24028:2020, artificial intelligence (AI) is the ”capability of an engineered system to acquire, process and apply knowledge and skills” [23]. Machine learning is a subfield of AI that develops and studies the statistical approaches to solve tasks by processing training data instead of using formal instructions. Deep learning is a subfield of machine learning that utilizes features created by the artificial neural network during the model training process. With the continuous advancement of computing capabilities, particularly with the “graphical processing units” (GPUs) and the growth of global data, deep learning has demonstrated high performance in various tasks. Modern neural networks exhibit the ability to accomplish image [24] and speech [25] recognition, generate lifelike data [26], comprehend natural language [27], and learning autonomy [28]. With computational power reaching billions of parameters, these models possess the capacity to autonomously learn complex processes from the available data. As a result, these models find applications not only in decision support but also in intermediate steps, such as data synthesis or labeling, further enhancing their potential impact across various domains.

In general, medical data-driven AI uses data from prior cases to address clinical inquiries for new patients. Transforming existing clinical data into actionable information would empower healthcare

professionals to make better-informed decisions for future patient care.
Radiomics is a methodology that combines imaging data with machine learning to facilitate quantitative imaging [29]. The main hypothesis of radiomics is that medical images are not just visual representations of the human body but valuable data [30]. This approach involves extracting numerous quantitative features from medical images and linking them to relevant clinical outcomes. There are two main approaches within radiomics: handcrafted and deep radiomics [31, 32]. Handcrafted radiomics relies on the computation of predefined mathematical features, while deep radiomics utilizes deep learning techniques to automatically extract and select model-defined features. Handcrafted radiomics can be seen as a manually engineered precursor to the more automated deep learning pipeline. While handcrafted radiomics boasts interpretability and requires less data, it is limited by pipeline design and may fail to uncover sophisticated patterns in the data. The success of deep radiomics relies on vast amounts of training data, and its black-box nature poses interpretability challenges for clinicians [33]. Nevertheless, methods such as saliency mapping or GradCam [34] offer opportunities to interpret the outcomes of deep neural networks, thus promoting transparency and enhancing their clinical application. Radiomics has proven to be highly successful in oncology, yielding promising results in improving the diagnosis, prognosis, and treatment planning for cancer patients [35, 36, 37]. It is driven by a large amount of labeled data available due to the need for treatment protocols. Over the past years, researchers have investigated the main challenges and limitations associated with this approach [38]. As the field of radiomics continues to progress, there is a rising interest in exploring its potential beyond oncology, particularly in the domain of neurology. In this context, radiomics holds the promise of playing a pivotal role in early disease detection and optimizing patient care for neurological conditions. To achieve this goal, a comprehensive analysis of good practices and well-established pipeline steps in the oncological radiomics approach must be conducted. By examining

the successful methodologies and techniques utilized in oncological studies, valuable insights can be gleaned to guide the implementation of radiomics in the neurological field. However, given the unique characteristics and demands of neurological diseases, it is essential to identify and address the additional requirements specific to this domain. By integrating the existing neuroimaging data, the potential benefits include improved model training and validation, enhanced generalization of results, and accelerated progress in neurological radiomics research. Ultimately, harnessing the vast repository of neuroimaging data is expected to yield valuable insights and enhance the clinical utility of radiomics in the neurological context. As we delve into this new frontier, several critical questions arise concerning the applicability of radiomics in neurology. Identifying potential pitfalls and proposing effective solutions becomes imperative to ensure the successful integration of this methodology in the neurological domain and gaining its full potential in improving patient outcomes.

1.2 Background: Imaging of the most common neurological disorders

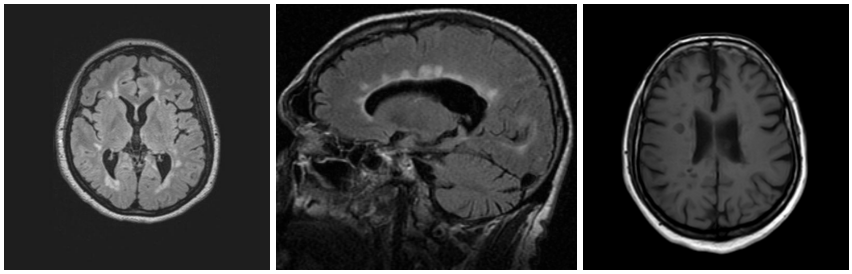
1.2.1 Multiple sclerosis

Multiple sclerosis is a demyelinating disorder that primarily affects the central nervous system [39]. It exhibits a characteristic pattern of dissemination, both in space (with multiple plaques occurring in different regions of the central nervous system) and in time (with plaques appearing at different time points). Clinical symptoms can vary significantly depending on the affected brain regions, leading to the application of various tests to assess motor and cognitive functions.
The diagnosis of multiple sclerosis relies on the McDonald diagnostic criteria [40], which encompass a range of assessments, including clinical history, cerebrospinal fluid analysis, immunoglobulin G

levels in serum, evoked potentials, and MRI. Radiographic features primarily revolve around the distinctive characteristics of the plaques, which typically manifest as ovoid shapes distributed around veins.
CT features tend to be non-specific and may include uniformly hypoattenuating lesions, brain atrophy (particularly in advanced stages), and contrast enhancement during active phases [41, 42].
In contrast, MRI plays a crucial role in diagnosis and follow-up. The MRI protocol typically includes T1-weighted (T1w), T2-weighted (T2w), fluid-attenuated inversion recovery (FLAIR), and diffusion-weighted imaging (DWI) sequences.
Specific MRI findings include hypointense lesions (also called ”black hole lesions”) on T1w images, thinned corpus callosum (brain region beneath the cerebral cortex), and hyperintense lesions in advanced disease cases. On T2w images, lesions appear hyperintense. FLAIR sequences are the most sensitive among the other typical sequences, and they reveal hyperintense lesions. Active lesions, on T1w images with gadolinium contrast enhancement, show increased intensity. However, routine asymptomatic follow-up does not necessarily require contrast enhancement. DWI can display active plaques with either high or low apparent diffusion coefficient values, signifying altered diffusion properties [42, 43].
A distinctive MS-specific feature is the presence of ”Dawson’s fingers” — demyelinating plaques distributed perpendicular to the lateral ventricles of the brain [43].
The typical radiographic features of multiple sclerosis are presented on Figure 1.1.

1.2.2 Stroke

A stroke is a clinical condition characterized by the abrupt onset of a focal neurological deficit, presumed to be of vascular origin [44]. Strokes are typically categorized into two main types: ischemic stroke and hemorrhagic stroke. The most common ischemic stroke occurs when there is a sudden episode of neurological dysfunction due to

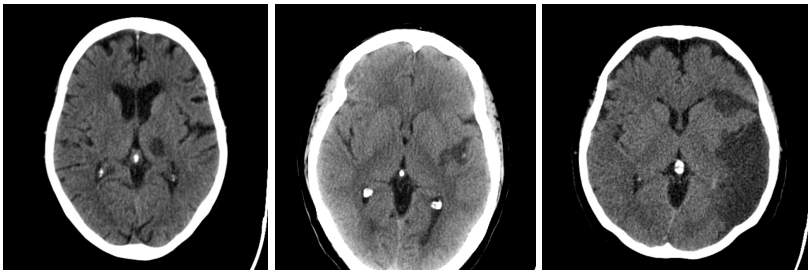


(a) (b) (c)

Figure 1.1: Some of the typical radiographic features of multiple sclerosis observed in three different patients on MR scans: (a) multi-focal areas of hyperintensity on FLAIR (image source: case courtesy of Henry Knipe, radiopaedia.org, rID: 30252), (b) hyperintense Dawson's fingers on FLAIR (image source: case courtesy of Frank Gaillard, radiopaedia.org, rID: 35916), (c) T1w "black hole" lesions (image source: case courtesy of Ahmed Abdrabou, radiopaedia.org, rID: 35195)

localized tissue infarction in the central nervous system, attributed to arterial thrombosis, embolization, or severe hypoperfusion. In common usage, ischemic stroke mainly refers to cerebral infarctions.
Ischemic strokes usually present with a rapid onset of neurological deficits, the nature of which depends on the specific area of the brain affected. These symptoms may evolve over hours and can either worsen or improve [45].
For diagnosing acute strokes, several imaging techniques are commonly employed, mostly including non-contrast CT scans, CT perfusion scans, and CT angiography. The earliest visible sign on CT imaging is a hyperdense segment within a blood vessel, representing the direct visualization of intravascular thrombus or embolus and is immediately apparent, as well as a loss of differentiation between gray and white matter [46]. In acute cases, hypoattenuation and swelling become more pronounced, and subacute cases exhibit subsiding swelling and the emergence of small cortical petechial hemorrhages, resulting in increased cortex attenuation [47]. In chronic

cases, a region of low density with a negative mass effect is present [47].
CT perfusion scans are used to identify the core of the infarct (the irreversibly damaged area) and the penumbra (the affected but potentially recoverable area) [48]. CT angiography helps identify thrombi, assess the condition of the carotid and vertebral arteries, and evaluate collateral vessels [49].
MRI is frequently employed for follow-up imaging to monitor the recovery process and identify potential necrotic tissue [50].
Stroke imaging provides crucial additional information about the stroke patient's condition, aiding in the selection of appropriate treatment strategies and assessing the extent of damage [45].
The typical radiographic features of stroke are presented on Figure 1.2.



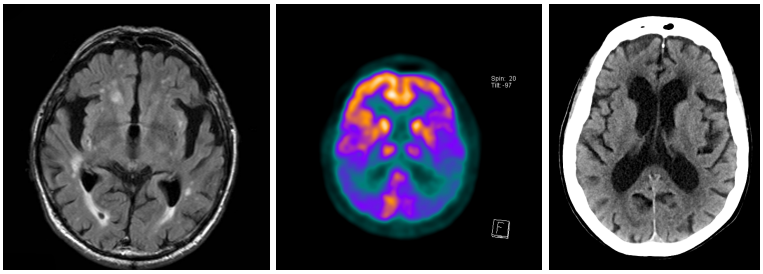
(a) (b) (c)

Figure 1.2: Some of the typical radiographic features of stroke observed in three different patients on CT scans: (a) left thalamic hypodense focus developed during 24 hours after stroke onset (image and description source: case courtesy of David Cuete, radiopaedia.org, rID: 36507), (b) a small irregular hypodense area seen at the left temporal region with loss of grey-white matter differentiation (image and description source: case courtesy of Rania Adel Anan, radiopaedia.org, rID: 99808), (c) extensive infarction in the territory of the left middle cerebral artery (image and description source: case courtesy of David Cuete, radiopaedia.org, rID: 35732)

1.2.3 Alzheimer's disease

Alzheimer's disease is a neurodegenerative condition characterized by the accumulation and deposition of cerebral amyloid-β [51]. It is primarily associated with memory deficits and is diagnosed based on specific criteria, such as the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association Alzheimer's criteria [52], which involve both clinical and histological signs.
These criteria include the use of neuroimaging and laboratory examinations, including blood and cerebrospinal fluid analyses, to rule out other potential causes. However, it's essential to note that the only definitive diagnostic test for Alzheimer's disease is a brain biopsy, which is rarely performed in practice due to its invasiveness [53]. Additionally, certain fluid biomarkers, such as beta-amyloid, total tau, and hyperphosphorylated tau, are employed to provide further diagnostic support [54].
The most commonly used imaging modalities for Alzheimer's disease are MRI and PET scans. Braak's staging system is employed to distinguish between different progression stages: I-II involves early cortical involvement, III-IV involves the limbic system and hippocampus, and V-VI involves the cortex with temporal lobes [55]. During MRI examinations, gray matter atrophy is measured both overall and in various cortical regions.
To assess patients with mild cognitive impairment and predict whether it will progress to Alzheimer's disease, nuclear imaging techniques like F-18 fluorodeoxyglucose (FDG), amyloid, and tau PET scans are used. These techniques reveal hypometabolism, increased amyloid deposition, and heightened tau protein activity [56]. However, it's important to note that these features are not specific to Alzheimer's disease and can be observed in other conditions as well.

The typical radiographic features of Alzheimer's disease are presented on Figure 1.3.



(a) (b) (c)

Figure 1.3: Some of the typical radiographic features of Alzheimer's disease observed in three different patients on multimodal data: (a) cerebral volume loss on FLAIR, T2w hyperintense regions suggest a concurrent component of chronic small vessel ischaemic change (image and description source: case courtesy of Frank Gaillard, radiopaedia.org, rID: 10738), (b) decreased metabolism activity on the bilateral parietotemporal cortex (image and description source: case courtesy of Bruno Di Muzio, radiopaedia.org, rID: 22715), (c) volume loss and large ventricles on CT (image and description source: case courtesy of Frank Gaillard, radiopaedia.org, rID: 33753)

1.3 Objectives, aims, and outline of the thesis

The primary aim of this thesis was to explore the feasibility of utilizing the radiomics approach in clinical neuroimaging while identifying the key factors and challenges within the field (neuroimaging and radiomics workflows connected by this thesis are illustrated on Figure 1.4). An additional objective was to test the hypothesis that radiomic features extracted from the brain regions provide information about brain tissue pathology not detectable by the human eye. To achieve these goals, the specific objectives of the study were as follows: (i) to conduct a comprehensive review of the current radiomic workflow as applied as a transition from oncological to clinical neuroimaging tasks and uncover the main challenges (Chapters 2 and 3), (ii) to assess the predictive capability of radiomic features in neuroimaging tasks (Chapter 4), (iii) to propose potential solutions for addressing some of the challenges encountered in applying the radiomics ap-

proach to neuroimaging (Chapters 4, 5, and 6).

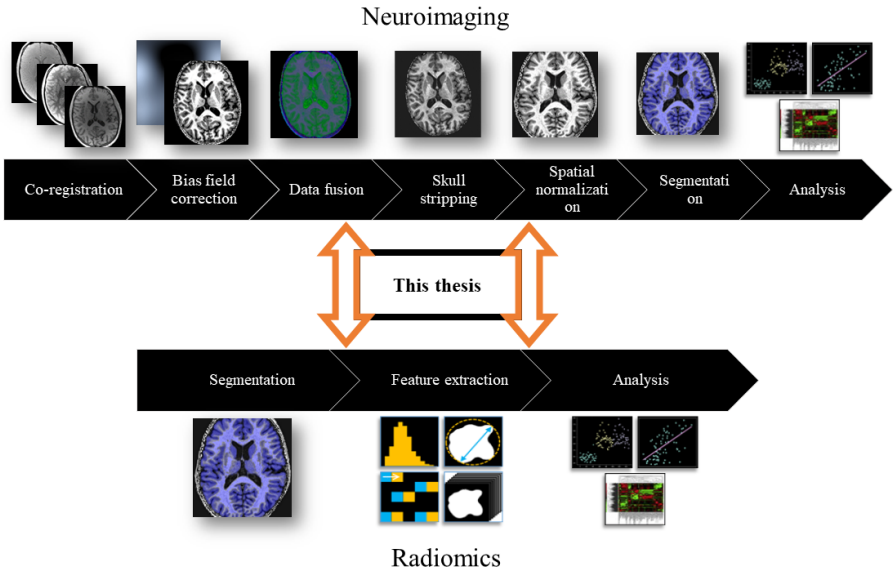


Figure 1.4: The goal of this thesis is to explore the feasibility of utilizing the radiomics approach in clinical neuroimaging inheriting the best practices from both radiomics and neuroimaging pipelines (as illustrated by connecting neuroimaging and radiomics workflows by orange arrows).

This thesis is composed of 7 chapters.
Chapter 1: General introduction and outline of the thesis.
In this introductory section, I outline the motivation behind this thesis and present its main objectives. The chapter provides an overview of the entire thesis, highlighting the structure and content of the subsequent chapters. Chapter 2: Radiomics Capability in Neuro-Oncological Tasks [57].
This chapter explores the established radiomics methodology in the context of neuro-oncological tasks. Specifically, in this published peer-reviewed article we demonstrate how brain MRI radiomics can

estimate the chromosome 1p/19q co-deletion status, which serves as a diagnostic biomarker for low-grade glioma. This work serves as a strong foundation and motivation for extending the application of radiomics to non-oncological brain studies.

Chapter 3: Radiomics Application in Non-Oncological Neurology: Overview and Challenges (under the peer review).

Here, we present a comprehensive review of the application of radiomics in non-oncological neurology. Focusing on neurology-specific steps of the radiomics pipeline, we explore potential clinical applications and discuss the existing challenges in implementing radiomics in non-oncological brain studies.

Chapter 4: Exploratory Study: Radiomic Biomarkers in Multiple Sclerosis [58].

This chapter presents an exploratory study that identifies potential radiomic biomarkers of multiple sclerosis. In this published peer-reviewed article we put our thesis hypothesis to the test, demonstrating the potential of qMRI radiomics in the detection of pathological information in normal-appearing tissues. To address challenges related to MRI expressed in arbitrary units and enable multi-center multiple-sclerosis studies, we propose the use of unique qMRI sequences.

Chapter 5: Deep Learning for Carotid Artery Segmentation in Stroke At-Risk Patients [59].

In this chapter, we address a critical challenge in radiomics by proposing a deep learning method for automated carotid artery segmentation in patients at risk of stroke. While automated plaque characterization research exists, carotid artery localization is typically performed manually or in a semi-automated manner. In this published peer-reviewed article we introduce a U-Net implementation with a consensus patching strategy, trained on partially labeled data, that effectively detects and segments carotid arteries on neck MRI with performance comparable to state-of-the-art models. This chapter presents an innovative auto-segmentation method as a solution to the manual data labeling challenge.

Chapter 6: An Open-Source Software Package for Medical Imaging

Data Curation and Exploration [60].
in this published peer-reviewed article we propose an open-source software tool that links the various steps of the radiomics pipeline. By offering a solution to pipeline and implementation standardization challenges, this software package facilitates the integration and standardization of radiomics methods. Moreover, we advocate for the use of an open-source tool for data preparation and exploratory analysis to democratize the implementation of radiomics.
Chapter 7: General Discussion and Perspective.
This final chapter offers a comprehensive discussion of the work presented in this thesis. We delve into the implications of our findings and recommendations for future applications of radiomics in neurological decision-making. Additionally, we explore the potential for enabling radiomics as a valuable tool in neurological clinical decision aid.
The structure of this thesis is presented in Table 1.1.

Table 1.1: Thesis structure

Part 1 – Introduction	
Chapter 1	General introduction and outline of the thesis
Part 2 – Radiomic methodology transfer from oncology to neurology	
Chapter 2	Radiomics capability in neuro-oncological tasks
Chapter 3	Radiomics application in non-oncological neurology: overview and challenges
Part 3 – Predictive power of radiomics in neuroimaging	
Chapter 4	Exploratory study: radiomic biomarkers in multiple sclerosis
Part 4 – Addressing challenges in radiomics for neurology	
Chapter 5	Deep learning for carotid artery segmentation in stroke-at-risk patients
Chapter 6	An open-source software package for medical imaging data curation and exploration
Part 5 – Discussion	
Chapter 7	General discussion and prospectives
Appendices	

References

[1] Valery L Feigin et al. "The global burden of neurological disorders: translating evidence into policy". In: *The Lancet Neurology* 19.3 (2020), pp. 255–265. DOI: 10.1016/S1474-4422(19)30411-9.

[2] Bruno Dubois et al. "Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria". en. In: *Alzheimers. Dement.* 12.3 (Mar. 2016), pp. 292–323. DOI: 10.1016/j.jalz.2016.02.002.

[3] Chenyu Ding et al. "Global, regional, and national burden and attributable risk factors of neurological disorders: The Global Burden of Disease study 1990–2019". en.

In: *Front. Public Health* 10 (Nov. 2022), p. 952161. DOI: 10.3389/fpubh.2022.952161.

[4] Marianna Cortese et al. "Preclinical disease activity in multiple sclerosis: A prospective study of cognitive performance prior to first symptom". en. In: *Ann. Neurol.* 80.4 (Oct. 2016), pp. 616–624. DOI: 10.1002/ana.24769.

[5] Philipp Mahlknecht et al. "Prodromal Parkinson's disease: hype or hope for disease-modification trials?" en. In: *Transl. Neurodegener.* 11.1 (Feb. 2022), p. 11. DOI: 10.1186/s40035-022-00286-1.

[6] Harald Hampel et al. "The foundation and architecture of precision medicine in neurology and psychiatry". en. In: *Trends Neurosci.* 46.3 (Mar. 2023), pp. 176–198. DOI: 10.1016/j.tins.2022.12.004.

[7] Gavin Giovannoni et al. "Brain health: time matters in multiple sclerosis". en. In: *Mult. Scler. Relat. Disord.* 9 (Sept. 2016), S5–S48. DOI: 10.1016/j.msard.2016.07.003.

[8] A P Porsteinsson et al. "Diagnosis of Early Alzheimer's Disease: Clinical Practice in 2021". en. In: *The Journal of Prevention of Alzheimer's Disease* 8.3 (June 2021), pp. 371–386. DOI: 10.14283/jpad.2021.23.

[9] Abdur Rauf and Md Mominur Rahman. "Potential therapeutics against neurological disorders: Natural products-based drugs". en. In: *Front. Pharmacol.* 13 (2022). DOI: 10.3389/fphar.2022.950457.

[10] Benjamin Blanco, Toni Mora, and Marta Trapero-Bertran. "Direct health care costs associated to neurological diseases and different degrees of malnutrition". en. In: *Clin Nutr ESPEN* 44 (Aug. 2021), pp. 297–305. DOI: 10.1016/j.clnesp.2021.05.028.

[11] Alberto Lleó. "Biomarkers in neurological disorders: a fast-growing market". en. In: *Brain Commun* 3.2 (May 2021), fcab086. DOI: 10.1093/braincomms/fcab086.

[12] Beomsue Kim et al. "A brief review of non-invasive brain imaging technologies and the near-infrared optical bioimaging". en. In: *Appl Microsc* 51.1 (June 2021), p. 9. DOI: 10.1186/s42649-021-00058-7.

[13] M Symms et al. "A review of structural magnetic resonance neuroimaging". en. In: *J. Neurol. Neurosurg. Psychiatry* 75.9 (Sept. 2004), pp. 1235–1244. DOI: 10.1136/jnnp.2003.032714.

[14] Yuhui Du, Zening Fu, and Vince D Calhoun. "Classification and Prediction of Brain Disorders Using Functional Connectivity: Promising but Challenging". en. In: *Front. Neurosci.* 12 (Aug. 2018), p. 525. DOI: 10.3389/fnins.2018.00525.

[15] Alexander Seiler et al. "Multiparametric Quantitative MRI in Neurological Diseases". en. In: *Front. Neurol.* 12 (Mar. 2021), p. 640239. DOI: 10.3389/fneur.2021.640239.

[16] Michelle M Miller-Thomas and Tammie L S Benzinger. "Neurologic Applications of PET/MR Imaging". en. In: *Magn. Reson. Imaging Clin. N. Am.* 25.2 (May 2017), pp. 297–313. DOI: 10.1016/j.mric.2016.12.003.

[17] William C Kreisl et al. "PET imaging of neuroinflammation in neurological disorders". en. In: *Lancet Neurol.* 19.11 (Nov. 2020), pp. 940–950. DOI: 10.1016/S1474-4422(20)30346-X.

[18] Pieter-Jan Buyck et al. "Diagnostic accuracy of noncontrast CT imaging markers in cerebral venous thrombosis". en. In: *Neurology* 92.8 (Feb. 2019), e841–e851. DOI: 0.1212/WNL.0000000000006959.

[19] Akmal Sabarudin, Cantiriga Subramaniam, and Zhonghua Sun. "Cerebral CT angiography and CT perfusion in acute stroke detection: a systematic review of diagnostic value". en. In: *Quant. Imaging Med. Surg.* 4.4 (Aug. 2014), pp. 282–290. DOI: 10.3978/j.issn.2223-4292.2014.07.10.

[20] Jelle Demeestere et al. "Review of Perfusion Imaging in Acute Ischemic Stroke". en. In: *Stroke* (Mar. 2020). DOI: 10.1161/STROKEAHA.119.028337.

[21] Achim Seeger et al. "Acute stroke imaging: feasibility and value of MR angiography with high spatial and temporal resolution for vessel assessment and perfusion analysis in patients with wake-up stroke". en. In: *Acad. Radiol.* 22.4 (Apr. 2015), pp. 413–422. DOI: 10.1016/j.acra.2014.11.013.

[22] Kelly Rootes-Murdy et al. "Federated Analysis of Neuroimaging Data: A Review of the Field". en. In: *Neuroinformatics* 20.2 (Apr. 2022), pp. 377–390. DOI: 10.1007/s12021-021-09550-7.

[23] *Information technology – Artificial intelligence – Overview of trustworthiness in artificial intelligence*. Standard. Geneva, CH: International Organization for Standardization, May 2020.

[24] Alexey Dosovitskiy et al. "An image is worth 16x16 words: Transformers for image recognition at scale". In: *arXiv preprint arXiv:2010.11929* (2020). DOI: 10.48550/arXiv.2010.11929.

[25] Ambuj Mehrish et al. "A review of deep learning techniques for speech processing". In: (2023). DOI: 10.48550/arXiv.2305.00359.

[26] Aditya Ramesh et al. "Zero-shot text-to-image generation". In: (2021). DOI: 10.48550/arXiv.2102.12092.

[27] Sakib Shahriar and Kadhim Hayawi. "Let's have a chat! A Conversation with ChatGPT: Technology, Applications, and Limitations". In: (2023). DOI: 10.47852/bonviewAIA3202939.

[28] Randall Balestriero et al. “A cookbook of self-supervised learning”. In: (2023). DOI: 10.48550/arXiv.2304.12210.

[29] Hugo J W L Aerts et al. “Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach”. en. In: *Nat. Commun.* 5 (June 2014), p. 4006. DOI: 10.1038/ncomms5006.

[30] Philippe Lambin et al. “Radiomics: extracting more information from medical images using advanced feature analysis”. In: *European journal of cancer* 48.4 (2012), pp. 441–446. DOI: 10.1016/j.ejca.2011.11.036.

[31] Parnian Afshar et al. “From handcrafted to deep-learning-based cancer radiomics: Challenges and opportunities”. In: *IEEE Signal Process. Mag.* 36.4 (July 2019), pp. 132–160. DOI: 10.1109/MSP.2019.2900993.

[32] Ana Rodrigues et al. “Value of handcrafted and deep radiomic features towards training robust machine learning classifiers for prediction of prostate cancer disease aggressiveness”. en. In: *Sci. Rep.* 13.1 (Apr. 2023), pp. 1–10. DOI: 10.1038/s41598-023-33339-0.

[33] Zohaib Salahuddin et al. “Transparency of deep neural networks for medical image analysis: A review of interpretability methods”. en. In: *Comput. Biol. Med.* 140 (Dec. 2021), p. 105111. DOI: 10.1016/j.compbiomed.2021.105111.

[34] Ramprasaath R Selvaraju et al. “Grad-cam: Visual explanations from deep networks via gradient-based localization”. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 618–626. DOI: 10.1109/ICCV.2017.74.

[35] Kaustav Bera et al. “Predicting cancer outcomes with radiomics and artificial intelligence in radiology”. en. In: *Nat. Rev. Clin. Oncol.* 19.2 (Feb. 2022), pp. 132–146. DOI: 10.1038/s41571-021-00560-7.

[36] Philipp Lohmann et al. “Radiomics in neuro-oncological clinical trials”. In: *Lancet Digit Health* 4.11 (Nov. 2022), e841–e849. DOI: 10.1016/S2589-7500(22)00144-3.

[37] Sumeet Hindocha et al. “Gross tumour volume radiomics for prognostication of recurrence & death following radical radiotherapy for NSCLC”. en. In: *NPJ Precis Oncol* 6.1 (Oct. 2022), p. 77. DOI: 10.1038/s41698-022-00322-3.

[38] Janita E van Timmeren et al. “Radiomics in medical imaging-“how-to” guide and critical reflection”. en. In: *Insights Imaging* 11.1 (Aug. 2020), p. 91. DOI: 10.1186/s13244-020-00887-2.

[39] Nicolae Sarbu et al. “White matter diseases with radiologic-pathologic correlation”. In: *Radiographics* 36.5 (2016), pp. 1426–1447. DOI: 10.1148/rg.2016160031.

[40] Alan J Thompson et al. “Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria”. In: *The Lancet Neurology* 17.2 (2018), pp. 162–173. DOI: 10.1016/S1474-4422(17)30470-2.

[41] Yulin Ge et al. “Brain atrophy in relapsing-remitting multiple sclerosis: fractional volumetric analysis of gray matter and white matter”. In: *Radiology* 220.3 (2001), pp. 606–610. DOI: 10.1148/radiol.2203001776.

[42] Valery N Kornienko and Igor Nicolaevich Pronin. *Diagnostic neuroradiology*. Springer Science & Business Media, 2008.

[43] Ruth Geraldes et al. “The current role of MRI in differentiating multiple sclerosis from its imaging mimics”. In: *Nature Reviews Neurology* 14.4 (2018), pp. 199–213. DOI: 10.1038/nrneurol.2018.39.

[44] Ralph L Sacco et al. “An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke

Association". In: *Stroke* 44.7 (2013), pp. 2064–2089. DOI: 10.1161/STR.0b013e318296aeca.

[45] Chandril Chugh. "Acute ischemic stroke: management approach". In: *Indian journal of critical care medicine: peer-reviewed, official publication of Indian Society of Critical Care Medicine* 23.Suppl 2 (2019), S140. DOI: 10.5005/jp-journals-10071-23192.

[46] Shinichi Nakano et al. "Correlation of early CT signs in the deep middle cerebral artery territories with angiographically confirmed site of arterial occlusion". In: *American journal of neuroradiology* 22.4 (2001), pp. 654–659.

[47] Christopher A Potter et al. "CT for treatment selection in acute ischemic stroke: a code stroke primer". In: *Radiographics* 39.6 (2019), pp. 1717–1738. DOI: 10.1148/rg.2019190142.

[48] Julia Hopyan et al. "Certainty of stroke diagnosis: incremental benefit with CT perfusion over noncontrast CT and CT angiography". In: *Radiology* 255.1 (2010), pp. 142–153. DOI: 10.1148/radiol.09091021.

[49] Vanja Douglas, Michel Shamy, and Pratik Bhattacharya. *Should CT angiography be a routine component of acute stroke imaging?* 2015. DOI: 10.1177/1941874415588393.

[50] Laura M Allen et al. "Sequence-specific MR imaging findings that are useful in dating ischemic stroke". In: *Radiographics* 32.5 (2012), pp. 1285–1297. DOI: 10.1148/rg.325115760.

[51] Chen Ma, Fenfang Hong, and Shulong Yang. "Amyloidosis in Alzheimer's disease: Pathogeny, etiology, and related therapeutic directions". In: *Molecules* 27.4 (2022), p. 1210. DOI: 10.3390/molecules27041210.

[52] Marie Sarazin et al. "Clinical and research diagnostic criteria for Alzheimer's disease". In: *Neuroimaging Clinics* 22.1 (2012), pp. 23–32. DOI: 10.1016/j.nic.2011.11.004.

[53] JD Warren et al. “Brain biopsy in dementia”. In: *Brain* 128.9 (2005), pp. 2016–2025. DOI: 10.1093/brain/awh543.

[54] Kaj Blennow. “A review of fluid biomarkers for Alzheimer’s disease: moving from CSF to blood”. In: *Neurology and therapy* 6 (2017), pp. 15–24. DOI: 10.1007/s40120-017-0073-9.

[55] Heiko Braak et al. “Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry”. In: *Acta neuropathologica* 112.4 (2006), pp. 389–404. DOI: 10.1007/s00401-006-0127-z.

[56] Freddie Márquez and Michael A Yassa. “Neuroimaging biomarkers for Alzheimer’s disease”. In: *Molecular neurodegeneration* 14 (2019), pp. 1–14. DOI: 10.1186/s13024-019-0325-5.

[57] Roberto Casale et al. “Development and external validation of a non-invasive molecular status predictor of chromosome 1p/19q co-deletion based on MRI radiomics analysis of Low Grade Glioma patients”. In: *European Journal of Radiology* 139 (June 2021), p. 109678. DOI: 10.1016/j.ejrad.2021.109678.

[58] Elizaveta Lavrova et al. “Exploratory radiomic analysis of conventional vs. quantitative brain MRI: toward automatic diagnosis of early multiple sclerosis”. In: *Frontiers in neuroscience* 15 (2021), p. 679941. DOI: 10.3389/fnins.2021.679941.

[59] Elizaveta Lavrova et al. “UR-CarA-Net: A Cascaded Framework With Uncertainty Regularization for Automated Segmentation of Carotid Arteries on Black Blood MR Images”. In: *IEEE Access* 11 (2023), pp. 26637–26651. DOI: 10.1109/ACCESS.2023.3258408.

[60] Elizaveta Lavrova et al. “Precision-medicine-toolbox: An open-source python package for the quantitative medical image analysis”. In: *Software Impacts* 16 (2023), p. 100508. DOI: 10.1016/j.simpa.2023.100508.

2

Radiomics capability in neuro-oncological tasks

Roberto Casale, Elizaveta Lavrova, Sebastian Sanduleanu, Henry C. Woodruff, Philippe Lambin

Adapted from: Roberto Casale et al. "Development and external validation of a non-invasive molecular status predictor of chromosome 1p/19q co-deletion based on MRI radiomics analysis of Low Grade Glioma patients". In: *European Journal of Radiology* 139 (June 2021), p. 109678. DOI: 10.1016/j.ejrad.2021.109678.

Abstract

Purpose. The 1p/19q co-deletion status has been demonstrated to be a prognostic biomarker in lower grade glioma (LGG). The objective of this study was to build a magnetic resonance (MRI)-derived radiomics model to predict the 1p/19q co-deletion status.

Method. 209 pathology-confirmed LGG patients from 2 different datasets from The Cancer Imaging Archive were retrospectively reviewed; one dataset with 159 patients as the training and discovery dataset and the other one with 50 patients as validation dataset. Radiomics features were extracted from T2- and T1-weighted post-contrast MRI resampled data using linear and cubic interpolation methods. For each of the voxel resampling methods a three-step approach was used for feature selection and a random forest (RF) classifier was trained on the training dataset. Model performance was evaluated on training and validation datasets and clinical utility indexes (CUIs) were computed. The distributions and intercorrelation for selected features were analyzed.

Results. Seven radiomics features were selected from the cubic interpolated features and five from the linear interpolated features on the training dataset. The RF classifier showed similar performance for cubic and linear interpolation methods in the training dataset with accuracies of 0.81 (0.75-0.86) and 0.76 (0.71-0.82) respectively; in the validation dataset the accuracy dropped to 0.72 (0.6-0.82) using cubic interpolation and 0.72 (0.6-0.84) using linear resampling. CUIs showed the model achieved satisfactory negative values (0.605 using cubic interpolation and 0.569 for linear interpolation).

Conclusions. MRI has the potential for predicting the 1p/19q status in LGGs. Both cubic and linear interpolation methods showed similar performance in external validation.

2.1 Introduction

Gliomas are tumors of the central nervous system and are the most frequent primary tumors arising in the brain [1]. They are classified into four grades based on their aggressiveness by The World Health Organization (WHO). WHO grade II (low grade) and grade III (anaplastic) diffuse gliomas form a heterogeneous group of neoplasms, also known as Low Grade Gliomas (LGGs), characterized by a wide range of malignant potential affecting mostly young adults [2]; LGG is potentially a fatal disease, with an median overall survival of around 7 years [3]. LGG finally advances to higher grades, with a significantly lower survival rate [3].

Treatment choices for LGG are based on WHO grades, molecular profiles, and patient characteristics (e.g. age and Karnofsky performance status) [4]. The co-deletion of chromosome arms 1p and 19q has an important role in choosing the right treatment, indeed co-deletion is a useful prognostic molecular marker as it can be used for the prediction of response to radiotherapy and chemotherapy, and it is associated with longer survival [5, 6, 7, 8]. Thus, efficient treatment planning necessitates proper classification of WHO grade and 1p/19q co-deletion status.

The 1p/19q status can be determined by different techniques: fluorescence in situ hybridization (FISH), polymerase chain reaction, array comparative genomic hybridization, or multiplex ligation-dependent probe amplification [9]. This molecular classification is achieved through histopathologic examination; albeit being the reference standard for this task, it has some limits, such as limited surgical accessibility and heterogeneity of the sampled tissue. Furthermore, biopsy samples are not representative of the whole neoplasm [10].

The unmet clinical need is to find a non-invasive and robust classification method of 1p/19q status of the entire tumor volume in order to effectively direct treatment planning of LGG [5, 6, 7, 8] for cases when complete resection cannot be performed and/or where the biopsy cannot be obtained from the tumor. Most notably, in

childhood tumors around 30–50% of LGGs are inoperable as a result of their position in highly eloquent areas of the brain [11]. Currently, MRI is a useful technique in order to obtain helpful data for therapy decisions, and for pre-therapeutic noninvasive diagnosis.
Radiomics is a research field whose scope is to extract imaging features from radiographic images (including MRI) that can potentially capture phenotypic, genomic, proteomics patterns having prognostic value and clinical significance. The underlying hypothesis of radiomics is that medical imaging may express additional data correlating with genomic and proteomics patterns and can be manifested in macroscopic image-based features, not visible by the unaided eye and thus not used [12, 13, 14].
In the last few years different studies have demonstrated that 1p19q status can be predicted using MRI [15, 16, 17, 18, 19, 20, 21, 22]. Furthermore, Branzoli et al. [23] recently identified elevated levels of cystathionine in 1p/19q co-deleted gliomas compared to non co-deleted gliomas, using in vivo magnetic resonance spectroscopy. In our analysis, routine MRI sequences were used, without additional experimental or expensive MRI sequences.
The main purpose of this study was to develop and to validate a non-invasive method to predict the 1p/19q status of LGG from T2-weighted and T1-weighted post-contrast MRI images using texture analysis as an alternative to surgical biopsy. The secondary aim was to compare two voxel resampling methods: radiomics features calculated from images resampled using cubic and linear interpolation methods. Cubic spline and convolution interpolation are third-order methods that typically interpolate smoother surfaces than linear methods, while they are known to be slower in implementation [24]. Linear interpolation is a commonly used algorithm since it is computationally cheap and leads neither to rough blocking artifacts images that are generated by nearest neighbor techniques, nor will it cause out-of-range gray levels that might be produced by higher order interpolation [25].

2.2 Materials and methods

2.2.1 Data

The training dataset consisted of 159 LGG patients with pre-operative MRI images and 1p/19q status proven by biopsy. They were identified within the LGG-1p19q Deletion dataset [15, 26, 27] on The Cancer Imaging Archive (TCIA). The validation dataset consisted of similar patient data of 50 randomly selected patients, also from TCIA, albeit in the TCGA-LGG dataset [27, 28]. For TCGA, the 1p/19q status for validation dataset was derived from a previous study based on this dataset [16]. Patients were selected according to the following inclusion criteria: exams with a slice thickness ≤ 7.5 mm, artifacts in less than 50% of the slices containing the gross tumor volume (GTV) visually assessed by a radiologist with 3 years' experience (RC), and the presence of T2-weighted and contrast enhanced T1-weighted images and 1p/19q status. The GTV was delineated using MIM software version 6.9.0 (MIM, Cleveland, United States) by one radiologist (RC).

2.2.2 Image pre-processing and radiomic feature extraction

In order to somewhat account for inter-scanner variability, Z-score normalization was applied to the GTVs in each image series (per patient). The formula for Z-score normalization for GTV intensities is: $\frac{OriginalIntensityValue-\mu}{\sigma}$, where μ is the mean intensity inside each GTV and σ is the intensity standard deviation in each GTV.
Voxel size resampling was performed before feature extraction using cubic and linear interpolation separately. Images were resampled to a voxel size of 3 mm $\times$3 mm$\times$3 mm.
To reduce noise and computational burden, grayscale values were aggregated into the same number of bins (50 bins) for all MRI exams. The fixed bin number method was used to achieve a better normalizing effect as intensity units are not absolute in MRI [25]. Radiomics features compliant with the International Biomarker Standardization Initiative (IBSI), as well as non-IBSI features were

extracted with the RadiomiX research software (supported by Oncoradiomics, Liège, Belgium).
Radiomics features were extracted consisting of five main groups: 1) fractal features, 2) first order statistics, 3) shape and size, 4) texture descriptors including gray level co-occurrence (GLCM), gray level run-length (GLRLM) and gray level size-zone texture matrices (GLSZM), 5) features from groups 1, 3 and 4 after wavelet decomposition. GLCM distance was 1.

2.2.3 Feature selection and statistical analysis

The Figure 2.1 illustrates the 3 steps that were performed only on the training dataset for feature selection; the second step was repeated 300 times, with different sample groupings. All this procedure was performed twice, for cubic and linear interpolation respectively. The first step used correlation-based feature subset selection (CfsSubsetEval function, Weka software version 3.8.3) [29, 30, 31] to eliminate irrelevant and redundant features. In the second step, a table was created that ordered and ranked features according to their importance using a 10-fold cross validation treebag recursive feature elimination algorithm (RFE) (Python 3.7.6 version, scikit-learn 0.21.2 package). Finally, in the third step, a learning curve was computed (AUC vs. the incremental number of features obtained from the ranked feature table).
Inter-correlation among selected features and with volume were calculated with the Spearman correlation coefficient. Moreover, the Mann-Whitney test was applied in order to check statistically significant differences in GTV values in co-deleted/non co-deleted groups in the training dataset.
Statistical analysis was performed with Python 3.7.6 version (scipy 1.4.1 package, pandas 1.0.0 package).

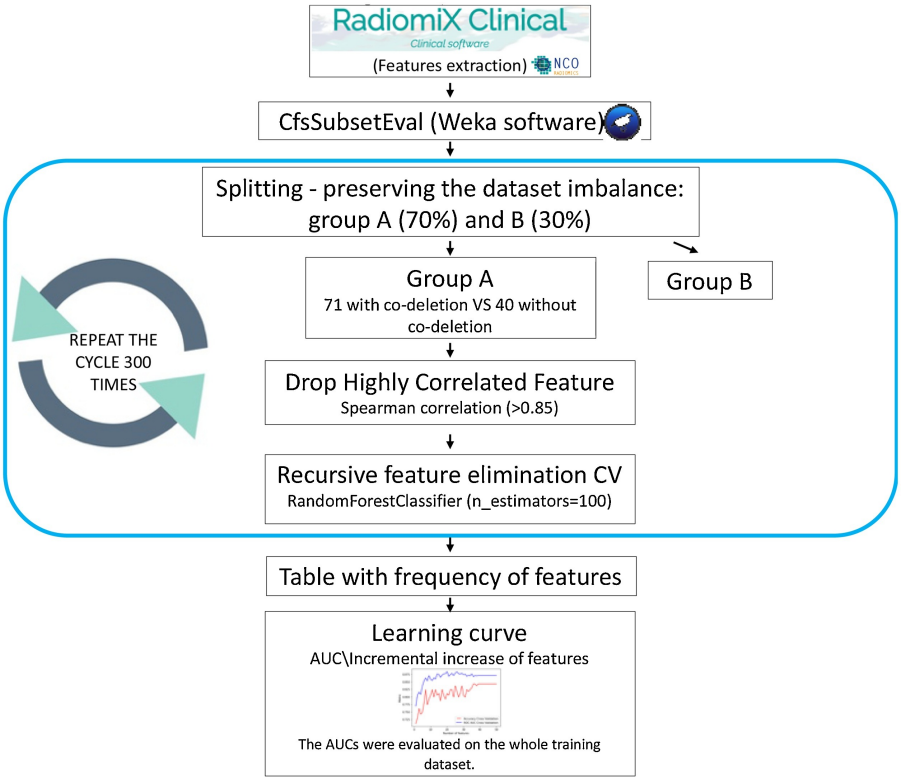


Figure 2.1: Feature selection (only training dataset).

2.2.4 Classification

A random forest (RF) classification model was trained on the training dataset with the selected features, and performance metrics calculated when applied to both datasets without further adjustments. To mitigate the effect of the unbalanced outcomes, the training dataset was balanced using an adaptive synthetic (ADASYN) resampling approach, which creates artificial patients for the minority class, before the RF model was trained [32]. On the training dataset, internal 10-fold cross-validation was performed, followed by a

bootstrap method (n=10000) to have an evaluation of the error of the performance metrics (median, 2.5 and 97.5 percentiles). On the validation dataset, a bootstrap method (n=10000) was implemented, and the median values and 2.5 and 97.5 percentiles calculated. During the cross-validation procedure, each set preserved roughly the same ratio of samples for each class (co-deleted/non co-deleted) as the complete training dataset and ADASYN applied to the training fold.
Accuracy, sensitivity, specificity, receiver operating characteristic curve (ROC) and AUC were computed. All these steps of the workflow were repeated twice (for cubic interpolation and linear interpolation). Classification performance was compared for cubic and linear interpolation-based data for both training dataset cross-validation results and validation results; the DeLong test was used to compare AUCs obtained from each model.
This segment and statistical analysis were performed with Python 3.7.6 version (scikit-learn 0.21.2 package, scipy 1.4.1 package), and R 3.6.1 version (pROC 1.14.0 package).

2.2.5 TRIPOD and Radiomics Quality Score

This study followed the instruction of Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD), and the Radiomics quality score (RQS) was used to evaluate the radiomics workflow [12, 33, 34]. The RQS score for this specific study was 44%. The RQS maximum score is 100% and it is based on a 36 points system; a high value reveals a higher methodological quality research and reporting [33].

2.2.6 Clinical utility index (CUI)

Clinical utility indexes were computed for the RF model tested on external validation dataset. CUI was developed in 2007 and aimed to take into account both occurrence and discrimination [35, 36, 37, 38, 39, 40]. The value for CUI ranges from 0 to 1: excellent utility (CUI ≥ 0.81),

good utility (CUI ≥ 0.64), satisfactory/fair utility (CUI ≥ 0.49), and otherwise poor [36].

2.3 Results

2.3.1 Data

Training dataset

One hundred and fifty-nine consecutive LGG patients with pre-operative MRI images collected between 01-10-2002 and 01-09-2011 and biopsy proven 1p/19q status were identified within the LGG-1p19q Deletion archive. The data included 102 patients with co-deleted 1p/19q arms and 57 with non-co-deleted arms. The grades of the LGG lesions were II ($n = 104$) and III ($n = 55$). The types of LGG were oligoastrocytoma ($n = 97$), oligodendrogliomas ($n = 45$), and astrocytomas ($n = 17$). Median age was 42 (range 13–84) and this dataset included 76 women and 83 men. Post-contrast T1- and T2-weighted images were available for all selected patients. All images were acquired with 1.5T or 3T scanners, slice thicknesses ranged from 1 to 7.5 mm and isotropic pixel size in the axial plane ranged from 0.43 to 1.09 mm.

External validation dataset

No significant differences in gender ($\frac{M}{F} = 1.1$ in training set vs. $\frac{M}{F} = 1.3$ in validation set) and WHO grade ratios ($\frac{II}{III} = 1.9$ in training set vs. $\frac{II}{III} = 1.4$) were observed between the training and validation sets. There were significant differences in histology and age (mean age 46.5 in training set vs. 41.6 in validation set). Level of significance was $\alpha = 0.05$ for Chi-square tests and Mann-Whitney test (age comparison). Demographic and clinical data description are presented in Table 2.1.

Table 2.1: Data description

	Training dataset	Validation dataset	p-value
Number of patients	159	60	—
Age in years, mean (SD)	41.6 (13.8)	46.5 (13.0)	0.026
Gender ratio (M/F)	83/76	28/22	0.759
Grade ratio (II/III)	104/55	29/21	0.435
Histology ratio (astrocytoma/ oligoastrocytoma/ oligodendroglioma)	17/97/45	8/14/28	0.000
Outcome ratio (co-deletion/ non co-deletion)	102/57	25/25	—

The p-values for statistically significant differences of value distribution in training and validation datasets were calculated with the following statistical tests: age — Mann-Whitney, gender ratio — chi-square, grade ratio — chi-square, histology ratio – chi square.

2.3.2 Radiomic feature extraction, selection, and statistical analysis

In total, 5352 radiomics features per patient were extracted from both T1- and T2-weighted images; 2676 features extracted with each of cubic and linear interpolation voxel resampling methods.
After correlation-based feature subset selection a total of 48 features remained for cubic interpolation and 51 features for linear interpolation.
These remaining features were fed into 300 loops of 10-fold cross-validation RFE. GTV volume was not chosen among the selected features.
With these ranked features, two learning curves were computed (AUC vs. incremental increase of features) respectively for cubic interpolation and linear interpolation using only the training dataset.

Table 2.2: Selected features for cubic interpolation with frequency.

Features	Frequency
GLCM_average (T2w)	293
Wavelet_LHL_Stats_median (T1w)	289
Wavelet_LLH_Stats_median (T1w)	276
GLCM_clusShade (T1w)	258
Wavelet_LHH_Fractal_lacunarity (T2w)	223
Wavelet_HLL_GLCM_correl1 (T2w)	212
Wavelet_LLL_Stats_p10 (T2w)	206

Frequency is number of times the feature was selected during the 300 loops.

Table 2.3: Selected features for linear interpolation with frequency.

