



Universitat de Girona

ATLAS-BASED SEGMENTATION OF MULTIPLE SCLEROSIS LESIONS IN MAGNETIC RESONANCE IMAGING

Mariano CABEZAS GREBOL

Dipòsit legal: Gi. 1121-2013

<http://hdl.handle.net/10803/119608>

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. Access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.



Universitat de Girona

PhD Thesis

Atlas-based segmentation of multiple
sclerosis lesions in magnetic resonance
imaging

Mariano Cabezas Grebol

2013



Universitat de Girona

PhD Thesis

**Atlas-based segmentation of multiple
sclerosis lesions in magnetic resonance
imaging**

Mariano Cabezas Grebol

2013

Doctoral Programme in Technology

Supervised by:

Xavier Lladó and Arnau Oliver

Work submitted to the University of Girona in partial fulfilment of the
requirements for the degree of Doctor of Philosophy

Agraïments

I després de moltes hores, dies, mesos i anys, arriba el moment definitiu. El moment final. El punt en el que ja tot està escrit, però encara queden coses a dir. Agraïments. Aquella part que sempre costa d'escriure, i no només per el fet d'haver d'agrair. Aquella secció que tots deixem pel final, però que apareix al principi de la majoria de tesis al passar la portada. Aquell text que sovint serveix per repassar el camí que ens ha portat fins a aquest document que avui tens entre mans.

Com tot camí personal comença amb un naixement, el meu i una família que sempre m'ha fet costat i a qui dono les gràcies per haver-me acompanyat fins aquí i haver-me aguantat tants anys seguits. El viatge segueix a l'escola que va despertar en mi l'interès per la informàtica, la ciència i les matemàtiques que a dia d'avui han fet possible que aquest document estigui complet. Per tant, agraeixo a tots els meus companys i professors d'escola i institut tot el que m'han aportat.

Però, el viatge de debó no comença fins la meva arribada a la Universitat de Girona. Aquí és on vaig conèixer els meus actuals companys amb els quals hem compartit diversos anys d'estudis i el grup VICOROB amb aquella primera classe de Visió per Computador amb el meu actual director, el Dr. Xavier Lladó. Des d'aquí per tant, m'agradaria també agrair un per un als meus companys: en Ricard, que ha hagut de patir-me com a company de pràctiques i de grup segurament massa vegades; en Pablo, amb qui hem discutit durant hores absurditats que després han quedat oblidades amb unes cerveses i/o whiskys; l'Albert (Gubern) que ens ha abandonat per viure a les meravelloses terres holandeses; en Pla, amb el seu entranyable humor negre; en Pons, amb qui hem parlat hores sobre teories a series que al final han resultat ser millors que el que els guionistes ens tenien preparat; en Massi, que tot i començar un any més tard el doctorat s'ha fet indispensable en els dinars dels dijous (fins al punt de ser l'únic fidel!); l'Enric, perdut per la piscina però sempre a punt per apuntar-se a tot; i els meus companys de cervells, l'Onur, amb qui he compartit grans moments a l'ECTRIMS (sobretot Amsterdam), l'Eloy, la jove promesa expert en negocis,

en Sergi, aquell noi de lletres que puja muntanyes i en Yago, que tot i també treballar en pit, s'ha hagut de barallar amb imatges de cervell. També voldria agrair a en Bertu per ajudar-me amb el Boosting i compartir aquesta recta final cap al títol de doctor.

