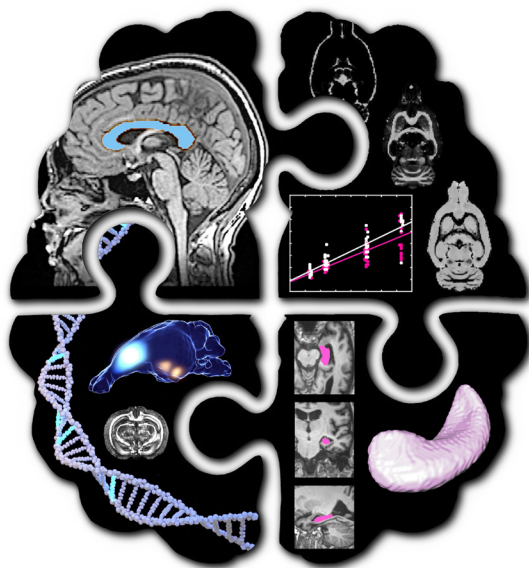


Doctoral Thesis in Medical Technology

# Methods for the analysis and characterization of brain morphology from MRI images

IRENE BRUSINI



Stockholm, Sweden 2022



**Karolinska  
Institutet**

# Methods for the analysis and characterization of brain morphology from MRI images

IRENE BRUSINI

Academic Dissertation which, with due permission of the KTH Royal Institute of Technology, is submitted for public defence for the Degree of Doctor of Philosophy on Friday the 25th of March 2022, at 1:00 p.m. in T2, Hälsovägen 11C, Huddinge, Sweden.

Doctoral Thesis in Medical Technology  
KTH Royal Institute of Technology  
Karolinska Institutet  
Stockholm, Sweden 2022

© Irene Brusini

ISBN 978-91-8040-138-8  
TRITA-CBH-FOU 2022:9

Printed by: Universitetsservice US-AB, Sweden 2022

# Abstract

Brain magnetic resonance imaging (MRI) is an imaging modality that produces detailed images of the brain without using any ionizing radiation. From a structural MRI scan, it is possible to extract morphological properties of different brain regions, such as their volume and shape. These measures can both allow a better understanding of how the brain changes due to multiple factors (e.g., environmental and pathological) and contribute to the identification of new imaging biomarkers of neurological and psychiatric diseases. The overall goal of the present thesis is to advance the knowledge on how brain MRI image processing can be effectively used to analyze and characterize brain structure.

The first two works presented in this thesis are animal studies that primarily aim to use MRI data for analyzing differences between groups of interest. In Paper I, MRI scans from wild and domestic rabbits were processed to identify structural brain differences between these two groups. Domestication was found to significantly reshape brain structure in terms of both regional gray matter volume and white matter integrity. In Paper II, rat brain MRI scans were used to train a brain age prediction model. This model was then tested on both controls and a group of rats that underwent long-term environmental enrichment and dietary restriction. This healthy lifestyle intervention was shown to significantly affect the predicted brain age trajectories by slowing the rats' aging process compared to controls. Furthermore, brain age predicted on young adult rats was found to have a significant effect on survival.

Papers III to V are human studies that propose deep learning-based methods for segmenting brain structures that can be severely affected by neurodegeneration. In particular, Papers III and IV focus on U-Net-based 2D segmentation of the corpus callosum (CC) in multiple sclerosis (MS) patients. In both studies, good segmentation accuracy was obtained and a significant correlation was found between CC area and the patient's level of cognitive and physical disability. Additionally, in Paper IV, shape analysis of the segmented CC revealed a significant association between disability and both CC thickness and bending angle. Conversely, in Paper V, a novel



method for automatic segmentation of the hippocampus is proposed, which consists of embedding a statistical shape prior as context information into a U-Net-based framework. The inclusion of shape information was shown to significantly improve segmentation accuracy when testing the method on a new unseen cohort (i.e., different from the one used for training). Furthermore, good performance was observed across three different diagnostic groups (healthy controls, subjects with mild cognitive impairment and Alzheimer’s patients) that were characterized by different levels of hippocampal atrophy.

In summary, the studies presented in this thesis support the great value of MRI image analysis for the advancement of neuroscientific knowledge, and their contribution is mostly two-fold. First, by applying well-established processing methods on datasets that had not yet been explored in the literature, it was possible to characterize specific brain changes and disentangle relevant problems of a clinical or biological nature. Second, a technical contribution is provided by modifying and extending already-existing brain image processing methods to achieve good performance on new datasets.

**Keywords:** Brain MRI, Image Segmentation, Machine Learning, Deep Learning, Shape Analysis, Aging, Neurodegeneration

## Sammanfattning

Magnetresonansbilder (MR-bilder) används för att framställa detaljerade bilder av hjärnan utan joniserande strålning. Från en strukturell MR-bild är det möjligt att extrahera morfologiska egenskaper hos hjärnans olika regioner, såsom deras volym och form. Dessa egenskaper kan ge bättre förståelse för förändringar som hjärnan utsätts för på grund av en mängd faktorer (exempelvis miljö eller sjukdom) samt bidra till att identifiera nya bildbaserade biomarkörer för neurologiska och psykiatriska sjukdomar. Den här avhandlingens huvudsakliga mål är att bidra till kunskapen om hur bildbehandling av MR-bilder kan användas för att analysera och karaktärisera hjärnstrukturer.

De två första delarbetena som ingår i avhandlingen är djurstudier som primärt avser att använda MR-data för att analysera skillnaderna mellan två kohorter. I Artikel I behandlas MR-bilder från domesticerade och vilda kaniner för att identifiera skillnader i hjärnstruktur mellan de två grupperna. Domesticering visade sig förändra hjärnstrukturen signifikant, både den gråa hjärnsubstansens volym och den vita hjärnsubstansens integritet. I Artikel II användes MR-bilder på råttor för att träna en datadriven modell att predicera hjärnålder. Modellen testades sedan på en kontrollgrupp och en grupp råttor som under flera månader utsattes för en mer stimulerande miljö samt fick en diet med restriktioner. Den mer hälsosamma livsstilen visade sig bidra till en lägre predicerad hjärnålder genom att sakta ner råttornas åldringsprocess, jämfört med kontrollgruppen. Hjärnåldern hos unga, vuxna råttor visade sig signifikant påverka råttornas överlevnad.

Artikel III, IV och V är människostudier som föreslår djupinlärningsbaserade metoder för att segmentera (avgränsa) hjärnstrukturer som kan påverkas av neurodegeneration. Artikel III och IV i synnerhet fokuserar på U-Net-baserad 2D-segmentering av corpus callosum (CC) hos patienter med multipel skleros. I båda studierna uppmättes god träffsäkerhet för segmenteringsalgoritmen och signifikant korrelation mellan CC:s area och patientens kognitiva och fysiska nedsättning. Utöver detta visar Artikel IV genom geometrisk analys av den segmenterade CC ett signifikant samband mellan sjukdom och CC:s tjocklek och böjvinkel. I Artikel V introduceras en ny metod för automatisk segmentering av hippocampus. Metoden kombinerar U-Net-baserad segmentering med en inbyggd statistisk representation av hippocampus' form. Metoden visade sig ge en signifikant förbättring av segmenteringskvaliteten när metoden utvärderades på en ny, tidigare osedd, kohort. Goda resultat uppmättes även i tre olika diagnosgrupper (en frisk kontrollgrupp, patienter med milda kognitiva symptom och en grupp patienter med Alzheimers sjukdom) som särskilde sig genom tre olika nivåer av atrofi av hippocampus.

Sammanfattningsvis bidrar studierna som ingår i avhandlingen till att förstärka värdet av MR-bildanalys för framsteg inom neurovetenskapen, och detta på två sätt. Genom att applicera väletablerade bildbehandlingsmetoder på dataset som ännu inte utforskats i litteraturen var det möjligt att karaktärisera specifika förändringar i hjärnans geometri och därmed lösa relevanta kliniska eller biologiska utmaningar. Vidare har studierna bidragit till den teknologiska metodutvecklingen genom att modifiera och utvidga existerande bildbehandlingsmetoder för hjärnbilder för att uppnå goda resultat på nya dataset.

**Nyckelord:** MRT av hjärnan, Bildsegmentering, Maskininlärning, Djupinlärning, Formanalys, Åldrande, Neurodegeneration

# Acknowledgments

It is hard for me to believe that my PhD journey has already come to an end. These years have marked one of the most enriching, rewarding and fun periods of my life. But it has not been all roses: there were many stressful days when I struggled with difficult research questions, never-ending code debugging and challenging course-work. On top of that, a pandemic broke out too! Despite these challenges, I feel proud for maintaining a positive attitude and having accomplished my objectives. But this would have not been possible without the support of all the great people that accompanied me during this journey.

The first people that I want to thank are my supervisors. Chunliang, you guided me throughout my PhD and inspired me with your love for research. Above that, I am very grateful for your ongoing dedication in helping me grow as an independent researcher. Örjan, I look up to you for your great experience and I could always count on your help in times of need. Especially when it came to statistical analysis or scientific writing, your feedback and support was irreplaceable. Eric, your enthusiasm in highly interdisciplinary research has been a key motivator for me. I have always been inspired by your ideas and advice, and it is thanks to you that I got the chance to spend part of my PhD in London. E grazie anche per le nostre chiacchierate in italiano! Lars-Olof, thank you for welcoming me with enthusiasm in your division at Karolinska Institute and for supporting this research collaboration.

I am also very lucky to have found other great colleagues at KTH. First of all, Rodrigo: everything started thanks to you in 2016, when I worked on my Master's thesis with you as supervisor. This experience made me become passionate about brain image analysis and I am grateful for your support throughout the years. I would also like to thank the PhD students of the group: the talented programmer and “graphic designer” Fabian, my fellow Unipd alumnus Simone, my teaching companion Mehdi and my office-mate Jingru. I could always count on you whenever I needed help or just an honest opinion and, thanks to you, my days in the office were never boring! And thank you also to my former KTH colleagues Gabriel, Daniel

and Yongjun: I have so many good memories of our work days together in Flemingsberg, back when my PhD defense still seemed so far away in the future. Gabriel, I am very grateful also for your continued support with my Swedish.

A big thanks also to Sebastian, Daniel, Olof, Konstantinos and the rest of the colleagues at the Division of Clinical Geriatrics at KI: our meetings and recurrent seminars have been an important part of my PhD journey. A special thanks goes to Michael too: it was great to work with you over the past two years, and I am also thankful for your help in understanding the intricate rules of our joint PhD program! And thank you, Tobias, for inviting me to collaborate with you and Michael and for the valuable feedback provided on our work.

My collaborators at King's College London also played a crucial role in my PhD experience. Diana, thank you for welcoming me into your group and for your tremendous support throughout my stay in London. I am very thankful for having been involved in such an interesting project. Eilidh, thank you for your patience and support, especially in the first stages of the project. Your passion for research is really inspiring. Mattia, thanks for always giving me valuable feedback. Your positive attitude and enthusiasm are contagious and made me feel very happy about my work. And thanks also to Eugene and Federico for all the useful scientific discussions and for your involvement in my work.

A big thank you also to Leif, Mats and the rest of the team who worked on the rabbit brain image analysis project, which then became the first study of my PhD thesis. Thank you for involving me in this project and for valuing my contribution.

I also want to thank all the friends that I met here in Stockholm and that made these past years unforgettable. A big thank you goes especially to the “the Ducks”: Yuri, the first friend that I made in Sweden, for being my “Dutch brother”; Sara, for being the best neighbor I could ask for, for your amazing pizzas and cakes and for understanding my struggles with the Swedish cold; Jairam, for being firstly a great flatmate and then also a fantastic bodybalance-mate and soccer-watching-mate; Shaghayegh, for being always the best host and for sharing my love for sunny holiday destinations; Marcel, for being the life of the party and my “student”; and finally Lieve and Vassilis, for the amazing dinners, brunches and parties you have hosted and for being part of so many special occasions throughout the years. I also want to give a special thank you to my former flatmates Sandra, Tove and Elin. I have so many good memories of our big dinners and parties, as well as of our nice chats in front of a cup of coffee or tea after a busy work day.

Despite the kilometers separating us, I am very thankful for the support that I have always received from my old friends from Italy. Thank you

to the Padova group: Jacopo, Sara, Valeria, Giulia, Camilla, Laura. No matter when we had last seen each other, every time we met or called it was like time never passed. The same can also be said about my friends from university. Alessia, we studied together and you know how much I was looking forward to this moment. And I hope that there will be many more special moments to celebrate together. Andrea and Marco, thanks for being the funniest people I know, for sharing the joys and sorrows of the expat life and for all our past (and future!) reunions around Europe.

A very special thank you also to Xavier. For your very positive and calm attitude, which is exactly what I needed in the most stressful times. For never failing to make me laugh. For moving to Europe and for the time living together in London. And in general for being my number one supporter!

Finally, words are not enough to express my gratitude towards my parents, thanks to whom I had the privilege to study and to follow my dreams. I hope that my achievements will always make you happy and proud.



# List of scientific papers

This doctoral thesis is based on the five scientific papers listed below.

## Paper I

### **Changes in brain architecture are consistent with altered fear processing in domestic rabbits**

I. Brusini, M. Carneiro, C. Wang, C.J. Rubin, H. Ring, S. Afonso, J.A. Blanco-Aguilar, N. Ferrand, N. Rafati, R. Villafuerte, Ö. Smedby, P. Damberg, F. Hallböök, M. Fredrikson, L. Andersson.

In *Proceedings of the National Academy of Sciences*, 115 (28), pp. 7380-7385. 2018.

## Paper II

### **MRI-derived brain age as a biomarker of ageing in rats: validation using a healthy lifestyle intervention**

I. Brusini, E. MacNicol, E. Kim, Ö. Smedby, C. Wang, E. Westman, M. Veronese, F. Turkheimer, D. Cash.

In *Neurobiology of Aging*, 109, pp. 204-215. 2021.

## Paper III

### **Automatic deep learning multicontrast corpus callosum segmentation in multiple sclerosis**

I. Brusini\*, M. Platten\*, R. Ouellette, F. Piehl, C. Wang, T. Granberg.

\* Shared first authorship.

In *Journal of Neuroimaging*. 2022.



## **Paper IV**

### **Deep Learning Corpus Callosum Segmentation as a Neurodegenerative Marker in Multiple Sclerosis**

M. Platten, I. Brusini, O. Andersson, R. Ouellette, F. Piehl, C. Wang, T. Granberg.

In *Journal of Neuroimaging*, 31 (3), pp. 493-500. 2021.

## **Paper V**

### **Shape information improves the cross-cohort performance of deep learning-based segmentation of the hippocampus**

I. Brusini, O. Lindberg, J. Muehlboeck, Ö. Smedby, E. Westman, C. Wang.

In *Frontiers in Neuroscience*, 14, p. 15. 2020.

## Division of work between authors

The contribution of every author of each paper appended to this thesis is reported below.

### Paper I

MC, FH and LA formulated the main scientific hypothesis.

MC, RV, FH, LA designed the overall research to be conducted. PD designed the image acquisition protocols. **IB** designed the image processing methodology, with frequent input from CW and ÖS.

MC and PD provided the resources for the study (animal breeding and scanning). CW performed the preliminary skull-stripping of the images, while **IB** conducted all other image processing tasks. The results were analyzed jointly by **IB**, MC, CW, C-J R, NR, ÖS, PD FH, MF and LA. **IB** performed the statistical analysis (with help from C-J R) and received input from the other co-authors regarding the interpretation of the results.

**IB**, MC, CW, MF and LA jointly wrote the original draft of the paper. All co-authors provided critical feedback.

### Paper II

**IB**, EM, EW, MV, FT and DC formulated the scientific hypothesis.

**IB**, EM and EK designed the data preprocessing steps. **IB** designed the machine learning-based part of the project, with inputs from MV, CW and FT. **IB** and ÖS designed the statistical tests and survival analysis.

EM provided the data resources. **IB** implemented all data processing and machine learning pipelines. **IB** and ÖS performed the statistical analyses.

**IB** wrote the original draft of the manuscript. All co-authors provided critical feedback.

### Paper III

The scientific hypothesis was jointly formulated by **IB**, MP, CW and TG.

**IB** and MP designed the study while receiving input from CW and TG.

MP annotated the data. **IB** implemented both the deep learning-based algorithm. **IB** and MP jointly performed all data analysis.

**IB** and MP jointly wrote the original manuscript draft. All co-authors provided critical feedback.

## Paper IV

MP, TG and CW formulated the main scientific hypothesis. **IB** and CW formulated the hypothesis regarding the shape analysis part of the work, with input from MP.

The segmentation study design was proposed by MP, TG and CW. **IB** and CW designed the shape analysis part. **IB** implemented the shape analysis methods. MP, OA and CW developed the deep learning-based segmentation algorithm.

MP performed the analysis of the segmentation results, while **IB** analyzed the outcomes of shape analysis.

MP wrote the original draft of the manuscript, except for the shape analysis part that was drafted jointly by **IB** and MP. All co-authors provided critical feedback.

## Paper V

**IB**, CW, ÖS and EW formulated the main scientific hypothesis.

**IB**, CW, ÖS and EW designed the study. **IB** and CW designed how to integrate shape models into the pipeline.

**IB** and CW developed the segmentation pipeline and shape context integration. **IB** developed the shape model. OL provided the manual segmentations. J-SM organized the database and provided the FreeSurfer segmentations. **IB**, CW, ÖS, EW, and OL worked on the final method evaluation and analysis of the results.

**IB** wrote the original draft of the manuscript. All co-authors provided critical feedback.

# Other scientific contributions

During her doctoral studies, the student has also contributed to the following publications that are not included in this thesis.

## Journal papers

- **Disrupted *Cacna1c* Gene Expression Perturbs Spontaneous  $\text{Ca}^{2+}$  Activity Causing Abnormal Brain Development and Increased Anxiety.**  
E. Smedler, L. M. Louhivuori, R. Romanov, D. Masini, I. Dehnisch Ellström, C. Wang, M. Caramia, Z. West, S. Zhang, P. Rebellato, S. Malmersjö, I. Brusini, S. Kanatani, G. Fisone, T. Harkany, P. Uhlén.  
In *Proceedings of the National Academy of Sciences*, 119(7).2022.
- **Age-Specific Adult Rat Brain MRI Templates and Tissue Probability Maps.**  
E. MacNicol, P. Wright, E. Kim, I. Brusini, O. Esteban, C. Simmons, F. Turkheimer, D. Cash.  
In *Frontiers in Neuroinformatics*, vol. 15, p. 74. 2022.

## Conference papers

- **Fully automatic estimation of the waist of the nerve fiber layer at the optic nerve head angularly resolved.**  
I. Brusini, G. Carrizo, S. Bendazzoli, C. Wang, Z. Yu, C. Sandberg-Melin, K. Kisonaite, F. Raeme, P. G. Söderberg.  
In *Ophthalmic Technologies XXXI*, vol. 11623, p. 116231D. International Society for Optics and Photonics. 2021.
- **Automatic rat brain segmentation from MRI using statistical shape models and random forest.**

S. Bendazzoli, I. Brusini, P. Damberg, Ö. Smedby, L. Andersson, C. Wang.

In *Medical Imaging 2019: Image Processing*, vol. 10949, p. 109492O. International Society for Optics and Photonics. 2019.

- **Voxel-Wise Clustering of Tractography Data for Building Atlases of Local Fiber Geometry.**

I. Brusini, D. Jörgens, Ö. Smedby, R. Moreno.

In *Computational Diffusion MRI: International MICCAI Workshop, Granada, Spain, September 2018*, p. 345. Springer. 2019.

## Conference abstracts

- **A deep learning-based pipeline for error detection and quality control of brain MRI segmentation results.**

I. Brusini, D. F. Padilla, J. Barroso, I. Skoog, Ö. Smedby, E. Westman, C. Wang.

In *Medical Imaging with Deep Learning*. 2020.

- **Influence of tractography algorithms and settings on local curvature estimations.**

I. Brusini, D. Jörgens, Ö. Smedby, R. Moreno.

In *Proceedings of OHBM Annual Meeting*. 2017.

## Preprints

- **Humanized *CYP2C19* transgenic mouse as an animal model of cerebellar ataxia.**

F. Milosavljević, I. Brusini, A. Atanasov, M. Manojlović, M. Novalen, M. Vučić, Z. Oreščanin Dušić, J. Brkljačić, S. Miksys, Č. Miljević, A. Nikolić, D. Blagojević, C. Wang, P. Damberg, V. Pešić, R. F. Tyndale, M. Ingelman-Sundberg, M. M. Jukić.

*bioRxiv preprint 2022.01.10.475612*. 2022.

- **Development and evaluation of a 3D annotation software for interactive COVID-19 lesion segmentation in chest CT.**

S. Bendazzoli, I. Brusini, M. Astaraki, M. Persson, J. Yu, B. Connolly, S. Nyrén, F. Strand, Ö. Smedby, C. Wang.

*arXiv preprint arXiv:2012.14752*. 2020.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Research aims</b>	<b>5</b>
<b>3</b>	<b>Background</b>	<b>9</b>
3.1	Basics of neuroscience . . . . .	9
3.1.1	Organization of the nervous system . . . . .	9
3.1.2	Brain anatomy . . . . .	11
3.1.3	Human versus rodent brains . . . . .	14
3.1.4	Brain aging and neurodegeneration . . . . .	15
3.2	Magnetic resonance imaging . . . . .	20
3.2.1	Nuclear magnetic resonance . . . . .	20
3.2.2	Image reconstruction . . . . .	22
3.2.3	T1- and T2-weighted MRI . . . . .	23
3.2.4	Diffusion MRI . . . . .	24
3.3	Medical image processing . . . . .	27
3.3.1	Image registration . . . . .	27
3.3.2	Image segmentation . . . . .	33
3.3.3	Other relevant tools for brain MRI image processing . . . . .	38
3.4	Machine learning in medical image processing . . . . .	40
3.4.1	An overview of relevant models . . . . .	41
3.4.2	Introduction to deep learning . . . . .	45
3.4.3	Convolutional neural networks . . . . .	46
3.4.4	Image segmentation with U-Net . . . . .	48
3.4.5	Incorporating context information into a DNN . . . . .	50
3.4.6	Relevant applications in brain MRI image processing . . . . .	51

CONTENTS

<b>4</b>	<b>Methods</b>	<b>57</b>
4.1	Investigating group differences from brain MRI: Two animal studies . . . . .	57
4.1.1	Overview of the studies . . . . .	57
4.1.2	Brain extraction . . . . .	58
4.1.3	Image registration and segmentation . . . . .	59
4.1.4	Voxel-based morphometry . . . . .	60
4.1.5	Tract-based spatial statistics . . . . .	61
4.1.6	Brain age prediction . . . . .	61
4.1.7	Statistical analyses . . . . .	62
4.2	Deep learning and shape analysis of the degenerating human brain . . . . .	63
4.2.1	Overview of the studies . . . . .	64
4.2.2	Automatic midslice selection . . . . .	65
4.2.3	U-Net-based segmentation . . . . .	65
4.2.4	Shape analysis . . . . .	68
4.2.5	Method evaluation . . . . .	70
<b>5</b>	<b>Results</b>	<b>73</b>
5.1	Investigating group differences from brain MRI . . . . .	73
5.1.1	Paper I . . . . .	73
5.1.2	Paper II . . . . .	76
5.2	Deep learning and shape analysis of the degenerating human brain . . . . .	77
5.2.1	Paper III . . . . .	77
5.2.2	Paper IV . . . . .	81
5.2.3	Paper V . . . . .	81
<b>6</b>	<b>Discussion</b>	<b>87</b>
6.1	The important role of brain MRI image analysis . . . . .	87
6.1.1	MRI image analysis for better understanding of the brain . . . . .	87
6.1.2	MRI-based biomarkers . . . . .	91
6.2	Machine learning: The future of neuroradiology? . . . . .	96
6.2.1	Advantages of machine learning approaches . . . . .	96
6.2.2	Obstacles and ethical concerns . . . . .	98
6.2.3	Aiming for balance between human and machine decision making . . . . .	100
6.3	Limitations . . . . .	100

<b>7</b>	<b>Future work</b>	<b>103</b>
7.1	Larger and heterogeneous datasets . . . . .	103
7.2	Validation of MRI-derived biomarkers . . . . .	104
7.3	Testing new architectures . . . . .	106
<b>8</b>	<b>Conclusions</b>	<b>107</b>
	<b>Bibliography</b>	<b>109</b>
	<b>Appended papers</b>	<b>129</b>





# List of Abbreviations

<b>2D</b>	Two-dimensional
<b>3D</b>	Three-dimensional
<b>AD</b>	Alzheimer’s disease
<b>ADC</b>	Apparent diffusion coefficient
<b>AUROC</b>	Area under the receiver operating characteristic curve
<b>BrainAGE</b>	Brain age gap estimation
<b>CC</b>	Corpus callosum
<b>CI</b>	Confidence interval
<b>CNN</b>	Convolutional neural network
<b>CNS</b>	Central nervous system
<b>CSF</b>	Cerebrospinal fluid
<b>CT</b>	Computed tomography
<b>DL</b>	Deep learning
<b>dMRI</b>	Diffusion magnetic resonance imaging
<b>DNN</b>	Deep neural network
<b>DT</b>	Diffusion tensor
<b>DTI</b>	Diffusion tensor imaging
<b>EDSS</b>	Expanded disability status scale

## *CONTENTS*

<b>EEDR</b>	Environmental enrichment and dietary restriction
<b>FA</b>	Fractional anisotropy
<b>FCN</b>	Fully convolutional network
<b>FFD</b>	Free-form deformation
<b>FID</b>	Free induction decay
<b>FLAIR</b>	Fluid-attenuated inversion recovery
<b>GM</b>	Gray matter
<b>GP</b>	Gaussian process
<b>GPR</b>	Gaussian process regression
<b>IC</b>	Intracranial
<b>ICC</b>	Intraclass correlation coefficient
<b>LR</b>	Logistic regression
<b>MAP</b>	Maximum a posteriori
<b>MCI</b>	Mild cognitive impairment
<b>ML</b>	Machine learning
<b>MLE</b>	Maximum likelihood estimation
<b>MRI</b>	Magnetic resonance imaging
<b>MS</b>	Multiple sclerosis
<b>NN</b>	Neural network
<b>PC</b>	Principal component
<b>PCA</b>	Principal component analysis
<b>PNS</b>	Peripheral nervous system
<b>ROI</b>	Region of interest
<b>SDMT</b>	Symbol digit modality test

<b>T1w</b>	T1-weighted
<b>T2w</b>	T2-weighted
<b>TBSS</b>	Tract-based spatial statistics
<b>TPM</b>	Tissue probability map
<b>VBM</b>	Voxel-based morphometry
<b>WM</b>	White matter



## Chapter 1

# Introduction

Brain magnetic resonance imaging (MRI) is a noninvasive imaging modality. It can provide good contrast between soft tissues and high spatial resolution without using any ionizing radiation. For these reasons, it is widely used in the clinic for both diagnostics and disease monitoring. Nonetheless, brain MRI image processing research is extremely active with the aim to expand the range of clinical applications of brain MRI by introducing so-called imaging biomarkers. In a commentary by the Biomarkers Definitions Working Group (2001) of the American National Institute of Health, a biomarker was defined as “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses.” Neuroimaging biomarkers are examples of such objective measurements, and they are specifically extracted from brain imaging data by applying a set of predefined image processing techniques.

Important information on brain morphology can be extrapolated from structural MRI scans. The segmentation of certain brain regions of interest can allow the study of their morphological properties (e.g., volume and shape), which represent potential biomarkers of some diseases. For instance, significant volumetric reductions (atrophy) in some brain regions (e.g., the hippocampus and entorhinal cortex) have been demonstrated to correlate with Alzheimer’s disease (AD) progression (Scheltens et al., 2002; Varghese et al., 2013). Thus, early detection of this atrophy can aid in the early diagnosis of AD and its treatment planning. Furthermore, structural MRI can be used to identify other markers of brain damage and/or degeneration, for example the presence of brain lesions and their extent, as well as level

## CHAPTER 1. INTRODUCTION

of vascular burden (Young et al., 2020).

Advanced image processing methods can be employed to improve understanding of brain structure and how it changes due to multiple factors (e.g., underlying pathology, genetics, environment or normal aging). Several software tools for brain image processing as well as rich image datasets are now available to the scientific community with the aim of advancing research in this field. Moreover, in the last decades, the explosion of machine learning and, more recently, deep learning has attracted the attention of the neuroimaging community given the strong potential and accuracy of these tools, especially when it comes to automatic image classification and segmentation. A prime example of this phenomenon is the popularity of the U-Net model for medical image segmentation (Ronneberger et al., 2015), which presented an alternative to other deep learning solutions based on convolutional neural networks. This novel fully convolutional architecture, trained together with extensive data augmentation, was originally shown to achieve excellent performance on cell segmentation from 2D light microscopy images, outperforming all other previously proposed methods. To date, the original U-Net paper by Ronneberger et al. (2015) has received over 37 000 citations and has been used in applications that go far beyond the segmentation of microscopy images. It has since been revisited and modified for various medical image segmentation tasks including brain MRI segmentation (Kermi et al., 2019; Mehta and Arbel, 2019; Wu et al., 2019; Hwang et al., 2019; Dong et al., 2017; Lee et al., 2020), and it constitutes the baseline model in multiple studies.

This thesis aims to add a new piece to the brain image analysis puzzle by advancing knowledge on how image processing tools can be used to analyze and characterize brain morphology. The present work includes both animal and human brain MRI studies, and its contribution is generally two-fold. The first general goal is to show how the use of well-established image processing methods can help disentangle problems of a broader clinical or biological nature by analyzing the properties of brain MRI datasets that have not yet been explored in the literature. Secondly, this thesis aims to provide a technical contribution to the field by proposing new ways to revisit, modify and extend already-existing methods in order to achieve good performance on the datasets of interest for the presented studies.

In the next chapter, the specific aims of each of the five papers appended to this thesis are outlined. Subsequently, the theoretical and research background that constitute the foundation of these five papers is presented. In

Chapters 4 and 5, the methods and results of these research studies are summarized. The results are then thoroughly discussed in Chapter 6, while the possible future outlook of the presented research and the final conclusions to be drawn from each paper are outlined in Chapters 7 and 8, respectively. Finally, the last part of this thesis comprises the full texts of the appended papers.





## Chapter 2

# Research aims

The overall aim of the present doctoral thesis is to strengthen the importance of MRI image analysis for analyzing brain structure and, more specifically, characterizing the changes that it undergoes due to multiple factors (evolutionary, environmental or pathological). This is done, first of all, by using well-established image processing tools on new, unique datasets with the aim of solving clinically or biologically relevant questions using MRI. Furthermore, some of the already-existing image analysis methods are revisited and modified in order to provide new technical contributions to the brain image analysis literature, with a special focus on the computation of clinically relevant neuroimaging biomarkers.

This thesis is made up of five different scientific papers, which are appended at the end of this work. Below, the motivations and aims of each paper are outlined.

### Paper I

**Motivation:** Through evolution, domestic animals have developed very different behavioral traits compared to their wild counterparts (Jensen, 2014), allowing them to live in contact with humans and suppress fight and flight responses. The rabbit constitutes an ideal case for studying the differences between domestic and wild animals since many wild rabbits still exist in Southern France (where their domestication originally began), and rabbits are already used and suited for genotypic and phenotypic studies. A previous work by Carneiro et al. (2014) reported genetic differences between wild and domestic rabbits, especially in the vicinity of genes responsible for

## CHAPTER 2. RESEARCH AIMS

neural development. This suggests the presence of brain alterations between the two groups.

**Aim:** To analyze high-resolution ex vivo brain MRI scans from wild and domestic rabbits in order to identify structural brain differences between the two groups.

### Paper II

**Motivation:** Previous studies on humans have shown that brain MRI scans can be used as input to machine learning-based models to predict the chronological age of healthy participants with high accuracy. If the model predicts an age that is greater than the subject's chronological age, this might reflect an underlying disease or neurodegeneration. Conversely, a younger predicted age could indicate a positive trend in the aging process. Therefore, the BrainAGE score (defined as the difference between predicted and chronological age) has been proposed as an aging biomarker. However, validations of this biomarker on laboratory animals are currently lacking.

**Aim:** To implement a new MRI-based brain age prediction model for rats and to test it on a cohort including both controls and a group of rats that have undergone long-term environmental enrichment and dietary restriction. The BrainAGE scores of controls are compared against those from the healthy lifestyle group in order to investigate whether this lifestyle intervention has a significant effect on BrainAGE.

### Papers III and IV

**Motivation:** Multiple sclerosis (MS) is an autoimmune demyelinating disease characterized by neuroinflammation and neurodegeneration. The corpus callosum (CC), a large fiber bundle that connects the two brain hemispheres, is known for being highly susceptible to MS progression. Therefore, the identification of CC atrophy from brain MRI can constitute a relevant biomarker of neurodegeneration in MS.

**Aim:** To develop efficient methods for automatic CC segmentation on the mid-sagittal slice of T1-weighted (Paper III), FLAIR (Paper III) and T2-weighted (Paper IV) MRI scans of MS patients. Moreover, in both papers, the correlation between the normalized CC area and disability in MS is investigated. Additionally, in Paper IV, the relation between disability and CC shape is analyzed.

## Paper V

**Motivation:** Atrophy of the hippocampus is one of the earliest signs of Alzheimer’s disease (AD). Thus, accurate hippocampal segmentation tools are needed in order to reliably analyze the structural integrity of this brain structure as a potential neuroimaging biomarker of AD. However, this can be an extremely challenging task, especially when segmenting images from new unseen cohorts (i.e., differing from those used for training the segmentation models) and from patients with a high degree of hippocampal atrophy.

**Aim:** To propose a novel, deep learning framework for hippocampal segmentation that embeds statistical shape models as an additional input context layer of a traditional U-Net architecture. The model accuracy is investigated both within a training cohort (through cross-validation) and on a new independent cohort. All used datasets include healthy controls, subjects with mild cognitive impairment and AD patients.



## Chapter 3

# Background

### 3.1 Basics of neuroscience

#### 3.1.1 Organization of the nervous system

The nervous system controls voluntary and involuntary actions by transmitting or receiving signals within the body (Swanson, 2013; Brown, 2012). The units of the nervous system are the nerve cells, also called neurons, which interact with each other at membrane specializations (gap junctions and chemical synapses). The receptive surface of a neuron usually includes the dendrites, the cell body and its axon terminals. Based on the anatomical location of the neurons, the overall nervous system is divided into two parts: the central nervous system (CNS) and peripheral nervous system (PNS). The CNS is made up of the brain and the spinal cord. In contrast, the PNS comprises different types of nerve cells: some are located completely outside the CNS; some have their cell bodies in the CNS but innervate muscles or peripheral neurons, while others have cell bodies outside the CNS but originate from sense organs and innervate the CNS.

An alternative way of dividing the nervous system is based on its microarchitecture, that is on cellular organization across the system (Swanson, 2013; Purves, 2018). This approach is most often applied to the CNS only, and it consists in dividing it broadly into gray matter (GM) and white matter (WM), which are represented in Figure 3.1. The first mainly consists of neural cell bodies, which can have different shapes, sizes, locations and packing densities. These cells can be organized as nuclei, which are accumulations of neurons having similar functions and connections, and can be found, for example, in the spinal cord, the cerebrum and the brainstem.

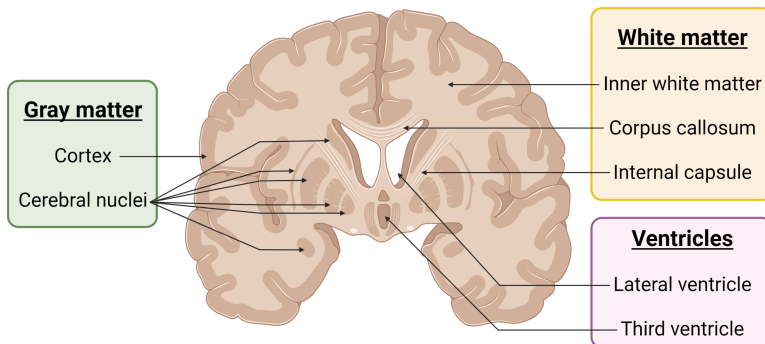


Figure 3.1: Schematic representation of a coronal brain slice. Gray matter and white matter structures, as well as the ventricles (which are filled with cerebrospinal fluid) are indicated. (Figure created with BioRender.com)

GM cells can also be arranged to make up cortices, which have a sheet-like structure and are found both in the two cerebral hemispheres and the cerebellum. WM is made up of axons that—within the CNS—form neural tracts, which are more or less similar to the nerves that characterize the PNS. All neural tracts that cross the midline of the brain are also referred to as commissures.