Features	Frequency
Wavelet_LLH_Stats_median (T1w)	275
Wavelet_LHL_Stats_median (T1w)	271
Wavelet_LLL_IH_p10 (T1w)	257
GLCM_clusShade (T1w)	213
Wavelet_LHH_Fractal_lacunarity (T2w)	205

Frequency is number of times the feature was selected during the 300 loops.

The classifier used to generate the curve was RF, with co-deleted/ non co-deleted outcome and 10-fold cross validation.
The number of features for the final model was chosen near the first salient point of the learning curve for AUC score. All features that went into the model satisfy the condition that they were selected more than 68% (greater than or equal to 205 times) in the RFE loops to ensure a certain level of robustness.
Finally, the selected features were 7 for cubic interpolation (Table 2.2) and 5 for linear interpolation (Table 2.3).

Table 2.4: Classification performance scores obtained on training dataset (10-fold cross validation).

	Cubic interpolation, median [2.5-97.5 percentile]	Linear interpolation, median [2.5-97.5 percentile]
Accuracy	0.81 [0.75-0.86]	0.76 [0.71-0.82]
Sensitivity	0.77 [0.69-0.85]	0.72 [0.63-0.80]
Specificity	0.85 [0.78-0.92]	0.81 [0.74-0.88]
AUC	0.86 [0.81-0.91]	0.82 [0.75-0.87]

2.3.3 Classification

Results on training dataset

All results are reported as the median [2.5th-97.5th percentile]. For cubic interpolation, the RF model achieved an AUC of 0.86 [0.81-0.91] and for linear interpolation an AUC of 0.82 [0.75-0.87] (Table 2.4).
The DeLong test was used to compare model performances obtained from models trained on data that underwent cubic and linear interpolation. According to the results of this test, there was no statistically significant difference between the two AUCs ($p = 0.073$).
The Mann-Whitney, applied to GTV values in co-deleted/non co-deleted groups, shows no statistical difference between the two groups ($p = 0.149$; $\alpha = 0.05$).

Results on validation dataset

The AUC for features extracted for cubic interpolation was 0.87 [0.76-0.95] and for linear interpolation was 0.77 [0.61-0.89] (Table 2.5). According to the DeLong test there was no statistically significant difference between the two models ($p = 0.178$).

The ROC curves are shown in Figure 2.2 (for cubic interpolation) and Figure 2.3 (for linear interpolation).

Table 2.5: Classification performance scores obtained on external validation dataset.

	Cubic interpolation, median [2.5-97.5 percentile]	Linear interpolation, median [2.5-97.5 percentile]
Accuracy	0.72 [0.60-0.82]	0.72 [0.60-0.84]
Sensitivity	0.52 [0.32-0.72]	0.60 [0.40-0.80]
Specificity	0.92 [0.80-1.00]	0.84 [0.68-0.96]
AUC	0.87 [0.76-0.95]	0.77 [0.61-0.89]

2.3.4 Clinical utility index (CUI)

The positive CUI, calculated for the RF model with cubic interpolation features and tested on validation dataset, was 0.451 (CI: 0.203-0.698); the negative CUI was 0.605 (CI: 0.483-0.762). The positive and negative CUI values obtained with cubic interpolation features had a poor and a satisfactory/fair utility value respectively.
The positive CUI for the RF model, obtained with linear interpolation features and tested on validation dataset, was 0.474 (CI: 0.238-0.709), so with a poor utility value; the negative CUI for the model obtained with linear interpolation features was 0.569 (CI: 0.435-0.703), so with a satisfactory/fair utility value.
These results showed that the RF model, trained both with cubic and linear interpolation features, achieved a satisfactory negative CUI, meaning that this algorithm can be reasonably useful for screening patients with 1p-19q non-co-deletion status. On the other side, the RF model, trained both with cubic and linear interpolation features, achieved a poor positive CUI, meaning that this method has low utility to confirm patients with non-co-deleted status; in practical terms, if a patient obtains a result that suggest having non-co-deleted status, this patient should be further studied to confirm the non-co-deleted status.

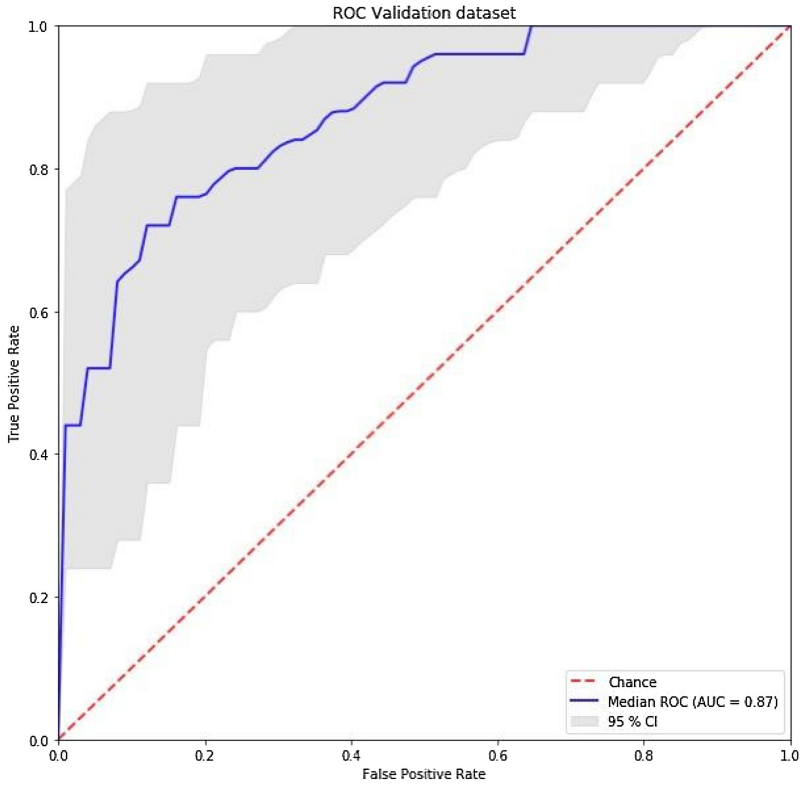


Figure 2.2: ROC AUC for features calculated with cubic interpolation - results obtained on validation dataset.

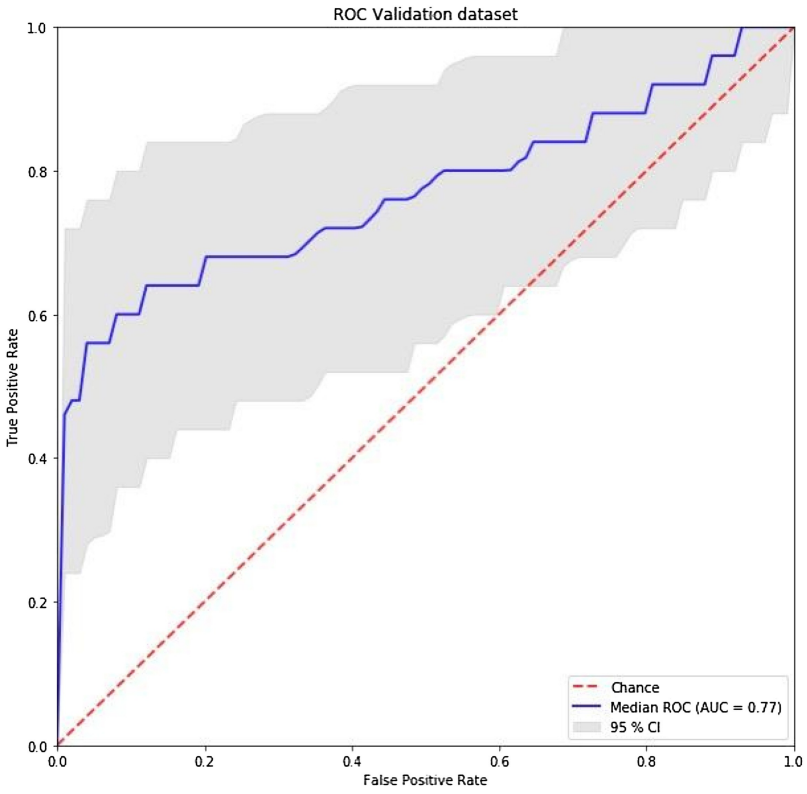


Figure 2.3: ROC AUC for features calculated with linear interpolation - results obtained on validation dataset.

2.4 Discussion

In this study we explored the ability of radiomics features extracted from the GTV on preoperative MRI (acquired with T1-weighted contrast enhanced and T2-weighted sequences) to predict molecular status of chromosome 1p/19q co-deletion in LGG patients. To investigate the influence of the resampling method on the classification models' performance, we used both cubic and linear interpolation kernels for further comparison. After feature selection, the feature vectors contained 5 and 7 features for cubic and linear interpolation-based data, respectively. These feature vectors only had 4 common features (Wavelet_LLH_Stats_median (T1w), Wavelet_LHL_Stats_median (T1w), GLCM_clusShade (T1w) and Wavelet_LHH_Fractal_lacunarity (T2w)). Therefore, we conclude that the method presented is not completely robust to the resampling method and additional studies on features reproducibility are needed. According to Spearman correlation coefficient, within the training dataset, the feature vectors consisted of statistically independent features. In validation dataset, some of these features are correlated to each other (Spearman correlation coefficient 0.76 and 0.77 for T1-weighted Wavelet_LHL_Stats˙median (T1w) and Wavelet_LLH_-Stats˙median (T1w) for cubic and linear interpolation-based features, respectively).
According to DeLong test, there were no statistically significant differences between AUCs obtained from the cubic interpolation model and linear interpolation model both on training dataset ($p = 0.178$) and validation dataset ($p = 0.073$).
The advantages of the present study are its non-invasiveness, the analysis of the entire volume of the lesion, and the ubiquitous availability, as it is based on simple conventional MRI sequences.
Other studies try to predict 1p/19q status, some of which aim to solve the same problem using MRI [15, 16, 17, 18, 19, 20, 21, 22]. They all are using multimodal conventional MRI data, most often combining T2-weighted and contrast enhanced T1-weighted data together. Classification performance of the present study did not

exceed the results, obtained in [15, 16, 17, 19, 20]. Nevertheless, the present study has some benefits over previously mentioned studies [15, 16, 17, 20]: (1) the potential reproducibility, achieved with open source data usage and utilization of an automated pipeline, (2) the potential interpretability of results, as input features are known and understood, (3) the presence of clinical utility evaluation, (4) the evaluation of two different resampling methods.

The present study has some limitations. The main limitation is the relatively small sample size, which decreases statistical power of the classification results. For this reason, to test the model, we performed cross-validation on the training dataset and then we trained it on the whole training dataset to perform validation on external dataset. Also, for this reason, to estimate model performance and its error on external validation dataset, we performed a bootstrapping approach, which produces multiple instances of the same observations and omits other original observations. The second limitation was related to data balance within and between training and validation datasets. Outcomes in the training dataset were significantly unbalanced (102 cases of co-deletion vs. 57 cases of non-co-deletion); to partially overcome this limitation, the ADASYN method was used, which is not without uncertainties. The third limitation was related to significant differences in histology and age distribution in training and validation datasets. Histology effect and age have not been investigated and included into models and they could be explored in further studies. The fourth limitation was related to different MRI field strengths, values of slice thickness (0.9-7.5 mm) and isotropic pixel spacing (0.39-1.09 mm); these differences could be a source of batch effects, modifying radiomics features significantly, but also could be an opportunity to test the stability of methods across different image acquisition parameters. The fifth limitation arises from possible bias stemming from the random selection for 50 patients inside the validation dataset.

In summary, the proposed non-invasive method is able to predict molecular status of chromosome 1p/19q co-deletion in LGG patients, based on multi-scanner multi-field MRI data. Although there is

still room for improvement in accuracy metrics, its usefulness was indicated for the estimation of prognostic molecular markers. Results of its validation on external data demonstrated its generalizability. According to the results of statistical tests, there were no statistically significant differences between the AUCs obtained with different spatial resampling interpolation methods (cubic and linear).
Regarding the diagnostic utility of this method, the CUI demonstrated that the RF model (trained both with cubic and linear interpolation features) achieved a satisfactory negative CUI, while the RF model (trained both with cubic and linear interpolation features) achieved a poor positive CUI. Therefore, linear and cubic models can be reasonably helpful for ruling out non-co-deleted status, but they can be poorly useful for confirming non-co-deleted status. This difference can be explained by the different accuracy metrics: indeed, both algorithms had specificity and positive predictive values higher than sensitivity and negative predictive values; moreover, the unbalanced class in the training dataset could affect the performance. These results should be considered in future studies and should be taken into account in a future clinical scenario.
This approach may be an opportunity to help medical decision. Despite the dataset was limited, ADASYN increased the number of cases in the training phase. However, further studies based on more heterogeneous and larger patient population are mandatory to confirm and validate our current results.

2.5 Conclusions

MRI radiomics analysis, based on T2-weighted and T1-weighted post-contrast images, could supply a reliable noninvasive technique for the prediction of 1p/19q status in LGGs, giving useful information for personalized therapy assessment and pretreatment prediction. Regarding the two different voxel resampling methods, no statistically significant differences were found.

2.6 Data availability statement

The dataset and GTV used in this article can be provided upon contact with the corresponding author. The python code used for the feature selection, classification model and evaluation of the algorithm is available on GitHub `https://github.com/roberto-casale/LGG-1p-19q-deletion`. Supplementary material is available via the link `https://www.ejradiology.com/article/S0720-048X(21)00158-3/fulltext#supplementaryMaterial`. Supplementary material contains:

- outcomes for both training and external validation datasets (Tables 1 and 2 corespondingly), detailed feature list (Table 3), definitions of texture matrices (Table 4), results for feature selection with RFE for cubic and linear interpolation implementation of the pipeline (Tables 5 and 6), confusion matrices for cubic and linear interpolation implementation of the pipeline (Tables 7 and 8),
- learning curves for cubic and linear interpolation implementation of the pipeline (Figures 1 and 2), correlation matrices for cubic and linear interpolation implementation of the pipeline in training and external validation datasets (Figures 3, 4, 5, and 6), distributions of the values of the selected features for cubic and linear interpolation implementation of the pipeline in training and external validation datasets (Figures 7 and 8),
- description of voxel size selection, detailed description of the features selection pipeline, definition and interpretation of CUI.

References

[1] S Cha. "Update on brain tumor imaging: from anatomy to physiology". In: *American Journal of Neuroradiology* 27.3 (2006), pp. 475–487.

[2] Andrea Lanese, Enrico Franceschi, and Alba A Brandes. "The risk assessment in low-grade gliomas: an analysis of the European Organization for Research and Treatment of Cancer (EORTC) and the Radiation Therapy Oncology Group (RTOG) criteria". In: *Oncology and Therapy* 6.2 (2018), pp. 105–108. DOI: 10.1007/s40487-018-0063-9.

[3] Elizabeth B Claus et al. "Survival and low-grade glioma: the emergence of genetic information". In: *Neurosurgical focus* 38.1 (2015), E6. DOI: 10.3171/2014.10.FOCUS12367.

[4] Alberto Picca, Giulia Berzero, and Marc Sanson. "Current therapeutic approaches to diffuse grade II and III gliomas". In: *Therapeutic advances in neurological disorders* 11 (2018), p. 1756285617752039. DOI: 10.1177/1756285617752039.

[5] Slim Fellah et al. "Multimodal MR imaging (diffusion, perfusion, and spectroscopy): is it possible to distinguish oligodendroglial tumor grade and 1p/19q codeletion in the pretherapeutic diagnosis?" In: *American Journal of Neuroradiology* 34.7 (2013), pp. 1326–1333. DOI: 10.3174/ajnr.A3352.

[6] Nathalie L Jansen et al. "Prediction of oligodendroglial histology and LOH 1p/19q using dynamic [18F] FET-PET imaging in intracranial WHO grade II and III gliomas". In: *Neuro-oncology* 14.12 (2012), pp. 1473–1480. DOI: 10.1093/neuonc/nos259.

[7] Yasuo Iwadate et al. "Molecular imaging of 1p/19q deletion in oligodendroglial tumours with 11C-methionine positron emission tomography". In: *Journal of Neurology, Neurosurgery & Psychiatry* 87.9 (2016), pp. 1016–1021. DOI: 10.1136/jnnp-2015-311516.

[8] Pierre Bourdillon et al. "Prediction of anaplastic transformation in low-grade oligodendrogliomas based on magnetic resonance spectroscopy and 1p/19q codeletion status". In: *Journal of neuro-oncology* 122 (2015), pp. 529–537. DOI: 10.1007/s11060-015-1737-x.

[9] Adelheid Woehrer et al. “FISH-based detection of 1p 19q codeletion in oligodendroglial tumors: procedures and protocols for neuropathological practice-a publication under the auspices of the Research Committee of the European Confederation of Neuropathological Societies (Euro-CNS)”. In: *Clinical neuropathology* 30.2 (2011), pp. 47–55. DOI: 10.5414/npp30047.

[10] Hikaru Sasaki et al. “Histopathological-molecular genetic correlations in referral pathologist-diagnosed low-grade “oligodendroglioma””. In: *Journal of Neuropathology & Experimental Neurology* 61.1 (2002), pp. 58–63. DOI: 10.1093/jnen/61.1.58.

[11] Maximilian I Ruge et al. “Stereotactic brachytherapy with iodine-125 seeds for the treatment of inoperable low-grade gliomas in children: long-term outcome”. In: *Journal of clinical oncology* 29.31 (2011), pp. 4151–4159. DOI: 10.1200/JCO.2011.37.3381.

[12] Philippe Lambin et al. “Radiomics: extracting more information from medical images using advanced feature analysis”. In: *European journal of cancer* 48.4 (2012), pp. 441–446. DOI: 10.1016/j.ejca.2011.11.036.

[13] Hugo J W L Aerts et al. “Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach”. en. In: *Nat. Commun.* 5 (June 2014), p. 4006. DOI: 10.1038/ncomms5006.

[14] Philippe Lambin et al. “Radiomics: the bridge between medical imaging and personalized medicine”. In: *Nature reviews Clinical oncology* 14.12 (2017), pp. 749–762. DOI: 10.1038/nrclinonc.2017.141.

[15] Zeynettin Akkus et al. “Predicting deletion of chromosomal arms 1p/19q in low-grade gliomas from MR images using machine intelligence”. In: *Journal of digital imaging* 30 (2017), pp. 469–476. DOI: 10.1007/s10278-017-9984-3.

[16] Chia-Feng Lu et al. “Machine learning–based radiomics for molecular subtyping of gliomas”. In: *Clinical Cancer Research* 24.18 (2018), pp. 4429–4436. DOI: 10.1158/1078-0432.CCR-17-3445.

[17] Shun Zhang et al. “MRI based texture analysis to classify low grade gliomas into astrocytoma and 1p/19q codeleted oligodendroglioma”. In: *Magnetic resonance imaging* 57 (2019), pp. 254–258. DOI: 10.1016/j.mri.2018.11.008.

[18] Sebastian R van der Voort et al. “Predicting the 1p/19q codeletion status of presumed low-grade glioma with an externally validated machine learning algorithm”. In: *Clinical Cancer Research* 25.24 (2019), pp. 7455–7462. DOI: 10.1158/1078-0432.CCR-19-1127.

[19] Burak Kocak et al. “Radiogenomics of lower-grade gliomas: Machine learning–based MRI texture analysis for predicting 1p/19q codeletion status”. In: *European radiology* 30 (2020), pp. 877–886. DOI: 10.1007/s00330-019-06492-2.

[20] Ziren Kong et al. “Thin-slice magnetic resonance imaging-based radiomics signature predicts chromosomal 1p/19q co-deletion status in grade II and III gliomas”. In: *Frontiers in neurology* 11 (2020), p. 551771. DOI: 10.3389/fneur.2020.551771.

[21] Zeina A Shboul, James Chen, and Khan M. Iftekharuddin. “Prediction of molecular mutations in diffuse low-grade gliomas using MR imaging features”. In: *Scientific reports* 10.1 (2020), p. 3711. DOI: 10.1038/s41598-020-60550-0.

[22] Abhishta P Bhandari et al. “Noninvasive determination of IDH and 1p19q status of lower-grade gliomas using MRI radiomics: a systematic review”. In: *American Journal of Neuroradiology* 42.1 (2021), pp. 94–101. DOI: 10.3174/ajnr.A6875.

[23] Francesca Branzoli et al. “Cystathionine as a marker for 1p/19q codeleted gliomas by in vivo magnetic resonance spectroscopy”. In: *Neuro-oncology* 21.6 (2019), pp. 765–774. DOI: 10.1093/neuonc/noz031.

[24] Ruijiang Li et al. *Radiomics and radiogenomics: technical basis and clinical applications*. CRC press, 2019.

[25] Alex Zwanenburg et al. “The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping”. en. In: *Radiology* 295.2 (May 2020), pp. 328–338. DOI: 10.1148/radiol.2020191145.

[26] Bradley Erickson et al. “Data from LGG-1p19qDeletion”. In: *The Cancer Imaging Archive* 76.10.7937 (2017), K9. DOI: 10.7937/K9/TCIA.2017.DWEHTZ9V.

[27] Kenneth Clark et al. “The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository”. In: *Journal of digital imaging* 26 (2013), pp. 1045–1057. DOI: 10.1007/s10278-013-9622-7.

[28] Nancy Pedano et al. “Radiology data from the cancer genome atlas low grade glioma [TCGA-LGG] collection”. In: *The cancer imaging archive* 2 (2016). DOI: 10.7937/K9/TCIA.2016.L4LTD3TK.

[29] Mark A Hall. “Correlation-based feature subset selection for machine learning”. In: *Thesis submitted in partial fulfilment of the requirements of the degree of Doctor of Philosophy at the University of Waikato* (1988).

[30] M Hall et al. *The WEKA data mining software. ACM SIGKDD Explor Newsl 11: 10–18*. 2009.

[31] Ian H Witten et al. “Practical machine learning tools and techniques”. In: *Data Mining. Fourth Edition, Elsevier Publishers* (2017).

[32] Haibo He et al. "ADASYN: Adaptive synthetic sampling approach for imbalanced learning". In: *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*. Ieee. 2008, pp. 1322–1328. DOI: 10.1109/IJCNN.2008.4633969.

[33] Sebastian Sanduleanu et al. "Tracking tumor biology with radiomics: a systematic review utilizing a radiomics quality score". In: *Radiotherapy and Oncology* 127.3 (2018), pp. 349–360. DOI: 10.1016/j.radonc.2018.03.033.

[34] Karel GM Moons et al. "Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration". In: *Annals of internal medicine* 162.1 (2015), W1–W73. DOI: 10.7326/M14-0698.

[35] Alex J Mitchell. "The clinical significance of subjective memory complaints in the diagnosis of mild cognitive impairment and dementia: a meta-analysis". In: *International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences* 23.11 (2008), pp. 1191–1202. DOI: 10.1002/gps.2053.

[36] Michael Pentzek et al. "Apart from nihilism and stigma: what influences general practitioners' accuracy in identifying incident dementia?" In: *The American Journal of Geriatric Psychiatry* 17.11 (2009), pp. 965–975. DOI: 10.1097/JGP.0b013e3181b2075e.

[37] Alex J Mitchell. *How do we know when a screening test is clinically useful?* OUP Oxford, UK, 2009.

[38] Daniela C Gonçalves et al. "Case finding in dementia: comparative utility of three brief instruments in the memory clinic setting". In: *International Psychogeriatrics* 23.5 (2011), pp. 788–796. DOI: 0.1017/S1041610210002292.

[39] Alex J Mitchell. "Sensitivity$\times$ PPV is a recognized test called the clinical utility index (CUI+)". In: *European journal of epidemiology* 26 (2011), pp. 251–252. DOI: 10.1007/s10654-011-9561-x.

[40] R Rhys Davies and Andrew J Larner. "Addenbrooke's cognitive examination (ACE) and its revision (ACE-R)". In: *Cognitive Screening Instruments: A Practical Approach* (2013), pp. 61–77. DOI: 10.1007/s10072-015-2410-z.

3

Radiomics application in non-oncological neurology: overview and challenges

Elizaveta Lavrova, Henry C. Woodruff, Hamza Khan, Eric Salmon, Philippe Lambin*, Christophe Phillips*; "*" — equal contribution

---

Adapted from: Elizaveta Lavrova et al. "Handcrafted and deep radiomics in neurodegenerative diseases: a transition from oncology to clinical neurology". In: *Submitted to Clinical NeuroImage* (2023).

Abstract

Medical imaging technologies have undergone extensive development, enabling non-invasive visualization of clinical information. The traditional review of medical images by clinicians remains subjective, time-consuming, and prone to human error. With the recent availability of medical imaging data, quantification has become an important goal in the field. Radiomics, a methodology aimed at extracting quantitative information from imaging data, has emerged as a promising approach to uncover hidden biological information and support decision-making in clinical practice. This paper presents a review of the radiomic pipeline from the clinical neuroimaging perspective, providing a detailed overview of each step with practical advice. It discusses the application of handcrafted and deep radiomics in neuroimaging, stratified by neurological diagnosis. Although radiomics shows great potential for increasing diagnostic precision and improving treatment quality in neurology, several limitations hinder its clinical implementation. Addressing these challenges requires collaborative efforts, advancements in image harmonization methods, and the establishment of reproducible and standardized pipelines with transparent reporting. By overcoming these obstacles, radiomics can significantly impact clinical neurology and enhance patient care.

3.1 Introduction

Since the discovery of X-rays [1], medical imaging has advanced significantly. However, the conventional manual review of medical images by clinicians is subjective, time-consuming, and costly. With the increasing availability of medical imaging data, there is a growing opportunity for quantitative analysis in this field.
Radiomics is a methodology aimed at retrieving quantitative information from imaging data [2]. It is based on extracting numerous descriptors from medical images and finding a link between features and clinical outcomes. Handcrafted radiomics utilizes termed features, which are mathematically defined during the pipeline development, whereas deep radiomics utilizes features created by the artificial neural network during the model training process. The radiomics approach hypothesizes that medical imaging data contains hidden, complementary biological information that can be used for decision support in clinical practice [3]. Therefore, this method is of high interest for application in individualized diagnosis and treatment.
As handcrafted radiomics workflow requires a segmented region of interest (ROI), this methodology has been extensively developed in the oncological field where tumors and organs are routinely delineated for treatment planning purposes [2]. Since its inception, pioneer studies have revealed the connection between imaging biomarkers and histology [4] and have matured to produce externally validated and clinically relevant predictive models [5, 6].
Whereas in oncology a large amount of segmented imaging data is accumulated mostly for radiotherapy needs, other branches of medicine have collected imaging data and could potentially benefit from the application of radiomics. Thus, it is essential to perform an early diagnosis of neurological diseases since symptoms appear after the disease progresses considerably. Often there are no formal reliable biomarkers, and the diagnosis is based on the regularly reviewed diagnostic criteria [7]. Therefore, the differential diagnosis between the diseases and handling the atypical cases might be challenging.

Thus, radiomics is an emerging methodology in medical imaging research expanding from oncology to other branches of medicine. However, the review of radiomics in non-oncological neurology is needed to analyze the current state of the art, identify the methodological pitfalls, and suggest possible solutions for the future progress of quantitative clinical neuroimaging. In this review, we present a typical workflow of radiomics analysis regarding neuroimaging. We provide a broad overview of the currently published works stratified by neurological diagnosis. We discuss the current limitations of radiomics in neurology, suggesting potential improvements.

3.2 Workflow

The following section considers the practical implementation of radiomics in the neuroimaging field combining some common steps of radiomics and neuroimaging workflows (illustrated in Figure 3.1). After the steps are described, the list of the corresponding software is provided.

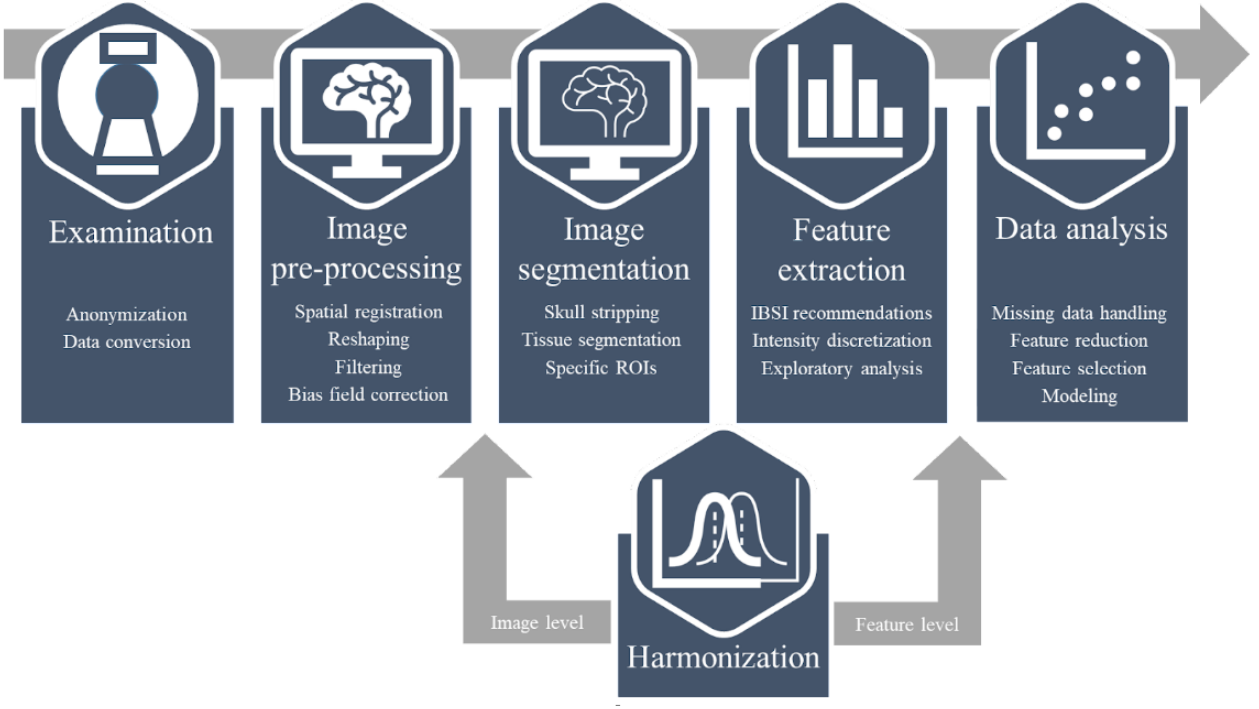


Figure 3.1: Radiomics pipeline for neuroimaging.

3.2.1 Data curation

In hospitals, the data is saved in Picture Archiving and Communications Systems (PACS) in Digital Imaging and Communications in Medicine (DICOM) format [8, 9]. It stores imaging data together with metadata. In research, open file formats are preferred, such as Nifti, Analyse, MNC, and NRRD [10].
To read and write the imaging and metadata, an application programming interface (API) for the currently relevant programming languages is recommended to make the pipeline fully automated and avoid manually introduced mistakes.
Clinical and metadata need to be anonymized or pseudo-anonymized, considering the possible need for follow-up acquisitions [11]. Brain scans usually include the facial features of the patient or teeth. Since facial features or teeth can be used to identify a person, it is necessary to remove them as well. A simple procedure for defacing is skull stripping [12].
A good practice is a sanity check of the data. It might include linking imaging and non-imaging samples to reveal missing or unwanted data, acquisition time point check for longitudinal studies and image quality check.
In neuroimaging research, the Nifti format is preferred as it is standardized and constructed specifically for neuroimaging data. In the case of data conversion to Nifti with a custom code, it is important to correctly transfer the geometrical parameters of the scan, as described in `https://nipy.org/nibabel/coordinate_systems.html`.
For automated data analysis, maintaining a uniform data structure is crucial. Data structure for different patients, imaging modalities, and potential acquisition timeframes should be established together with naming conventions for files and folders. For neuroimaging specifically, the Brain Imaging Data Structure (BIDS) [13] is recommended. It offers a standardized approach that is suited for multi-modal data and its derivatives and is supported by the community.

3.2.2 Data pre-processing

After the proper curation, data is considered ready for use. The next step is image pre-processing which is described in [14].
Since many imaging modalities or image acquisition time points can be combined in neurology, brain scan co-registration is needed. It means that multiple brain scans should be co-aligned to achieve the closest spatial position. The data to be co-registered can belong to different imaging modalities or sequences. Besides co-registration at the patient level, registration to the tissue probability maps in the standardized space can be performed [15]. Co-registration can be performed in both rigid (only the head position and orientation are changed) and non-rigid (additional scaling and elastic deformations) ways. Even though non-rigid co-registration allows for the best correspondence of the anatomy and regions of interest, it changes visualized tissue texture. Therefore, for clinical tasks, mostly rigid co-registration on a patient level is applied.
Image re-shaping is required to obtain the same voxel shape within the dataset. It allows for the same input image shape in the pipeline. While changing the voxel size, it is important to consider the interpolation effects introduced. In [16], different interpolation methods are described. The detailed recommendations are given in [14].
Brain scans contain intensity inhomogeneities due to the presence of the bias field. In MRI, the bias field is caused by the MR field inhomogeneity of the scanner originating from the equipment [17] and the patient disturbing the magnetic field. To reduce the effect of the bias field, bias field correction (BFC) can be applied. While performing BFC, it is important to consider that it might reduce the contrast and remove critical abnormality information. The most popular method is N4 BFC [18]. However, there are recent works on deep learning-based BFC [19]. CT is not affected by bias because it represents attenuation of the X-ray beam through the body, therefore, in general, this procedure is not recommended for CT scans.
Since CT images are expressed in HU, with a well-defined range, more advanced reliable pre-processing is possible. Knowing the characteris-

tic HU for the tissue of interest, it is possible to exclude all the objects on the scan that are not relevant to the analysis. Signal clipping can be applied to the intensities outside of the range of interest [20].
Even though "hard pre-processing" is not recommended in the quantitative image analysis to prevent a signal loss [14], some filtering can be applied to decrease the noise level. The most popular filters among smoothing filters are Gaussian and median filters. Gaussian filter is effective in removing high-frequency noises whereas a median filter is applied to remove impulse noise [21].

3.2.3 Image segmentation

In neurology, ROIs can be anatomically or physiologically derived and vary from application to application. Since brain structures have a systematic organization and traditional computer vision techniques can be applied for segmentation, there are many computer vision-based auto-segmentation tools recognized by the neuroimaging community. Nevertheless, the development of neural networks brings new solutions which are gaining more interest. The deep learning models are trained on different data and do not contain mathematical constraints about anatomy or the expected distribution of intensities.
Brain extraction narrows the image size and removes the surrounding tissues. In some studies, radiomics analysis was performed over the whole brain mask [22]. But since the brain includes different structures, whole-brain radiomics do not give comprehensive information about particular shapes and textures. Nevertheless, this kind of analysis is prospective in the discovery of healthy and pathological brain signatures for screening. In some works, features are extracted from the right and left hemispheres to be compared [23].
A lower level of defining the ROIs is presented with the nervous tissues. The human brain is composed of white and gray matter. Analysis of the radiomics features extracted from the separate brain tissues is closer to the in-vivo histology and allows for the

interpretation of texture and density abnormalities.
Many studies focus on analyzing certain areas of the cortex and deep gray matter. In this case, it is possible to build connections between imaging data and functional outcomes since particular cortical areas are responsible for the specific functions. Moreover, analyzing certain areas can lead to early disease diagnosis before more severe symptoms appear [24].
ROIs described above can contain both healthy and pathological tissues. In most cases, neurological pathology appears in textural changes. But when the pathological tissue is compact (hematoma or tissue lesion), it can be selected as ROI.
Physiologically-derived ROIs are obtained from functional imaging such as PET or fMRI. They represent the delineated tracer or function activity areas. No manual delineation is needed, and ROI contours depend on the selected binarization method.

3.2.4 Feature extraction

Radiomic features are usually divided into shape-, intensity-, and texture-based features [4]. The first category contains mathematical descriptors of ROI geometry, both 2D and 3D. The second category contains intensity statistics and histogram-derived descriptors. The third category contains descriptors of the spatial distribution of image intensity values and their mutual orientation. Shape-based features are not relevant in most neuroimaging studies since the shape of the analyzed ROIs is either standardized or complicated. Nevertheless, volume is always important since it can represent the dystrophy of brain structures, lesion load, or size of the affected area. Intensity- and texture-based features describe tissue properties including tissue homogeneity and density, therefore these groups of features are used in neurological studies.
Additional features to the ones obtained from the original image can be extracted when the same values are calculated from the transformed images. Image transformations may include, but are

not limited to, square, square root, logarithm, exponential, Gaussian, Laplacian, Laplace of Gaussian, wavelet, local binary pattern, and Gabor filters.
While performing feature extraction, besides image and ROI mask, feature extraction parameters are needed. Different feature extraction tools provide different levels of customization. IBSI has some recommendations on the most common feature extraction parameters. This includes intensities re-scaling with normalization or z-scoring. Intensity re-scaling for the images expressed in arbitrary units (MRI) is recommended but it is not for (semi) quantitative images, such as CTs. Another feature extraction parameter is intensity discretization before extracting texture features [25].

3.2.5 Data analysis

After the feature extraction step, data analysis as well as model development and validation are performed. For this, there are a large number of publications on good practices in AI [26, 27]. Therefore, in this section, we will focus on radiomics-specific steps.
Every case of missing data raises a question of feature or patient elimination or data imputation. In [28], a histological data imputation approach was suggested relying on the present features. Excluding patients will limit the population. Excluding features will limit the amount of diagnostic information. However, crucial clinical or demographic information missing should lead to record exclusion. Since the radiomic features are highly intercorrelated, in case of the absence of a radiomic feature, it can be both eliminated or imputed.
While developing a radiomics signature, it is important to assess feature stability and exclude non-reproducible features. To detect reproducible features, test-retest studies should be performed [29, 30]. Additionally, stability does not mean informativity, therefore, further steps on feature selection are needed.
After the data is split into train and test sets, the test set should be kept apart and used only in model evaluation so that they do not

interfere with the model-building process. In some cases, when the data size is not large or to show model robustness to data deviation, cross-validation is performed. This means that the data is split in one of the common cross-validation schemes multiple times and the whole training process is performed from scratch for every split. To show model generalizability, a good practice is to perform external validation —- to demonstrate model performance on external data coming from different acquisition equipment or hospitals.

To get rid of the redundant information in the data, inter-correlated features should be excluded preserving the information content. Additionally, it is necessary to exclude features with zero and low variance since they may contain little signal. To detect non-variant features, the standard deviation is usually calculated followed by scaling to the mean value. This approach gives unstable results if feature values have significantly different ranges and mean values. One of the popular methods is implemented as a nearZeroVar function of Caret package (`https://topepo.github.io/caret/`) in R.

The final step before modeling is feature selection to only retain the informative features. Usually, feature selection steps are model-based. Therefore, for different machine learning models, different features might be selected. The feature set should be reported together with the model performance: if the resulting model performs with low scores, we cannot conclude that the features are strongly linked to the outcome. Feature selection can be performed based on the univariate feature performance [31], feature weights in the model, or "recursive feature elimination" (RFE) based on recursively decreasing the feature set size and comparing model performances [32]. The number of the features in the final feature set can be estimated based on the sample size using several rules of thumb [33, 34], or empirically based on the saliency point of the dependency of the model score from the number of features.

Dimensionality reduction methods such as principal component analysis (PCA) [35], independent component analysis (ICA) [36], or linear discriminant analysis (LDA) [37] can also be used to decrease

the model complexity. This group of methods is based on the features decomposition resulting in the input matrix transformation into a lower-dimensionality matrix containing only distinctive information. However, it is not commonly used in radiomics since it transforms transparent radiomic features into abstract values losing the interpretability of the final method.
In deep radiomics, feature reduction, and selection procedures are performed by the neural network in a data-driven manner.
At this stage, overfitting as one of the major problems in machine learning should be considered. Comparison of the training and testing set scores enables assessment of an overfitting effect. A way to examine the model for overfitting is a permutation test [38]. To prevent overfitting, several techniques can be applied. These techniques are based on introducing random components into the data or the model and include data augmentation, regularization, ensembling, early stopping, or dropout layer for deep radiomics.

3.2.6 Harmonization

After the whole radiomics pipeline is established, the next steps are larger multi-center studies or clinical trials. One of the challenges here is feature instability caused by variations in population or acquisition equipment. It leads to situations where models are performing poorly on the data from the unseen domain. Harmonization is data alignment ensuring its compatibility and consistency. As it was shown in [39], harmonization can be performed at different steps of the radiomics pipeline, but globally in the image or feature domain.
In multi-center studies, data harmonization can start at the beginning of the study by standardizing image acquisition. Nevertheless, even when the protocols are standardized, the intensity distribution in scans can vary, especially while dealing with MRI data. To identify stable features, test-retest and phantom studies can be performed [30]. If the raw sensor data is available, scans can be reconstructed with the same parameters [40]. It is possible to implement traditional image

processing methods (intensity normalization, z-scoring) as well as deep-learning-based style transfer. Nevertheless, while changing the appearance of the scans and their intensity distribution, it is not clear whether it will improve feature analysis. In [41], the U-Net is trained to produce the MRI scans with consistent contrast. In [42], the generative adversarial network was trained to harmonize MRI scans aiming to preserve the consistency of feature values.
Finally, feature values can be harmonized. The most popular feature harmonization method is ComBat originally developed for harmonization of the gene expression data [43]. It is an empirical Bayesian method aimed at removing batch-specific bias and preserving the influence of biologically significant components. For the neuroimaging data, DeepCombat combining conditional variational autoencoder architecture with ComBat methodology is suggested [44].

3.2.7 Handcrafted vs. deep radiomics

Deep radiomics automatically learns representative image features from the high-dimensional data by using non-linear modules of the neural network [45]. Handcrafted radiomics represents a "hard-coded" version of deep radiomics. Whereas in handcrafted radiomics ROI, feature formulas, and mathematical models are defined by the user, in deep radiomics, these instances are learned from the data. Therefore, feature extraction and selection can be replaced with the neural network. However, a neural network can be a supplementary component for the handcrafted workflow to segment the ROI. In the most general case, the whole scan can be used as a neural network input, as illustrated in Figure 3.2. In this case, the model will learn to identify the informative features of the scan.

Both handcrafted and deep radiomics have advantages and limitations. Handcrafted radiomics can be trained with less data and is more transparent due to the interpretable features. However, it is limited in capturing complex patterns, not robust to variations

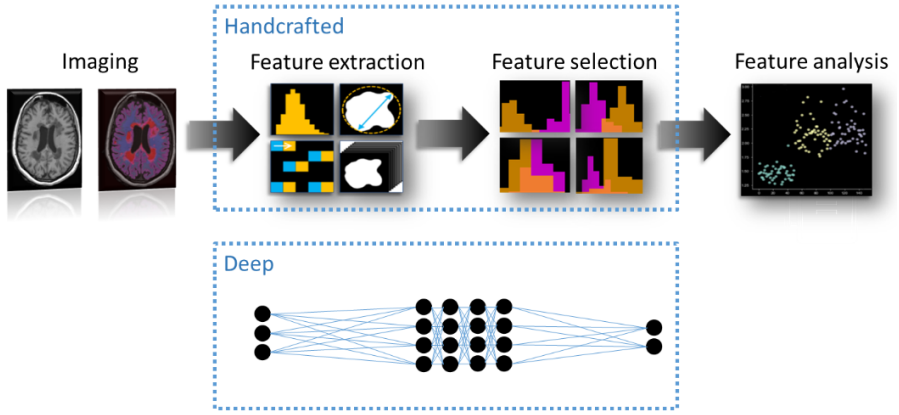


Figure 3.2: Radiomics pipeline: steps of the handcrafted radiomics represent the ”hard-coded” implementation of the deep radiomics.

of imaging parameters, and requires image preprocessing and segmentation. Deep radiomics learns relevant features automatically, captures more complex dependencies, and shows impressive results. It can adapt to the different imaging modalities and tasks with minimal changes in architecture. Moreover, additional domain knowledge can be utilized with transfer learning. However, deep radiomics is greedy for training data and computational resources and is challenging to interpret [46, 47].

Not all of the handcrafted features are necessarily linked to the outcome. In contrast, deep radiomics generates features during the training process. Recently attention in AI has shifted towards self-supervised learning and foundation models [48, 49]. Self-supervised learning aims to provide pseudo-labels to the data by deriving supervisory signals from the data. After pre-training, models can be fine-tuned for specific downstream tasks. This approach allows for training with a substantial amount of unlabeled data and enables utilization of the same pseudo-labels for multiple downstream tasks creating a new concept of imaging features.

3.2.8 Neuroimaging software packages

To sum up, the most acknowledged open-source neuroimaging software packages are presented in Table 3.1. At the moment, the most common language for open-source research software is Python. Table 3.2 presents Python packages for neuroimaging.

Table 3.1: Neuroimaging software.

Name	Link	Platform/environment	Functionality
3D Slicer [50]	https://www.slicer.org/	Desktop, API	Images visualization, processing, segmentation, registration, and analysis; planning and navigating image-guided procedures
ITK-Snap [51]	http://www.itksnap.org/	Desktop, API	Image segmentation
RadiAnt	https://www.radiantviewer.com/	Desktop	Viewer
MicroDicom	https://www.microdicom.com/	Desktop	Viewer
SPM12 [52]	https://www.fil.ion.ucl.ac.uk/spm/software/spm12/	MATLAB	Complete analysis package
Dcm2nii	https://www.nitrc.org/projects/dcm2nii/	Desktop, CLI	DICOM to Nifti
Dcm2niix	https://github.com/rordenlab/dcm2niix	CLI	DICOM to Nifti conversion
Dcm2niir	https://github.com/muschellij2/dcm2niir	R	DICOM to Nifti conversion
FSL [53]	https://fsl.fmrib.ox.ac.uk/fsl/fslwiki	Desktop, CLI, API	Complete analysis package
Freesurfer [54]	https://surfer.nmr.mgh.harvard.edu/	Desktop, CLI, API	Complete analysis package
ANTs [55]	https://github.com/ANTsX/ANTs	Unix scripting	Complete analysis package (dependent on fMRI packages)
Brainsuit [56]	https://brainsuite.org/	MATLAB	Complete analysis package
LST [57]	https://www.applied-statistics.de/lst.html	MATLAB + SPM	WM lesion segmentation
Camino [58]	http://camino.cs.ucl.ac.uk&CLI	Diffusion MRI processing	
QuickNAT [59]	https://github.com/ai-med/QuickNATv2	MATLAB	Neuroanatomy segmentation
AFNI [60]	https://afni.nimh.nih.gov/	Desktop, CLI, API	Complete analysis package
CAT [61]	https://neuro-jena.github.io/cat//	MATLAB + SPM	Computational anatomy analysis

Table 3.2: Neuroimaging Python packages.

Name	Link	Functionality
Pydicom [62]	https://pydicom.github.io/	Reading, modifying and writing DICOM data
Medpy [63]	https://github.com/loli/medpy	Medical image manipulation
Nibabel [64]	https://nipy.org/nibabel/	Reading, modifying and writing common neuroimaging data formats
Pydeface [65]	https://github.com/poldracklab/pydeface	Defacing
Mridefacer	https://github.com/mih/mridefacer	Defacing
Mriqc [66]	https://github.com/nipreps/mriqc	Image quality assessment
Miqa	https://github.com/OpenImaging/miqa	Image quality assessment
ImageQC	https://github.com/EllenWasbo/ImageQC	Image quality assessment
Precision medicine toolbox [67]	https://github.com/primakov/precision-medicine-toolbox	Image format conversion, metadata collection, basic quality check and pre-processing, basic exploratory analysis of the tabular data
Pybids [68]	https://github.com/bids-standard/pybids	BIDS data management
Deepbrain	https://github.com/iitzco/deepbrain	Brain extraction
DeepBleed [69]	https://github.com/msharrock/deepbleed	Haemorrhage segmentation
Pyradiomics [70]	https://github.com/AIM-Harvard/pyradiomics	Radiomic features extraction
Nipype	https://github.com/nipy/nipype	Pipeline implementation
TorchiO [71]	https://torchio.readthedocs.io/	Medical image augmentation and transformations including noise and artefacts generation
HippoDeep [72]	https://github.com/bthyreau/hippodeep	Hippocampus segmentation

3.3 Applications

The number of papers in non-oncological neurology is growing starting from the first publication in 2017 [9], which explicitly mentions radiomics. Figure 3.3 shows the number of papers per year found in the PubMed database with the search query "radiomics AND neurology NOT oncology". In this section, we will review the non-oncological neurological studies utilizing radiomics.