Ja que hi som, faig extensible l'agraïment a tot el grup VICOROB, els companys de sala i a la joventut que cada setmana dóna suport a les meves sessions de bogeria i cervesa. Unes sessions que no serien possibles sense la Meritxell, la meva companya del Beer Team i tesorera oficial. Òbviament no m'oblido dels meus directors de tesi, que després de 3 anys encara tenen ganes d'enfrontar-se a la meva tossuderia i deixar que aquesta tesi sigui defensada. Em refereixo al Dr. Xavier Lladó i el Dr. Arnau Oliver. Gràcies per tot, per que la paciència que heu demostrat al llarg dels anys, a mi em manca (com ja ho sabeu!). Gràcies també al Dr. Jordi Freixenet i en Jordi Gich, ja que sense aquella primera oportunitat per treballar amb el projecte EM-Line seguint els passos d'en Quintana res d'això hauria estat possible. I ja que hi som, gràcies també a tot l'equip d'en Lluís Ramió de l'Hospital Josep Trueta, l'equip de l'Àlex Rovira de l'hospital Vall d'Hebrón i l'equip d'en Joan C. Vilanova de la Clínica Girona per fer possible el projecte SALEM que ens ha permès obrir aquesta meravellosa línia de recerca i assistir any rere any a l'ECTRIMS.

Finalment, agraeixo a la meva parella, la Tina, tot el suport que em dóna dia a dia, ja que sense ella les coses serien molt diferents. Durant aquests més de tres anys junts, ella ha estat una part molt important i indispensable del viatge!

Acknowledgments

If you didn't feel acknowledged, you missed your name or you didn't understand anything this text here has been written exclusively for you. My trip through computer vision may have started in catalan in Girona, but it wouldn't have been possible without international help. Of course I'm talking about Lausanne and my beautiful research stay at the EPFL under the magnificent supervision of Dr. Meritxell Bach Cuadra. I loved my stay there, I met a lot of interesting people and I learned more than I could have ever expected. Therefore I thank all the LTS groups (even though I know some of you don't have an LTS name anymore) and all the people that crossed my path there. I have great memories that helped build this book.

I would also like to thank all the VICOROB team, once again in English for those that couldn't understand (yet, I know you're learning catalan) the previous pages. I have more or less interacted with quite a few people during the different events the group carried out and I'm glad I did. These events might not have inspired my research or point me to a great direction but they made the walk enjoyable.

Obviously, I should also be thanking you if you are reviewing this text. I just lacked the words to make these acknowledgments worthwhile before the first deposit, but I thank both the reviewers and the jury for taking the time to read (and care about) my work. I hope you enjoy it as much as I did.

This is quite an special part because with this lines I would also like to thank all the producers, writers, showrunners, actors, etc. that made possible all the TV shows and movies I've been enjoying for the past three years. They have been an inspiration (of sorts) and they've helped improve my English. As you will see through each chapter my appreciation does not stop in this acknowledgments as I try to subtly use quotes from some of my favorite shows to *introduce* in an obscure way each topic.

And finally, if you were not mentioned before and you are reading this work out of

interest, first why are you doing it? And second, thank you for supporting my work.

So, again, I want to thank you who are reading this (and allow me to break the fourth wall once more) for the support during this personal trip. Now it's your time to experience it as I did through the pages of this document. Take my hand and follow me down the rabbit hole...

Publications

Journals

- **[INFSCI 2013]** Mariano Cabezas, Arnau Oliver, Jordi Freixenet, Joan C. Vilanova, Lluís Ramió-Torrentà, Àlex Rovira, Xavier Lladó. “BOOST: Boosting with outliers and other spatial tools to segment multiple sclerosis lesion”. *Information Sciences*, submitted.
- **[NR 2013]** Mariano Cabezas, Arnau Oliver, Jordi Freixenet, Joan C. Vilanova, Lluís Ramió-Torrentà, Àlex Rovira, Xavier Lladó. “Tissue inference by statistical segmentation using expectation-maximisation and lesion outlier thresholding”. *Neuroradiology*, submitted.
- **[HBM 2013]** Sergi Valverde, Arnau Oliver, Yago Díez, Mariano Cabezas, Joan C. Vilanova, Àlex Rovira, Lluís Ramió-Torrentà, Xavier Lladó. “Evaluating the Effects of White Matter Lesions on Six Brain Tissue Segmentation methods”. *Human Brain Mapping*, submitted.
- **[INFSCI 2012]** X. Lladó, A. Oliver, M. Cabezas, J. Freixenet, J.C. Vilanova, A. Quiles, L. Valls, Ll. Ramió-Torrentà, A. Rovira. “Segmentation of multiple sclerosis lesions in brain MRI: a review of automated approaches”. *Information Sciences*, 186(1), pp. 164-185. 2012.
- **[CMPB 2011]** M. Cabezas, A. Oliver, X. Lladó, J. Freixenet, M. Bach-Cuadra. “A review of atlas-based segmentation for magnetic resonance brain images” *Computer Methods and Programs in Biomedicine*, 104(3), pp e158-e177. 2011.