However, neurons are not the only cellular components of the nervous system. Glial cells are at least as abundant as neurons and can play multiple roles such as supporting the propagation of the neural signals, repair after neural injury and response to toxins by aiding communication between the immune system and the brain (Purves, 2018). Glial cells can be divided into three groups. Astrocytes make up the blood-brain barrier that separates blood flow from the brain, and they also aid in the efficient creation of synapses. Microglia have a phagocytic function that allows eliminating cellular debris and can actively regulate synaptogenesis. Finally, oligodendrocytes generate myelin around axons, which is a lipid-rich substance that facilitates the isolation of the neural signal inside the axons, and therefore a more efficient communication between neurons.

In addition to the approximately 86 billion neurons and at least as many glia, other types of cells can be identified within the nervous system. For example, normal brain function relies not only on correct functioning of neurons and glia, but also on vasculature (Zeisel et al., 2015). Therefore,

it is possible to identify multiple mural and endothelial cells that control vascular contraction and play an important role in the blood-brain barrier. Ependymal cells are also present and are particularly important for generating and regulating the cerebrospinal fluid (CSF), a fluid that surrounds both the brain and the spinal cord. CSF is formed from blood plasma and provides the basic mechanical and immunological protection to the organs of the nervous system.

#### 3.1.2 Brain anatomy

The most complex organ of not only the nervous system but the whole human body is the brain. It provides humans with the ability to process sensory stimuli as well as to perform cognitive functions (e.g., memory, attention and language). Moreover, thanks to the brain, many visceral, endocrine and musculoskeletal functions are regulated. Therefore, the brain is an irreplaceable organ in the human body, and the understanding of its functioning as well as of its pathological processes has always been a focus of attention in scientific research.

From an anatomical perspective, the brain can be divided into three main parts: cerebrum, brainstem and cerebellum (see Figure 3.2A). The cerebrum consists of the left and right cerebral hemispheres (also referred to as the telencephalon), the diencephalon and subcortical nuclei. The brainstem comprises the midbrain (mesencephalon), the pons (which is part of the metencephalon, together with the cerebellum) and the medulla oblongata (myelencephalon). The content of this section is based on the *Brain Anatomy* chapter by Sharma and Majsak (2014), as well as the book *Neuroscience* by Purves (2018).

#### Cerebrum

The two cerebral hemispheres are mainly made of GM cerebral cortex and of a dense network of WM fibers that allow communication between different brain regions and with the periphery. Moreover, inside each of the two hemispheres, a lateral ventricle can be identified. These two lateral ventricles are the largest of the four cerebral ventricles, which are cavities containing CSF.

The cortical part of the cerebrum can be 1.5 to 5.0 mm thick. It is organized into layers that can be divided in two groups: the allocortex, which is deeper and phylogenetically older, and the neocortex, which takes



## CHAPTER 3. BACKGROUND

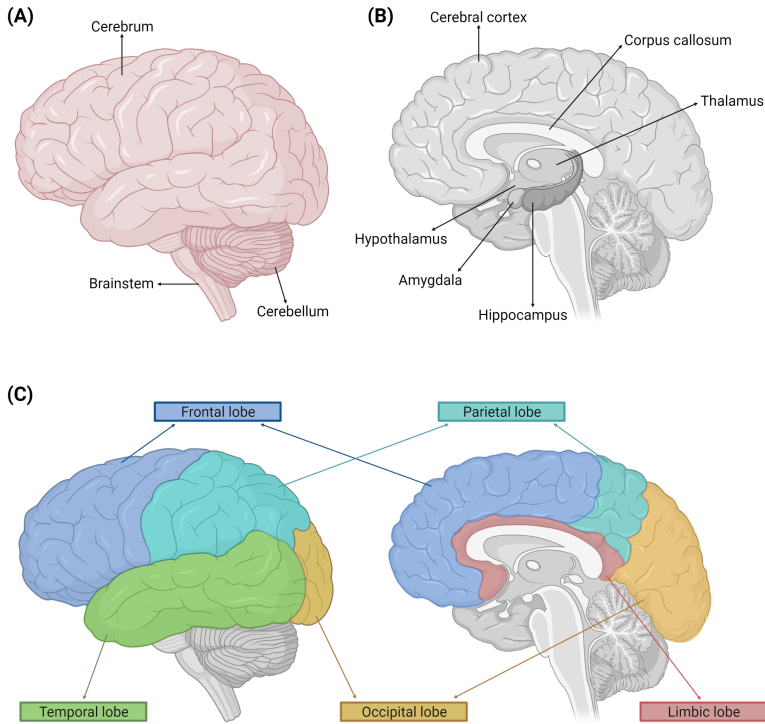


Figure 3.2: Basics of brain anatomy. In (A), the subdivision into cerebrum, brainstem and cerebellum is shown. (B) represents a sagittal section of the brain, and the location of some important brain structures is highlighted. In (C), the five cerebral lobes are shown. (Figure created with BioRender.com)

up 90% of the cortex and is responsible for higher-order brain functions. The brain cortex is extremely convoluted into different ridges (gyri) and grooves (sulci). Gyri and sulci also define the edges of the five main lobes of the cerebral hemispheres: frontal, parietal, temporal, occipital and limbic (see Figure 3.2C). Each lobe is associated with different functions. For example, the frontal lobe participates in decision making, motor planning, attention and voluntary movements. The temporal lobe is more involved in learning and memory as well as in the processing of auditory and visual information. In contrast, the limbic lobe (which is the only lobe situated in the medial as opposed to the lateral brain surface) constitutes a very important area

for both memory and emotional responses.

Regarding the WM portion of the hemispheres, which is made of neural fiber bundles, its main role is to coordinate and connect different cortical regions as well as the diencephalon, brainstem and spinal cord. Within each hemisphere, the major WM fiber bundles are the superior longitudinal fasciculus, the superior and inferior occipitofrontal fasciculi and the cingulum. The main and fundamental bundle connecting the two hemispheres with each other is the corpus callosum (CC), which is made up of commissural fibers.

Between the two hemispheres and above the brainstem is the diencephalon (from the Greek: di[a]-“across” + enkephalos “brain”). The largest region of the diencephalon is the thalamus, which receives information from multiple sources in the CNS: cortical areas responsible for mental processes, neuronal centers involved in emotions, as well as sensory (e.g., auditory and visual) and motor pathways (from the cerebellum and the basal ganglia). These inputs are processed by the thalamus and later terminate in the cerebral cortex. Another important structure of the diencephalon is the hypothalamus, whose roles include the control of blood circulation and body temperature, the production of hormones regulating water and food intake as well as emotional expression.

Finally, located deep between the two hemispheres, three paired (i.e., one per hemisphere) brain structures are found. One consists of the basal ganglia, which are involved in the regulation of motor performance. Second, there is the hippocampus, which plays a fundamental role in the formation and storage of memory. Finally, there is the amygdala, which participates in several attentional and emotional processes as well as in episodic-autobiographical memory.

## **Brainstem**

The medulla is the most caudal part of the brainstem, and it is directly connected with the spinal cord. It controls fundamental and primordial body functions such as breathing, digestion and heart rate. Superior to the medulla is the pons, which mostly comprises neural pathways connecting different areas of the brain with the cerebellum. Finally, the midbrain is responsible for multiple processes including motor functions (e.g., eye movement), sensory functions (e.g., auditory and visual reflexes), pain modulation, attention and alertness.

### Cerebellum

Although the cerebellum accounts for only 10% of total brain volume, it comprises approximately 50% of the CNS neurons. It plays a fundamental role in equilibrium and learning of motor tasks thanks to a closed loop of neural circuits that allow correction of motor errors. It is also responsible for controlling the timing, magnitude and direction of movements. Moreover, the cerebellum is involved in some cognitive functions, including nondeclarative memory, attention and emotional control.

#### 3.1.3 Human versus rodent brains

In brain research, it is common to use small-animal models (especially mice and rats) to better understand how the brain is affected by specific genetic and environmental conditions, diseases or therapies. Moreover, animal models have the great advantage of being characterized by group homogeneity (Hoyer et al., 2014). However, the differences between human and animal brains must be considered when carrying out neuroimaging research, which is the focus of the present thesis.

The first clear difference between a human and a rodent brain is of course the size: the volume of a human brain is about 1450 ml, while that of a mouse ranges from 0.5 to 0.6 ml (Hoyer et al., 2014). In brain imaging, size differences must be accounted for when it comes to image resolution: a resolution of 1 mm is normally satisfactory for humans, but the same cannot be said for mouse or rat data. This is why animal MRI scanners are usually characterized by higher magnetic field magnitude (in the range of 4.7–11.7T), specific radiofrequency coils and optimized imaging sequences, leading to higher resolutions (Denic et al., 2011). Furthermore, animals are most often imaged either under anesthesia or *ex vivo*, since it is extremely challenging—if not impossible—to habituate them to the scanning procedures. While this can be a problem for functional MRI studies (where anesthesia can confound the results), it does not normally constitute a limitation for structural MRI. More theoretical background on MRI imaging can be found in Section 3.2.

Although human and rodent brains have common anatomical divisions (i.e., those outlined in Section 3.1.2), these structures have evolved differently across species (Northcutt, 2002). For example, as it can be seen in Figure 3.3, the surface of a mouse (or a rat) brain is smooth, and thus it does not have the sulci and gyri found on the human cortex that further

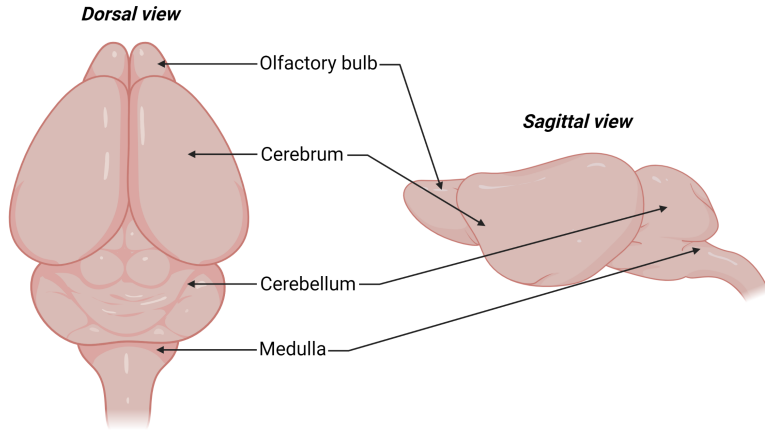


Figure 3.3: Representation of a mouse brain from a dorsal (left) and sagittal (right) view. (Figure created with BioRender.com)

increase its surface area. The increased cortical surface allows the formation of very different and specialized areas on the human cortex related to higher mental functions that rodents have not developed (e.g., language and cognition). The olfactory bulbs constitute an example of a region that is much larger in rodents (with respect to total brain size); unlike humans, rodents interact with the surrounding environment by relying mostly on their sense of smell. In contrast, the cerebellum has developed similarly in both humans and rodents; the motor functions controlled by this structure are highly important for both species.

Therefore, given the multiple anatomical differences existing across species, it is important to use and adapt appropriate image analysis tools in order to properly translate from preclinical (i.e., on laboratory animals) to clinical (on humans) neuroimaging research and vice versa. More details on brain image processing methods can be found in Section 3.3.

#### 3.1.4 Brain aging and neurodegeneration

The number of elderly individuals across the world is rapidly growing: in 2020, around 9.3% of the world's population was aged 65 years or older, and it has been estimated that this percentage could increase to 16% by 2050 (the United Nations Department of Economic and Social Affairs, 2020). However, the penalty of this longer life expectancy is an increased risk for

## CHAPTER 3. BACKGROUND

elderly individuals to develop diseases related to aging, and the brain—like many other parts of the human body—can be considerably affected by this (Denver and McClean, 2018). Aging is indeed an important risk factor for developing dementia, which is a clinical syndrome that comprises a range of recognizable symptoms that impair independent living, for example language disturbances, memory loss, damaged perceptual-motor function and impaired attention (Burns and Iliffe, 2009; Irwin et al., 2018). When dementia is associated with an underlying deterioration of brain cells, neurodegenerative dementia is involved, which is both progressive and irreversible (Ripich and Horner, 2004). The four main types of neurodegenerative dementia are Alzheimer’s disease (AD, which accounts for about 60% of all dementias), vascular dementia, dementia with Lewy bodies and frontotemporal dementia. It is also important to point out that neurodegenerative dementias do not constitute the only brain degenerative diseases. Other diseases characterized by neurodegeneration include multiple sclerosis (MS), Huntington’s disease and Parkinson’s disease (Hardiman et al., 2011).

In this section, three topics of relevance to the present thesis are presented. First, the main characteristics of normal brain aging are summarized and later compared with the main features of AD progression. Finally, neurodegeneration caused by MS is discussed.

### Normal brain aging

Even when no underlying disease is present, normal brain aging can be associated with impaired cognitive functions and presence of neuropathological lesions (Denver and McClean, 2018). These conditions can indeed be considered benign when they can still be compensated for and do not relevantly affect a subject’s daily life. In particular, normal brain aging can be characterized by decreased working memory and processing speed, symptoms that usually become more evident above 60 years of age.

From an anatomical perspective, these cognitive complaints have been associated with increased atrophy in the prefrontal lobes (Swerdlow, 2011), although this is not the only area affected by age-related changes. Multiple studies have shown that the GM undergoes a general shrinkage, but the magnitude of this phenomenon changes across different brain areas (Good et al., 2001; Walhovd et al., 2005). For example, the cortex is normally affected by more atrophy than subcortical structures. Moreover, the relationship between GM volume reduction and age can be linear or nonlinear according to the brain region being analyzed. Similarly, small WM changes can

be observed in healthy elderly subjects, for example WM volume reduction or disrupted microstructural integrity (Gunning-Dixon et al., 2009). This loss of WM integrity is believed to be associated with age-related cognitive decline.

In order to promote healthy brain aging, it is important to investigate evidence-based strategies that could help prevent the occurrence of neurodegenerative diseases. Such strategies may include the use of specific medications and the control of subject-specific risk factors (e.g., diabetes or high blood pressure) but also lifestyle and environmental changes (Desai et al., 2010). Multiple studies have indeed shown that cognitive functions are positively influenced by environmental enrichment and physical exercise in both rodents and humans (Speisman et al., 2013; Hötting and Röder, 2013). A healthier lifestyle has been shown to enhance neurogenesis, synaptic remodeling and brain plasticity. Moreover, brain plasticity, together with neuroprotection, has been shown to be augmented by applying dietary changes (Mattson, 2010; Martin et al., 2006). Thus, multidomain lifestyle changes (i.e., the combination of better diet, physical activity and a stimulating environment) can help positively influence healthy brain aging.

#### **Alzheimer’s disease**

Some of the clinical and neuropathological features of AD can also be present during the healthy aging process described in the previous paragraph. Therefore, it can be challenging to distinguish AD from normal aging, especially during the first stages of the disease (Denver and McClean, 2018). While healthy aging can be characterized by moderate impairment of some cognitive abilities, AD is associated with a more severe decline of a broader range of cognitive functions that can affect the independent living of the patient as well as their personal, social and professional life. However, the exact mechanism behind the initiation of AD neurodegeneration is still unknown, as are the causes of AD diverging from healthy aging (Irwin et al., 2018).

According to the AD staging criteria defined by Braak and Braak (1995), the initiation of the disease itself takes place in the transentorhinal cortex (Stages I and II), later reaching the hippocampus (Stages III and IV) and finally involving the neocortex (Stage V). These stages were originally defined based on the accumulation of intracellular tau neurofibrillary tangles, which, together with the loss of synapses and the deposition of extracellular amyloid beta plaques, are the main hallmarks of an AD brain (see

## CHAPTER 3. BACKGROUND

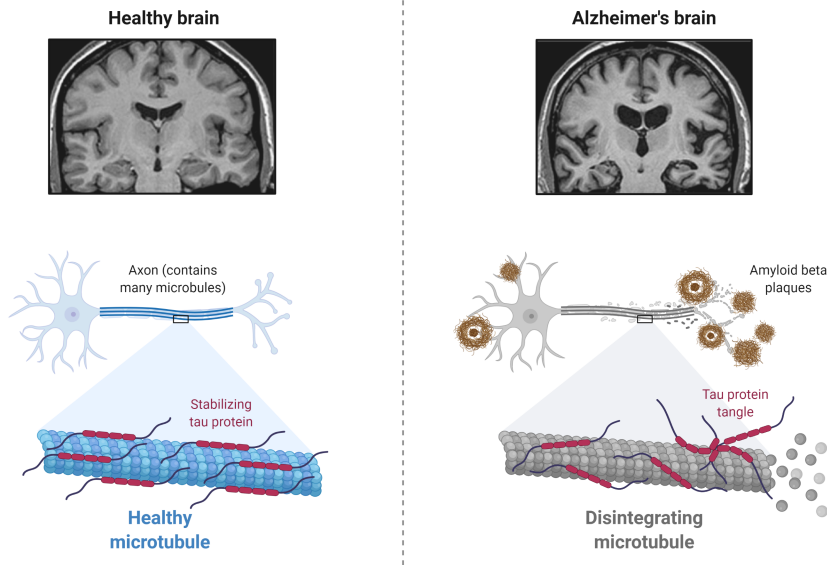


Figure 3.4: Representation of microscopic and macroscopic hallmarks of Alzheimer's disease. Macroscopically, the Alzheimer's brain is characterized by extensive atrophy, which can be observed from MRI scans. Microscopically, accumulation of tau neurofibrillary tangles as well as amyloid beta plaques can be observed. [Reprinted from "Pathology of Alzheimer's Disease 2", by BioRender.com (2022). Retrieved from <https://app.biorender.com/biorender-templates>]

Figure 3.4). However, similar steps in the progression of brain atrophy can be identified from neuroimaging data. In particular, reductions in both the volume of the hippocampus and the thickness of the entorhinal cortex can be observed from brain MRI as early signs of AD (Scheltens et al., 2002). It has been shown that these geometric measures can be used to predict conversion to AD in subjects with mild cognitive impairment (MCI) (Liu et al., 2010). MCI is a condition in which a subject's cognitive impairment is greater than what would be expected considering their age and education level but not significant enough to severely affect daily life and thus be classified as dementia (Irwin et al., 2018). Not all MCI subjects eventually develop dementia. However, they still constitute an important study group since many MCI cases evolve towards AD.

As discussed in the previous section, cortical atrophy is a characteristic

also found in healthy elderly subjects. However, once AD progresses to affecting the brain cortex, the associated GM reductions differ from those typical of healthy aging, which is characterized by a stronger decline in frontal brain regions (Terribilli et al., 2011), while AD tends to progress by affecting more the medial temporal lobe (DeCarli, 2000).

Finally, AD patients have been shown to be affected by a lower WM integrity compared to healthy subjects. Previous studies have reported AD-related alterations in several WM fiber tracts including the CC (Di Paola et al., 2010; Liu et al., 2011), longitudinal fasciculus (Liu et al., 2011), cingulum (Liu et al., 2011; Kiuchi et al., 2009), uncinate fasciculus (Liu et al., 2011; Kiuchi et al., 2009) and fornix (Liu et al., 2011; Oishi et al., 2012).

#### Multiple sclerosis

MS is a chronic autoimmune demyelinating disease. In contrast to dementia, MS does not develop at older ages: its onset is usually between 20 and 40 years of age, and it is the most common cause of nontraumatic neurological disability (Hardiman et al., 2011). In the CNS of MS patients, an inflammatory process damages myelin and in turn the axons, leading to gliosis, which is a proliferation of glial cells in response to damage (Liu et al., 2017). This chronic extensive neuroinflammation ultimately leads to neuronal death, causing atrophy of several structures, which is why MS is also considered to be a neurodegenerative disorder.

In MS patients, it is common to identify WM atrophy, which has also been shown to be directly associated with cognitive impairment (Sacco et al., 2015). In particular, the CC is a WM structure that is greatly affected by MS lesions (Hardiman et al., 2011), and its atrophy has been shown to significantly correlate with two different types of disability measures (Granberg et al., 2015b). One is the Expanded Disability Status Scale (EDSS), which is a score of physical disability ranging from 0 to 10 in 0.5 units: the lower the score, the lower the level of disability (Kurtzke, 1983). A patient with an EDSS below 4.5 is normally able to walk without any aid but may have other types of functional impairments in one or more of the following functional systems: pyramidal (related to muscle weakness), cerebellar (impaired balance and coordination), brainstem (impaired speech and swallowing), sensory, bowel and bladder, visual (impaired sight), cerebral (impaired cognitive functions) or other. When the EDSS score is greater than or equal to 5, problems with walking are also present.



## CHAPTER 3. BACKGROUND

The second disability measure is obtained by performing the Symbol Digit Modalities Test (SDMT), which tests for cognitive disability (Smith, 1973). The test is based on giving a task to the patient that consists of matching numerical digits with a set of predefined abstract symbols. In this case, the score is inversely related to the level of disability: the higher the score, the higher the level of attention and processing speed.

### 3.2 Magnetic resonance imaging

In this section, basic principles of MRI image acquisition and conventional structural MRI sequences are presented. The content is mainly based on the book by Kuperman (2000) and the topical review by Gossuin et al. (2010), in which more technical details about MRI physics—that are outside the scope of this thesis—can be found.

#### 3.2.1 Nuclear magnetic resonance

The imaging technology MRI is based on the physical phenomenon of nuclear magnetic resonance (NMR), which was independently discovered by Bloch (1946) and Purcell et al. (1946). This phenomenon describes the interaction between nuclei (in this specific case, hydrogen nuclei) and an external static magnetic field  $\mathbf{B}_0$ . Such interaction consists of a precession of each nuclear magnetic moment around  $\mathbf{B}_0$  at an angular frequency  $\omega_0$ , which is known as the Larmor frequency. In the case of MRI, it is always in the radio frequency (RF) domain. If  $\mathbf{B}_0$  is applied along the  $z$  direction (i.e., the head-feet direction),  $\omega_0$  is proportional to the strength  $B_0$  of  $\mathbf{B}_0$ :  $\omega_0 = \gamma B_0$ , where  $\gamma$  is a constant called the *gyromagnetic ratio*. In an equilibrium condition, the nuclear magnetic moments are aligned in parallel and antiparallel to the external field, with a slight predominance in the parallel alignment. For this reason, their summation results in a total nonzero macroscopic magnetization  $\mathbf{M}$  (see Figure 3.5A).

The system can be put out of equilibrium if another magnetic field  $\mathbf{B}_1$  rotating at the Larmor frequency is applied perpendicularly to  $\mathbf{B}_0$  (resonance phenomenon). The additional field introduces another rotation of  $\mathbf{M}$  around the  $\mathbf{B}_1$  direction with an angular frequency of  $\omega_1 = \gamma B_1$ . This results in a spiral trajectory of  $\mathbf{M}$ , with the nuclear magnetic moments rotating around both  $\mathbf{B}_0$  and  $\mathbf{B}_1$ . Thus, as soon as  $\mathbf{B}_1$  is applied,  $\mathbf{M}$  is no longer parallel to  $\mathbf{B}_0$  and, at time  $t$ , is tilted at angle  $\theta$ , which is called the

### 3.2. MAGNETIC RESONANCE IMAGING

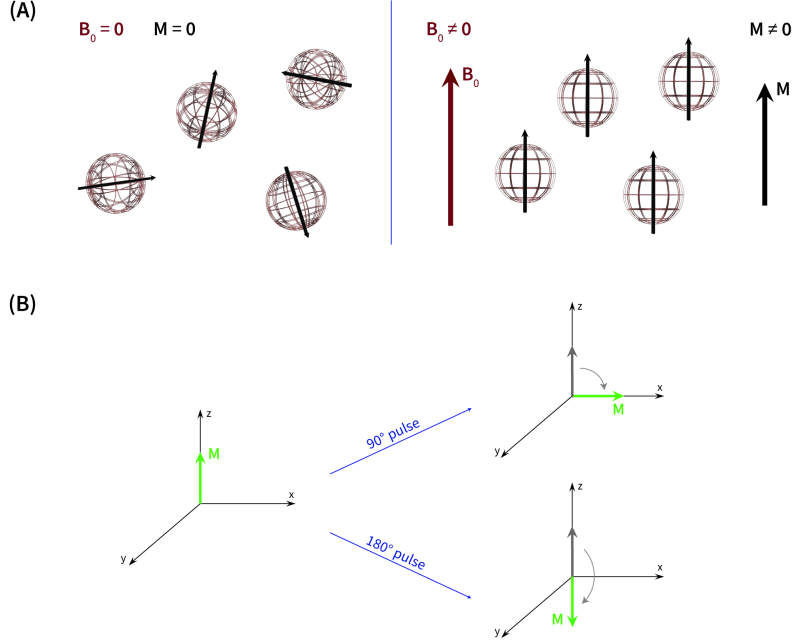


Figure 3.5: (A) Left: in the absence of an external magnetic field, the nuclei are randomly oriented. Right: when a magnetic field  $\mathbf{B}_0$  is applied, the majority of the nuclear moments are aligned in parallel to the field. (B) The effect of a  $90^\circ$  and  $180^\circ$  pulse on the macroscopic magnetization  $\mathbf{M}$  is illustrated.

*flip angle:*

$$\theta = \int_0^t \omega_1(\tau) d\tau = \gamma \int_0^t B_1(\tau) d\tau.$$

Flip angles used in MRI can range between  $0^\circ$  and  $180^\circ$  (see Figure 3.5B). After applying this RF pulse, the system goes back to equilibrium, that is  $\mathbf{M}$  is again aligned to  $\mathbf{B}_0$ .

In the case of a  $90^\circ$  pulse,  $\mathbf{M}$  lies on the  $xy$  plane and rotates around  $\mathbf{B}_0$  at the Larmor frequency. Such a rotation causes magnetic flux variations that induce a voltage in a coil placed on the  $xy$  plane that results in an electric signal oscillating at the Larmor frequency. These oscillations are usually removed by a detection system, obtaining only an exponentially decaying signal called free induction decay (FID). The maximum amplitude of the FID depends on the proton density (related to the amplitude of  $M$ ),

## CHAPTER 3. BACKGROUND

while its time course depends on two time constants referred to as  $T_1$  and  $T_2$  ( $T_1 \geq T_2$ ).  $T_1$  is the longitudinal relaxation time and describes the speed at which the  $z$  component of the magnetization is recovered after the RF pulse.  $T_2$  is the transverse relaxation time and describes the return to equilibrium of the  $x$  and  $y$  components of  $\mathbf{M}$ . If  $M_0$  is the initial amplitude of  $\mathbf{M}$ , the relaxation process can be described by the following equations:

$$M_z(t) = M_0(1 - e^{-\frac{t}{T_1}}) + M_z(0)e^{-\frac{t}{T_1}}, \quad M_{xy}(t) = M_{xy}(0)e^{-\frac{t}{T_2}}.$$

In practice, though, the transverse FID decays more rapidly, with a time constant  $T_2^*$  that is lower than  $T_2$ . This is mainly due to inhomogeneities of  $\mathbf{B}_0$  that cause a dephasing of the proton magnetic moments in the  $xy$  plane after the pulse. This dephasing can be reduced using a particular MRI sequence (i.e., a set of subsequent RF pulses) called *spin echo* that consists of applying, after the  $90^\circ$  pulse, several  $180^\circ$  pulses, one every  $T_E$  (echo time). At each  $T_E$ , the magnetic moments are refocused, and the signal reaches its maximum value. The estimated curve for the transverse relaxation can then be obtained by fitting these maximum signal amplitudes. This sequence is then repeated at every  $T_R$  (repetition time), that is the time between the two  $90^\circ$  pulses. The overall intensity of the measured signal is given by the following formula:

$$S = K \cdot [H] \cdot (1 - e^{-\frac{T_R}{T_1}}) \cdot e^{-\frac{T_E}{T_2}},$$

where  $K$  is a scaling factor, and  $[H]$  indicates the proton density.

### 3.2.2 Image reconstruction

Lauterbur (1973) introduced for the first time a technique for image formation based on the use of gradients to identify the spatial position of a measured MR signal. This method consists of three main steps:

1. **Slice selection.** A magnetic field gradient  $G_z$  is applied along the  $z$  direction such that the Larmor frequency is different in every axial slice:  $\omega(z) = \gamma(B_0 + G_z z) = \omega_0 + \gamma G_z z$ . Thus, it is possible to excite only the protons belonging to the desired slice  $z_p$  by applying an excitation pulse gradient with frequency  $\omega(z_p)$ .
2. **Frequency encoding.** A gradient  $G_x$  is applied along the  $x$  direction and the magnetic resonance moments at the same  $x$  position rotate with a frequency  $\omega_x = \omega_0 + \gamma G_x x$ .

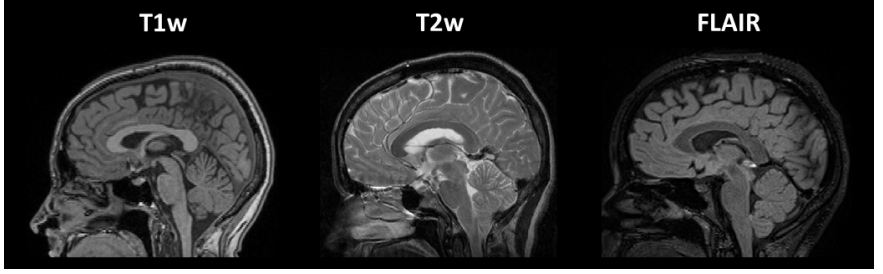


Figure 3.6: Comparison of T1w, T2w and FLAIR MRI scans.

**3. Phase encoding.** A gradient  $G_y$  is applied along the  $y$  direction. After a time interval  $t_y$ , the moments at the same  $y$  position shown a phase given by  $\phi_y = (\omega_0 + \gamma G_y y) \cdot t_y$ .

Each voxel in the slice  $z_p$  is then associated with its pair  $(\omega_x, \phi_y)$ . The information about frequency and phase can be extracted from the detected MR signal, which has been shown to be the Fourier transform of the effective proton density  $\rho$ , that is the true density corrected by all the phenomena that can affect the quality of the measured signal (Ljunggren, 1983; Twieg, 1983). Subsequently, it is possible to reconstruct the image (which is a representation of the proton density) by taking the inverse Fourier transform of the sampled data.

#### 3.2.3 T1- and T2-weighted MRI

The above-described principles can be used to reconstruct images of body tissues by targeting hydrogen nuclei. Since the human body is made of 70% water, and every water molecule contains two hydrogen protons, hydrogen nuclei produce an excellent NMR signal, and different types of soft tissues can be distinguished. For example, in the brain, it is possible to distinguish WM and GM from their different relaxation times. By applying a magnetic field of 3.0 T (common values are from 1.5 T to 3.0 T), Wansapura et al. (1999) obtained average  $T_1$  values of 1331 and 832 ms from GM and WM, respectively, and average  $T_2$  of 80 and 110 ms, respectively. The simple proton density variations between tissues are usually within a few percent. This is why the difference in relaxation times is more often employed to reconstruct detailed brain MRI images.

When imaging different tissues using a spin-echo sequence, the longitudinal magnetization is restored faster and the transverse magnetization

## CHAPTER 3. BACKGROUND

more slowly in tissues having a shorter  $T_1$ . Tissues with short  $T_1$  and long  $T_2$  result in being bright in so-called T1-weighted (T1w) images. This dominant  $T_1$  contrast is obtained by setting both a very short  $T_E$  (i.e., lower than the shortest  $T_2$  of the tissues of interest) and a short  $T_R$  ( $T_R \approx T_1$ ).

In a similar manner, images with predominant  $T_2$  contrast can also be produced by using the differences in  $T_2$  relaxation times between different tissues. T2-weighted (T2w) images can be obtained by using a  $T_E$  that is equal to or longer than the shortest  $T_2$  of the tissues of interest and a long  $T_R$  ( $T_R > T_1$ ).

When employing a T2w MRI sequence, the signal from the brain's GM and WM decays much more rapidly compared to that of the CSF. Therefore, in some cases, the use of a T2w fluid attenuated inversion recovery sequence—the FLAIR sequence—is preferred as it allows suppressing the signal from the CSF (Hajnal et al., 1992). This is done by using, first of all, very long  $T_R$  and  $T_E$ , thus obtaining a very heavily T2-weighted scan. Next, a so-called *inversion recovery* sequence is applied; it consists in applying an initial  $180^\circ$  pulse followed by a series of slice-selective  $90^\circ$  pulses (applied after a time interval  $T_i$ , i.e., *inversion time*) and spin echo acquisitions (Hajnal et al., 1992; Valli and Coppini, 2005). Therefore, after initially inverting the magnetization with the  $180^\circ$  pulse, an appropriate choice of  $T_i$  can null the signal from a specific tissue of interest such as the CSF. Thanks to the combination of high T2-weighting with the reduction of artefacts normally caused by a high CSF signal, FLAIR images produce high contrast corresponding to brain lesions, especially in regions close to the CSF. For this reason, this sequence is particularly useful for detecting MS lesions that result in hyperintensities (Bakshi et al., 2001).

A comparison between the contrasts of a T1w, T2w and FLAIR brain MRI image is shown in Figure 3.6.

### 3.2.4 Diffusion MRI

While T1w and T2w imaging constitute the most common techniques for imaging the general structure of the brain, diffusion MRI (dMRI) represents an extremely useful method for imaging WM connectivity. The basic concept behind dMRI is that MR images are sensitive to the random Brownian motion of water molecules in the direction of the applied magnetic field gradients (Jellison et al., 2004). The diffusion rate of the molecules can be described by the equation  $\langle r^2 \rangle = 6Dt$ , where  $\langle r^2 \rangle$  is the mean squared molecular displacement,  $t$  is the diffusion time, and  $D$  is the diffu-

### 3.2. MAGNETIC RESONANCE IMAGING

sion constant, which is typically expressed in  $\text{mm}^2/\text{s}$ . The higher  $D$  is, the more mobile are the water molecules. In biological tissues, such diffusion is impeded and deviated by cellular structures, resulting in anisotropic diffusion. This is particularly evident in the WM, while GM and CSF show more isotropic diffusion (Mukherjee et al., 2008).

#### The diffusion-weighted signal

In dMRI experiments, the standard gradient pulse sequence is the Stejskal-Tanner sequence (Stejskal and Tanner, 1965), which consists of adding two magnetic gradients, one each for a time  $\delta$ , to a spin-echo sequence: the first gradient is applied before the  $180^\circ$  RF excitation, and the second is applied immediately after. The two pulse gradients are separated by a time interval  $\Delta$ . After the first pulse, there is a phase shift that depends on the position of the nuclei in the direction of that gradient. The  $180^\circ$  excitation inverts this shift, and then the second gradient introduces an additional dephasing that is equal to that of the first pulse. If the nuclei do not move along the direction of the gradients during  $\Delta$ , the second gradient undoes the effect of the first, bringing the magnetic moments in phase again. If instead there is diffusion in some voxels, the changes in position will be reflected in an attenuation of the MR signal in those voxels.

The diffusion-weighted signal can be described by the following equation:

$$S_i = S_0 \cdot e^{-b \cdot \text{ADC}_i}.$$

$S_i$  is the signal intensity when the pulse gradients are applied along the direction  $i$ , while  $S_0$  is the signal intensity measured without any gradient.  $\text{ADC}_i$  (i.e., the apparent diffusion coefficient) is the diffusion constant measured in clinical settings; it includes not only the actual diffusion constant  $D$  but also any other source of water mobility. Regarding the parameter  $b$ , it is called b-value and is given by

$$b = \gamma^2 G^2 \delta^2 \left( \Delta - \frac{\delta}{3} \right),$$

where  $\gamma$  is the gyromagnetic ratio, and  $G$  is the amplitude of the gradients. By acquiring an image with a b-value of  $0 \text{ s/mm}^2$ , we can derive the signal intensity  $S_0$ . The ADC values can then be obtained as

$$\text{ADC}_i = -\frac{\ln\left(\frac{S_i}{S_0}\right)}{b}.$$

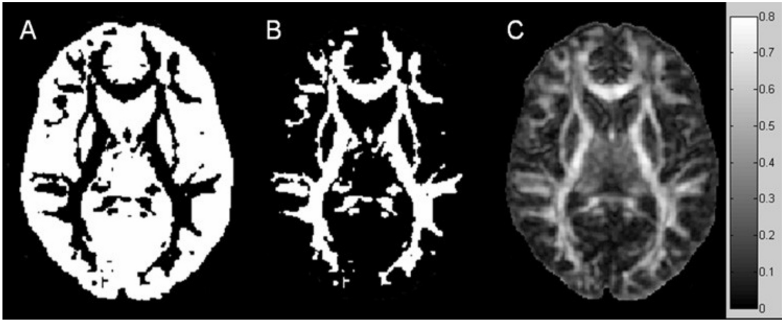


Figure 3.7: Comparison of (A) binary mask of low FA ( $0 \leq \text{FA} \leq 0.3$ ), (B) binary mask of high FA ( $\text{FA} > 0.3$ ) and (C) the original FA map on one slice of one subject. [© Giannelli et al. (2010), reprinted with permission]

Once the ADC values are derived for each voxel, it is possible to represent the diffusion information in a diffusion-weighted image.