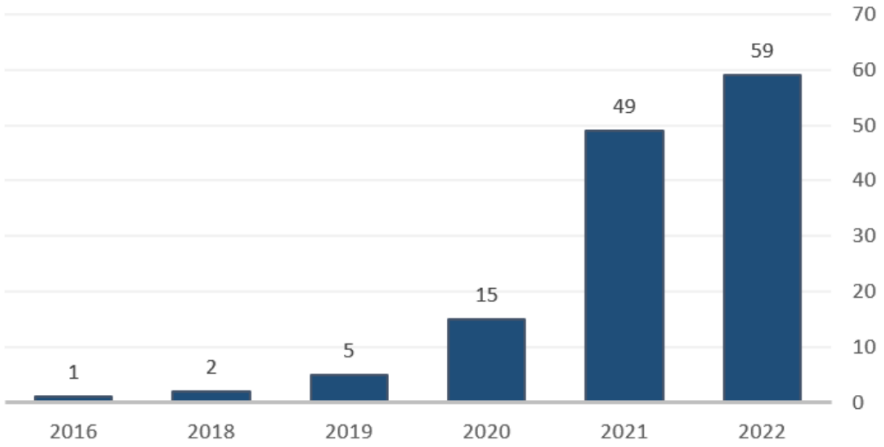


Figure 3.3: Number of publications, by year, containing keywords "radiomics AND neurology NOT oncology" in PubMed database (access date: 22.09.2023).

3.3.1 Alzheimer's disease

Alzheimer's disease (AD) is the leading cause of dementia worldwide, and dementia is the leading cause of disability among the elderly population [73]. Since research is extensively performed in this field, and much experience and data have been accumulated, the first neurological radiomic studies were performed in this area.
As the disease develops gradually, causing different levels of

disability, the first work [74] classified disease stages with deep radiomics using data from Alzheimer's Disease Neuroimaging Initiative (ADNI) database [75]. This first study pointed out the need for more data and for establishing the connection between the radiomic features and pathological processes. There are some later works on AD stages classification as well [76]. Many later works performed binary classification between AD patients and normal controls (NC) to show that MRI- and PET-derived features can be associated with AD [77, 78, 79, 80, 81, 82, 83, 84, 85]. Whereas in AD, pathological changes might be well visible in scans and be accompanied by strong clinical symptoms, mild cognitive impairment (MCI) is harder to diagnose. Moreover, MCI individuals can be confused with AD patients and NC. There are works on distinguishing between NC and MCI patients [86, 87]. Most of the AD radiomics studies are performed on classification between NC, MCI, and AD patients utilizing either pair-wise binary or multi-class classification [88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99]. As MCI is considered an early stage of AD, for these patients it is crucial to know whether their impairment will progress to AD. In [100], the classification model is trained to estimate amyloid positivity status in MCI patients. In several works, classification models are built to directly predict the conversion from MCI to AD [101, 102, 103, 104, 105, 106, 107]. There were also attempts to predict the speed of disease progression [108]. Nevertheless, AD is not the only cause of dementia, and focusing on AD patients only will lead to the lack of specificity of the methods. There are works on the classification of the different dementia diagnoses: dementia with Lewy bodies vs AD [108, 109] and idiopathic normal pressure hydrocephalus vs. AD [110].

Nevertheless, besides clinical radiomics, there are traditional neuroimaging features to characterize the brain. Functional connectivity plays an important role here. Even though several studies relied on fMRI [111, 87, 76, 105, 83], the connection between these biomarkers has yet to be revealed. In [111], the correlation between connectivity and radiomic features is studied.

3.3.2 Multiple sclerosis

Multiple sclerosis (MS) is the leading cause of disability among the young population [112]. The disease progression is fast, therefore early diagnosis and prognosis for the patient are important.
However, the first deep radiomic study on inflammatory degeneration was performed on survival classification of the amyotrophic lateral sclerosis patients [113]. MS-related studies were performed later and covered simple binary classification between MS and NC [114, 115, 116]. These works show the utility of the approach but lack the specificity of MS among other inflammatory neurodegenerative diseases. Neuromyelitis optica-spectrum disorder (NOSD) can be easily confused with MS therefore a precise diagnosis is needed to enable the correct treatment. There are some works on binary classification between MS and NOSD [117, 118, 119, 120]. There are studies where the classification models are built to distinguish between MS and other diagnoses such as neuropsychiatric systemic lupus erythematosus [121] and ischemic vasculopathy [122]. The multi-class classifier was built to distinguish between MS, NOSD, migraine, and vasculitis [123]. Besides diagnosis, it is essential to grade the disease severity within the MS cohort. In [124], the radiomics model was trained to estimate the relapse rate in MS. In [125], the EDSS score was predicted with deep radiomics. In [126], MS types were classified as well as NC.
Several models are trained to characterize the neurodegeneration process. In [127], the deep learning model is built to perform the detection of the demyelinated voxels on PET. In [128], MS lesion classification is performed. In [129, 130], lesion rim status classification is performed. In [131], a handcrafted radiomic model was built to predict lesion growth.

3.3.3 Parkinson's disease

Parkinson's disease (PD) is a degenerative condition of the central nervous system, which develops disability faster than any other neurological disorder [132]. Since it develops gradually, exploration

of the early and specific biomarkers is essential.
The first work was on the motor function assessment in PD [9]. To show the utility of radiomics, the models were built to distinguish between PD patients and NC [133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144]. As PD is heterogeneous in terms of its clinical phenotype, for treatment and severity estimation, PD variants classification is important. The models to classify the patients between parkinsonism subtypes were built in some works [145, 146, 147, 148, 149, 150, 151, 152]. In PD there is a need for differentiation from other similar diseases. In [153, 154], binary classification models were trained to distinguish between PD and multiple system atrophy and progressive supranuclear palsy. Radiomics approach was also used for the treatment response prediction [155].
Transfer learning attempts are performed in the PD field trying to fine-tune the AD diagnostic model for PD diagnosis [156]. Diagnostic support solutions are suggested as in [155], where the nigrosome 1 abnormalities detector is trained.

3.3.4 Stroke

Stroke is the second leading cause of death, and third leading cause of disability worldwide [157]. The main task in stroke management is patient outcome prediction.
As a first task in stroke management, stroke areas should be identified. In [158], the deep learning model was trained to detect the stroke area on non-contrast CT scans. In [159], the model was trained to distinguish between hyperperfusion areas from normal ones. In [160], primary and secondary hemorrhages were classified. The other works were devoted to the prognosis of stroke area development [161, 162, 163, 164]. Besides the pathology development prediction, there were studies focused on the prediction of the biological recovery processes. In [165], collateral circulation was classified. In [166, 167], models were trained to predict recanalization. Most works were focused on the treatment outcome prediction, including functional and cognitive,

for both thrombolysis and mechanical thrombectomy [168, 169, 170, 171, 172, 173]. There are studies on stroke onset time estimation with radiomics [174]. Finally, some studies are performed to train the radiomic models to predict post-stroke events such as recurrent stroke or epilepsy [175, 176, 177].

3.3.5 Epilepsy

Epilepsy is the most common neurological disease [178]. In [179], the model was trained to predict epilepsy laterality. In [180], the epileptic foci detection method was suggested. In [181], the radiomic model was trained to detect focal cortical dysplasia lesions. In [182], the binary classifier was built to distinguish between Juvenile Myoclonic Epilepsy and NC.

3.3.6 Mental disorders

Mental disorders affect behavior and quality of life significantly and are observed in 12.5% of the population [183]. Most studies in radiomics in mental disorders are devoted to the binary classification between the NC and schizophrenia patients [184, 185, 186, 187, 188, 189, 190, 191, 192], bipolar disorder patients [193], or first episode psychosis [194, 195, 196]. In [195], a more difficult deep radiomics classifier was built to distinguish between schizophrenia patients, major depressive disorder, and NC. In [196], the classifier was built to distinguish between first-episode psychosis, bipolar disorder, and NC. Finally, there are some works on radiomics implementation for treatment response prediction [197, 198].

3.3.7 Neurodevelopmental disorders

The most common neurodevelopmental disorders are autism spectrum disorder (ASD) and attention-deficit/hyperactivity disorder (ADHD). They affect cognitive and behavioral functions and might

Table 3.3: Some of the open-source neuroimaging datasets.

Name	Modality	Cohort
ADNI [75]	MRI, PET	AD
OASIS [203]	MRI, PET, CT	NC, AD
ATLAS [204]	T1w MRI	Stroke
IXI Dataset [205]	T1w, T2w, PDw, MRA, dMRI	NC
Yale Test-Retest Data [206]	Anatomical and functional MRI	NC
ISLES [207]	DWI, ADC and FLAIR MRI, CT perfusion	Stroke
MS Lesion Segmentation Challenge 2015 [208]	T1w, T2w, FLAIR, PDw MRI	MS
NFBS skull stripped repository [209]	T1w MRI	Psychiatric symptoms
Calgary-Campinas-359 [210]	T1w MRI	NC
CT-ICH [211]	CT	TBI
RSNA Intracranial Hemorrhage [212]	CT	Stroke
PPMI [213]	Clinical and biological data	PD

require life-long care and support. In most radiomic studies, binary classification models were built to distinguish between NC and ASD [199], and ADHD [200, 201]. In [202], ASD-linked radiomic features were revealed.

3.3.8 Open-source datasets

We believe that to enable extensive research in some diagnostic areas it is essential to have access to a sufficient amount of medical imaging data. Table 3.3 presents some popular open medical imaging datasets containing neuroimaging data.

3.4 Discussion

In this review, we gave an overview of the radiomics pipeline in clinical non-oncological neuroimaging and gave some recommendations

on each step of the pipeline. The variety of open-source tools for neuroimaging analysis as well as the expanding amount of radiomics studies are bringing optimism to the development of artificial intelligence (AI) in clinical neurology. Moreover, the amount of accumulated data is still growing, pushing the quantitative neuroimaging development forward. Therefore, the approaches, that previously existed in oncology, can be landed in neurology. However, there are some limitations present in the majority of the papers and characterizing the current challenges in the field.

3.4.1 Data availability

Many current studies are cross-sectional and performed on small private datasets. These datasets usually are not sufficient to demonstrate pathological patterns and represent the target cohort. One-center training severely decreases the generalizability of the model. To enable clinical implementation of the model, it needs to confirm its performance on various demographics as well as acquisition equipment set-ups.
External validation on samples coming from the different data domains is recommended. It is desirable to have external validation data to be prospectively collected to show robustness to the potential data drift. However, external validation results should be interpreted carefully. Sample size, heterogeneity, and data balance should be considered. External validation performance should be explainable considering the sample properties. The external validation does not give absolute information about model generalizability. There are some industry-inspired suggestions to perform regular and recurrent validation every time the model is deployed to evaluate the generalizability of the predictive model [214, 215].
Another consequence of the limited data accessibility is the low reproducibility of the published studies. Experiments on private data cannot be repeated. Additionally, it is not possible to check the labeling correctness. Moreover, if one wants to compare some models, he needs to test them on the same cohort. Nevertheless, there are

established datasets used by multiple research groups. These are 1) open-source datasets, such as ADNI, 2) challenge datasets (`https://grand-challenge.org/`) which are much smaller and usually do not have extensive multi-modal data, 3) clinical trials datasets (for example, [216]). However, using a single dataset without any external data introduces overfitting across the community, which limits the usefulness of the dataset itself. To make the data and data-driven solutions sustainable and therefore trustful, the following four principles have to be maintained: Findability, Accessibility, Interoperability, and Reusability (FAIR) [217].
Lately, the attention is drawn to the latest generation of models – foundation models. They have a huge potential in medicine [49, 218]. However, the main challenge in their implementation is the high demand for the amount of the training data. This fact gives an additional motivation for data sharing and aggregation.

3.4.2 Data harmonization

Another limitation resulting from limited data availability is the lack of data harmonization. It is affecting the radiomic methods themselves because of the heterogeneity of the population as well as acquisition equipment. If the development data is not population and equipment representative, every new inference data point can be out of the distribution which leads to the wrong model outcomes. The CT data is presented in HU which gives it quantification and stability. MRI data is expressed in arbitrary units and acquired with a large variety of MR sequences and hardware. One of the promising directions in the MRI research is in multi-echo qMRI maps reconstruction. It gives stability and quantification to the MRI data, but at the moment this approach is far from being used in the clinical set-up.

3.4.3 Clinical relevance of the data

Reliable and stable imaging biomarkers for neurological disorders should be not only sensitive but also specific for every neurological condition. In the current studies, most of the models are developed to distinguish between the disorder and NC. Therefore, these approaches are not applicable in clinical practice, where more than one neurological condition can be suspected. Moreover, co-existing conditions are also possible. Therefore, broader studies and intra-diagnosis tests are needed to develop disease-specific radiomic signatures.

Most neurological diseases develop gradually, and the patients are diagnosed in already chronic stages of the disease. This brings bias to the study data since it almost does not contain non-symptomatic patients at early stages. These patients are highly important in developing methods for early diagnosis and disease prevention. However, there are some longitudinal studies (for example, ADNI-based) that are performed with the early-stage cases.

3.4.4 Study design

Current studies are mostly proof-of-concept. Therefore, the study design is highly simplified. Prediction tasks are solved as classification tasks in most cases and the outcomes are limited by the predicted event timeframe.

Imaging modality should be selected correctly, based on domain knowledge, existing clinical protocols, and its availability.

The fusion of data of different natures should not be performed before the predictive power of every baseline data source is studied. However, data fusion can be justified by showing the added value in model performance obtained with the fused data compared to the separate baseline models.

For handcrafted radiomics, ROI selection should be based on domain knowledge. However, for the radiomic studies of different clinical

applications, ROIs are selected based on the brain areas affected the most by the corresponding diseases.

3.4.5 Pipeline implementation

Different research groups perform radiomic pipeline steps differently and in different order. This results in inconsistency and low reproducibility of the results. To overcome this issue, transparent reporting is essential following TRIPOD [219] and RQS [220]. Recently, the CheckList for EvaluAtion of Radiomics Research (CLEAR) checklist for radiomics was out [221].
Additionally, code sharing has become a common practice in scientific reporting in the last few years making the results transparent and reproducible.
Another challenge is caused by the fact that multiple data processing tools are implemented in different platforms and environments breaking the consistency of the data flow. However, for every common neuroimaging tool multiple APIs exist enabling a single infrastructure for the study implementation.
To justify the selection of the model, its design, and hyperparameters in a reproducible environment, experiment tracking tools such as MLFlow (`www.mlflow.org`) are useful. They do not only inform the researcher about the best-performing setup but also the protocol of all the experiments.
Another implementation challenge is related to image segmentation which is traditionally performed manually. Since intra- and inter-reader agreement is never absolute, development and improvement of the automated segmentation methods is needed.
For the clinical application of AI in medical imaging, FUTURE-AI guiding principles are developed [222] to ensure that AI solutions are effective, trustworthy, ethical, and safe.

3.4.6 Interpretation

Even though the reported models perform with high scores, there is still a lack of interpretation. For this, behavioral analysis should be performed and aligned with the clinical knowledge. Connections between the predictive and stable radiomic features and clinical parameters should be studied. Additionally, pathological mechanisms are not revealed by the radiomic studies, and large work should be done in this field supported by extensive clinical and histological data.
Since medical imaging analysis involves high-stakes decisions, information is needed about which influence inputs have on a final decision of the model. Due to a simple implementation, handcrafted radiomics is more transparent compared to deep radiomics. However, for acceptance in clinical practice, implementation of explainable AI (XAI) is needed for overcoming a "black box problem" [223, 224].
While every AI model is accompanied by its performance scores, which provide insights into its efficiency and facilitate comparisons with other models [225], it is imperative to remember that the significance lies not in the AI scores themselves but in the impact on clinical outcomes. Consequently, for more advanced models, the inclusion of supplementary metrics is vital to elucidate how they enhance the clinical pipeline.

3.5 Conclusion

We gave a review of the radiomic pipeline from the clinical neuroimaging perspective. The amount of the collected data and the high performance of the published models have shown that the application of radiomics in neuroimaging will increase diagnostic precision and quality of treatment. However, there are some important limitations preventing the implementation of this methodology in clinical practice. To overcome these limitations, it is necessary to set data exchange and

collaborations, work on data harmonization methods, and implement reproducible pipelines and transparent reporting.

References

[1] Eyal Bercovich and Marcia C Javitt. "Medical Imaging: From Roentgen to the Digital Revolution, and Beyond". en. In: *Rambam Maimonides Med J* 9.4 (Oct. 2018). DOI: 10.5041/RMMJ.10355.

[2] Philippe Lambin et al. "Radiomics: extracting more information from medical images using advanced feature analysis". In: *European journal of cancer* 48.4 (2012), pp. 441–446. DOI: 10.1016/j.ejca.2011.11.036.

[3] Julien Guiot et al. "A review in radiomics: Making personalized medicine a reality via routine imaging". en. In: *Med. Res. Rev.* 42.1 (Jan. 2022), pp. 426–440. DOI: 10.1002/med.21846.

[4] Hugo J W L Aerts et al. "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach". en. In: *Nat. Commun.* 5 (June 2014), p. 4006. DOI: 10.1038/ncomms5006.

[5] Wyanne A Noortman et al. "Development and External Validation of a PET Radiomic Model for Prognostication of Head and Neck Cancer". en. In: *Cancers* 15.10 (May 2023). DOI: 10.3390/cancers15102681.

[6] Philipp Guevorguian et al. "External validation of a CT-based radiomics signature in oropharyngeal cancer: Assessing sources of variation". en. In: *Radiother. Oncol.* 178 (Jan. 2023), p. 109434. DOI: 10.1016/j.radonc.2022.11.023.

[7] Vivek Kumar Sharma et al. "Biomarkers: Role and Scope in Neurological Disorders". en. In: *Neurochem. Res.* 48.7 (Feb. 2023), pp. 2029–2058. DOI: 10.1007/s11064-023-03873-4.

[8] Charles Parisot. "The DICOM standard". en. In: *Int. J. Card. Imaging* 11.3 (Sept. 1995), pp. 171–177. DOI: 10.1007/BF01143137.

[9] Arman Rahmim et al. "Improved prediction of outcome in Parkinson's disease using radiomics analysis of longitudinal DAT SPECT images". en. In: *Neuroimage Clin* 16 (Aug. 2017), pp. 539–544. DOI: 10.1016/j.nicl.2017.08.021.

[10] Xiangrui Li et al. "The first step for neuroimaging data analysis: DICOM to NIfTI conversion". In: *Journal of neuroscience methods* 264 (2016), pp. 47–56. DOI: 10.1016/j.jneumeth.2016.03.001.

[11] Khaled El Emam, Sam Rodgers, and Bradley Malin. "Anonymising and sharing individual patient data". en. In: *BMJ* 350 (2015). DOI: 10.1136/bmj.h1139.

[12] Athena E Theyers et al. "Multisite Comparison of MRI Defacing Software Across Multiple Cohorts". en. In: *Front. Psychiatry* 12 (Feb. 2021), p. 617997. DOI: 10.3389/fpsyt.2021.617997.

[13] Krzysztof Gorgolewski et al. "Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python". In: *Frontiers in neuroinformatics* (2011), p. 13. DOI: 10.3389/fninf.2011.00013.

[14] Alex Zwanenburg et al. "The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping". en. In: *Radiology* 295.2 (May 2020), pp. 328–338. DOI: 10.1148/radiol.2020191145.

[15] David Alexander Dickie et al. "Whole Brain Magnetic Resonance Image Atlases: A Systematic Review of Existing Atlases and Caveats for Use in Population Imaging". en. In: *Front. Neuroinform.* 11 (Jan. 2017), p. 226915. DOI: 10.3389/fninf.2017.00001.

[16] Thomas Martin Lehmann, Claudia Gonner, and Klaus Spitzer. "Survey: Interpolation methods in medical image processing". In: *IEEE transactions on medical imaging* 18.11 (1999), pp. 1049–1075. DOI: 10.1109/42.816070.

[17] P Padilla, J F Valenzuela-Valdes, and J L Padilla. "Inhomogeneity reduction for near field acquisition in high resolution MRI systems". In: *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*. Valencia, Spain: IEEE, June 2017, pp. 1513–1516. DOI: 10.1109/ACCESS.2017.2685079.

[18] Nicholas J Tustison et al. "N4ITK: improved N3 bias correction". In: *IEEE transactions on medical imaging* 29.6 (2010), pp. 1310–1320. DOI: 10.1109/TMI.2010.2046908.

[19] Kai-Hsiang Chuang et al. "Deep learning network for integrated coil inhomogeneity correction and brain extraction of mixed MRI data". en. In: *Sci. Rep.* 12.1 (May 2022), pp. 1–14. DOI: 10.1038/s41598-022-12587-6.

[20] Pankaj Kandhway and Ashish Kumar Bhandari. "Modified clipping based image enhancement scheme using difference of histogram bins". en. In: *IET Image Proc.* 13.10 (Aug. 2019), pp. 1658–1670. DOI: 10.1049/iet-ipr.2019.0111.

[21] Arvind Kumar and Sartaj Singh Sodhi. "Comparative analysis of gaussian filter, median filter and denoise autoenocoder". In: *2020 7th International Conference on Computing for Sustainable Global Development (INDIACom)*. IEEE. 2020, pp. 45–51. DOI: 10.23919/INDIACom49435.2020.9083712.

[22] Yingping Li et al. "Radiomics-based method for predicting the glioma subtype as defined by tumor grade, IDH mutation, and 1p/19q codeletion". In: *Cancers* 14.7 (2022), p. 1778. DOI: 10.3390/cancers14071778.

[23] Qingguo Ren et al. "A Pilot Study of Radiomic Based on Routine CT Reflecting Difference of Cerebral Hemispheric Perfusion". en. In: *Front. Neurosci.* 16 (Mar. 2022), p. 851720. DOI: 10.3389/fnins.2022.851720.

[24] Y Lakshmisha Rao et al. "Hippocampus and its involvement in Alzheimer's disease: a review". en. In: *3 Biotech* 12.2 (Feb. 2022), p. 55. DOI: 10.1007/s13205-022-03123-4.

[25] Loic Duron et al. "Gray-level discretization impacts reproducible MRI radiomics texture features". In: *PLoS One* 14.3 (Mar. 2019), e0213459. DOI: 10.1371/journal.pone.0213459.

[26] Sujay Kakarmath et al. "Best practices for authors of healthcare-related artificial intelligence manuscripts". en. In: *npj Digital Medicine* 3.1 (Oct. 2020), pp. 1–3. DOI: 10.1038/s41746-020-00336-w.

[27] Gaël Varoquaux and Veronika Cheplygina. "Machine learning for medical imaging: methodological failures and recommendations for the future". In: *NPJ digital medicine* 5.1 (2022), p. 48. DOI: 10.1038/s41746-022-00592-y.

[28] Ming Fan et al. "A deep matrix completion method for imputing missing histological data in breast cancer by integrating DCE-MRI radiomics". en. In: *Med. Phys.* 48.12 (Dec. 2021), pp. 7685–7697. DOI: 10.1002/mp.15316.

[29] Janna E van Timmeren et al. "Test–Retest Data for Radiomics Feature Stability Analysis: Generalizable or Study-Specific?" en. In: *Tomography* 2.4 (Dec. 2016), pp. 361–365. DOI: 10.18383/j.tom.2016.00208.

[30] A K Jha et al. "Repeatability and reproducibility study of radiomic features on a phantom and human cohort". en. In: *Sci. Rep.* 11.1 (Jan. 2021), pp. 1–12. DOI: 10.1038/s41598-021-81526-8.

[31] Chintan Parmar et al. “Machine Learning methods for Quantitative Radiomic Biomarkers”. en. In: *Sci. Rep.* 5.1 (Aug. 2015), pp. 1–11. DOI: 10.1038/srep13087.

[32] Isabelle Guyon et al. “Gene selection for cancer classification using support vector machines”. In: *Machine learning* 46 (2002), pp. 389–422. DOI: 0.1023/A:1012487302797.

[33] Jianping Hua et al. “Optimal number of features as a function of sample size for various classification rules”. In: *Bioinformatics* 21.8 (2005), pp. 1509–1515. DOI: 10.1093/bioinformatics/bti171.

[34] Yaser S Abu-Mostafa, Malik Magdon-Ismail, and Hsuan-Tien Lin. *Learning from Data: A Short Course*. en. AMLBook New York, Jan. 2012.

[35] Deleu Anne-Leen et al. “Principal component analysis of texture features derived from FDG PET images of melanoma lesions”. en. In: *EJNMMI Physics* 9.1 (Sept. 2022), pp. 1–10. DOI: 10.1186/s40658-022-00491-x.

[36] Sheng Zhang and Chiang-Shan R Li. “Functional connectivity parcellation of the human thalamus by independent component analysis”. In: *Brain Connectivity* 7.9 (2017), pp. 602–616. DOI: 10.1089/brain.2017.0500.

[37] Yimeng Fan et al. “Radiomics-Based Machine Learning Technology Enables Better Differentiation Between Glioblastoma and Anaplastic Oligodendroglioma”. en. In: *Front. Oncol.* 9 (Nov. 2019), p. 478633. DOI: 10.3389/fonc.2019.01164.

[38] Bárbara M de Andrade et al. “Comparison of the performance of multiclass classifiers in chemical data: Addressing the problem of overfitting with the permutation test”. In: *Chemometrics and Intelligent Laboratory Systems* 201 (2020), p. 104013. DOI: 10.1016/j.chemolab.2020.104013.

[39] Shruti Atul Mali et al. “Making Radiomics More Reproducible across Scanner and Imaging Protocol Variations: A Review of Harmonization Methods”. en. In: *J Pers Med* 11.9 (Aug. 2021). DOI: 10.3390/jpm11090842.

[40] Leticia Gallardo-Estrella et al. “Normalizing computed tomography data reconstructed with different filter kernels: effect on emphysema quantification”. en. In: *Eur. Radiol.* 26.2 (Feb. 2016), pp. 478–486. DOI: 10.1007/s00330-015-3824-y.

[41] Blake E Dewey et al. “DeepHarmony: A deep learning approach to contrast harmonization across scanner changes”. In: *Magnetic resonance imaging* 64 (2019), pp. 160–170. DOI: 10.1016/j.mri.2019.05.041.

[42] Md Selim et al. “CT image harmonization for enhancing radiomics studies”. In: *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE. 2021, pp. 1057–1062. DOI: 10.1109/BIBM52615.2021.9669448.

[43] Fanny Orlhac et al. “A Guide to ComBat Harmonization of Imaging Biomarkers in Multicenter Studies”. en. In: *J. Nucl. Med.* 63.2 (Feb. 2022), pp. 172–179. DOI: 10.2967/jnumed.121.262464.

[44] Fengling Hu et al. “DeepComBat: A Statistically Motivated, Hyperparameter-Robust, Deep Learning Approach to Harmonization of Neuroimaging Data”. en. Apr. 2023. DOI: 10.1101/2023.04.24.537396.

[45] Jürgen Schmidhuber. “Deep learning in neural networks: an overview”. en. In: *Neural Netw.* 61 (Jan. 2015), pp. 85–117. DOI: 10.1016/j.neunet.2014.09.003.

[46] William Rogers et al. “Radiomics: from qualitative to quantitative imaging”. In: *The British journal of radiology* 93.1108 (2020), p. 20190948. DOI: 10.1259/bjr.20190948.

[47] Matthias W Wagner et al. “Radiomics, machine learning, and artificial intelligence-what the neuroradiologist needs to know”. en. In: *Neuroradiology* 63.12 (Dec. 2021), pp. 1957–1967. DOI: 10.1007/s00234-021-02813-9.

[48] Rayan Krishnan, Pranav Rajpurkar, and Eric J Topol. “Self-supervised learning in medicine and healthcare”. en. In: *Nature Biomedical Engineering* 6.12 (Aug. 2022), pp. 1346–1352. DOI: 10.1038/s41551-022-00914-1.

[49] Michael Moor et al. “Foundation models for generalist medical artificial intelligence”. In: *Nature* 616.7956 (2023), pp. 259–265. DOI: 10.1038/s41586-023-05881-4.

[50] Ron Kikinis, Steve D Pieper, and Kirby G Vosburgh. “3D Slicer: A Platform for Subject-Specific Image Analysis, Visualization, and Clinical Support”. en. In: *Intraoperative Imaging and Image-Guided Therapy* (2014), pp. 277–289. DOI: 10.1007/978-1-4614-7657-3_19.

[51] Matthew McCormick et al. “ITK: enabling reproducible research and open science”. In: *Frontiers in neuroinformatics* 8 (2014), p. 13. DOI: 10.3389/fninf.2014.00013.

[52] William D Penny et al. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. en. Elsevier, Apr. 2011.

[53] Mark Jenkinson et al. “FSL”. In: *Neuroimage* 62.2 (2012), pp. 782–790. DOI: 10.1016/j.neuroimage.2011.09.015.

[54] Anders M Dale, Bruce Fischl, and Martin I Sereno. “Cortical surface-based analysis: I. Segmentation and surface reconstruction”. In: *Neuroimage* 9.2 (1999), pp. 179–194. DOI: 10.1006/nimg.1998.0395.

[55] Brian B Avants et al. “A reproducible evaluation of ANTs similarity metric performance in brain image registration”. In: *Neuroimage* 54.3 (2011), pp. 2033–2044. DOI: 10.1016/j.neuroimage.2010.09.025.

[56] David W Shattuck and Richard M Leahy. "BrainSuite: an automated cortical surface identification tool". In: *Medical image analysis* 6.2 (2002), pp. 129–142. DOI: 10.1016/s1361-8415(02)00054-3.

[57] Paul Schmidt et al. "An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis". In: *Neuroimage* 59.4 (2012), pp. 3774–3783. DOI: 10.1016/j.neuroimage.2011.11.032.

[58] Peter J Basser, James Mattiello, and Denis LeBihan. "Estimation of the effective self-diffusion tensor from the NMR spin echo". In: *Journal of Magnetic Resonance, Series B* 103.3 (1994), pp. 247–254. DOI: 10.1006/jmrb.1994.1037.

[59] Abhijit Guha Roy et al. "QuickNAT: A fully convolutional network for quick and accurate segmentation of neuroanatomy". In: *NeuroImage* 186 (2019), pp. 713–727. DOI: 10.1016/j.neuroimage.2018.11.042.

[60] Robert W Cox. "AFNI: software for analysis and visualization of functional magnetic resonance neuroimages". In: *Computers and Biomedical research* 29.3 (1996), pp. 162–173. DOI: 10.1006/cbmr.1996.0014.

[61] Christian Gaser et al. "CAT–A computational anatomy toolbox for the analysis of structural MRI data". In: *biorxiv* (2022), pp. 2022–06. DOI: 10.1101/2022.06.11.495736.

[62] D Mason. "SU-E-T-33: Pydicom: An Open Source DICOM Library". In: *Med. Phys.* 38.6Part10 (June 2011), pp. 3493–3493. DOI: 10.1118/1.3611983.

[63] Oskar Maier et al. *loli/medpy: MedPy 0.4.0*. Version 0.4.0. Feb. 2019. DOI: 10.5281/zenodo.2565940.

[64] Matthew Brett et al. *nipy/nibabel: 5.1.0*. Version 5.1.0. Apr. 2023. DOI: 10.5281/zenodo.7795644.

[65] Omer Faruk Gulban et al. *poldracklab/pydeface: PyDeface v2.0.2*. Version v2.0.2. July 2022. DOI: 10.5281/zenodo.6856482.

[66] Oscar Esteban et al. “MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites”. In: *PloS one* 12.9 (2017), e0184661. DOI: 10.1371/journal.pone.0184661.

[67] Elizaveta Lavrova et al. “Precision-medicine-toolbox: An open-source python package for the quantitative medical image analysis”. In: *Software Impacts* 16 (2023), p. 100508. DOI: 10.1016/j.simpa.2023.100508.

[68] Tal Yarkoni et al. “PyBIDS: Python tools for BIDS datasets”. In: *Journal of open source software* 4.40 (2019). DOI: 10.21105/joss.01294.

[69] Matthew F Sharrock et al. “3D deep neural network segmentation of intracerebral hemorrhage: development and validation for clinical trials”. In: *Neuroinformatics* 19 (2021), pp. 403–415. DOI: 10.1007/s12021-020-09493-5.

[70] Joost JM Van Griethuysen et al. “Computational radiomics system to decode the radiographic phenotype”. In: *Cancer research* 77.21 (2017), e104–e107. DOI: 10.1158/0008-5472.CAN-17-0339.

[71] Fernando Pérez-García, Rachel Sparks, and Sébastien Ourselin. “TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning”. In: *Computer Methods and Programs in Biomedicine* 208 (2021), p. 106236. DOI: 10.1016/j.cmpb.2021.106236.

[72] Artemis Zavaliangos-Petropulu et al. “Testing a convolutional neural network-based hippocampal segmentation method in a stroke population”. In: *Human Brain Mapping* 43.1 (2022), pp. 234–243. DOI: 10.1002/hbm.25210.

[73] World health Organization. *Dementia*. 2023. URL: https://www.who.int/news-room/fact-sheets/detail/dementia (visited on 09/28/2023).

[74] Siqi Liu et al. "Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease". en. In: *IEEE Trans. Biomed. Eng.* 62.4 (Apr. 2015), pp. 1132–1140. DOI: 10.1109/TBME.2014.2372011.

[75] R C Petersen et al. "Alzheimer's Disease Neuroimaging Initiative (ADNI)". en. In: *Neurology* 74.3 (Jan. 2010), pp. 201–209. DOI: 10.1212/WNL.0b013e3181cb3e25.

[76] Farheen Ramzan et al. "A Deep Learning Approach for Automated Diagnosis and Multi-Class Classification of Alzheimer's Disease Stages Using Resting-State fMRI and Residual Neural Networks". en. In: *J. Med. Syst.* 44.2 (Dec. 2019), p. 37. DOI: 10.1007/s10916-019-1475-2.

[77] Ahmad Chaddad, Christian Desrosiers, and Tamim Niazi. "Deep radiomic analysis of MRI related to Alzheimer's disease". In: *Ieee Access* 6 (2018), pp. 58213–58221. DOI: 10.1109/ACCESS.2018.2871977.

[78] Shangran Qiu et al. "Fusion of deep learning models of MRI scans, Mini-Mental State Examination, and logical memory test enhances diagnosis of mild cognitive impairment". en. In: *Alzheimers. Dement.* 10 (Sept. 2018), pp. 737–749. DOI: 10.1016/j.dadm.2018.08.013.

[79] Kun Zhao et al. "Independent and reproducible hippocampal radiomic biomarkers for multisite Alzheimer's disease: diagnosis, longitudinal progress and biological basis". en. In: *Sci Bull (Beijing)* 65.13 (July 2020), pp. 1103–1113. DOI: 10.1016/j.scib.2020.04.003.

[80] Shangran Qiu et al. "Development and validation of an interpretable deep learning framework for Alzheimer's disease classification". en. In: *Brain* 143.6 (June 2020), pp. 1920–1933. DOI: 10.1093/brain/awaa137.

[81] Ahsan Bin Tufail, Yong-Kui Ma, and Qiu-Na Zhang. “Binary Classification of Alzheimer’s Disease Using sMRI Imaging Modality and Deep Learning”. en. In: *J. Digit. Imaging* 33.5 (Oct. 2020), pp. 1073–1090. DOI: 10.1007/s10278-019-00265-5.

[82] Taeho Jo et al. “Deep learning detection of informative features in tau PET for Alzheimer’s disease classification”. en. In: *BMC Bioinformatics* 21.Suppl 21 (Dec. 2020), p. 496. DOI: 10.1186/s12859-020-03848-0.

[83] Nguyen Thanh Duc et al. “3D-Deep Learning Based Automatic Diagnosis of Alzheimer’s Disease with Joint MMSE Prediction Using Resting-State fMRI”. en. In: *Neuroinformatics* 18.1 (Jan. 2020), pp. 71–86. DOI: 10.1007/s12021-019-09419-w.

[84] Shui Liu et al. “Investigation of Underlying Association Between Whole Brain Regions and Alzheimer’s Disease: A Research Based on an Artificial Intelligence Model”. en. In: *Front. Aging Neurosci.* 14 (June 2022), p. 872530. DOI: 10.3389/fnagi.2022.872530.

[85] Sijia Du et al. “Deep learning-based PET/MR radiomics for the classification of annualized relapse rate in multiple sclerosis”. In: *Multiple Sclerosis and Related Disorders* 75 (2023), p. 104750. DOI: 10.1016/j.msard.2023.104750.

[86] Qi Feng et al. “Hippocampus Radiomic Biomarkers for the Diagnosis of Amnestic Mild Cognitive Impairment: A Machine Learning Method”. en. In: *Front. Aging Neurosci.* 11 (Nov. 2019), p. 323. DOI: 10.3389/fnagi.2019.00323.

[87] Ronghui Ju et al. “Early Diagnosis of Alzheimer’s Disease Based on Resting-State Brain Networks and Deep Learning”. en. In: *IEEE/ACM Trans. Comput. Biol. Bioinform.* 16.1 (2019), pp. 244–257. DOI: 10.1109/TCBB.2017.2776910.

[88] Jun Shi et al. “Multimodal Neuroimaging Feature Learning With Multimodal Stacked Deep Polynomial Networks for Diagnosis of Alzheimer’s Disease”. en. In: *IEEE J*

Biomed Health Inform 22.1 (Jan. 2018), pp. 173–183. DOI: 10.1109/JBHI.2017.2655720.

[89] Yiming Ding et al. "A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using F-FDG PET of the Brain". en. In: *Radiology* 290.2 (Feb. 2019), pp. 456–464. DOI: 10.1148/radiol.2018180958.

[90] Feng Feng et al. "Radiomic features of hippocampal subregions in Alzheimer's disease and amnestic mild cognitive impairment". In: *Frontiers in aging neuroscience* 10 (2018), p. 290. DOI: 10.3389/fnagi.2018.00290.

[91] Nicola Amoroso et al. "Deep learning reveals Alzheimer's disease onset in MCI subjects: Results from an international challenge". en. In: *J. Neurosci. Methods* 302 (May 2018), pp. 3–9. DOI: 10.1016/j.jneumeth.2017.12.011.

[92] Yupeng Li et al. "Radiomics: a novel feature extraction method for brain neuron degeneration disease using 18F-FDG PET imaging and its implementation for Alzheimer's disease and mild cognitive impairment". en. In: *Ther. Adv. Neurol. Disord.* (Mar. 2019). DOI: 10.1177/1756286419838682.

[93] Sara Ranjbar et al. "Brain MR Radiomics to Differentiate Cognitive Disorders". en. In: *J. Neuropsychiatry Clin. Neurosci.* 31.3 (Jan. 2019), pp. 210–219. DOI: 10.1176/appi.neuropsych.17120366.

[94] Silvia Basaia et al. "Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks". en. In: *Neuroimage Clin* 21 (2019), p. 101645. DOI: 10.1016/j.nicl.2018.101645.

[95] Dan Pan et al. "Early Detection of Alzheimer's Disease Using Magnetic Resonance Imaging: A Novel Approach Combining Convolutional Neural Networks and Ensemble Learning". en. In: *Front. Neurosci.* 14 (May 2020), p. 259. DOI: 10.3389/fnins.2020.00259.

[96] Yanhui Ding et al. “Quantitative Radiomic Features as New Biomarkers for Alzheimer’s Disease: An Amyloid PET Study”. en. In: *Cereb. Cortex* 31.8 (July 2021), pp. 3950–3961. DOI: 10.1093/cercor/bhab061.

[97] Qi Feng et al. “Comprehensive classification models based on amygdala radiomic features for Alzheimer’s disease and mild cognitive impairment”. en. In: *Brain Imaging Behav.* 15.5 (Oct. 2021), pp. 2377–2386. DOI: 10.1007/s11682-020-00434-z.

[98] Yang Du et al. “Radiomic Features of the Hippocampus for Diagnosing Early-Onset and Late-Onset Alzheimer’s Disease”. en. In: *Front. Aging Neurosci.* 13 (2021), p. 789099. DOI: 10.3389/fnagi.2021.789099.

[99] Kun Zhao et al. “Regional Radiomics Similarity Networks Reveal Distinct Subtypes and Abnormality Patterns in Mild Cognitive Impairment”. en. In: *Adv. Sci.* 9.12 (Apr. 2022), e2104538. DOI: 10.1002/advs.202104538.

[100] Jun Pyo Kim et al. “Predicting amyloid positivity in patients with mild cognitive impairment using a radiomics approach”. en. In: *Sci. Rep.* 11.1 (Mar. 2021), p. 6954. DOI: 10.1038/s41598-021-86114-4.

[101] Kexin Huang et al. “A multipredictor model to predict the conversion of mild cognitive impairment to Alzheimer’s disease by using a predictive nomogram”. en. In: *Neuropsychopharmacology* 45.2 (Jan. 2020), pp. 358–366. DOI: 10.1038/s41386-019-0551-0.

[102] Hucheng Zhou et al. “Dual-Model Radiomic Biomarkers Predict Development of Mild Cognitive Impairment Progression to Alzheimer’s Disease”. en. In: *Front. Neurosci.* 12 (2018), p. 1045. DOI: 10.3389/fnins.2018.01045.

[103] Simeon Spasov et al. “A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer’s disease”. en. In: *Neuroimage* 189 (Apr. 2019), pp. 276–287. DOI: 10.1016/j.neuroimage.2019.01.031.

[104] Garam Lee et al. "Predicting Alzheimer's disease progression using multi-modal deep learning approach". en. In: *Sci. Rep.* 9.1 (Feb. 2019), p. 1952. DOI: 10.1038/s41598-018-37769-z.

[105] Tao-Ran Li et al. "Radiomics Analysis of Magnetic Resonance Imaging Facilitates the Identification of Preclinical Alzheimer's Disease: An Exploratory Study". en. In: *Front Cell Dev Biol* 8 (Dec. 2020), p. 605734. DOI: 10.3389/fcell.2020.605734.

[106] Anees Abrol et al. "Deep residual learning for neuroimaging: An application to predict progression to Alzheimer's disease". en. In: *J. Neurosci. Methods* 339 (June 2020), p. 108701. DOI: 10.1016/j.jneumeth.2020.108701.

[107] Fan Yang et al. "Combining PET with MRI to improve predictions of progression from mild cognitive impairment to Alzheimer's disease: an exploratory radiomic analysis study". en. In: *Ann Transl Med* 10.9 (May 2022), p. 513. DOI: 10.21037/atm-21-4349.

[108] Yupeng Li et al. "Radiomics features as predictors to distinguish fast and slow progression of Mild Cognitive Impairment to Alzheimer's disease". en. In: *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2018 (July 2018), pp. 127–130. DOI: 10.1109/EMBC.2018.8512273.

[109] Tomomichi Iizuka, Makoto Fukasawa, and Masashi Kameyama. "Deep-learning-based imaging-classification identified cingulate island sign in dementia with Lewy bodies". en. In: *Sci. Rep.* 9.1 (June 2019), p. 8944. DOI: 10.1038/s41598-019-45415-5.

[110] Ryusuke Irie et al. "A Novel Deep Learning Approach with a 3D Convolutional Ladder Network for Differential Diagnosis of Idiopathic Normal Pressure Hydrocephalus and Alzheimer's Disease". en. In: *Magn. Reson. Med. Sci.* 19.4 (2020), pp. 351–358. DOI: 10.2463/mrms.mp.2019-0106.

[111] Qi Feng et al. “Correlation Between Hippocampus MRI Radiomic Features and Resting-State Intrahippocampal Functional Connectivity in Alzheimer’s Disease”. en. In: *Front. Neurosci.* 13 (May 2019), p. 435. DOI: 10.3389/fnins.2019.00435.

[112] Hans Lassmann. “Multiple sclerosis pathology”. In: *Cold Spring Harbor perspectives in medicine* 8.3 (2018). DOI: 10.1101/cshperspect.a028936.

[113] Hannelore K van der Burgh et al. “Deep learning predictions of survival based on MRI in amyotrophic lateral sclerosis”. en. In: *Neuroimage Clin* 13 (2017), pp. 361–369. DOI: 10.1016/j.nicl.2016.10.008.

[114] Youngjin Yoo et al. “Deep learning of joint myelin and T1w MRI features in normal-appearing brain tissue to distinguish between multiple sclerosis patients and healthy controls”. In: *NeuroImage: Clinical* 17 (2018), pp. 169–178. DOI: 10.1016/j.nicl.2017.10.015.

[115] Fabian Eitel et al. “Uncovering convolutional neural network decisions for diagnosing multiple sclerosis on conventional MRI using layer-wise relevance propagation”. en. In: *Neuroimage Clin* 24 (Sept. 2019), p. 102003. DOI: 10.1016/j.nicl.2019.102003.

[116] Elizaveta Lavrova et al. “Exploratory radiomic analysis of conventional vs. quantitative brain MRI: toward automatic diagnosis of early multiple sclerosis”. In: *Frontiers in neuroscience* 15 (2021), p. 679941. DOI: 10.3389/fnins.2021.679941.

[117] Yaou Liu et al. “Radiomics in multiple sclerosis and neuromyelitis optica spectrum disorder”. en. In: *Eur. Radiol.* 29.9 (Sept. 2019), pp. 4670–4677. DOI: 10.1007/s00330-019-06026-w.

[118] Xiaoxiao Ma et al. "Quantitative radiomic biomarkers for discrimination between neuromyelitis optica spectrum disorder and multiple sclerosis". In: *Journal of Magnetic Resonance Imaging* 49.4 (2019), pp. 1113–1121. DOI: 10.1002/jmri.26287.

[119] Zhuo Wang et al. "3D Compressed Convolutional Neural Network Differentiates Neuromyelitis Optical Spectrum Disorders From Multiple Sclerosis Using Automated White Matter Hyperintensities Segmentations". en. In: *Front. Physiol.* 11 (Dec. 2020), p. 612928. DOI: 10.3389/fphys.2020.612928.

[120] Akifumi Hagiwara et al. "Differentiation between multiple sclerosis and neuromyelitis optica spectrum disorders by multiparametric quantitative MRI using convolutional neural network". en. In: *J. Clin. Neurosci.* 87 (May 2021), pp. 55–58. DOI: 10.1016/j.jocn.2021.02.018.

[121] Xiao Luo et al. "Multi-lesion radiomics model for discrimination of relapsing-remitting multiple sclerosis and neuropsychiatric systemic lupus erythematosus". en. In: *Eur. Radiol.* 32.8 (Aug. 2022), pp. 5700–5710. DOI: 10.1007/s00330-022-08653-2.

[122] Ting He et al. "MS or not MS: T2-weighted imaging (T2WI)-based radiomic findings distinguish MS from its mimics". en. In: *Mult. Scler. Relat. Disord.* 61 (May 2022). DOI: 10.1016/j.msard.2022.103756.

[123] Maria A Rocca et al. "Deep Learning on Conventional Magnetic Resonance Imaging Improves the Diagnosis of Multiple Sclerosis Mimics". en. In: *Invest. Radiol.* 56.4 (Apr. 2021), pp. 252–260. DOI: 10.1097/RLI.0000000000000735.

[124] Yang Du et al. "The effect of hippocampal radiomic features and functional connectivity on the relationship between hippocampal volume and cognitive function in Alzheimer's disease". en. In: *J. Psychiatr. Res.* 158 (Feb. 2023), pp. 382–391. DOI: 10.1016/j.jpsychires.2023.01.024.

[125] P Roca et al. “Artificial intelligence to predict clinical disability in patients with multiple sclerosis using FLAIR MRI”. In: *Diagnostic and Interventional Imaging* 101.12 (2020), pp. 795–802. DOI: 10.1016/j.diii.2020.05.009.

[126] Yunyan Zhang et al. “Grad-CAM helps interpret the deep learning models trained to classify multiple sclerosis types using clinical brain magnetic resonance imaging”. en. In: *J. Neurosci. Methods* 353 (Apr. 2021), p. 109098. DOI: 10.1016/j.jneumeth.2021.109098.

[127] Wen Wei et al. “Predicting PET-derived myelin content from multisequence MRI for individual longitudinal analysis in multiple sclerosis”. en. In: *Neuroimage* 223 (Dec. 2020), p. 117308. DOI: 10.1016/j.neuroimage.2020.117308.

[128] Zezhong Ye et al. “Deep learning with diffusion basis spectrum imaging for classification of multiple sclerosis lesions”. en. In: *Ann Clin Transl Neurol* 7.5 (May 2020), pp. 695–706. DOI: 10.1002/acn3.51037.

[129] Germán Barquero et al. “RimNet: A deep 3D multimodal MRI architecture for paramagnetic rim lesion assessment in multiple sclerosis”. en. In: *Neuroimage Clin* 28 (Sept. 2020), p. 102412. DOI: 10.1016/j.nicl.2020.102412.

[130] Hang Zhang et al. “QSMRim-Net: Imbalance-aware learning for identification of chronic active multiple sclerosis lesions on quantitative susceptibility maps”. en. In: *Neuroimage Clin* 34 (Mar. 2022), p. 102979. DOI: 10.1016/j.nicl.2022.102979.

[131] Yuling Peng et al. “Prediction of unenhanced lesion evolution in multiple sclerosis using radiomics-based models: a machine learning approach”. en. In: *Mult. Scler. Relat. Disord.* 53 (Aug. 2021), p. 102989. DOI: 10.1016/j.msard.2021.102989.

[132] World health Organization. *Parkinson disease*. 2023. URL: https://www.who.int/news-room/fact-sheets/detail/parkinson-disease (visited on 09/28/2023).

[133] Zenghui Cheng et al. "Radiomic Features of the Nigrosome-1 Region of the Substantia Nigra: Using Quantitative Susceptibility Mapping to Assist the Diagnosis of Idiopathic Parkinson's Disease". en. In: *Front. Aging Neurosci.* 11 (July 2019), p. 167. DOI: 10.3389/fnagi.2019.00167.