Conferences

- **[IbPRIA2013]** M. Cabezas, A. Oliver, J. Freixenet and X. Lladó. “A supervised approach for multiple sclerosis lesion segmentation using context features and an outlier map”. Iberian Conference on Pattern Recognition and Image Analysis. LNCS, To appear. Madeira, Portugal, 2013.
- **[SEN 2012]** X. Lladó, M. Cabezas, O. Ganiler, A. Oliver, Y. Donoso, J. Freixenet, L. Valls, A. Quiles, G. Laguillo, D. Pareto, J.C. Vilanova, A. Rovira, Ll. Ramió-Torrentà. “SALEM: Herramientas informáticas para la detección de lesiones de esclerosis múltiple en estudios longitudinales mediante imágenes de resonancia magnética del cerebro”. Sociedad Española de Neurología. Neurología, 27(Num. Esp.Congreso), pp 195-196. Barcelona, Spain, 2012.
- **[ECTRIMS 2012]** M. Cabezas, A. Oliver, X. Lladó, Y. Díez, J. Freixenet, J.C. Vilanova, A. Quiles, G. Laguillo, Ll. Ramió-Torrentà, D. Pareto, and A. Rovira. “A supervised approach to segment multiple sclerosis lesions using context-rich features and a boosting classifier”. European Committee for Treatment and Research in Multiple Sclerosis conference. Multiple Sclerosis, 18(4S). Lyon, France. October 2012.
- **[ECTRIMS 2012]** Y. Díez, X. Lladó, A. Oliver, R. Martí, E. Roura, M. Cabezas, O. Ganiler, J. Freixenet, J.C. Vilanova, L. Valls, Ll. Ramió-Torrentà, D. Pareto, and A. Rovira. “Registration of serial brain MRI scans from multiple sclerosis patients. Analysis of 3D intensity-based methods”. European Committee for Treatment and Research in Multiple Sclerosis conference. Multiple Sclerosis, 18(4S). Lyon, France. October 2012.
- **[SPIE 2012]** G. Pons, J. Martí, R. Martí, M. Cabezas, A. Di Battista and J.A. Noble. “Lesion Segmentation and Bias Correction in Breast Ultrasound B-mode Images Including Elastography Information”. SPIE Conference on Medical Imaging, 8314, pp. 83141E1-1E6. San Diego, California, USA, February 2012.
- **[ECTRIMS 2011]** X. Lladó, O. Ganiler, A. Oliver, M. Cabezas, J. Freixenet, J.C. Vilanova, L. Valls, Ll. Ramió-Torrentà, and A. Rovira. “Computer-assisted strategies to automated quantification of multiple sclerosis lesion evolution on brain magnetic resonance imaging”. European Committee for Treatment and Research in Multiple

Sclerosis conference. Multiple Sclerosis, 17(10S), pp 161-162. Amsterdam, Holland, 2011