### Diffusion tensor imaging

Many of the WM fiber tracts are organized in macroscopic structures that are coherently oriented in the direction of the fibers. The diffusion of water molecules is much faster along WM fibers than perpendicularly to them, so dMRI can provide voxel-wise information on the orientation of these large groups of neuronal axons (Alexander, 2009). This information can be extracted using different techniques, of which the most commonly used is diffusion tensor imaging (DTI) because of its simple model and good scanning efficiency.

DTI infers a 3D description of the direction of diffusion of water molecules within a voxel by representing the diffusion as a  $3 \times 3$  symmetric positive semidefinite matrix  $\mathbf{D}$  (Chowdhury et al., 2014). This matrix is also called the *diffusion tensor* (DT):

$$\mathbf{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix}.$$

Since six out of nine values of the matrix are independent, at least six different gradient directions are required to derive the DT. If the number of gradient directions is increased, the estimate will be more accurate but at

the cost of increased scanning time (Zhang et al., 2009). The DT is usually represented by a 3D ellipsoid, which easily describes the anisotropic nature of water diffusion in the WM.

The principal eigenvector of the DT that corresponds to its highest eigenvalue is assumed to be aligned along the local fiber direction. The degree of directionality of the diffusivity within each voxel can be measured with an index called fractional anisotropy (FA), which ranges from 0 (random motion) to 1 (unidirectional diffusion) (Mukherjee et al., 2008):

$$\text{FA} = \frac{\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}}{\sqrt{2} \cdot \sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}},$$

where  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$  are the eigenvalues of the DT. These eigenvalues, together with the respective eigenvectors, describe the directions and lengths of the axes of the diffusion ellipsoid. The FA values of each voxel can be represented in 3D images referred to as FA maps (see Figure 3.7).

## 3.3 Medical image processing

Medical image processing research focuses on the development of methods for enhancing raw medical image data in order to analyze and visualize them according to specific clinical or research needs. Furthermore, medical image processing comprises multiple research topics ranging from very problem-specific applications to more general techniques that can be applied in different contexts (Zhu, 2003). In this section, two of the main research subfields of medical image processing that are also relevant for the present doctoral thesis are presented: image registration and segmentation. What is presented in this section is mainly based on the *Handbook of Medical Imaging* by Beutel et al. (2000) as well as two chapters of the *Handbook of Medical Image Processing and Analysis* by Woods (2009) and Bankman (2009). It must be noted that the described methods constitute traditional approaches that are currently often replaced by machine learning (ML)-based—in particular, deep learning (DL)-based—methods, especially for what concerns image segmentation. These latter approaches are addressed in Section 3.4.

### 3.3.1 Image registration

Image registration is the process of aligning one image (usually referred to as the *moving image*) to another (*fixed image*). This is done by applying



a specific geometrical transformation to the moving image so that it can match as closely as possible the fixed image according to a given criterion.

### Linear transformation models

The type of transformation applied depends on the choice of a spatial transformation model. The simplest, but also the most constrained, is the *rigid-body* model, which consists in applying only global rotations and translations to the image. In 3D images, this model can be defined by specifying six independent parameters: three for the translations (along the  $x$ ,  $y$  and  $z$  axes) and three for the rotations (around the  $x$ ,  $y$  and  $z$  axes). This type of transformation can be used for registering body parts that are subject to deformations only in a limited manner, for example bones.

Another type of model is the *global rescaling* transformation, which adds a seventh parameter to the rigid-body transformation: a magnification factor. This means that the moving image is subject not only to rotations and translations but also to scaling. A slightly more complex alternative consists in performing an anisotropic rescaling rather than using the same scaling factor for all three axes. This type of model is usually referred to as a *nine-parameter affine* model because of the number of parameters required to describe it.

The most general type of linear transformation model is the *affine* model, which, in three dimensions, is described by a total of 12 independent parameters. It imposes a geometric constraint according to which lines that are parallel in the original image will still be parallel after applying the transformation. This type of geometrical transformation can be described as

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} e_{11} & e_{12} & e_{13} & e_{14} \\ e_{21} & e_{22} & e_{23} & e_{24} \\ e_{31} & e_{32} & e_{33} & e_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}.$$

This notation allows the representation in a  $4 \times 4$  form of all the transformations that have been applied to the image: translation, rotation, scaling and shearing.

### Nonlinear transformation models

In brain imaging as well as other medical imaging subfields, the structures to be aligned are often so complex and heterogeneous that the use of affine

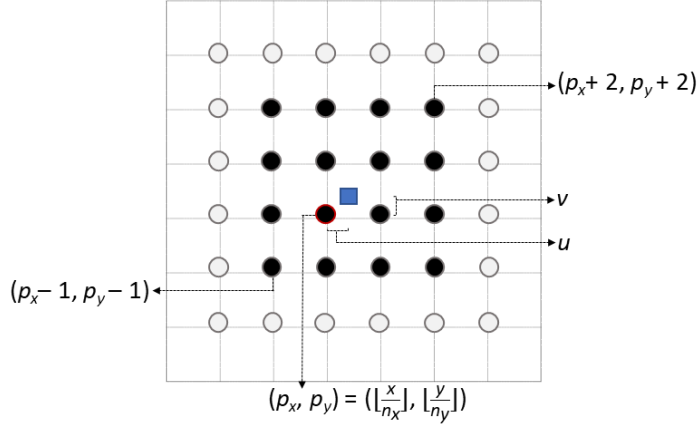


Figure 3.8: 2D representation of the neighborhood of 16 control points (black circles) employed for the deformation of a pixel/voxel (blue square) using the FFD algorithm. The white circles correspond to the points that are not used since they do not belong to the voxel's neighborhood.  $(p_x, p_y)$ , which is outlined in red, is the reference grid point that is closest to the voxel. The control points used in the algorithm range from the index  $(p_x - 1, p_y - 1)$  to the index  $(p_x + 2, p_y + 2)$ .

models is not sufficient to obtain a good registration. This is why warping algorithms need to be applied to perform different deformations that are no longer global but depend on the location in the image. In the literature, a large variety of warping algorithms have been proposed, but they are mainly divided into two groups: parametric and nonparametric algorithms (Dwivedi Chenna et al., 2018). The first are based on the definition of a grid of control points that correspond to the parts of the image that are actually employed by the deformation algorithm. In contrast, nonparametric methods associate a displacement vector to every single voxel in the moving image.

Some of the most commonly used parametric methods are those based on B-splines, including the so-called *free-form deformation* (FFD) algorithm (Rueckert et al., 1999). Given an image volume  $\Omega = \{(x, y, z) \mid 0 \leq x < X, 0 \leq y < Y, 0 \leq z < Z\}$  and an  $n_x \times n_y \times n_z$  mesh of control points  $\varphi_{i,j,k}$  with uniform spacing  $\delta$ , the displacement of every voxel  $(x, y, z)$  is determined by the displacement of a neighborhood of control points around the voxel (see Figure 3.8). This displacement can be defined as the following tensor product using a neighborhood of 64 control points (or 16 control

### CHAPTER 3. BACKGROUND

points in the 2D case, as depicted in Figure 3.8):

$$\mathbf{T}_{\text{local}}(x, y, z) = \sum_{l=0}^3 \sum_{m=0}^3 \sum_{n=0}^3 B_l(u) B_m(v) B_n(w) \varphi_{i+l, j+m, k+n},$$

where  $i = \lfloor \frac{x}{n_x} \rfloor - 1$ ,  $j = \lfloor \frac{y}{n_y} \rfloor - 1$ ,  $k = \lfloor \frac{z}{n_z} \rfloor - 1$ ,  $u = \frac{x}{n_x} - \lfloor \frac{x}{n_x} \rfloor$ ,  $v = \frac{y}{n_y} - \lfloor \frac{y}{n_y} \rfloor$ ,  $w = \frac{z}{n_z} - \lfloor \frac{z}{n_z} \rfloor$ .  $B_l$  (as well as  $B_m$  and  $B_n$ ) is the  $l^{\text{th}}$  basis function of the cubic B-spline, which is defined as

$$\begin{aligned} B_0(u) &= \frac{(1-u)^3}{6}, \\ B_1(u) &= \frac{(3u^3 - 6u^2 + 4)}{6}, \\ B_2(u) &= \frac{(-3u^3 + 3u^2 + 3u + 1)}{6}, \\ B_3(u) &= \frac{u^3}{6}. \end{aligned}$$

The control points  $\varphi_{i,j,k}$  constitute the parameters of the algorithm; by increasing the resolution of the mesh, the degrees of freedom of the deformation also increase. In medical imaging, deformations should usually be smooth in order to achieve a realistic and well-proportioned image after applying the transformations. Therefore, a regularization term is usually integrated and defined as

$$\begin{aligned} C_{\text{smooth}} = \frac{1}{V} \int_0^X \int_0^Y \int_0^Z \left[ \left( \frac{\partial^2 \mathbf{T}}{\partial x^2} \right)^2 + \left( \frac{\partial^2 \mathbf{T}}{\partial y^2} \right)^2 + \left( \frac{\partial^2 \mathbf{T}}{\partial z^2} \right)^2 + \right. \\ \left. + 2 \left( \frac{\partial^2 \mathbf{T}}{\partial xy} \right)^2 + 2 \left( \frac{\partial^2 \mathbf{T}}{\partial xz} \right)^2 + 2 \left( \frac{\partial^2 \mathbf{T}}{\partial yz} \right)^2 \right] dx dy dz, \end{aligned}$$

where  $V$  indicates the image volume.

A method like the above-described FFD algorithm has the limitation of assuming that both moving and fixed images present very similar structures. However, if this is not the case, the topology of the transformed image may change uncontrollably, leading to results that are difficult to interpret from an anatomical perspective (Avants et al., 2011a). For this reason, registration methods that make use of diffeomorphic transformations (i.e., smooth and invertible) have been introduced.

One example is the symmetric normalization (SyN) algorithm proposed

by Avants et al. (2008). In this work, a diffeomorphism  $\phi$  is defined in the image domain  $\Omega$  and maps an image  $I$  into a new space  $\phi I = I(\phi(\mathbf{x}, t = 1))$ , where  $t$  is a time parameter (which can be seen as the image registration optimization time,  $t \in [0, 1]$ ), and  $\mathbf{x}$  is a spatial coordinate. A velocity field  $\mathbf{v}(\mathbf{x}, t)$ , which is square integrable, can be defined on  $\Omega$ . The integration of the velocity field in time can be used to generate the diffeomorphism:  $\phi(\mathbf{x}, 1) = \phi(\mathbf{x}, 0) + \int_0^1 \mathbf{v}(\phi(\mathbf{x}, t), t) dt$ , where  $\phi(\mathbf{x}, 0)$  is the identity. Additionally, the fact that a diffeomorphism  $\phi$  can be decomposed in two parts ( $\phi_1$  and  $\phi_2$ ) that have the same value in the middle of the registration can be exploited. In other words, if the registration defines a transformation from image  $I$  to image  $J$ , then the same diffeomorphic path would be computed from  $J$  to  $I$  (no matter which one is the moving and which one is the reference image). Two new velocities fields  $\mathbf{v}_1$  and  $\mathbf{v}_2$  can then be defined so that  $\mathbf{v}(\mathbf{x}, t) = \mathbf{v}_1(\mathbf{x}, t)$  for  $t \in [0, 0.5]$  and  $\mathbf{v}(\mathbf{x}, t) = \mathbf{v}_2(\mathbf{x}, t)$  for  $t \in [0.5, 1]$ . Therefore, it is possible to split the above-mentioned integral into the two different components of the velocity field, allowing the optimization algorithm to deform both  $I$  and  $J$  along  $\phi$  in such a way that they will always meet at a fixed point midway. This optimization process with respect to both  $\phi_1$  and  $\phi_2$ —the details of which are outside the scope of the present thesis but can be found in the paper by Avants et al. (2008)—provide the symmetric solution that is characteristic of the SyN algorithm.

### Similarity metrics

In addition to defining a transformation model, a similarity metric also needs to be set. The role of the registration algorithm consists in maximizing the similarity between the moving image and the fixed image according to the selected similarity metric.

If two images are acquired using the same imaging modality, they should have similar brightness and contrast, and thus—after registration—the intensity values in corresponding voxels should be (nearly) equal. When this is the case, the simplest similarity metric to be applied is the least square difference, which should reach its minimum value when the images are aligned.

Another more advanced and commonly used metric is cross-correlation, which consists in first multiplying the image intensities at each voxel and summing these products. The obtained value is then divided by the product of the root mean squared intensities of each image. The maximum value of the cross-correlation is 1, which is achieved when two images are identical. The use of cross-correlation assumes the existence of a linear relationship

### CHAPTER 3. BACKGROUND

between the intensities of the two images. This assumption is less strict than that of the least square difference, which assumes an identity relation between the two aligned images.

When dealing with intermodality registration, the intensity values of corresponding voxels of aligned images can be very different from one another. Therefore, metrics such as the least square difference will most likely fail in these cases. The assumption of a linear relationship (cross-correlation) between the two images is also often untrue for multimodality problems. This is why other metrics are usually employed that are based on information theory: registration is seen as a process for reducing the amount of information present in the two images combined. The most common way of describing this amount information is the *joint entropy*:

$$H(A, B) = - \sum_{a \in A} \sum_{b \in B} p_{AB}(a, b) \cdot \log_2(p_{AB}(a, b)),$$

where  $A$  and  $B$  are the two images, while  $p_{AB}$  is their joint probability distribution. Such distribution is obtained by deriving the two images' joint histogram and dividing it by the total number of voxels in the images. The value of  $H(A, B)$  can be minimized for performing image registration, although a more reliable and commonly used metric is the *mutual information*, which consists in not only considering the joint information but also the information contributed by each image to the overlapping volume:

$$\begin{aligned} MI(A, B) &= H(A) + H(B) - H(A, B) = \\ &= \sum_{a \in A} \sum_{b \in B} p_{AB}(a, b) \cdot \log_2 \left( \frac{p_{AB}(a, b)}{p_A(a) \cdot p_B(b)} \right), \end{aligned}$$

where  $H(A) = - \sum_{a \in A} p_A(a) \cdot \log_2 p_A(a)$  and  $H(B) = - \sum_{b \in B} p_B(b) \cdot \log_2 p_B(b)$ .

During the image registration process, each transformation step results in different degrees of overlap between the two images, and the mutual information metric is highly dependent on this overlap. However, intermodality images often differ in terms of both spatial resolution and field of view. This is, for example, often seen when comparing computed tomography (CT) with MRI data: CT images usually contain a smaller volume in order to limit the amount of radiation. Therefore, the overlapping volume between the two images can be much lower than the actual volume represented in either modality. For this reason, another similarity metric that

is more overlap-independent can be used. This is the *normalized mutual information*, which can be defined as

$$NMI(A, B) = \frac{H(A) + H(B)}{H(A, B)}.$$

### 3.3.2 Image segmentation

Image segmentation consists in separating an image into regions of interest (ROI) that can be analyzed for a specific task. In brain MRI, this could involve, for example, the calculation of the volumes of the three main tissue types (GM, WM and CSF) or the labeling of deeper structures—such as the hippocampus or the amygdala—to study their volume and shape.

#### Segmentation methods

*Gray-level thresholding* is the simplest segmentation method. It consists in assigning a certain label to each voxel having an intensity value higher than a given threshold. With just one threshold, the image can simply be segmented into two regions (i.e., foreground and background). If more regions need to be segmented, the number of threshold values can increase accordingly.

A more complex segmentation method that takes into consideration the similarity (in terms of intensity) of neighboring voxels is *region growing*. It starts by manually defining different seeds that are known to belong to the ROI. Then, nearby voxels are automatically added to the segmentation mask if they are similar enough to the seeds according to a so-called homogeneity criterion. Such a criterion could, for example, consist in the comparison of the difference between voxel intensity and the mean intensity over a region. However, many different criteria exist in literature, and choosing them appropriately is fundamental to avoid leaking into contiguous areas of the image.

The initialization of the segmentation by defining seed regions is also common to other algorithms, such as *watershed* segmentation. With this method, once the seeds are specified, they are grown by applying a morphological watershed transformation. An intuitive way of describing this transformation consists in seeing the bright voxels of the image as mountaintops and the darker regions as valleys. This “landscape” is punctured at some points, and from each puncture, water comes out, but water from different punctures is not allowed to mix. Therefore, once valleys start to be

### CHAPTER 3. BACKGROUND

submerged, dams need to be added to separate water coming from different punctures. These dams constitute the boundaries of the segmented ROIs, while the submerged valleys represent the ROIs themselves. Edge operators are often used in combination with watershed to generate images with high intensities on the edges and low intensities inside the ROIs to be segmented.

Edge detection operators are highly employed in another class of segmentation approaches called *edge based*. These consist in describing the image in terms of the edges existing between different regions, which can be defined by the local voxel intensity gradient. Edges can be identified by applying a wide range of possible edge operators, the most popular of which include the Sobel, Canny, Laplacian-of-Gaussian and Prewitt operators. Once the edges are detected, several region boundary candidates can be extracted from the image. However, the edge detection process is often quite sensitive to noise, so several pre and postprocessing steps are required, such as preliminary image smoothing using a Gaussian filter or semi-automatic edge linking to identify relevant closed boundaries.

Segmentation has always constituted a challenging task due to several issues including poor contrast and spatial resolution as well as noise and artifacts in the image. For this reason, many segmentation methods do not rely only on gray-level information. Such approaches include, for example, *level set* methods. The idea behind these consists in representing the contours of a ROI as a propagating wavefront. Such propagation takes place along the contour's normal direction, and its speed is controlled by a function  $\varphi$  that incorporates both surface features (e.g., normal direction and curvature) and image features (e.g., intensity and gradient). The boundary of the ROI to be segmented is the zero level set of  $\varphi$ , while the inside of the ROI is identified by all points in which  $\varphi$  is positive. The level set function  $\varphi$  satisfies the following equation:

$$\frac{\partial \varphi}{\partial t} = v |\nabla \varphi|,$$

where  $v$  is the speed,  $t$  is time, and  $|\cdot|$  denotes the Euclidean norm. The solution of the above equation provides the segmentation result of interest.

Additional a priori information on the ROI to be segmented can be integrated into the above-described model. One example that is particularly relevant for the present thesis is the incorporation of *statistical shape models* into the level set segmentation as originally proposed by Leventon et al. (2000). In this study, a training dataset of  $N$  manual segmentations of

the structures of interest is selected, and a shape model is built over the distribution of these  $N$  surfaces. This is done by first computing a signed distance map for each surface that has negative values inside, positive values outside, and zero in correspondence to the zero level set. Afterwards, the mean surface  $\mu$  is computed, and principal component analysis (PCA) is performed to extract the first  $k$  orthogonal models of shape variation, which can be represented as columns of a matrix  $U_k$ . Any novel shape  $u$  can therefore be approximated with an estimate  $\tilde{u}$  by computing a  $k$ -dimensional vector of coefficients  $\alpha$ :

$$\tilde{u} = U_k \alpha + \mu.$$

We can then assume that all shapes have a Gaussian distribution. Therefore, the probability of a certain shape, represented by its  $\alpha$  vector, can be estimated as:

$$P(\alpha) = \frac{1}{\sqrt{(2\pi)^k |\Sigma_k|}} e^{-\frac{1}{2} \alpha^T \Sigma_k^{-1} \alpha}.$$

Once a curve representation  $\alpha$  and its probability distribution are obtained, it is possible to incorporate shape information into the level set segmentation process. If  $u$  is the shape at a certain step of the curve evolution, and  $\nabla I$  is the image gradient, the respective shape parameters  $\alpha$  as well as the pose  $p$  of the evolving curve with respect to the shape model can be estimated by applying a maximum a posteriori (MAP) approach:

$$\alpha_{\text{MAP}}, p_{\text{MAP}} = \operatorname{argmax}_{\alpha, p} P(\alpha, p \mid u, \nabla I).$$

Since no prior information on the exact location of the object is retained, the pose parameters are assumed to have a uniform distribution. After estimating the appropriate values of  $\alpha$  and  $p$ , a final evolving surface  $u^*(t)$  can be extracted starting from the surface  $u(t)$  at time  $t$ . Finally, the level set update expression including shape information can be expressed as:

$$u(t+1) = \lambda_1 (g(c + \kappa) |\nabla u(t)| + \nabla u(t) \cdot \nabla g) + \lambda_2 (u^*(t) - u(t)) + u(t).$$

The first term refers to the typical level set expression, where  $g$  is a function of the image gradient,  $c$  is a balloon force that makes the contour flow outward, and  $\kappa$  is the curvature. By contrast, the second term refers to the MAP estimation discussed above. By choosing appropriate  $\lambda_1$  and  $\lambda_2$ , it is possible to balance the influence of the gradient-curvature model and the shape model (see Figure 3.9).



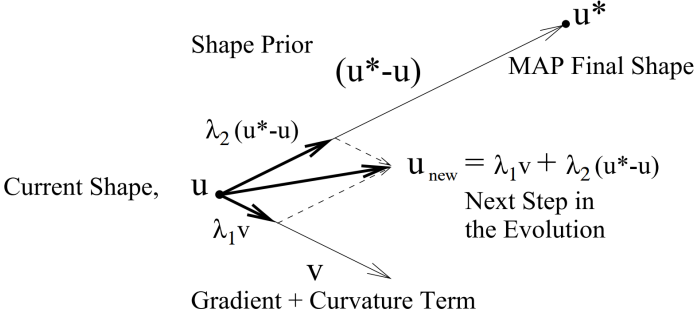


Figure 3.9: Schematic representation of the contribution of the gradient-curvature model (weighed by the parameter  $\lambda_1$ ) and the statistical shape model (weighed by  $\lambda_2$ ) in the level set curve evolution (from  $u$  to  $u_{\text{new}}$ ). [From Leventon et al. (2002), © 2002 IEEE, reprinted with permission]

Another way of incorporating a priori knowledge into the segmentation process is based on applying *atlas-based* methods. First, a linear registration step aligns a reference atlas volume to the image to be segmented. Next, nonlinear transformations are applied to deform the atlas to achieve local correspondence with the target image. The atlas is originally provided together with a corresponding segmentation mask that has been performed—or at least checked and corrected—manually. Therefore, the same transformation that was applied to the atlas volume can be applied to the mask in order to obtain a new segmentation mask for the target image. Atlas-based segmentation is a successful method that is widely applied for brain segmentation, for example to parcellate the brain into several subcortical and cortical structures. Sometimes, such methods are further extended by using more robust multi-atlas approaches that consist in registering different atlases (and thus segmentation masks) to the target image and then obtaining the final segmentation by merging the results together (e.g., through majority voting).

### Evaluation metrics

When developing a semi or fully automatic segmentation algorithm, it is important to investigate its performance by using standardized evaluation metrics. These metrics are usually based on a comparison of the algorithm's output segmentation with one (or more) reference manual segmentation(s) performed by expert annotators. Such reference segmentations are usually

referred to as *ground truth*.

One of the most commonly used evaluation metrics is the *Dice coefficient* (Dice, 1945), which measures the degree of overlap between a predicted segmentation mask and the ground truth segmentation. Given two segmentations  $A$  and  $B$ , the Dice coefficient is defined as

$$\text{Dice}(A, B) = \frac{2|A \cap B|}{|A| + |B|},$$

where the operator  $|\cdot|$  refers to the number of nonzero elements in the given segmentation. Alternatively, by defining as  $FP$  the number of false positives in the predicted segmentation (i.e., non-zero elements that are instead marked as background in the ground truth), as  $FN$  the number of false negatives (zero elements that are labeled as part of the ROI in the ground truth) and as  $TP$  the number of true positives (nonzero elements in both the prediction and the ground truth), the Dice coefficient can be computed as

$$\text{Dice}(A, B) = \frac{2TP}{2TP + FP + FN}.$$

Using the same notation, two other important evaluation metrics can be defined: *precision* (often referred to as *positive predictive value*) and *recall* (or *sensitivity*). These provide additional information on the proportion of, respectively, false positives and false negatives in the predicted segmentation masks:

$$\text{precision}(A, B) = \frac{TP}{TP + FP}, \quad \text{recall}(A, B) = \frac{TP}{TP + FN}.$$

Useful evaluation metrics also exist for estimating the distance in the image space between the predicted and the ground truth masks. One relevant example is the Hausdorff distance  $H(A, B)$  (Huttenlocher et al., 1993), which is defined as

$$H(A, B) = \max(d(A, B), d(B, A)),$$

where  $d(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|$ , and  $\|\cdot\|$  usually refers to either the  $L_2$  or the Euclidean norm.

### 3.3.3 Other relevant tools for brain MRI image processing

The above-described registration and segmentation methods are often only the first steps of more extensive image analyses that allow the investigation of brain morphology in more detail. In this section, a few more methods as well as software tools for performing these analyses are presented. All methods presented here are relevant for the present doctoral thesis.

#### Voxel-based morphometry

The principle of nonlinear registration can be applied not only to simply align different 3D volumes but also to analyze local brain differences across different subject groups. This is the case with the neuroimage processing method called *voxel-based morphometry* (VBM), which aims to compare local GM volume between two subject groups (Ashburner and Friston, 2000). This is done by first segmenting the GM of each subject from structural MRI images. Subsequently, these GM masks are affine registered to a pre-defined GM template. A further nonlinear registration step is added to identify for each subject which local deformations need to be performed in order to match the reference template. Such deformations consist in local expansions if a subject has a GM concentration that is locally lower than that of the template or contractions if the GM concentration of the subject is higher than that of the template. All information on local expansions and contractions can be stored in modulated GM volumes, which are obtained by multiplying all the registered GM segmentations by the Jacobian of the deformation field. These modulated images are often then blurred with a Gaussian kernel in order to both compensate for any inaccuracies in the registration and make the data more normally distributed so that they are valid for parametric statistical testing (the residuals after fitting a general linear model should be both independent and normally distributed). The final step consists in the use of generalized linear models to compare voxel-wise local deformations between the two groups of interest.

#### Tract-based spatial statistics

Voxel-wise statistics for identifying group differences can be applied in other neuroimaging contexts, including dMRI analysis. This is the case for a technique called *tract-based spatial statistics* (TBSS), which aims to identify local differences in FA between subject groups (Smith et al., 2006). FA maps are first extracted from every subject's dMRI scan and aligned to a reference

FA template by applying nonlinear registration. Next, a mean FA image is generated by averaging all the aligned images, and then it is thinned to identify a so-called FA skeleton, which represents the centers of the main WM bundles in the average brain. Finally, each subject's FA is projected onto the skeleton by assigning to each skeleton voxel the maximum FA value that is found in that subject by searching perpendicularly to the skeleton voxel. Voxel-wise statistics can then be carried out on the FA values in the skeleton space.

#### Brain MRI image processing software

Both VBM and TBSS for human brain image analysis have been implemented in one of the most commonly used software packages for neuroimage processing: FSL (Jenkinson et al., 2012). Another useful tool implemented in FSL is FAST (FMRIB's Automated Segmentation Tool), which corrects for possible field inhomogeneities that affect image intensity and segments a 3D brain image into its three main tissue types (GM, WM and CSF). Each tissue parcellation is provided both as a hard binary segmentation (i.e., every voxel is classified into only one tissue class) and as a tissue probability map (i.e., every voxel is associated with a value from 0 to 1 corresponding to the probability of that voxel belonging to the specified tissue).

When it comes to brain parcellation into different cortical and sub-cortical regions, one of the most successful and used software packages is FreeSurfer (Fischl, 2012). The full FreeSurfer segmentation pipeline starts by constructing surface models corresponding to the boundary between WM and GM as well as the pial surface (i.e., the outer GM layer). This is followed by volumetric cortical and subcortical parcellations, which are based on both a subject-independent probabilistic atlas (which is built from a set of manually labeled images) and subject-specific measured values (including voxel intensity and curvature).

Another popular software package for structural brain MRI image analysis is ANTs (<http://stnava.github.io/ANTs/>), which is based on the medical image processing library Insight ToolKit (ITK, <https://itk.org/>). Similarly to FSL's FAST, ANTs can be used for probabilistic tissue segmentation by applying the *Atropos* tool (Avants et al., 2011b), which combines the use of prior reference segmentations with finite mixture modeling. ANTs also provides a reliable tool for image registration, the SyN algorithm (Avants et al., 2008), which is based on diffeomorphic mapping (as described in Section 3.3.1).

Within the context of brain MRI image registration, another robust alternative is the NiftyReg open-source software (<http://cmictig.cs.ucl.ac.uk/wiki/index.php/NiftyReg>). It provides an implementation of a fast FFD algorithm for nonlinear registration (Modat et al., 2010) that constitutes a refactoring of the FFD method described in Section 3.3.1.

### 3.4 Machine learning in medical image processing

ML is a subset of artificial intelligence consisting in the development of algorithms that modify computer actions so the system can improve with experience, that is it can achieve better accuracy when performing a specific task of interest (Alzubi et al., 2018). Depending on the type of task to be performed, different types of ML algorithms exist, and they can be divided into the following categories:

- **Classification.** The output of the algorithm can only be one of a predefined number of classes, which are known a priori.
- **Anomaly detection.** The algorithm aims to analyze specific patterns in the input data and detect any change or anomaly in such patterns. Therefore, this type of task is more focused on identifying outliers in the input data.
- **Regression.** Here, the output is numeric and continuous, and therefore these algorithms aim to answer questions such as “How many?” or “How much?”.
- **Clustering.** It aims to identify relevant structures in the input data and use the similarity between these structures in order to group data into different clusters. Data that are grouped into the same cluster are meant to have a high level of similarity, while data belonging to different clusters should show lower similarity.
- **Reinforcement learning.** Here, past learning experience is constantly used by the algorithm to produce an output. The algorithm makes use of an agent that is in contact with a constantly changing environment and learns a behavior using trial and error.

In the papers appended to this thesis, supervised learning-based methods for classification and regression are applied. For these problems, given an input vector  $\mathbf{x}$  ( $\mathbf{x} \in \mathbb{R}^n$ , where  $n$  is the number of input features), the

### 3.4. MACHINE LEARNING IN MEDICAL IMAGE PROCESSING

algorithm should predict an output variable  $y = f(\mathbf{x})$ , where  $y$  is a scalar for single-class and regression problems. The function  $f$  is then identified using a set of training data for which all  $\mathbf{x}$  and their respective outputs  $y$  are already known (Wernick et al., 2010). This trained model is then tested on new unseen data in order to validate its performance. Often, such performance validation is carried out using  $k$ -fold cross-validation, which consists in training the model  $k$  times: at each iteration,  $\frac{1}{k}$  of the available (and already labeled) data is used as validation set (i.e., to check the performance of the algorithm on these new unseen data), while the remaining samples constitute the training set.

In the section below, more detailed examples of traditional ML algorithms of relevance to the present doctoral thesis are outlined. Moreover, in the following sections, DL-based methods are presented. Finally, some relevant examples of applications of ML in brain MRI image processing are described in Section 3.4.6.

#### 3.4.1 An overview of relevant models

##### Neural networks

Learning with neural networks (NNs) is often seen as the archetypal ML method (Erickson et al., 2017). As schematized in Figure 3.10, in an NN, all features of the input vector  $\mathbf{x}$  are first multiplied by a weight vector  $\mathbf{w}$ , whose number of features is equal to the number of nodes in Layer 1 of the NN. Every node is also characterized by a nonlinear activation function  $f(\mathbf{x}, \mathbf{w})$ . The outputs of the activation functions are then fed as input to the next network's layers. Finally, in the last layer, the activation functions' outputs are used to obtain the network's output. The closer this value is to the true output  $y$  of  $\mathbf{x}$ , the more accurate is the algorithm. The accuracy of the network can be improved by iteratively calculating a loss function (which estimates the level of discrepancy between true and predicted output) and updating accordingly the network's weights, aiming to obtain a lower loss value at the next iteration.

##### Logistic regression

Logistic regression (LR) classification models constitute another popular ML method that may be seen as a simplified version of an NN (Dreiseitl and Ohno-Machado, 2002). An LR model is indeed identical to a one-layer NN having a sigmoid function as the activation function. In particular, in a

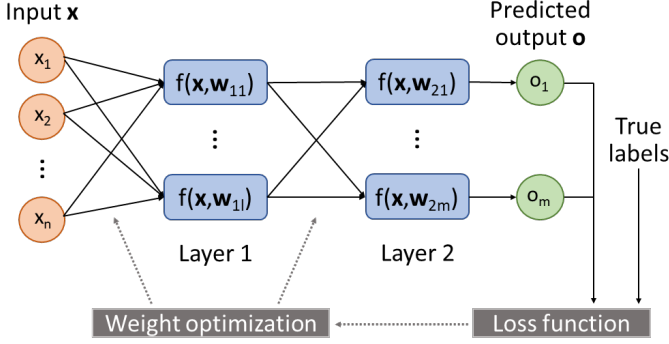


Figure 3.10: Schematic representation of a neural network with two hidden layers. A weighted sum of the  $n$  features of input vector  $\mathbf{x}$  is fed into each of the  $l$  nodes of Layer 1. Their output  $\mathbf{y}$  is also weighed and fed into Layer 2. In the training phase, the output of Layer 2 is compared with the true labels associated with  $\mathbf{x}$ . The weights of the network are then adjusted in order to reduce the loss function, that is the discrepancy between predicted output and true labels.

binary classification task (where 1 and 0 are the two classes), the LR model aims to compute the class membership probability of its input  $\mathbf{x}$  using a weight vector  $\boldsymbol{\alpha}$ :

$$P(y = 1) = \frac{1}{1 + e^{-\boldsymbol{\alpha} \cdot \mathbf{x}}}, \quad P(y = 0) = 1 - P(y = 1) = \frac{e^{-\boldsymbol{\alpha} \cdot \mathbf{x}}}{1 + e^{-\boldsymbol{\alpha} \cdot \mathbf{x}}}.$$

In the hyperplane of points  $\mathbf{x}$  for which  $\boldsymbol{\alpha} \cdot \mathbf{x} = 0$ ,  $P(y = 0) = P(y = 1) = 0.5$ . From the formulas above, it is also possible to derive the following relationship between the probabilities of Class 1 and Class 0:

$$\frac{P(y = 1)}{P(y = 0)} = \frac{1}{e^{-\boldsymbol{\alpha} \cdot \mathbf{x}}} \Rightarrow \ln \left( \frac{P(y = 1)}{P(y = 0)} \right) = \boldsymbol{\alpha} \cdot \mathbf{x}.$$

When dealing with multiclass problems, multinomial LR can be used. If  $K$  is the number of classes, the multinomial LR model for class  $k$  (with  $k \in [2, K]$ ) given an input vector  $\mathbf{x} \in \mathbb{R}^n$  can be defined by the formula

$$\ln \left( \frac{P(y = k)}{P(y = 1)} \right) = \sum_{q=1}^n \beta_{q,k} x_q = \boldsymbol{\beta}_k \cdot \mathbf{x} \Rightarrow P(y = k) = P(y = 1) e^{\boldsymbol{\beta}_k \cdot \mathbf{x}},$$

### 3.4. MACHINE LEARNING IN MEDICAL IMAGE PROCESSING

where  $\beta_k \in \mathbb{R}^n$  is the vector of regression coefficients associated with class  $k$  (Vermunt, 2010). Please note that in these formulas, Class 1 was chosen as the reference class in the LR model. However, any class can be chosen as reference. Moreover, given that the probabilities of all classes should sum to 1, it is possible to derive that:

$$P(y = 1) = \frac{1}{1 + \sum_{l=2}^K e^{\beta_l \cdot \mathbf{x}}} \Rightarrow P(y = k) = \frac{e^{\beta_k \cdot \mathbf{x}}}{1 + \sum_{l=2}^K e^{\beta_l \cdot \mathbf{x}}}.$$

To summarize, the idea behind LR (both in a binary and multiclass context) consists in modeling the class probabilities as the result of nonlinear functions, which are in turn applied on a linear combination of the input features.