[134] Yue Wu et al. "Use of radiomic features and support vector machine to distinguish Parkinson's disease cases from normal controls". en. In: *Ann Transl Med* 7.23 (Dec. 2019), p. 773. DOI: 10.21037/atm.2019.11.26.

[135] Bin Xiao et al. "Quantitative susceptibility mapping based hybrid feature extraction for diagnosis of Parkinson's disease". en. In: *Neuroimage Clin* 24 (Nov. 2019), p. 102070. DOI: 10.1016/j.nicl.2019.102070.

[136] Xuan Cao et al. "A Radiomics Approach to Predicting Parkinson's Disease by Incorporating Whole-Brain Functional Activity and Gray Matter Structure". en. In: *Front. Neurosci.* 14 (July 2020), p. 751. DOI: 10.3389/fnins.2020.00751.

[137] Panshi Liu et al. "Parkinson's Disease Diagnosis Using Neostriatum Radiomic Features Based on T2-Weighted Magnetic Resonance Imaging". en. In: *Front. Neurol.* 11 (Apr. 2020), p. 248. DOI: 10.3389/fneur.2020.00248.

[138] Zhenyu Shu et al. "An Integrative Nomogram for Identifying Early-Stage Parkinson's Disease Using Non-motor Symptoms and White Matter-Based Radiomics Biomarkers From Whole-Brain MRI". en. In: *Front. Aging Neurosci.* 12 (Dec. 2020), p. 548616. DOI: 10.3389/fnagi.2020.548616.

[139] Sabyasachi Chakraborty, Satyabrata Aich, and Hee-Cheol Kim. “Detection of Parkinson’s Disease from 3T T1 Weighted MRI Scans Using 3D Convolutional Neural Network”. en. In: *Diagnostics (Basel)* 10.6 (June 2020). DOI: 10.3390/diagnostics10060402.

[140] Xuan Cao, Kyoungjae Lee, and Qingling Huang. “Bayesian variable selection in logistic regression with application to whole-brain functional connectivity analysis for Parkinson’s disease”. en. In: *Stat. Methods Med. Res.* 30.3 (Mar. 2021), pp. 826–842. DOI: 10.1177/0962280220978990.

[141] Koichiro Yasaka et al. “Parkinson’s disease: deep learning with a parameter-weighted structural connectome matrix for diagnosis and neural circuit disorder investigation”. en. In: *Neuroradiology* 63.9 (Sept. 2021), pp. 1451–1462. DOI: 10.1007/s00234-021-02648-4.

[142] Benedetta Tafuri et al. “The impact of harmonization on radiomic features in Parkinson’s disease and healthy controls: A multicenter study”. en. In: *Front. Neurosci.* 16 (Oct. 2022), p. 1012287. DOI: 10.3389/fnins.2022.1012287.

[143] Takuro Shiiba et al. “Dopamine transporter single-photon emission computed tomography-derived radiomics signature for detecting Parkinson’s disease”. en. In: *EJNMMI Res.* 12.1 (June 2022), p. 39. DOI: 10.1186/s13550-022-00910-1.

[144] Xiaoming Sun et al. “Use of deep learning-based radiomics to differentiate Parkinson’s disease patients from normal controls: a study based on [F]FDG PET imaging”. en. In: *Eur. Radiol.* 32.11 (Nov. 2022), pp. 8008–8018. DOI: 10.1007/s00330-022-08799-z.

[145] Yu Zhao et al. “A 3D Deep Residual Convolutional Neural Network for Differential Diagnosis of Parkinsonian Syndromes on F-FDG PET Images”. en. In: *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2019 (July 2019), pp. 3531–3534. DOI: 10.1109/EMBC.2019.8856747.

[146] Shigeru Kiryu et al. "Deep learning to differentiate parkinsonian disorders separately using single midsagittal MR imaging: a proof of concept study". en. In: *Eur. Radiol.* 29.12 (Dec. 2019), pp. 6891–6899. DOI: 10.1007/s00330-019-06327-0.

[147] Huize Pang et al. "MRI-Based Radiomics of Basal Nuclei in Differentiating Idiopathic Parkinson's Disease From Parkinsonian Variants of Multiple System Atrophy: A Susceptibility-Weighted Imaging Study". en. In: *Front. Aging Neurosci.* 12 (Nov. 2020), p. 587250. DOI: 10.3389/fnagi.2020.587250.

[148] Dong Sun et al. "Differentiating Parkinson's disease motor subtypes: A radiomics analysis based on deep gray nuclear lesion and white matter". en. In: *Neurosci. Lett.* 760 (Aug. 2021), p. 136083. DOI: 10.1016/j.neulet.2021.136083.

[149] Joseph Jankovic et al. *Principles and Practice of Movement Disorders E-Book*. en. Elsevier Health Sciences, June 2021.

[150] Dafa Shi et al. "Classification of Parkinson's disease using a region-of-interest- and resting-state functional magnetic resonance imaging-based radiomics approach". en. In: *Brain Imaging Behav.* 16.5 (Oct. 2022), pp. 2150–2163. DOI: 10.1007/s11682-022-00685-y.

[151] Huize Pang et al. "Multimodal striatal neuromarkers in distinguishing parkinsonian variant of multiple system atrophy from idiopathic Parkinson's disease". en. In: *CNS Neurosci. Ther.* 28.12 (Dec. 2022), pp. 2172–2182. DOI: 10.1111/cns.13959.

[152] Yu Zhao et al. "Decoding the dopamine transporter imaging for the differential diagnosis of parkinsonism using deep learning". en. In: *Eur. J. Nucl. Med. Mol. Imaging* 49.8 (July 2022), pp. 2798–2811. DOI: 10.1007/s00259-022-05804-x.

[153] Xuehan Hu et al. "Multivariate radiomics models based on F-FDG hybrid PET/MRI for distinguishing between Parkinson's disease and multiple system atrophy". en. In: *Eur. J. Nucl. Med.*

Mol. Imaging 48.11 (Oct. 2021), pp. 3469–3481. DOI: 10.1007/s00259-021-05325-z.

[154] Priyanka Tupe-Waghmare et al. "Radiomics on routine T1-weighted MRI can delineate Parkinson's disease from multiple system atrophy and progressive supranuclear palsy". en. In: *Eur. Radiol.* 31.11 (Nov. 2021), pp. 8218–8227. DOI: 10.1007/s00330-021-07979-7.

[155] Dong Hoon Shin et al. "Automated assessment of the substantia nigra on susceptibility map-weighted imaging using deep convolutional neural networks for diagnosis of Idiopathic Parkinson's disease". en. In: *Parkinsonism Relat. Disord.* 85 (Apr. 2021), pp. 84–90. DOI: 10.1016/j.parkreldis.2021.03.004.

[156] Hongyoon Choi et al. "Cognitive signature of brain FDG PET based on deep learning: domain transfer from Alzheimer's disease to Parkinson's disease". en. In: *Eur. J. Nucl. Med. Mol. Imaging* 47.2 (Feb. 2020), pp. 403–412. DOI: 10.1007/s00259-019-04538-7.

[157] World health Organization. *World Stroke Day*. 2023. URL: https://www.who.int/southeastasia/news/detail/28-10-2021-world-stroke-day (visited on 09/28/2023).

[158] Mizuho Nishio et al. "Automatic detection of acute ischemic stroke using non-contrast computed tomography and two-stage deep learning model". en. In: *Comput. Methods Programs Biomed.* 196 (Nov. 2020), p. 105711. DOI: 10.1016/j.cmpb.2020.105711.

[159] Yingwei Guo et al. "Radiomics features of DSC-PWI in time dimension may provide a new chance to identify ischemic stroke". en. In: *Front. Neurol.* 13 (Nov. 2022), p. 889090. DOI: 10.3389/fneur.2022.889090.

[160] Jianbo Lyu et al. "Machine learning-based CT radiomics model to discriminate the primary and secondary intracranial hemorrhage". en. In: *Sci. Rep.* 13.1 (Mar. 2023), p. 3709. DOI: 10.1038/s41598-023-30678-w.

[161] Anne Nielsen et al. "Prediction of Tissue Outcome and Assessment of Treatment Effect in Acute Ischemic Stroke Using Deep Learning". en. In: *Stroke* 49.6 (June 2018), pp. 1394–1401. DOI: 10.1161/STROKEAHA.117.019740.

[162] Qian Chen et al. "Clinical-radiomics Nomogram for Risk Estimation of Early Hematoma Expansion after Acute Intracerebral Hemorrhage". en. In: *Acad. Radiol.* 28.3 (Mar. 2021), pp. 307–317. DOI: 10.1016/j.acra.2020.02.021.

[163] Gang Xie et al. "Radiomics-based infarct features on CT predict hemorrhagic transformation in patients with acute ischemic stroke". en. In: *Front. Neurosci.* 16 (Sept. 2022), p. 1002717. DOI: 10.3389/fnins.2022.1002717.

[164] Lei Song et al. "Predicting Hemorrhage Progression in Deep Intracerebral Hemorrhage: A Multicenter Retrospective Cohort Study". en. In: *World Neurosurg.* 170 (Feb. 2023), e387–e401. DOI: 10.1016/j.wneu.2022.11.022.

[165] Mumu Aktar et al. "A Radiomics-Based Machine Learning Approach to Assess Collateral Circulation in Ischemic Stroke on Non-contrast Computed Tomography". en. In: *Multimodal Learning for Clinical Decision Support and Clinical Image-Based Procedures* (2020), pp. 24–33. DOI: 10.1007/978-3-030-60946-7_3.

[166] W Qiu et al. "Radiomics-Based Intracranial Thrombus Features on CT and CTA Predict Recanalization with Intravenous Alteplase in Patients with Acute Ischemic Stroke". en. In: *AJNR Am. J. Neuroradiol.* 40.1 (Jan. 2019), pp. 39–44. DOI: 10.3174/ajnr.A5918.

[167] Jeremy Hofmeister et al. “Clot-Based Radiomics Predict a Mechanical Thrombectomy Strategy for Successful Recanalization in Acute Ischemic Stroke”. en. In: *Stroke* 51.8 (Aug. 2020), pp. 2488–2494. DOI: 10.1161/STROKEAHA.120.030334.

[168] A Hilbert et al. “Data-efficient deep learning of radiological image data for outcome prediction after endovascular treatment of patients with acute ischemic stroke”. en. In: *Comput. Biol. Med.* 115 (Dec. 2019), p. 103516. DOI: 10.1016/j.compbiomed.2019.103516.

[169] Sucheta Chauhan et al. “A Comparison of Shallow and Deep Learning Methods for Predicting Cognitive Performance of Stroke Patients From MRI Lesion Images”. en. In: *Front. Neuroinform.* 13 (July 2019), p. 53. DOI: 10.3389/fninf.2019.00053.

[170] Hidehisa Nishi et al. “Deep Learning–Derived High-Level Neuroimaging Features Predict Clinical Outcomes for Large Vessel Occlusion”. en. In: *Stroke* (May 2020). DOI: 10.1161/STROKEAHA.119.028101.

[171] Stephen Bacchi et al. “Deep Learning in the Prediction of Ischaemic Stroke Thrombolysis Functional Outcomes: A Pilot Study”. en. In: *Acad. Radiol.* 27.2 (Feb. 2020), e19–e23. DOI: 10.1016/j.acra.2019.03.015.

[172] Yan Li et al. “Combining machine learning with radiomics features in predicting outcomes after mechanical thrombectomy in patients with acute ischemic stroke”. en. In: *Comput. Methods Programs Biomed.* 225 (Oct. 2022), p. 107093. DOI: 10.1016/j.cmpb.2022.107093.

[173] Manon L Tolhuisen et al. “Outcome Prediction Based on Automatically Extracted Infarct Core Image Features in Patients with Acute Ischemic Stroke”. en. In: *Diagnostics (Basel)* 12.8 (July 2022). DOI: 10.3390/diagnostics12081786.

[174] Yi-Qun Zhang et al. "MRI radiomic features-based machine learning approach to classify ischemic stroke onset time". en. In: *J. Neurol.* 269.1 (Jan. 2022), pp. 350–360. DOI: 10.1007/s00415-021-10638-y.

[175] Hongxia Li et al. "Identification of high-risk intracranial plaques with 3D high-resolution magnetic resonance imaging-based radiomics and machine learning". en. In: *J. Neurol.* 269.12 (Dec. 2022), pp. 6494–6503. DOI: 10.1007/s00415-022-11315-4.

[176] Hao Wang et al. "Diffusion-weighted imaging-based radiomics for predicting 1-year ischemic stroke recurrence". en. In: *Front. Neurol.* 13 (Oct. 2022), p. 1012896. DOI: 10.3389/fneur.2022.1012896.

[177] Ru Lin et al. "Development and validation of a novel radiomics-clinical model for predicting post-stroke epilepsy after first-ever intracerebral haemorrhage". en. In: *Eur. Radiol.* 33.7 (July 2023), pp. 4526–4536. DOI: 10.1007/s00330-023-09429-y.

[178] World health Organization. *Epilepsy*. 2023. URL: https://www.who.int/news-room/fact-sheets/detail/epilepsy (visited on 09/28/2023).

[179] E-Nae Cheong et al. "Extrahippocampal Radiomics Analysis Can Potentially Identify Laterality in Patients With MRI-Negative Temporal Lobe Epilepsy". en. In: *Front. Neurol.* 12 (Aug. 2021), p. 706576. DOI: 10.3389/fneur.2021.706576.

[180] Qinming Zhang et al. "A deep learning framework for F-FDG PET imaging diagnosis in pediatric patients with temporal lobe epilepsy". en. In: *Eur. J. Nucl. Med. Mol. Imaging* 48.8 (July 2021), pp. 2476–2485. DOI: 10.1007/s00259-020-05108-y.

[181] Cuixia Feng et al. "Improved detection of focal cortical dysplasia in normal-appearing FLAIR images using a Bayesian classifier". en. In: *Med. Phys.* 48.2 (Feb. 2021), pp. 912–925. DOI: 10.1002/mp.14646.

[182] Kyung Min Kim et al. “Development and Validation of MRI-Based Radiomics Models for Diagnosing Juvenile Myoclonic Epilepsy”. en. In: *Korean J. Radiol.* 23.12 (Dec. 2022), pp. 1281–1289. DOI: 10.3348/kjr.2022.0539.

[183] World health Organization. *Mental disorders*. 2023. URL: https://www.who.int/news-room/fact-sheets/detail/mental-disorders (visited on 09/28/2023).

[184] Junghoe Kim et al. “Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia”. In: *Neuroimage* 124 (2016), pp. 127–146. DOI: 10.1016/j.neuroimage.2015.05.018.

[185] Long-Biao Cui et al. “Disease Definition for Schizophrenia by Functional Connectivity Using Radiomics Strategy”. en. In: *Schizophr. Bull.* 44.5 (Feb. 2018), pp. 1053–1059. DOI: 10.1093/schbul/sby007.

[186] Ling-Li Zeng et al. “Multi-Site Diagnostic Classification of Schizophrenia Using Discriminant Deep Learning with Functional Connectivity MRI”. en. In: *EBioMedicine* 30 (Apr. 2018), pp. 74–85. DOI: 10.1016/j.ebiom.2018.03.017.

[187] Weizheng Yan et al. “Discriminating schizophrenia using recurrent neural network applied on time courses of multi-site FMRI data”. en. In: *EBioMedicine* 47 (Sept. 2019), pp. 543–552. DOI: 10.1016/j.ebiom.2019.08.023.

[188] Muhammad Naveed Iqbal Qureshi, Jooyoung Oh, and Boreom Lee. “3D-CNN based discrimination of schizophrenia using resting-state fMRI”. en. In: *Artif. Intell. Med.* 98 (July 2019), pp. 10–17. DOI: 10.1016/j.artmed.2019.06.003.

[189] Yae Won Park et al. “Differentiating patients with schizophrenia from healthy controls by hippocampal subfields using radiomics”. en. In: *Schizophr. Res.* 223 (Sept. 2020), pp. 337–344. DOI: 10.1016/j.schres.2020.09.009.

[190] Gang Li et al. "Application of deep canonically correlated sparse autoencoder for the classification of schizophrenia". en. In: *Comput. Methods Programs Biomed.* 183 (Jan. 2020), p. 105073. DOI: 10.1016/j.cmpb.2019.105073.

[191] Jihoon Oh et al. "Identifying Schizophrenia Using Structural MRI With a Deep Learning Algorithm". en. In: *Front. Psychiatry* 11 (Feb. 2020), p. 16. DOI: 10.3389/fpsyt.2020.00016.

[192] Mengjiao Hu et al. "Structural and diffusion MRI based schizophrenia classification using 2D pretrained and 3D naive Convolutional Neural Networks". en. In: *Schizophr. Res.* 243 (May 2022), pp. 330–341. DOI: 10.1016/j.schres.2021.06.011.

[193] Ying Wang et al. "Classification of Unmedicated Bipolar Disorder Using Whole-Brain Functional Activity and Connectivity: A Radiomics Analysis". en. In: *Cereb. Cortex* 30.3 (Mar. 2020), pp. 1117–1128. DOI: 10.1093/cercor/bhz152.

[194] Sandra Vieira et al. "Using Machine Learning and Structural Neuroimaging to Detect First Episode Psychosis: Reconsidering the Evidence". en. In: *Schizophr. Bull.* 46.1 (Feb. 2019), pp. 17–26. DOI: 10.1093/schbul/sby189.

[195] Jianlong Zhao et al. "Functional network connectivity (FNC)-based generative adversarial network (GAN) and its applications in classification of mental disorders". In: *Journal of neuroscience methods* 341 (2020), p. 108756. DOI: 10.1016/j.jneumeth.2020.108756.

[196] Zhuangzhuang Li et al. "Deep learning based automatic diagnosis of first-episode psychosis, bipolar disorder and healthy controls". In: *Computerized Medical Imaging and Graphics* 89 (2021), p. 101882. DOI: 10.1016/j.compmedimag.2021.101882.

[197] Long-Biao Cui et al. “Thalamus Radiomics-Based Disease Identification and Prediction of Early Treatment Response for Schizophrenia”. en. In: *Front. Neurosci.* 15 (July 2021), p. 682777. DOI: 10.3389/fnins.2021.682777.

[198] Long-Biao Cui et al. “Baseline structural and functional magnetic resonance imaging predicts early treatment response in schizophrenia with radiomics strategy”. en. In: *Eur. J. Neurosci.* 53.6 (Mar. 2021), pp. 1961–1975. DOI: 10.1111/ejn.15046.

[199] Ahmad Chaddad et al. “Hippocampus and amygdala radiomic biomarkers for the study of autism spectrum disorder”. en. In: *BMC Neurosci.* 18.1 (July 2017), pp. 1–12. DOI: 10.1186/s12868-017-0373-0.

[200] Huaiqiang Sun et al. “Psychoradiologic Utility of MR Imaging for Diagnosis of Attention Deficit Hyperactivity Disorder: A Radiomics Analysis”. en. In: *Radiology* (Nov. 2017). DOI: 10.1148/radiol.2017170226.

[201] Guanlu Liu et al. “Identifying individuals with attention-deficit/hyperactivity disorder based on multisite resting-state functional magnetic resonance imaging: A radiomics analysis”. en. In: *Hum. Brain Mapp.* 44.8 (June 2023), pp. 3433–3445. DOI: 10.1002/hbm.26290.

[202] Ahmad Chaddad, Christian Desrosiers, and Matthew Toews. “Multi-scale radiomic analysis of sub-cortical regions in MRI related to autism, gender and age”. en. In: *Sci. Rep.* 7 (Mar. 2017), p. 45639. DOI: 10.1038/srep45639.

[203] Weifeng Chen et al. “Oasis: A large-scale dataset for single image 3d in the wild”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 679–688.

[204] Sook-Lei Liew et al. “A large, curated, open-source stroke neuroimaging dataset to improve lesion segmentation algorithms”. en. Dec. 2021. DOI: 10.1038/s41597-022-01401-7.

[205] Imperial College London Biomedical Image Analysis Group. *IXI Dataset*. Accessed: 2023-6-30. URL: http://brain-development.org/%20ixi-dataset/.

[206] Stephanie Noble et al. "Influences on the Test–Retest Reliability of Functional Connectivity MRI and its Relationship with Behavioral Utility". en. In: *Cereb. Cortex* 27.11 (Sept. 2017), pp. 5415–5429. DOI: 10.1093/cercor/bhx230.

[207] Arsany Hakim et al. "Predicting Infarct Core From Computed Tomography Perfusion in Acute Ischemia With Machine Learning: Lessons From the ISLES Challenge". en. In: *Stroke* 52.7 (July 2021), pp. 2328–2337. DOI: 10.1161/STROKEAHA.120.030696.

[208] Aaron Carass et al. "Longitudinal multiple sclerosis lesion segmentation: resource and challenge". In: *NeuroImage* 148 (2017), pp. 77–102. DOI: 10.1016/j.neuroimage.2016.12.064.

[209] Benjamin Puccio et al. "The preprocessed connectomes project repository of manually corrected skull-stripped T1-weighted anatomical MRI data". en. In: *Gigascience* 5.1 (Oct. 2016), s13742–016–0150–5. DOI: 10.1186/s13742-016-0150-5.

[210] Roberto Souza et al. "An open, multi-vendor, multi-field-strength brain MR dataset and analysis of publicly available skull stripping methods agreement". In: *NeuroImage* 170 (2018), pp. 482–494. DOI: 10.1016/j.neuroimage.2017.08.021.

[211] Murtadha D Hssayeni et al. "Intracranial Hemorrhage Segmentation Using a Deep Convolutional Model". en. In: *Brown Univ. Dig. Addict. Theory Appl.* 5.1 (Feb. 2020), p. 14. DOI: 10.3390/data5010014.

[212] Radiological Society of North America. *RSNA Intracranial Hemorrhage Detection*. Accessed: 2023-6-30. 2019. URL: https://kaggle.com/competitions/rsna-intracranial-hemorrhage-detection.

[213] Kenneth Marek et al. “The Parkinson progression marker initiative (PPMI)”. In: *Progress in neurobiology* 95.4 (2011), pp. 629–635. DOI: 10.1016/j.pneurobio.2011.09.005.

[214] Jean Feng et al. “Clinical artificial intelligence quality improvement: towards continual monitoring and updating of AI algorithms in healthcare”. In: *npj Digital Medicine* 5.1 (2022), p. 66. DOI: 10.1038/s41746-022-00611-y.

[215] Alex Youssef et al. “All models are local: time to replace external validation with recurrent local validation”. In: *arXiv preprint arXiv:2305.03219* (2023). DOI: 10.48550/arXiv.2305.03219.

[216] Olvert A Berkhemer et al. “A randomized trial of intraarterial treatment for acute ischemic stroke”. en. In: *N. Engl. J. Med.* 372.1 (Jan. 2015), pp. 11–20. DOI: 10.1056/NEJMoa1411587.

[217] Mark D Wilkinson et al. “The FAIR Guiding Principles for scientific data management and stewardship”. In: *Scientific data* 3.1 (2016), pp. 1–9. DOI: 10.1038/sdata.2016.18.

[218] Shaoting Zhang and Dimitris Metaxas. “On the Challenges and Perspectives of Foundation Models for Medical Image Analysis”. In: *arXiv preprint arXiv:2306.05705* (June 2023). DOI: 10.48550/arXiv.2306.05705.

[219] Gary S Collins et al. “Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement”. en. In: *BMC Med.* 13.1 (Jan. 2015), pp. 1–10. DOI: 10.1136/bmj.g7594.

[220] Philippe Lambin et al. “Radiomics: the bridge between medical imaging and personalized medicine”. In: *Nature reviews Clinical oncology* 14.12 (2017), pp. 749–762. DOI: 10.1038/nrclinonc.2017.141.

[221] Burak Kocak et al. "CheckList for EvaluAtion of Radiomics research (CLEAR): a step-by-step reporting guideline for authors and reviewers endorsed by ESR and EuSoMII". In: *Insights into Imaging* 14.1 (2023), pp. 1–13. DOI: 10.1186/s13244-023-01415-8.

[222] Karim Lekadir et al. "FUTURE-AI: Guiding Principles and Consensus Recommendations for Trustworthy Artificial Intelligence in Medical Imaging". In: *arXiv preprint arXiv:2109.09658* (Sept. 2021). DOI: 10.48550/arXiv.2109.09658.

[223] Bas H.M. van der Velden et al. "Explainable artificial intelligence (XAI) in deep learning-based medical image analysis". In: *Medical Image Analysis* 79 (2022), p. 102470. DOI: 10.1016/j.media.2022.102470.

[224] Zohaib Salahuddin et al. "Transparency of deep neural networks for medical image analysis: A review of interpretability methods". In: *Computers in Biology and Medicine* 140 (2022), p. 105111. DOI: 10.1016/j.compbiomed.2021.105111.

[225] Lena Maier-Hein, Bjoern Menze, et al. "Metrics reloaded: Pitfalls and recommendations for image analysis validation". In: *arXiv. org* 2206.01653 (2022). DOI: 10.48550/arXiv.2206.01653.

4

Exploratory study: radiomic biomarkers in multiple sclerosis

Elizaveta Lavrova*, Emilie Lommers*, Henry C Woodruff*, Avishek Chatterjee, Pierre Maquet, Eric Salmon, Philippe Lambin, Christophe Phillips**; "*", "**" – equal contribution.**

Adapted from: Elizaveta Lavrova et al. "Exploratory radiomic analysis of conventional vs. quantitative brain MRI: toward automatic diagnosis of early multiple sclerosis". In: *Frontiers in neuroscience* 15 (2021), p. 679941. DOI: 10.3389/fnins.2021.679941.

Abstract

Conventional magnetic resonance imaging (cMRI) is poorly sensitive to pathological changes related to multiple sclerosis (MS) in normal-appearing white matter (NAWM) and gray matter (GM), with the added difficulty of not being very reproducible. Quantitative MRI (qMRI), on the other hand, attempts to represent the physical properties of tissues, making it an ideal candidate for quantitative medical image analysis or radiomics. We therefore hypothesized that qMRI-based radiomic features have added diagnostic value in MS compared to cMRI. This study investigated the ability of cMRI (T1w) and qMRI features extracted from white matter (WM), NAWM, and GM to distinguish between MS patients (MSP) and healthy control subjects (HCS). We developed exploratory radiomic classification models on a dataset comprising 36 MSP and 36 HCS recruited in CHU Liège, Belgium, acquired with cMRI and qMRI. For each image type and region of interest, qMRI radiomic models for MS diagnosis were developed on a training subset and validated on a testing subset. Radiomic models based on cMRI were developed on the entire training dataset and externally validated on open-source datasets with 167 HCS and 10 MSP. The best diagnostic performance was achieved in the whole WM. Here the model based on magnetization transfer imaging features yielded a median area under the receiver operating characteristic curve (AUC) of 1.00 in the testing sub-cohort. Ranked by image type, the best performance was achieved by the magnetization transfer models, with median AUCs of 0.79 (0.69–0.90, 90% CI) in NAWM and 0.81 (0.71–0.90) in GM. The external validation of the T1w models yielded an AUC of 0.78 (0.47–1.00) in the whole WM, demonstrating a large 95% CI and a low sensitivity of 0.30 (0.10–0.70). This exploratory study indicates that qMRI radiomics could provide efficient diagnostic information using NAWM and GM analysis in MSP. T1w radiomics could be useful for a fast and automated check of conventional MRI for WM abnormalities once acquisition and reconstruction heterogeneities have been overcome. Further prospective validation is needed, involving more data.

4.1 Introduction

Multiple sclerosis (MS) is an inflammatory disorder of the central nervous system, responsible for focal and diffuse damages, including both demyelination and neurodegeneration, and often leading to physical and mental disability [1, 2]. In 2016, there were more than two million prevalent cases globally [3]. In Europe, the overall mean cost per patient was more than €50K (adjusted to 2015 purchasing power parity) in a severe disease [4].

Early diagnosis in MS is challenging because the pathology mechanisms are not yet completely understood, and disease biomarker discovery is still ongoing. The McDonald criteria is currently used for diagnosis [5]. It assimilates information about clinical relapses and focal white matter (WM) lesions (plaques) visualized with conventional magnetic resonance imaging (cMRI) and cerebrospinal fluid (CSF) analysis [6, 7, 8, 5]. If the patient does not meet the diagnostic criteria, the diagnosis of MS is provisionally not retained. Although cMRI is playing a valuable role in routine clinical practice, it merely captures a very small proportion of MS-related pathological processes [9, 10]. It is particularly not sensitive to detect and track diffuse pathological changes occurring both in the normal appearing white matter (NAWM) and gray matter (GM). These changes appear in the early stages of the disease and better correlate with clinical outcomes than only the WM focal lesion load [11, 12, 13, 14, 15]. Additionally, routine cMRI voxel intensities are expressed in arbitrary units, which vary based on a large number of factors, including the patient being examined, equipment, and protocol being used. This makes MRI analysis strongly dependent on the expertise of the medical specialist and hinders data reproducibility and comparison in follow-up and cross-sectional studies. Therefore, there is an unmet clinical need for the development and automated detection of quantitative and objective early MS biomarkers.

Quantitative MRI (qMRI) potentially overcomes these limitations by quantifying the physical micro-structural properties of brain tissues in standardized units. Commonly, some of the following parameters

are estimated: longitudinal and effective transverse relaxation rates (R1 and R2*, respectively) or times (T1 and T2*, respectively), proton density (PD), magnetization transfer (MT) saturation, and a number of diffusion MRI (dMRI) metrics. Values in qMRI maps are linked to the physical properties of biological tissues, such as axonal myelination (MT, R1, R2*, T1, and dMRI), iron accumulation (R2* and T2*), and free water proportion (PD) [16, 17, 18]. It has been shown that qMRI data are fairly reproducible between different scanners and attractive for multi-center studies [19]. Current MS research compares the qMRI properties of brain between healthy control subjects (HCS) and MS patients (MSP) [20, 21, 22, 13, 23, 24]. It has been shown that, with specific qMRI sequences, more MS-related damages can be detected compared with cMRI using similar acquisition times [25]. Furthermore, it has been shown that qMRI reveals pathological GM alterations [26] and early MS-related GM changes [27].
The discovery of quantitative imaging biomarkers is currently experiencing a large increase in research interest, and radiomics is rapidly emerging as a major tool in radiology. Radiomics is a high-throughput imaging data quantification approach aimed to calculate the quantitative descriptors of medical images to characterize the underlying biology and establish a correlation with clinical endpoints [28, 29, 30]. Radiomics has shown promise in personalized medicine for cancer treatment [31, 32, 33, 34] and is already applied in neurology to predict epilepsy in patients with low-grade gliomas [35], to distinguish between MS and neuromyelitis optica spectrum disorders on spine MRI [36, 37], and to differentiate Alzheimer's disease from mild cognitive impairment on MRI and positron emission tomography [38, 39]. The standard pipeline for radiomic analysis is presented in Figure 4.1.

Within the present study, we hypothesized that cMRI- and qMRI-based radiomic models have a diagnostic value in MS, while qMRI-based features have an advantage in the detection of diffuse damages. The objective of the study was to investigate the ability of radiomic features found in WM, NAWM, and GM, extracted from cMRI and qMRI maps, to distinguish between HCS and MSP.

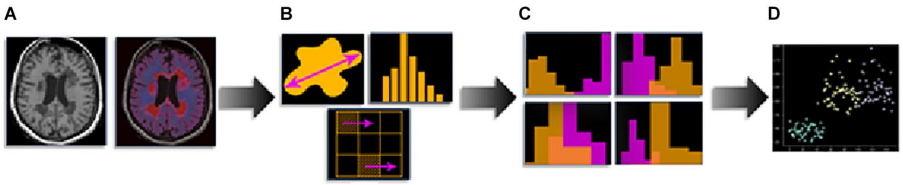


Figure 4.1: Radiomics pipeline: (A) medical imaging and segmentation, (B) feature extraction, (C) feature selection, and (D) modeling.

Radiomic classification models were developed and tested, and cMRI models were validated on external publicly available datasets.

4.2 Materials and methods

4.2.1 Study design

This study was performed on three datasets: dataset 1 (DS1) contains both cMRI (T1w and FLAIR) and four types of qMRI maps (PD, MT, R1, and R2*) of both MSP and HCS, dataset 2 (DS2) contains cMRI (T1W) of HCS, and dataset 3 (DS3) contains the cMRI of MSP (T1w and FLAIR) (see Table 4.1). DS2 and DS3 were combined into one validation dataset (DSV) using data selection and additional pre-processing to minimize any mismatch with DS1 regarding demographics and image acquisition parameters. For each participant, the same brain tissue segmentation method was applied. DS1 was randomly split and used to train and test multi-channel qMRI models as well as used for training of cMRI models, while DSV was used to validate the cMRI models. The observations from test subsets were kept apart from those of train subsets and were used only to test the models. For each participant, radiomic features were independently extracted from whole WM, NAWM, and GM regions from all available image types. For MSP, WM volume included combined NAWM and focal WM lesions. Since HCS do not have focal

WM lesions, for them WM and NAWM volumes are matching.

With the addition of models combining features extracted from all four qMRI maps, a total of 18 models were trained on DS1 [three regions of interest (ROIs), five image types, and a combination thereof], of which three models (three ROIs, one image type) were validated on DSV. All feature selection and model training were performed in the respective training datasets. The testing and/or validation datasets were kept apart and were used only for evaluation purposes. The study design is detailed in Figure 4.2. For each step, workflow execution times were recorded, and the averages reported.

4.2.2 Data description

DS1 is a private dataset consisting of 72 participants, 36 MSP with relapsing-remitting and progressive forms (CHU Liège, Belgium), and 36 HCS (GIGA-CRC in vivo imaging, University of Liège, Liège, Belgium) acquired within an MS cross-sectional study (local ethic committee approval B707201213806) retrospectively collected between 2013 and 2017 [23]. It contains cMRI data (T1w for all the participants and FLAIR only for the MSP) and qMRI maps (PD, MT, R1, and R2*; see Figure 4.3). The inclusion criteria were as follows: (1) age between 18 and 65 years, (2) Expanded Disability Status Scale (EDSS) not more than 6.5, (3) no relapse in the previous 4 weeks, and (4) MRI compatibility. The details of the MPM protocol are available in [23]. MS status was estimated by CHU Liège neurology specialists based on McDonald's criteria 2010 [40]. This dataset was used for all the exploratory analyses, including feature selection and model parameter tuning. Before the feature selection and subsequent steps, DS1 was randomly split into training and testing subsets (80/20%), attempting to maintain distributions of outcome, age, gender, and scanner variables.

DS2 is the Calgary–Campinas-359 dataset – an open, multi-vendor, multi-field-strength brain MRI dataset [41]. It is composed of volumetric T1w images of 359 presumed healthy adults, scanned between

2009 and 2016. In the dataset description, there is no information about the neurological status assessment.
DS3 is a subset of the MICCAI 2016 MS lesions segmentation (MSSEG) challenge dataset. The MSSEG challenge dataset contains MRI data for 53 MSP, but only 15 participants from the training subset are publicly available [42, 43]. The data were acquired not later than 2016 in three different sites in France on four different multi-field multi-vendor scanners with different sequences, including T1w and FLAIR. We used the unprocessed data from DS2 to implement the same image pre-processing protocol for all the datasets.
There are some differences between DS1 and DSV, the main difference being the different image acquisition equipment and protocols (see Table 4.1). Other differences are the lack of information about how HCS and MSP status, respectively, was assessed in DS2 and DS3, and the lack of MS stage of EDSS in DS3, making a comparison between DS1 and DS3 difficult. To minimize those differences and any potential bias, DS2 and DS3 were combined and filtered to match the age range and field strength present in DS1. Within the datasets, there were no incomplete data.
A summary of the datasets is presented in Table 4.1.

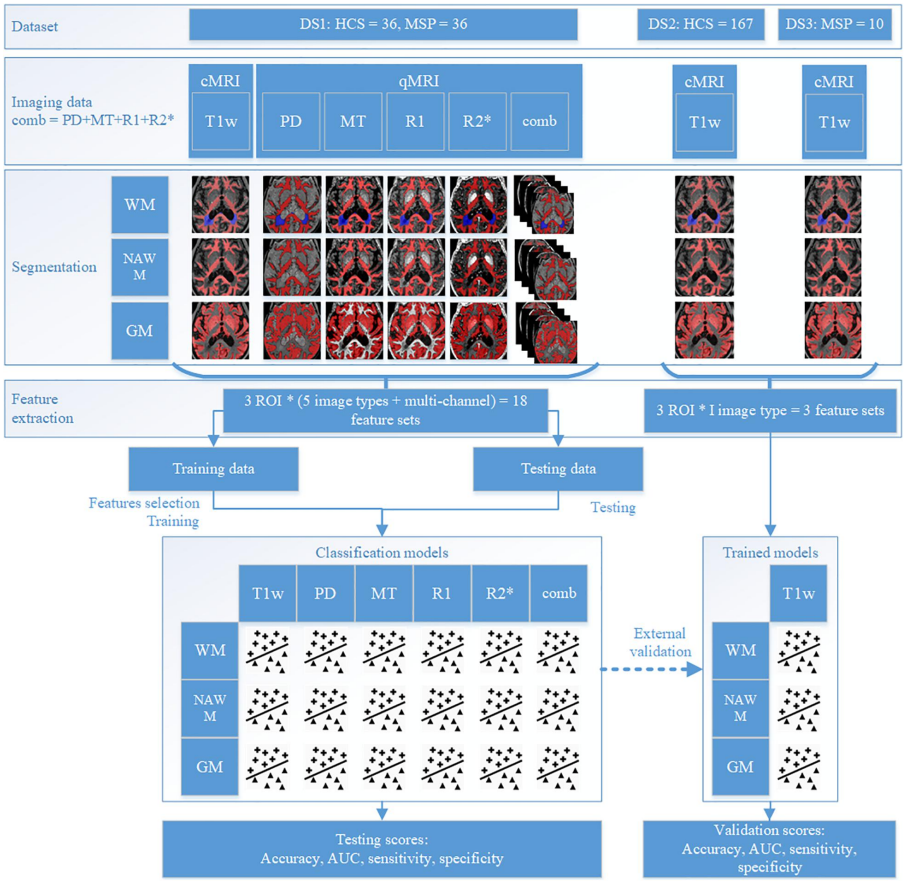


Figure 4.2: Study design.

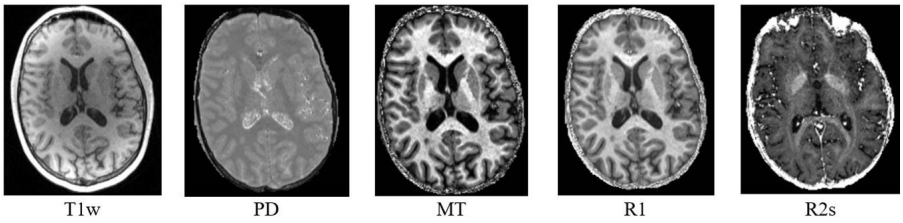


Figure 4.3: Example of MRI data presented in DS1.

Table 4.1: Data summary details.

	Dataset 1	Dataset 2	Dataset 3
Dataset	Private CHU Liège	CC-359	MICCAI 2016 MSSEG challenge (training subset)
Participants	MSP (15 relapsing-remitting, 21 progressive), HCS (36)	HCS (359)	MSP (15)
Age, $\mu \pm \sigma$ (years)	45.8 ± 12.1	52.7 ± 7.3	40.5 ± 10.8
Gender (M/F)	0.76	0.96	1.00
Image types	T1w, PD, MT, R1, R2*, FLAIR	T1w	T1w, FLAIR
Sites	CHU (Liège, Belgium); GIGA CRC In vivo imaging, University of Liège (Liège, Belgium)	Campinas (San Paulo, Brasil); Calgary (Alberta, Canada)	CHU Rennes (Rennes, France); CHU Lyon (Lyon, France)
Equipment	3T Siemens Magnetom Allegra (37) and Prisma (35)	3T and 1.5T Siemens (120), Philips (119), GE Healthcare (120) MRI scanners	3T Siemens Magnetom Verio (5); 1.5T Siemens Magnetom Aera (5); 3T Philips Ingenia (5)
Protocol	MPM protocol with FLASH sequences	3D MP-RAGE (Philips, Siemens), comparable 3D T1w spoiled gradient echo sequence (GE Healthcare)	Sagittal 3D FLAIR, sagittal 3D T1w
Matrix	256×224	224×224, 240×240, 256×256	256 256 (Siemens), 336 336 (Philips)
Slices	176	164–224	176 (Siemens), 200 (Philips)
Voxel resolution, mm×mm× mm	1×1×1	1×1×1 (Siemens)	1.08×1.08×0.9 (1.5T Siemens), 1×1×1 (3T Siemens), 0.74×0.74×0.85 (Philips)

We used the following abbreviations: μ — average, σ —standard deviation, MSP — multiple sclerosis patients, HCS — healthy control subjects, M — male, F — female.

4.2.3 MRI data pre-processing

All the data processing and analysis hereafter were performed on a system containing 4×10 core 2.40 GHz Intel Xeon CPU and 64 GB RAM.
The qMRI maps were generated in MATLAB 2017b (The MathWorks Inc., Natick, MA, United States) with the use of the hMRI toolbox, v0.2.0 [18], an extension of SPM12. In the absence of radiofrequency field sensitivity bias map acquisition, the radiofrequency field bias was corrected with a unified segmentation approach. The radiofrequency transmit field (B1) bias was corrected using B1 and B0 maps, which were acquired with 3D echo-planar imaging mapping protocols. The B1 data was processed with parameters which were identical to the standard default ones. The multiparameter input images included six MT-, eight PD-, and six T1-weighted images.
All images within DS1 were reconstructed with a resolution of 1 mm $\times$ 1 mm $\times$ 1 mm; hence, we decided to resample the scans within DS2 and DS3 to the same resolution. We used cubic spline interpolation as it performs well in terms of its Fourier properties, visual image quality, and interpolation errors [44].
Following this step, tissue masks for CSF, GM, NAWM, and lesions within DS1 were estimated. Tissue segmentation in HCS was performed with a multi-channel unified segmentation protocol [45], using multiple qMRI maps (PD, MT, R2*, and R1). It was performed in MATLAB using hMRI for SPM12 with light regularization (regularization coefficient, 0.001) and 60-mm cutoff for full-width at half-maximum of Gaussian smoothness of bias. The outputs were tissue probability maps for CSF, GM, and WM, with the voxel values between 0 (background) and 1 (corresponding brain tissue). In order to ensure the inclusion of only the relevant tissue class, binary masks for each tissue were obtained by thresholding the tissue probability maps at a high level of 0.9. For MSP, lesion masks were generated from the combination of T1w and FLAIR images with LST [46] for SPM12 by the lesion growth algorithm and corrected manually by a qualified MS specialist (ELo) when necessary. Multi-channel tissue

segmentation was performed using multiple qMRI maps (PD, MT, R2*, and R1) with unified segmentation protocol in US-with-Lesion [47], adding an extra lesion tissue class. In DSV, brain tissue segmentation was performed with a single channel (T1w) unified segmentation protocol in MATLAB with SPM12, using T1w images.
After segmentation, total intracranial volume (TIV) was estimated for each patient as the morphological sum of the CSF, GM, NAWM, and lesion volumes (where applicable). This combined ROI was used for intensity normalization, as described below.
As the magnetic field inside an MRI scanner is not ideally homogeneous and is affected by objects within it, a bias field signal is introduced, degrading image quality as a smooth, low-frequency signal that distorts segmentation results and feature values. To partially correct for this in T1w images, N4 bias field correction [48] was performed in TIV.
As cMRI voxel intensities are expressed in arbitrary units, the Image Biomarker Standardization Initiative (IBSI) recommends using normalization for raw MR data [49]. Therefore, within each T1w scan, the intensities were normalized to arrive at a mean of 0 and a standard deviation of 1. Normalization was performed within the TIV, considering only TIV intensities.

4.2.4 Radiomic feature extraction and exploration

Radiomic features that quantitatively characterize the ROI, e.g., intensity histogram, simple statistics, and texture [28, 50], were extracted from pre-processed cMRI and qMRI data using PyRadiomics 2.2.0 [51] in python v.3.7.1. Due to their small volumes, features from lesion ROIs were not extracted, and they were used only as an additional tissue class for brain segmentation. The radiomic features of the following classes were extracted from original images: FO statistics, gray-level co-occurrence matrix (GLCM) [52], gray-level run length matrix [53], gray-level size zone matrix [54], neighboring gray tone difference matrix [55], and gray-level dependence matrix

Table 4.2: Overview of the independent feature sets per participant.

ROI		Image type
WM (for MSP, NAWM + WM lesions)	cMRI	T1w
NAWM	qMRI	PD
GM		MT
		R1
		R2*
In total, three ROIs	In total, five image types	

(GLDM) [56]. Contrary to oncological radiomic studies where shape features are usually involved [28, 29, 50], here only FO and texture features were considered. Many neurodegenerative disorders have reported volumetric brain changes, showing disease-specific patterns in brain substructures [57], which were not delineated in the present study. Moreover, WM volumetric atrophy changes are mostly explained with the presence of lesions [58], which also influence first-order and texture features. Therefore, to further reduce the ratio of the number of features vs. the number of samples, shape features were excluded. Before gray-level texture matrices were calculated, intensity discretization was performed with a fixed number of bins $N_{bins} = 50$, in line with IBSI recommendations [49]. The fixed bin number approach groups voxel intensities before discretization, which additionally harmonizes multi-scanner multi-vendor multi-site data.

No feature harmonization methods, such as ComBat [59], were applied across the different datasets because of the small sample sizes and considerable heterogeneity of scanners and protocols. To speed up feature extraction, the ROI was pre-cropped into a bounding box with 5-voxel-width padding. A separate feature set was calculated for each ROI and image type. An overview of the feature sets is presented in Table 4.2.

Feature analysis was performed in the whole DS1 to describe the data;

its results were not included into model building. Statistical tests were performed to gauge diagnostic efficacy in such a small dataset. A univariate Mann–Whitney test was carried out using Bonferroni correction, and $p \leq 0.01$ for two-sided hypothesis was considered statistically significant. Point-biserial correlation coefficients r_{pb} and p-values were calculated between radiomic feature values and MS status; a correlation was considered statistically significant if $|r_{pb}| \geq 0.85$ and $p \leq 0.05$. Spearman correlations between the features and age and the feature ROI volume were computed to gauge the added value of radiomic features compared to age and volumetry, with $|r_S| > 0.85$ considered highly correlated for each test. Additionally, the univariate area under the receiver operating characteristic curve (AUC) was calculated for each feature.

4.2.5 Radiomic feature selection

In order to remove redundant and non-informative features, feature reduction and selection were performed on DS1, using the MS status as the binary outcome where applicable. Feature selection was independently carried out for the T1w, PD, MT, R1, and R2* maps to arrive at a subset of N features each, attempting to adhere to published rules of thumb to estimate the optimal number [60, 61]. We chose the following approach to estimate the number of features $N_{features} = int(\frac{N_S}{10})$, as outlined in [61], where N_S is the number of samples in the minor class.

Since DS1 is relatively small, especially after the train/test split, feature selection as described below was performed 100 times on an extended and balanced cohort of 100 participants created by randomly sampling (with replacement) observations from the training subset. In each of the 100 iterations, a fixed number N of the highest-ranking features was retained, and at the end the features were ranked according to how often they were selected.

The feature selection pipeline starts with excluding features with zero or low variance. A feature was considered of low variance if the per-

centage of its distinct values out of the number of observations was less than 10% and the ratio of its most frequent values was more than $95/5$. Next, features with high inter-correlation were excluded by calculating the pairwise Spearman correlation between all the features. From each pair of highly correlated features ($|r_S| > 0.85$), we excluded a feature having the highest correlation on average with all the remaining features. The final selection was performed with recursive feature elimination [62] using random forest classifier [63] models [100 trees, as recommended by [64]; a number of features to consider when looking for the best split $int(\sqrt{N_{features}})$, where $\sqrt{N_{features}}$ is changing during recursive feature elimination iterations, as recommended by [65]]. Random forest (RF) classifiers allow for robust variable importance computation and do not need normalization. Moreover, the number of available features exceeds the number of samples, and a random forest classifier is still able to deal with such data. For each selected feature, a distribution map was generated by calculating the feature value within each 26 connected neighborhood of each voxel within the image ROIs.

4.2.6 Model training and testing

Models were trained and tested on independent subsets of DS1. Observations from the training and testing subsets were randomly sampled with replacement for 100 times, resulting in the creation of extended and balanced training and testing cohorts. Every cohort contained 100 participants.
Separate binary classification models were trained on DS1 for different image types, T1w, PD, MT, R1, and R2*, and for a combination of features from PD + MT + R1 + R2* (composed of qMRI) to investigate the value of each image type and ROI in the estimation of the MS status. For each image type, three binary classification models were trained using the same features from each image type and ROI: (i) random forest (RF), (ii) support vector machine (SVM) [66], and (iii) logistic regression (LR). For the RF model, the same settings as for the recursive

feature elimination were used; for SVM, a radial basis function kernel was used with regularization parameter C = 1.0, kernel coefficient $\gamma = 1/(N_{features} \times Var(X)))$, where $Var(X)$ is the variance of the input feature X (since we did not have any a priori expert knowledge about the classification problem and did not perform any empirical validation of the model parameters, these are the default parameters for the SVM, keeping a balance between classification accuracy and tolerance to misclassification errors), and for LR, L2 penalty was used since this regularization does not lead to high values among the regression coefficients, with dual formulation, as recommended when the amount of observations exceeds the amount of features, and a liblinear solver, which is recommended for small datasets; inverse of regularization strength $C = 1.0$, which is optimal in terms of balance between accuracy and model complexity. Due to the small dataset sizes, DS1 was used again as an exploratory dataset.