- **[ECTRIMS 2011]** M. Cabezas, M. Bach-Cuadra, A. Oliver, X. Lladó, J. Freixenet, J.C. Vilanova, L. Valls, Ll. Ramió-Torrentà, E. Huerga, D. Pareto, and A. Rovira. “A pipeline approach with spatial information for segmenting multiple sclerosis lesions on brain magnetic resonance imaging”. European Committee for Treatment and Research in Multiple Sclerosis conference. Multiple Sclerosis, 17(10S), pp 381. Amsterdam, Holland, 2011.
- **[SEN 2011]** X. Lladó, O. Ganiler, A. Oliver, M. Cabezas, J. Freixenet, J.C. Vilanova, L. Valls, Ll. Ramió-Torrentà, and A. Rovira. “Técnicas automáticas de segmentación de lesiones de EM y de cuantificación volumétrica en estudios temporales”. Sociedad Española de Neurología. Neurología, 26(Num. Esp.Congreso), pp 181-182. Barcelona, Spain, 2011.
- **[SEN 2011]** J. Freixenet, M. Cabezas, M. Bach-Cuadra, A. Oliver, X. Lladó, J.C. Vilanova, L. Valls, Ll. Ramió-Torrentà, E. Huerga, D. Pareto, and A. Rovira. “Segmentación de lesiones de esclerosis múltiple en resonancia magnética del cerebro mediante información espacial”. Sociedad Española de Neurología. Neurología, 26(Num. Esp.Congreso), pp 182. 15-19. Barcelona, Spain, Noviembre 2011.
- **[ECTRIMS 2010]** X. Lladó, M. Cabezas, O. Ganiler, A. Oliver, J. Freixenet, J.C. Vilanova, A. Quiles, Ll. Ramió-Torrenta, A. Rovira. “Strategies for Automated Segmentation of Multiple Sclerosis Lesions on Brain Magnetic Resonance Imaging”. European Committee for Treatment and Research in Multiple Sclerosis conference. Multiple Sclerosis, 16(10), pp S256. Gothenburg, Sweden, 2010.

Workshops

- **[JEMGI 2012]** J. Freixenet, X. Lladó, A. Oliver, M. Cabezas, O. Ganiler, Y. Diez, E. Roura, S. Valverde. “Anàlisi d’imatge automatitzat en Esclerosi Múltiple: Identificació de lesions i atròfia”. Ponència invitada a la II Jornada d’Esclerosi Múltiple de Girona celebrada a Pals (Girona), el 13/07/2012.
- **[TIC Salut 2012]** X. Lladó, A. Oliver, J. Freixenet, M. Cabezas, O. Ganiler. “Projecte SALEM: Segmentació de lesions d’esclerosi múltiple”. In III Jornada R+D+i

en TIC Salut realitzada en el Parc Científic i Tecnològic de la Universitat de Girona.
7 i 8 de juny 2012.

- **[MICCAT 2011]** M. Cabezas, A. Oliver, J. Freixenet, X. Lladó. “Segmentation of multiple sclerosis lesions in MRI using spatial information”. In Medical Image Computing in Catalunya: Graduate Student Workshop (non-indexed). Barcelona, October 2011.
- **[JEMGI 2011]** X. Lladó, A. Oliver, J. Freixenet, M. Cabezas, O. Ganiler. “Técnicas de segmentación automática de imágenes de RM Neurológicas”. Ponència invitada a la I Jornada d’Esclerosi Múltiple de Girona celebrada a Pals (Girona), el 15/07/2011.
- **[MICCAT 2010]** M. Cabezas, X. Lladó, A. Oliver, and J. Freixenet. “Strategies for Automated Segmentation of Multiple Sclerosis Lesions on Brain MRI”. In Medical Image Computing in Catalunya: Graduate Student Workshop (non-indexed). Girona, October 2010.

Abstract

This thesis deals with the segmentation of brain magnetic resonance imaging applied to multiple sclerosis patients. This disease is characterised by the presence of white matter lesions in this image modality. After a thorough analysis of the state-of-the-art on this topic, pointing out the importance of prior knowledge, and a subsequent review of atlas-based segmentation of brain imaging, we propose two different multiple sclerosis lesion segmentation pipelines based on the conclusions of these studies. The first one provides an initial tissue classification using a modified expectation-maximisation algorithm, which is later on refined with a lesion segmentation step based on thresholding and a regionwise false positive reduction approach. The second one focuses only on the segmentation of lesions and uses an ensemble classifier alongside a rich feature pool including image intensities, probabilistic atlas maps, an outlier map and contextual information. Both approaches are tested against a novel database comprising imaging data from three different hospitals with a variable lesion load per case. The evaluation, carried out in a quantitative and qualitative manner, includes a comparison and uses several metrics for detection and segmentation. The analysis of the results points out a better performance relative to state-of-the-art approaches, with a clear improvement on the first pipeline in terms of detection, and a clear improvement on the second pipeline in terms of segmentation.