#### Gaussian process regression

The above-described NN and LR models are fitted by assuming that the training data samples are independent and drawn from a true distribution that is unknown (Simeone, 2017). Thus, their parameters  $\theta$  are commonly obtained from the input data  $\mathbf{x}$  by using maximum likelihood estimation (MLE), that is  $\theta_{\text{MLE}} = \text{argmax}_{\theta} \ln(P(\mathbf{x} | \theta))$ . However, other types of algorithms exist that employ instead a Bayesian approach based on the Bayes formula:

$$P(\theta | \mathbf{x}) = \frac{P(\mathbf{x} | \theta)P(\theta)}{P(\mathbf{x})}.$$

This type of problem can be solved using a MAP approach:

$$\begin{aligned} \theta_{\text{MAP}} &= \text{argmax}_{\theta} P(\theta | \mathbf{x}) = \text{argmax}_{\theta} \frac{P(\mathbf{x} | \theta)P(\theta)}{P(\mathbf{x})} = \\ &= \text{argmax}_{\theta} \frac{P(\mathbf{x} | \theta)P(\theta)}{\int P(\mathbf{x} | \theta)P(\theta)d\theta} = \text{argmax}_{\theta} P(\mathbf{x} | \theta)P(\theta). \end{aligned}$$

Gaussian process regression (GPR) constitutes an example of a Bayesian approach that can be applied to regression problems. It is based on modeling the output  $y = f(\mathbf{x})$  of the model as a Gaussian process (GP), which is a collection of random variables having a joint Gaussian distribution and can thus be fully described by its mean function  $m(\mathbf{x})$  and covariance function  $k(\mathbf{x}, \mathbf{x}')$  (Rasmussen, 2003):

$$f(\mathbf{x}) \sim \text{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')).$$



### CHAPTER 3. BACKGROUND

It is important to stress the fact that every input  $\mathbf{x}$  is associated with a specific random variable  $f(\mathbf{x})$ . This introduces a certain margin of uncertainty in the model prediction. Moreover, in practice, the mean  $m(\mathbf{x})$  is usually set to 0. The covariance  $k(\mathbf{x}, \mathbf{x}')$ , which describes the relationship between any pair of points  $\mathbf{x}$  and  $\mathbf{x}'$ , is estimated in a limited set of points. The described GPR can be used as prior in the Bayesian inference: it does not depend on the training data but only on predefined properties of the process.

If  $\mathbf{X}$  is the training data and  $\mathbf{f}$  its associated function, while  $\mathbf{X}_*$  is the test data and  $\mathbf{f}_*$  its associated function, their joint distribution  $P(\mathbf{f}, \mathbf{f}_* \mid \mathbf{X}, \mathbf{X}_*)$  is described by

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu}_* \end{bmatrix}, \begin{bmatrix} \Sigma & \Sigma_* \\ \Sigma_*^T & \Sigma_{**} \end{bmatrix} \right),$$

where  $\boldsymbol{\mu}$  and  $\boldsymbol{\mu}_*$  are, respectively, the training and test means, while  $\Sigma$ ,  $\Sigma_*$  and  $\Sigma_{**}$  are the training, training-test and test covariances, respectively. Therefore, the distribution  $P(\mathbf{f}_* \mid \mathbf{f}, \mathbf{X}, \mathbf{X}_*)$  of  $\mathbf{f}_*$  conditioned by  $\mathbf{f}$  can be modeled by

$$\mathbf{f}_* \mid \mathbf{f}, \mathbf{X}, \mathbf{X}_* \sim \mathcal{N}(\boldsymbol{\mu}_* + \Sigma_*^T \Sigma^{-1}(\mathbf{f} - \boldsymbol{\mu}), \Sigma_{**} - \Sigma_*^T \Sigma^{-1} \Sigma_*).$$

Furthermore, we can take into account the fact that we normally observe noisy versions of the true function values, that is  $y = f(\mathbf{x}) + \epsilon$  (Wang, 2020). This noise can be modeled as additive Gaussian noise with zero mean and covariance  $\sigma^2 I$ , and therefore the joint distribution  $P(\mathbf{y}, \mathbf{f}_* \mid \mathbf{X}, \mathbf{X}_*)$  of the noisy observed values and the test data's function values becomes

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu}_* \end{bmatrix}, \begin{bmatrix} \Sigma + \sigma^2 I & \Sigma_* \\ \Sigma_*^T & \Sigma_{**} \end{bmatrix} \right),$$

while the conditional distribution  $P(\mathbf{f}_* \mid \mathbf{y}, \mathbf{X}, \mathbf{X}_*)$  is given by

$$\mathbf{f}_* \mid \mathbf{y}, \mathbf{x}, \mathbf{x}_* \sim \mathcal{N}(\boldsymbol{\mu}_* + \Sigma_*^T (\Sigma + \sigma^2 I)^{-1}(\mathbf{y} - \boldsymbol{\mu}), \Sigma_{**} - \Sigma_*^T (\Sigma + \sigma^2 I)^{-1} \Sigma_*).$$

During training, the parameters defining the mean and covariance function are tuned in order to implement an appropriate GP model in light of the available training data. If  $m(\mathbf{x})$  and  $k(\mathbf{x}, \mathbf{x}')$  are parameterized in terms of a set of hyperparameters  $\theta_m$  and  $\theta_k$ , respectively, the log marginal likelihood of the observed values (which are assumed to have a Gaussian

### 3.4. MACHINE LEARNING IN MEDICAL IMAGE PROCESSING

distribution) given the hyperparameters can be computed as

$$\ln(p(\mathbf{y} \mid \mathbf{X}, \theta_m, \theta_k)) = -\frac{1}{2}\ln(|\Sigma|) - \frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{y} - \boldsymbol{\mu}) - \frac{n}{2}\ln(2\pi).$$

The log marginal likelihood can finally be maximized using partial derivatives together with a numerical optimization routine:

$$\begin{aligned} \frac{\partial L}{\partial \theta_m} &= -(\mathbf{y} - \boldsymbol{\mu})^T \Sigma^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_m}, \\ \frac{\partial L}{\partial \theta_k} &= \frac{1}{2}\text{trace}(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_k}) + \frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^T \frac{\partial \Sigma}{\partial \theta_k} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_k} (\mathbf{y} - \boldsymbol{\mu}). \end{aligned}$$

Thus, during model training, the hyperparameters are adjusted according to the observed training input and output pairs. Once the GPR model is trained, probabilistic outputs are produced on inference phase by using the optimized conditional distribution.

#### 3.4.2 Introduction to deep learning

DL, also referred to as *deep structured learning* or *hierarchical learning*, is a sub-set of ML techniques that has recently impacted every sector of data processing. In the book *Deep Learning: Methods and Application* by Li and Dong (2014), DL is defined as “a class of machine learning techniques that exploit many layers of non-linear information processing for supervised and unsupervised feature extraction and transformation, and for pattern analysis and classification.” Therefore, a deep neural network (DNN) is often seen as an NN with a large (usually higher than three) number of hidden layers.

One of the main differences between DL and the rest of ML consists in DL’s ability to process input data directly in their raw form (LeCun et al., 2015). Traditional ML-based systems strongly rely on the use of a feature extractor that transforms the raw data into a specific feature vector, which can later be used as input for the ML algorithm. The definition of this feature extractor is a fundamental part of the process and requires expertise and engineering skills. With DL, this step can be eliminated. A DNN is trained to automatically determine which features should be extracted from the input data that are more strongly associated with the task of interest. This is done by combining multiple (“deep”) layers of representation: each layer extracts simple but nonlinear features from the previous layer. The more layers, the more abstract is the information being extracted from the

input data. For example, in the case of image classification, the first layer usually detects the presence of edges, while higher layers assemble particular arrangements of those edges that correspond to specific objects.

In the following sections, the main DL architectures of interest for the present doctoral thesis are described.

### 3.4.3 Convolutional neural networks

In the field of medical imaging, the most popular types of DL architectures are *convolutional neural networks* (CNNs) (Lundervold and Lundervold, 2019). The layers of a CNN are linked with very few connections (as opposed to the so-called *fully connected layers*), and they preserve the spatial relationships present in the input data, which makes them appropriate for image-oriented problems.

Like other standard artificial NNs, CNNs are typically trained using *backward propagation* and *gradient descent*. The first refers to the practice of iteratively calculating the gradient of the objective function with respect to the network's weights by the chain rule. This gradient is calculated for one layer at a time, starting from the last layer and going back to the first. Gradient descent is an optimization algorithm that allows identifying the weights that minimize the objective function.

A schematic representation of a typical CNN is shown in Figure 3.11. The most common types of layers in a CNN are described below.

- **Convolutional layers**

The input (either the raw image or the activations from the previous layer) is convolved with a set of filters, usually of size  $3 \times 3$  for 2D and  $3 \times 3 \times 3$  for 3D inputs. These filters are stored in tensors  $W_{ij}$ , where  $i$  is the layer number, and  $j$  is the filter number. Therefore, the same weights are shared across the whole input, leading to a reduction in the number of parameters that have to be learned. The results of these convolutions are referred to as *feature maps*.

- **Activation layers**

The feature maps are fed into nonlinear activation functions, which most often are simple rectified linear units (ReLUs):  $\text{ReLU}(z) = \max(0, z)$ .

- **Pooling layers**

These layers down-sample the feature maps by performing pooling

### 3.4. MACHINE LEARNING IN MEDICAL IMAGE PROCESSING

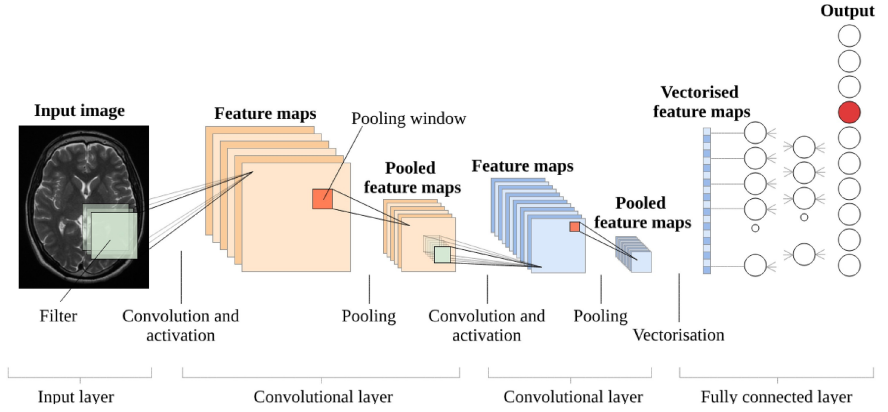


Figure 3.11: Representation of a simple CNN-based architecture for image classification. [From Lundervold and Lundervold (2019), © 2019 Elsevier, reprinted with permission]

operations, which take as input small regions of the map and assign one single output number to each of these regions. Common pooling operations are *max pooling*, which assigns the maximum value of the selected region, and *average pooling*, which performs an average function in the region. These operations provide the CNN with some translational invariance since small shifts in the image result only in small changes in their activation maps.

- **Dropout regularization**

A certain portion (referred to as *dropout rate*) of neurons is removed from some of the layers during training. Therefore, a slightly different network is trained for every new batch of training data, and the weights are optimized according to all these different variations. In this way, the model tends to overfit less.

- **Batch normalization**

These regularization layers normalize the activation maps by subtracting the mean and dividing them by the standard deviation for each of the training batches. In addition to reducing overfitting, the use of batch normalization also speeds up the training and makes the network less dependent on the initial parameter initialization.

A conventional CNN is thus made up of sequences of convolutional and

## CHAPTER 3. BACKGROUND

activation layers, which are often intermingled with pooling layers and regularization layers. Finally, the last layers of a CNN are typically fully connected layers in which every neuron is connected with all the activations of the previous layer.

One of the most popular uses of CNNs in medical image processing is segmentation. Early CNN-based segmentation methods were based on patch-wise classification in which each voxel is classified depending only on the local context around it (Milletari et al., 2016). However, the use of only local information made these networks prone to failure. Moreover, the higher the number of extracted patches, the higher the computational time. These issues led to the implementation of more efficient schemes, referred to as *fully convolutional networks* (FCNs) (Long et al., 2015). These models are trained end-to-end to take an image of any size as input and provide a correspondingly sized output segmentation. In the original FCN implementation, the last fully connected layer of a conventional CNN is replaced with a convolutional layer followed by an up-sampling operator that allows the final pixel-wise classification. Moreover, a skip-connection architecture is integrated in which high resolution feature maps are summed with the up-sampled outputs from deeper and more abstract layers. This sum is then fed into the last dense pixel-wise classification layer. The skip-connection trick combines coarse semantic information with more detailed spatial information in order to obtain a more accurate segmentation output.

### 3.4.4 Image segmentation with U-Net

One of the most famous DL-based architectures for medical image segmentation is the U-Net (see Figure 3.12), originally proposed by Ronneberger et al. (2015). Similarly to a traditional FCN, the U-Net consists of a *contracting path* (also referred to as *encoder*) made of subsequent convolutional and pooling layers, as well as an *expansive path* (also called *decoder*) where pooling is replaced by up-sampling operators that increase the resolution of deep feature maps. A skip-connection architecture is also implemented: the contracting path's high-resolution features are combined with up-sampled feature maps, and this combination is then fed into a successive convolutional layer. However, some modifications make the U-Net different from a more traditional FCN.

In the U-Net's expansive path, a large number of feature channels are used, so that the expansive path becomes symmetric to the contracting path. This allows the network to propagate much more contextual informa-

### 3.4. MACHINE LEARNING IN MEDICAL IMAGE PROCESSING

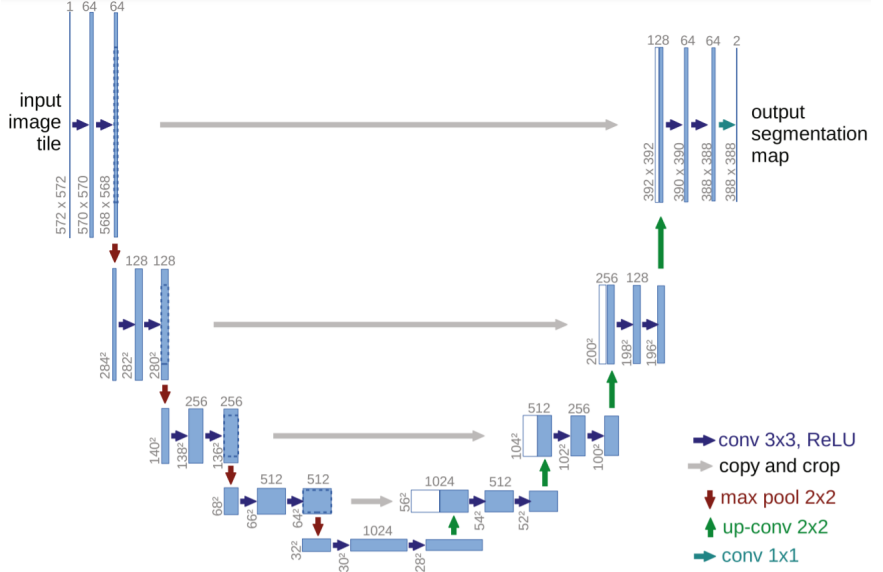


Figure 3.12: The original U-Net architecture. On top of each blue block (representing a multichannel feature map), the respective number of channels is presented. In the bottom of each block, the  $xy$  input size is represented. White blocks in the expansive path represent the feature maps copied from the contracting path. [Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Ronneberger et al. (2015) © 2015]

tion when up-sampling back to higher resolution layers. In particular, the contracting path is a simple CNN in which the number of feature channels is doubled at each down-sampling step. Each step in the expansive path consists of the following:

1. A transpose convolution made of learnable filters that replace the simple bilinear interpolation used in traditional FCNs for up-sampling. This operation halves the number of feature channels.
2. A concatenation with the symmetric (i.e., having the same image size) feature map from the contracting path. The concatenation replaces the sum of a traditional FCN.
3. Two convolutional layers, each followed by a ReLU activation.

In the original U-Net paper, the architecture was proposed for segmentation of neuronal structures in electron microscopy images. However, because of its extremely successful performance, this method has been repurposed—and often modified—for several other medical image segmentation tasks, as mentioned in the Introduction of this thesis. Since medical images are often 3D, one relevant extension of the U-Net is the 3D U-Net model developed by Çiçek et al. (2016). Similarly to the original U-Net, this model is made of a contracting and an expansive path, but all convolution, pooling and transpose convolution operations are performed using 3D kernels.

An alternative way of segmenting 3D images using U-Nets consists in implementing three different 2D U-Nets, each one trained on input slices from a specific view (axial, coronal or sagittal). Since the three views are orthogonal to one another, the results from the three 2D U-Nets can easily be combined back together (e.g., by averaging them) in a 3D volume.

### 3.4.5 Incorporating context information into a DNN

The U-Net architecture described in the previous section as well as most of the traditional image segmentation frameworks usually receive as input the simple raw image to be segmented. However, the segmentation accuracy can potentially be improved by adding some context information as input to the network. The incorporation of context information consists in adding priors and likelihood to the segmentation pipeline, and it is a widely used practice in computer vision (Tu and Bai, 2010).

One of the simplest examples of context information is the *autocontext*, which consists in first training a network and then feeding its output as second input of a second network (Tu and Bai, 2010; Chen et al., 2018; Mirikharaji et al., 2018). This strategy has also been employed to improve the accuracy of brain MRI segmentation results (Chen et al., 2018).

In a study by Mahbod et al. (2018), shape context was tested as context information to improve the performance of artificial NNs for brain segmentation; this was inspired by the work by Leventon et al. (2000) described in Section 3.3.2. The idea of incorporating shape context information was later tested on DNNs in several studies, which showed that the use of shape priors in the segmentation pipeline can increase the network’s performance. Shape context has been integrated into DNNs, for example, by adding a convolutional autoencoder to a U-Net as a shape regularization network (Ravishankar et al., 2017), by feeding a statistical shape model as a second input channel to a U-Net (Wang and Smedby, 2017) and by jointly training

### 3.4. MACHINE LEARNING IN MEDICAL IMAGE PROCESSING

an FCN with a level set (Tang et al., 2017).

The inclusion of a priori information into the segmentation pipeline can be relevant specifically in medical imaging since we often suffer from lack of training data, but extensive prior knowledge on human anatomy is available.

#### 3.4.6 Relevant applications in brain MRI image processing

In the appended Papers II to V, ML is employed as the main tool to process and analyze structural brain data. In this section, relevant background literature on the topics addressed by these four papers is presented.

##### Brain age prediction models

Previous studies have shown that brain MRI scans can be used for predicting chronological age in healthy subjects with high accuracy. The first neuroimaging-based approaches for brain age prediction proposed ML models fed with different MRI modalities as inputs, including T1w MRI (Dosenbach et al., 2010; Franke et al., 2010). The T1w-based approach consists of first processing the raw images using VBM to standardize the input data. Second, data dimensionality reduction is automatically performed to avoid overfitting and to increase computational efficiency. Finally, a regression model such as GPR (described in Section 3.4.1) or relevance vector regression (RVR) (Tipping, 2000, 2001) is fitted in order to model brain age using the processed MRI images. However, in recent years, DL has also been employed for this purpose. A DL-based approach was tested for the first time by Cole et al. (2017b), who developed 3D CNNs for age prediction. More recently, more advanced DL-based approaches, for example residual CNNs (Jónsson et al., 2019) and the Inception-ResNet-v2 architecture (Bashyam et al., 2020), have also been proposed and have obtained accurate brain age prediction results in humans.

These models are meant to be trained on healthy individuals and tested on new unseen subjects. If during testing the predicted brain age is greater than the subject’s chronological age, this might reflect an underlying disease or neurodegeneration (Cole et al., 2017b). Conversely, a predicted age that is lower than the chronological age could indicate a positive trend in the aging process. In one of the first brain age prediction studies by Franke et al. (2010), the difference between predicted and chronological age was defined as brain age gap estimation, also referred to as BrainAGE score.

Brain age prediction models have been tested on patients with several



### CHAPTER 3. BACKGROUND

neuropsychiatric diseases as well as other clinical conditions. An increase in predicted age was found to be associated with several conditions, such as epilepsy (Pardoe et al., 2017), schizophrenia (Nenadić et al., 2017), traumatic brain injury (Cole et al., 2015), HIV (Cole et al., 2017c) and Down’s syndrome (Cole et al., 2017a). Studies performed on subjects with MCI have also shown that a higher BrainAGE score is found in individuals who will develop Alzheimer’s disease within three years (Gaser et al., 2013; Löwe et al., 2016). Increased BrainAGE scores have also been found in the presence of diabetes (Franke et al., 2013) and midlife obesity (Ronan et al., 2016). Conversely, measures of physical activity as well as number of years of education were found to correlate with lower brain age (Steffener et al., 2016). Moreover, a study by Cole et al. (2018) found an association between higher brain age prediction and higher mortality risk. The use of these MRI and ML-based models is thus becoming more relevant for gaining a better understanding of brain aging. Furthermore, the above-mentioned research results regarding the association between BrainAGE scores and clinical conditions suggest that such neuroimaging tools are promising for establishing a new aging biomarker.

Standardized criteria exist for qualifying what can be considered an aging biomarker. According to the American Federation for Aging Research (AFAR, from the *Infoaging Guides*, 2016), one of the requirements for an aging biomarker is that it must work on both humans and laboratory animals so that it can be tested extensively in a laboratory before being validated in humans. This highlights the importance of performing brain age studies in a preclinical framework. Franke et al. (2016) proposed two species-specific adaptations of the brain age prediction model: one for baboons and one for rats. In both cases, the model achieved accurate results, thus showing how it can effectively be adapted to animal studies. In particular, the rat-specific model showed the best results, also thanks to its richer dataset: a correlation coefficient of 0.95 was obtained between chronological and predicted age with a mean absolute error of 49 days (corresponding to 6% of the total age range).

Preclinical studies on BrainAGE can be particularly useful not only to investigate the value of these predictive models across species but also to study brain aging in more controlled laboratory environments. The aging process is indeed highly heterogeneous and influenced by both genetic and environmental factors. In a laboratory setting, it is possible to better control for such factors and therefore investigate which interventions could

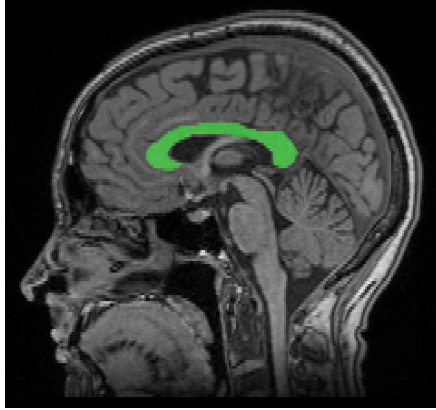


Figure 3.13: Segmentation of the corpus callosum (in green) manually performed on the midsagittal slice of a T1w scan.

potentially slow down—or even reverse—brain aging.

#### Segmentation of the corpus callosum

As mentioned in Section 3.1.4, the CC is a structure of great importance as a proxy for neurodegeneration in MS, since it is greatly affected by lesions and atrophy. Although the CC can be identified as a 3D structure, multiple studies have rather focused on the analysis of the 2D CC that can be segmented from a scan’s midsagittal slice (see Figure 3.13). In such a slice, the CC is indeed more easily identifiable and distinguishable with respect to the surrounding brain tissue (Park et al., 2018b; Joshi et al., 2013). Moreover, moving from 3D to 2D representations can decrease complexity while still providing useful measures related to atrophy (such as area and thickness) and reflecting the overall properties of the CC fibers (Joshi et al., 2013). Once both the CC and the intracranial (IC) area are segmented in 2D, the normalized CC can be computed as the ratio between CC and IC area. This metric has been shown to significantly correlate with MS burden in previous studies (Granberg et al., 2015a,b).

In order to accurately segment the CC from a midsagittal slice, several methods have been proposed in the literature, especially as applied to T1w scans (Cover et al., 2018). In a study by Adamson et al. (2014), an atlas-based approach, followed by a refinement step based on morphological operations and vein/vessel removal, was employed on T1w images of both

healthy controls and Alzheimer’s patients. Model-based methods have also been tested; for instance, in a study by Liu and Ruan (2015), high accuracy was obtained by integrating shape priors into a level set-based segmentation framework. High performance was also obtained in other works that employed ML methods such as artificial neural networks (Magnotta et al., 1999) as well as Gaussian mixture modeling and fuzzy C-means (İçer, 2013).

With the successful advent of DL, some studies have also applied CNNs to segment the CC. Park et al. (2018a) proposed two patch-based CNN methods: one without any prior information (i.e., just using the MRI slice as input) and one with prior information (i.e., adding two input channels: a CC probability map and an average image, both generated from an atlas-based segmentation approach). Both methods showed a mean Dice coefficient of approximately 0.95. Additionally, Chandra et al. (2019) explored an encoder-decoder network for the same purpose.

However, to the best of the author’s knowledge, current literature is still largely lacking DL-based CC segmentation methods specifically trained and tested on MS patients, although such a study would provide a valuable contribution to the understanding of MS progression as well as an advancement in research for neuroimaging biomarkers of MS.

### **Segmentation of the hippocampus**

As described in Section 3.1.4, hippocampal atrophy is one of the early signs of AD (Scheltens et al., 2002). An accurate segmentation of the hippocampus from MRI scans can allow measuring its volume, which is useful in order to separate healthy subjects from AD patients as well as to potentially identify cases that may soon convert into AD (Liu et al., 2010). Moreover, hippocampal shape analysis was also shown to provide useful indicators for AD diagnosis and for its differentiation from other types of dementia (Lindberg et al., 2012).

Various software tools performing automatic hippocampal segmentation already exist and are widely used. These include, most notably, FreeSurfer and FSL (see Section 3.3.3). However, their computational time is rather long (in the order of a few hours) and not suitable for use in the clinical setting. Therefore, many other methods for automatic hippocampal segmentation have been proposed in the literature (Zhu et al., 2021; Dill et al., 2015). Both single-atlas (Barnes et al., 2007; Carmichael et al., 2005; Kwak et al., 2013) and multi-atlas (Heckemann et al., 2006; Sabuncu et al., 2010; Wenyan et al., 2011) based methods have been shown to be particularly

### 3.4. MACHINE LEARNING IN MEDICAL IMAGE PROCESSING

successful at this task. Furthermore, with the advent of DL, a large portion of the literature is based on the use of DNNs that perform quicker automatic segmentation of the hippocampus. Such methods are mostly divided into two groups (Zhu et al., 2021): (1) DL methods that integrate multi-atlas approaches into the pipeline and (2) end-to-end DL models. The first ones aim to use DNNs to further improve the performance obtained from traditional atlas-based approaches (Fang et al., 2019; Yang et al., 2018). The second group of studies aims to perform the segmentation by directly receiving the raw MRI scan as input, thus proposing novel DL-based strategies for this purpose (Chen et al., 2017; Milletari et al., 2017; Chen et al., 2018; Guha Roy et al., 2019; Thyreau et al., 2018).

Independently from the method being used, performing an accurate segmentation of the hippocampus is particularly challenging, especially in the presence of pathology (i.e., when the hippocampus is atrophic) as well as when dealing with different MRI scanners and protocols or images affected by artifacts (Akkus et al., 2017). Therefore, it is important to thoroughly investigate suitable DL methods that are also robust across different diagnostic groups and across images acquired from different scanners and centers.



## Chapter 4

# Methods

### 4.1 Investigating group differences from brain MRI: Two animal studies

The first two works presented in this thesis are animal studies that—despite having different aims—are both based on the use of similar image processing methods for analyzing differences between animal groups of interest.

#### 4.1.1 Overview of the studies

In Paper I, brain MRI scans of eight wild and eight domestic rabbits were used to identify differences in brain morphology between these two groups. Both T1-weighted (T1w) and diffusion MRI (dMRI) images were employed in order to study both gray matter (GM) and white matter (WM) differences, respectively.

Paper II aimed to investigate differences in MRI-derived brain age between a group of 24 control rats and a group of 24 rats that underwent long-term environmental enrichment and dietary restriction (EEDR). First, a machine learning (ML)-based age prediction model was trained using processed T1w images from a cohort of 31 normal aging rats. Subsequently, this model was tested on the test cohort of interest to study the effect of the two different lifestyles on brain age. All rats (from both the training and test cohorts) were scanned at four specific time points in their life: at about 3, 5, 11 and 17 months of age.

Both Papers I and II rely on the use of appropriate processing pipelines that allow extracting and analyzing relevant information from the brain scans. These pipelines include multiple steps that are presented in the

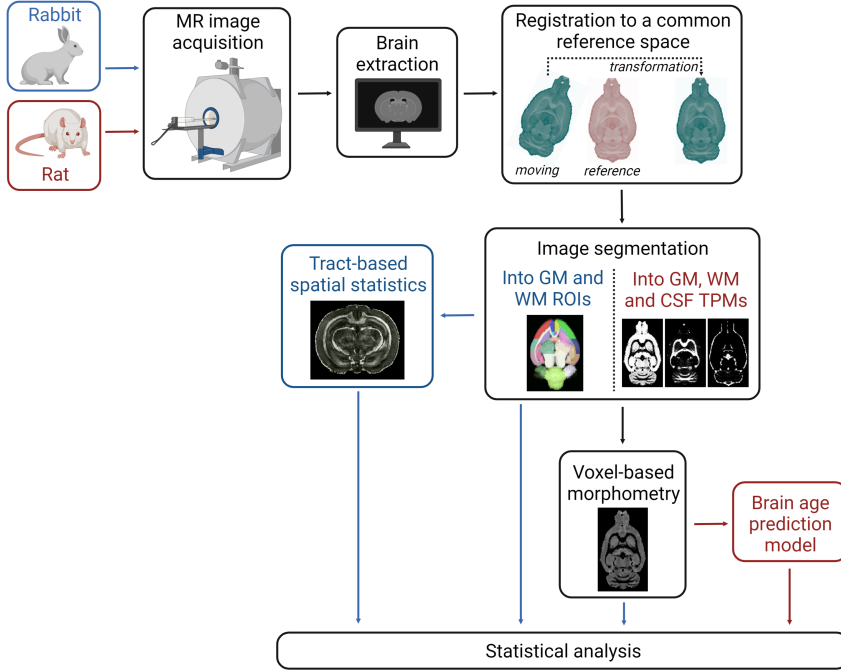


Figure 4.1: Schematic representation of the image processing workflow implemented for Papers I and II of this thesis. These two studies share some common processing steps that are represented with black boxes and arrows, whereas study-specific steps are shown in blue for Paper I and in red for Paper II. (Figure created with BioRender.com)

following sections (see also Figure 4.1). Information on animal breeding and MRI acquisition protocols is not included in this thesis as these tasks had been previously performed by other collaborators.

#### 4.1.2 Brain extraction

In both Papers I and II, the first step consisted in brain extraction from the analyzed T1w scans, that is the removal of the skull and any other nonbrain tissue from the MRI image. When working with human brain scans, multiple software exists for robust skull-stripping on T1w images, for example FSL’s Brain Extraction Tool (BET) (Jenkinson et al., 2012; Smith, 2002) and the skull-stripping step of the FreeSurfer processing pipeline (Fischl,

#### 4.1. INVESTIGATING GROUP DIFFERENCES FROM BRAIN MRI: TWO ANIMAL STUDIES

2012; Ségonne et al., 2004). However, when processing animal data, such popular tools would obviously fail. Therefore, tailored species-specific solutions are needed.

In Paper I, rabbit skull-stripping was performed by an expert collaborator using an interactive segmentation method based on level set (Wang et al., 2014). Once all T1w images were skull-stripped, they were manually aligned to the dMRI data of their respective subjects. This procedure allowed identifying the brain area in dMRI data too, and thus obtaining skull-stripped FA maps by running the `dtifit` function of FMRIB’s Diffusion Toolbox (Jenkinson et al., 2012; Smith et al., 2004).

For Paper II, an atlas-based approach was employed by running a modified implementation of the `artsBrainExtraction` algorithm proposed by MacNicol et al. (2020). In order to run this algorithm, a reference atlas with its associated brain mask must be selected. For this purpose, the 11-month-old rat brain template created by MacNicol et al. (2021) was employed since it corresponds to a “middle age” that could minimize the differences across all rats analyzed in the study.

##### 4.1.3 Image registration and segmentation

In order to properly analyze and compare the extracted animal brains, it is necessary to normalize images to a common space. Within this image space, the brain images can be parcellated into different brain structures.

In Paper I, the rabbit brain template by Muñoz-Moreno et al. (2013) was nonlinearly registered to each T1w image. More specifically, the NiftyReg software was used to run a free-form deformation algorithm (Modat et al., 2010) with the template as moving image. The same transformation was then applied to the corresponding rabbit atlas, which is defined in the template’s space and includes 60 brain regions of interest (ROIs). This atlas-based segmentation allowed parcellating each subject’s brain into 60 ROIs. However, due to the difference in resolution between the atlas and the MRI images used in the study ( $0.15 \times 0.15 \times 7 \text{ mm}^3$  vs.  $0.16 \times 0.16 \times 0.17 \text{ mm}^3$ ), the segmentations of some ROIs had to be manually refined in order to improve their quality and obtain more reliable results. Finally, since the atlas includes both GM and WM regions, it was possible to use this information to obtain both GM and WM tissue maps for each subject.

Similarly to Paper I and as already touched upon in Section 4.1.2, a reference atlas was employed for tissue segmentation also in Paper II. The previously mentioned `artsBrainExtraction` algorithm includes a nonlinear



registration of each individual rat brain to the reference template. This registration step is based on the diffeomorphic image registration algorithm implemented within ANTs Syn (Avants et al., 2008). Subsequently, ANTs **Atropos** tool (Avants et al., 2011b) was employed to extract subject-specific tissue probability maps (TPMs) for GM, WM and cerebrospinal fluid (CSF).

### 4.1.4 Voxel-based morphometry

Differences in brain tissue volume—with respect to a reference template, as well as between groups—can be investigated not only from ROI-based analyses but also at the voxel level using voxel-based morphometry (VBM) (Ashburner and Friston, 2000), as described also in the background Section 3.3.3.

In Paper I, the FSL-VBM tool (Douaud et al. (2007), <http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSLVBM>) was employed as a basis to perform VBM on the rabbit data. However, the human-specific pipeline implemented within FSL had to be slightly adjusted in order to use it with the rabbit data of interest. The original FSL-VBM relies on an affine registration of each subject’s GM segmentation to the human GM ICBM-152 template (Lancaster et al., 2007). In Paper I, this template was replaced with the GM segmentation of a randomly selected wild rabbit, which was blurred using a Gaussian filter ( $\sigma^2 = 0.1 \text{ mm}^2$ ). Subsequently, a study-specific template was generated by nonlinearly registering the GM segmentations of every subject to the created reference template (Andersson et al., 2007a,b) and then averaging them. Nonlinear registrations were then run again, but this time using the study-specific template as reference image. The deformation fields from these latter registrations were finally used for deriving the Jacobians, from which each subject’s modulated GM image could be computed. A Gaussian filter was applied to smooth each modulated image before computing voxel-wise statistics.

Modulated tissue segmentations—or, more precisely, TPMs (for GM, WM and CSF)—were also obtained in Paper II. However, they were not used for finding differences between groups but rather as input to the implemented brain age prediction model (see Section 4.1.6 for details). Moreover, the Jacobians obtained from the nonlinear registration to the reference 11-month-old template (Section 4.1.3) could be multiplied to all the TPMs to obtain modulated TPMs. Here, no smoothing was applied to the modulated images, but the used registration algorithm (ANTs Syn) applies a Gaussian smoothing to the deformation field after every iteration. Furthermore,

#### 4.1. INVESTIGATING GROUP DIFFERENCES FROM BRAIN MRI: TWO ANIMAL STUDIES

TPMs are themselves smoother compared to binary tissue segmentations.

##### 4.1.5 Tract-based spatial statistics

In Paper I, not only T1w but also dMRI images were available; therefore, fractional anisotropy (FA) maps could be extracted (see Section 3.2.4). FSL’S tract-based spatial statistics (TBSS) (Smith et al., 2006; Jenkinson et al., 2012) was employed to study local differences in FA between domestic and wild rabbits (see Section 3.3.3). Similarly to what was mentioned regarding FSL-VBM, some adjustments had to be performed in order to run the full pipeline on rabbit images.