The performances of the models were estimated in terms of the following metrics: accuracy, sensitivity, specificity, and AUC, with the corresponding 90% confidence intervals (CI); for each model, learning and curves were plotted. Since all the scores were estimated on the data subsets, containing equal numbers of HCS and MSP, the imbalanced data correction was not needed. The best model was selected based on these performance metrics for different ROIs and tissue types, giving the AUC score more weight and excluding models with median AUC scores below the threshold of 0.7, which is considered an underperforming classification model. In order to select the best model type (RF, SVM, or LR), the number of highest AUC scores was used.

The final models with the original coefficients were subsequently validated on DS2 and DS3. As the combined dataset containing DS2 and DS3 was highly unbalanced regarding the outcome, bootstrapping with balanced sampling was implemented. The models for qMRI were not validated externally due to the unavailability of similar datasets.

To examine the models and methodology for overfitting, a permutation test was performed on DS1. The class labels in both training and testing subsets were randomized, maintaining the same distributions as in the original subsets. Without modifying the pipeline, feature se-

lection was performed, models were trained and tested, and performance metrics were calculated to ascertain whether the pipeline detects patterns in randomly generated outcomes.

4.3 Results

4.3.1 Data description and MRI data pre-processing

Participants were drawn from DS2, aiming to match DS1 regarding age and magnetic field strength. Participants with MRI quality, which was not sufficient for robust automatic segmentation, were excluded after a visual check (ELa). Finally, 167 participants were selected from this dataset. Another 10 participants were selected from DS3, again trying to match the age and field strength distributions with those of DS1. An overview of the resulting feature sets is presented in Table 4.3.

4.3.2 Radiomic feature extraction and description

For each T1w and qMRI image and ROI combination, 93 features were extracted, resulting in 1395 features per participant. The Mann–Whitney test revealed that 16% of the features (220 features out of 1395) were sampled from significantly different distributions in the HCS and MSP cohorts, mostly originating from WM in all image types but also from NAWM in MT and R2*. In the entire feature set, there was only one feature (R1 first-order minimum in WM) that was highly correlated with the outcome, no feature was highly correlated with age, and 10 features out of 1395 were highly correlated with ROI volume. A univariate analysis showed that 28% of the features (395 features out of 1395) had an area under the receiver operating characteristic curve (ROC AUC) score > 0.75, most of which were obtained from the PD, MT, and R2* maps (see Table 4.4).

Table 4.3: Dataset summary details for the included participants.

	Dataset 1	Dataset 2	Dataset 3
Participants	MSP (15 RR, 21 progressive), HCS (36)	HCS (167)	MSP (10)
Equipment	3T Siemens Magnetom Allegra (37) and Prisma (35)	3T Siements (53), Philips (54), GE Healthcare (60) MRI scanners	3T Siemens Magnetom Verio (5), 3T Philips Ingenia (5)
Age, $\mu \pm \sigma$ (years)	45.8 ± 12.1	52.7 ± 7.3	40.5 ± 10.8
Gender (M/F)	0.76	0.96	1.00

We used the following abbreviations: μ — average, σ — standard deviation, MSP — multiple sclerosis patients, RR — relapsing-remitting, HCS — healthy control subjects, M — male, F — female.

Table 4.4: Number of features out of 1395 with age, volume, and outcome correlations having $|r_S| > 0.85$ as well as univariate AUC >0.75 and corrected Mann-Whitney $p < 0.01$.

	ROI	T1w	PD	MT	R1	R2*
$\|r_S^{age}\| > 0.85$	WM	0	0	0	0	0
	NAWM	0	0	0	0	0
	GM	0	0	0	0	0
$\|r_S^{volume} > 0.85\|$	WM	0	3	1	1	0
	NAWM	0	3	1	1	0
	GM	0	0	0	0	0
$\|r_{pb}^{outcome}\| > 0.85$	WM	0	0	0	1	0
	NAWM	0	0	0	0	0
	GM	0	0	0	0	0
$AUC_{univar} > 0.75$	WM	13	62	21	45	52
	NAWM	8	28	57	9	37
	GM	3	7	26	5	22
$p_{Mann-Whitney}^{Bonferroni} < 0.01$	WM	9	41	10	37	7
	NAWM	0	12	42	5	26
	GM	1	0	18	2	10

4.3.3 Radiomic feature selection

In the training subset of DS1, on average among all the image types and ROIs, 7% from the initial feature set were excluded by the low variance step, followed by 79% exclusion by the high correlation step. The RF-based recursive feature elimination using data sampling with replacement yielded the final feature vectors for each ROI and MRI image type. To make the models easier to compare across ROI and MRI image types, the three ($N_{features} = \frac{N_S}{10} = \frac{28}{10} = 3$) top-ranking features were left in each final feature vector.
No high correlations were discovered between the selected features, age, and ROI volume. For the selected features, the univariate AUC was below a threshold of 0.7 for PD, MT, and R2* in NAWM and T1w and for PD in GM. According to the Mann–Whitney test, the highest number of features with significant differences in means in HCS and MSP is discovered in WM (15 features out of 15) when ranking by ROIs and on R1 (eight features out of nine) when ranking by image types. A list of the selected features with their univariate ROC AUC scores is presented in Figure 4.4. For the best features in each ROI and image type, saliency maps were obtained by calculation of the feature value in the neighborhood of each voxel. Examples of the normalized saliency maps are presented in Figure 4.5.

4.3.4 Model training and testing

According to the DeLong test with use of the Bonferroni correction, different ML models had significantly different ($p = 0.01$) AUC scores in all the cases, with the exception of MT and $qMRI_{comb}$ in WM, R1 in NAWM, and PD in GM. Among all the ROI and image types, in most cases, the median values of the RF classifier performance scores dropped below a threshold of 0.7. Having the highest number of top AUC values, the LR model was selected. Results from the LR model will be shown in the main body of the text. The performance metrics are presented in Table 4.5.

Models using features extracted from WM achieved the best classification performance, with the best performance achieved by the MT data. There were no statistical differences ($p \leq 0.01$) in AUC scores obtained for WM in MT, R1, and $qMRI_{comb}$. The highest median performance across all metrics was achieved with the MT model, all of which yielded a value of 1.00. The T1w model performed generally lower than the MT and combined qMRI models but outperformed the PD model in median specificity, the R1 model in median sensitivity, and the R2* model in median accuracy and AUC.
In NAWM, there were no significant differences in AUC scores obtained for highest scoring R2* and $qMRI_{comb}$ models. The highest overall performance was achieved with the R1 model. The PD model yielded a median specificity of 0.00 (no true negatives were achieved). The T1w model performed generally poorer than the MT and R1 models but better than the PD, R2*, and $qMRI_{comb}$ models.
In GM, there were no significant differences in AUC scores obtained for MT and R1 and for R2* and $qMRI_{comb}$. The highest overall performance was achieved with the MT-based model, which yielded a median accuracy of 0.88.
The permutation test results showed a significant ($p \leq 0.01$) drop in AUC for all the models, except for PD and MT in NAWM and for T1w in GM.
The classification performance metrics T1w models using the WM, NAWM, and GM validated on the external DSV are presented in Table 4.6. Since DS2 comprises only MS negative outcomes and DS3 only MS positive outcomes, the separate accuracies for DS2 and DS3 are equal to specificity and sensitivity, respectively, on the whole validation data. The medial validation model accuracy for DS2 is 1.00 in WM and NAWM and 0.00 in GM; the medial validation model accuracy for DS2 is 0.30 in WM, 0.20 in NAWM, and 0.90 in GM.

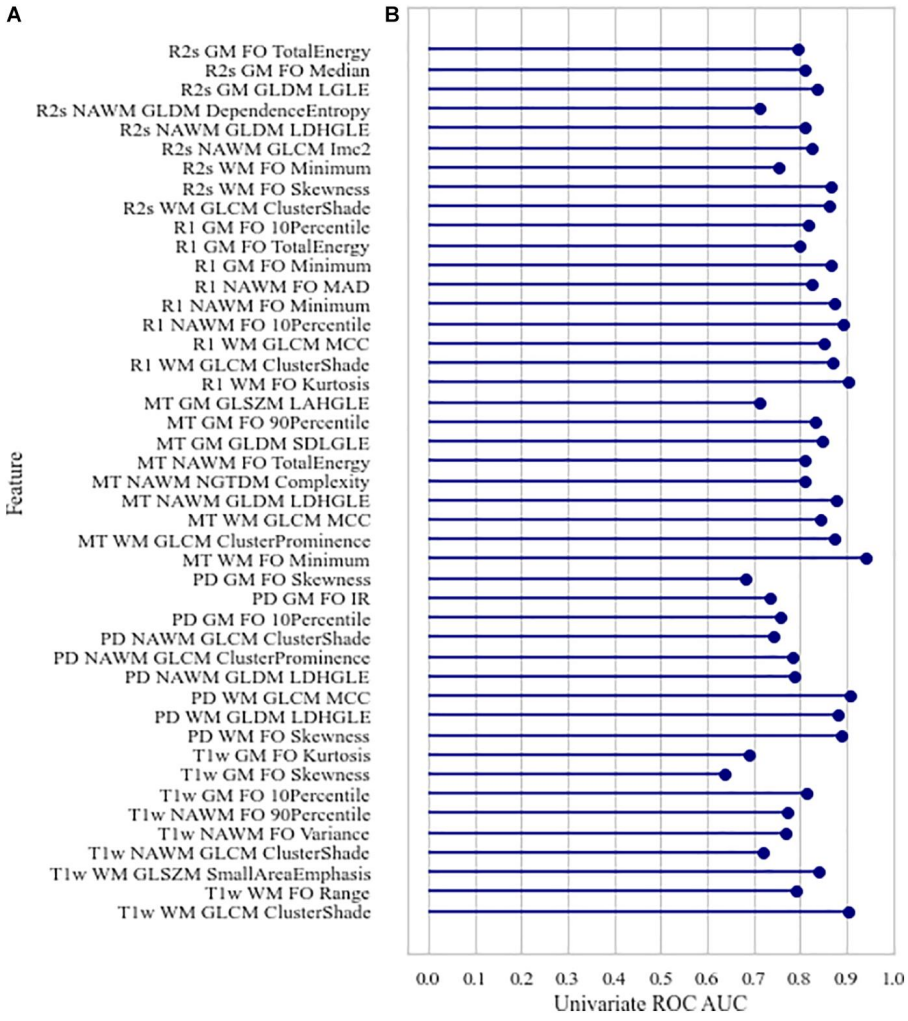


Figure 4.4: Univariate area under the receiver operating characteristic curve for the selected features; we used the following abbreviations: FO — first order, LDHGLE — large dependence high gray level emphasis, SDLGLE — small dependence low gray level emphasis, LAHGLE — large area high gray level emphasis, MAD — mean absolute deviation, LGLE — low gray level emphasis.

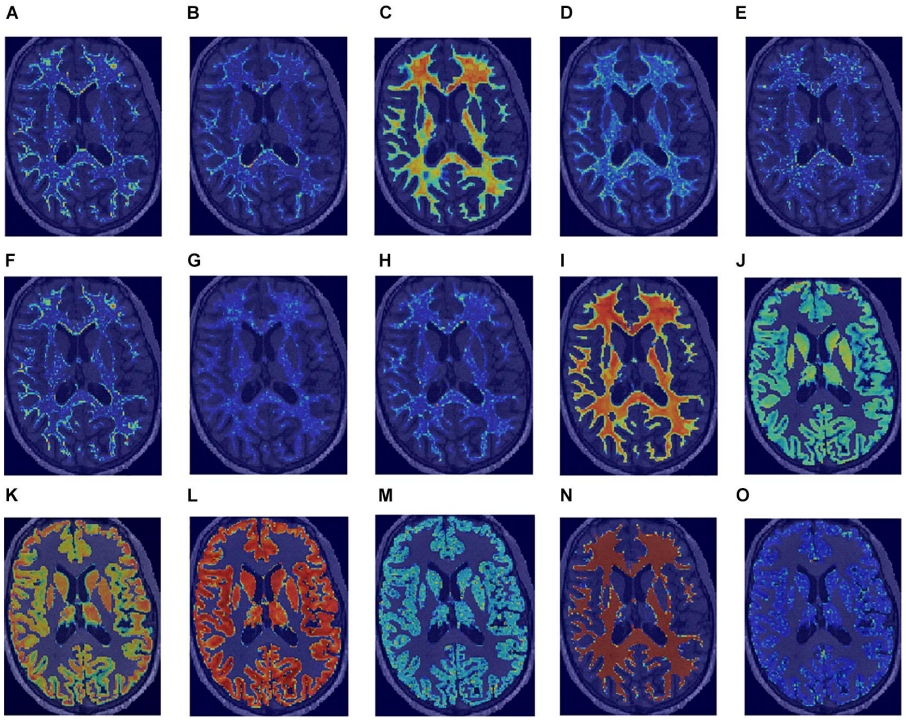


Figure 4.5: Normalized saliency maps for the best selected features for each region of interest and image type highlight the areas with the highest feature values: (A) T1w GLCM cluster shade in WM, (B) PD first-order skewness in WM, (C) MT first-order minimum in WM, (D) R1 first-order kurtosis in WM, (E) R2* GLCM cluster shade in WM, (F) T1w GLCM cluster shade in NAWM, (G) PD GLDM large dependence high gray level emphasis in NAWM, (H) MT GLDM large dependence high gray level emphasis in NAWM, (I) R1 first-order 10-percentile in NAWM, (J) R2* GLCM Imc2 in NAWM, (K) T1w first-order 10-percentile in GM, (L) PD first-order 10-percentile in GM, (M) MT GLDM small dependence low gray level emphasis in GM, (N) R1 first-order minimum in GM, and (O) R2* GLDM low gray level emphasis in GM.

Table 4.5: Logistic regression model performances on testing data showing the median (90% CI) for each image and tissue type (ROI) (median values above 0.7 for all the performance metrics for the same model are highlighted with **bold** font).

ROI	Image	Accuracy	AUC	Sensititity	Specificity
WM	T1w	**0.74** (0.66, 0.82)	**0.90** (0.84, 0.95)	**0.76** (0.67, 0.86)	**0.72** (0.59, 0.82)
	PD	0.64 (0.58, 0.71)	**0.98** (0.95, 1.00)	**1.00** (1.00, 1.00)	0.28 (0.17, 0.42)
	MT	**1.00** (1.00, 1.00)	**1.00** (1.00, 1.00)	**1.00** (1.00, 1.00)	**1.00** (1.00, 1.00)
	R1	**0.82** (0.76, 0.88)	**1.00** (1.00, 1.00)	0.64 (0.52, 0.75)	**1.00** (1.00, 1.00)
	R2*	**0.73** (0.63, 0.83)	**0.86** (0.78, 0.93)	**0.76** (0.62, 0.86)	**0.72** (0.58, 0.84)
	$qMRI_{comb}$	**0.93** (0.88, 0.97)	**1.00** (1.00, 1.00)	**1.00** (1.00, 1.00)	**0.86** (0.77, 0.94)
NAWM	T1w	**0.73** (0.66, 0.82)	**0.86** (0.77, 0.93)	**0.76** (0.64, 0.87)	**0.70** (0.59, 0.81)
	PD	0.37 (0.30, 0.44)	0.67 (0.55, 0.81)	**0.74** (0.60, 0.87)	0.00 (0.00, 0.00)
	MT	**0.81** (0.74, 0.89)	**0.79** (0.69, 0.90)	**0.76** (0.64, 0.87)	**0.86** (0.77, 0.94)
	R1	**0.87** (0.80, 0.93)	**0.97** (0.93, 0.99)	**0.88** (0.77, 0.98)	**0.86** (0.77, 0.94)
	R2*	0.66 (0.56, 0.76)	**0.83** (0.73, 0.94)	**0.76** (0.64, 0.87)	0.56 (0.40, 0.72)
	$qMRI_{comb}$	**0.74** (0.67, 0.81)	**0.82** (0.73, 0.90)	0.62 (0.48, 0.77)	**0.86** (0.77, 0.94)
GM	T1w	0.41 (0.32, 0.52)	0.60 (0.47, 0.73)	0.26 (0.16, 0.40)	0.56 (0.43, 0.71)
	PD	0.69 (0.61, 0.79)	**0.83** (0.74, 0.91)	0.51 (0.38, 0.66)	**0.86** (0.77, 0.94)
	MT	**0.88** (0.82, 0.94)	**0.81** (0.71, 0.90)	**0.76** (0.64, 0.87)	**1.00** (1.00, 1.00)
	R1	**0.82** (0.75, 0.87)	**0.81** (0.72, 0.88)	0.64 (0.50, 0.74)	**1.00** (1.00, 1.00)
	R2*	**0.73** (0.65, 0.83)	**0.86** (0.78, 0.95)	**0.76** (0.64, 0.87)	**0.71** (0.58, 0.84)
	$qMRI_{comb}$	**0.81** (0.73, 0.88)	**0.86** (0.78, 0.93)	**0.76** (0.64, 0.87)	**0.84** (0.77, 0.95)

Table 4.6: Logistic regression model performances on external validation dataset showing the median (90% CI) for each tissue type for T1w images.

ROI	Accuracy	AUC	Sensitivity	Specificity
WM	0.65 (0.55, 0.85)	0.78 (0.47, 1.00)	0.30 (0.10, 0.70)	1.00 (0.90, 1.00)
NAWM	0.60 (0.55, 0.95)	0.65 (0.29, 1.00)	0.20 (0.10, 1.00)	1.00 (0.90, 1.00)
GM	0.45 (0.15, 0.45)	0.24 (0.05, 0.56)	0.90 (0.10, 0.90)	0.00 (0.00, 0.30)

4.3.5 TRIPOD Statement and Radiomics Quality Assurance

This study was evaluated with the Radiomics Quality Score (RQS) [29], which yielded a final result of 39%. We likewise evaluated it with the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) [67] checklist score, which was in the range of 0.71-0.77%.

4.4 Discussion

In this exploratory brain tissue MRI and qMRI radiomics study based on a unique dataset, we report on several hypothesis-generating findings for HCS vs. MSP classification. Previous studies on radiomics in MS have been focused on T2w cMRI data and aimed to distinguish between MS and neuromyelitis optica spectrum disorder [36, 37] without external validation, hence the importance of this work.

Of the three machine learning models (RFC, SVM, and LR) tested, LR was the most stable with median accuracy, AUC, sensitivity, and specificity all exceeding a value of 0.7 while achieving the highest performance in terms of AUC. The fact that LR outperformed the other models could be due to the small number of observations, where the simplest models might perform best since they are less likely to overfit. The selected radiomics features were not correlated with age and volume (also a radiomic feature), which indicates that radiomics could provide additional information to those simple variables.

The best LR model performance concerning tissue type was achieved using features extracted from WM. This was expected since focal WM lesions (plaques) in the WM of MSPs affect the intensity distribution [6]. In NAWM classification, which is more challenging, good classification is achieved not only with MT and R1 maps but also with T1w data. This result was not expected since this MRI sequence is not sensitive to pathological NAWM changes within, as reported in [6] and [21]. Nevertheless, it could be explained by the fact that qMRI voxel values have a physical meaning, reflecting the water and myelin contents. Furthermore, the qMRI map generation pipeline contains image co-registration and B0 and B1 fields correction steps, leading to interpolation and, therefore, smoothing of the qMRI map. Consequently, T1w images have a higher spatial resolution, leading to a more detailed texture analysis. In GM, the T1w-based model underperforms, as it was expected, according to previous publications [6, 21].

Among the image types, the best performance was achieved with MT maps, which corroborates the findings of [23], where statistical tests showed considerable differences between HCS and MSP. In WM, the MT model demonstrated median accuracy, AUC, sensitivity, and specificity of 1.00, which means that all the training observations were classified correctly. As far as training observations did not enter model training, we can conclude that, in our relatively small dataset, the presence of focal WM lesions (plaques) makes the selected MT features distinctive from the ones extracted from the healthy brain. The PD maps showed the poorest performance with at least one of the performance metrics crossing below a value of 0.7 in each tissue type. This could be due to the potential residual T2* weighting, as mentioned previously [23]. The results obtained with T1w and R1 data were significantly different, although both these image types represent longitudinal relaxation. The main difference between them is that T1w demonstrates the relative level of longitudinal relaxation at some moment, expressed in arbitrary units, whereas the R1 map represents the actual physical property of the tissue and is expressed in standardized physical units (Hz). Furthermore, unlike for T1w

data, reconstruction of the qMRI images is always performed with the correction of instrumental biases and receive fields [18].
Although the T1w models are non-quantitative, they outperformed some of the qMRI models in WM and NAWM yet had the poorest performance in GM. Among all the T1w models, the WM model yielded the highest median AUC of 0.78 and an underperforming sensitivity with a median value of 0.30 on the testing subset of the development dataset. On the external validation for T1w-based models, all showed a poor performance. Nevertheless, among these models, the best performance was achieved in WM, mainly due to the presence of focal WM lesions, which are easily captured in the radiomic analysis. In NAWM and GM, the differences between HCS and MSP are presented on the microstructural level. The T1w data is expressed in arbitrary units, and it is not consistent enough to detect these changes within different scanners and centers. As the T1w-based model in GM underperformed on the testing data, a good performance on the validation dataset was not expected. Thus, even though T1w data can perform well on the development dataset, its application is challenging for multi-centric studies. The explanation can be due to differences in imaging data, lack of sensitivity of T1w contrast for these applications, low predictive ability of the corresponding features, and their susceptibility to data effects. Additionally, we suspect a bias that can be introduced by the clinical differences in the cohorts in DS1, DS2, and DS3. Whereas MS status assessment details, EDSS, and MS stage are known for DS1, there is no such information about the participants from DS3, and there is no information about the tests carried out for DS2 participants to determine them as HCS.
The strengths of the current study include the use of unique quantitative and reproducible imaging data, the use of an external validation open-source data, and in-depth investigation of the features in traditionally challenging tissues such as NAWM and GM, which can have potential in early MS diagnosis.
This study has some limitations, too. The first stems from the small number of observations in the DS1. Consequently, for external

validation, we excluded participants which did not correspond to the participants from DS1 in terms of age or MRI magnetic field strength. All participants with insufficient MRI data quality, rendering it unsuitable for robust automatic brain tissue segmentation, were also excluded, introducing more bias. Another limitation is related to the uniqueness of qMRI data, which means that there are no available similar qMRI brain datasets for external validation, especially for MSP. However, it was reported that qMRI is reproducible between different scanner models, and multi-center studies can be expected [19]. The third limitation is the absence of data harmonization performed across datasets involved in this study. It results in non-uniformity of non-quantitative MRI data between datasets and thus leads to model performance degradation. The next limitation is related to the analysis of only HCS and MSP data. Although the exploratory analysis of the features demonstrated that some had very high univariate AUC scores (¿0.99), absence of data for other neurodegenerative diseases, and relatively small amounts of observations, we cannot conclude that these features themselves can be reliable biomarkers in MS. Thus, an analysis of other neurodegenerative disorders is needed to distinguish between different diagnoses. The fifth limitation pertains to the cMRI sequence analyzed in this study: even though focal WM lesions are noticeable on T1w, this image type is not the leading one in MS investigation. Among cMRI modalities, T2w, FLAIR, and contrast-enhanced T1w provide appropriate contrast. These modalities were not available for all the participants of DS1 (with qMRI acquisition): FLAIR scans were available for MSP only. Therefore, an analysis of another cMRI and qMRI could be a subject of future research. Finally, different brain segmentation approaches were used for DS1 and external validation data. Even though the same method was implemented for all the MRIs, segmentation in DS1 was performed with qMRI data, while segmentation for external validation was performed with cMRI data. It could affect the values of radiomic features, as cMRI-based segmentation leads to an inaccurate delineation of deep GM regions [16, 23].

Within the present study, we used standard open-source tools for data

pre-processing and analysis. Thus, the diagnostic support workflow execution times obtained within this study are indicative. Moreover, they strongly depend on the hardware and software used, original medical image parameters, pre-processing and analysis settings, and radiomic features, composing the final signature. We did not implement any optimization of computational resource consumption; therefore, the obtained execution times represent the upper bound of a workflow duration. Within the present study, cMRI- and qMRI-based workflows took approximately up to 26 and 38 min per participant, excluding the image acquisition time. This difference is due to the relatively long time of qMRI map reconstruction. This shows that the cMRI workflow can be implemented into the brain scanning protocols as a screening for WM abnormalities. The qMRI workflow requires a particular scanning protocol [16] and a relatively long analysis time. Therefore, it can be implemented for diagnostic support for patients with suspicious medical evidence.

This study indicated the potential of cMRI and qMRI radiomics in MS-related biomarker development. In differentiating between MSP and HCS, qMRI showed the advantage over cMRI in NAWM and GM regions. Therefore, application of qMRI is promising in early MS diagnosis. We believe that qMRI radiomic signatures can contribute to multi-center studies, as also indicated in previous works [16, 17, 23, 18]. For this, the reproducibility of qMRI features is to be investigated in the future. T1w WM analysis could potentially be applied for a rapid check of cMRI for WM abnormalities. For research purposes, 7 T MRI is often applied to study NAWM and GM [15, 68], but it is not widely used in clinical practice yet. We believe that 7 T MRI radiomic analysis is a potential research field in MS diagnosis.

Our next step is to validate those findings in a prospective qMRI study and test the hypothesis that those signatures are sensitive to neurodegenerative changes in the early stages of MS and have a diagnostic value for subjects at risk (e.g., clinically isolated syndrome).

4.5 Conclusion

This study demonstrates that brain cMRI and qMRI radiomic features have the potential to distinguish between MSP and HCS. In NAWM and GM analysis, having a potential in early automated diagnosis, stable results are achieved with qMRI-based data. This is a proof-of-concept clinical study demonstrating a strong signal in brain imaging, but further research is needed to develop and approve radiomic signatures for MS.
Nevertheless, future large-scale studies should evaluate the reproducibility and generalizability of the proposed method and create an MS-specific radiomic signature. Because of fully automated pipeline and imaging data quantification, the proposed approach shows its potential in relevance to time-saving and reproducibility in MS diagnosis.

4.6 Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions DS2 and DS3 are public datasets, the accession details can be found in [41] and [39]. DS1 MRI data cannot be shared publicly. The code to perform the analysis and radiomic features values are publically available from GitHub URL: https://github.com/CyclotronResearchCentre/brain-tissue-radiomics-on-clinical-and-quantitative-MRI-for- Requests to work with the DS1 on a collaborative basis should be directed to ELo, elommers@chuliege.be. The Supplementary Material for this article contains intermediate and negative results as well as TRIPOD and RQS checklists and it can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.679941. Supplementary material contains: detailed image acquisition parameters for DS1 (Table 1), demographic data for DS1 (Table 2), qMRI parameters calculated in DS1 (Table 3, [23]), comparison of age and sex distributions in HCS and MSP as well

as in development and validation groups (Table 4), list of extracted radiomic features (Table 5), participant distributions in development and validation groups (Table 6), number of features per feature set kept after every feature selection step (Table 7), list of the selected features (Table 8), Spearman correlation coefficients of the selected features with age and ROI volume (Figure 1), univariate ROC AUC scores and correlations with the mean ROI intensity for the selected features (Table 9), DeLong p-values for LR, SV, and RF classifiers (Table 10), performance comparison for LR, SV, and RF classifiers (Table 11), LR coefficients for the trained models (Table 12), DeLong test p-values for different qMRI maps (Table 13), permutation test results (Table 14), permutation test p-values (Table 15), univariate ROC AUC scores for demographic and clinical variables (Table 16), RQS checklist (Table 17), TRIPOD checklist (Table 18), and list of the software packages used (Table 19).

References

[1] Hans Lassmann. "Multiple sclerosis pathology". In: *Cold Spring Harbor perspectives in medicine* 8.3 (2018). DOI: 10.1101/cshperspect.a028936.

[2] Jing Chen et al. "Estimating MS-related work productivity loss and factors associated with work productivity loss in a representative Australian sample of people with multiple sclerosis". In: *Multiple Sclerosis Journal* 25.7 (2019), pp. 994–1004. DOI: 10.1177/1352458518781971.

[3] Mitchell T Wallin et al. "Global, regional, and national burden of multiple sclerosis 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016". In: *The Lancet Neurology* 18.3 (2019), pp. 269–285. DOI: 10.1016/S1474-4422(18)30443-5.

[4] Gisela Kobelt et al. “New insights into the burden and costs of multiple sclerosis in Europe”. In: *Multiple Sclerosis Journal* 23.8 (2017), pp. 1123–1136. DOI: 10.1177/1352458517694432.

[5] Alan J Thompson et al. “Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria”. In: *The Lancet Neurology* 17.2 (2018), pp. 162–173. DOI: 10.1016/S1474-4422(17)30470-2.

[6] SA Trip and David H Miller. “Imaging in multiple sclerosis”. In: *Journal of Neurology, Neurosurgery & Psychiatry* 76.suppl 3 (2005), pp. iii11–iii18. DOI: 10.1136/jnnp.2005.073213.

[7] Ulrike W Kaunzner and Susan A Gauthier. “MRI in the assessment and monitoring of multiple sclerosis: an update on best practice”. In: *Therapeutic advances in neurological disorders* 10.6 (2017), pp. 247–261. DOI: 10.1177/1756285617708911.

[8] Jiwon Oh, Angela Vidal-Jordana, and Xavier Montalban. “Multiple sclerosis: clinical aspects”. In: *Current opinion in neurology* 31.6 (2018), pp. 752–759. DOI: 10.1097/WCO.0000000000000622.

[9] Robert Zivadinov and Thomas P Leist. “Clinical–magnetic resonance imaging correlations in multiple sclerosis”. In: *Journal of Neuroimaging* 15 (2005), 10S–21S. DOI: 10.1177/1051228405283291.

[10] Massimo Filippi et al. “Association between pathological and MRI findings in multiple sclerosis”. In: *The Lancet Neurology* 18.2 (2019), pp. 198–210. DOI: 10.1016/S1474-4422(18)30451-4.

[11] Colette M Griffin et al. “The relationship between lesion and normal appearing brain tissue abnormalities in early relapsing remitting multiple sclerosis”. In: *Journal of neurology* 249 (2002), pp. 193–199. DOI: 10.1007/pl00007864.

[12] Guillaume Bonnier et al. "Advanced MRI unravels the nature of tissue alterations in early multiple sclerosis". In: *Annals of clinical and translational neurology* 1.6 (2014), pp. 423–432. DOI: 10.1002/acn3.68.

[13] Youngjin Yoo et al. "Deep learning of joint myelin and T1w MRI features in normal-appearing brain tissue to distinguish between multiple sclerosis patients and healthy controls". In: *NeuroImage: Clinical* 17 (2018), pp. 169–178. DOI: 10.1016/j.nicl.2017.10.015.

[14] Nikunj Davda, Emma Tallantyre, and Neil P Robertson. "Early MRI predictors of prognosis in multiple sclerosis". In: *Journal of Neurology* 266 (2019), pp. 3171–3173. DOI: 10.1007/s00415-019-09589-2.

[15] Constantina A Treaba et al. "Longitudinal characterization of cortical lesion development and evolution in multiple sclerosis with 7.0-T MRI". In: *Radiology* 291.3 (2019), pp. 740–749. DOI: 10.1148/radiol.2019181719.

[16] Nikolaus Weiskopf et al. "Quantitative multi-parameter mapping of R1, PD*, MT, and R2* at 3T: a multi-center validation". In: *Frontiers in neuroscience* 7 (2013), p. 95. DOI: 10.3389/fnins.2013.00095.

[17] Nikolaus Weiskopf et al. "Advances in MRI-based computational neuroanatomy: from morphometry to in-vivo histology". In: *Current opinion in neurology* 28.4 (2015), pp. 313–322. DOI: 10.1097/WCO.0000000000000222.

[18] Karsten Tabelow et al. "hMRI–A toolbox for quantitative MRI in neuroscience and clinical research". In: *Neuroimage* 194 (2019), pp. 191–210. DOI: 10.1016/j.neuroimage.2019.01.029.

[19] René-Maxime Gracien et al. "How stable is quantitative MRI?–Assessment of intra-and inter-scanner-model reproducibility using identical acquisition sequences and data

analysis programs". In: *NeuroImage* 207 (2020), p. 116364. DOI: 10.1016/j.neuroimage.2019.116364.

[20] A Hagiwara et al. "Analysis of white matter damage in patients with multiple sclerosis via a novel in vivo MR method for measuring myelin, axons, and g-ratio". In: *American Journal of Neuroradiology* 38.10 (2017), pp. 1934–1940. DOI: 10.3174/ajnr.A5312.

[21] Sarah C Reitz et al. "Multi-parametric quantitative MRI of normal appearing white matter in multiple sclerosis, and the effect of disease activity on T2". In: *Brain imaging and behavior* 11 (2017), pp. 744–753. DOI: 10.1007/s11682-016-9550-5.

[22] C Andica et al. "Automated brain tissue and myelin volumetry based on quantitative MR imaging with various in-plane resolutions". In: *Journal of neuroradiology* 45.3 (2018), pp. 164–168. DOI: 10.1016/j.neurad.2017.10.002.

[23] Emilie Lommers et al. "Multiparameter MRI quantification of microstructural tissue alterations in multiple sclerosis". In: *NeuroImage: Clinical* 23 (2019), p. 101879. DOI: 10.1016/j.nicl.2019.101879.

[24] Laetitia Saccenti et al. "Brain tissue and myelin volumetric analysis in multiple sclerosis at 3T MRI with various in-plane resolutions using synthetic MRI". In: *Neuroradiology* 61 (2019), pp. 1219–1227. DOI: 0.1007/s00234-019-02241-w.

[25] A Hagiwara et al. "Synthetic MRI in the detection of multiple sclerosis plaques". In: *American journal of Neuroradiology* 38.2 (2017), pp. 257–263. DOI: 10.3174/ajnr.A5012.

[26] Emilie Lommers et al. "Voxel-Based quantitative MRI reveals spatial patterns of grey matter alteration in multiple sclerosis". In: *Human Brain Mapping* 42.4 (2021), pp. 1003–1012. DOI: 10.1002/hbm.25274.

[27] René-Maxime Gracien et al. "Assessment of cortical damage in early multiple sclerosis with quantitative T2 relaxometry". In: *NMR in Biomedicine* 29.4 (2016), pp. 444–450. DOI: 10.1002/nbm.3486.

[28] Philippe Lambin et al. "Radiomics: extracting more information from medical images using advanced feature analysis". In: *European journal of cancer* 48.4 (2012), pp. 441–446. DOI: 10.1016/j.ejca.2011.11.036.

[29] Philippe Lambin et al. "Radiomics: the bridge between medical imaging and personalized medicine". In: *Nature reviews Clinical oncology* 14.12 (2017), pp. 749–762. DOI: 10.1038/nrclinonc.2017.141.

[30] William Rogers et al. "Radiomics: from qualitative to quantitative imaging". In: *The British journal of radiology* 93.1108 (2020), p. 20190948. DOI: 10.1259/bjr.20190948.

[31] Thibaud P Coroller et al. "CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma". In: *Radiotherapy and Oncology* 114.3 (2015), pp. 345–350. DOI: 10.1016/j.radonc.2015.02.015.

[32] Prateek Prasanna et al. "Radiomic features from the peritumoral brain parenchyma on treatment-naive multi-parametric MR imaging predict long versus short-term survival in glioblastoma multiforme: preliminary findings". In: *European radiology* 27 (2017), pp. 4188–4197. DOI: 10.1007/s00330-016-4637-3.

[33] Philippe Lambin et al. "Decision support systems for personalized and participative radiation oncology". In: *Advanced drug delivery reviews* 109 (2017), pp. 131–153. DOI: 10.1016/j.addr.2016.01.006.

[34] Janna E van Timmeren et al. "Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam CT images". In: *Radiotherapy and Oncology* 123.3 (2017), pp. 363–369. DOI: 10.1016/j.radonc.2017.04.016.

[35] Zhenyu Liu et al. “Radiomics analysis allows for precise prediction of epilepsy in patients with low-grade gliomas”. In: *NeuroImage: Clinical* 19 (2018), pp. 271–278. DOI: 10.1016/j.nicl.2018.04.024.

[36] Yaou Liu et al. “Radiomics in multiple sclerosis and neuromyelitis optica spectrum disorder”. en. In: *Eur. Radiol.* 29.9 (Sept. 2019), pp. 4670–4677. DOI: 10.1007/s00330-019-06026-w.

[37] Xiaoxiao Ma et al. “Quantitative radiomic biomarkers for discrimination between neuromyelitis optica spectrum disorder and multiple sclerosis”. In: *Journal of Magnetic Resonance Imaging* 49.4 (2019), pp. 1113–1121. DOI: 10.1002/jmri.26287.

[38] Feng Feng et al. “Radiomic features of hippocampal subregions in Alzheimer’s disease and amnestic mild cognitive impairment”. In: *Frontiers in aging neuroscience* 10 (2018), p. 290. DOI: 10.3389/fnagi.2018.00290.

[39] Ruijiang Li et al. *Radiomics and radiogenomics: technical basis and clinical applications*. CRC press, 2019.

[40] Chris H Polman et al. “Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria”. In: *Annals of neurology* 69.2 (2011), pp. 292–302. DOI: 10.1002/ana.22366.

[41] Roberto Souza et al. “An open, multi-vendor, multi-field-strength brain MR dataset and analysis of publicly available skull stripping methods agreement”. In: *NeuroImage* 170 (2018), pp. 482–494. DOI: 10.1016/j.neuroimage.2017.08.021.

[42] F Cotton et al. “OFSEP, a nationwide cohort of people with multiple sclerosis: Consensus minimal MRI protocol”. In: *Journal of Neuroradiology* 42.3 (2015), pp. 133–140. DOI: 10.1016/j.neurad.2014.12.001.

[43] Olivier Commowick et al. "Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure". In: *Scientific reports* 8.1 (2018), p. 13650. DOI: 10.1038/s41598-018-31911-7.

[44] Thomas Martin Lehmann, Claudia Gonner, and Klaus Spitzer. "Survey: Interpolation methods in medical image processing". In: *IEEE transactions on medical imaging* 18.11 (1999), pp. 1049–1075. DOI: 10.1109/42.816070.

[45] John Ashburner and Karl J Friston. "Unified segmentation". In: *Neuroimage* 26.3 (2005), pp. 839–851. DOI: 10.1016/j.neuroimage.2005.02.018.

[46] Paul Schmidt et al. "An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis". In: *Neuroimage* 59.4 (2012), pp. 3774–3783. DOI: 10.1016/j.neuroimage.2011.11.032.

[47] Christophe Phillips and Cyril Pernet. "Unifying lesion masking and tissue probability maps for improved segmentation and normalization". In: *23rd Annual Meeting of the Organization for Human Brain Mapping*. 2017.

[48] Nicholas J Tustison et al. "N4ITK: improved N3 bias correction". In: *IEEE transactions on medical imaging* 29.6 (2010), pp. 1310–1320. DOI: 10.1109/TMI.2010.2046908.

[49] Alex Zwanenburg et al. "The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping". en. In: *Radiology* 295.2 (May 2020), pp. 328–338. DOI: 10.1148/radiol.2020191145.

[50] Stefania Rizzo et al. "Radiomics: the facts and the challenges of image analysis". In: *European radiology experimental* 2.1 (2018), pp. 1–8. DOI: 10.1186/s41747-018-0068-z.

[51] Joost JM Van Griethuysen et al. “Computational radiomics system to decode the radiographic phenotype”. In: *Cancer research* 77.21 (2017), e104–e107. DOI: 10.1158/0008-5472.CAN-17-0339.

[52] Robert M Haralick, Karthikeyan Shanmugam, and Its' Hak Dinstein. “Textural features for image classification”. In: *IEEE Transactions on systems, man, and cybernetics* 6 (1973), pp. 610–621.

[53] Mary M Galloway. “Texture analysis using gray level run lengths”. In: *Computer graphics and image processing* 4.2 (1975), pp. 172–179. DOI: 10.1016/S0146-664X(75)80008-6.

[54] Guillaume Thibault et al. “Shape and texture indexes application to cell nuclei classification”. In: *International Journal of Pattern Recognition and Artificial Intelligence* 27.01 (2013), p. 1357002. DOI: 10.1142/S0218001413570024.

[55] Moses Amadasun and Robert King. “Textural features corresponding to textural properties”. In: *IEEE Transactions on systems, man, and Cybernetics* 19.5 (1989), pp. 1264–1274. DOI: 10.1109/21.44046.

[56] Chengjun Sun and William G Wee. “Neighboring gray level dependence matrix for texture classification”. In: *Computer Vision, Graphics, and Image Processing* 23.3 (1983), pp. 341–352. DOI: 10.1016/0734-189X(83)90032-4.

[57] Dejan Jakimovski et al. “Long-standing multiple sclerosis neurodegeneration: volumetric magnetic resonance imaging comparison to Parkinson's disease, mild cognitive impairment, Alzheimer's disease, and elderly healthy controls”. In: *Neurobiology of Aging* 90 (2020), pp. 84–92. DOI: 10.1016/j.neurobiolaging.2020.02.002.

[58] Ewelina Marciniewicz et al. “The role of MR volumetry in brain atrophy assessment in multiple sclerosis: A review of the literature”. In: *Advances in Clinical & Experimental Medicine* 28.7 (2019). DOI: 10.17219/acem/94137.

[59] W Evan Johnson, Cheng Li, and Ariel Rabinovic. "Adjusting batch effects in microarray expression data using empirical Bayes methods". In: *Biostatistics* 8.1 (2007), pp. 118–127. DOI: 10.1093/biostatistics/kxj037.

[60] Jianping Hua et al. "Optimal number of features as a function of sample size for various classification rules". In: *Bioinformatics* 21.8 (2005), pp. 1509–1515. DOI: 10.1093/bioinformatics/bti171.

[61] Yaser S Abu-Mostafa, Malik Magdon-Ismail, and Hsuan-Tien Lin. *Learning from data*. Vol. 4. AMLBook New York, 2012.

[62] Isabelle Guyon et al. "Gene selection for cancer classification using support vector machines". In: *Machine learning* 46 (2002), pp. 389–422. DOI: 0.1023/A:1012487302797.

[63] Leo Breiman. "Random forests". In: *Machine learning* 45 (2001), pp. 5–32. DOI: 10.1023/A:1010933404324.

[64] Thais Mayumi Oshiro, Pedro Santoro Perez, and José Augusto Baranauskas. "How many trees in a random forest?" In: *Machine Learning and Data Mining in Pattern Recognition: 8th International Conference, MLDM 2012, Berlin, Germany, July 13-20, 2012. Proceedings 8*. Springer. 2012, pp. 154–168. DOI: 10.1007/978-3-642-31537-4_13.

[65] Trevor Hastie et al. *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer, 2009.

[66] John Platt et al. "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods". In: *Advances in large margin classifiers* 10.3 (1999), pp. 61–74.

[67] Gary S Collins et al. "Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement". en. In: *BMC Med.* 13.1 (Jan. 2015), pp. 1–10. DOI: 10.1136/bmj.g7594.

[68] Jonathan Zurawski et al. “7T MRI cerebral leptomeningeal enhancement is common in relapsing-remitting multiple sclerosis and is associated with cortical and thalamic lesions”. In: *Multiple Sclerosis Journal* 26.2 (2020), pp. 177–187. DOI: 10.1177/1352458519885106.

5

Deep learning for carotid artery segmentation in stroke-at-risk patients

Elizaveta Lavrova, Zohaib Salahuddin, Henry C. Woodruff, Mohamed Kassem, Robin Camarasa, Anja G. van der Kolk, Paul J. Nederkoorn, Daniel Bos, Jeroen Hendrikse, M. Eline Kooi, Philippe Lambin

---

Adapted from: Elizaveta Lavrova et al. "UR-CarA-Net: A Cascaded Framework With Uncertainty Regularization for Automated Segmentation of Carotid Arteries on Black Blood MR Images". In: *IEEE Access* 11 (2023), pp. 26637–26651. DOI: 10.1109/ACCESS.2023.3258408.

Abstract

We present a fully automated method for carotid artery (CA) outer wall segmentation in black blood MRI using partially annotated data and compare it to the state-of-the-art reference model. Our model was trained and tested on multicentric data of patients (106 and 23 patients, respectively) with a carotid plaque and was validated on different MR sequences (24 patients) as well as data that were acquired with MRI systems of a different vendor (34 patients). A 3D nnU-Net was trained on pre-contrast T1w turbo spin echo (TSE) MR images. A CA centerline sliding window approach was chosen to refine the nnU-Net segmentation using an additionally trained 2D U-Net to increase agreement with manual annotations. To improve segmentation performance in areas with semantically and visually challenging voxels, Monte-Carlo dropout was used. To increase generalizability, data were augmented with intensity transformations. Our method achieves state-of-the-art results yielding a Dice similarity coefficient (DSC) of 91.7% (interquartile range (IQR) 3.3%) and volumetric intraclass correlation (ICC) with ground truth of 0.90 on the development domain data and a DSC of 91.1% (IQR 7.2%) and volumetric ICC with ground truth of 0.83 on the external domain data outperforming top-ranked methods for open-source CA segmentation. The uncertainty-based approach increases the interpretability of the proposed method by providing an uncertainty map together with the segmentation.

5.1 Introduction

Stroke is the second leading cause of death and a leading cause of disability in adults worldwide [1]. Carotid artery (CA) atherosclerosis is one of the major causes of stroke as it can lead to the formation of an embolus from an atherosclerotic plaque or hypoperfusion due to narrowing of the CA lumen. Current risk assessment and treatment decision strategies for patients with carotid artery stenosis due to the presence of a carotid plaque who suffered a recent transient ischemic attack (TIA) or stroke are based on the degree of CA stenosis [2]. However, recent studies have shown that plaque morphology and composition can improve stroke prediction [3, 4, 5, 6].

Modern medical imaging techniques, such as ultrasound (US), computed tomography angiography (CTA), and magnetic resonance imaging (MRI) help shed light on CA plaque characteristics in a non- or minimally-invasive manner. MRI, especially when combining multiple contrast weightings with a reference black blood T1-weighted scan, can provide extensive information about the CA, plaque morphology, and even plaque subcomponents in 3D without the adverse effects of radiation dose on the patient [5, 7]. One of the challenges of MRI is the fact that manual CA plaque characterization is time-consuming and subjective [8]. Therefore, there is a demand to automate this process.

In the last decades, several research groups reported on various methods to enable automated plaque characterization on multi-contrast MRI by segmenting plaque components. Computer vision approaches such as shape fitting, active contours, and level sets, in combination with simple machine learning methods such as classification and clustering, were attempted early on [9, 10, 11, 12, 13, 14, 15, 8, 16, 17, 18, 19, 20, 21]. In more recent years, convolutional neural networks (CNN), including U-Net, have gained increasing attention [22, 23, 24, 25, 26, 27, 28]. Whereas in early works, mostly area or volume differences were used as segmentation evaluation metrics, [17] was one of the first to report traditional segmentation scores such as Dice Score Coefficient and Hausdorff distance.

When analyzing the many techniques that have attempted to characterize CA plaque and its subcomponents, it becomes clear that the levels of automation are widely disparate. Even though plaque characterization methods are automated, they require CA localization first, and this step is performed manually. In some studies, readers need to delineate the outer wall and lumen on every MRI slice [10, 12, 11, 8, 22, 21]. In other studies, plaque characterization needs manually pre-cropped regions containing the CA or lumen seed points in every slice as an input [9, 23, 26, 27]. Sometimes even additional delineation of a muscle region is needed for intensity re-scaling [9, 10, 11]. While there are some studies where the CA was located using the lumen seed points in the distal slices alone, or in different CA branches [13, 18, 14, 19, 20], user interaction is still necessary, and no studies show robustness to seed positioning. A couple of publications report on automated CA localization, but the detection area is limited to manually selected slices [17]. These approaches are time-consuming and introduce inter- and intra-reader variability. Moreover, with a multi-contrast approach, different MRI scans should be co-registered, which is at the moment performed in a semi-automated manner. Therefore, to increase the speed, robustness, reliability, and reproducibility of advanced quantitative CA analysis, while decreasing cost and clinical burden, automated detection and segmentation of the CA on black blood MRI sequences remains an unmet clinical need.