The same wild rabbit used as reference for FSL-VBM was also employed as reference for TBSS: all rabbits’ FA maps were first nonlinearly registered to it (Andersson et al., 2007a,b) and then averaged. The mean FA image was later thinned in order to obtain an FA skeleton that would represent the main WM fiber tracts identified in the rabbits’ brains. Finally, voxel-wise cross-subject statistics were computed on the FA images projected onto the mean FA skeleton.

##### 4.1.6 Brain age prediction

###### Description of the GPR and LR models

One of the main goals of Paper II was to use the available rat brain MRI scans to train and test a novel rat brain age prediction model. The proposed model combines Gaussian process regression (GPR) and a logistic regression (LR) classifier. Both models can receive as input the modulated GM, WM and CSF TPMs of one subject after they have undergone principal component analysis (PCA). In particular, the projections onto the first 77 principal components were used as inputs. The output of each model represents the predicted rat brain age expressed in weeks.

The GPR model (fitted using a linear kernel) was the baseline model for this work since it has previously shown accurate results in human studies (Cole et al., 2015). The addition of an LR classifier into the pipeline, and thus the formulation of the age prediction problem as a classification task, constituted the main methodological novelty of this study. This extension was originally implemented with the aim to investigate whether it could help reduce the “regression towards the mean” that intrinsically affects GPR (Liang et al., 2019). For this purpose, the whole available age range (spanning from 14 to 70 weeks) was divided into 40 bins (or classes),

and each subject was assigned to its respective “age class”. The model was then trained to predict the correct age class for every input subject. When fitting the model, class weights were used to correct for data imbalance: each age class was associated with a weight that was inversely proportional to the class’s frequency in the training dataset.

Finally, the predictions from the two independent GPR and LR models were combined by computing their weighted average. Each GPR prediction had a weight of 1. The weight of each LR prediction was set equal to its classification probability estimate, that is the probability (produced as output by the classifier) that the input subject actually belongs to the predicted age class.

### Cross-validation and testing strategy

Both the GPR and LR models were first trained in a leave-one-out cross-validation fashion using the 31 normally aging rats belonging to the study’s training cohort. This means that every model was trained 31 times in total: at every iteration, the longitudinal data from one rat were used as the validation set, while the data from all remaining 30 rats constituted the training set.

Once the models had been optimized and their performance had been evaluated with cross-validation, they were retrained once more using all 31 rats as the training set. The newly fitted pipeline was then tested on a separate cohort of rats (the aging cohort), including 24 controls and 24 EEDR rats. For each rat at each available time point, the BrainAGE score was computed as the difference between predicted and chronological age.

#### 4.1.7 Statistical analyses

As mentioned at the beginning of Section 4.1, the common denominator between the first two studies is the aim to use MRI for investigating differences between two groups of animals (domestic and wild rabbits for Paper I, control and EEDR rats for Paper II). In order to achieve this, appropriate statistical analysis methods must be applied. For all the applied methods described below, the significance threshold was set to  $p = 0.05$ .

In Paper I, a ROI-based analysis was first performed by studying the difference in ROI volumes between the two rabbit groups. This was accomplished by first dividing each ROI volume by the total subject’s cerebrum volume (made up of 54 ROIs in total). Then, permutation  $t$ -tests (with

## 4.2. DEEP LEARNING AND SHAPE ANALYSIS OF THE DEGENERATING HUMAN BRAIN

5 000 permutations) were performed to find statistically significant differences between groups for each GM ROI (21 in total). Bonferroni correction was carried out to correct for these 21 independent statistical tests.

The same 5 000-permutation strategy was followed for the analysis of the VBM and TBSS results in order to reduce the chance of finding false positives. However, in these cases, the design of the statistical tests was chosen according to what was already implemented within the FSL workflow. In particular, voxel-wise generalized linear models and subsequent permutation-based nonparametric testing were used for identifying statistically significant differences in both local GM volume and local FA between domestic and wild rabbits.

In Paper II, other types of group differences were investigated. First, we were interested in analyzing whether controls and EEDR rats showed statistically significant differences in their BrainAGE scores throughout their lifespan. For this purpose, we fitted a linear mixed-effects model with BrainAGE score as the dependent variable. The fixed effects were chronological age, lifestyle group (controls vs. EEDR) and their interaction. Moreover, since it was a longitudinal study with repeated measures, subject identity was considered a random effect.

Finally, we aimed to investigate whether lifestyle group and/or BrainAGE score had any significant effect on survival. In particular, we focused only on the BrainAGE scores measured at the first two time-points of the study observation period (i.e., at about three and five months of age), when no test rats were yet excluded from the study. The survival analysis was performed by fitting five Cox regression models, each having different independent variables: (1) lifestyle group only, (2) BrainAGE at the first session only, (3) BrainAGE at the second session only, (4) lifestyle and BrainAGE at the first session, and (5) lifestyle and BrainAGE at the second session. When performing this analysis, seven rats were considered as having terminal events since they died for health-related reasons before the end of the study's observation period. The survival data from the other 41 rats was considered censored.

## 4.2 Deep learning and shape analysis of the degenerating human brain

Papers III to V constitute the human studies of the present thesis and introduce additional methods that were not investigated in Papers I and II.

## CHAPTER 4. METHODS

First, these works are all based on deep learning (DL) methods for both 2D image classification (Paper III) and segmentation (Papers III, IV and V). Furthermore, they introduce the use of statistical shape models based on PCA. These models are employed both to analyze the results of DL segmentation frameworks (Paper IV) and to further improve their accuracy (Paper V). Finally, these studies aim to create robust processing pipelines that can work on brain images not only in healthy subjects but also in patients with multiple sclerosis (MS) (Papers III and IV), Alzheimer’s disease (AD) and mild cognitive impairment (MCI) (Paper V).

### 4.2.1 Overview of the studies

Papers III and IV focus on U-Net-based 2D segmentation of the corpus callosum (CC) in MS patients. This segmentation is performed on the midsagittal slice of the acquired 3D MRI scans since the CC is well-defined and reflects the overall characteristics of the 3D CC (see background Section 3.4.6). In Paper III, an automatic method for midsagittal slice selection on T1w and FLAIR images is also presented. In Paper IV, segmentation is instead performed directly on manually selected midslices of T2w images. Metrics related to the CC area, as well as shape in the case of Paper IV, were also investigated with respect to physical and cognitive disability scores.

In Paper V, a novel framework for automatic segmentation of the hippocampus from T1w images is presented. The proposed approach consists in embedding shape context information into the segmentation pipeline. The method was tested on three different diagnostic groups (healthy controls [HC], AD and MCI), as well as in two different cohorts (AddNeuroMed and ADNI).

In the following sections, the methods implemented in these three studies are described. Detailed information on the datasets and MRI acquisition protocols is not presented: for Papers III and IV, these tasks had been previously performed by other collaborators from the Karolinska Institute; for Paper V, the data were obtained from the AddNeuroMed Consortium (Lovestone et al., 2009; Simmons et al., 2009) and the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (Jack et al., 2008). Furthermore, in Paper IV, it should be noted that the author of this thesis was only responsible for the shape analysis part of the work. However, some details on the DL-based segmentation method are provided in this section for completeness.

### 4.2.2 Automatic midslice selection

The automatic midsagittal slice selection method proposed in Paper III is based on a convolutional neural network (CNN) that was trained to predict whether an input sagittal slice is (Class 1) or is not (Class 0) a midslice. The CNN was implemented with four 2D convolutional layers (with  $3 \times 3$  kernels; rectified linear unit activation function; 16, 32, 64 and 64 filters, respectively) intermingled with max-pooling layers and finally followed by two dense layers (with 128 and 1 unit, respectively). The last dense layer has a sigmoid activation function that will produce an output from 0 to 1 that can be interpreted as the estimated probability of the input image being a midsagittal slice.

The above-described architecture (referred to as *midCNN*) was trained three times, depending on the type of MRI sequences provided as input: *midCNN<sub>T1</sub>* for T1w inputs only, *midCNN<sub>FLAIR</sub>* for FLAIR only and *midCNN<sub>T1/FLAIR</sub>* for both T1w and FLAIR. All three architectures were trained both with 10-fold cross-validation and with scanner-wise cross-validation. In the first case, all available 3D scans were randomly split into 10 folds. By contrast, the scanner-wise approach consisted in a three-fold cross-validation method in which the images of each fold all belonged to the same scanner (i.e., Siemens Aera, Avanto or Trio).

During each of the training phases, only half of the available non-middle slices were used. This was done to address the problem of data imbalance since only one middle slice per scan is available. In the validation phase, though, all available sagittal slices of every subject were fed into the trained model: the slice associated with the highest *midCNN*'s output was considered to be the predicted midslice.

For all the proposed CNNs, the same image preprocessing was performed: each input slice was resized to  $256 \times 256$  pixels and then normalized by subtracting its mean pixel intensity and dividing it by its standard deviation.

### 4.2.3 U-Net-based segmentation

In all three human studies, the popular U-Net architecture originally proposed by Ronneberger et al. (2015) was used to perform 2D image segmentation (see background Section 3.4.4).

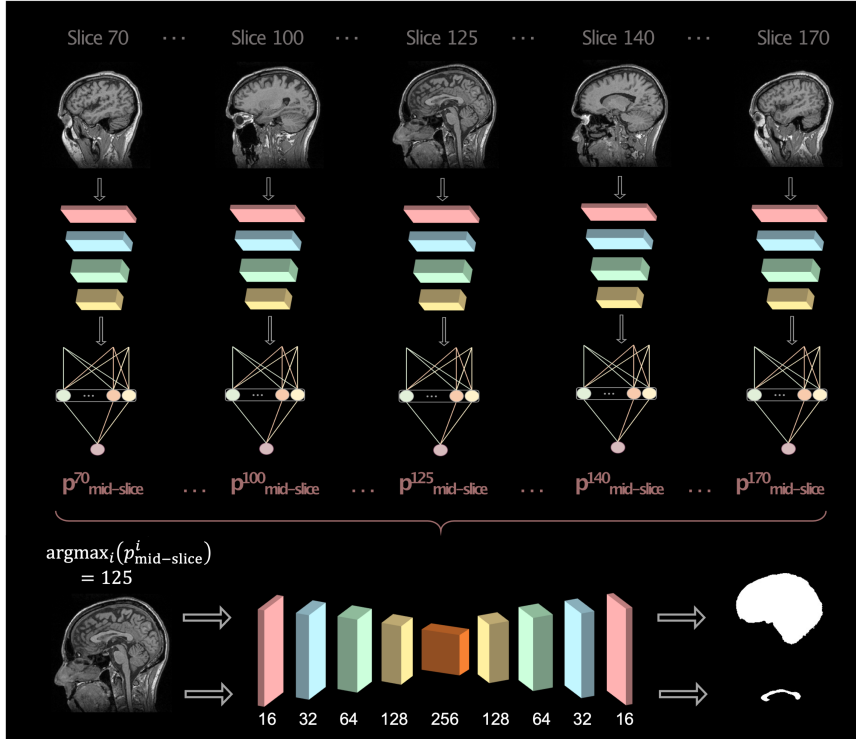


Figure 4.2: The pipeline proposed in Paper III for automatic midslice selection (upper part) and subsequent 2D segmentation of both CC and IC (lower part). All slices of an input 3D scan are fed into a CNN and classified as either midslice or not. The slice  $i$  having the highest probability  $p_{\text{mid-slice}}$  of being a midsagittal slice is fed into two U-Net-like architectures, which produce an IC and a CC binary segmentation as output. [© Brusini et al. (2022b). *Journal of Neuroimaging* published by Wiley Periodicals LLC on behalf of American Society of Neuroimaging. Reprinted with permission]

### Segmentation of the corpus callosum

In both Papers III and IV, two main types of segmentation networks were proposed: one for segmenting the CC and one for segmenting the intracranial (IC) area from an input MRI slice. The latter segmentation was performed in order to compute the normalized CC area, which is defined as the ratio between CC and IC area.

In Paper III, we focused on two types of input MRI sequences

## 4.2. DEEP LEARNING AND SHAPE ANALYSIS OF THE DEGENERATING HUMAN BRAIN

(T1w and FLAIR) and trained different networks according to the type of sequence that can be provided as input:  $CC\text{-}Net_{T1}$ ,  $CC\text{-}Net_{FLAIR}$  and  $CC\text{-}Net_{T1/FLAIR}$  for CC segmentation;  $IC\text{-}Net_{T1}$ ,  $IC\text{-}Net_{FLAIR}$  and  $IC\text{-}Net_{T1/FLAIR}$  for IC. These networks receive as input a slice that can be automatically selected by running the midslice selection algorithm described in Section 4.2.2 (see also Figure 4.2). In Paper IV, slices from only T2w images were used, and therefore only one CC and one IC segmentation framework were trained.

In Paper III, the original U-Net architecture was used, except for the addition of batch normalization layers after every convolutional layer. For  $IC\text{-}Net_{T1/FLAIR}$ , though, batch normalization was removed since it led to less accurate results. In Paper IV, besides batch normalization, dropout layers were also added into the framework to prevent overfitting. Moreover, other techniques were employed to improve the inference performance, including online data augmentation (in Papers III and IV) and the use of different sample weights (for  $CC\text{-}Net_{T1/FLAIR}$  and  $IC\text{-}Net_{T1/FLAIR}$  in Paper III in order to increase the accuracy on FLAIR images, which are more challenging to segment).

In both studies, the U-Nets were first trained in a 10-fold cross-validation fashion. Furthermore, in Paper III, a scanner-wise cross-validation approach was investigated (as outlined above in Section 4.2.2).

Finally, in Paper III, additional test sets were used to investigate both the reproducibility of the proposed framework and the correlation between normalized CC and two types of disability scores: Expanded Disability Status Scale (EDSS) and Symbol Digit Modality Test (SDMT) scores (see background Section 3.1.4).

### Segmentation of the hippocampus

In Paper V, the hippocampus was segmented from 3D T1w scans using three separate 2D U-Nets, making up an architecture referred to as *MRI U-Net*. These three U-Nets were trained independently for each image view: axial, coronal and sagittal. Once the network outputs were obtained for each slice from each view, they were averaged together in a final 3D volume that constitutes the final segmentation output.

The centers of gravity of the left and right hippocampal segmentations were estimated. Subsequently, the input T1w scan was cropped around the two centers of gravity, generating two new smaller 3D volumes with a pre-defined size. These cropped 3D images were used to train a new architecture



made up of other three 3D U-Nets (one per view), which is referred to as *Cropped MRI U-Net*.

Finally, one last ensemble of three orthogonal U-Nets was implemented: the *Shape MRI U-Net*. In this framework, two input images are provided as input: the cropped MRI image (already mentioned above) and its respective shape context, which was obtained by using a statistical shape model fitted to the segmentation output of Cropped MRI U-Net. In Section 4.2.4, the methods for the generation of such shape context information are described in more detail.

All the proposed network architectures are schematized in Figure 4.3. They were first trained and evaluated on 54 subjects from the ADNI cohort using nine-fold stratified cross-validation stratifying by magnetic field strength (1.5T and 3.0T) and diagnosis (HC, MCI and AD). Finally, these trained networks were tested first on 37 scans from the AddNeuroMed cohort and then on a separate dataset of 5 948 unlabeled ADNI subjects.

### 4.2.4 Shape analysis

In Papers IV and V, a crucial role is played by statistical shape models based on PCA. However, they are employed in different ways. In Paper IV, the principal components (PCs) of the CC shape model were analyzed to identify which shape variations correlated with disability. In Paper V, the hippocampal shape model was used to generate the input context layer of Shape MRI U-Net.

### Shape analysis of corpus callosum segmentations

The output CC segmentations of Paper IV (see Section 4.2.3) were used to generate a statistical shape model. According to the method proposed by Leventon et al. (2000) and described in the background Section 3.3.2, distance transforms were first computed from the segmentations and used for generating the model. Only the first PCs representing 99% of the total variance were preserved. It was then possible to analyze the projections of each subject's CC distance transform onto the PCs' space and compute their correlations with EDSS and SDMT.

In addition to the analysis described above, we investigated how the thickness and bending angle of the CC may change with disability. CC thickness was obtained by averaging the values of the CC distance transform along the CC midline and then computing its absolute value. The bending

## 4.2. DEEP LEARNING AND SHAPE ANALYSIS OF THE DEGENERATING HUMAN BRAIN

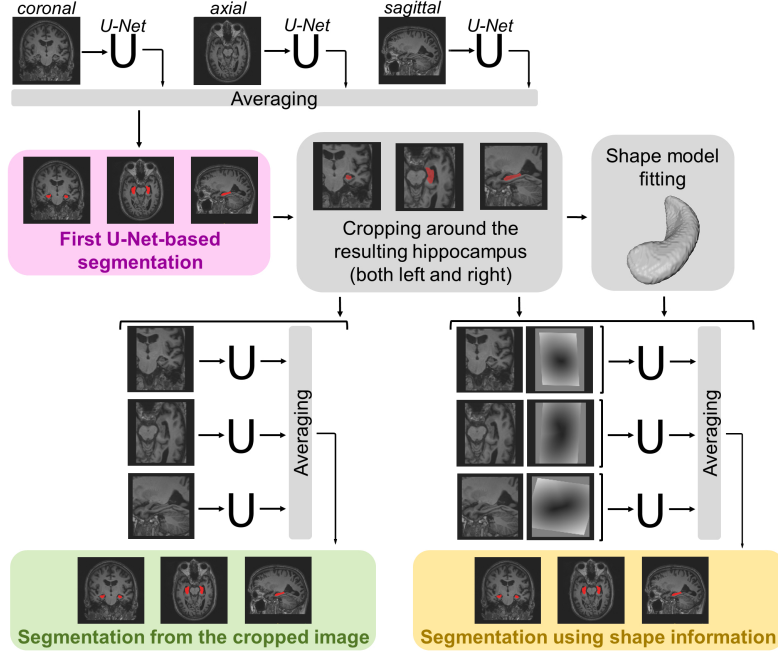


Figure 4.3: Summary of the methods for hippocampal segmentation proposed in Paper V. The segmentation results are highlighted in pink for MRI U-Net, in green for Cropped MRI U-Net and in yellow for Shape MRI U-Net. All three architectures are made up of three orthogonal 2D U-Nets (one per view) and receive T1w MRI slices as input. For Cropped and Shape MRI U-Net, though, such slices are cropped around the hippocampus segmented by MRI U-Net. Shape MRI U-Net also includes an additional input channel that embeds shape information obtained by fitting a hippocampal shape model. [© Brusini et al. (2020), reprinted with permission]

angle was obtained by calculating the angle between the two vectors that connect the endpoints of the midline to its center.

### Shape context for hippocampus segmentation

The hippocampal shape model in Paper V was generated using segmentations of four HC, four MCI and four AD subjects from the ADNI cohort, all scanned using 3.0T scanners. Also in this case, the PCA-based method described by Leventon et al. (2000) was used, but performed on 3D distance

transforms instead. Moreover, a unique model for both right and left hippocampus was generated. Therefore, all 12 right hippocampi were mirrored in order to be modelled. The images used to generate the model were also up-sampled to  $0.5 \times 0.5 \times 0.5 \text{ mm}^3$  voxel size (from  $1 \times 1 \times 1 \text{ mm}^3$ ) and cropped around the hippocampi's centers of gravity. Finally, PCA was performed on these 24 distance transforms, and the five main PCs of variation were preserved.

The above-described model can be fitted to the distance transform of the output of Cropped MRI U-Net. This fitted shape model constitutes the input shape context of Shape MRI U-Net.

#### 4.2.5 Method evaluation

Several data analysis methods were applied to evaluate the performance of the proposed tools as well as to find associations with clinical and diagnostic parameters.

In all three human studies, the Dice coefficient was utilized as the main metric of segmentation accuracy. In Paper V, precision, recall and Hausdorff distance were also computed. By performing Student's *t*-tests (in Paper III) and mixed-effects analysis of variance (ANOVA, in Paper V), it was possible to compare the resulting evaluation metrics across different approaches. For example, the difference in segmentation performance using T1w data versus FLAIR data was investigated in Paper III; in Paper V, single-cohort performance (using nine-fold cross-validation on ADNI data) was compared to the one obtained in the cross-cohort analysis (testing the trained networks on the AddNeuroMed cohort). In Paper III, ANOVA tests were also carried out to evaluate the accuracy of the implemented algorithms across the three available scanners.

Since ground truth was not available for the large ADNI test dataset used in the inference phase of Paper V, this was the only case in which segmentation accuracy could not be precisely estimated. Therefore, in this case, the correlation between the hippocampal volumes predicted by the proposed methods and those obtained using FreeSurfer 6.0 was instead computed.

Finally, changes in the performance across patients with different levels of neurodegeneration were investigated. In Paper III, this was done by dividing subjects into three different levels of CC atrophy (low, medium and high) and investigating differences in segmentation accuracy across these three groups using *t*-tests. Similarly, in Paper V, the segmentation performance was compared between HC, MCI and AD subjects. Moreover,

#### *4.2. DEEP LEARNING AND SHAPE ANALYSIS OF THE DEGENERATING HUMAN BRAIN*

differences in predicted hippocampal volume as well as their diagnostic prediction power were investigated between these three groups using one-way ANOVA tests and binary LR classification models, respectively.



## Chapter 5

# Results

### 5.1 Investigating group differences from brain MRI

#### 5.1.1 Paper I

CHANGES IN BRAIN ARCHITECTURE ARE CONSISTENT WITH ALTERED FEAR PROCESSING IN DOMESTIC RABBITS

In the first part of the study, the normalized volumes of all gray matter (GM) regions of interest (ROIs) were computed for each rabbit, and differences between domestic and wild rabbits were investigated. After Bonferroni correction, a significant difference between the two groups was found in two ROIs (see Figure 5.1). The amygdala showed a significant volumetric reduction ( $p < 0.01$ ) of  $-10.1\%$  and  $-8.7\%$  in domestic rabbits for the right and left amygdala, respectively. The medial frontal cortex presented a significant volumetric increase ( $p < 0.01$ ) of  $+12.1\%$  and  $+11.1\%$  in the right and left hemisphere, respectively.

Voxel-based morphometry (VBM) was later employed to analyze local voxel-level volumetric changes, which are independent from the ROI parcelations. Nevertheless, consistent results with the ROI-based analysis were observed. As shown in Figure 5.2, two main GM clusters were affected by significant group differences ( $p < 0.05$ ): one centered around the amygdala (with a domestication-induced GM loss) and one around the frontal cortex (with a domestication-induced GM increase). In particular, the amygdalar cluster was found to overlap mostly with the basolateral, central and lateral nuclei of the amygdala. Moreover, it partially extended into the hippocampus and both the entorhinal and piriform cortices. The cluster showing

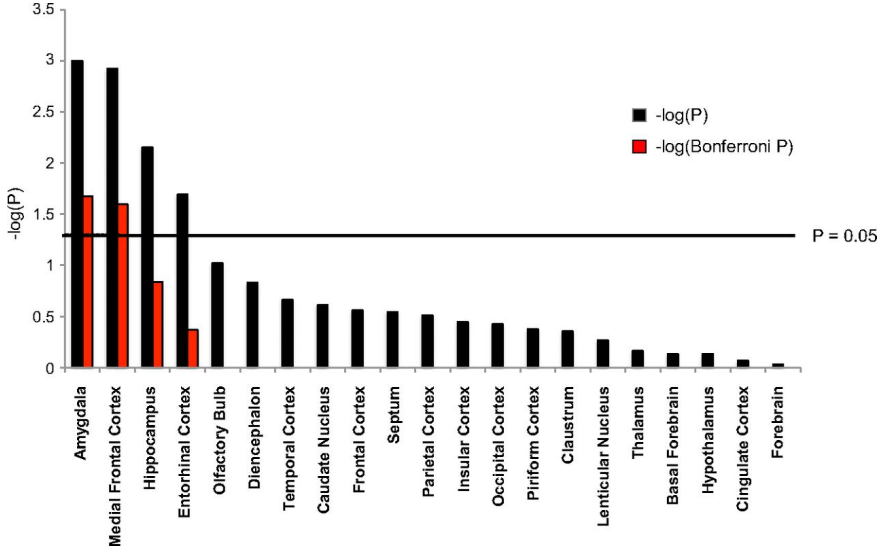


Figure 5.1: Results of the ROI-based analysis of Paper I, investigating differences in GM ROI volumes between domestic and wild rabbits. The results from the two hemispheres are merged together, while separate results for the left and right hemispheres can be found in the Supplementary Material of Paper I. The black bars show  $p$ -values based on 5 000 permutations  $t$ -tests. The red bars show Bonferroni-corrected  $p$ -values, which take the total number of comparisons into account. [© Brusini et al. (2018). Published by PNAS and reprinted with permission]

GM gain included mostly the medial frontal cortex as well as parts of the temporal and parietal cortices.

Finally, white matter (WM) changes were analyzed by running tract-based spatial statistics (TBSS) to identify local differences in fractional anisotropy (FA). In contrast to the GM analysis, very dispersed (rather than localized) significant changes were observed across the brain. In particular, wild rabbits showed a significantly higher FA ( $p < 0.05$ ) in 86% of the corona radiata volume, 69% of the subcortical WM and 71% of the corpus callosum (CC).

## 5.1. INVESTIGATING GROUP DIFFERENCES FROM BRAIN MRI

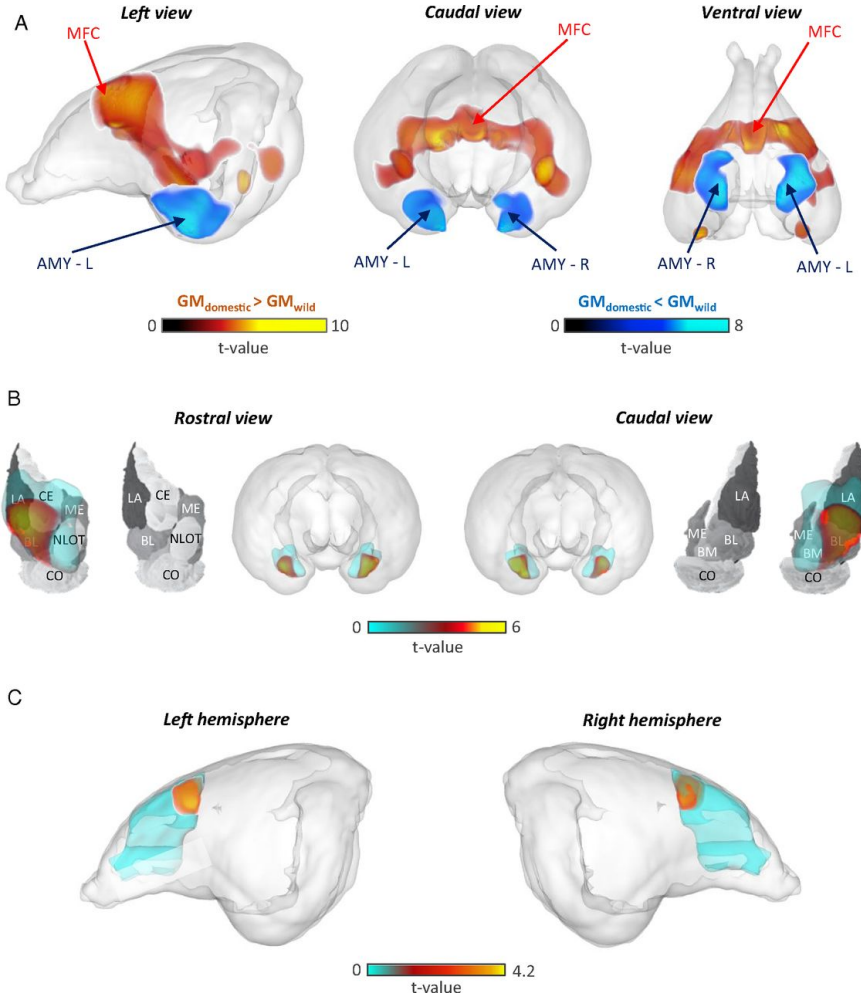


Figure 5.2: GM volume differences between domestic and wild rabbits obtained from VBM.  $T$ -values were estimated using threshold-free cluster enhancement (Smith and Nichols, 2009). (A) Smaller and larger GM volumes in domestic rabbits (compared to wild) are shown in blue and red, respectively. Two blue clusters intersect the left and right amygdala (AMY-L and AMY-R). The largest red cluster intersects the medial frontal cortex (MFC). Two smaller red clusters are not located completely inside the brain and intersect mostly superficial vessels, so they were disregarded. (B) Within the amygdala (in blue), the regions showing significant volume differences (in red) intersect the basolateral (BL), lateral (LA) and central (CE) nuclei. [Modified with permission from Równiak et al. (2007)]. (C) Within the MFC (in blue), significant differences (in red) were found dorsally. [© Brusini et al. (2018). Published by PNAS and reprinted with permission] 75



### 5.1.2 Paper II

#### MRI-DERIVED BRAIN AGE AS A BIOMARKER OF AGING IN RATS: VALIDATION USING A HEALTHY LIFESTYLE INTERVENTION

First, the performance of the three proposed brain age prediction models—based on Gaussian process regression (GPR) only, logistic regression (LR) classification only and their ensemble—was evaluated on a training cohort using leave-one-out cross-validation. The ensemble model showed the best performance, with a mean absolute error (MAE) of 4.87 weeks in the predictions and a linear correlation coefficient  $r = 0.92$  between predicted and chronological age, as opposed to  $\text{MAE} = 5.69$  and  $r = 0.92$  for GPR only, and  $\text{MAE} = 5.47$  and  $r = 0.80$  for LR only.

By analyzing the input features with the highest weights in the LR classifier, it was also found that the aging rat brain is affected by spread and unspecific variations of various magnitudes rather than very localized changes. Still, a few regions did seem to show slightly higher variations across different ages compared to others. For example, older rats were associated with a general increase in GM volume (especially in parts of the cerebellum, amygdala and hippocampus) as well as a simultaneous WM volume decrease in the cerebellar area and increase in large portions of the cerebrum.

In Figure 5.3, the predictions of the GPR+LR ensemble model on the aging cohort (including controls and EEDR rats) are presented. In general, a loss in accuracy was observed compared to the training cohort, with a MAE of 9.89 weeks and a correlation coefficient of 0.86. It is also noticeable how the age predictions on the EEDR cohort are generally lower, especially starting with the second scanning session. This result was analytically confirmed by fitting a linear mixed-effects model, which showed a statistically significant influence of both chronological age ( $p < 0.001$ ) and the interaction between lifestyle group (control or EEDR) and chronological age ( $p = 0.015$ ) on the subjects' BrainAGE scores.

Finally, survival analysis was performed by fitting Cox regression models. A significant effect on survival was found only in one of the analyzed cases: when using the BrainAGE score at Session 2 as the independent variable of the survival prediction model. In particular, higher BrainAGE scores at about five months of age were significantly associated with higher risk of mortality ( $p = 0.03$ ).

## 5.2. DEEP LEARNING AND SHAPE ANALYSIS OF THE DEGENERATING HUMAN BRAIN

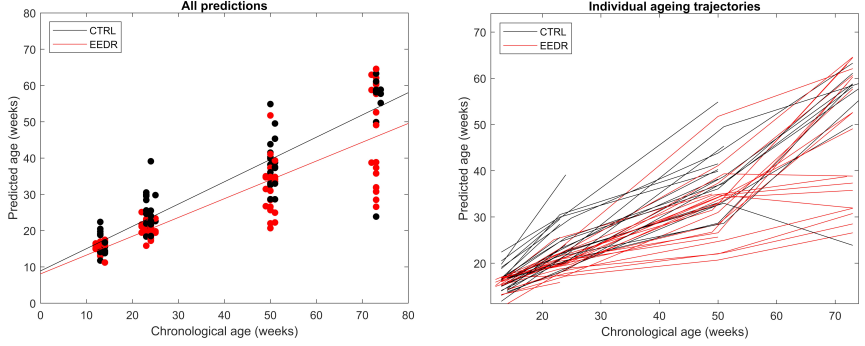


Figure 5.3: Brain age predictions obtained on the test cohort, including 24 control rats (in black) and 24 EEDR rats (in red). Left plot: each single prediction is shown as a black or red dot, while the lines represent the linear fitting for the two groups (slope of 0.61 and 0.52 for controls and EEDR, respectively). Right plot: the predicted aging trajectories are shown for all 48 subjects using the same color code. [From Brusini et al. (2022a), © 2022, *Elsevier*. Reprinted with permission]

## 5.2 Deep learning and shape analysis of the degenerating human brain

### 5.2.1 Paper III

#### AUTOMATIC DEEP LEARNING MULTICONTRAST CORPUS CALLOSUM SEGMENTATION IN MULTIPLE SCLEROSIS

In this work, three convolutional neural networks (CNNs) for automatic midsagittal slice selection were first developed: one specific for T1-weighted (T1w) data, one for FLAIR and one for both. For all three algorithms, over 98% of the predictions fell within 3 mm from the ground truth using both the 10-fold and scanner-wise cross-validation strategy. In the latter case, no significant performance difference was found between scanners.

Intrarater variability in the midslice selection task was also evaluated by computing the intraclass correlation coefficient (ICC) between two sets of manual annotations performed by the same rater on two different occasions. This resulted in an ICC of 0.999 and 0.993 for T1w and FLAIR images, respectively. ICCs of 0.991, 0.968 and 0.998 were observed between

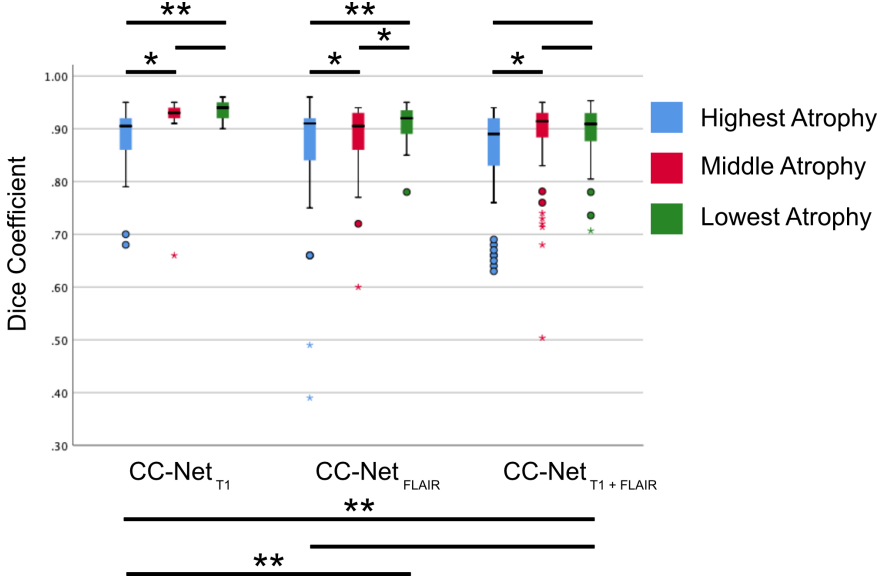


Figure 5.4: Clustered boxplot showing how the CC segmentation accuracy, expressed as Dice coefficient, significantly decreases in patients with higher atrophy using all three implemented architectures (CC-Net<sub>T1</sub>, CC-Net<sub>FLAIR</sub> and CC-Net<sub>T1/FLAIR</sub>). These results were obtained with 10-fold cross-validation on the training cohort. All patients were divided into three atrophy levels, depending on whether their normalized CC area was in the top, middle or bottom third of the cohort. \* refers to  $p < 0.05$ , \*\* to  $p < 0.01$ . [© Brusini et al. (2022b). *Journal of Neuroimaging* published by Wiley Periodicals LLC on behalf of American Society of Neuroimaging. Reprinted with permission]

ground truth annotations and the results from midCNN<sub>T1</sub>, midCNN<sub>FLAIR</sub> and midCNN<sub>T1+FLAIR</sub> using 10-fold cross-validation.

Three intracranial (IC) and three CC segmentation networks were also developed. All IC alternatives showed consistent accuracy. Using 10-fold cross-validation, the overall mean Dice coefficient ranged from 0.970 (for IC-Net<sub>T1</sub>) to 0.978 (for IC-Net<sub>T1/FLAIR</sub>). With scanner-wise cross-validation, the Dice ranged from 0.965 (for IC-Net<sub>FLAIR</sub>) to 0.974 (for IC-Net<sub>T1</sub>), and no significant difference was found between scanners.

Different patterns were observed for CC segmentation. When performing 10-fold cross-validation, the T1-specific architecture showed a significantly higher ( $p < 0.05$ ) Dice coefficient (equal to  $0.91 \pm 0.05$ ) compared to both

## 5.2. DEEP LEARNING AND SHAPE ANALYSIS OF THE DEGENERATING HUMAN BRAIN

Table 5.1: Correlations between disability scores (EDSS and SDMT) and normalized CC measures from both FreeSurfer (applied on T1w data only) and the proposed DL-based pipelines. Since FreeSurfer provides a 3D segmentation of the CC, normalized CC volume (nCCV) was computed in that case. For all other configurations using the proposed pipelines for 2D segmentation, normalized CC area was used instead.  $\rho$  is the Spearman’s correlation coefficient and  $r$  the Pearson’s correlation coefficient. Significant correlations are indicated with \* (for  $p < 0.05$ ) and \*\* (for  $p < 0.01$ ). [Adapted from © Brusini et al. (2022b) with permission. *Journal of Neuroimaging published by Wiley Periodicals LLC on behalf of American Society of Neuroimaging*]

Pipeline		EDSS (+/- 6 months) n=252	EDSS future (6.7 +/- 2.6 years) n = 331	SDMT (+/- 6 months) n=172	SDMT future (5.8 +/- 2.8 years) n = 304
FreeSurfer (nCCV)		$\rho = -0.19$ **	$\rho = -0.18$ **	$r = 0.18$ *	$r = 0.24$ **
T1w-specific		$\rho = -0.15$ *	$\rho = -0.18$ **	$r = 0.12$	$r = 0.18$ **
FLAIR-specific		$\rho = -0.13$ *	$\rho = -0.19$ **	$r = 0.18$ **	$r = 0.28$ **
Combined T1w and FLAIR	On T1w data	$\rho = -0.12$	$\rho = -0.18$ **	$r = 0.12$	$r = 0.21$ **
	On FLAIR data	$\rho = -0.21$ **	$\rho = -0.24$ **	$r = 0.25$ **	$r = 0.29$ **

CC-Net<sub>FLAIR</sub> ( $0.881 \pm 0.091$ ) and CC-Net<sub>T1/FLAIR</sub> ( $0.88 \pm 0.074$ ). A superior performance on T1w scans compared to FLAIR was observed also with scanner-wise cross-validation: Dice coefficients of  $0.902 \pm 0.065$  and  $0.894 \pm 0.062$  were obtained on T1w data using CC-Net<sub>T1</sub> and CC-Net<sub>T1/FLAIR</sub>, respectively, compared to  $0.828 \pm 0.11$  and  $0.808 \pm 0.149$  on FLAIR scans using CC-Net<sub>FLAIR</sub> and CC-Net<sub>T1/FLAIR</sub>, respectively. Furthermore, when segmenting FLAIR data using CC-Net<sub>T1/FLAIR</sub>, a significant difference in performance ( $p < 0.01$ ) was found between scanners, with the Trio scanner showing the highest accuracy (mean Dice of 0.852).

In the 10-fold cross-validation analysis, the CC segmentation performance was also compared between patients having different levels of atrophy. The segmentations were found to be significantly more accurate in subjects with lower CC atrophy in all three network implementations (see Figure 5.4).

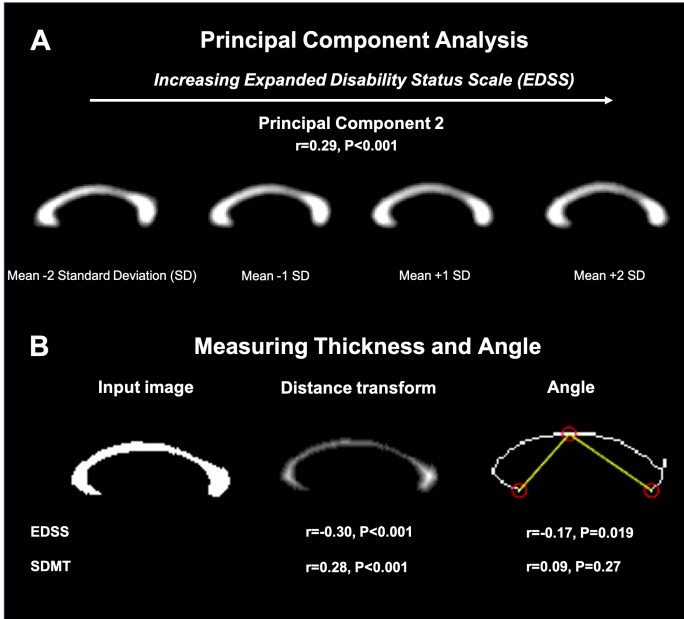


Figure 5.5: CC shape changes that are significantly correlated with disability. (A) The second principal component of shape variation (from PCA) significantly correlates with the EDSS physical disability score. (B) Mean CC thickness (computed as the average intensity along the midline of the CC distance transform) significantly correlates with both EDSS (physical) and future SDMT (cognitive) disability scores. The CC bending angle (i.e., the angle between the two yellow vectors in the figure) significantly correlates with EDSS. [© Platten et al. (2021). *Journal of Neuroimaging* published by Wiley Periodicals LLC on behalf of American Society of Neuroimaging. Reprinted with permission]

The reproducibility of the proposed pipeline (including both midslice selection and subsequent segmentation on the selected slice) was tested on nine subjects who were scanned on three different scanners on the same day. Excellent reproducibility was achieved with the T1-specific pipeline ( $ICC = 0.942$ ) as well as with the pipeline tailored for both T1w and FLAIR data when tested on T1w data only ( $ICC = 0.908$ ). The reliability of the algorithms on FLAIR data was moderate to good ( $ICC$  of 0.739 and 0.753 for the FLAIR-specific and the T1/FLAIR pipelines, respectively).

Finally, the proposed algorithms were tested again on a separate set of

## 5.2. DEEP LEARNING AND SHAPE ANALYSIS OF THE DEGENERATING HUMAN BRAIN

patients for which physical (EDSS) and cognitive (SDMT) disability scores were recorded both at baseline (within six months from the scan) and in the future (with follow-up time of  $6.7 \pm 2.6$  and  $5.8 \pm 2.8$  years for EDSS and SDMT, respectively). The correlations between normalized CC measures and all available disability scores are presented in Table 5.1.

### 5.2.2 Paper IV

#### DEEP LEARNING CORPUS CALLOSUM SEGMENTATION AS A NEURODEGENERATIVE MARKER IN MULTIPLE SCLEROSIS

In this study, U-Net-based architectures similar to those presented in Paper III were employed for CC and IC segmentation, although from only T2-weighted images. Additionally, shape analysis of the CC segmentation results—more specifically, of their distance transforms—was performed in order to investigate the correlation between shape changes and disability.

First, the first 13 principal components (PCs) obtained from principal component analysis (PCA) were preserved as they represented 99% of the total variance. The second PC was found to significantly correlate with EDSS after Bonferroni correction (correlation coefficient  $r = 0.29$ ,  $p < 0.001$ ). In particular, this component seems to reflect a change towards a thinner and more bent CC with higher degrees of disability (see Figure 5.5).

Both mean thickness and bending angle of the subjects' CC were also extracted. EDSS scores were found to significantly correlate with both those shape properties:  $r = -0.30$  and  $p < 0.001$  for thickness,  $r = -0.17$  and  $p = 0.043$  for bending angle (see Figure 5.5). Moreover, normalized future SDMT  $z$ -scores significantly correlated with CC thickness ( $r = 0.28$ ,  $p < 0.001$ ).

### 5.2.3 Paper V

#### SHAPE INFORMATION IMPROVES THE CROSS-COHORT PERFORMANCE OF DEEP LEARNING-BASED SEGMENTATION OF THE HIPPOCAMPUS

#### Single-cohort versus cross-cohort performance

In the first part of the work, a single-cohort analysis (i.e., on ADNI data only using cross-validation) was carried out to compare the performance of

## CHAPTER 5. RESULTS

the three proposed networks for hippocampal segmentation. As shown in Table 5.2, no relevant differences were observed between the presented MRI U-Net, Cropped MRI U-Net and Shape MRI U-Net. Rather consistent accuracy, although slightly numerically inferior, was achieved also when using two other context-based networks that were implemented for comparison: Tissue MRI U-Net and Autocontext MRI U-Net (see Paper V for further details). However, all these deep learning-based architectures showed a superior accuracy compared to FreeSurfer 6.0. Furthermore, using the three proposed networks, a small loss in performance was observed on the images from Alzheimer’s disease (AD) patients, with an average decrease in the Dice coefficient between -1% and -2% with respect to healthy controls (HC) and subjects with mild cognitive impairment (MCI).

Subsequently, a cross-cohort analysis was carried out by testing the above-described networks on data from the AddNeuroMed cohort. In this case, MRI U-Net was affected by a dramatic loss in performance, reaching an accuracy close to that of FreeSurfer 6.0. Compared to MRI U-Net, an improvement was obtained using Cropped MRI U-Net (+5.35% and +2.7% in the average Dice coefficient for the left and right hippocampus, respectively) and even more using Shape MRI U-Net (+5.83% and +4.04%). Autocontext MRI U-Net achieved instead an accuracy between that of MRI U-Net and Cropped MRI U-Net, while Tissue MRI U-Net performed similarly to Cropped MRI U-Net (except for the recall, which was consistently lower compared to Cropped MRI U-Net). Additionally, in this cross-cohort analysis, AD cases showed the lowest Dice coefficients on average.

In order to further support the hypothesis that shape context information can improve cross-cohort performance, a new cross-cohort analysis was performed by switching the two available cohorts, that is by using the AddNeuroMed dataset as the training set and the ADNI one as the test set. The results were consistent with those presented above for the first cross-cohort investigation, with Shape MRI U-Net showing superior segmentation accuracy compared to the other methods (see Paper V for further details).

### Validation on a larger dataset

The MRI U-Net and Shape MRI U-Net trained on the 54 ADNI subjects were tested one last time on a new dataset made up of 5948 additional ADNI images (not used during training). On these data, rather high and consistent correlation coefficients (from 0.879 to 0.892) were observed between FreeSurfer-based hippocampal volumes and those obtained using both

## 5.2. DEEP LEARNING AND SHAPE ANALYSIS OF THE DEGENERATING HUMAN BRAIN

Table 5.2: Comparison of the performance of six different segmentation methods analyzed in Paper V: MRI U-Net, Cropped MRI U-Net, Shape MRI U-Net, Tissue MRI U-Net, Autocontext MRI U-Net and FreeSurfer 6.0. The performance is reported for both the left and the right hippocampus and expressed as mean  $\pm$  standard deviation of four different evaluation metrics: Dice coefficient, precision, recall and 95% percentile Hausdorff distance. For each metric, the results from both the single cohort and the first cross-cohort analysis are reported for comparison. The single cohort metrics were obtained on the ADNI dataset and all five U-Net-based architectures were trained using 9-fold cross-validation. The cross-cohort metrics were obtained on the AddNeuroMed dataset and all five U-Net-based architectures were trained previously on the ADNI dataset. [Modified from © Brusini et al. (2020) with permission]

Region of interest	Segmentation method	Dice coefficient		Precision		Recall		Hausdorff distance (in voxels)	
		Single cohort	Cross-cohort	Single cohort	Cross-cohort	Single cohort	Cross-cohort	Single cohort	Cross-cohort
Left hippocampus	MRI U-Net	90.17 $\pm$ 1.44%	79.09 $\pm$ 2.63%	89.46 $\pm$ 2.20%	74.72 $\pm$ 4.27%	90.96 $\pm$ 2.29%	84.23 $\pm$ 3.15%	2.33 $\pm$ 0.55	3.44 $\pm$ 0.74
	Cropped MRI U-Net	90.28 $\pm$ 1.30%	84.44 $\pm$ 2.32%	89.36 $\pm$ 2.20%	78.47 $\pm$ 4.17%	91.28 $\pm$ 1.85%	91.60 $\pm$ 2.47%	2.22 $\pm$ 0.53	3.19 $\pm$ 0.64
	Shape MRI U-Net	90.01 $\pm$ 1.41%	84.92 $\pm$ 2.56%	88.55 $\pm$ 2.69%	79.46 $\pm$ 5.03%	91.60 $\pm$ 1.87%	91.57 $\pm$ 3.60%	2.35 $\pm$ 0.67	3.16 $\pm$ 0.77
	Tissue MRI U-Net	88.79 $\pm$ 1.61%	84.32 $\pm$ 2.16%	86.79 $\pm$ 2.72%	79.04 $\pm$ 4.12%	91.01 $\pm$ 2.94%	90.59 $\pm$ 2.90%	2.39 $\pm$ 0.60	3.33 $\pm$ 0.85
	Autocontext MRI U-Net	89.45 $\pm$ 1.46%	80.55 $\pm$ 2.61%	86.69 $\pm$ 2.77%	73.99 $\pm$ 4.23%	92.53 $\pm$ 2.75%	88.67 $\pm$ 3.50%	2.30 $\pm$ 0.57	3.33 $\pm$ 0.72
Right hippocampus	FreeSurfer 6.0	79.52 $\pm$ 3.14%	79.41 $\pm$ 3.77%	82.94 $\pm$ 5.01%	78.89 $\pm$ 5.46%	76.60 $\pm$ 3.94%	80.20 $\pm$ 4.39%	4.34 $\pm$ 1.08	4.24 $\pm$ 1.25
	MRI U-Net	90.12 $\pm$ 1.41%	80.15 $\pm$ 2.25%	89.59 $\pm$ 2.48%	74.54 $\pm$ 3.12%	90.77 $\pm$ 2.72%	86.80 $\pm$ 3.08%	2.39 $\pm$ 0.53	3.92 $\pm$ 1.14
	Cropped MRI U-Net	90.26 $\pm$ 1.41%	82.85 $\pm$ 2.52%	89.29 $\pm$ 2.62%	75.97 $\pm$ 3.91%	91.35 $\pm$ 2.39%	91.27 $\pm$ 2.31%	2.47 $\pm$ 0.59	3.80 $\pm$ 1.05
	Shape MRI U-Net	90.08 $\pm$ 1.67%	84.19 $\pm$ 2.50%	88.50 $\pm$ 3.39%	77.94 $\pm$ 4.49%	91.86 $\pm$ 2.34%	91.83 $\pm$ 3.28%	2.54 $\pm$ 0.79	3.62 $\pm$ 1.04
	Tissue MRI U-Net	88.74 $\pm$ 1.50%	82.88 $\pm$ 2.35%	86.40 $\pm$ 2.84%	76.86 $\pm$ 3.60%	91.00 $\pm$ 3.16%	90.08 $\pm$ 2.71%	2.63 $\pm$ 0.70	3.68 $\pm$ 1.11
	Autocontext MRI U-Net	89.63 $\pm$ 1.32%	80.51 $\pm$ 2.20%	87.25 $\pm$ 2.81%	73.08 $\pm$ 3.27%	92.30 $\pm$ 2.89%	89.79 $\pm$ 3.03%	2.39 $\pm$ 0.53	3.88 $\pm$ 1.16
	FreeSurfer 6.0	80.21 $\pm$ 3.86%	79.57 $\pm$ 3.54%	83.63 $\pm$ 4.35%	77.71 $\pm$ 5.53%	77.31 $\pm$ 5.36%	81.78 $\pm$ 3.40%	4.50 $\pm$ 1.23	4.61 $\pm$ 1.11



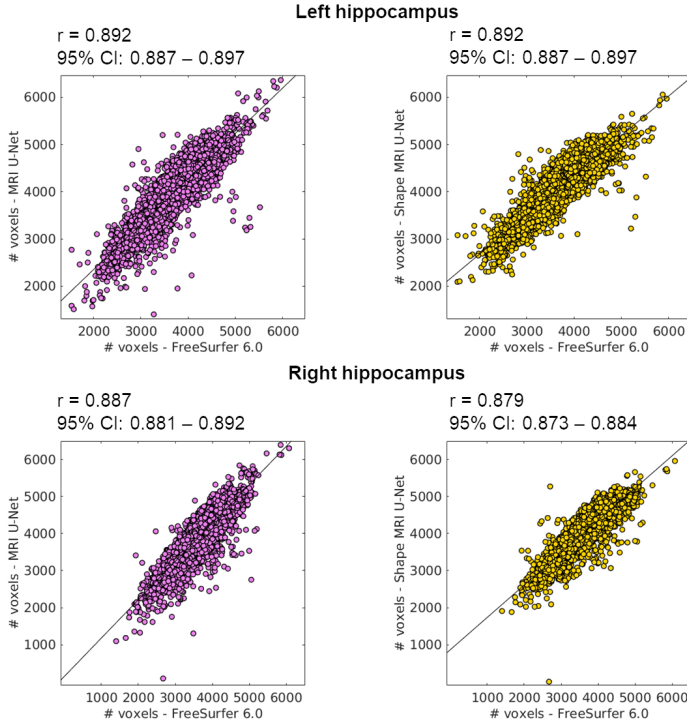


Figure 5.6: The hippocampal volumes (expressed as number of voxels) obtained using the proposed methods (MRI U-Net on the right and Shape MRI U-Net on the left) are plotted against those from FreeSurfer 6.0, for left (top) and right (bottom) hippocampus separately. The correlation coefficient  $r$  between these measures is also provided, as well as its 95% confidence interval (CI). [© Brusini et al. (2020). Reprinted with permission]

deep learning-based architectures (see Figure 5.6). Higher correlation coefficients of 0.952 and 0.958 for left and right hippocampus, respectively, were achieved between MRI U-Net and Shape MRI U-Net.

An expert rater was then asked to compare the quality of deep learning-based versus FreeSurfer-based segmentations on 14 images, which were all from either AD or MCI subjects. These images represented cases in which a high difference in hippocampal volume was observed for both MRI U-Net and Shape MRI U-Net compared to FreeSurfer. In particular, the expert rater compared FreeSurfer results with those from MRI U-Net, which was chosen as the reference deep learning-based method for this evaluation.

## 5.2. DEEP LEARNING AND SHAPE ANALYSIS OF THE DEGENERATING HUMAN BRAIN

In all 14 cases, the FreeSurfer segmentations presented evident segmentation errors. With MRI U-Net, three out of 14 cases were considered good segmentations, five out of 14 were inaccurate but better than FreeSurfer, while the remaining six segmentations were considered as erroneous as those from FreeSurfer. Some examples of these examined segmentation results are shown in Figure 6 of Paper V.

The results from the large ADNI test dataset were also employed to investigate differences in the estimated normalized hippocampal volumes (i.e., hippocampal volume divided by IC volume) across the three available diagnostic groups (HC, MCI and AD). As shown in Table 5.3, statistically significant differences between these groups were found for all tested segmentation methods (MRI U-Net, Shape MRI U-Net and FreeSurfer), with AD patients presenting the most atrophic hippocampi, followed by MCI and HC. Moreover, the diagnostic prediction power of the computed normalized volumes was investigated by fitting binary LR models. All three segmentation methods showed a good prediction power in distinguishing AD subjects from HC (AUROC consistently above 0.80). However, lower performance was achieved for automatic classification of AD versus MCI (AUROC = 0.68 for all three segmentation methods) and for MCI versus HC (AUROC of 0.67, 0.65 and 0.64 for MRI U-Net, Shape MRI U-Net and FreeSurfer, respectively).

Table 5.3: Comparison of hippocampal volume across the three diagnostic groups of interest: Alzheimer’s disease (AD), mild cognitive impairment (MCI) and healthy controls (HC). The hippocampal volume of each subject was divided by their total intracranial volume (ICV) and later multiplied by the average ICV of the present dataset. Volumetric results were calculated at baseline on patients who did not change diagnosis within two years, and they are reported as mean  $\pm$  standard deviation. For each row (i.e., each method and region of interest), a one-way ANOVA test was performed to check for significance in the differences between the three diagnostic groups. [© Brusini et al. (2020). Reprinted with permission]

Region of interest	Segmentation method	AD (n = 93)	MCI (n = 267)	HC (n = 154)	p-value (one-way ANOVA)
Left hippocampus	MRI U-Net	$3.54 \pm 0.66\text{cm}^3$	$3.95 \pm 0.60\text{cm}^3$	$4.30 \pm 0.54\text{cm}^3$	$p < 0.001$
	Shape MRI U-Net	$3.67 \pm 0.60\text{cm}^3$	$4.04 \pm 0.59\text{cm}^3$	$4.39 \pm 0.51\text{cm}^3$	$p < 0.001$
	FreeSurfer 6.0	$3.25 \pm 0.60\text{cm}^3$	$3.60 \pm 0.63\text{cm}^3$	$3.99 \pm 0.58\text{cm}^3$	$p < 0.001$
Right hippocampus	MRI U-Net	$3.38 \pm 0.66\text{cm}^3$	$3.81 \pm 0.66\text{cm}^3$	$4.25 \pm 0.54\text{cm}^3$	$p < 0.001$
	Shape MRI U-Net	$3.56 \pm 0.61\text{cm}^3$	$3.93 \pm 0.62\text{cm}^3$	$4.34 \pm 0.50\text{cm}^3$	$p < 0.001$
	FreeSurfer 6.0	$3.16 \pm 0.59\text{cm}^3$	$3.50 \pm 0.63\text{cm}^3$	$3.89 \pm 0.54\text{cm}^3$	$p < 0.001$



## Chapter 6

# Discussion

### 6.1 The important role of brain MRI image analysis

The studies presented in this thesis all revolve around a common thread: the use of MRI image analysis methods to analyze and characterize brain structure. Thanks to their good contrast and spatial resolution, structural brain MRI scans were successfully used as input to multiple processing pipelines, which allowed extracting important morphological information. The results of these studies support the great value of MRI image analysis as an essential tool for the advancement of neuroscientific knowledge.

#### 6.1.1 MRI image analysis for better understanding of the brain

As described in the background Section 3.1.2, the brain is the most complex organ in the human body, and there is still much to discover about it on multiple levels. Structural MRI image analysis can provide further insights on the changes the brain undergoes due to multiple factors, such as environmental, evolutionary, genetic and pathological.

#### **The effect of domestication on the rabbit brain**

In Paper I, brain changes of an evolutionary nature were investigated by analyzing a unique dataset comprising scans from domestic and wild rabbits. In a previous study, domestication was found to induce genetic changes in the vicinity of genes that modulate brain and/or neural development (Carneiro et al., 2014). Therefore, brain MRI scans of these animals could be employed to investigate whether any macroscopic phenotypic changes can be identified.

## CHAPTER 6. DISCUSSION

Very localized gray matter (GM) changes were found in the domestic group, which showed a decreased amygdalar volume and an enlarged medial frontal cortex compared to their wild counterparts. The amygdala is known for being involved in fear conditioning. Therefore, an enlarged amygdala can facilitate fear expression in wild rabbits when exposed to environmental threats. Furthermore, the amygdalar areas most subject to changes corresponded to the basolateral nucleus, which is responsible for receiving and processing sensory stimuli (Agren et al., 2012), as well as parts of the central and lateral nuclei, which control efferent and afferent processing, respectively. This suggests that domestication has mostly affected areas responsible not only for detecting fear stimuli but also for expressing fear responses. This finding is in agreement with the simultaneous enlargement of the medial prefrontal cortex since this part of the brain controls emotions and reduces fear responses (Urry et al., 2006; Winkelmann et al., 2016). Moreover, aggressive behavior is known for being strongly modulated by the prefrontal-amygdala fear network (Davidson et al., 2000). Therefore, the findings of Paper I reflect well the behavioral components characteristic of domestic rabbits. It is reasonable to think that constant contact with humans and a quieter life in captivity are responsible for these evolutionary changes. Conversely, wild rabbits must maintain an attentive and aggressive behavior that is more suitable for wildlife.

In contrast to what was observed for the GM, domestication-induced white matter (WM) changes were far more dispersed across the brain, with a general reduction of fractional anisotropy (FA) in the brain of domestic rabbits. This affects largely all types of fibers: association fibers (connecting different areas of the brain), projection fibers (connecting the brain with other parts of the central nervous system) and commissural fibers (connecting the two hemispheres). This finding suggests that domestication may compromise the communication between these areas in terms of information processing capacity and speed.

As stated by Charles Darwin, “Hardly any animal is more difficult to tame than the young of the wild rabbit; scarcely any animal is tamer than the young of the tame rabbit” (Darwin, 1859). Thanks to the use of brain MRI image analysis, it was possible to gain a better understanding of why this phenomenon occurs on a brain structural level.

### **The effect of a healthy lifestyle on rat brain aging**

In Paper II, rat brain MRI scans were analyzed to understand the influence of lifestyle factors on brain aging. In particular, the effect of environmental enrichment and dietary restriction (EEDR) was investigated. This type of intervention had already shown a positive impact on the overall health of the EEDR subjects compared to controls, since they exhibited significantly lower levels of frailty on average (see Paper II for more details on these results). Brain MRI analysis can allow focusing specifically on the impact that this lifestyle change has on the brain.

After applying a pretrained age prediction model, the two lifestyle groups of interest showed diverging aging trajectories, with the EEDR group having younger predicted ages on average compared to controls. This difference between groups increased over time, a result that was further confirmed by fitting a linear mixed-effects model that showed that the interaction between lifestyle group and chronological age had a significant effect on the BrainAGE score (i.e., the difference between predicted and chronological age). As mentioned in Section 3.1.4, the combination of multiple lifestyle changes (a good diet, physical exercise and a stimulating environment) is believed to have a positive effect on brain aging. However, to the author's knowledge, Paper II constitutes the first study on rats—and on animals in general—that attempts to investigate the effect of a healthy lifestyle intervention on brain age prediction. The results of this work reflect well what has already been speculated in previous studies on humans. In particular, they are in accordance with very recent findings by Bittner et al. (2021) that show how multidomain lifestyle behaviors on humans (i.e., physical exercise, social interactions, smoking and alcohol consumption) have a significant effect on the BrainAGE score.

Furthermore, the analysis of the coefficients of the logistic regression (LR) model for age prediction shed light on what areas of the brain most influence the model predictions and thus are affected by more changes throughout the study observation period. Rather than very specific and localized variations, widespread changes were found to affect the whole brain. This is consistent with findings from previous studies on humans (Good et al., 2001; Walhovd et al., 2005), and it is believed that these complex and dispersed variations are the result of multiple factors (e.g., environmental, metabolic or immune) that vary across time and influence the brain in a nonlinear way.

To conclude, in Paper II, MRI images were successfully used for mod-

eling rat brain aging, analyzing how their brains change through time and understanding whether lifestyle interventions can have a direct impact on brain aging. The findings of this study strengthen the idea that lifestyle-related prevention methods can slow brain aging in rats, and this hypothesis may possibly be extended to other species (including humans).

### **The association between corpus callosum shape and disability in multiple sclerosis patients**

As described in Section 3.1.4, the corpus callosum (CC) is a WM structure that is greatly affected by multiple sclerosis (MS). CC atrophy is commonly observed in MS patients, and it can be numerically estimated by calculating the ratio between CC area and intracranial (IC) area. However, other features can also be explored. In Paper IV, shape analysis was carried out on the automatic CC segmentations of a dataset of MS patients. It was found that the CC became thinner with greater cognitive and physical disability. Moreover, more angled CCs were observed in patients with higher levels of physical disability.

In a previous study by Sigirli et al. (2012), disease progression was found to be associated with changes in CC shape. Shape differences were indeed identified not only between healthy controls and MS patients but also between two different clinical manifestations of MS: relapsing-remitting MS (where patients have periods of stability between different episodes of severe MS symptoms) and secondary progressive MS (disability worsens in a more continuous manner). For the latter condition, the CC was shown to present more pronounced deformations, especially in its anterior part. Moreover, in a work by Van Schependom et al. (2017), significant differences were found between healthy controls and MS patients in terms of both thickness profile and CC index, which is a measure that corresponds to the ratio between thickness and curvature. Therefore, the outcome of Paper IV further supports the existence of CC shape changes in MS that relate to the patient's clinical condition and disease progression.

On one hand, the identification of CC thinning with increasing disability matches well with the established knowledge on neurodegeneration and CC atrophy in MS. On the other hand, this work provides new insights on the relation between CC shape and disability with the identification of a significant correlation between the structure's bending angle and physical disability scores. With MS progression and the subsequent atrophy of the brain, ventricular enlargement may occur. Thus, it may be speculated that

## 6.1. THE IMPORTANT ROLE OF BRAIN MRI IMAGE ANALYSIS

this phenomenon can displace the CC upwards, making it appear more angled. Such a hypothesis may be further addressed in future MRI image analysis studies in order to provide more useful insights on the progression of this disease and on how to better interpret MRI scans from MS patients.

### 6.1.2 MRI-based biomarkers

In addition to being a useful tool for gaining new knowledge about the brain, another important role of MRI image analysis consists in the extraction of imaging biomarkers, which are measurable indicators of normal or pathogenic processes as well as pharmacologic responses. As described in Sections 3.1.4 and 3.4.6, extensive literature exists on the computation of features significantly associated with the presence of disease, including Alzheimer’s disease (AD) and MS, as well as with positive and healthy trends in an individual’s aging process. Such research opens up multiple possibilities for future clinical use of MRI-derived biomarkers for early diagnosis and monitoring of diseases. In this section, the results of the present thesis that relate to neuroimaging biomarker research are discussed.

### BrainAGE as a marker of healthy aging in rats

As discussed in Section 3.4.6, the BrainAGE score has been previously proposed as a brain aging biomarker, given the strong association between its value and multiple clinical conditions. In particular, higher and lower BrainAGE scores correspond to negative and positive trends in the aging process, respectively. To the best of the author’s knowledge, Paper II constitutes the first preclinical validation of BrainAGE as a biomarker of healthy aging.

The first important outcome of this study was the implementation of an accurate age prediction model from a training cohort of 31 normal rats (89 longitudinal scans in total). Similar accuracy was achieved on another rat study by Franke et al. (2016) in which a different dataset was employed (273 scans from 24 rats). These consistent outcomes support the potential of using brain age prediction models not only on humans but also on laboratory animals. This can allow investigating BrainAGE as a potential biomarker in a preclinical framework, which is of fundamental importance for both validating BrainAGE as an aging biomarker and studying how it is affected by active interventions performed during long observation periods.

The second part of the study consisted in analyzing the effect of a long-



## CHAPTER 6. DISCUSSION

term EEDR intervention on the BrainAGE score. As mentioned previously, the brains of the EEDR rats were found to age at a slower pace compared to controls. This finding is of high relevance not only to understand the effects of a healthy lifestyle on the brain (as discussed in Section 6.1.1) but also to support the value of BrainAGE as a sensitive marker of healthy aging and lifestyle. In other words, when a relevantly young brain age is predicted in a subject, it could be inferred not only that a neuropsychiatric disorder and/or neurodegeneration may be absent, but also that the brain is aging in a slower and healthier way compared to the normal population. However, many more preclinical studies should be performed to confirm this hypothesis.

The potential of BrainAGE as an important aging biomarker was strengthened by finding a significant effect of the BrainAGE score at about five months of age (approximately two months after the intervention started) on the rats' survival. This result was in agreement with a previous work by Cole et al. (2018), who also found a significant association between BrainAGE and survival in humans. This constitutes more evidence of the similarities between animal and human studies and therefore of the importance of conducting thorough preclinical research while controlling for multiple genetic, environmental or lifestyle factors that are more difficult to account for in human studies.

Finally, one of the strengths of Paper II is the fact that brain age predictions are analyzed longitudinally. At least two scans per test subject were available, and this allowed investigating differences in rate of aging between the two examined lifestyle groups. It has already been observed—both in previous studies (Franke et al., 2016; Liang et al., 2019; Cole et al., 2017b; Pardoe and Kuzniecky, 2018) and in Paper II itself—that the accuracy of the model can vary across chronological ages, with a general tendency to overestimate brain age at younger chronological ages and underestimate it at older ages. Prediction errors at older chronological ages are also characterized by higher variability. For this reason, it is crucial to control at least for age when studying the value of BrainAGE as an aging biomarker. An even better alternative would consist in examining aging trajectories across time for every subject. Similar observations were drawn in a recent work by Vidal-Pineiro et al. (2021), who compared results from cross-sectional and longitudinal brain age. No significant association was found between these two approaches, and the readers are advised to interpret with caution the results from cross-sectional studies (which can be influenced by

## 6.1. THE IMPORTANT ROLE OF BRAIN MRI IMAGE ANALYSIS

subject-specific congenital factors), while longitudinal approaches should be preferred.

### Segmentation area and volume as markers of neurodegeneration

Given the association between CC atrophy and MS as well as between hippocampal atrophy and AD, measures of their size can be extracted from their segmentations and investigated as possible imaging biomarkers of neurodegeneration. Papers III to V aim to contribute to this area of study.

In Paper III, deep learning (DL)-based architectures are proposed for automatic CC segmentation from both T1w and FLAIR data. The models were used to calculate normalized CC areas on a dataset of MS patients for which EDSS (physical) and SDMT (cognitive) disability scores were available both at baseline and at follow-up (measured about six years after the MRI scan). The areas measured from all implemented algorithms on both types of MRI sequences consistently showed significant correlations with both of the future disability scores. A similar analysis was carried out in Paper IV using T2w scans instead. Also in this case, the normalized CC area showed a significant correlation with both future EDSS and SDMT (measured about 8-10 years later on average) as well as with baseline EDSS.

As discussed in Section 6.1.1, the significant correlations between CC atrophy and disability are in agreement with previous studies, as well as with general background knowledge on the neurodegeneration that characterizes MS progression. However, the most remarkable—and not necessarily expected—results are the significant correlations found with *future* disability scores, which were calculated several years after the patients were scanned. These results are also consistent with the findings of another study by Ouellette et al. (2018) in which the normalized CC area—together with the lesion area—was found to significantly correlate with cognitive disability after 8.5 years. These findings are of great relevance because they strongly support the potential of using neuroimaging for predicting disease outcome. This would also allow better investigating prevention methods and designing personalized treatment plans.

The topic of hippocampal segmentation is addressed in Paper V, where a novel DL-based model (i.e., the Shape MRI U-Net) is proposed for obtaining accurate segmentations and, subsequently, measures of hippocampal atrophy from 3D T1w scans. The algorithm was tested on a test dataset of over 500 cases including healthy controls (HC), subjects with mild cognitive impairment (MCI) and AD patients. The derived hippocampal volumes were

## CHAPTER 6. DISCUSSION

shown to be significantly different between these three diagnostic groups, supporting the possibility of using automatic tools for identifying patterns of atrophy.

LR classifiers were trained using the subjects' hippocampal volumes as input in order to investigate their power in predicting the diagnostic group of every subject. The highest performance was achieved when discriminating HC from AD cases (AUROC of 0.84), for which the difference in atrophy is the largest. This metric dropped to 0.68 and 0.65 for AD vs. MCI and MCI vs. HC, respectively. This discrepancy in classification performance, which varied according to the class pairs to be discriminated, has also been reported in previous literature (Westman et al., 2011; Voevodskaya et al., 2014). These results are expected since they reflect the main characteristics of AD progression: the difference in hippocampal atrophy between HC and AD is inevitably larger on average compared to the other two pairs, and this affects the classifier's accuracy. Moreover, it is important to note that Paper V only uses volumetric information from the hippocampus, which may be too limited in some cases. In a previous study by Poulakis et al. (2018), it was indeed found that about 6% of AD patients present a form of AD characterized by hippocampal sparing (i.e., the hippocampus is not relevantly atrophied), while about 18% exhibit minimal atrophy across the brain in general. Therefore, volumetric information from such patients can drastically affect classification performance. For this reason, even though hippocampal volumetry has been shown to have great value in AD prediction, it is also important to combine it with other types of imaging biomarkers (e.g., based on other degenerating brain regions) in order to obtain more reliable diagnostic tools.

Similar reasoning may be also applied to the field of MS monitoring. Although statistically significant, the correlation coefficients between CC area and disability scores reported in Papers III and IV never reached high values. Therefore, the measure of CC area alone may be informative in itself but not sufficiently robust for clinical decision making. However, it may be investigated in combination with other imaging biomarkers, which together could potentially correlate more strongly with disease state.

### **Robustness and reliability of segmentation-derived biomarkers**

The potential of the above-described markers of neurodegeneration strongly relies on the outcome of the automatic segmentation algorithms. Therefore, in order to better trust the clinical relevance of these results, it is important

### 6.1. THE IMPORTANT ROLE OF BRAIN MRI IMAGE ANALYSIS

to ensure that the applied segmentation methods are sufficiently robust. This topic is addressed especially in Papers III and V.