There are a number of challenges related to CA segmentation. First of all, currently popular deep learning-based semantic segmentation approaches show robust performance but require large amounts of fully annotated training data [29]. In most datasets, only the symptomatic CA is delineated. Moreover, in most cases, only the internal CA is delineated meaning that in addition to the external branch of the symptomatic CA and the entire asymptomatic CA, vertebral arteries are also not delineated. Even though those arteries are not of high clinical interest, for a deep learning network they are essentially the same objects as the CA. Other neck arteries are also present on the slice, and if their external walls and lumens are clear on

the image, being unlabelled, they can confuse a segmentation model. Second, multi-contrast MRI often experiences a domain shift caused by different acquisition protocols [30]. Therefore, models trained on the data from some particular scanner and acquisition protocol might not be performing well on the data from slightly different acquisition settings. Third, MRI is expressed in arbitrary units [31]. Therefore, segmentation by thresholding characteristic physical density units, such as Hounsfield units, is problematic. Fourth, ground-truth segmentation is usually performed on the reference MR sequence, and the other black blood sequences are rigidly coregistered in a semi-automated manner, which might cause co-registration issues related to patient movement. Finally, Most modern deep learning applications do not provide uncertainty estimations of the segmentations and retail "black boxes" regarding the interpretability of the outputs. Nevertheless, at the moment, the demand for interpretable methods is growing [32]. At the moment, there are several studies aimed to perform the whole CA segmentation [27, 28, 33], but they all used the data from the same domain for both training and evaluation of their models.

To address these issues, we aimed to develop an automated method for common and internal CA detection on black blood MRI which would be robust against image quality and acquisition protocol variations and could be trained on partially segmented data, as well as providing uncertainty estimates of the segmentations generated. For this, we trained a nnU-Net [34], known for its high semantic segmentation performance in similar tasks [35], on lateral halves of T1-weighted (T1w) scans containing CA outer wall contours. After an nnU-Net model was trained to perform semantic segmentation of CA on 3D MRI scans, an additional U-Net was trained on 2D patches, containing the CA contour, to refine the output contours in anatomically challenging slices such as bifurcation areas and increase the generalizability of the method on the external data. To refine segmentation in visually challenging voxels, we incorporated an uncertainty component into the loss function using Monte-Carlo dropout during training. Within this study, we hypothesized that

the proposed workflow would be able to segment CA on black blood MRI robustly regardless of black blood sequence, contrast media, or MRI vendor. Our secondary hypothesis was that these nnU-Net refinements will improve the segmentation. We refer to the refined segmentation network as "U-CarA-Net" (short for "U-Net-based carotid artery segmentation"), and the network with added uncertainty regularization as "UR-CarA-Net" (short for "U-Net-based carotid artery segmentation with regularization").
We tested U-CarA-Net and UR-CarA-Net on the training data as well as on data from an external center. Additionally, we tested the method on the regular T2w TSE scans and post-contrast T1w TSE scans. We performed an ablation study comparing the segmentations produced by the backbone nnU-Net, U-CarA-Net, and UR-CarA-Net. Our contribution in this work lies in the implementation of a fully automated model for the segmentation of CA in two clinically relevant MRI sequences which can be trained on partially labeled data as only the symptomatic CA was delineated - and the addition of an uncertainty estimate of the segmentations. The code is open-source and available at GitHub: `https://github.com/lavrovaliz/ur-cara-net` [36].

5.2 Methods

5.2.1 Method overview

The main steps of the proposed method are:

1. Splitting the image of the neck through the central medial plane and keeping only the halves containing ground-truth CA segmentation, keeping right side halves and reflecting left sides to preserve the anatomy;

2. Training a 3D nnU-Net on lateral halves of T1w TSE images, containing ground-truth CA contours, for CA segmentation;

3. Training the U-CarA-Net on 2-channel 2D patches localized along the CA centerline localized by the nnU-Net;
4. Training the UR-CarA-Net on 2-channel 2D patches localized along the CA centerline using U-CarA-Net weights as initial weights.

The pipeline is illustrated in Figure 5.1. A detailed description of the steps is presented in the next sections.

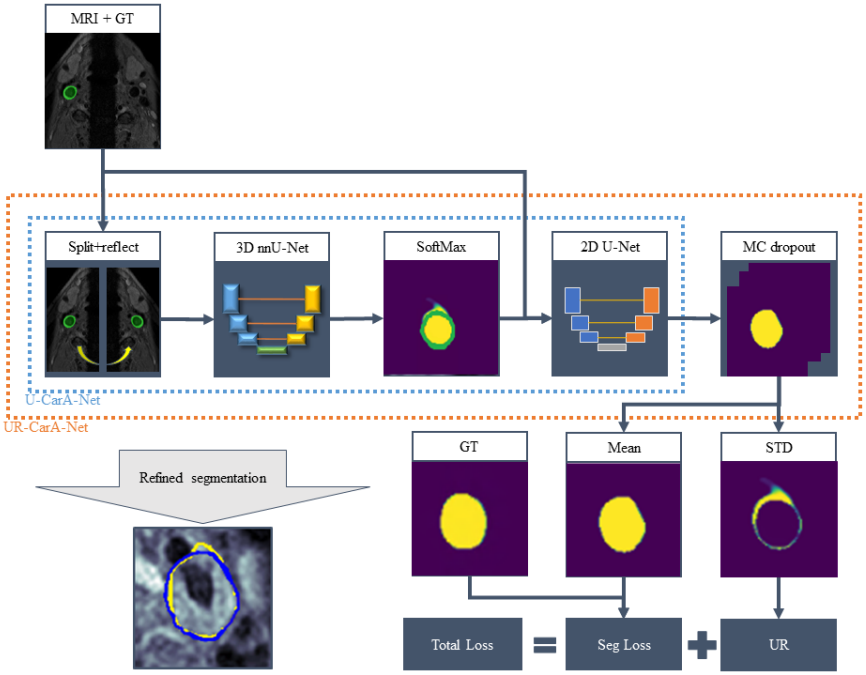


Figure 5.1: Method overview; MRI - magnetic resonance imaging, GT - ground-truth segmentation, SoftMax - nnU-Net SoftMax output, MC dropout - Monte-Carlo dropout, STD - standard deviation, Seg Loss - segmentation loss, UR - uncertainty regularization; all the models have binary outputs: CA and non-CA, models outputs are presented in purple-yellow segmentation map; the presented contours have several colors: green - ground truth, yellow - nnU-Net, blue - UR-CarA-Net.

5.2.2 MRI scan split

As ground-truth segmentations were available only for the symptomatic side, we split all the scans with the central median plane to deal with the missing contra-lateral segmentation. For training and validation, we selected the halves containing the ground-truth segmentations. To preserve the anatomy for the nnU-Net model, we keep the right side halves untouched and the left side halves are reflected relative to the median plane assuming lateral symmetry of the neck is not disturbed significantly with the minor deviations in the positioning of the patients.

5.2.3 CA segmentation with nnU-Net

To perform the initial CA segmentation, we trained a 3D full-resolution nnU-Net [34] on T1w TSE data. The method described performs data pre-processing and adapts its training and data augmentation parameters according to the properties of the training data. We used the standard settings from the original implementation including a combination of Dice similarity coefficient (DSC) and cross-entropy as a loss function, stochastic gradient descent with Nesterov momentum as an optimizer, poly learning rate schedule with an initial learning rate of 0.01, and training on 1000 epochs.

5.2.4 Segmentation refinement

Using the nnU-Net trained to segment CA in black blood T1w TSE carotid scans, we were able to localize the CA centerline and get the CA contour information. To refine the contours, given there could be cases of slight co-registration errors in the data from the different sequences, we trained a 2D U-Net [37] to segment the CA on square MRI slice patches, moving along the CA centerline.
For every slice, the patch center was placed into the center mass of the nnU-Net outcome. As a 2-channel input for the U-CarA-Net, we

used corresponding patches of the T1w slice and the nnU-Net softmax output. The nnU-Net softmax output gives information about the reference segmentation as well as uncertainty information by labeling voxels with values between 0 and 1.
After the U-CarA-Net is trained, we re-trained it in the following manner to improve its performance, calling it UR-CarA-Net. Besides utilizing a 2-channel input, we incorporated uncertainty information into the loss function as suggested in [38] to improve probability calibration in the "difficult" region. We generated the uncertainty estimate using Monte-Carlo dropout during training. We trained the model with dropout and after every epoch, we sampled outcomes obtained on the validation data from the current epoch and previous $N-1$ epochs. Then we calculated the value of the loss function composed of the weighted segmentation loss and uncertainty regularization component, where segmentation loss had DSC and cross-entropy components.
In the following, $\mathcal{P} \subset \mathrm{R}^2$ denotes the set pixels of a patch, $\theta \in \Omega$ are the parameters of the model, y_p represents the "ground-truth" label of the pixel $p \in \mathcal{P}$, $y_{p,\theta}$ the segmentation prediction for the pixel $p \in \mathcal{P}$ with a model with parameters θ, and $\overline{y}_p = \int_{\theta \in \Omega} \mathbb{P}(y_p|\theta)\mathbb{P}(\theta|\mathcal{D})d\theta$ is the Bayesian average estimate for the pixel $p \in \mathcal{P}$ approximated drawing sets of weights $\Theta \in \Omega$ using Monte-Carlo dropout with n_Θ times sampling on the training dataset $\mathcal{D}$:

$$\begin{aligned}\mathcal{L}_{TOTAL} = \underbrace{w_{DSC}\mathcal{L}_{DSC} + w_{BCE}tanh(\frac{\gamma_{BCE}\mathcal{L}_{BCE}}{2}))}_{\text{Segmentation}} + \\ + \underbrace{w_{UNC}tanh(\frac{\gamma_{UNC}\mathcal{R}_{UNC}}{2}))}_{\text{Uncertainty}}\end{aligned} \tag{5.1}$$

where $tanh : x \to \frac{e^x - e^{-x}}{e^x + e^{-x}}$ is an hyperbolic tangent, $\mathcal{L}_{TOTAL}$ is the final loss function, the segmentation loss contains weighted DSC loss $\mathcal{L}_{DSC}$

and cross-entropy loss $\mathcal{L}_{BCE}$, $\mathcal{R}_{UNC}$ is our proposed uncertainty regularization component, the corresponding weights are w_{DSC}, w_{BCE}, and w_{UNC}, and scaling factors are γ_{BCE} and γ_{UNC}. The loss functions and regularization component are calculated as the following:

$$\mathcal{L}_{DSC} = 1 - \frac{s + 2\sum_{p\in\mathcal{P}} y_p \overline{y}_p}{s + \sum_{p\in\mathcal{P}} y_p + \sum_{p\in\mathcal{P}} \overline{y}_p}, \tag{5.2}$$

$$\mathcal{L}_{BCE} = -\sum_{p\in\mathcal{P}} y_p log(\overline{y}_p) + (1 - y_p) log(1 - \overline{y}_p)), \tag{5.3}$$

$$\mathcal{R}_{UNC} = \sum_{p\in\mathcal{P}} \frac{1}{n_\Theta} \sum_{\theta\in\Theta} (\overline{y}_p - y_{p,\theta})^2. \tag{5.4}$$

where s is a smoothing factor.

We implemented hyperbolic tangent as a scaling function to cross-entropy and uncertainty components since they have ranges of values different from each other and DSC. Both components have 0 as a minimal possible value, and the highest value depends on the image size in the extreme case. We used hyperbolic tangent to scale the values of these components in the range of [0, 1].

A new loss was implemented for training the U-CarA-Net model since we assumed that the learning curve plateau was reached. As we intended to improve the performance of a model trained to solve a particular task, we assumed that the U-CarA-Net weights were close to the optimal point in the parameter hyperspace. Therefore, for UR-CarA-Net, we decreased the learning rate to find a solution close to the U-CarA-Net weights.

5.3 Experimental set-up

5.3.1 Imaging data

This model-building procedure was registered at OSF.io (10.17605/OSF.IO/VPT2B). We used imaging data acquired within the PARISK study (clinical trials.gov NCT01208025) in Amsterdam Medical Center (center 1), Erasmus Medical Center (center 2), Maastricht University Medical Center+ (center 3), and University Medical Center Utrecht (center 4) [2]. PARISK is a large prospective multicenter study to improve recurrent stroke risk stratification based on multimodality carotid imaging in symptomatic patients with mild to moderate CA stenosis. Inclusion criteria were a transient ischemic attack (TIA), amaurosis fugax, or minor stroke (modified Rankin scale $\leq$ 3) of the CA territory, CA NASCET stenosis ¡70% of the ipsilateral internal CA detected on Doppler US or CTA, and no indication for a revascularization procedure. Exclusion criteria were a probable cardiac source of embolism, a clotting disorder, severe comorbidity, standard contraindications for MRI, . Written informed consent was obtained from all patients before enrolment. MRI was performed on 3T whole-body scanners. Centers 1, 3, and 4 used an Achieva TX scanner (Philips Healthcare, Best, The Netherlands) with an eight-channel phased-array coil (Shanghai Chenguang Medical Technologies Co., Shanghai, China). Center 2 used a Discovery MR 750 system (GE Healthcare, Milwaukee, MI, USA) with a four-channel phased-array coil with an angulated setup (Machnet B.V., Roden, Netherlands). Apart from the difference in hardware and MRI protocols, the main difference is in acquired and reconstructed voxel sizes. More information is available in [2].
For this study, we used pre-contrast T1w TSE MRI. Additionally, we validated the models on T2w and contrast-enhanced T1w (T1w CE) TSE MRI. The scans were acquired with the same reconstructed slice thickness of 2 mm and contained up to 15 slices, centered on the CA bifurcation. The acquisition plane for all the protocols was transversal. The MRI protocols were described previously [2] and are

summarized in Table Table 5.1. We selected patients for whom the described sequences were available. Therefore, our data contained 13, 115, 25, and 34 patients from centers 1, 2, 3, and 4, respectively.

Table 5.1: **MRI scan parameters.**

Sequence	T1w QIR TSE	T1w DIR FSE	T2w TSE
Center	1, 3, 4	2	1, 3, 4
TR (ms)	800	1 RR	4800
TE (ms)	10	5.2	49
Acquired voxel size (mm×mm)	0.62×0.67	0.55×0.71	0.62×0.63
Reconstructed voxel size (mm×mm)	0.30×0.30	0.55×0.63	0.30×0.30
Acquisition matrix (pixels×pixels)	260×240	256×224	260×252
Reconstruction matrix (pixels×pixels)	528×528	256×256	528×528

We used the following abbreviations: TR — repetition time, TE — echo time, RR — R wave to R wave interval (1 heartbeat), QIR — quadruple inversion recovery, TSE — turbo spin echo, DIR — double inversion recovery, FSE — fast spin echo.

5.3.2 Image pre-processing

The slices from centers 1, 3, and 4 were cropped to a 512 pixels × 512 pixels matrix by removing 8 border pixels from each side. The slices from center 2 were resampled with cubic interpolation to an in-plane 0.3 mm × 0.3 mm pixel size and reshaped to a 512 × 512 matrix. The reshaping was performed by padding. Intensity normalization was performed by subtracting the minimum intensity and dividing it by the intensity range for every slice.

5.3.3 Experiments

We trained and evaluated three CA detection and segmentation models: the baseline nnU-Net and U-CarA-Net with and without uncertainty component in the loss function. All the models were trained, validated, and tested on the same data, for both U-CarA-Net models the same data augmentation transformations were applied to the same slices.
The patients from centers 1, 3, and 4 were split in a center-stratified manner into training, validation, and test sets in the proportions of 0.70, 0.15, and 0.15, respectively. Training and validation sets were used while training the models. Data from the test set were used to evaluate the performance scores. Data from center 2 were used for external testing. The summary of the resulting data split can be seen in Table Table 5.2.

Table 5.2: Patients split into training, validation, and test sets.

Center	1	2	3	4	Total
Total (N)	13	34	115	25	187
Training (N)	9	0	80	17	106
Validation (N)	2	0	17	4	23
Test (N)	2	0	18	4	23
External test (N)	0	34	0	0	34

All the data we had, we could separate into 4 domains. The pre-contrast T1w TSE data from centers 1, 3, and 4 were split into training, validation, and test sets and represented the development domain (referred to as DD). The scans from the training and validation sets were used to train the CA detection models and tune hyperparameters. The models were evaluated on the data from 3 domains different from the development domain:

1. post-contrast T1w TSE MRI from the patients of centers 1, 3, and 4 test sets (referred to as D1),

2. T2w MRI from the patients of centers 1, 2, and 3 test set (referred to as D2),

3. T1w FSE from center 2 (referred to as external domain, ED).

To investigate the generalizability of the models trained on the DD, they were evaluated on D1, D2, and ED to assess the center-specific impact of a different vendor, and different acquisition and reconstruction protocols. A summary of the data domains we used in this study is presented in Figure 5.2.

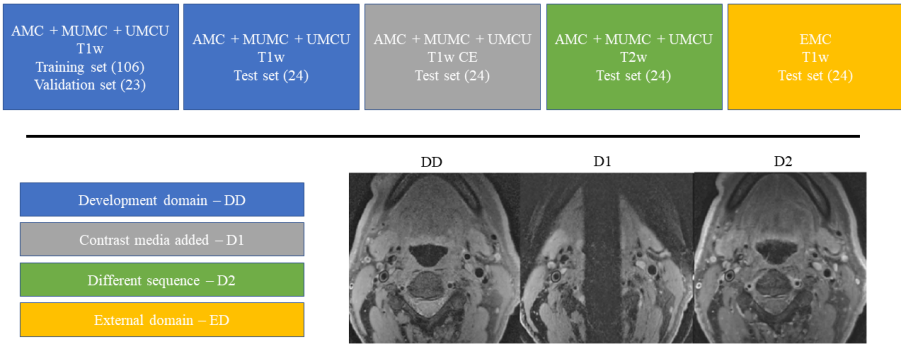


Figure 5.2: Data domain description.

5.3.4 Evaluation

To evaluate the results, we used common segmentation metrics such as the DSC and Hausdorff distance (HD). These scores are traditionally reported in segmentation studies and give an understanding of the overlap and distance between the segmented areas. Additionally, we used problem-specific metrics. As for our task, it is important to preserve the centerline of the segmented area, so we used the center line Dice similarity coefficient (clDSC) to evaluate the models [39]. As we are dealing with uncertain reference annotations in D1 and D2, we utilized the normalized surface distance (NSD) [39, 40] as an uncertainty-aware score with a tolerance of 2 voxels as this is the average contour-

ing disagreement we calculated from the manual annotations. Even though we suggest a fully automated method, we still assume an expert interaction in the cases where the method fails. To assess a value of a possible expert correction, we evaluate the relative added path length (relAPL), which is the length of the contour that has to be drawn while editing a segmentation [41] reported relative to the ground-truth contour length.

Finally, we computed the clinical metric used in cardiovascular clinical practice, the root-mean-square error between the ground truth and automatically segmented volumes (VRMSE). To measure the agreement between ground-truth and automated segmentations, we calculated the intraclass correlation (ICC) and performed Bland-Altman analysis for the CA volumes. For the non-normal distributed scores, we reported median values and interquartile ranges (IQR). For the differences in the scores, statistical significance was assessed using the two-sided Wilcoxon test.

To better gauge the performance of our model in different conditions, we investigated the influence of the slice location and image quality on 2D segmentation scores. As previously described evaluation scores assess segmentation quality in the whole 3D scan, to compare the segmentation performance in different anatomical parts of the CA, we calculated 2D DSCs in common and internal CA. We were dealing with multiple data domains originating from contrast media presence, different acquisition protocols, and equipment, resulting in different levels of noise and intensity bias, which, from the digital imaging point of view, resulted in different image contrast. Therefore, we compared 3D DSCs obtained for D1, D2, and ED scans, having the contrast within and beyond DD contrast values. As a contrast evaluation metric, we chose Michelson contrast, which characterizes areas with non-uniform textures and is used in medical imaging [42].

To evaluate the effects of adding uncertainty information, we compared the baseline model (nnU-Net) with U-CarA-Net and UR-CarA-Net. All the models were trained on the whole training and validation datasets with only the symptomatic CA labeled. All the pipelines used the same trained nnU-Net model, both U-CarA-Net and UR-CarA-Net

models had the same architectures. Evaluations were performed on the test data.

5.3.5 Implementation details

We used the same architecture of U-Net as a backbone model for both U-CarA-Net and UR-CarA-Net. The input shape was 64×64×2 since we had 2 channels and a square patch size of 64 pixels × 64 pixels. A patch size of 64×64 was selected since 64 is the smallest power of two which exceeds the root square of the maximum CA area in training data slices. Convolutional layers (Conv2D) with a kernel size of 3 pixels ×3 pixels were followed by batch normalization for faster training and to reduce overfitting. The first Conv2D consisted of 16 filters. In the contracting path, a number of filters were duplicated in every other Conv2D, resulting in 256 filters in the bottleneck Conv2D layer. In the expanding path, the number of filters was halved in every Conv2D. All the activation layers after batch normalization layers were exponential linear units (ELUs) for simplicity and generalizability, and to avoid the vanishing gradient and dying node problems, except for sigmoid activation in the last conv2D for pixel-wise CA probability prediction for a non-linearly separable problem. Activation layers were followed by 2D max-pooling layers downsampling the input with a 2×2 window and a stride equal to the pool size selecting the max value from the window. Every max-pooling layer was followed by a dropout layer to prevent overfitting and enable uncertainty regularization component in UR-CarA-Net. Every transposed Conv2D (Conv2DTranspose) had a kernel size of 3 pixels × 3 pixels, strides of 2 pixels × 2 pixels, and the same number of filters as an upcoming Conv2D. The initial learning rate of the Adam optimizer was reduced by a factor of 0.1 after 3 epochs of non-improvement of the loss function.
We trained the U-CarA-Net for 100 epochs with a batch size of 64. The dropout rate was 0.05. The initial rate of the Adam optimizer was 0.001, it was reduced by the factor of 0.1 while the learning curve was on a plateau for 3 epochs, and the lowest bound for the learning rate

was 10-5. We trained UR-CarA-Net using U-CarA-Net weights as the initial weights. The dropout rate was 0.1 to increase its influence on the uncertainty regularization component. We used the same batch size and learning rate-reducing strategy as for U-CarA-Net. However, the initial learning rate of the Adam optimizer was set to 10^{-9}, and the lower bound for the learning rate was 10^{-15}. Small values of the learning rate were used since the baseline model (U-CarA-Net) was already trained to solve the task, hence, only refinement was needed. Therefore, it was important not to move far in the hyperparameter space. Weights for the loss functions components and scaling function factor were set to $w_{DSC} = 0.5$, $w_{BCE} = 0.5$, $w_{UNC} = 1$, $\gamma_{BCE} = 1$, and $\gamma_{UNC} = 1$. These values were estimated empirically while experimenting with different weighting factors using weights from [38] as a reference.
To avoid overfitting and increase the robustness of the model to different hardware, and reconstruction and acquisition protocols, data augmentation was applied to all the training slices. The transformations had a probability of 0.5 each and mostly simulated differences in intensity distributions. The following transformations were performed: horizontal and vertical flips, blurring with a kernel size ranging from 3 to 7 pixels, Gaussian noise with 0 mean and variance in the range from 10 to 50, brightness and contrast variation by 25%, and gamma transformation with gamma ranging from 0.8 to 1.2. Every transformation parameter was a random number from the uniform distribution of the corresponding range. From every original slice used for data augmentation, 10 augmented slices were generated.
We utilized Keras 2.2.4 with a TensorFlow 1.14.0 backend. All the training and testing were performed on one NVIDIA GeForce RTX 2080 Ti.

5.4 Results

The resulting nnU-Net configuration can be found in the debug file in the project GitHub repository [36].

As can be seen from Table Table 5.3, the overlap-based metrics, such as DSC, clDSC, and NSD, are relatively high for all the models and data domains. The highest medium DSC is 91.9% for the U-CarA-Net in DD, the lowest is 88.3% for the U-CarA-Net in ED. The clDSC is above 97% for all the models and all the data domains, even yielding a median value of 100.0% for nnU-Net in D2, the U-CarA-Net in DD and D2, and the UR-CarA-Net in D2. NSD scores are significantly higher than the corresponding DSC, exceeding 97% for all the models and data domains.

Table 5.3: Comparison of carotid artery (CA) segmentation performance of the baseline (nnU-Net), improved (U-CarA-Net), and proposed (UR-CarA-Net) methods calculated on development domain (DD), first domain (D1), second domain (D2), and external domain (ED).

Domain	DSC (%)	HD (mm)	clDSC (%)	NSD (%)	relAPL	VRMSE (ml)	ICC (–)
				nnU-Net			
DD	91.6 (4.0)	2.5 (5.3)	98.4 (5.6)	99.3 (3.1)	**0.28 (0.12)**	**0.22**	**0.94**
D1	**89.6 (12.2)**	3.6 (9.6)	**99.2 (27.5)**	98.7 (16.2)	**0.34 (0.15)**	0.72	0.55
D2	**89.9 (4.9)**	2.6 (7.0)	**100.0 (8.1)**	**99.6 (2.2)**	0.32 (0.09)	0.30	0.83
ED	88.7 (11.6)	3.0 (3.1)	98.9 (7.2)	97.5 (11.7)	0.47 (0.21)	1.26	0.53
				U-CarA-Net			
DD	**91.9 (5.0)**	2.3 (1.3)	**100.0 (1.7)**	**99.5 (2.8)**	0.30 (0.11)	0.28	0.88
D1	88.6 (13.2)	**3.5 (6.1)**	97.8 (19.0)	98.9 (16.7)	0.35 (0.19)	**0.44**	**0.74**
D2	89.7 (5.0)	**2.5 (4.8)**	**100.0 (7.8)**	99.4 (6.4)	0.33 (0.06)	0.28	0.85
ED	88.3**** (8.5)	**2.9 (3.0)**	98.7 (9.8)	98.2 (6.0)**	0.44 (0.19)****	0.82	0.77
				UR-CarA-Net			
DD	91.7 (3.3)	**2.2 (2.3)**	99.6 (4.5)	99.3 (2.6)	**0.28 (0.07)**	0.24	0.91
D1	89.2 (11.4)	3.6 (6.3)	98.8 (26.7)	**98.9 (15.7)**	**0.34 (0.14)**	0.45	**0.74**
D2	89.8 (5.1)	2.5 (5.9)	**100.0 (9.4)**	99.5 (3.6)	**0.31 (0.07)**	**0.27**	**0.86**
ED	**91.1 (7.2)****	**2.9 (2.6)**	**99.2 (7.2)**	**98.4 (3.9)****	**0.39 (0.18)****	**0.68**	**0.83**

Comparison of carotid artery (CA) segmentation performance of the baseline (nnU-Net), improved (U-CarA-Net), and proposed (UR-CarA-Net) methods calculated on development domain (DD), first domain (D1), second domain (D2), and external domain (ED). The values mentioned are median (IQR). We used the following abbreviations: DSC — Dice similarity coefficient, HD — Hausdorff distance, clCSC — centerline Dice Similarity Coefficient, NSD — normalized surface distance, VRMSE — volume root-mean-square error, ICC — volume intraclass correlation. The best scores are in bold. For DSC, JSC, clDSC, NSD, and ICC the higher the value the better whereas for HD, relAPL, and VRMSE the lower the value the better. Segmentation scores between the proposed and baseline and proposed and improved methods were compared on ED: p-values were calculated using a two-sided Wilcoxon test and referred as follows: "ns" refers to the p value in the range $5 \times 10^{-2} < p \leq 1$, "*" refers to the p value in the range $1 \times 10^{-2} < p \leq 5 \times 10^{-2}$, "**" refers to the p value in the range $1 \times 10^{-3} < p \leq 1 \times 10^{-2}$, "***" refers to the p value in the range $1 \times 10^{-4} < p \leq 1 \times 10^{-3}$, "****" refers to the p value in the range $p \leq 1 \times 10^{-4}$.

The best performance among the data domains is achieved on the DD. The common segmentation metrics (DSC, clDSC, and NSD) yielded by the baseline model were relatively high in D1, D2, and ED. Nevertheless, clinically relevant volume-related scores, such as VRMSE, showed insufficient segmentation in these domains, especially in ED yielding a VRMSE of 1.26 compared to a VRMSE of 0.22 in DD. Low values of ICC between ground-truth and auto-segmented contours in D1 and ED show a poor agreement with the manual segmentation. Contour refinement with the U-CarA-Net improved segmentation scores for the DD and distance-based metric HD for the D1 and D2. This resulted in the improvement of VRMSE and ICC yielding better agreement with the manual segmentation. Finally, contour refinement with the UR-CarA-Net improved distance-based HD and relAPL for all the data domains. It resulted in a lower VRMSE in D2 and ED and in higher ICC for D1, D2, and ED. Moreover, for ED, compared to nnU-Net, ICC increased by 0.30 and VRMSE decreased by a factor of 2. Additionally, the UR-CarA-Net application results in IQR drop for HD, NSD, and relAPL in all the domains, as well as DSC and clDSC for DD, D1, and ED. For the ED, all the metrics obtained with the UR-CarA-Net show the best performance of this model and the lowest IQRs. Nevertheless, for the DD data, the best volume-based scores are still yielded with nnU-Net. The most challenging data domain, even for the UR-CarA-Net, was post-contrast MRI, where we could achieve an ICC of 0.74.
Even though we noticed an improvement in the distance- and volume-based metrics while using the proposed model, overlap-based metrics have high values which are close to each other. Also a statistical comparison of the interval-based estimations of the metrics is shown. We compared the proposed method (UR-CarA-Net) with the baseline (nnU-Net) and improved (U-CarA-Net) ones on the ED. As we can see from results table, UR-CarA-Net significantly improved DSC, NSD, and relAPL.
Bland-Altman plots in Figure 5.3 show, that even though all the models are yielding absolute bias values close to 0 ml, the proposed method could decrease the absolute systematic bias in DD by 0.07 ml, in D2 by 0.01 ml, and in ED by 0.40 ml. Nevertheless, for DD and D1 the lowest

bias values were yielded with U-CarA-Net. We can also see that the implementation of U-CarA-Net and UR-CarA-Net increased the segmented volumes in DD, D2, and ED.
Figure 5.4 shows that 2D DSCs calculated in CCA are higher than in ICA. The UR-CarA-Net application decreases the range of DSC deviation, especially in ICA. Nevertheless, the lowest 2D DSC values and the widest DSC IQR are observed in D1 ICA, where U-CarA-Net and UR-CarA-Net do not improve the segmentation performance.
According to Figure 5.5, segmentation performance drops when the Michelson contrast is not within the range presented in the DD. In the D1, DSC is decreased by more than 20%, but in the D2 and ED, UR-CarA-Net application reduces the DSC variance for the scans with Michelson contrast values outside of the DD values.
Figure 5.6 shows segmentation results in the "simple" cases. These slices are taken from the different patients and data domains. On these slices, the CA is clearly visible, it has a sufficient area, and the anatomy is simple, without any bifurcations and other CA branches present. The DSC values calculated for the same slice but with different models are similar.
Figure 5.7 shows segmentation results in "challenging" cases. The slices are taken from the different patients with 2 representations from each data domain.
The segmentation challenges were caused by complicated anatomy (DD, D1), low image quality (ED), or different intensity distribution (D2). In these examples, the nnU-Net contours do not cover the whole CA area, especially in the bifurcation area or in the CA with stenosis. This results in the non-smooth contours not covering the whole CA area. In the DD case, the ICA was not a part of the contour. In the D2 case, the CA had only some pixels segmented; in the case of ED, it was not segmented. Implementation of the U-CarA-Net and UR-CarA-Net increased the segmented areas and DSCs. For case A, even though the highest DSC was yielded with UR-CarA-Net, the model still did not segment the whole CA area, and the contour was not smooth. However, the uncertainty map for this slice has high values in the pixels, classified as false negatives, showing that additional atten-

tion is needed. Finally, the UR-CarA-Net segmentation in the case ED overlaps the CA area. Still, it has a curved shape, which is reflected in the corresponding uncertainty map, having many bright pixels along the predicted segmentation edges. Nevertheless, DSC increased from 0.0% to 79.9%.

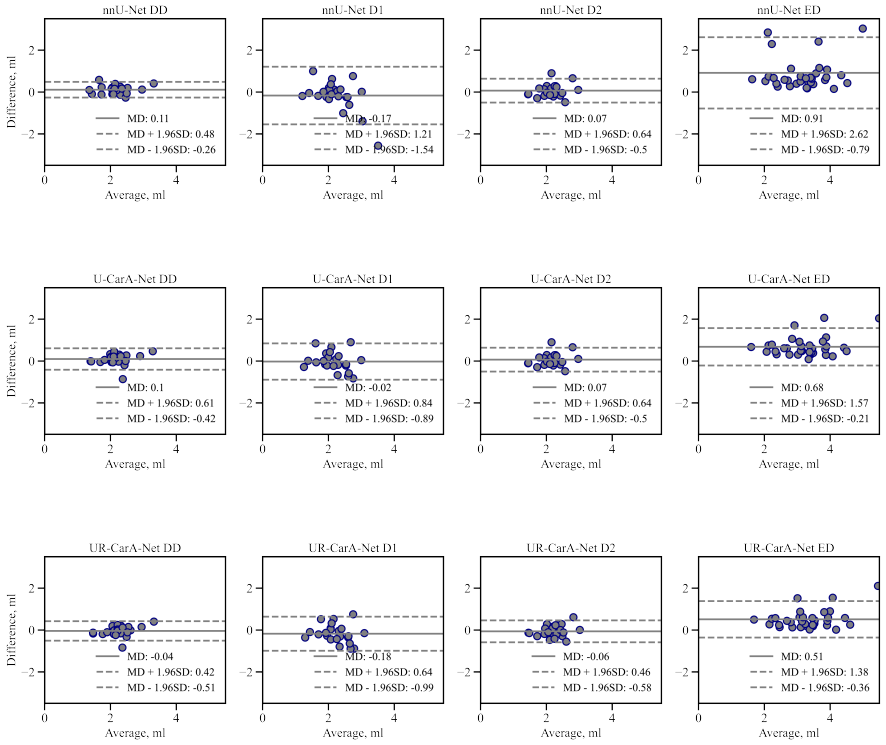


Figure 5.3: Bland-Altman plots for the models obtained on DD, D1, D2, and ED. The solid line represents the mean difference. The dashed lines represent +1.96 standard deviations (top) and -1.96 standard deviations (bottom) from the mean.

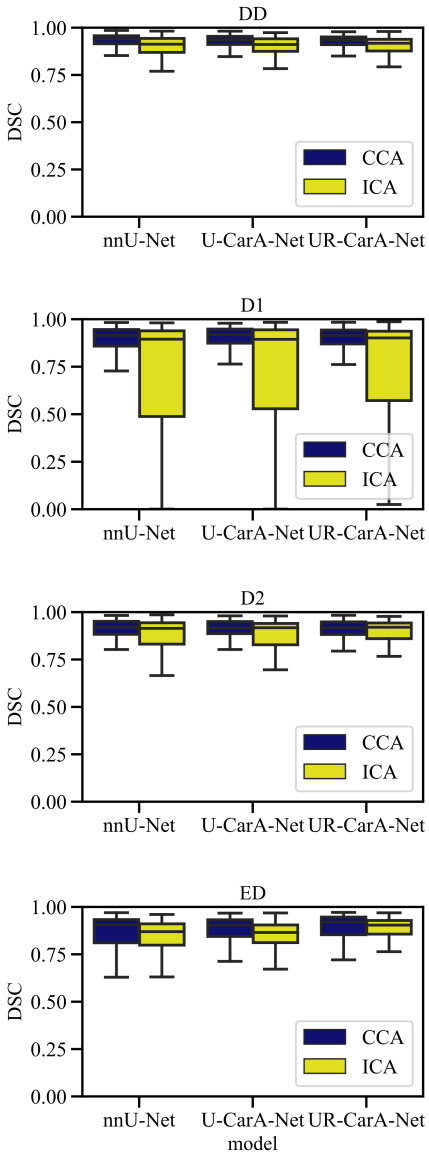


Figure 5.4: Segmentation performance with regards to carotid artery (CA) anatomy. Segmentation performance is measured in 2D Dice similarity coefficient (DSC). CA anatomy is defined as common CA (CCA) or internal CA (ICA). Data were presented as box plots, where boxes are representing the interquartile range (IQR), extending from Q1 to Q3 and centered on the median value. Upper whiskers represent the highest data point that is less than $Q3 + 1.5 \times IQR$. Lower whiskers represent the smallest data point that is greater than $Q3 - 1.5 \times IQR$.

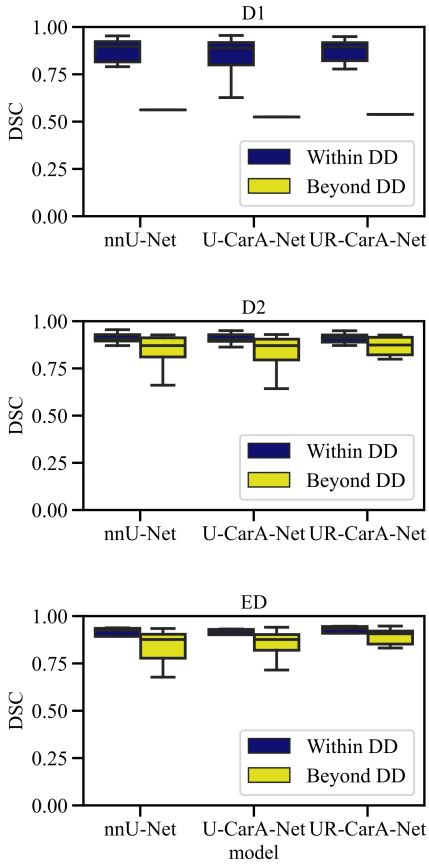


Figure 5.5: Segmentation performance with regards to the image quality corresponding or not to image quality in development domain (DD) data. Segmentation performance is measured in 3D Dice similarity coefficient (DSC). Image quality is defined in terms of Michelson contrast lying within and beyond DD values. Data were presented as box plots, where boxes are representing the interquartile range (IQR), extending from Q1 to Q3 and centered on the median value. Upper whiskers represent the highest data point that is less than $Q3 + 1.5 \times IQR$. Lower whiskers represent the smallest data point that is greater than $Q3 - 1.5 \times IQR$.

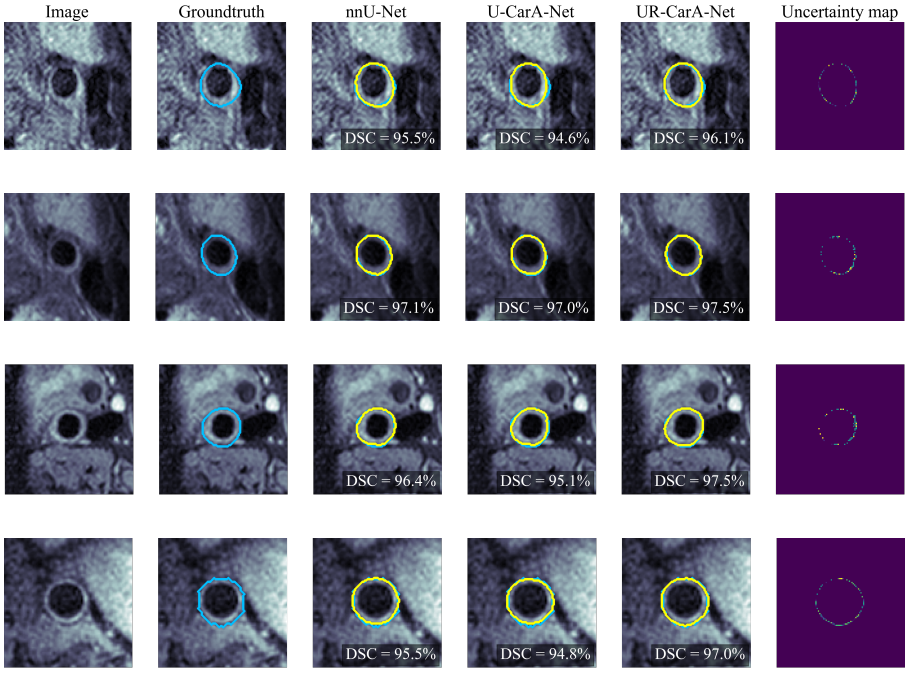


Figure 5.6: Segmentation of the 'simple' slices: blue contour — ground-truth, yellow contour — segmentation model, DSC — 2D Dice similarity coefficient for 2D contour, GT — ground-truth segmentation; first line — T1w TSE from the development domain, second line — post-contrast T1w TSE, third line — T2w, fourth line — external domain T1w FSE.

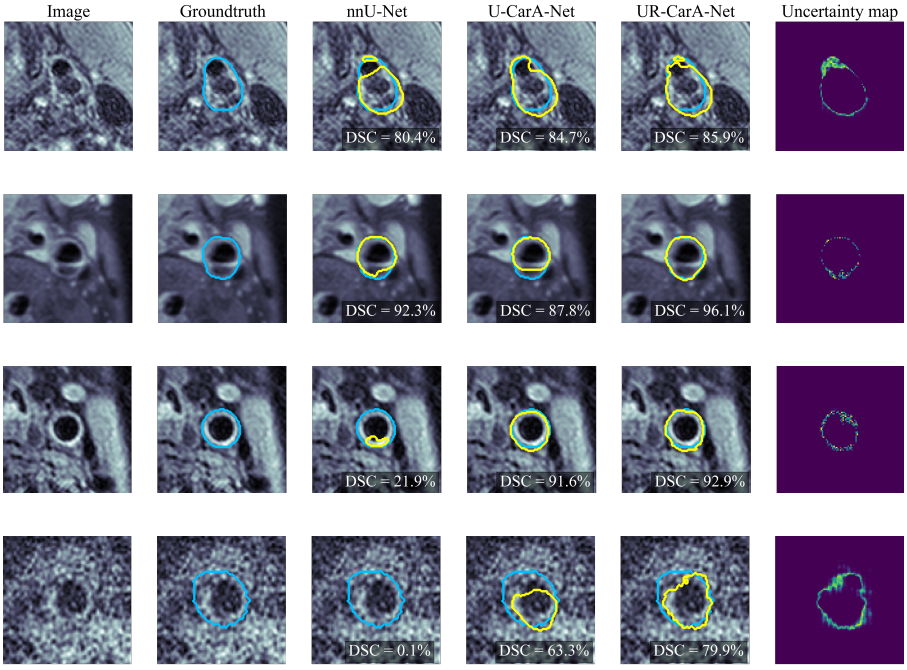


Figure 5.7: Segmentation of the 'challenging' slices: blue contour — ground-truth, yellow contour — segmentation model, DSC — 2D Dice similarity coefficient for 2D contour in percentage; first line — T1w TSE from the development domain, second line — post-contrast T1w TSE, third line — T2w, fourth line — external domain T1w FSE.

5.5 Discussion and conclusion

In this work, we presented a method for CA segmentation on BB MRI. The robustness of Our proposed method is more robust for different MRI protocols and acquisition equipment. Even though we did not observe any significant improvement testing our model on the data from the development domain, volumetric ICC between ground-truth and automated segmentations improved by 0.19, 0.03, and 0.30 in the contrast-enhanced, T2w, and different center MRI, respectively. Moreover, in the external domain data, DSC, NSD, and relAPL improved significantly. Even though the proposed method brought segmentation performance scores in the external domain more in alignment with the ones in the development domain, in the contrast-enhanced and T2w MRI, the volumetric scores were lower. This can be explained by the fact that according to the protocol, the ground-truth contours are drawn on the T1w TSE data. Therefore, they were drawn on development domain data and then projected on contrast-enhanced and T2w MRI. Due to the minor patient movement, some slight misregistrations were possible. Additionally, in the T1w CE scans, contrast accumulates in the highly vascularised outer vessel wall (adventitial layer), which can cause a bias in the quantification of the wall volumes. High clDSC and NSD scores were observed, which did not change much between the models. Therefore, CA centerlines and main areas segmented by the nnU-Net did not change much, and contours were refined at the edges resulting in the difference between the segmented volumes, which proves our secondary hypothesis.
In comparison to similar published work, [27] used a 3D U-Net to segment CA outer wall and lumen on the PARISK data excluding EMC scans. The authors applied Monte-Carlo dropout as well to assess the segmentation uncertainty, yielding DSCs of 76.4% for the vessel wall and 88.5% for the lumen. The authors limited the scans by a $128 \times 128 \times 16$ bounding box placed in the center of the ground-truth segmentations. Zhu et al. [33] and Alblas et al. [28] applied 3D U-Net for CA localization and 2D CNN for CA segmentation method similar to ours that employs the sequential application of two deep

learning models for 1) 3D CA centerline detection and 2) 2D CA wall contouring. They achieved a median DSC of 81.3% for the vessel wall on the open-source dataset acquired with the same MR protocol. Zhu et al. [33] achieved DSCs of 89.68% and 80.29% for the lumen and wall segmentation, respectively. They combined deep learning and graph-based approaches and applied them to the multi-sequence MRI acquired with the same protocols. DSCs reported in the current study, obtained on the PARISK data, exceed the values reported in the literature. But for a fair comparison, the same dataset should be used.

Although several studies to develop CA segmentation methods have been previously performed, we believe our work still stands out as a fully automated pipeline. Moreover, to our knowledge, it is the first study performing validation in multi-domain and multi-center data. We showed that the implementation of our method improves the segmentation performance on the external data domain to the level of the performance achieved on the development data domain. To improve our segmentation performance, we introduced three additional steps to the high-performing baseline model (nnU-Net): 1) carotid MRI lateral split to deal with one-sided annotations, 2) 2D contour refinement in a patch put into the localized CA area using the original image and nnU-Net softmax output, 3) introduction of an uncertainty component to the loss function using Monte-Carlo dropout. The reproducibility of this study is guaranteed by using carotid MRI data from a national multicenter PARISK study which is a highly recognized dataset. As all the patients had mild to moderate CA stenosis, the models were trained on the target patient cohort data, so they are robust to the pathological CA shape variations. We separately applied the method to the different BB MR sequences from the multi-contrast dataset instead of merging all these data into different channels which makes our method less demanding for the input data. Finally, UR-CarA-Net does not utilize the uncertainty information only in the training process but also enables uncertainty map generation by activating the dropout layers during the inference. Therefore, scans with high uncertainty scores can be reviewed separately.

There are some limitations of the presented approach. First of all, it is

highly dependent on the baseline model (nnU-Net) used for the initial CA segmentation. After nnU-Net, U/UR-CarA-Net scans the patches along the primarily segmented CA. Therefore, if nnU-Net does not segment the vessel or segments the wrong object in the scan, this error cannot be corrected by our method. Moreover, as presented in Figure 5.4, whereas CCA segmentation DSCs are relatively high, there might be errors in ICA segmentation. Nevertheless, Figure 5.4 also shows that the proposed method improves the ICA segmentation. However, these errors can be alleviated by increasing the dataset size. The second limitation is related to the blind lateral split of the MR scans into the right and left sides. If the patient is not positioned perfectly in the scanner, it results in a significant disturbance of the anatomy. Nevertheless, we rely on the carotid MRI data corresponding to the acquisition protocols where a patient is carefully centered in the scan. The third limitation is in the application of the 2D approach in contour refinement. We selected this approach for several reasons: 1) by using a 2D input, the training data size is increasing, 2) scanning 3D scans with a cubic 3D U-Net window will require more computational capacity, 3) reconstructed z-axis resolution is much lower than in-plane resolution (2 mm vs. 0.30 mm $\times$ 0.30 mm). Finally, it is necessary to be careful in selecting the BB MR sequences since the enhancement of the outer layer of the vessel wall can cause a bias in volumetric measurements.
As the proposed method is the first step in CA plaque characterization, in future work, we aim to build a model for cardiovascular event prediction in patients with CA stenosis to be able to stratify patients based on their risk of stroke. Additionally, the results show that it is possible to use our method for BB MR sequence automated co-registration by segmenting CA on the scans of the same patient and minimizing the distance between the segmentations obtained on the different sequences. The achieved state-of-the-art segmentation metrics of our method together with its interpretability due to the uncertainty maps generation means the approach can be used as an initial step in CA plaque analysis. It can be followed by automated plaque components segmentation with one of the existing methods and handcrafted or deep radiomics applications for clinical outcomes prediction.

To conclude, in this first externally validated multi-center fully automated CA segmentation study, our model showed good segmentation performance (DSC of 91.7% (IQR 3.3%) on the development domain scans and 91.1% (IQR 7.2%) on the external domain scans), as well as an agreement with the manual segmentation (volume ICC of 0.91 on the development domain scans and 0.83 on the external domain scans). Its application is also feasible for the other BB MRI, obtained with other equipment or sequences. The suggested approach can be used for the other tasks on partially labeled data. The code is available on GitHub [36].

References

[1] Emelia J Benjamin et al. "Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association". en. In: *Circulation* 139.10 (Mar. 2019), e56–e528. DOI: 10.1161/CIR.0000000000000659.

[2] M T Truijman et al. "Plaque At RISK (PARISK): prospective multicenter study to improve diagnosis of high-risk carotid plaques". In: *Int. J. Stroke* 9.6 (Aug. 2014). DOI: 10.1111/ijs.12167.

[3] Kelly P H Nies et al. "Emerging Role of Carotid MRI for Personalized Ischemic Stroke Risk Prediction in Patients With Carotid Artery Stenosis". en. In: *Front. Neurol.* 0 (2021). DOI: 10.3389/fneur.2021.718438.