In Paper III, scanner-wise cross-validation was performed to analyze the CC segmentation accuracy on data from scanners that differed from those used for training. By doing so, the average Dice coefficient dropped by approximately 6% on FLAIR images compared to the 10-fold cross-validation approach (where data from different scanners were randomly distributed across folds). Conversely, for T1w data, the accuracy remained high also with scanner-wise cross-validation, supporting the high level of robustness and generalization of the proposed methods when applied on T1w data. This generalization discrepancy between FLAIR and T1w was partially expected: as presented in Section 5.2.1, CC segmentation on FLAIR scans was consistently more challenging in terms of both Dice coefficient and scan-rescan reproducibility. This is most likely caused by the different contrast of FLAIR images compared to T1w. Even for the manual annotator who labeled the data used in Paper III, it was perceived as more difficult to segment CC on FLAIR with high accuracy and confidence compared to T1w. Thus, it is expected that automatic algorithms would reflect this trend, especially when the discrepancy between training and test set is higher.

A relatively similar strategy was followed in Paper V, where the comparison between single- and cross-cohort performance was carried out. This analysis revealed that Shape MRI U-Net could provide more robust hippocampal segmentation results when tested on a cohort that differed from the one used for training as opposed to the standard U-Net. This result is of high relevance since it supports the choice of Shape MRI U-Net—which benefits from the use of a priori shape information—over other more standard approaches.

Furthermore, in both Papers III and V, consistency in segmentation accuracy was analyzed between different levels of disease progression. This is a crucial aspect when it comes to using these tools for clinical applications and research: the more accurate the segmentation results are across the population, the more reliable the computed biomarkers will be, and the higher the chances are of eventually using them in the clinic.

More specifically, in Paper III, the CC segmentation performance was investigated across different atrophy levels, and patients with a high degree of CC atrophy showed significantly lower Dice scores. This was expected since extensive and heterogeneous atrophy patterns can make it very challenging for the network to segment the CC properly. However, it should

be noted that this loss in accuracy was not dramatic, and very few outliers (i.e., subjects with particularly low Dice scores compared to average) were found. Thus, this constitutes a good starting point for a potential future use of these segmentation tools in clinical settings.

Similarly, in Paper V, the performance of the proposed DL-based architectures was compared between the three diagnostic groups of interest (HC, MCI and AD). As expected, due to the different levels of hippocampal atrophy, a performance loss was observed for MCI and, even more, AD patients in both the single- and cross-cohort analyses. However, the accuracy observed on these patients was higher than that obtained using the widely used software FreeSurfer.

To conclude, the observations presented above are very important to keep in mind when testing segmentation algorithms on different MRI sequences as well as on data from new unseen scanners or patients. A consistent test accuracy may not be obtained at all times since multiple factors (e.g., image contrast, disease progression, etc.) come into play and may affect accuracy. This is a crucial point that both clinical and machine learning (ML) researchers should always keep in mind when approaching imaging biomarker research.

## 6.2 Machine learning: The future of neuroradiology?

In the previous sections, the importance of MRI image analysis for neuroscience was presented by discussing the main results of the papers appended to this thesis. In four out of five papers (i.e., Papers II to V), ML and DL represent the core approaches employed to analyze the brain images of interest. As described in Section 3.4, both traditional ML and DL methods currently play a fundamental role in medical imaging research and dominate a large portion of the scientific literature in this field. For this reason, it is difficult to imagine a future of neuroradiology without the integration of automated ML-based tools for clinical decision making.

### 6.2.1 Advantages of machine learning approaches

One of the strengths of ML-based systems is their time efficiency. While the training process can be lengthy (especially when dealing with particularly complex DL networks and large datasets), the inference phase is much quicker. For example, in Paper III, the full pipeline (including automatic midslice selection and both CC and IC segmentation) was shown to take less

## 6.2. MACHINE LEARNING: THE FUTURE OF NEURORADIOLOGY?

than one minute for each input 3D MRI scan. The same process would take much longer (in the order of tens of minutes) if performed fully manually and possibly even longer with other automatic and more established—but not ML-based—neuroimaging software. For example, using FreeSurfer, the full processing was shown to take up to 10 hours per input 3D scan. Similarly, also in Paper V, a huge time difference was reported compared to FreeSurfer since the full Shape MRI U-Net pipeline for hippocampal segmentation took approximately two-and-a-half minutes per subject.

In addition to time efficiency, another advantage of these state-of-the-art approaches—especially when it comes to novel DL methods—is their higher accuracy compared to more traditional automatic methods. This aspect is addressed in Section 3.4 and explored in part of the works appended to this thesis. For example, in Paper V, another comparison with FreeSurfer is carried out by analyzing the Dice coefficients achieved on FreeSurfer-based segmentations as opposed to those obtained with the proposed Shape MRI U-Net. The latter method was shown to have a consistently superior accuracy, even when applied on a new unseen cohort and on AD patients presenting hippocampal atrophy. Similarly, in Paper III, the CC segmentation accuracy obtained using a relatively simple U-Net-like architecture was reported to be comparable, if not superior, to other similar segmentation studies, especially when working with T1w data.

Researchers are also extremely active in pushing the performance of these ML methods even further to achieve higher and higher accuracy. Also in this thesis, some technical novelties are presented with the aim of proposing alternative tools that may be tested by other researchers to improve the state of the art. For example, in Paper II, a well-established age prediction method based on Gaussian process regression (GPR) was extended by creating a novel ensemble of GPR and an LR classifier. This novel combination showed a lower prediction error on average by taking advantage of two aspects of LR: (1) its high accuracy when a relatively large number of samples is available (i.e., at younger ages in the case of Paper II); (2) the absence of “regression towards the mean” (i.e., the tendency of over- and under-estimating the predictions at younger and older chronological ages, respectively). Another technical innovation was presented in the context of hippocampal segmentation in Paper V by adding fitted statistical shape models to the inputs of standard U-Nets. As discussed in Section 6.1.2, this approach was found to significantly improve segmentation accuracy when testing the architecture on new unseen data. These examples from Papers

II and V show how the field of ML-based image analysis can continuously be improved. Thus, it is reasonable to follow this research direction in order to improve also the existing clinical routines.

Finally, one other aspect in favor of using ML models consists in the possibility of obtaining objective outputs, which is something that cannot be achieved when the same task is performed by a human annotator. The manual annotations of expert neuroradiologists are of extremely high value, but they will always be relatively subjective (i.e., affected by both inter- and intra-rater variability). They can indeed be influenced by multiple factors such as the expert's level of experience and/or concentration in the moment in which the annotations are performed. However, once an ML model has been successfully trained and optimized, it will always show consistent outputs when tested on the same input image.

### 6.2.2 Obstacles and ethical concerns

Given the above-examined advantages of ML-based methods, would it be acceptable to integrate them fully into the clinical routine and have them replace part of the work currently performed by radiologists? Where should the boundary be drawn between decision making performed by a medical doctor and that performed by a machine? These are probably the most significant ethical questions when it comes to the employment of artificial intelligence in healthcare. Despite the evidence that a human will never be as quick and objective as a machine in analyzing brain images and using them to compute useful diagnostic measurements, we are probably still far from being able to fully replace the experience and clinical judgment of a human expert. The author believes that this is mostly due to a series of obstacles that characterize ML when applied to medical image analysis problems.

First, once an automatic tool has been implemented and optimized, it will most likely not show perfect accuracy. In the papers appended to this thesis, no average Dice scores of 100% or mean absolute errors equal to 0 are reported, and the same can be claimed with confidence regarding the rest of the neuroimaging literature (or at least the vast majority of it) that is currently available. Part of the reason for this has been addressed frequently throughout this thesis: the discrepancy between training and test image datasets in terms of contrast, disease progression, image resolution, etc. Moreover, the achievement of only suboptimal model accuracy can arise from inter- and/or intra-rater disagreement among the annotators who

## 6.2. MACHINE LEARNING: THE FUTURE OF NEURORADIOLOGY?

generate the ground truth. This issue can make it even more difficult to assess what the optimal performance of an ML model should really be: if there is not 100% agreement among human raters, it is not possible to aim for 100% accuracy using fully automated tools. Despite this, the available manual annotations are still what can be considered closest to the truth in neuroimaging studies. Therefore, while machine performance is below this supposed truth, it will always be difficult to replace the radiologist's work.

Secondly, for some tasks, the sensitivity and caution of a human expert are of essential importance, no matter which algorithm may be used instead and what level of accuracy it is claimed to reach. For example, nowadays, a diagnosis of AD can be made with absolute certainty only after death, when the brain tissue can be examined microscopically. Simultaneously, neuroimaging biomarker research is focusing more and more on providing an AD diagnosis as early as possible. This is an extremely important goal to achieve, but at the same time, such an early diagnosis should be done with extreme caution since both false positives and false negatives can be very harmful. The first can cause patients to live with an extreme but unwarranted fear, and false negatives can prevent patients from seeking care as early as possible. Therefore, subjective and cautious evaluation from a medical professional would still be preferred in certain situations.

This latter issue becomes even more critical when considering that many of the state-of-the-art DL algorithms are “black boxes”, meaning that the reason they produce a specific output cannot be easily deciphered by the user. By using the same example as above, a trained deep neural network (DNN) may perform an early diagnosis of AD by simply receiving a brain MRI scan as input, which may (still) look healthy to the eye of a trained physician. If it is not possible to understand exactly *why* that output has been produced by the network, it will be challenging to introduce such a tool in a clinical routine. Furthermore, even if a DNN has always shown excellent accuracy in previous trials, it is well known that DL architectures can be vulnerable to adversarial attacks, that is input images that are slightly perturbed (e.g., by additive noise) and, because of that, can “fool” the network and greatly affect its output (Akhtar and Mian, 2018). More extensive research should be carried out to address these problems in order to improve the explainability and reliability of ML for clinical applications.



### 6.2.3 Aiming for balance between human and machine decision making

In the sections above, some of the advantages and disadvantages of employing fully automated ML-based neuroimaging methods are examined. Both should always be considered. Therefore, we should find an optimal balance between the contribution that can be provided by a medical doctor and that provided as output of an ML model.

Fully automated algorithms will probably not replace physicians any time soon. However, highly accurate, quick and objective ML-based methods may (and should) soon become an important part of the clinical workflow. Given the high potential of the neuroimaging and ML-based biomarkers discussed throughout this thesis, the hope is for them to be further evaluated in the future and eventually employed by medical professionals who would still have the last word on clinical decisions. In this way, these methods can find their perfect fit in the clinical routine as an additional, but still highly important, aid for improving disease diagnosis and monitoring. Aiming for such an improvement of the current clinical practices is also in line with the third of the 17 goals for global sustainable development, which is focused on promoting the good health and well-being of the world population (United Nations, 2022).

## 6.3 Limitations

The five studies presented in this thesis provide relevant contributions to the neuroimaging field. However, they are also affected by a number of limitations, which are addressed in this section and should be taken into account for improvement in the future (see also Chapter 7 for suggested future work).

**Sample size** In all five papers, the size of the available datasets was rather limited. This is a very common and well-documented issue in medical imaging as opposed to other computer vision applications that make use of similar image processing techniques (Varoquaux and Cheplygina, 2021). This problem is even more present in animal studies, where an increased sample size can highly affect the costs of the project as well as potentially cause ethical problems in case of redundant samples (Serdar et al., 2021).

In Paper I, the small sample size (eight rabbits per group), combined with a conservative Bonferroni correction in the statistical analysis, might

have led to an increased risk for Type II errors, meaning that some additional brain differences may be present between wild and domestic rabbits. In the studies using ML (i.e., Papers II to V), the limited size of the training datasets could affect the model accuracy. This issue has partially been controlled for by performing cross-validation and, in the papers using DL (Papers III to V), by adding data augmentation. However, the results of these studies also showed that the model performance decreases in the inference phase on datasets from different cohorts than the one(s) used for training. Therefore, an increased sample size would be desirable in order to further reduce this gap between training and testing performance.

**Data heterogeneity** Another limitation that characterizes the datasets used in the appended papers is their limited heterogeneity, which is known to cause additional biases (Varoquaux and Cheplygina, 2021).

In Paper I, only female rabbits were employed, but it would be interesting to study the effects of domestication on male rabbits as well. The opposite issue affects Paper II, where only male rats were used. Moreover, in this longitudinal study, the rats were scanned only at four specific time points throughout their lifespan. More frequent scans and a longer observation period could be beneficial for both implementing a more robust age prediction model and drawing more reliable conclusions on the value of BrainAGE as a marker of healthy aging. Another aspect that may affect the current model’s robustness and future applicability is the fact that all images were acquired from the same scanner. Therefore, it is unknown what the performance would be on data acquired from different sites. The same limitation affects also Paper IV, in which the CC segmentation networks were trained and tested on T2w data from only one scanner. In Paper III, the issue was addressed by using T1w and FLAIR data from three different scanners, supporting more the generalizability of the models on new unseen data. However, all three scanners were from the same manufacturer, so the performance of the model is still unknown on scans from different manufacturers.

**Processing and modeling choices** For the sake of time and simplicity, specific processing and modeling choices were made in the presented papers, but more options could be explored in order to both improve these works and further validate their main findings.

In Paper II, GPR with a linear covariance function was employed as

a baseline model in order to mimic the strategy used in human studies in the past (Cole et al., 2015, 2017b, 2018). However, considering also the nonlinearity in the aging trajectories that was observed in this and other studies, alternative nonlinear models could also be suitable. It should also be noted, though, that the use of more complex nonlinear models may not be appropriate with the limited amount of available training samples. Furthermore, a limitation of the used linear covariance function is that it does not allow to directly extract importance measures for each of the input features of the model. Therefore, this aspect was investigated only for the LR classifier, but the use of alternative covariance functions (e.g., kernels with automatic relevance determination) could allow performing this analysis for the regressor as well (Caywood et al., 2017).

When it comes to the workflows implemented for the three human studies (Papers III to V), one common drawback consists in the limited image preprocessing strategies tested. In all three works, the DL networks' inputs only underwent resizing and intensity normalization. This simple preprocessing has the advantage of being extremely time efficient. However, some additional data normalization steps could also be tested, for example a preliminary registration onto a reference space.

Another common feature of all three human studies is the use of 2D U-Nets for image segmentation. As described in Section 3.4.4, this is a very well-established architecture within the medical image analysis community, and it has the advantage of being simple and relatively quick to train. Therefore, on one hand, opting for this network is particularly advantageous in view of potential future implementation by other researchers on new data. On the other hand, remaining with just one network can be limiting, since other alternative options could be tested and potentially be equally or more successful.

## Chapter 7

# Future work

The scientific contributions provided by the studies presented in this thesis can be seen as a promising starting point for future improvement and validation. Below, some of the possible future directions for continuing this work are outlined.

### 7.1 Larger and heterogeneous datasets

As discussed in Section 6.3, a limitation that all five appended papers have in common is the limited size and heterogeneity of the employed datasets. Although larger and more diverse datasets do not necessarily correspond to better model performance (Varoquaux and Cheplygina, 2021), working with more data would at least be useful to validate the findings of the presented works. Below are listed the additions to the datasets that should be performed before reproducing and/or extending the present studies.

#### **Domestication-induced changes in the rabbit brain (Paper I)**

Male rabbits should be included in the dataset. As mentioned in the paper, female rabbits are particularly subjective to stress, so this behavioral difference between male and female animals may also reflect some underlying structural brain differences. By adding male subjects to the already available dataset, it would be possible to validate the previous findings and reduce the risk for Type II errors.

### **Rat brain age prediction models (Paper II)**

The generalization power of the proposed prediction model should be improved by training it with additional image scans acquired as follows:

1. From both male and female animals.
2. At more frequent time-points as well as before and after the current limits of the study observation period (i.e., 3 and 17 months).
3. From multiple sites and scanners.

### **Segmentation of the degenerating corpus callosum and hippocampus (Papers III to V)**

The three human studies of this thesis would benefit from expanding the available training datasets by manually labeling new scans based on the following:

1. More sites and scanners. This would increase the possibility of successfully using the proposed methods on data acquired by other researchers and clinicians.
2. More subjects affected by high levels of neurodegeneration (in the corpus callosum for Papers III and IV and in the hippocampus for Paper V). The appended papers show a general tendency of the proposed methods to perform better on healthy participants. While this is expected, it would still be important to focus on reducing the gap in performance between healthy and disease cases. This would increase the reliability of the MRI-derived biomarkers.

## **7.2 Validation of MRI-derived biomarkers**

In this thesis, brain image analysis methods for extracting potential biomarkers were explored. In some cases, new processing pipelines were proposed for deriving imaging biomarkers that are already established in literature. This is the case of the hippocampal volume (Paper V) and the corpus callosum (CC) area (Papers III and IV), which had already been recognized as potential indicators of disease progression in previous studies (see background Section 3.1.4). This thesis also presented novel insights into MRI-derived information that could be of high relevance for biomarker research. This is the case for the investigation of BrainAGE as a marker

## 7.2. VALIDATION OF MRI-DERIVED BIOMARKERS

of aging in rats (Paper II)—since previously it had only been analyzed in humans—as well as the CC shape deformations identified in multiple sclerosis (MS) patients (Paper IV), which correlate with disability in a similar way as the CC area. Given the high novelty of these two latter results, the author believes that future work on these specific applications should have the highest priority.

### **Preclinical applications of BrainAGE**

Given the promising results of Paper II, which showed that a healthy lifestyle intervention does have a significant effect on BrainAGE, future work should involve testing the proposed age prediction algorithm on other rat models of aging. In particular, it would be interesting to analyze the BrainAGE score on cohorts of pathological aging, such as rat models of Alzheimer’s disease. In such applications, the main hypothesis is that higher BrainAGE scores would be observed compared to controls as this would match the results reported in human studies (see Section 3.4.6). If this is the case, it would constitute an important validation of the BrainAGE biomarker in preclinical settings.

Moreover, this study may also be extended by developing brain age prediction models for mice. Indeed, mice are often preferred over rats as rodent models given their smaller size, their lower maintenance costs and the higher availability of tools for genetic manipulation (Bryda, 2013). Therefore, analyzing BrainAGE on mice would be extremely useful not only to strengthen its potential across species but also to investigate additional models of abnormal aging.

### **Geometric features of the corpus callosum in multiple sclerosis**

Paper IV provided interesting insights on the fact that with MS progression, not only a reduction in CC area can be observed but also a change in its shape. The relevance of this result is two-fold. The shape features not only represent interesting MS biomarkers; they can also provide new knowledge on the macroscopic neurodegeneration process in these patients. However, it is important to further validate these findings.

First, a natural way of continuing the present work would consist in performing shape analysis on CC segmented from larger datasets. For this purpose, it would be interesting to include CC segmentations obtained by applying the T1w- and FLAIR-specific pipelines proposed in Paper III. In

this study, a significant correlation was identified between the CC area and disability (similarly to what was reported in Paper IV). Therefore, similar correlations between shape features and disability would also be expected.

Furthermore, given the availability of software tools for CC segmentation in 3D (such as FreeSurfer, which was tested in Papers III and IV), shape analysis could be performed in 3D. This approach might shed new light on the shape changes that the CC undergoes.

Finally, longitudinal studies should be designed to investigate how the geometry of the CC changes through time. In Papers III and IV, both area and shape features of the CC were shown to correlate with future disability scores. This is an important result in the context of disease diagnosis and monitoring. Thus, by running longitudinal studies, it would be possible to closely monitor the disease progression of the participants through time and more thoroughly investigate the prediction power of the potential biomarkers presented in this thesis.

### 7.3 Testing new architectures

In all three human studies presented in this thesis, the 2D U-Net was chosen as the preferred architecture for approaching all segmentation tasks (both for the CC in 2D and for the hippocampus in 3D). Despite the positive results, it would still be beneficial to test new architectures and compare their performance with that presented in this thesis.

For example, the knowledge gained from Paper V on the use of shape context for improving cross-cohort accuracy could be applied for CC segmentation (Papers III and IV). This could improve the segmentation performance on those data that were shown to be particularly challenging, such as FLAIR scans and scans from patients with higher degrees of atrophy (see Paper III).

The hippocampal segmentation method presented in Paper V can also be further improved. While the shape input layer can be preserved, alternative segmentation networks could be tested. For instance, the three orthogonal 2D U-Nets might be replaced by only one 3D network, for example the 3D U-Net (Çiçek et al., 2016) or the V-Net (Milletari et al., 2016). In this way, the network would learn from 3D information instead of considering one image view at a time, and this may improve the final segmentation accuracy.

## Chapter 8

# Conclusions

The studies included in this thesis showed how the development and use of appropriate MRI image processing tools can help characterize and analyze brain morphology. MRI scans could be employed to extract important information on the size and shape of brain structures, which in turn could be used for accomplishing two main goals.

First, this morphological information could provide new knowledge for better understanding the brain and how its structure changes due to multiple factors. Morphological changes of an evolutionary nature were explored in Paper I, where domestication was found to significantly reshape brain structure in rabbits in terms of both regional gray matter volume and white matter integrity. Lifestyle-induced modifications were investigated in Paper II, where a healthy lifestyle was found to significantly slow rat brain aging. New insights were also gained regarding the changes in the human brain affected by multiple sclerosis (MS) thanks to the shape analysis of the corpus callosum (CC) conducted in Paper IV, which revealed a significant correlation between disability and CC thickness and bending angle.

Second, data on brain structure, together with the implementation of appropriate machine learning-based models, could be used to validate and expand the current knowledge on neuroimaging biomarkers. A preclinical validation of BrainAGE as a potential marker of healthy brain aging was successfully carried out in Paper II, whereas methods for obtaining reliable markers of neurodegeneration were explored in the three human studies of this thesis. In particular, morphological measures extracted from the automatic segmentations of the CC and the hippocampus were associated with disease state in MS and Alzheimer's disease (AD), respectively.





# Bibliography

- Adamson, C., Beare, R., Walterfang, M., Seal, M., 2014. Software pipeline for midsagittal corpus callosum thickness profile processing. *Neuroinformatics* 12, 595–614. URL: <https://doi.org/10.1007/s12021-014-9236-3>, doi:10.1007/s12021-014-9236-3.
- Agren, T., Engman, J., Frick, A., Björkstrand, J., Larsson, E.M., Furmark, T., Fredrikson, M., 2012. Disruption of reconsolidation erases a fear memory trace in the human amygdala. *Science* 337, 1550–2. doi:10.1126/science.1223006.
- Akhtar, N., Mian, A., 2018. Threat of adversarial attacks on deep learning in computer vision: A survey. *IEEE Access* 6, 14410–14430. doi:10.1109/ACCESS.2018.2807385.
- Akkus, Z., Galimzianova, A., Hoogi, A., Rubin, D.L., Erickson, B.J., 2017. Deep learning for brain MRI segmentation: State of the art and future directions. *J Digit Imaging* 30, 449–459. doi:10.1007/s10278-017-9983-4.
- Alexander, D.C., 2009. Modelling, fitting and sampling in diffusion MRI. Springer. pp. 3–20.
- Alzubi, J., Nayyar, A., Kumar, A., 2018. Machine learning from theory to algorithms: an overview, in: *Journal of physics: conference series*, IOP Publishing. p. 012012.
- Andersson, J.L., Jenkinson, M., Smith, S., 2007a. Non-linear registration, aka spatial normalisation FMRIB technical report TR07JA2. FMRIB Analysis Group of the University of Oxford 2, e21.
- Andersson, J.L., Jenkinson, M., Smith, S., Andersson, J., 2007b. Non-linear optimisation. FMRIB technical report TR07JA1. Practice .
- Ashburner, J., Friston, K.J., 2000. Voxel-based morphometry—the methods. *NeuroImage* 11, 805–821.

## BIBLIOGRAPHY

- Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis* 12, 26–41.
- Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C., 2011a. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage* 54, 2033–2044. URL: <https://pubmed.ncbi.nlm.nih.gov/20851191><https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3065962/>, doi:10.1016/j.neuroimage.2010.09.025.
- Avants, B.B., Tustison, N.J., Wu, J., Cook, P.A., Gee, J.C., 2011b. An open source multivariate framework for n-tissue segmentation with evaluation on public data. *Neuroinformatics* 9, 381–400.
- Bakshi, R., Ariyaratana, S., Benedict, R.H.B., Jacobs, L., 2001. Fluid-attenuated inversion recovery magnetic resonance imaging detects cortical and juxtacortical multiple sclerosis lesions. *Archives of Neurology* 58, 742–748. URL: <https://doi.org/10.1001/archneur.58.5.742>, doi:10.1001/archneur.58.5.742.
- Bankman, I.N., 2009. Segmentation. Academic Press, Burlington. pp. 71–72. URL: <http://www.sciencedirect.com/science/article/pii/B978012373904950012X>, doi:<https://doi.org/10.1016/B978-012373904-9.50012-X>.
- Barnes, J., Boyes, R.G., Lewis, E.B., Schott, J.M., Frost, C., Schill, R.I., Fox, N.C., 2007. Automatic calculation of hippocampal atrophy rates using a hippocampal template and the boundary shift integral. *Neurobiology of Aging* 28, 1657–1663. URL: <https://www.sciencedirect.com/science/article/pii/S0197458006002624>, doi:<https://doi.org/10.1016/j.neurobiolaging.2006.07.008>.
- Bashyam, V.M., Erus, G., Doshi, J., Habes, M., Nasrallah, I., Truelove-Hill, M., Srinivasan, D., Mamourian, L., Pomponio, R., Fan, Y., 2020. MRI signatures of brain age and disease over the lifespan based on a deep brain network and 14 468 individuals worldwide. *Brain* 143, 2312–2324.
- Beutel, J., Kundel, H.L., Van Metter, R.L., Fitzpatrick, J.M., 2000. Handbook of Medical Imaging: Medical image processing and analysis. volume 2. Spie Press.
- Biomarkers Definitions Working Group, 2001. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 69, 89–95. doi:10.1067/mcp.2001.113989.

- Bittner, N., Jockwitz, C., Franke, K., Gaser, C., Moebus, S., Bayen, U.J., Amunts, K., Caspers, S., 2021. When your brain looks older than expected: combined lifestyle risk and brainage. *Brain Structure and Function* URL: <https://doi.org/10.1007/s00429-020-02184-6>, doi:10.1007/s00429-020-02184-6.
- Bloch, F., 1946. Nuclear induction. *J Physical review* 70, 460.
- Braak, H., Braak, E., 1995. Staging of Alzheimer's disease-related neurofibrillary changes. *Neurobiology of Aging* 16, 271–278. URL: <https://www.sciencedirect.com/science/article/pii/0197458095000216>, doi:[https://doi.org/10.1016/0197-4580\(95\)00021-6](https://doi.org/10.1016/0197-4580(95)00021-6).
- Brown, A.G., 2012. *Nerve cells and nervous systems: an introduction to neuroscience*. Springer Science Business Media.
- Brusini, I., Carneiro, M., Wang, C., Rubin, C.J., Ring, H., Afonso, S., Blanco-Aguilar, J.A., Ferrand, N., Rafati, N., Villafuerte, R., Smedby, Ö., Damberg, P., Hallböök, F., Fredrikson, M., Andersson, L., 2018. Changes in brain architecture are consistent with altered fear processing in domestic rabbits 115, 7380–7385. URL: <https://www.pnas.org/content/115/28/7380>, doi:10.1073/pnas.1801024115.
- Brusini, I., Lindberg, O., Muehlboeck, J.S., Smedby, Ö., Westman, E., Wang, C., 2020. Shape information improves the cross-cohort performance of deep learning-based segmentation of the hippocampus. *Frontiers in Neuroscience* 14. URL: <https://www.frontiersin.org/article/10.3389/fnins.2020.00015>, doi:10.3389/fnins.2020.00015.
- Brusini, I., MacNicol, E., Kim, E., Smedby, Ö., Wang, C., Westman, E., Veronese, M., Turkheimer, F., Cash, D., 2022a. MRI-derived brain age as a biomarker of ageing in rats: validation using a healthy lifestyle intervention. *Neurobiology of Aging* 109, 204–215. URL: <https://www.sciencedirect.com/science/article/pii/S019745802100316X>, doi:<https://doi.org/10.1016/j.neurobiolaging.2021.10.004>.
- Brusini, I., Platten, M., Ouellette, R., Piehl, F., Wang, C., Granberg, T., 2022b. Automatic deep learning multicontrast corpus callosum segmentation in multiple sclerosis. *Journal of Neuroimaging* URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/jon.12972>, doi:<https://doi.org/10.1111/jon.12972>.
- Bryda, E.C., 2013. The mighty mouse: the impact of rodents on advances in biomedical research. *Mo Med* 110, 207–11.
- Burns, A., Iliffe, S., 2009. Alzheimer's disease. *Bmj* 338, b158. doi:10.1136/bmj.b158.

## BIBLIOGRAPHY

- Carmichael, O.T., Aizenstein, H.A., Davis, S.W., Becker, J.T., Thompson, P.M., Meltzer, C.C., Liu, Y., 2005. Atlas-based hippocampus segmentation in Alzheimer's disease and mild cognitive impairment. *NeuroImage* 27, 979–990. URL: <https://www.sciencedirect.com/science/article/pii/S1053811905003162>, doi:<https://doi.org/10.1016/j.neuroimage.2005.05.005>.
- Carneiro, M., Rubin, C.J., Di Palma, F., Albert, F.W., Alföldi, J., Martinez Barrio, A., Pielberg, G., Rafati, N., Sayyab, S., Turner-Maier, J., Younis, S., Afonso, S., Aken, B., Alves, J.M., Barrell, D., Bolet, G., Boucher, S., Burbano, H.A., Campos, R., Chang, J.L., Duranthon, V., Fontanesi, L., Garreau, H., Heiman, D., Johnson, J., Mage, R.G., Peng, Z., Queney, G., Rogel-Gaillard, C., Ruffier, M., Searle, S., Villafuerte, R., Xiong, A., Young, S., Forsberg-Nilsson, K., Good, J.M., Lander, E.S., Ferrand, N., Lindblad-Toh, K., Andersson, L., 2014. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* 345, 1074–1079. doi:[10.1126/science.1253714](https://doi.org/10.1126/science.1253714).
- Caywood, M.S., Roberts, D.M., Colombe, J.B., Greenwald, H.S., Weiland, M.Z., 2017. Gaussian process regression for predictive but interpretable machine learning models: An example of predicting mental workload across tasks. *Frontiers in Human Neuroscience* 10. URL: <https://www.frontiersin.org/article/10.3389/fnhum.2016.00647>, doi:[10.3389/fnhum.2016.00647](https://doi.org/10.3389/fnhum.2016.00647).
- Chandra, A., Verma, S., Raghuvanshi, A.S., Bodhey, N.K., Londhe, N.D., Subham, K., 2019. SegNet-based corpus callosum segmentation for brain magnetic resonance images (MRI), in: 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), pp. 317–322. doi:[10.1109/ICCCIS48478.2019.8974559](https://doi.org/10.1109/ICCCIS48478.2019.8974559).
- Chen, H., Dou, Q., Yu, L., Qin, J., Heng, P.A., 2018. VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images. *Neuroimage* 170, 446–455. doi:[10.1016/j.neuroimage.2017.04.041](https://doi.org/10.1016/j.neuroimage.2017.04.041).
- Chen, Y., Shi, B., Wang, Z., Sun, T., Smith, C.D., Liu, J., 2017. Accurate and consistent hippocampus segmentation through convolutional LSTM and view ensemble, Springer International Publishing. pp. 88–96.
- Chowdhury, F.K., Jacobs, E., Hossen, J., Salan, T., 2014. Diffusion tensor based global tractography of human brain fiber bundles, in: 8th International Conference on Electrical and Computer Engineering, IEEE. pp. 5–8.
- Cole, J.H., Annus, T., Wilson, L.R., Remtulla, R., Hong, Y.T., Fryer, T.D., Acosta-Cabronero, J., Cardenas-Blanco, A., Smith, R., Menon, D.K., 2017a. Brain-predicted age in Down syndrome is associated with beta amyloid deposition and cognitive decline. *Neurobiology of aging* 56, 41–49.

- Cole, J.H., Leech, R., Sharp, D.J., Initiative, A.D.N., 2015. Prediction of brain age suggests accelerated atrophy after traumatic brain injury. *Annals of neurology* 77, 571–581.
- Cole, J.H., Poudel, R.P., Tsagkrasoulis, D., Caan, M.W., Steves, C., Spector, T.D., Montana, G., 2017b. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage* 163, 115–124.
- Cole, J.H., Ritchie, S.J., Bastin, M.E., Hernández, M.V., Maniega, S.M., Royle, N., Corley, J., Pattie, A., Harris, S.E., Zhang, Q., 2018. Brain age predicts mortality. *Molecular psychiatry* 23, 1385–1392.
- Cole, J.H., Underwood, J., Caan, M.W., De Francesco, D., van Zoest, R.A., Leech, R., Wit, F.W., Portegies, P., Geurtsen, G.J., Schmand, B.A., 2017c. Increased brain-predicted aging in treated HIV disease. *Neurology* 88, 1349–1357.
- Cover, G.S., Herrera, W.G., Bento, M.P., Appenzeller, S., Rittner, L., 2018. Computational methods for corpus callosum segmentation on MRI: A systematic literature review. *Computer Methods and Programs in Biomedicine* 154, 25–35. URL: <https://www.sciencedirect.com/science/article/pii/S0169260716314432>, doi:<https://doi.org/10.1016/j.cmpb.2017.10.025>.
- Darwin, C., 1859. *The origin of species by means of natural selection*.
- Davidson, R.J., Putnam, K.M., Larson, C.L., 2000. Dysfunction in the neural circuitry of emotion regulation—a possible prelude to violence. *Science* 289, 591–4. doi:10.1126/science.289.5479.591.
- DeCarli, C., 2000. Part IV. Neuroimaging in dementing disorders. *Dis Mon* 46, 706–24. doi:10.1016/s0011-5029(00)90031-2.
- Denic, A., Macura, S.I., Mishra, P., Gamez, J.D., Rodriguez, M., Pirko, I., 2011. MRI in rodent models of brain disorders. *Neurotherapeutics* 8, 3–18. doi:10.1007/s13311-010-0002-4.
- Denver, P., McClean, P.L., 2018. Distinguishing normal brain aging from the development of Alzheimer’s disease: inflammation, insulin signaling and cognition. *J Neural regeneration research* 13, 1719.
- Desai, A.K., Grossberg, G.T., Chibnall, J.T., 2010. Healthy brain aging: a road map. *Clinics in Geriatric Medicine* 26, 1–16.
- Di Paola, M., Di Iulio, F., Cherubini, A., Blundo, C., Casini, A.R., Sancesario, G., Passafiume, D., Caltagirone, C., Spalletta, G., 2010. When, where, and how the corpus callosum changes in MCI and AD: a multimodal MRI study. *Neurology* 74, 1136–42. doi:10.1212/WNL.0b013e3181d7d8cb.

## BIBLIOGRAPHY

- Dice, L.R., 1945. Measures of the amount of ecologic association between species. *Ecology* 26, 297–302.
- Dill, V., Franco, A.R., Pinho, M.S., 2015. Automated methods for hippocampus segmentation: the evolution and a review of the state of the art. *Neuroinformatics* 13, 133–150.
- Dong, H., Yang, G., Liu, F., Mo, Y., Guo, Y., 2017. Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks, Springer International Publishing. pp. 506–517.
- Dosenbach, N.U., Nardos, B., Cohen, A.L., Fair, D.A., Power, J.D., Church, J.A., Nelson, S.M., Wig, G.S., Vogel, A.C., Lessov-Schlaggar, C.N., 2010. Prediction of individual brain maturity using fMRI. *Science* 329, 1358–1361.
- Douaud, G., Smith, S., Jenkinson, M., Behrens, T., Johansen-Berg, H., Vickers, J., James, S., Voets, N., Watkins, K., Matthews, P.M., James, A., 2007. Anatomically related grey and white matter abnormalities in adolescent-onset schizophrenia. *Brain* 130, 2375–86. doi:10.1093/brain/awm184.
- Dreiseitl, S., Ohno-Machado, L., 2002. Logistic regression and artificial neural network classification models: a methodology review. *Journal of biomedical informatics* 35, 352–359.
- Dwivedi Chenna, Y., Ghassemi, P., Pfefer, T., Casamento, J., Wang, Q., 2018. Free-form deformation approach for registration of visible and infrared facial images in fever screening. *Sensors* 18, 125.
- Erickson, B.J., Korfiatis, P., Akkus, Z., Kline, T.L., 2017. Machine learning for medical imaging. *Radiographics* 37, 505–515.
- Fang, L., Zhang, L., Nie, D., Cao, X., Rekik, I., Lee, S.W., He, H., Shen, D., 2019. Automatic brain labeling via multi-atlas guided fully convolutional networks. *Medical Image Analysis* 51, 157–168. URL: <https://www.sciencedirect.com/science/article/pii/S1361841518308600>, doi:<https://doi.org/10.1016/j.media.2018.10.012>.
- Fischl, B., 2012. Freesurfer. *NeuroImage* 62, 774–81. doi:10.1016/j.neuroimage.2012.01.021.
- Franke, K., Dahnke, R., Clarke, G., Kuo, A., Li, C., Nathanielsz, P., Schwab, M., Gaser, C., 2016. MRI based biomarker for brain aging in rodents and non-human primates, in: 2016 international workshop on pattern recognition in neuroimaging (PRNI), IEEE. pp. 1–4.

- Franke, K., Gaser, C., Manor, B., Novak, V., 2013. Advanced BrainAGE in older adults with type 2 diabetes mellitus. *Frontiers in aging neuroscience* 5, 90.
- Franke, K., Ziegler, G., Klöppel, S., Gaser, C., Alzheimer's Disease Neuroimaging Initiative, 2010. Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: exploring the influence of various parameters. *Neuroimage* 50, 883–892.
- Gaser, C., Franke, K., Klöppel, S., Koutsouleris, N., Sauer, H., Alzheimer's Disease Neuroimaging Initiative, 2013. BrainAGE in mild cognitive impaired patients: predicting the conversion to Alzheimer's disease. *PloS one* 8.
- Giannelli, M., Cosottini, M., Michelassi, M.C., Lazzarotti, G., Belmonte, G., Bartolozzi, C., Lazzeri, M., 2010. Dependence of brain DTI maps of fractional anisotropy and mean diffusivity on the number of diffusion weighting directions. *Journal of Applied Clinical Medical Physics* 11, 176–190. URL: <https://aapm.onlinelibrary.wiley.com/doi/abs/10.1120/jacmp.v11i1.2927>, doi:<https://doi.org/10.1120/jacmp.v11i1.2927>.
- Good, C.D., Johnsrude, I.S., Ashburner, J., Henson, R.N.A., Friston, K.J., Frackowiak, R.S.J., 2001. A voxel-based morphometric study of ageing in 465 normal adult human brains. *NeuroImage* 14, 21–36. URL: <https://www.sciencedirect.com/science/article/pii/S1053811901907864>, doi:<https://doi.org/10.1006/nimg.2001.0786>.
- Gossuin, Y., Hocq, A., Gillis, P., Vuong, Q.L., 2010. Physics of magnetic resonance imaging: from spin to pixel. *Journal of Physics D: Applied Physics* 43, 213001.
- Granberg, T., Bergendal, G., Shams, S., Aspelin, P., Kristoffersen-Wiberg, M., Fredrikson, S., Martola, J., 2015a. MRI-defined corpus callosal atrophy in multiple sclerosis: A comparison of volumetric measurements, corpus callosum area and index. *Journal of Neuroimaging* 25, 996–1001. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/jon.12237>, doi:<https://doi.org/10.1111/jon.12237>.
- Granberg, T., Martola, J., Bergendal, G., Shams, S., Damangir, S., Aspelin, P., Fredrikson, S., Kristoffersen-Wiberg, M., 2015b. Corpus callosum atrophy is strongly associated with cognitive impairment in multiple sclerosis: Results of a 17-year longitudinal study. *Multiple Sclerosis Journal* 21, 1151–1158. URL: <https://journals.sagepub.com/doi/abs/10.1177/1352458514560928>, doi:[10.1177/1352458514560928](https://doi.org/10.1177/1352458514560928).
- Guha Roy, A., Conjeti, S., Navab, N., Wachinger, C., 2019. QuickNAT: A fully convolutional network for quick and accurate segmentation of neuroanatomy. *NeuroImage* 186, 713–727. doi:[10.1016/j.neuroimage.2018.11.042](https://doi.org/10.1016/j.neuroimage.2018.11.042).



## BIBLIOGRAPHY

- Gunning-Dixon, F.M., Brickman, A.M., Cheng, J.C., Alexopoulos, G.S., 2009. Aging of cerebral white matter: a review of MRI findings. *International Journal of Geriatric Psychiatry* 24, 109–117.
- Hajnal, J.V., De Coene, B., Lewis, P.D., Baudouin, C.J., Cowan, F.M., Pennock, J.M., Young, I.R., Bydder, G.M., 1992. High signal regions in normal white matter shown by heavily T2-weighted CSF nulled IR sequences. *J Comput Assist Tomogr* 16, 506–513.
- Hardiman, O., Doherty, C.P., Elamin, M., Bede, P., 2011. *Neurodegenerative disorders*. Springer.
- Heckemann, R.A., Hajnal, J.V., Aljabar, P., Rueckert, D., Hammers, A., 2006. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage* 33, 115–126. URL: <https://www.sciencedirect.com/science/article/pii/S1053811906006458>, doi:<https://doi.org/10.1016/j.neuroimage.2006.05.061>.
- Hoyer, C., Gass, N., Weber-Fahr, W., Sartorius, A., 2014. Advantages and challenges of small animal magnetic resonance imaging as a translational tool. *Neuropsychobiology* 69, 187–201. doi:10.1159/000360859.
- Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J., 1993. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15, 850–863. doi:10.1109/34.232073.
- Hwang, H., Rehman, H.Z.U., Lee, S., 2019. 3D U-Net for skull stripping in brain MRI. *Applied Sciences* 9, 569. URL: <https://www.mdpi.com/2076-3417/9/3/569>.
- Hötting, K., Röder, B., 2013. Beneficial effects of physical exercise on neuroplasticity and cognition. *Neuroscience Biobehavioral Reviews* 37, 2243–2257.
- Irwin, K., Sexton, C., Daniel, T., Lawlor, B., Naci, L., 2018. Healthy aging and dementia: Two roads diverging in midlife? *Frontiers in aging neuroscience* 10.
- Jack, C. R., J., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., J, L.W., Ward, C., Dale, A.M., Felmlee, J.P., Gunter, J.L., Hill, D.L., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C.S., Krueger, G., Ward, H.A., Metzger, G.J., Scott, K.T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J.P., Fleisher, A.S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., Weiner, M.W., 2008. The Alzheimer’s Disease Neuroimaging Initiative (ADNI): MRI methods. *J Magn Reson Imaging* 27, 685–91. doi:10.1002/jmri.21049.

- Jellison, B.J., Field, A.S., Medow, J., Lazar, M., Salamat, M.S., Alexander, A.L., 2004. Diffusion tensor imaging of cerebral white matter: a pictorial review of physics, fiber tract anatomy, and tumor imaging patterns. *American Journal of Neuroradiology* 25, 356–369.
- Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. FSL. *NeuroImage* 62, 782–90. doi:10.1016/j.neuroimage.2011.09.015.
- Jensen, P., 2014. Behavior genetics and the domestication of animals. *Annu Rev Anim Biosci* 2, 85–104. doi:10.1146/annurev-animal-022513-114135.
- Joshi, S.H., Narr, K.L., Philips, O.R., Nuechterlein, K.H., Asarnow, R.F., Toga, A.W., Woods, R.P., 2013. Statistical shape analysis of the corpus callosum in schizophrenia. *NeuroImage* 64, 547–559. URL: <http://www.sciencedirect.com/science/article/pii/S1053811912009342>, doi:<https://doi.org/10.1016/j.neuroimage.2012.09.024>.
- Jónsson, B.A., Bjornsdottir, G., Thorgeirsson, T., Ellingsen, L.M., Walters, G.B., Gudbjartsson, D., Stefansson, H., Stefansson, K., Ulfarsson, M., 2019. Brain age prediction using deep learning uncovers associated sequence variants. *Nature communications* 10, 1–10.
- Kermi, A., Mahmoudi, I., Khadir, M.T., 2019. Deep convolutional neural networks using U-Net for automatic brain tumor segmentation in multimodal MRI volumes, Springer International Publishing. pp. 37–48.
- Kiuchi, K., Morikawa, M., Taoka, T., Nagashima, T., Yamauchi, T., Makinodan, M., Norimoto, K., Hashimoto, K., Kosaka, J., Inoue, Y., Inoue, M., Kichikawa, K., Kishimoto, T., 2009. Abnormalities of the uncinate fasciculus and posterior cingulate fasciculus in mild cognitive impairment and early Alzheimer's disease: a diffusion tensor tractography study. *Brain Res* 1287, 184–91. doi:10.1016/j.brainres.2009.06.052.
- Kuperman, V., 2000. Magnetic resonance imaging: physical principles and applications. Elsevier.
- Kurtzke, J.F., 1983. Rating neurologic impairment in multiple sclerosis. *Neurology* 33, 1444. URL: <http://n.neurology.org/content/33/11/1444.abstract>, doi:10.1212/WNL.33.11.1444.
- Kwak, K., Yoon, U., Lee, D.K., Kim, G.H., Seo, S.W., Na, D.L., Shim, H.J., Lee, J.M., 2013. Fully-automated approach to hippocampus segmentation using a graph-cuts algorithm combined with atlas-based segmentation and morphological opening. *Magnetic Resonance Imaging* 31, 1190–1196. URL: <https://www.sciencedirect.com/science/article/pii/S0730725X13001513>, doi:<https://doi.org/10.1016/j.mri.2013.04.008>.

## BIBLIOGRAPHY

- Lancaster, J.L., Tordesillas-Gutiérrez, D., Martinez, M., Salinas, F., Evans, A., Zilles, K., Mazziotta, J.C., Fox, P.T., 2007. Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Human brain mapping* 28, 1194–1205.
- Lauterbur, P.C., 1973. Image formation by induced local interactions: examples employing nuclear magnetic resonance .
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436. URL: <https://doi.org/10.1038/nature14539>, doi:10.1038/nature14539.
- Lee, B., Yamanakkanavar, N., Choi, J.Y., 2020. Automatic segmentation of brain MRI using a novel patch-wise U-net deep architecture. *Plos one* 15, e0236493.
- Leventon, M.E., Grimson, W.E.L., Faugeras, O., 2000. Statistical shape influence in geodesic active contours, in: *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)*, pp. 316–323 vol.1. doi:10.1109/CVPR.2000.855835.
- Leventon, M.E., Grimson, W.E.L., Faugeras, O., 2002. Statistical shape influence in geodesic active contours, in: *5th IEEE EMBS International Summer School on Biomedical Imaging, 2002.*, IEEE. p. 8 pp.
- Li, D., Dong, Y., 2014. Deep Learning: Methods and Applications. *Deep Learning: Methods and Applications*, now. URL: <http://ieeexplore.ieee.org/document/8187230>.
- Liang, H., Zhang, F., Niu, X., 2019. Investigating systematic bias in brain age estimation with application to post-traumatic stress disorders. *Human Brain Mapping* 40, 3143–3152. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.24588>, doi:<https://doi.org/10.1002/hbm.24588>.
- Lindberg, O., Walterfang, M., Looi, J.C., Malykhin, N., Ostberg, P., Zandbelt, B., Styner, M., Paniagua, B., Velakoulis, D., Orndahl, E., Wahlund, L.O., 2012. Hippocampal shape analysis in Alzheimer’s disease and frontotemporal lobar degeneration subtypes. *J Alzheimers Dis* 30, 355–65. doi:10.3233/jad-2012-112210.
- Liu, H., Yang, Y., Xia, Y., Zhu, W., Leak, R.K., Wei, Z., Wang, J., Hu, X., 2017. Aging of cerebral white matter. *Ageing Research Reviews* 34, 64–76. URL: <https://www.sciencedirect.com/science/article/pii/S1568163716301192>, doi:<https://doi.org/10.1016/j.arr.2016.11.006>.
- Liu, W., Ruan, D., 2015. A unified variational segmentation framework with a level-set based sparse composite shape prior. *Physics in Medicine and Biology* 60, 1865–1877. URL: <http://dx.doi.org/10.1088/0031-9155/60/5/1865>, doi:10.1088/0031-9155/60/5/1865.

- Liu, Y., Paaajanen, T., Zhang, Y., Westman, E., Wahlund, L.O., Simmons, A., Tunnard, C., Sobow, T., Mecocci, P., Tsolaki, M., Vellas, B., Muehlboeck, S., Evans, A., Spenger, C., Lovestone, S., Soininen, H., 2010. Analysis of regional MRI volumes and thicknesses as predictors of conversion from mild cognitive impairment to Alzheimer's disease. *Neurobiol Aging* 31, 1375–85. doi:10.1016/j.neurobiolaging.2010.01.022.
- Liu, Y., Spulber, G., Lehtimäki, K.K., Könönen, M., Hallikainen, I., Gröhn, H., Kivipelto, M., Hallikainen, M., Vanninen, R., Soininen, H., 2011. Diffusion tensor imaging and tract-based spatial statistics in Alzheimer's disease and mild cognitive impairment. *Neurobiol Aging* 32, 1558–71. doi:10.1016/j.neurobiolaging.2009.10.006.
- Ljunggren, S., 1983. A simple graphical representation of fourier-based imaging methods. *Journal of Magnetic Resonance* (1969) 54, 338–343. URL: <http://www.sciencedirect.com/science/article/pii/0022236483900604>, doi:[https://doi.org/10.1016/0022-2364\(83\)90060-4](https://doi.org/10.1016/0022-2364(83)90060-4).
- Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440.
- Lovestone, S., Francis, P., Kloszewska, I., Mecocci, P., Simmons, A., Soininen, H., Spenger, C., Tsolaki, M., Vellas, B., Wahlund, L.O., Ward, M., 2009. AddNeuroMed—the European collaboration for the discovery of novel biomarkers for Alzheimer's disease. *Ann N Y Acad Sci* 1180, 36–46. doi:10.1111/j.1749-6632.2009.05064.x.
- Lundervold, A.S., Lundervold, A., 2019. An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik* 29, 102–127.
- Löwe, L.C., Gaser, C., Franke, K., Initiative, A.D.N., 2016. The effect of the APOE genotype on individual BrainAGE in normal aging, mild cognitive impairment, and Alzheimer's disease. *PloS one* 11.
- MacNicol, E., Ciric, R., Kim, E., DiCenso, D., Cash, D., Poldrack, R., Esteban, O., 2020. Atlas-based brain extraction is robust across rat MRI studies .
- MacNicol, E., Wright, P., Kim, E., Brusini, I., Esteban, O., Simmons, C., Turkheimer, F., Cash, D., 2021. Age-specific adult rat brain MRI templates and tissue probability maps .
- Magnotta, V.A., Heckel, D., Andreasen, N.C., Cizadlo, T., Corson, P.W., Ehrhardt, J.C., Yuh, W.T.C., 1999. Measurement of brain structures with artificial neural networks: Two- and three-dimensional applications. *Radiology*

## BIBLIOGRAPHY

- 211, 781–790. URL: <https://pubs.rsna.org/doi/abs/10.1148/radiology.211.3.r99ma07781>, doi:10.1148/radiology.211.3.r99ma07781.
- Mahbod, A., Chowdhury, M., Smedby, Ö., Wang, C., 2018. Automatic brain segmentation using artificial neural networks with shape context. *Pattern Recognition Letters* 101, 74–79. URL: <http://www.sciencedirect.com/science/article/pii/S0167865517304312>, doi:<https://doi.org/10.1016/j.patrec.2017.11.016>.
- Martin, B., Mattson, M.P., Maudsley, S., 2006. Caloric restriction and intermittent fasting: two potential diets for successful brain aging. *Ageing research reviews* 5, 332–353.
- Mattson, M.P., 2010. The impact of dietary energy intake on cognitive aging. *Frontiers in aging neuroscience* 2, 5.
- Mehta, R., Arbel, T., 2019. 3D U-Net for brain tumour segmentation, Springer International Publishing. pp. 254–266.
- Milletari, F., Ahmadi, S.A., Kroll, C., Plate, A., Rozanski, V., Maiostre, J., Levin, J., Dietrich, O., Ertl-Wagner, B., Bötzel, K., Navab, N., 2017. Hough-CNN: Deep learning for segmentation of deep brain regions in mri and ultrasound. *Computer Vision and Image Understanding* 164, 92–102. URL: <https://www.sciencedirect.com/science/article/pii/S1077314217300620>, doi:<https://doi.org/10.1016/j.cviu.2017.04.002>.
- Milletari, F., Navab, N., Ahmadi, S.A., 2016. V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision (3DV), IEEE. pp. 565–571.
- Mirikharaji, Z., Izadi, S., Kawahara, J., Hamarneh, G., 2018. Deep auto-context fully convolutional neural network for skin lesion segmentation, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 877–880. doi:10.1109/ISBI.2018.8363711.
- Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C., Ourselin, S., 2010. Fast free-form deformation using graphics processing units. *Comput Methods Programs Biomed* 98, 278–84. doi:10.1016/j.cmpb.2009.09.002.
- Mukherjee, P., Berman, J., Chung, S., Hess, C., Henry, R., 2008. Diffusion tensor MR imaging and fiber tractography: theoretic underpinnings. *American Journal of Neuroradiology* 29, 632–641.

- Muñoz-Moreno, E., Arbat-Plana, A., Batalle, D., Soria, G., Illa, M., Prats-Galino, A., Eixarch, E., Gratacos, E., 2013. A magnetic resonance image based atlas of the rabbit brain for automatic parcellation. *PLoS One* 8, e67418.
- Nenadić, I., Dietzek, M., Langbein, K., Sauer, H., Gaser, C., 2017. BrainAGE score indicates accelerated brain aging in schizophrenia, but not bipolar disorder. *Psychiatry Research: Neuroimaging* 266, 86–89.
- Northcutt, R.G., 2002. Understanding vertebrate brain evolution. *Integrative and Comparative Biology* 42, 743–756. URL: <https://doi.org/10.1093/icb/42.4.743>, doi:10.1093/icb/42.4.743.
- Oishi, K., Mielke, M.M., Albert, M., Lyketsos, C.G., Mori, S., 2012. The fornix sign: a potential sign for Alzheimer’s disease based on diffusion tensor imaging. *J Neuroimaging* 22, 365–74. doi:10.1111/j.1552-6569.2011.00633.x.
- Ouellette, R., Bergendal, Å., Shams, S., Martola, J., Mainero, C., Kristoffersen Wiberg, M., Fredrikson, S., Granberg, T., 2018. Lesion accumulation is predictive of long-term cognitive decline in multiple sclerosis. *Multiple Sclerosis and Related Disorders* 21, 110–116. URL: <https://www.sciencedirect.com/science/article/pii/S2211034818300907>, doi:<https://doi.org/10.1016/j.msard.2018.03.002>.
- Pardoe, H.R., Cole, J.H., Blackmon, K., Thesen, T., Kuzniecky, R., Human Epilepsy Project Investigators, 2017. Structural brain changes in medically refractory focal epilepsy resemble premature brain aging. *Epilepsy research* 133, 28–32.
- Pardoe, H.R., Kuzniecky, R., 2018. NAPR: a cloud-based framework for neuroanatomical age prediction. *Neuroinformatics* 16, 43–49. URL: <https://doi.org/10.1007/s12021-017-9346-9>, doi:10.1007/s12021-017-9346-9.
- Park, G., Hong, J., Lee, J.M., 2018a. Corpus callosum segmentation using deep neural networks with prior information from multi-atlas images. volume 10579 of *SPIE Medical Imaging*. SPIE. URL: <https://doi.org/10.1117/12.2293568>.
- Park, G., Kwak, K., Seo, S.W., Lee, J.M., 2018b. Automatic segmentation of corpus callosum in midsagittal based on bayesian inference consisting of sparse representation error and multi-atlas voting. *Frontiers in Neuroscience* 12. URL: <https://www.frontiersin.org/article/10.3389/fnins.2018.00629>, doi:10.3389/fnins.2018.00629.
- Platten, M., Brusini, I., Andersson, O., Ouellette, R., Piehl, F., Wang, C., Granberg, T., 2021. Deep learning corpus callosum segmentation as a neurodegenerative marker in multiple sclerosis. *Journal of Neuroimaging* 31, 493–

## BIBLIOGRAPHY

500. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/jon.12838>, doi:<https://doi.org/10.1111/jon.12838>.
- Poulakis, K., Pereira, J.B., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Soininen, H., Lovestone, S., Simmons, A., Wahlund, L.O., Westman, E., 2018. Heterogeneous patterns of brain atrophy in Alzheimer's disease. *Neurobiology of Aging* 65, 98–108. URL: <http://www.sciencedirect.com/science/article/pii/S0197458018300174>, doi:<https://doi.org/10.1016/j.neurobiolaging.2018.01.009>.
- Purcell, E.M., Torrey, H.C., Pound, R.V., 1946. Resonance absorption by nuclear magnetic moments in a solid. *J Physical review* 69, 37.
- Purves, D., 2018. *Neuroscience*. 6th edition ed., Sinauer Associates, New York.
- Rasmussen, C.E., 2003. Gaussian processes in machine learning, in: *Summer school on machine learning*, Springer. pp. 63–71.
- Ravishankar, H., Venkataramani, R., Thiruvankadam, S., Sudhakar, P., Vaidya, V., 2017. Learning and incorporating shape models for semantic segmentation, in: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (Eds.), *Medical Image Computing and Computer Assisted Intervention MICCAI 2017*, Springer International Publishing. pp. 203–211.
- Ripich, D.N., Horner, J., 2004. The neurodegenerative dementias: Diagnoses and interventions. *The ASHA Leader* 9, 4–15.
- Ronan, L., Alexander-Bloch, A.F., Wagstyl, K., Farooqi, S., Brayne, C., Tyler, L.K., Fletcher, P.C., 2016. Obesity associated with increased brain age from midlife. *Neurobiology of aging* 47, 63–70.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional networks for biomedical image segmentation, in: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Springer International Publishing. pp. 234–241.
- Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J., 1999. Nonrigid registration using free-form deformations: application to breast mr images. *IEEE Transactions on Medical Imaging* 18, 712–721.
- Równiak, M., Robak, A., Sztejn, S., Bogus-Nowakowska, K., Wasilewska, B., Najdzion, J., 2007. The morphometric study of the amygdala in the rabbit. *Folia Morphol (Warsz)* 66, 44–53.
- Sabuncu, M.R., Yeo, B.T.T., Leemput, K.V., Fischl, B., Golland, P., 2010. A generative model for image segmentation based on label fusion. *IEEE Transactions on Medical Imaging* 29, 1714–1729. doi:10.1109/TMI.2010.2050897.

- Sacco, R., Bisecco, A., Corbo, D., Della Corte, M., d'Ambrosio, A., Docimo, R., Gallo, A., Esposito, F., Esposito, S., Cirillo, M., Lavorgna, L., Tedeschi, G., Bonavita, S., 2015. Cognitive impairment and memory disorders in relapsing-remitting multiple sclerosis: the role of white matter, gray matter and hippocampus. *J Neurol* 262, 1691–7. doi:10.1007/s00415-015-7763-y.
- Scheltens, P., Fox, N., Barkhof, F., De Carli, C., 2002. Structural magnetic resonance imaging in the practical assessment of dementia: beyond exclusion. *Lancet Neurol* 1, 13–21.
- Serdar, C.C., Cihan, M., Yücel, D., Serdar, M.A., 2021. Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. *Biochemia medica* 31, 010502–010502. URL: <https://pubmed.ncbi.nlm.nih.gov/33380887><https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7745163/>, doi:10.11613/BM.2021.010502.
- Sharma, S., Majsak, M.J., 2014. *Brain Anatomy*. Academic Press, Oxford. pp. 463–466. URL: <http://www.sciencedirect.com/science/article/pii/B9780123851574011210>, doi:<https://doi.org/10.1016/B978-0-12-385157-4.01121-0>.
- Sigirli, D., Ercan, I., Ozdemir, S.T., Taskapilioglu, O., Hakyemez, B., Turan, O.F., 2012. Shape analysis of the corpus callosum and cerebellum in female MS patients with different clinical phenotypes. *Anat Rec (Hoboken)* 295, 1202–11. doi:10.1002/ar.22493.
- Simeone, O., 2017. A brief introduction to machine learning for engineers. arXiv preprint arXiv:1709.02840 .
- Simmons, A., Westman, E., Muehlboeck, J.S., Mecocci, P., Vellas, B., Tsolaki, M., Kłoszewska, I., Wahlund, L.O., Soininen, H., Lovestone, S., Evans, A., Spenger, C., 2009. MRI Measures of Alzheimer's Disease and the AddNeuroMed Study. volume 1180 of *Annals of the New York Academy of Sciences*. doi:10.1111/j.1749-6632.2009.05063.x.
- Smith, A., 1973. Symbol digit modalities test. Western Psychological Services Los Angeles.
- Smith, S.M., 2002. Fast robust automated brain extraction. *Human brain mapping* 17, 143–155.
- Smith, S.M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T.E., Mackay, C.E., Watkins, K.E., Ciccarelli, O., Cader, M.Z., Matthews, P.M., 2006. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *NeuroImage* 31, 1487–1505.



## BIBLIOGRAPHY

- Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* 23 Suppl 1, S208–19. doi:10.1016/j.neuroimage.2004.07.051.
- Smith, S.M., Nichols, T.E., 2009. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* 44, 83–98. doi:10.1016/j.neuroimage.2008.03.061.
- Speisman, R.B., Kumar, A., Rani, A., Pastoriza, J.M., Severance, J.E., Foster, T.C., Ormerod, B.K., 2013. Environmental enrichment restores neurogenesis and rapid acquisition in aged rats. *Neurobiology of aging* 34, 263–274.
- Steffener, J., Habeck, C., O’Shea, D., Razlighi, Q., Bherer, L., Stern, Y., 2016. Differences between chronological and brain age are related to education and self-reported physical activity. *Neurobiology of aging* 40, 138–144.
- Stejskal, E.O., Tanner, J.E., 1965. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *The Journal of Chemical Physics* 42, 288–292.
- Swanson, L.W., 2013. Basic principles of nervous system organization , 1255–1288.
- Swerdlow, R.H., 2011. Brain aging, Alzheimer’s disease, and mitochondria. *J Biochimica et Biophysica Acta - Molecular Basis of Disease* 1812, 1630–1639.
- Ségonne, F., Dale, A.M., Busa, E., Glessner, M., Salat, D., Hahn, H.K., Fischl, B., 2004. A hybrid approach to the skull stripping problem in MRI. *Neuroimage* 22, 1060–1075.
- Tang, M., Valipour, S., Zhang, Z., Cobzas, D., Jagersand, M., 2017. A deep level set method for image segmentation, in: Cardoso, M.J., Arbel, T., Carneiro, G., Syeda-Mahmood, T., Tavares, J.M.R.S., Moradi, M., Bradley, A., Greenspan, H., Papa, J.P., Madabhushi, A., Nascimento, J.C., Cardoso, J.S., Belagiannis, V., Lu, Z. (Eds.), *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer International Publishing. pp. 126–134.
- Terribilli, D., Schaufelberger, M.S., Duran, F.L., Zanetti, M.V., Curiati, P.K., Menezes, P.R., Scazufca, M., Amaro Jr, E., Leite, C.C., Busatto, G.F., 2011. Age-related gray matter volume changes in the brain during non-elderly adulthood. *Neurobiology of aging* 32, 354–368.

- the United Nations Department of Economic and Social Affairs, 2020. “World Population Ageing 2020 Highlights”. Report.
- Thyreau, B., Sato, K., Fukuda, H., Taki, Y., 2018. Segmentation of the hippocampus by transferring algorithmic knowledge for large cohort processing. *Med Image Anal* 43, 214–228. doi:10.1016/j.media.2017.11.004.
- Tipping, M.E., 2000. The relevance vector machine, in: *Advances in neural information processing systems*, pp. 652–658.
- Tipping, M.E.J.J.o.m.l.r., 2001. Sparse bayesian learning and the relevance vector machine 1, 211–244.
- Tu, Z., Bai, X., 2010. Auto-context and its application to high-level vision tasks and 3D brain image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, 1744–1757. doi:10.1109/TPAMI.2009.186.
- Twieg, D.B., 1983. The k-trajectory formulation of the NMR imaging process with applications in analysis and synthesis of imaging methods. *J Medical physics* 10, 610–621.
- United Nations, 2022. Goal 3: Ensure healthy lives and promote well-being for all at all ages. URL: <https://www.un.org/sustainabledevelopment/health/>. Accessed on February 17, 2022.
- Urry, H.L., van Reekum, C.M., Johnstone, T., Kalin, N.H., Thurow, M.E., Schaefer, H.S., Jackson, C.A., Frye, C.J., Greischar, L.L., Alexander, A.L., Davidson, R.J., 2006. Amygdala and ventromedial prefrontal cortex are inversely coupled during regulation of negative affect and predict the diurnal pattern of cortisol secretion among older adults. *J Neurosci* 26, 4415–25. doi:10.1523/jneurosci.3215-05.2006.
- Valli, G., Coppini, G., 2005. Bioimmagini. Pàtron.
- Van Schependom, J., Gielen, J., Laton, J., Sotiropoulos, G., Vanbinst, A.M., De Mey, J., Smeets, D., Nagels, G., 2017. The effect of morphological and microstructural integrity of the corpus callosum on cognition, fatigue and depression in mildly disabled MS patients. *Magnetic Resonance Imaging* 40, 109–114. URL: <https://www.sciencedirect.com/science/article/pii/S0730725X17300814>, doi:<https://doi.org/10.1016/j.mri.2017.04.010>.
- Varghese, T., Sheelakumari, R., James, J.S., Mathuranath, P., 2013. A review of neuroimaging biomarkers of Alzheimer’s disease. *Neurology Asia* 18, 239–248. URL: <https://pubmed.ncbi.nlm.nih.gov/25431627><https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4243931/>.

## BIBLIOGRAPHY

- Varoquaux, G., Cheplygina, V., 2021. How I failed machine learning in medical imaging—shortcomings and recommendations. arXiv preprint arXiv:2103.10292 .
- Vermunt, J.K., 2010. Latent Class Models. Elsevier, Oxford. pp. 238–244. URL: <https://www.sciencedirect.com/science/article/pii/B9780080448947013403>, doi:<https://doi.org/10.1016/B978-0-08-044894-7.01340-3>.
- Vidal-Pineiro, D., Wang, Y., Krogsrud, S.K., Amlien, I.K., Baaré, W.F.C., Bartres-Faz, D., Bertram, L., Brandmaier, A.M., Drevon, C.A., Düzel, S., Ebmeier, K., Henson, R.N., Junqué, C., Kievit, R.A., Kühn, S., Leonardsen, E., Lindenberg, U., Madsen, K.S., Magnussen, F., Mowinckel, A.M., Nyberg, L., Roe, J.M., Segura, B., Smith, S.M., Sørensen, O., Suri, S., Westhausen, R., Zalesky, A., Zsoldos, E., Walhovd, K.B., Fjell, A., 2021. Individual variations in ‘brain age’ relate to early-life factors more than to longitudinal brain change. eLife 10, e69995. URL: <https://doi.org/10.7554/eLife.69995>, doi:10.7554/eLife.69995.
- Voevodskaya, O., Simmons, A., Nordenskjöld, R., Kullberg, J., Ahlström, H., Lind, L., Wahlund, L.O., Larsson, E.M., Westman, E., Alzheimer’s Disease Neuroimaging Initiative, 2014. The effects of intracranial volume adjustment approaches on multiple regional MRI volumes in healthy aging and Alzheimer’s disease. Frontiers in aging neuroscience 6, 264–264. URL: <https://www.ncbi.nlm.nih.gov/pubmed/25339897><https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4188138/>, doi:10.3389/fnagi.2014.00264.
- Walhovd, K.B., Fjell, A.M., Reinvang, I., Lundervold, A., Dale, A.M., Eilertsen, D.E., Quinn, B.T., Salat, D., Makris, N., Fischl, B., 2005. Effects of age on volumes of cortex, white matter and subcortical structures. Neurobiology of Aging 26, 1261–1270. URL: <https://www.sciencedirect.com/science/article/pii/S0197458005001673>, doi:<https://doi.org/10.1016/j.neurobiolaging.2005.05.020>.
- Wang, C., Frimmel, H., Smedby, Ö., 2014. Fast level-set based image segmentation using coherent propagation. Medical physics 41, 073501.
- Wang, C., Smedby, Ö., 2017. Automatic whole heart segmentation using deep learning and shape context, in: International Workshop on Statistical Atlases and Computational Models of the Heart, Springer. pp. 242–249.
- Wang, J., 2020. An intuitive tutorial to gaussian processes regression. arXiv preprint arXiv:2009.10862 .

- Wansapura, J.P., Holland, S.K., Dunn, R.S., Ball Jr, W.S., 1999. NMR relaxation times in the human brain at 3.0 tesla. *Journal of Magnetic Resonance Imaging* 9, 531–538.
- Wenyan, C., Shutao, L., Fucang, J., Xiaodong, Z., 2011. Segmentation of hippocampus based on ROI atlas registration, in: 2011 IEEE International Symposium on IT in Medicine and Education, pp. 226–230. doi:10.1109/ITiME.2011.6130821.
- Wernick, M.N., Yang, Y., Brankov, J.G., Yourganov, G., Strother, S.C., 2010. Machine learning in medical imaging. *IEEE signal processing magazine* 27, 25–38.
- Westman, E., Simmons, A., Zhang, Y., Muehlboeck, J.S., Tunnard, C., Liu, Y., Collins, L., Evans, A., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Soininen, H., Lovestone, S., Spenger, C., Wahlund, L.O., 2011. Multivariate analysis of MRI data for Alzheimer’s disease, mild cognitive impairment and healthy controls. *Neuroimage* 54, 1178–87. doi:10.1016/j.neuroimage.2010.08.044.
- Winkelmann, T., Grimm, O., Pohlack, S.T., Nees, F., Cacciaglia, R., Dinu-Biringer, R., Steiger, F., Wicking, M., Ruttorf, M., Schad, L.R., Flor, H., 2016. Brain morphology correlates of interindividual differences in conditioned fear acquisition and extinction learning. *Brain Struct Funct* 221, 1927–37. doi:10.1007/s00429-015-1013-z.
- Woods, R.P., 2009. Registration. Academic Press, Burlington. pp. 495–497. URL: <http://www.sciencedirect.com/science/article/pii/B9780123739049500386>, doi:<https://doi.org/10.1016/B978-012373904-9.50038-6>.
- Wu, J., Zhang, Y., Wang, K., Tang, X., 2019. Skip connection U-Net for white matter hyperintensities segmentation from MRI. *IEEE Access* 7, 155194–155202.
- Yang, H., Sun, J., Li, H., Wang, L., Xu, Z., 2018. Neural multi-atlas label fusion: Application to cardiac MR images. *Medical Image Analysis* 49, 60–75. URL: <https://www.sciencedirect.com/science/article/pii/S136184151830553X>, doi:<https://doi.org/10.1016/j.media.2018.07.009>.
- Young, P.N.E., Estarellas, M., Coomans, E., Srikrishna, M., Beaumont, H., Maass, A., Venkataraman, A.V., Lissaman, R., Jiménez, D., Betts, M.J., McGlinchey, E., Berron, D., O’Connor, A., Fox, N.C., Pereira, J.B., Jagust, W., Carter, S.F., Paterson, R.W., Schöll, M., 2020. Imaging biomarkers in

## BIBLIOGRAPHY

- neurodegeneration: current and future practices. *Alzheimer's Research Therapy* 12, 49. URL: <https://doi.org/10.1186/s13195-020-00612-7>, doi:10.1186/s13195-020-00612-7.
- Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., 2015. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* 347, 1138–1142.
- Zhang, N., Deng, Z.S., Wang, F., Wang, X.Y., 2009. The effect of different number of diffusion gradients on SNR of diffusion tensor-derived measurement maps. *Journal of Biomedical Science Engineering* 2, 96.
- Zhu, H., 2003. Medical image processing overview .
- Zhu, H., Wang, S., Qu, L., Shen, D., 2021. Chapter 10 - Hippocampus segmentation in MR images: Multiatlas methods and deep learning methods. Academic Press. pp. 181–215. URL: <https://www.sciencedirect.com/science/article/pii/B9780128228845000192>, doi:<https://doi.org/10.1016/B978-0-12-822884-5.00019-2>.
- Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3D U-Net: learning dense volumetric segmentation from sparse annotation, in: International conference on medical image computing and computer-assisted intervention, Springer. pp. 424–432.
- İçer, S., 2013. Automatic segmentation of corpus collasum using gaussian mixture modeling and fuzzy C means methods. *Computer Methods and Programs in Biomedicine* 112, 38–46. URL: <https://www.sciencedirect.com/science/article/pii/S0169260713001922>, doi:<https://doi.org/10.1016/j.cmpb.2013.06.006>.

## Appended papers