[4] A C van Dijk et al. "Intraplaque Hemorrhage and the Plaque Surface in Carotid Atherosclerosis: The Plaque At RISK Study (PARISK)". en. In: *AJNR Am. J. Neuroradiol.* 36.11 (Nov. 2015), pp. 2127–2133. DOI: 10.3174/ajnr.A4414.

[5] L Saba et al. *Carotid Artery Wall Imaging: Perspective and Guidelines from the ASNR Vessel Wall Imaging Study Group and Expert Consensus Recommendations of the American Society of Neuroradiology*. 2018. DOI: 10.3174/ajnr.A5488.

[6] Dianne HK van Dam-Nolen et al. “Carotid plaque characteristics predict recurrent ischemic stroke and TIA: the ParisK (Plaque At Risk) study”. In: *Cardiovascular Imaging* 15.10 (2022), pp. 1715–1726. DOI: 10.1016/j.jcmg.2022.04.003.

[7] Ashish Saxena, Eddie Yin Kwee Ng, and Soo Teik Lim. *Imaging modalities to diagnose carotid artery stenosis: progress and prospect*. 2019. DOI: 10.1186/s12938-019-0685-7.

[8] Arna van Engelen et al. “Multi-Center MRI Carotid Plaque Component Segmentation Using Feature Normalization and Transfer Learning”. en. In: *IEEE Trans. Med. Imaging* 34.6 (June 2015), pp. 1294–1305. DOI: 10.1109/TMI.2014.2384733.

[9] I M Adame et al. “Automatic segmentation and plaque characterization in atherosclerotic carotid artery MR images”. en. In: *MAGMA* 16.5 (Apr. 2004), pp. 227–234. DOI: 10.1007/s10334-003-0030-8.

[10] J M A Hofman et al. “Quantification of atherosclerotic plaque components using in vivo MRI and supervised classifiers”. en. In: *Magn. Reson. Med.* 55.4 (Apr. 2006), pp. 790–799. DOI: 10.1002/mrm.20828.

[11] R van ’t Klooster et al. “Automated versus manual in vivo segmentation of carotid plaque MRI”. en. In: *AJNR Am. J. Neuroradiol.* 33.8 (Sept. 2012), pp. 1621–1627. DOI: 10.3174/ajnr.A3028.

[12] Fei Liu et al. *Automated in vivo segmentation of carotid plaque MRI with Morphology-Enhanced probability maps*. 2006.

[13] William Kerwin et al. *Magnetic Resonance Imaging of Carotid Atherosclerosis*. 2007. DOI: 10.1097/rmr.0b013e3181598d9d.

[14] Hui Tang et al. “Semiautomatic carotid lumen segmentation for quantification of lumen geometry in multispectral MRI”. en. In: *Med. Image Anal.* 16.6 (Aug. 2012), pp. 1202–1215. DOI: 10.1016/j.media.2012.05.014.

[15] Arna van Engelen et al. "Atherosclerotic plaque component segmentation in combined carotid MRI and CTA data incorporating class label uncertainty". en. In: *PLoS One* 9.4 (Apr. 2014), e94840. DOI: 10.1371/journal.pone.0094840.

[16] Hui Tang et al. "Semi-automatic MRI segmentation and volume quantification of intra-plaque hemorrhage". en. In: *Int. J. Comput. Assist. Radiol. Surg.* 10.1 (Jan. 2015), pp. 67–74. DOI: 10.1007/s11548-014-1010-3.

[17] Danilo Samuel Jodas, Aledir Silveira Pereira, and João Manuel R S Tavares. "Lumen segmentation in magnetic resonance images of the carotid artery". en. In: *Comput. Biol. Med.* 79 (Dec. 2016), pp. 233–242. DOI: 10.1016/j.compbiomed.2016.10.021.

[18] Andrés M Arias-Lorza et al. "Carotid Artery Wall Segmentation in Multispectral MRI by Coupled Optimal Surface Graph Cuts". en. In: *IEEE Trans. Med. Imaging* 35.3 (Mar. 2016), pp. 901–911. DOI: 10.1109/TMI.2015.2501751.

[19] Shan Gao et al. "Learning-based automated segmentation of the carotid artery vessel wall in dual-sequence MRI using subdivision surface fitting". en. In: *Med. Phys.* 44.10 (Oct. 2017), pp. 5244–5259. DOI: 10.1002/mp.12476.

[20] Andrés M Arias-Lorza et al. "Cooperative carotid artery centerline extraction in MRI". en. In: *PLoS One* 13.5 (May 2018), e0197180. DOI: 10.1371/journal.pone.0197180.

[21] Ranying Zhang et al. "Identification of high-risk carotid plaque with MRI-based radiomics and machine learning". en. In: *Eur. Radiol.* 31.5 (May 2021), pp. 3116–3126. DOI: 10.1007/s00330-020-07361-z.

[22] Yuxi Dong et al. *Identifying Carotid Plaque Composition in MRI with Convolutional Neural Networks*. 2017. DOI: 10.1109/SMARTCOMP.2017.7947015.

[23] Jiayi Wu et al. *Deep morphology aided diagnosis network for segmentation of carotid artery vessel wall and diagnosis of carotid atherosclerosis on black-blood vessel wall MRI*. 2019. DOI: 10.1002/mp.13739.

[24] Li Chen et al. "Automated Artery Localization and Vessel Wall Segmentation using Tracklet Refinement and Polar Conversion". en. In: *IEEE Access* 8 (Nov. 2020), pp. 217603–217614. DOI: 10.1109/access.2020.3040616.

[25] Li Chen et al. "Domain adaptive and fully automated carotid artery atherosclerotic lesion detection using an artificial intelligence approach (LATTE) on 3D MRI". en. In: *Magn. Reson. Med.* 86.3 (Sept. 2021), pp. 1662–1673. DOI: 10.1002/mrm.28794.

[26] Daniel D Samber et al. "Segmentation of carotid arterial walls using neural networks". en. In: *World J. Radiol.* 12.1 (Jan. 2020), pp. 1–9. DOI: 10.4329/wjr.v12.i1.1.

[27] Robin Camarasa et al. *Quantitative Comparison of Monte-Carlo Dropout Uncertainty Measures for Multi-class Segmentation*. 2020. DOI: 10.1007/978-3-030-60365-6_4.

[28] Dieuwertje Alblas, Christoph Brune, and Jelmer M Wolterink. "Deep Learning-Based Carotid Artery Vessel Wall Segmentation in Black-Blood MRI Using Anatomical Priors". In: (Dec. 2021). DOI: 10.48550/arXiv.2112.01137.

[29] Gaël Varoquaux and Veronika Cheplygina. "Machine learning for medical imaging: methodological failures and recommendations for the future". In: *NPJ digital medicine* 5.1 (2022), p. 48. DOI: 10.1038/s41746-022-00592-y.

[30] Ekaterina Kondrateva et al. "Domain Shift in Computer Vision models for MRI data analysis: An Overview". In: (Oct. 2020). DOI: 10.48550/arXiv.2010.07222.

[31] Russell T Shinohara et al. "Statistical normalization techniques for magnetic resonance imaging". In: *NeuroImage: Clinical* 6 (2014), pp. 9–19. DOI: 10.1016/j.nicl.2014.08.008.

[32] Moloud Abdar et al. "A review of uncertainty quantification in deep learning: Techniques, applications and challenges". In: *Information Fusion* 76 (2021), pp. 243–297. DOI: 10.1016/j.inffus.2021.05.008.

[33] Chenglu Zhu et al. *Complex carotid artery segmentation in multi-contrast MR sequences by improved optimal surface graph cuts based on flow line learning*. 2022. DOI: 10.1007/s11517-022-02622-z.

[34] Fabian Isensee et al. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation". en. In: *Nat. Methods* 18.2 (Feb. 2021), pp. 203–211. DOI: 10.1038/s41592-020-01008-z.

[35] Shishuai Hu, Zehui Liao, and Yong Xia. "Label Propagation for 3D Carotid Vessel Wall Segmentation and Atherosclerosis Diagnosis". In: (Aug. 2022). DOI: 10.48550/arXiv.2208.13337.

[36] Lisa Lavrova. "lavrovaliz/plaq-u-net:" in: (Aug. 2022).

[37] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015. DOI: 10.48550/arXiv.1505.04597.

[38] Tewodros Weldebirhan Arega, Stéphanie Bricq, and Fabrice Meriaudeau. "Leveraging Uncertainty Estimates to Improve Segmentation Performance in Cardiac MR". en. In: *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Perinatal Imaging, Placental and Preterm Image Analysis* (2021), pp. 24–33. DOI: 10.1007/978-3-030-87735-4_3.

[39] Suprosanna Shit et al. *clDice - a Novel Topology-Preserving Loss Function for Tubular Structure Segmentation*. 2021. DOI: 10.1109/CVPR46437.2021.01629.

[40] Annika Reinke et al. "Common Limitations of Image Processing Metrics: A Picture Story". In: (Apr. 2021). DOI: 10.48550/arXiv.2104.05642.

[41] Femke Vaassen et al. “Evaluation of measures for assessing time-saving of automatic organ-at-risk segmentation in radiotherapy”. en. In: *Phys Imaging Radiat Oncol* 13 (Jan. 2020), pp. 1–6. DOI: 10.1016/j.phro.2019.12.001.

[42] Pauli Fält et al. *Spectral Image Enhancement for the Visualization of Dental Lesions*. 2018. DOI: 10.1007/978-3-319-94211-7_53.

6

An open-source software package for medical imaging data curation and exploration

Elizaveta Lavrova*, Sergey Primakov*, Zohaib Salahuddin, Manon Beuque, Damon Verstappen, Henry C. Woodruff , Philippe Lambin; "*" – equal contribution.

Adapted from: Elizaveta Lavrova et al. "Precision-medicine-toolbox: An open-source python package for the quantitative medical image analysis". In: *Software Impacts* 16 (2023), p. 100508. DOI: 10.1016/j.simpa.2023.100508.

Abstract

Medical image analysis plays a key role in precision medicine. Data curation and pre-processing are critical steps in quantitative medical image analysis that can have a significant impact on the resulting performance of machine learning models. In this work, we introduce the precision-medicine-toolbox, allowing clinical and junior researchers to perform data curation, image pre-processing, radiomics extraction, and feature exploration tasks with a customizable Python package. With this open-source tool, we aim to facilitate the crucial data preparation and exploration steps, bridge the gap between the currently existing packages, and improve the reproducibility of quantitative medical imaging research.

Current version	v0.11
Repository	`github.com/primakov/precision-medicine-toolbox`
Reproducible Capsule	`codeocean.com/capsule/0396992/tree/v1`
Legal Code License	BSD-3-Clause
Versioning system	Git
Code language	Python
Dependencies	numpy 1.16.2, SimpleITK 0.9.1, PyWavelets 0.4.0, pykwalify 1.6.0, six 1.10.0, tqdm 4.40.2, pydicom 1.3.0, pandas 0.25.1, pyradiomics 2.2.0, scikit-image 0.14.2, ipywidgets 7.4.2, matplotlib 3.0.3, Pillow 5.4.1, scikit-learn 0.21.3, scipy 1.2.1, plotly 4.8.1, mkdocstrings 0.18.0, statsmodels 0.12.2, opencv-python 4.1.2.30, seaborn 0.11.1, pickle-mixin 1.0.2, openpyxl 3.0.7
Developer documentation	`precision-medicine-toolbox.readthedocs.io/`

6.1 Introduction

Precision medicine (PM) aims to enhance individual patient care by identifying subgroups of patients within a disease group using genotypic and phenotypic data, consequently targeting the disease with more efficient treatment [1]. Medical image analysis plays a key role in PM as it allows the clinicians to non-invasively identify phenotypes [2].
The number of medical imaging data to analyze is rising rapidly. Hence, there is a need for medical image analysis tools that can aid clinicians in meeting the challenges of rising demand and better

clinical performance, while reducing variability and costs. At the heart of these tools will be advanced quantitative imaging analysis, such as handcrafted radiomics and deep learning. Handcrafted radiomics is the high-throughput extraction of pre-defined high-dimensional quantitative image features and their correlation with clinical outcomes using machine learning methods [3]. Deep learning automatically learns representative image features from the high dimensional image data without the need for feature engineering by using non-linear modules that constitute a neural network [4]. The field of quantitative image analysis is expanding [5, 6, 7]. Moreover, it has demonstrated promising results in various clinical applications [8, 9, 10, 11]. As with many nascent technologies, high-throughput quantitative image analysis suffers from a lack of standardization, e.g. in the image domain (different vendors, acquisition and reconstruction protocols, pre-processing), or different definitions of handcrafted features (such as shape, intensity, and texture features). The spread of widely used open-source software such as Pyradiomics, allows the extraction of standard handcrafted radiomics features [12].
Data curation and the pre-processing of medical images are time-consuming and critical steps in the radiomics workflow that can have a significant impact on the resulting model performance [13, 14, 15]. These steps may be performed manually or using lower level python libraries such as Numpy [16], Pandas [17], Pydicom [18], Scikit-image [19], Scikit-learn [20], SimpleITK [21], Nibabel [22], or Scipy [23]. As most current data curation workflows necessitate time-consuming human input, this step becomes an error-prone bottleneck and adds to the current reproducibility problem. Moreover, it is important to perform an exploratory analysis to understand the link between the data used as input in a machine learning model with the outcome it has to predict. While there are tools available for the implementation of the radiomics pipeline such as Nipype [24], Pymia [25], and MONAI [26], there is also the need for a tool that allows for the systematic and standardized data curation, image pre-processing, and feature exploration during the development phase of the study.
We introduce the open-source precision-medicine-toolbox that

facilitates data curation, image pre-processing, and feature exploration using customizable Python scripts.

6.2 Implementation and architecture

As illustrated in Figure 6.1, dedicated base classes have been implemented for each dataset type to extract the corresponding data, as well as the associated metadata. The functionality classes inherit from the base classes. This approach allows for the separation of reading and processing tasks and makes it readily available for new data formats or functions.

The imaging module allows for pre-processing and exploration of the imaging datasets. It consists of the base DataSet class and the inheriting ToolBox class. The DataSet class reads the imaging data and the corresponding metadata and initializes a dataset object. The ToolBox is an inheriting class that enables functions for working with raw computed tomography (CT) or magnetic resonance (MR) imaging data. Currently, the following functions are implemented: dataset parameter exploration by parsing of the DICOM metadata, dataset basic quality examination by comparing imaging parameters to the user-defined threshold, conversion of DICOM data into volumetric Nearly Raw Rusted Data (NRRD), image basic pre-processing, unrolling NRRD images and region of interest (ROI) masks into Joint Photographic Experts Group (JPEG) slices for a quick check of co-registration between imaging data and masks, radiomics feature extraction from NRRD/MHA data using Pyradiomics [12]. The image and mask co-alignment pre-view example is illustrated in Figure 6.2.

The features module allows for the exploration of the feature datasets. It consists of the base FeaturesSet class and the inheriting AnalysisBox class. The FeaturesSet class reads the features data and the corresponding metadata and initializes a FeaturesSet object. The AnalysisBox class allows for the primary analysis of the features. Currently, the following functions are implemented: visualization of feature value distributions in classes and mutual Spearman correlation

matrix, calculation of corrected p-values for Mann-Whitney U-test for features mean values in groups, visualization of univariate receiver operating characteristic (ROC) curves for each feature and calculation of the area under the curve (AUC), volumetric analysis, calculation of basic statistics for every feature. Features distribution in classes visualization is illustrated in Figure 6.3.
The binary classification metrics reporting module allows for the generation of binary classification performance metrics given true labels and predicted probabilities.

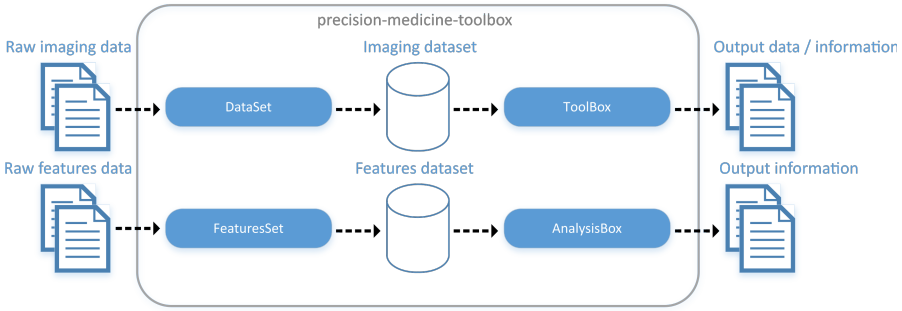


Figure 6.1: Organization of the precision-medicine-toolbox: The DataSet class takes an imaging dataset as in input and is inherited by the ToolBox class; the FeatiresSet class takes a features dataset as an input and is inherited by the AnalysisBox class.

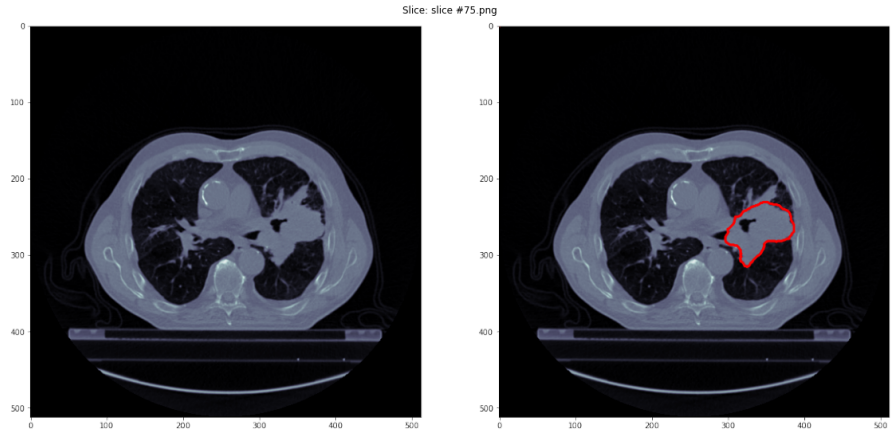


Figure 6.2: Example of the quick check of the segmentation alignment to the original scan by visualizing CT axial slices.

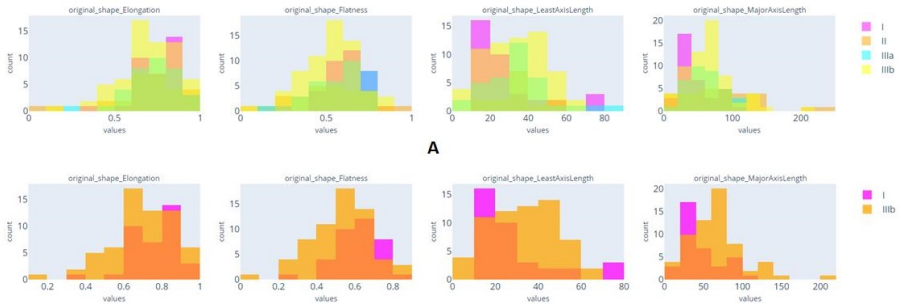


Figure 6.3: Feature value distributions in multiple classes: A - for all the presented classes, B - for the selected classes I and IIIb.

6.3 Quality control

To ensure that precision-medicine-toolbox meets the requirements, a continuous integration workflow is built in GitHub actions. Tests are run automatically after every new commit is pushed. Every time, the project is built and unit tests are performed for the latest Windows system on Python 3.7. Quick start and running software examples are described in the documentation. Additionally, code quality is reviewed with CodeFactor (`http://codefactor.io`). The API specifications for all the classes and methods are generated automatically from the source code annotations with Mkdocstrings (`https://mkdocstrings.github.io/`). This enables keeping documentation up to date with the latest developments of the package.

6.4 Software impacts

The functionality of the toolbox aims to meet some challenges that are specific to the radiomics field. One of these challenges is the lack of data and pipelines standardisation. Therefore, reproducibility is one of the key criterias for the radiomics studies.
The toolbox is mostly dedicated to radiomics analysis, as it allows for the handling of both raw imaging data and derivative features. Nevertheless, its modules can be used separately for other medical imaging research applications. The imaging module is applicable for deep learning tasks to prepare the imaging data and get information regarding the metadata. The features module can be used for any tabular data analysis, such as health records or histology-derived features.

The toolbox was utilized and tested during the development of multiple projects including automatic lung tumor segmentation on the CT [27], repeatability of breast MRI radiomic features [28], and radiomic-based diagnosis of multiple sclerosis [29].
The development of precision-medicine-toolbox aims for the democratization of the machine learning and deep learning pipelines

for researchers without strong programming skills. Additionally, it drives a programming community effort to improve this package and add its own variables and methods. Therefore, user contributions are very welcome.

6.5 Conclusions and future works

The development of the precision-medicine-toolbox aims to lower the entry barrier for researchers who are starting to work in medical imaging. Moreover, it provides an open-source solution for the researchers who already have their inhouse workflow of managing data to increase the reproducibility of the quantitative medical imaging research. We would also like to encourage the community to improve this open-source toolbox by contributing to it.

References

[1] Tianye Niu et al. "Pathways to radiomics-aided clinical decision-making for precision medicine". In: *Radiomics and Radiogenomics*. Chapman and Hall/CRC, 2019, pp. 193–201.

[2] U Rajendra Acharya et al. "Towards precision medicine: from quantitative imaging to radiomics". en. In: *J. Zhejiang Univ. Sci. B* 19.1 (Jan. 2018), pp. 6–24. DOI: 10.1631/jzus.B1700260.

[3] Philippe Lambin et al. "Radiomics: extracting more information from medical images using advanced feature analysis". In: *European journal of cancer* 48.4 (2012), pp. 441–446. DOI: 10.1016/j.ejca.2011.11.036.

[4] Jürgen Schmidhuber. "Deep learning in neural networks: an overview". en. In: *Neural Netw.* 61 (Jan. 2015), pp. 85–117. DOI: 10.1016/j.neunet.2014.09.003.

[5] Ohad Oren, Bernard J Gersh, and Deepak L Bhatt. "Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints". en. In: *Lancet Digit Health* 2.9 (Sept. 2020), e486–e488. DOI: 10.1016/S2589-7500(20)30160-6.

[6] Ravi Aggarwal et al. "Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis". en. In: *NPJ Digit Med* 4.1 (Apr. 2021), p. 65. DOI: 10.1038/s41746-021-00438-z.

[7] S Kevin Zhou et al. "A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises". In: *Proc. IEEE Inst. Electr. Electron. Eng.* 109.5 (May 2021), pp. 820–838. DOI: 10.1109/JPROC.2021.3054390.

[8] Alberto Stefano Tagliafico et al. "Overview of radiomics in breast cancer diagnosis and prognostication". en. In: *Breast* 49 (Feb. 2020), pp. 74–80. DOI: 10.1016/j.breast.2019.10.018.

[9] Wei Mu et al. "Non-invasive decision support for NSCLC treatment using PET/CT radiomics". en. In: *Nat. Commun.* 11.1 (Oct. 2020), p. 5228. DOI: 10.1038/s41467-020-19116-x.

[10] Yucheng Zhang et al. "Radiomics-based Prognosis Analysis for Non-Small Cell Lung Cancer". en. In: *Sci. Rep.* 7 (Apr. 2017), p. 46349. DOI: 10.1038/srep46349.

[11] Shouchao Wang et al. "Radiomics Analysis Based on Magnetic Resonance Imaging for Preoperative Overall Survival Prediction in Isocitrate Dehydrogenase Wild-Type Glioblastoma". en. In: *Front. Neurosci.* 15 (2021), p. 791776. DOI: 10.3389/fnins.2021.791776.

[12] Joost JM Van Griethuysen et al. "Computational radiomics system to decode the radiographic phenotype". In: *Cancer research* 77.21 (2017), e104–e107. DOI: 10.1158/0008-5472.CAN-17-0339.

[13] Xenia Fave et al. “Impact of image preprocessing on the volume dependence and prognostic potential of radiomics features in non-small cell lung cancer”. In: *Transl. Cancer Res.* 5.4 (Aug. 2016), pp. 349–363. DOI: 10.21037/tcr.2016.07.11.

[14] Ruiping Zhang et al. “Potential feature exploration and model development based on 18F-FDG PET/CT images for differentiating benign and malignant lung lesions”. en. In: *Eur. J. Radiol.* 121 (Dec. 2019), p. 108735. DOI: 10.1016/j.ejrad.2019.108735.

[15] Seyyed Ali Hosseini et al. *The Impact of Preprocessing on the PET-CT Radiomics Features in Non-small Cell Lung Cancer*. 2021.

[16] Stefan van der Walt, S Chris Colbert, and Gael Varoquaux. “The NumPy Array: A Structure for Efficient Numerical Computation”. In: *Computing in Science Engineering* 13.2 (Mar. 2011), pp. 22–30. DOI: 10.1109/MCSE.2011.37.

[17] Wes McKinney. “Data Structures for Statistical Computing in Python”. In: *Proceedings of the 9th Python in Science Conference*. Austin, Texas: SciPy, 2010.

[18] D Mason. “SU-E-T-33: Pydicom: An Open Source DICOM Library”. In: *Med. Phys.* 38.6Part10 (June 2011), pp. 3493–3493. DOI: 10.1118/1.3611983.

[19] Stéfan van der Walt et al. “scikit-image: image processing in Python”. en. In: *PeerJ* 2 (June 2014), e453. DOI: 10.7717/peerj.453.

[20] Oliver Kramer. “Scikit-Learn”. In: *Machine Learning for Evolution Strategies*. Ed. by Oliver Kramer. Cham: Springer International Publishing, 2016, pp. 45–53.

[21] Ziv Yaniv et al. “SimpleITK Image-Analysis Notebooks: a Collaborative Environment for Education and Reproducible Research”. en. In: *J. Digit. Imaging* 31.3 (June 2018), pp. 290–303. DOI: 10.1007/s10278-017-0037-8.

[22] Matthew Brett et al. *nipy/nibabel: 5.1.0*. Version 5.1.0. Apr. 2023. DOI: 10.5281/zenodo.7795644.

[23] Pauli Virtanen et al. "SciPy 1.0: fundamental algorithms for scientific computing in Python". In: *Nature methods* 17.3 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[24] Krzysztof Gorgolewski et al. "Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python". In: *Frontiers in neuroinformatics* (2011), p. 13. DOI: 10.3389/fninf.2011.00013.

[25] Alain Jungo et al. "pymia: A Python package for data handling and evaluation in deep learning-based medical image analysis". en. In: *Comput. Methods Programs Biomed.* 198 (Jan. 2021), p. 105796. DOI: 10.1016/j.cmpb.2020.105796.

[26] M Jorge Cardoso et al. "MONAI: An open-source framework for deep learning in healthcare". In: *arXiv preprint arXiv:2211.02701* (2022). DOI: 10.48550/arXiv.2211.02701.

[27] Sergey P Primakov et al. "Automated detection and segmentation of non-small cell lung cancer computed tomography images". In: *Nature communications* 13.1 (2022), pp. 1–12. DOI: 10.1038/s41467-022-30841-3.

[28] Renée WY Granzier et al. "Test–Retest Data for the Assessment of Breast MRI Radiomic Feature Repeatability". In: *Journal of Magnetic Resonance Imaging* 56.2 (2022), pp. 592–604. DOI: 10.1002/jmri.28027.

[29] Elizaveta Lavrova et al. "Exploratory radiomic analysis of conventional vs. quantitative brain MRI: toward automatic diagnosis of early multiple sclerosis". In: *Frontiers in neuroscience* 15 (2021), p. 679941. DOI: 10.3389/fnins.2021.679941.

7

General discussion and prospectives

7.1 General discussion

The field of radiomics has garnered significant interest and is primarily being explored in the context of oncology. Its primary objective in oncology is to enhance diagnostic precision and optimize treatment outcomes by associating quantitative imaging features with the tumor phenotype [1, 2]. Moreover, radiomics has shown promise in other medical domains through numerous proof-of-concept studies, highlighting its potential prognostic or predictive capabilities for various diseases beyond oncology.

This thesis represents a contribution to the application of radiomics in the realm of neurological diseases. Neurological conditions, both in the young and elderly population, are a leading cause of disability and the second leading [3] cause of death globally [4]. Despite the lack of effective treatment for most neurological diseases, early and accurate diagnosis is critical to slow down the disease progression and enhance the quality of life for patients and their families. The implementation of radiomics in the neurological field offers a promising tool to achieve these goals [3].

In the field of neurology, a majority of studies have focused on extracting image features after diagnosis to derive disease-specific characteristics, with some even aiming to distinguish between different diagnoses or subtypes [5, 6, 7]. However, a subset of studies has taken a longitudinal approach, extracting features from early-stage patients to predict disease progression outcomes during follow-up [8, 9, 6]. This longitudinal analysis may contain additional diagnostic information, potentially enabling the assessment of disease development or progression and facilitating the timely initiation of treatment. In this thesis, we investigated the following hypothesis stated in Chapter 1: Radiomic features extracted from brain regions can offer valuable information regarding brain tissue pathology that cannot be retrieved by the human eye. This investigation sought to demonstrate the potential of radiomics in providing crucial diagnostic information beyond traditional visually detectable biomarkers. The findings from this study could pave the way for enhanced early

diagnosis and intervention, ultimately improving patient outcomes and quality of life for individuals living with neurological diseases.

Chapter 2 of the thesis demonstrates the predictive power of radiomics in neuro-oncology, particularly in estimating the critical 1p/19q co-deletion status for low-grade glioma treatment selection. As the corresponding tests for 1p/19q co-deletion status are invasive and costly, the utilization of conventional handcrafted radiomics proves to be a promising non-invasive alternative. By analyzing T1w and T2w MR images, a proof-of-concept radiomic signature is derived, offering accurate estimations of the 1p/19q co-deletion status. The robustness of this signature is further validated using external datasets, solidifying its reliability for clinical application. In the pursuit of expanding oncological methodology to neurology, this work serves as a pivotal bridge. While adopting conventional oncological steps, such as dealing with the gross tumor volume, this study also incorporates essential neurological steps, such as image co-registration, bias field correction, and brain tissue segmentation. This chapter also has the potential to confirm the hypothesis of this thesis as the human eye cannot stratify the MR scans by 1p/19q co-deletion.

Subsequently, within the realm of glioma classification, novel approaches for feature selection have emerged [10]. In the domain of predicting the molecular expression status of gliomas, recent studies have focused on forecasting chromosome 7 gain and chromosome 10 loss in IDH wild-type histologically low-grade gliomas [11]. Additionally, fresh radiomics pipelines have been introduced for estimating 1p/19q co-deletion status, both without [12] and with [13] external validation.

Chapter 3 provides a comprehensive review of the implementation of handcrafted and deep radiomics in non-oncological clinical neuroimaging. The focus of the first part is on summarizing general recommendations in the field, particularly for neurological data, which is typically multi-modal and requires specific pre-processing steps such as data co-registration and intensity normalization. In the context of handcrafted radiomics, choosing the region of interest

is a complex task that demands expert knowledge of the brain regions affected by specific impairments. Nevertheless, radiomics has the potential to aid in localizing new regions of interest. Data segmentation, while challenging and requiring expertise, benefits from available automated tools.

The chapter proceeds to offer an overview of radiomics studies in major neurological diseases, including Alzheimer's disease, multiple sclerosis, Parkinson's disease, stroke, and psychiatric diseases. These studies primarily serve as a proof-of-concept, often with limited sample sizes, retrospective designs, and lacking external validation. While successful in binary classification between disease and normal controls, their lack of specificity is evident, as real biological populations may exhibit diverse diagnoses. However, the high classification scores, reaching ROC AUC of about 90%, showcase the potential of radiomics in distinguishing between disease and healthy individuals.

The review concludes by summarizing the current challenges hindering the widespread implementation of radiomics in clinical practice. One of the main challenges lies in data availability, as many studies rely on data acquired within the same hospital without an external hold-out testing subset. Despite the delicate nature of clinical data, the use of publicly available datasets can enhance trust and reproducibility. Additionally, data harmonization is crucial, considering the domain-specific biases present in medical imaging data from different hospitals and scanners. Another major challenge is the lack of standardization in the field of radiomics, resulting in a lack of reproducibility and comparability across studies. To address this, proposed solutions include careful reporting, adherence to reporting guidelines such as RQS and TRIPOD, and study standardization as suggested by IBSI. Furthermore, the availability of open-source tools for automated analysis can contribute to overcoming these challenges. The subsequent chapters delve into potential solutions for these challenges, paving the way for the broader application of radiomics in neurological clinical practice.

In Chapter 4, a proof-of-concept study was conducted to explore

the utility of radiomics in distinguishing between multiple sclerosis patients and normal controls. The study involved developing radiomic signatures using both T1w data and quantitative MRI (qMRI) maps. The analysis encompassed features extracted from gray matter, white matter, and normal-appearing white matter (NAWM) – white matter excluding lesions. Notably, the T1w-derived signature was externally validated using two open-source datasets, enhancing the robustness of the findings. The results of this study demonstrated the discriminative power of radiomics even in NAWM, which appears normal to the human eye. This outcome confirms the thesis hypothesis that radiomics can extract quantitative information about the underlying pathology from visually normal imaging data. This ability to unveil hidden disease-related features in NAWM is a promising step toward early detection and improved management of multiple sclerosis.

Subsequent publications in the field of radiomics in multiple sclerosis have primarily focused on grading disease severity rather than distinguishing between multiple sclerosis patients and healthy controls. In [14], features were meticulously analyzed to identify correlations with the disability status of the patients. In [15], a model was developed to predict the annualized relapse rate for individuals with multiple sclerosis.

Additionally, Chapter 4 shed light on a potential solution to one of the challenges highlighted in the thesis - the variability of imaging data across different sites, scanners, and protocols. By leveraging a unique dataset containing not only conventional clinical MR sequences but also unique qMRI mapping, the study attempted to overcome the data variability issue. The qMRI maps were obtained through a complex multi-image acquisition protocol, allowing for the reconstruction of physical properties in corresponding voxels. As a result, this approach not only provided a richer dataset and yielded higher classification scores but also produced potentially more standardized data across different equipment and sites due to the physical nature of the data [16]. It was shown, that neurology can benefit a lot from qMRI, but histological confirmation will build a stronger foundation

for the qMRI findings [17]. However, standardization is needed in the field of qMRI as well [18].
In Chapter 5, we addressed another challenge identified in Chapter 3 - the need for accurate and stable data segmentation. To overcome this challenge, we proposed and implemented a deep learning method to tackle the specific segmentation of the carotid arteries in neck scans of stroke patients. The model was trained on multi-center data originating from the same scanner model. To validate the model's performance, we tested it on data from different MR sequences, including after-contrast data, and validated it on an external dataset from a different hospital and scanner. Notably, manual annotations on some sequences had slight shifts due to minor co-registration issues related to biological motion. Therefore, our focus extended beyond achieving a perfect overlap with the ground truth to ensuring consistency in clinically relevant derivatives, such as carotid artery volume. To address co-registration issues in certain voxels, we introduced an uncertainty-based component in the loss function. This innovative approach enabled the model to produce robust and reliable segmentations, outperforming the current state-of-the-art model (nnU-Net) in MR sequences different from the development set and in the external dataset. The results of this work demonstrate the applicability of deep learning approaches in resolving data labeling challenges in a robust and automated manner. By leveraging the power of deep learning, we offer a valuable solution to enhance the accuracy and efficiency of data segmentation in clinical neuroimaging. The subsequent stage in carotid artery segmentation involves quality assurance for the segmentation process. This was achieved through uncertainty estimation in [19]. The anticipated future steps in the application of AI in carotid artery imaging include the development of dependable predictive models for forecasting recurrent stroke.
In Chapter 3, it became evident that radiomics faces a significant challenge related to the lack of reproducibility in current studies, mainly stemming from the diversity of pipeline solutions and undocumented pre-analysis steps, such as data curation and exploration. To address this critical issue, in Chapter 6, we

present the precision-medicine-toolbox, an open-source Python package specifically designed for medical imaging data curation and exploratory analysis. The precision-medicine-toolbox aims to bridge the gap between existing open-source packages for radiomics studies and improve the understanding of the underlying data. By providing a standardized and community-driven tool, we envision enhancing the reproducibility of radiomics studies and democratizing the implementation of radiomics in a cross-disciplinary environment. This toolbox compromises a comprehensive resource for researchers and clinicians, facilitating robust data curation and thorough exploratory analysis. It enables users to perform critical pre-processing steps, data quality checks, and data exploration in a standardized and transparent manner. Moreover, the open-source nature of the toolbox fosters collaboration and contributions from the radiomics community, ensuring continuous improvement and adaptability to emerging research needs. Through this collective effort, we seek to establish a more cohesive and standardized approach to radiomics, promoting data-driven advancements in neurology and beyond.

The infrastructure for radiomics research has witnessed continuous expansion, transitioning from individual tools to more integrated solutions. In [20], a physician-centered cloud platform was introduced to streamline the creation of radiomics-based predictive models.

7.2 Prospectives

Over the past decade, the implementation of artificial intelligence in medical image analysis, particularly in radiomics, has experienced remarkable growth. Initially, handcrafted radiomics and later deep radiomics found success in oncology, where relatively large and accurately labeled datasets from the oncological patient management pipeline enabled meaningful studies. However, critical revisions of the field over the last years have highlighted the current obstacles hindering radiomics from becoming part of clinical routine. Despite

these challenges, radiomics has garnered increasing interest in other medical domains, including neurology. In contrast to oncology, neurological pathologies exhibit greater diversity, yet datasets are often smaller and data exchange less prevalent. In the technology development cycle, oncological radiomics has gone through a phase of critical reflection and is now getting to the "slope of enlightenment," benefitting from the lessons learned as illustrated on Figure 7.1 [21]. Conversely, as indicated in Chapter 3, neurological radiomics is currently under academic interest and producing proof-of-concept studies with smaller single-center datasets of retrospective nature, still experiencing the "peak of inflated expectations".

In this thesis, the primary focus has been on comprehending

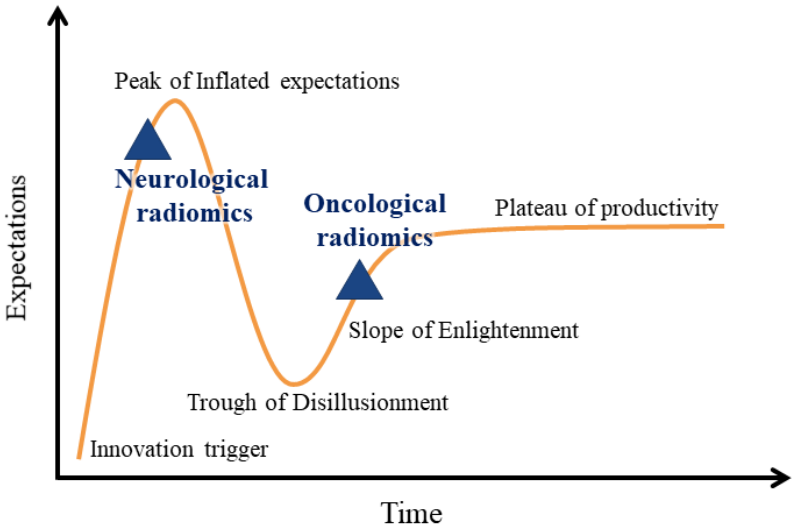


Figure 7.1: Phases of field development.

and addressing the limitations of radiomics in neuroimaging. By identifying and tackling the challenges unique to neurological data, we strive to move this field of research toward the "plateau of productivity". The work presented here delves into understanding the potential of radiomics in neurological diseases, investigating

predictive capabilities, proposing solutions to data labeling challenges, and offering tools for data curation and analysis.

7.2.1 Data availability

In the future, researching larger, heterogeneous, multi-centric prospective datasets will be crucial to improving the generalizability of radiomics methodology. This approach will allow testing the methodology across different populations, reducing potential biases, and providing reliable sensitivity and specificity scores in representative data. External validation is already standard practice in advanced studies, ensuring that the approaches are robust and applicable beyond the training data. However, external validation performance scores should be interpreted carefully since the external validation data is just a snapshot of a certain population acquired with a certain equipment at a certain time point. However, corresponding validation methods can be implemented for different validation goals such as internal validity, temporal generalizability, geographical and domain generalizability [22]. Lately, continuous recurrent validation has been proposed to deal with the data drift over time and locations [23]. Longitudinal studies will play a pivotal role, enabling researchers to delve into more relevant research tasks related to disease development and clinical outcomes prediction. This shift from current status estimation to early diagnosis and treatment selection will significantly enhance patient care and management. To achieve these research goals, stronger collaboration among institutions in the field is essential. By pooling resources and sharing data, researchers can access larger and more diverse datasets, further enhancing the accuracy and relevance of radiomics studies. In the ideal scenario, increased availability of research data will facilitate not only methodological development but also the reproducibility and reliability of reported results. Researchers will have access to a wealth of data to validate and refine their approaches, contributing to a more

robust and credible body of radiomics research.
However, data sharing is an essential subsequent challenge due to the legal and ethical aspects linked to sensitive medical data. A strategy to walk around and increase the amount of available training data is the generation of synthetic data [24]; however, this approach necessitates some attention to ensure the faithful preservation of realistic anatomical and biological features, as well as the accurate introduction of confounding factors inherent to the studies. Data privacy challenges might be addressed through the implementation of transfer or federated learning techniques, wherein data remains localized at respective sites while being employed to train models transmitted between these sites, as in [25]. In instances where data is available but relevant labels are missing, self-supervised learning is a potential method to mitigate the problem [26].

7.2.2 Study design

Improving the study design represents a crucial direction for the advancement of the radiomics field. Transitioning from cross-sectional to longitudinal studies will enable researchers to tackle predictive tasks, providing valuable insights into disease progression and clinical outcomes. Similarly, shifting from binary classification to multi-class classification will address real-world diagnostic challenges, where identifying the specific diagnosis is essential, enhancing the specificity and clinical relevance of radiomics. Furthermore, emphasizing regression tasks will quantitatively assess the patient's condition, offering a broader spectrum of diagnostic information beyond binary outcomes. Incorporating domain knowledge into study design will play a pivotal role in justifying the selection of the most appropriate imaging modality and region of interest. Given the wealth of domain-specific knowledge in neurology, this integration can significantly improve the accuracy and precision of radiomics-based analyses. Data fusion techniques

will further enhance the significance and performance of imaging biomarkers. By combining information from multiple modalities and sources, researchers can derive more comprehensive and insightful radiomic signatures, potentially capturing complex relationships and disease mechanisms. However, it is essential to critically investigate the added value of each radiomic signature. Comparing models with and without radiomics should be standard practice to ensure that the incorporation of radiomics truly contributes to improved diagnostic accuracy and clinical decision-making.

The enhancement of model performance does not always depend on new and more sophisticated architectures. The important step is the optimization of model parameters, as in some cases a seemingly less powerful model with better-tuned parameters can outperform a stronger but untailored model. Traditionally, ablation studies have provided a robust justification for hyperparameter selection. During the era of handcrafted radiomics, [27] suggested a framework that experimented with diverse model configurations and radiomic pipeline steps across various benchmark datasets, ultimately selecting the most effective combination. This concept has since evolved significantly within the domain of deep learning. Innovations like nnU-Net, which extends the traditional U-Net architecture by incorporating configuration parameters inferred from the training data, have emerged [28]. Consequently, this evolution has joined the field of AutoML, a methodology that leverages machine learning to autonomously determine the optimal AI model setup and hyperparameters, often employing techniques like neural architecture search [29, 30]. Nonetheless, it's imperative to acknowledge that while AutoML offers powerful capabilities for automating parameter selection, it comes at a cost. Both in terms of computational resources and data requirements, AutoML can be quite demanding.

In the context of implementing AI in medical imaging, our focus should not solely be on achieving higher AI-related technical scores for their own sake, but rather on the pursuit of improved diagnostic or predictive accuracy. While performance metrics like the Dice score, ROC AUC, or MAE are valuable and provide quantitative insights

into how well the AI model solves the statistical problem, it's crucial to understand that the ultimate goal is to enhance the model's ability to aid in clinical problems. By introducing clinically relevant scores, we align the development of AI models with their practical utility in the hospital setting.
As it was shown in, [31], there is no such thing as a fully validated model. Even with an abundance of data, its utility can be limited if strong domain shifts exist, leading to challenges in generalizing the radiomics methodology across different imaging parameters. To overcome these challenges, further research is needed in the field of data harmonization and feature reproducibility, particularly when dealing with a wide range of imaging parameter variations. Standardization efforts, such as following recommendations from organizations like IBSI [32], can play a vital role in ensuring consistency and comparability across radiomics studies. Phantom and test-retest studies are crucial to test the stability and reliability of radiomic signatures. These studies provide valuable insights into the robustness of the radiomics approach and help identify potential sources of variability. In this thesis (Chapter 3), it was demonstrated that radiomic signatures can benefit from the use of quantitative mapping, where each voxel represents a physical property of the corresponding volume. This approach not only enhances performance but also provides valuable insights into the relationship between the physical and biological properties of tissues. However, it is essential to continue the development of quantitative imaging methods to ensure that different data acquisition and image reconstruction protocols yield consistent and comparable data.

7.2.3 Technological advances

The advancement of radiomics is closely tied to progress in technology, encompassing developments in both hardware and software domains. Research has demonstrated that more powerful

MRI scanners can significantly reduce the acquisition time and acquire scans of higher quality [33].
The field of AI is advancing rapidly, with recent achievements such as visual transformers [34], ChatGPT [35], and Dall-E [36] capturing our attention and highlighting the significance of foundation models. These models are trained on extensive datasets in a self-supervised fashion, demonstrating their capability to handle multi-modal data and accommodate situations with limited task-specific labels. They can be fine-tuned for a wide array of downstream tasks. The potential applications of these models in medical AI, particularly in the realm of medical imaging, have been extensively discussed [37], and initial implementations have already been carried out [38, 39]. Even though foundation models are very promising, their development and implementation are associated with a large number of challenges, including interpretability, downstream tasking without losing knowledge, lack of effective benchmarking, and extensive data and computational demands [40]. Some proof-of-concept generalist biomedical AI systems are introduced together with multimodal biomedical benchmarks [41]. However, at the moment, supervised specialist models gain higher performance, and for widespread adoption, the primary requirement is access to large-scale medical datasets.

7.2.4 AI adoption

To increase AI acceptance in hospitals, models have to be not only effective but also transparent. For explainability of the model outcomes, XAI has to be implemented [42, 43, 44]. However, this step is necessary mostly for the research setup, during the model development and adoption. In a clinical setup, the demand for XAI explanations might be needed in non-trivial cases, such as clinical cases where the AI outcome contradicts clinical data.
For the continued advancement of radiomics in neuroimaging, stan-

dardization and transparent reporting throughout the pipeline implementation are imperative. To make the data and data-driven solutions sustainable and therefore trustful, the following four principles have to be maintained: Findability, Accessibility, Interoperability, and Reusability (FAIR) [45]. Adhering to guidelines such as RQS and TRIPOD will ensure a more consistent and reliable interpretation of study results. Transparent reporting practices will enable better collaboration among research groups, facilitating the exchange of both successful and unsuccessful practices. By embracing transparent reporting, researchers can collectively benefit from shared experiences and knowledge, enhancing the justification for different implementation steps within the radiomics pipeline. This collaborative effort will lead to more robust and reproducible radiomics studies, ultimately bolstering the credibility and applicability of radiomics in neurology and medical imaging. Community-driven tools will play a vital role in fostering standardization and transparency. Open-source tools will empower researchers with the resources to conduct radiomics studies in a standardized and reproducible manner.
Finally, while the accomplishments of AI in radiology are remarkable and have demonstrated that certain models can outperform radiologists [46], it is crucial to emphasize that AI is not meant to supplant hospital experts. Human radiologists possess a deeper understanding of atypical cases and a richer context than AI models. Additionally, humans are the ones to make decisions due to legal and ethical reasons. Instead, the objective of AI is to assume responsibility for routine tasks, enhance the capabilities of human radiologists, and reshape their roles [47, 48]. In essence, radiologists are transitioning into radiologists who leverage AI as a valuable tool in their practice.

7.2.5 Currently commercially available tools

Whereas academic research is aiming to bring radiomics into clinical practice, several AI-based commercial software products have already emerged within radiology. Information on these

products is aggregated at the www.aiforradiology.com website, encompassing details like imaging specialty, target disease, imaging modality, data format, processing time, pricing, and FDA and CE clearance [49]. As of 23.10.2023, the site documented 74 neuroradiology products. Notably, a majority of these tools (40 products) are primarily focused on image quantification, particularly involving brain or tissue segmentation and volume calculations. Additionally, some tools (15 products) specialize in pathology detection and classification, while a smaller subset (3 products) is dedicated to image processing and reporting tasks.

However, a limited number of tools address higher-level tasks such as disease grading (5 products), clinical predictions (2 products), and biomarker discovery (1 product). It's noteworthy that existing research highlights a lack of scientific evidence regarding the efficacy of many current tools, and transparency within the field remains an ongoing concern [49, 50]. However, [51] is an example of an extensive clinical validation of the MS monitoring tool. AI is progressively entering neurological clinical practices, at the moment with a predominant focus on enhancing scans, ROI segmentation, and grading.

The translation of AI advancements into clinical practice entails not only software and hardware engineering but also regulatory considerations and a diverse range of specialized expertise. The main medico-legal aspects to be addressed are data privacy, marketing authorization, and medical malpractice liability [52]. The field of medical AI has gained more government regulations in the last few years. In [53], an extensive overview of the risks and risk assessment methodologies is given. Finally, even though the potential of AI in radiology is large, we need to evaluate the added value of AI in real-world scenarios by assessing changes in the quality, efficiency, and costs of healthcare [54]. It is relevant for the detection and quantification tools as well, therefore development of the new products should include their integration into clinical workflow [55]. Collaborative efforts between academia and industry can significantly propel development by combin-

ing knowledge and infrastructure across this interdisciplinary domain.

7.3 Concluding remarks

Numerous studies have been conducted to explore the potential of radiomics in oncology, with some focus on neuro-oncology. However, due to data availability and the unmet clinical need for imaging biomarkers in other medical domains, radiomics has started gaining traction in other diseases, including neurology.
This thesis provides a comprehensive overview of the neurological radiomics pipeline, its clinical applications, and the current challenges in the field. Further, it demonstrates the proof-of-concept capability of radiomics to derive pathological information for neurological tasks, even from normal-appearing tissues. Despite these promising findings, neurological studies still face numerous limitations.
In an attempt to address some of these limitations, this thesis utilized quantitative mapping of the brain, providing an accurate measurement rather than conventional imaging. Additionally, we proposed an automated and robust method for data labeling through image segmentation, tackling the challenge of time-consuming and not stable data labeling. Finally, we introduced a solution to the pipeline standardization challenge, presenting an open-source software tool for imaging data curation and exploration.
However, as the field continues its development phase, there is still ample room for improvement. We outlined other challenges that remain, such as the need for extensive and representative data availability, data harmonization in the presence of diverse datasets, improvement in study design to utilize high-quality data, and standardization of the radiomics pipeline.
Addressing these challenges will drive the field of radiomics toward clinical studies and eventually clinical implementation, enabling early diagnosis and personalized treatment selection, ultimately improving patient care and outcomes in neurology. As radiomics

continues to evolve and overcome these hurdles, it holds the potential to revolutionize precision medicine in neurology.

References

[1] Hugo J W L Aerts et al. "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach". en. In: *Nat. Commun.* 5 (June 2014), p. 4006. DOI: 10.1038/ncomms5006.

[2] Johannes Müller et al. "Radiomics-based tumor phenotype determination based on medical imaging and tumor microenvironment in a preclinical setting". en. In: *Radiother. Oncol.* 169 (Apr. 2022), pp. 96–104. DOI: 10.1016/j.radonc.2022.02.020.

[3] K Rupabanta Singh and Sujata Dash. "Early detection of neurological diseases using machine learning and deep learning techniques: A review". In: *Artificial Intelligence for Neurological Disorders*. Elsevier, 2023, pp. 1–24. DOI: 10.1016/B978-0-323-90277-9.00001-8.

[4] Valery L Feigin et al. "The global burden of neurological disorders: translating evidence into policy". In: *The Lancet Neurology* 19.3 (2020), pp. 255–265. DOI: 10.1016/S1474-4422(19)30411-9.

[5] Maria A Rocca et al. "Deep Learning on Conventional Magnetic Resonance Imaging Improves the Diagnosis of Multiple Sclerosis Mimics". en. In: *Invest. Radiol.* 56.4 (Apr. 2021), pp. 252–260. DOI: 10.1097/RLI.0000000000000735.

[6] Fan Yang et al. "Combining PET with MRI to improve predictions of progression from mild cognitive impairment to Alzheimer's disease: an exploratory radiomic analysis study". en. In: *Ann Transl Med* 10.9 (May 2022), p. 513. DOI: 10.21037/atm-21-4349.

[7] Kun Zhao et al. “Regional Radiomics Similarity Networks Reveal Distinct Subtypes and Abnormality Patterns in Mild Cognitive Impairment”. en. In: *Adv. Sci.* 9.12 (Apr. 2022), e2104538. DOI: 10.1002/advs.202104538.

[8] Ruijiang Li et al. *Radiomics and radiogenomics: technical basis and clinical applications*. CRC press, 2019.

[9] Anees Abrol et al. “Deep residual learning for neuroimaging: An application to predict progression to Alzheimer’s disease”. en. In: *J. Neurosci. Methods* 339 (June 2020), p. 108701. DOI: 10.1016/j.jneumeth.2020.108701.

[10] Longfei Li et al. “Universal multi-factor feature selection method for radiomics-based brain tumor classification”. In: *Computers in Biology and Medicine* (2023), p. 107122. DOI: 10.1016/j.compbiomed.2023.107122.

[11] Xin Kong et al. “Development of a nomogram based on radiomics and semantic features for predicting chromosome 7 gain/chromosome 10 loss in IDH wild-type histologically low-grade gliomas”. In: *Frontiers in Oncology* 13 (2023), p. 1196614. DOI: 10.3389/fonc.2023.1196614.

[12] Yingping Li et al. “Radiomics-based method for predicting the glioma subtype as defined by tumor grade, IDH mutation, and 1p/19q codeletion”. In: *Cancers* 14.7 (2022), p. 1778. DOI: 10.3390/cancers14071778.

[13] Shingo Kihira et al. “Multi-Parametric Radiomic Model to Predict 1p/19q Co-Deletion in Patients with IDH-1 Mutant Glioma: Added Value to the T2-FLAIR Mismatch Sign”. In: *Cancers* 15.4 (2023), p. 1037. DOI: 10.3390/cancers15041037.

[14] Valentina Nepi et al. “MRI-Based Radiomics Analysis for Identification of Features Correlated with the Expanded Disability Status Scale of Multiple Sclerosis Patients”. In: *International Conference on Image Analysis and Processing*. Springer. 2022, pp. 362–373. DOI: 10.1007/978-3-031-13321-3_32.

[15] Sijia Du et al. “Deep learning-based PET/MR radiomics for the classification of annualized relapse rate in multiple sclerosis”. In: *Multiple Sclerosis and Related Disorders* 75 (2023), p. 104750. DOI: 10.1016/j.msard.2023.104750.

[16] René-Maxime Gracien et al. “How stable is quantitative MRI?–Assessment of intra-and inter-scanner-model reproducibility using identical acquisition sequences and data analysis programs”. In: *NeuroImage* 207 (2020), p. 116364. DOI: 10.1016/j.neuroimage.2019.116364.

[17] Alexander Seiler et al. “Multiparametric Quantitative MRI in Neurological Diseases”. en. In: *Front. Neurol.* 12 (Mar. 2021), p. 640239. DOI: 10.3389/fneur.2021.640239.

[18] Kathryn E Keenan et al. “Recommendations towards standards for quantitative MRI (qMRI) and outstanding needs”. In: *Journal of magnetic resonance imaging: JMRI* 49.7 (2019), e26. DOI: 10.1002/jmri.26598.

[19] Elina Thibeau-Sutre et al. “Uncertainty-based quality assurance of carotid artery wall segmentation in black-blood MRI”. English. In: *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging (UNSURE)*. Aug. 2023.

[20] Daniel Abler et al. “QuantImage v2: a comprehensive and integrated physician-centered cloud platform for radiomics and machine learning research”. In: *European Radiology Experimental* 7.1 (2023), pp. 1–13. DOI: 10.1186/s41747-023-00326-z.

[21] Doug Henton and Kim Held. “The dynamics of Silicon Valley: Creative destruction and the evolution of the innovation habitat”. en. In: *Soc. Sci. Inf.* 52.4 (Dec. 2013), pp. 539–557. DOI: 10.1177/0539018413497542.

[22] Anne AH de Hond et al. “Perspectives on validation of clinical predictive algorithms”. In: *NPJ Digital Medicine* 6.1 (2023), p. 86. DOI: 10.1038/s41746-023-00832-9.

[23] Alex Youssef et al. "All models are local: time to replace external validation with recurrent local validation". In: *arXiv preprint arXiv:2305.03219* (2023). DOI: 10.48550/arXiv.2305.03219.

[24] August DuMont Schütte et al. "Overcoming barriers to data sharing with medical image generation: a comprehensive evaluation". In: *NPJ digital medicine* 4.1 (2021), p. 141. DOI: 10.1038/s41746-021-00507-3.

[25] Yi Li et al. "A transfer learning-based multimodal neural network combining metadata and multiple medical images for glaucoma type diagnosis". In: *Scientific Reports* 13.1 (2023), p. 12076. DOI: 10.1038/s41598-022-27045-6.

[26] Shih-Cheng Huang et al. "Self-supervised learning for medical image classification: a systematic review and implementation guidelines". In: *NPJ Digital Medicine* 6.1 (2023), p. 74. DOI: 10.1038/s41746-023-00811-0.

[27] Martijn Starmans et al. "Reproducible radiomics through automated machine learning validated on twelve clinical applications". In: *arXiv preprint arXiv:2108.08618* (2021). DOI: 10.48550/arXiv.2108.08618.

[28] Fabian Isensee et al. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation". en. In: *Nat. Methods* 18.2 (Feb. 2021), pp. 203–211. DOI: 10.1038/s41592-020-01008-z.

[29] Dong Yang et al. "T-AutoML: Automated machine learning for lesion segmentation using transformers in 3d medical imaging". In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pp. 3962–3974. DOI: 10.48550/arXiv.2111.07535.

[30] Jianwei Zhang et al. "Auto machine learning for medical image analysis by unifying the search on data augmentation and neural architecture". In: *arXiv preprint arXiv:2207.10351* (2022). DOI: 10.48550/arXiv.2207.10351.

[31] Ben Van Calster et al. “There is no such thing as a validated prediction model”. In: *BMC medicine* 21.1 (2023), p. 70. DOI: 10.1186/s12916-023-02779-w.

[32] Alex Zwanenburg et al. “The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping”. en. In: *Radiology* 295.2 (May 2020), pp. 328–338. DOI: 10.1148/radiol.2020191145.

[33] David A Feinberg et al. “Next-generation MRI scanner designed for ultra-high-resolution human brain imaging at 7 Tesla”. In: *Nature Methods* (2023), pp. 1–10. DOI: 10.1038/s41592-023-02068-7.

[34] Alexey Dosovitskiy et al. “An image is worth 16x16 words: Transformers for image recognition at scale”. In: *arXiv preprint arXiv:2010.11929* (2020). DOI: 10.48550/arXiv.2010.11929.

[35] Tom Brown et al. “Language models are few-shot learners”. In: *Advances in neural information processing systems* 33 (2020), pp. 1877–1901. DOI: 10.48550/arXiv.2005.14165.

[36] Aditya Ramesh et al. “Hierarchical text-conditional image generation with clip latents”. In: *arXiv preprint arXiv:2204.06125* 1.2 (2022), p. 3. DOI: 10.48550/arXiv.2204.06125.

[37] Michael Moor et al. “Foundation models for generalist medical artificial intelligence”. In: *Nature* 616.7956 (2023), pp. 259–265. DOI: 10.1038/s41586-023-05881-4.

[38] Chaoyi Wu et al. “Towards Generalist Foundation Model for Radiology”. In: *arXiv preprint arXiv:2308.02463* (2023). DOI: 10.48550/arXiv.2308.02463.

[39] Yukun Zhou et al. “A foundation model for generalizable disease detection from retinal images”. In: *Nature* (2023), pp. 1–8. DOI: 10.1038/s41586-023-06555-x.

[40] Bobby Azad et al. "Foundational Models in Medical Imaging: A Comprehensive Survey and Future Vision". In: *arXiv preprint arXiv:2310.18689* (2023). DOI: 10.48550/arXiv.2310.18689.

[41] Tao Tu et al. "Towards generalist biomedical AI". In: *arXiv preprint arXiv:2307.14334* (2023). DOI: 10.48550/arXiv.2307.14334.

[42] Zohaib Salahuddin et al. "Transparency of deep neural networks for medical image analysis: A review of interpretability methods". In: *Computers in Biology and Medicine* 140 (2022), p. 105111. DOI: 10.1016/j.compbiomed.2021.105111.

[43] Bas H.M. van der Velden et al. "Explainable artificial intelligence (XAI) in deep learning-based medical image analysis". In: *Medical Image Analysis* 79 (2022), p. 102470. DOI: 10.1016/j.media.2022.102470.

[44] Melanie Champendal et al. "A scoping review of interpretability and explainability concerning artificial intelligence methods in medical imaging". In: *European Journal of Radiology* (2023), p. 111159. DOI: 10.1016/j.ejrad.2023.111159.

[45] Petros Kalendralis et al. "Making radiotherapy more efficient with FAIR data". In: *Physica Medica* 82 (2021), pp. 158–162. DOI: 10.1016/j.ejmp.2021.01.083.

[46] David Killock. "AI outperforms radiologists in mammographic screening". In: *Nature Reviews Clinical Oncology* 17.3 (2020), pp. 134–134. DOI: 10.1038/s41571-020-0329-7.

[47] Curtis P Langlotz. *Will artificial intelligence replace radiologists?* 2019.

[48] Curtis P Langlotz. *The Future of AI and Informatics in Radiology: 10 Predictions*. 2023. DOI: 10.1148/radiol.231114.

[49] Kicky G van Leeuwen et al. “Artificial intelligence in radiology: 100 commercially available products and their scientific evidence”. In: *European radiology* 31 (2021), pp. 3797–3804. DOI: 10.1007/s00330-021-07892-z.

[50] Zoe Mendelsohn et al. “Commercial volumetric MRI reporting tools in multiple sclerosis: a systematic review of the evidence”. In: *Neuroradiology* 65.1 (2023), pp. 5–24. DOI: 10.1007/s00234-022-03074-w.

[51] Michael Barnett et al. “A real-world clinical validation for AI-based MRI monitoring in multiple sclerosis”. In: *npj Digital Medicine* 6.1 (2023), p. 196. DOI: 10.1038/s41746-023-00940-6.

[52] Dennis M Hedderich et al. “Artificial intelligence tools in clinical neuroradiology: essential medico-legal aspects”. In: *Neuroradiology* (2023), pp. 1–9. DOI: 10.1007/s00234-023-03152-7.

[53] European Parliamentary Research Service. *Artificial intelligence in healthcare. Applications, risks, and ethical and societal impacts*. 2023.

[54] Kicky G van Leeuwen et al. “How does artificial intelligence in radiology improve efficiency and health outcomes?” In: *Pediatric Radiology* (2021), pp. 1–7. DOI: 10.1007/s00247-021-05114-8.

[55] Federico Spagnolo et al. “How far MS lesion detection and segmentation are integrated into the clinical workflow? A systematic review”. In: *NeuroImage: Clinical* 39 (2023), p. 103491. DOI: 10.1016/j.nicl.2023.103491.

Summary

The motivation behind this thesis is to explore the potential of ”radiomics” in the field of neurology, where early diagnosis and accurate treatment selection are crucial for improving patient outcomes. Neurological diseases are a major cause of disability and death globally, and there is a pressing need for reliable imaging biomarkers to aid in disease detection and monitoring. While radiomics has shown promising results in oncology, its application in neurology remains relatively unexplored. Therefore, this work aims to investigate the feasibility and challenges of implementing radiomics in the neurological context, addressing various limitations and proposing potential solutions.

The thesis begins with a demonstration of the predictive power of radiomics for identifying important diagnostic biomarkers in neuro-oncology. Building on this foundation, the research then delves into radiomics in non-oncological neurology, providing an overview of the pipeline steps, potential clinical applications, and existing challenges. Despite promising results in proof-of-concept studies, the field faces limitations, mostly data-related, such as small sample sizes, retrospective nature, and lack of external validation.

To explore the predictive power of radiomics in non-oncological tasks, a radiomics approach was implemented to distinguish between multiple sclerosis patients and normal controls. Notably, radiomic features extracted from normal-appearing white matter were found to contain distinctive information for multiple sclerosis detection, confirming the hypothesis of the thesis.

To overcome the data harmonization challenge, in this work quantitative mapping of the brain was used. Unlike traditional imaging methods, quantitative mapping involves measuring the physical properties of brain tissues, providing a more standardized and consistent data representation. By reconstructing the physical properties of each voxel based on multi-echo MRI acquisition,

quantitative mapping produces data that is less susceptible to domain-specific biases and scanner variability. Additionally, the insights gained from quantitative mapping are building the bridge toward the physical and biological properties of brain tissues, providing a deeper understanding of the underlying pathology.

Another crucial challenge in radiomics is robust and fast data labeling, particularly segmentation. A deep learning method was proposed to perform automated carotid artery segmentation in stroke at-risk patients, surpassing current state-of-the-art approaches. This novel method showcases the potential of automated segmentation to enhance radiomics pipeline implementation.

In addition to addressing specific challenges, the thesis also proposes a community-driven open-source toolbox for radiomics, aimed at enhancing pipeline standardization and transparency. This software package would facilitate data curation and exploratory analysis, fostering collaboration and reproducibility in radiomics research.

Through an in-depth exploration of radiomics in neuroimaging, this thesis demonstrates its potential to enhance neurological disease diagnosis and monitoring. By uncovering valuable information from seemingly normal brain tissues, radiomics holds promise for early disease detection. Furthermore, the development of innovative tools and methods, including deep learning and quantitative mapping, has the potential to address data labeling and harmonization challenges.

Looking to the future, embracing larger, diverse datasets and longitudinal studies will further enhance the generalizability and predictive power of radiomics in neurology. By addressing the challenges identified in this thesis and fostering collaboration within the research community, radiomics can advance toward clinical implementation, revolutionizing precision medicine in neurology.

Résumé

La motivation derrière cette thèse est d'explorer le potentiel de la "radiomique" dans le domaine de la neurologie, où le diagnostic précoce et la sélection précise du traitement sont cruciaux pour améliorer l'issue des patients. Les maladies neurologiques sont une cause majeure d'invalidité et de décès à l'échelle mondiale; il existe un besoin pressant de biomarqueurs d'imagerie fiables pour contribuer à la détection et à la surveillance des maladies. Bien que la radiomique ait montré des résultats prometteurs en oncologie, son application en neurologie reste relativement inexplorée. Par conséquent, ce travail vise à étudier la faisabilité et les défis de la mise en œuvre de la radiomique dans le contexte neurologique, en abordant diverses limitations et en proposant des solutions potentielles.

La thèse commence par la démonstration du pouvoir prédictif de la radiomique pour identifier d'importants biomarqueurs diagnostiques en neuro-oncologie. Sur cette base, la recherche se penche ensuite sur l'utilisation de la radiomique en neurologie non-oncologique, en fournissant un aperçu des étapes de la chaîne de traitements, des applications cliniques potentielles et des défis existants. Malgré des résultats prometteurs dans des études de "preuve de concept", le domaine présente des limitations, principalement liées aux données, telles que la petite taille d'échantillons, une nature rétrospective et un manque de validation externe.

Pour explorer le pouvoir prédictif de la radiomique dans des tâches non-oncologiques, une approche radiomique a été développée pour distinguer entre des patients atteints de sclérose en plaques et des sujets sains. Notamment, il a été constaté que les caractéristiques radiomiques extraites de la substance blanche d'apparence normale contenaient des informations distinctives pour la détection de la sclérose en plaques, confirmant l'hypothèse de la thèse.

Pour surmonter le défi de l'harmonisation des données, ce travail a utilisé la cartographie par IRM quantitative du cerveau.

Contrairement aux méthodes traditionnelles d'IRM, l'IRMquantitative implique l'estimation des propriétés physiques des tissus cérébraux, offrant une représentation des données plus normalisée et cohérente. En reconstruisant les propriétés physiques de chaque voxel en fonction de l'acquisition d'IRM multi-écho, l'IRM quantitative produit des données moins sensibles aux biais spécifiques du domaine et aux variations liées aux machines. De plus, les informations obtenues grâce à l'IRM quantitative contribuent à une meilleure compréhension de la pathologie sous-jacente.

Un autre défi crucial en radiomique est l'étiquetage robuste et rapide des données, en particulier la segmentation. Une méthode d'apprentissage profond a été proposée pour effectuer la segmentation automatisée de l'artère carotide chez des patients à risque d'accident vasculaire cérébral, dépassant les approches actuelles de pointe. Cette nouvelle méthode novatrice montre le potentiel de la segmentation automatisée pour améliorer la mise en œuvre de la chaîne de traitement radiomique.

En plus d'aborder des défis spécifiques, la thèse propose également une boîte à outils "open source" communautaire pour la radiomique, dans le but d'améliorer la normalisation et la transparence des chaînes de traitement. Ce logiciel facilitera la curation des données et les analyses exploratoires, favorisant la collaboration et la reproductibilité dans la recherche en radiomique.

Au travers d'une exploration approfondie de la radiomique en neuro-imagerie, cette thèse démontre son potentiel pour améliorer le diagnostic et la surveillance des maladies neurologiques. En extrayantdes informations précieuses à partir de tissus cérébraux apparemment normaux, la radiomique offre des perspectives de détection précoce de maladies. De plus, le développement d'outils et de méthodes innovantes, notamment l'apprentissage profond et l'IRM quantitative, ont le potentiel de résoudre les problèmes d'étiquetage et d'harmonisation des données.

En se tournant vers l'avenir, l'adoption de jeux de données plus vastes et diversifiés ainsi que d'études longitudinales améliorera encore la généralisabilité et la puissance prédictive de la radiomique

en neurologie. En abordant les défis identifiés dans cette thèse et en favorisant la collaboration au sein de la communauté de recherche, la radiomique peut progresser vers une mise en œuvre clinique, révolutionnant la médecine de précision en neurologie.

Samenvatting

De motivatie achter dit proefschrift is het verkennen van het potentieel van ”radiomics” op het gebied van neurologie, waarbij een vroege diagnose en nauwkeurige behandelingskeuze cruciaal zijn voor het verbeteren van de uitkomsten voor patiënten. Neurologische ziekten zijn wereldwijd een belangrijke oorzaak van invaliditeit en overlijden, en er is daarom een dringende behoefte aan betrouwbare beeldvormende biomarkers om bij te dragen aan de detectie en monitoring van deze ziekten. Hoewel radiomics veelbelovende resultaten heeft getoond in de oncologie, blijft de toepassing ervan in de neurologie relatief onontgonnen terrein. Daarom heeft dit proefschrift tot doel de haalbaarheid en uitdagingen van de implementatie van radiomics in de neurologische context te onderzoeken, waarbij verschillende beperkingen worden onderzocht en mogelijke oplossingen worden voorgesteld.

Dit proefschrift begint met een demonstratie van de voorspellende kracht van radiomics voor het identificeren van belangrijke diagnostische biomarkers in de neuro-oncologie. Op basis hiervan richt het onderzoek zich vervolgens op radiomics in niet-oncologische neurologie, waarbij een overzicht wordt gegeven van de stappen in de pijplijn, van potentiële klinische toepassingen en van bestaande uitdagingen. Ondanks veelbelovende resultaten in proof-of-concept studies kent het veld beperkingen, voornamelijk gerelateerd aan onderzoeksgegevens, zoals kleine steekproefgroottes, een retrospectief karakter en het gebrek aan externe validatie.

Om de voorspellende kracht van radiomics in niet-oncologische taken te onderzoeken, werd er een radiomics-benadering toegepast om multiple sclerosepatiënten te onderscheiden van een controlegroep. De voornaamste vondst is dat radiomics kenmerken die zijn geëxtraheerd uit normaal ogend witte stof in de hersenen onderscheidende informatie bleken te bevatten voor de detectie van multiple sclerose, wat de hypothese van dit proefschrift bevestigt.

Om de uitdaging van gegevensharmonisatie te overwinnen, werd er in dit werk kwantitatieve mapping van de hersenen gebruikt. In tegenstelling tot traditionele beeldvormingsmethoden meet kwantitatieve mapping fysische eigenschappen van het hersenweefsel, wat leidt tot een meer gestandaardiseerde en consistente gegevensrepresentatie. Door de fysische eigenschappen van elk voxel te reconstrueren op basis van multi-echo MRI-acquisitie, produceert kwantitatieve mapping gegevens die minder gevoelig zijn voor domeinspecifieke vertekeningen en scannervariaties. Daarnaast dragen inzichten die zijn verkregen uit kwantitatieve mapping bij aan een dieper begrip van de fysische en biologische eigenschappen van hersenweefsel, wat vervolgens resulteert in een dieper begrip van de onderliggende pathologie.
Een andere cruciale uitdaging in radiomics is robuuste en snelle gegevenslabeling, met name voor segmentatie. Er werd een deep learning methode voorgesteld om geautomatiseerde segmentatie van de halsslagader uit te voeren bij patiënten met een verhoogd risico op een beroerte, waarbij de huidige state-of-the-art benaderingen werden overtroffen. Deze nieuwe methode demonstreert het potentieel van geautomatiseerde segmentatie om de implementatie van radiomics-pijplijnen te verbeteren.
Naast het onderzoeken van specifieke uitdagingen, stelt dit proefschrift ook een door de gemeenschap gedreven open-source gereedschapskist voor radiomics voor, gericht op het verbeteren van de standaardisatie en transparantie van pijplijnen. Dit softwarepakket zou gegevenscuratie en exploratieve analyse vergemakkelijken, en de samenwerking en reproduceerbaarheid in radiomics-onderzoeken bevorderen.
Middels een diepgaande verkenning van radiomics in neurobeeldvorming toont dit proefschrift het potentieel om de diagnose en monitoring van neurologische ziekten te verbeteren. Door waardevolle informatie te onthullen uit ogenschijnlijk normaal hersenweefsel, biedt radiomics mogelijkheden om de ziekte in een vroeg stadium te detecteren. Bovendien heeft de ontwikkeling van innovatieve tools en methoden, waaronder deep learning en

kwantitatieve mapping, het potentieel om uitdagingen op het gebied van gegevenslabeling en harmonisatie aan te pakken.
Vooruitblikkend naar de toekomst zal het omarmen van grotere, diverse data en longitudinale studies de generaliseerbaarheid en voorspellende kracht van radiomics in de neurologie verder verbeteren. Door de uitdagingen die in dit proefschrift zijn geïdentificeerd op te lossen en samenwerking binnen de onderzoeksgemeenschap te bevorderen, kan radiomics vooruitgang boeken naar klinische implementatie, wat tot een revolutie kan leiden op het gebied van precisiegeneeskunde in de neurologie.

Valorisation

Neurological diseases pose a significant global health challenge, being a leading cause of disability and the second leading cause of death. These diseases encompass diverse pathologies but may present similar clinical symptoms, making accurate and early diagnosis crucial for effective treatment and patient outcomes. The application of radiomics, a methodology combining imaging data and machine learning, has the potential to transform neurology by facilitating quantitative imaging and providing valuable biomarkers for neurological conditions. Unlike its well-established use in oncology, radiomics in neurology faces unique challenges and requires careful evaluation of its clinical applications.

The focus of this thesis was to investigate the suitability of radiomics in neurology and identify potential clinical applications, while also addressing the existing challenges that hinder its implementation in clinical practice. By performing experiments in both neuro-oncological and neurological settings, we explored the predictive power of radiomics and its ability to extract informative features from brain tissues that may not be visually discernible, thus demonstrating the potential for early disease detection and monitoring.

The thesis also proposes solutions to tackle challenges related to data harmonization, segmentation, and pipeline standardization. By introducing a community-driven open-source software toolbox for data curation and exploration, we aim to enhance reproducibility and transparency in radiomics studies, fostering collaboration and standardization across research groups.

The technical findings and methodological advancements presented in this thesis lay a foundation for further research and development in the field of neurological radiomics. The utilization of radiomics and data-driven approaches in neurology holds the promise of personalized precision medicine, potentially changing patient care by enabling more accurate and tailored diagnoses, treatment plans, and patient manage-

ment strategies.
As we progress toward the era of precision medicine, the results and insights gained from this work have significant implications for the future of neurology.

1 Scientific impacts

1. The experiments and analysis in this thesis are published in well-cited journals, and all the publications are open-access, which increases knowledge exchange in the scientific community.

2. All the papers were accomplished by the corresponding code available at the open platforms (from Chapters 5 and 6 – submitted to `codeocean.com`) which confirmed the transparency of the results and can be used for other similar projects.

3. Chapter 2 shows a conventional radiomics pipeline in neuro-oncology and demonstrates the impact of image pre-processing (interpolation during resampling) on the model performance.

4. Chapter 3 provides an extensive up-to-date review of radiomics implementation in non-oncological clinical neuroimaging and identifies the current challenges in the field.

5. Chapter 4 shows proof-of-concept development of radiomic signature for multiple sclerosis; it demonstrates the results obtained on conventional T1w MRI and overperforming quantitative mapping, indicating that the second one has much more potential in quantitative medical image analysis.

6. Chapter 5 suggests an uncertainty-based deep learning approach for medical image segmentation on partially labeled data driven by clinically relevant derivable from the segmentation instead of traditionally applied segmentation performance scores.

7. Chapter 6 suggests an open-source Python toolbox for medical imaging data curation and exploration filling the gap between the currently existing open-source tools.

2 Social impacts

1. Identification of the current challenges in the field will determine the direction of future work for precision medicine empowerment in clinical neurology through radiomics.

2. Image quantification will non-invasively reveal diagnostic information previously not available for the subjective and qualitative perception of the human eye.

3. The application of quantitative imaging will lead to the development of objective and robust imaging biomarkers; additionally, it will create the data naturally harmonized between sites and scanners.

4. Deep learning methods will support healthcare specialists in solving time-consuming routine tasks such as medical image segmentation, revealing time for more challenging and expertise-demanding activities.

5. The precision-medicine-toolbox is not only available for every individual in society, but it is also democratizing artificial intelligence in medical imaging research by providing data curation and exploration functionality for researchers without strong programming backgrounds.

6. Precision medicine will facilitate early diagnosis and optimal treatment improving quality of life for the population and reducing healthcare costs.

3 Economical impacts

The average age of the population is on the rise, leading to an increase in the number of individuals affected by age-related conditions, primarily neurological disorders such as dementia, Alzheimer's, and Parkinson's diseases [1]. Additionally, there is a growing population affected by diseases that typically impact younger individuals, like multiple sclerosis [2]. As a consequence, the demand for healthcare services is steadily rising. Unfortunately, the healthcare system resources are already overstretched [3]. To effectively manage these patients, it is imperative to automate routine tasks and provide support to existing healthcare experts.
Furthermore, in many countries, healthcare costs for each patient are covered by healthcare insurance and treatment choices are influenced by the available budget for a particular treatment [4, 5]. Hence, there is a critical need for early diagnosis and optimized decision-making to allocate insurance funds efficiently.
In light of these challenges, our study identifies current obstacles in the field and proposes potential solutions for some of them. We aim to establish a foundation for future research and clinical applications that will aid neuroradiologists in making more informed decisions. The outcomes of this thesis contribute to solutions that bring precision medicine in neurology one step closer to reality.

References

[1] World Health Organization. *Ageing and Health*. 2023. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health (visited on 09/25/2023).

[2] Clare Walton et al. "Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS". In: *Multiple Sclerosis Journal* 26.14 (2020), pp. 1816–1821. DOI: 10.1177/1352458520970841.

[3] Valery L Feigin et al. “The global burden of neurological disorders: translating evidence into policy”. In: *The Lancet Neurology* 19.3 (2020), pp. 255–265. DOI: 10.1016/S1474-4422(19)30411-9.

[4] National Institute of Ageing. *How is Alzheimer's Disease Treated?* https://www.nia.nih.gov/health/how-alzheimers-disease-treated. Accessed: 25.09.2023. 2023/ 2023.

[5] Guoqiao Wang et al. “Health insurance affects the use of disease-modifying therapy in multiple sclerosis”. In: *Neurology* 87.4 (2016), pp. 365–374. DOI: 10.1212/WNL.0000000000002887.

List of publications

1. R. Casale, **E. Lavrova**, S. Sanduleanu, H.C. Woodruff, P. Lambin. "Development and external validation of a non-invasive molecular status predictor of chromosome 1p/19q co-deletion based on MRI radiomics analysis of Low Grade Glioma patients". In: European Journal of Radiology 139 (June 2021), p. 109678. DOI: 10.1016/j.ejrad.2021.109678,

2. **E. Lavrova**, E. Lommers, H.C. Woodruff, A. Chatterjee, P. Maquet, E. Salmon, P. Lambin, C. Phillips. "Exploratory radiomic analysis of conventional vs. quantitative brain MRI: toward automatic diagnosis of early multiple sclerosis". In: Frontiers in neuroscience 15 (2021), p. 679941. DOI: 10.3389/fnins.2021.679941,

3. S.P. Primakov, A. Ibrahim, J.E. van Timmeren, G. Wu, S.A. Keek, M. Beuque, R.W.Y. Granzier, **E. Lavrova**, M. Scrivener, S. Sanduleanu, E. Kayan, I. Halilaj, A. Lenaers, J. Wu, R. Monshouwer, X. Geets, H.A. Gietema, L.E.L. Hendriks, O. Morin, A. Jochems, H.C. Woodruff, P. Lambin. "Automated detection and segmentation of non-small cell lung cancer computed tomography images". In: Nature communications 13.1 (2022), pp. 1–12. DOI: 10.1038/s41467-022-30841-3,

4. I. Halilaj, C. Oberije, A. Chatterjee, Y. van Wijk, N. Mohammadian Rad, P. Galganebanduge, **E. Lavrova**, S. Primakov, Y. Widaatalla, A. Wind, P. Lambin. "Open Source Repository and Online Calculator of Prediction Models for Diagnosis and Prognosis in Oncology". In: Biomedicines 10.11 (2022), p. 2679. DOI: 10.3390/biomedicines10112679,

5. **E. Lavrova**, Z. Salahuddin, H.C. Woodruff, M. Kassem, R. Camarasa, A.G.D. Van Kolk, P.J. Nederkoorn, D. Bos, J. Hendrikse,

M.E. Kooi, P. Lambin. “UR-CarA-Net: A Cascaded Framework With Uncertainty Regularization for Automated Segmentation of Carotid Arteries on Black Blood MR Images”. In: IEEE Access 11 (2023), pp. 26637–26651. DOI: 10.1109/ACCESS.2023.3258408,

6. **E. Lavrova**, S. Primakov, Z. Salahuddin, M. Beuque, D. Verstappen, H.C. Woodruff, P. Lambin. “Precision-medicine-toolbox: An open-source python package for the quantitative medical image analysis”. In: Software Impacts 16 (2023), p. 100508. DOI: 10.1016/j.simpa.2023.100508,

7. W. Rogers, S.A. Keek, M. Beuque, **E. Lavrova**, S. Primakov, G. Wu, C. Yan, S. Sanduleanu, H.A. Gietema, R. Casale, M. Occhipinti, H.C. Woodruff, A. Jochems, P. Lambin. “Towards texture accurate slice interpolation of medical images using PixelMiner”. In: Computers in Biology and Medicine 161 (2023), p. 106701. DOI: 10.1016/j.compbiomed.2023.106701,

8. **E. Lavrova**, H.C. Woodruff, H. Khan, E. Salmon, P. Lambin, C. Phillips. “Handcrafted and deep radiomics in neurodegenerative diseases: a transition from oncology to clinical neurology”. In: Submitted to Clinical NeuroImage (2024).

List of software

1. **Chapter 2.** Imaging biomarkers exploration in neuro-oncology with radiomics: https://github.com/roberto-casale/LGG-1p-19q-deletion,
2. **Chapter 3.** Introduction into radiomics workflow: https://github.com/lavrovaliz/radiomics-training,
3. **Chapter 3.** Introduction into neuroimaging workflow: https://github.com/lavrovaliz/neuroimaging-training,
4. **Chapter 4.** Reliable imaging biomarkers exploration in MS with radiomics and qMRI: **Lisa Lavrova**. (2020). lavrovaliz/brain-tissue-radiomics-on-clinical-and-quantitative-MRI-for-MS 0 (Version 0). Zenodo. https://doi.org/10.5281/zenodo.4275545,
5. **Chapter 5.** Carotid artery segmentation on black blood MR images: **Lisa Lavrova**. (2023). lavrovaliz/ur-cara-net: 0.1 (0.1). Zenodo. https://doi.org/10.5281/zenodo.7741155,
6. **Chapter 6.** Quantitative medical imaging pipeline standardization through the open-source software: **Lisa Lavrova**, Sergey Primakov, mbeuque, and Damon Verstappen. (2022). primakov/precision-medicine-toolbox: pmtool v0.11 (0.11). Zenodo. https://doi.org/10.5281/zenodo.6982100.

List of presentations

1. **E. Lavrova**, H.C. Woodruff, C. Phillips, E. Lommers, P. Maquet, E. Salmon, P. Lambin. Quantitative neuroimaging with radiomics and deep learning. Presented at the CRC In vivo imaging Scientific Day, Liège, Belgium, October 2019.

2. **E. Lavrova**, H.C. Woodruff, C. Phillips, E. Lommers, P. Maquet, E. Salmon, P. Lambin. Multiple sclerosis detection with qMRI and Radiomics. Presented at the MUMC+ Research Day of the Department of Radiology and Nuclear Medicine, Maastricht, the Netherlands, February 2020.

3. **E. Lavrova**, H.C. Woodruff, C. Phillips, E. Lommers, P. Maquet, E. Salmon, P. Lambin. Development of an MR-based automatic diagnostic aid for multiple sclerosis using Radiomics. Presented at the 9th European Conference on Clinical Neuroimaging, Paris, France, March 2020.

4. **E. Lavrova**, H.C. Woodruff, C. Phillips, E. Lommers, P. Maquet, E. Salmon, P. Lambin. Development and external validation of automatic diagnostic aid for multiple sclerosis using a radiomics analysis of white matter on clinical and quantitative MRI. Presented at the European Congress of Radiology, virtual participation, July 2020.

5. **E. Lavrova**, H.C. Woodruff, C. Phillips, C. Bastin, P. Lambin, E. Salmon. Development of MR-based quantitative biomarkers for Alzheimer's disease using hippocampal Radiomics. Presented at the 10th European Conference on Clinical Neuroimaging, Geneva, Switzerland, March 2022.

6. **E. Lavrova**, S. Primakov. PM toolbox: open-source software development from scripts to project. Presented at the PM research seminar, Maastricht, the Netherlands, August 2022.

7. **E. Lavrova**, Z. Salahuddin, H.C. Woodruff, M. Kassem, P. Nederkoorn, D. Bos, J. Hendriske, P. Lambin, M.E. Kooi. Plaq-U-Net: a multi-patch consensus U-Net for automated detection and segmentation of carotid arteries on black blood MRI. Presented at the MUMC+ Radiology Research Day, Maastricht, the Netherlands, September 2022.
8. **E. Lavrova**, H.C. Woodruff, E. Lommers, P. Maquet, P. Lambin, C. Phillips. Cycle GANs for FLAIR brain scans synthesis from conventional T1w MRI and qMRI. Presented at the virtual conference of Quantitative Magnetic Resonance Imaging in Neurodegeneration, virtual participation, October 2022.

Ph.D. Portfolio

Department: CRC – In vivo imaging (Liège University), Precision Medicine (Maastricht University)
Ph.D. period: 01/02/2019 - 31/01/2023

Activities	Credits
Professional courses	
ULiege – Medical Imaging (2019)	3
ULiege – Siemens visit (2019)	1
ULiege – Neurosciences (2019)	3
Precision Medicine – AI4I (2019)	3
UMaastricht – Academic writing for PhD and MSc (2020)	0.5
ULiege – Introduction to intellectual property (2020)	0.5
ULiege – Software development good practices (2020)	0.5
ULiege – Responsible research data management (2021)	0.5
Conferences	
GIGA CRC Science day – oral presentation (2019)	3
GIGA CRC Seminar – oral presentation (2020)	3
WiDS Datathon – participation (2020)	1
MUMC+ Radiology research day – oral presentation (2020)	3
ECCN2020 – oral presentation (2020)	4
ECR2020 – oral presentation (2020)	5
OHBM2021 – attendance (2021)	3
ECCN2022 – oral presentation (2022)	5
AI4I2022 – organization lead of the practicals (2022)	3
MUMC+ Radiology research day – oral presentation (2022)	3
PM seminars – oral presentations (2022)	3
I qMRI conference – oral presentation (2022)	4
Teaching	
UMaastricht – thesis supervision of 2 BMS MSc students (2020, 2022)	2
UMaastricht – capstone supervision of AI+Neurosciences BSc student (2021)	1
UMaastricht – mentoring in career skills portfolio (2021-2022)	1
UMaastricht – practicals in AI in medicine for Paris-Maastricht BSc group (2022)	1
UMaastricht – practicals in Biomedical Approaches for 4 MSc groups (2022)	1

Acknowledgments

Five years ago, I was offered a Ph.D. position between ULiège and UMaastricht almost accidentally. But eventually, it turned out to be the most deliberate career choice I've ever made. All the diversified and sometimes incompatible skills I've gained during the previous years have finally set up the puzzle for this job. Therefore I would like to express my sincere gratitude to all the people accompanying me during this exciting adventure, through its ups and downs.

I would like to thank my promotor Professor Philippe Lambin for this opportunity. I grew from an unsure Ph.D. student to a fully independent researcher who can stand her ground.

I would not be able to accomplish my Ph.D. journey without my promotor Professor Christophe Phillips. He spent hours explaining the details of brain MR physics, data reconstruction, and analysis. Christophe gave me an immense support during the last year, when I left the labs and needed it the most. He was also a great neighbor in the beautiful countryside of Belgium. And most importantly, thanks to Christophe, I am proud to call myself an engineer.

I would like to thank my co-promotor Doctor Henry C. Woodruff for the freedom he gave me on my projects. He always encouraged my initiative and involvement, and it helped me to grow professionally.

I would like to express my gratitude to my co-promotor Professor Eric Salmon who was a clinical motivator for my research. His critical but always very delicate inputs not only improved my work but helped me to develop a lot (even though there is still a large room for improvement) in the field of neurology which became my passion.

It is important to note that from some of my promotors and co-promotors, I have inherited some bad habits, such as snowboarding and rock climbing. It was always cool to have a small chat about it after a long scientific discussion.

I am grateful to Doctor Arthur Jochems for interviewing me and offering me this Ph.D. position and to Doctor Isabelle Rausin for

taking care of the administrative process of hiring me. Without them, I would not be able to even start this journey.
Next, I would like to highlight some notable collaborations. Roberto Casale chose me as a partner-in-crime for his first research paper, and I think we nailed it by combining our medical and technical knowledge. Doctor Emilie Lommers and Professor Pierre Maquet have provided me with truly unique and valuable data, they gave me precious insights into multiple sclerosis pathology and supported me in my first first-author publication. Doctor Christine Bastin helped me to extend my knowledge from multiple sclerosis to Alzheimer's disease. I had hours of discussions with Professor Eline Kooi and Doctor Mohammed Kasseem. We talked in different languages about the segmentation problem but always came to a consensus. They expanded my expertise in MR acquisition protocols and challenging MR segmentation tasks and showed me that deep learning should be aimed at solving clinically relevant problems rather than chasing the common data science performance scores. Professor Wim van Zwam impressed me with his precise understanding of the unmet clinical needs in stroke and scientific ambitions. I am grateful to Susan Olthuis for always willing to help and always positive mindset. She made the data curation process less painful, and from her, I have learned a lot about stroke patients' management. To Sergey Primakov for a collaboration bringing us from a pile of Jupyter notebooks to a mature open-source Python package. I would like to thank Professor Liesbet Peeters for introducing me to the community of clinical neurology. And I am grateful to all the other co-authors who provided valuable feedback on my work over the years.
The next group of people I would like to offer my acknowledgements to is the assessment committee of this thesis: Professor Gilles Louppe, Professor Fabienne Collette, Doctor Johan van Soest, Professor Walter H. Backes, Professor Adrien Depeursinge, and Doctor Harini Veeraraghavan. I am honored to have my work to be evaluated in such an esteemed and multi-disciplinary round with an expertise in deep learning and AI, neurological diseases, medical physics, and medical software development.

All the logistics, paperwork, publishing papers, attending conferences, and submitting the thesis would not be possible without Annick Claes, Brigitte Herbillon, Floor Franssen, and Rianne Herben. I am grateful to them for always being here for me.
You learn a lot by teaching others. Therefore I am grateful to Dre, Stan, and Mart, the UM students, as well as my mentorship groups. It was an honor to supervise you.
Aside from the professional collaborations, I am grateful to all the people who created a welcoming work environment. And the double affiliation makes this list twice longer, so the risk of missing someone is dangerously high! For the vlaais, coffee breaks, and drinks at Thembis I am thankful to Abdalla, Anke, Anouk, Anshu, Cary, Damon, David, Elze, Esma, Guangyao, Hamza, Janita, Jennifer, Jionghui, Michael, Monideepa, Natacha, Prabash, Relinde, Sebastian, Shang, Shen, Shruti, Simone, Sithin, Xian, Yi, Yumeng, and Yvonka. To Fadila and Akshaya for turning my commute to Liège into a good experience, Yousif for the deep talks, and Nuwan for being into shameful dancing, Simon for the Zondag beers on Thursdays and literature discussions. For drinks at le Carre, lunches at Sart-Tilman, and French lessons I am grateful to Adrien, Camille, Christina, Daphne, Emma, Gabriel, Gilles, Justinas, Martin, Mathilde, Maxime, Nikita, Nora, Paolo, and Zoltan. To Katya and Vincenzo for sharing the evenings over a barbeque in summer and in front of a fireplace in winter.
I was lucky to have the best officemates at the both Maastricht and Liège offices. Thank you, Turkey for the motivation when I was running out of ideas. Thank you, Gregory, for the discussions on statistics (our whiteboard at CRC should be protected as a scientific and cultural heritage), physics, and Python methods, and all the support you gave me when I was settling in Liège. Thank you Siya not only for being the last person alive at every party, but also for the valuable pieces of career advice. I hope to visit you and Sanya in the UK one day.
I am grateful to my colleagues who eventually became my true friends. Thank you Zohaib not only for showing me how to be a

better AI researcher, but also for being a great companion for the lunch breaks. It was amazing to share different opinions, coming from different points of view, on the same topics. Thank you Manon for being an irreplaceable raclette, wine, and movie partner as well as an example of a person standing for her values. Will, all the fun things we did in Maastricht, we did with you, and I will always carry these memories in my heart. You live as long as you are remembered. Avi, it was fun having dinners with you and going on city trips. But the important thing is that you helped me to grow up. Iva, sweetie, my life in Maastricht would not be the same without you. I cannot imagine anyone else being with me in both fancy Vienna and wild Albanian mountains. You always were an example of strength and honesty. I am so lucky to have you and cannot express my gratitude, but the words are not needed.

I am grateful to all my friends outside of UM and ULg, supporting me remotely and in person, from Russia and all over the world. But I would like to offer my special gratitude to the people who made the Maastricht-Liège-Aachen triangle my home. Thank you Katha for being the best stable mate, offering support in parenting a large animal, and giving me a feeling of family. Thank you Kate for having a record-breaking amount of empathy. Thank you Ivan and Natasha for helping me to adapt to a new environment. And thank you Torsten for sharing the same passion. I hope many more cool mountaineering tours are awaiting.

During my last year in Maastricht, while wrapping up the thesis, I jumped into some new activities, and the MaasSAC membership helped me a lot to refresh my brain. Many thanks to Maastricht Hiking Group for a great company in short escapes to the Alps. My gratitude goes to the Ph.D. Academy fellows as well for the warm atmosphere and all the activities I would not have done on my own (acro-yoga and flunky ball, for sure).

And of course, I would like to thank my parents for all their support and trust in me. I can never thank you enough for your love and inspiration.

Elizaveta Lavrova
Maastricht
2024-05-31

About the author

Elizaveta (Lisa) Lavrova was born on September 6th 1992 in Moscow, Russia. She studied biomedical equipment engineering at Bauman Moscow State Technical University (BMSTU) with the specialization in digital signal processing, pattern recognition, and artificial intelligence. She obtained BSc and MSc with honors in 2014 and 2016, respectively. Her thesis was on biometric automated facial recognition with the Gabor features and machine learning.

During the university years, Lisa worked at GE Healthcare in the computed tomography department as a technical intern for sales. After this, she joined Research Center of Biometric Equipment as a regulatory and standardization specialist.

After graduation, Lisa used to work as a teaching assistant giving practicals in Medical Image Processing and Pattern Recognition at BMSTU combining it with the work for industry. In 2016, Lisa joined Microsoft as a technical evangelist specializing in tools for artificial intelligence. In 2018, she joined SAS as a junior data analyst.

To continue the career in the research field, in 2019 she started her PhD jointly between Liège and Maastricht Universities. Under the Liège-Maastricht Imaging Valley grant, she explored radiomics technology transfer from oncology to clinical neurology.

Over the course of four years, Lisa worked on diverse projects related to the different neurological diagnosis and data modalities. As a part of her initiatives, she was one of the main contributors to the open-source Python project "precision-medicine-toolbox," a tool for data curation and exploration.
During her Ph.D., Lisa was mentoring students and giving practicals in medical imaging and radiomics. She actively contributed to content preparation for AI4Imaging workshops and hackathons taking the lead in the hands-on activities in 2022.
In February 2023, Lisa joined Carl Zeiss Meditec, Munich, as a machine learning scientist, where she continued to apply her academic and industrial knowledge in engineering and medicine. After that, she moved into the international field of medtech startups, empowering precision medicine for neurological and ophtalmic patients.

To connect:

- Linkedin
- GitHub
- Twitter
- ResearchGate