Resum

Aquesta tesi es centra en la segmentació de imatges de ressonància magnètica del cervell aplicada a pacients d'esclerosi múltiple. Aquesta malaltia es caracteritza per l'aparició de lesions de matèria blanca, visibles en aquesta modalitat d'imatge. Després d'un anàlisi exhaustiu de l'estat de l'art en aquest tòpic, remarcant la importància de la informació prèvia, i també de la segmentació basada en atlas del cervell, proposem dues estratègies diferents per a la segmentació de lesions basades en les conclusions d'ambdós estudis. La primera proporciona una classificació inicial dels teixits mitjançant una extensió de l'algorisme de esperança-maximització, que es refina posteriorment amb un procés de segmentació de les lesions basat en una binarització inicial i una conseqüent estratègia de reducció de falsos positius a nivell de regió. La segona proposta es focalitza bàsicament en la segmentació de lesions i utilitza una combinació de classificadors febles entrenats amb un ric conjunt de característiques que inclou imatges d'intensitat, mapes probabilístics provinents d'un atlas, un mapa d'intensitats atípiques i informació contextual. Ambdues estratègies han estat provades amb una nova base de dades formada per imatges de tres hospitals diferents amb diferent càrrega lesional per cas. L'avaluació d'aquestes proves, que s'ha dut a terme de forma quantitativa i qualitativa, inclou una comparativa i utilitza diferents mètriques de detecció i segmentació. L'anàlisi d'aquests resultats apunta a un millor rendiment relatiu a l'estat de l'art actual, amb una millor detecció per part de la primera estratègia i una millor segmentació per part de la segona.

Resumen

Esta tesis se centra en la segmentación de imágenes de resonancia magnética del cerebro aplicada a pacientes de esclerosis múltiple. Esta enfermedad se caracteriza por la aparición de lesiones de materia blanca, visibles en esta modalidad de imagen. Después de un análisis exhaustivo del estado del arte en este tópico, remarcando la importancia de la información previa, y segmentación basada en atlas del cerebro, proponemos dos estrategias diferentes para la segmentación de lesiones basadas en las conclusiones aportadas por estos estudios. La primera proporciona una clasificación inicial de los tejidos mediante una extensión del algoritmo de esperanza-maximización, que es refinada con un proceso de segmentación de las lesiones basado en binarización y una estrategia de reducción de falsos positivos a nivel de región. La segunda se focaliza básicamente en la segmentación de lesiones y utiliza un agregado de clasificadores entrenado con un rico conjunto de características que incluyen imágenes de intensidad, mapas probabilísticos provenientes de un atlas, un mapa de intensidades atípicas e información contextual. Las dos estrategias han estado validadas con una nueva base de datos formada por imágenes de tres hospitales diferentes con diferente carga lesional por caso. La evaluación de estas pruebas, que se han realizado de forma cuantitativa y cualitativa, incluye una comparativa y utiliza diferentes métricas de detección y segmentación. El análisis de estos resultados apunta a un mejor rendimiento relativo al estado del arte actual, con una mejor detección por parte de la primera estrategia y una mejor segmentación por parte de la segunda.

List of Figures

1.1	Segmentation scheme	2
1.2	Lesion segmentation example	3
1.3	MRI scanner scheme	4
1.4	MRI examples	6
1.5	Bias field and non tissue example	7
1.6	Partial volume effect example.	8
2.1	Flowchart of the supervised strategies based on atlas.	16
2.2	Flowchart of the supervised strategies based on learning	21
2.3	Flowchart of the unsupervised strategies based on tissue properties.	24
2.4	Flowchart of the unsupervised strategies based on lesion properties.	25
2.5	Advantages and drawbacks of the different MS lesion segmentation strategies	28
2.6	MS lesion segmentation examples	32
2.7	Generated 3D volume with MS lesions segmented by two different experts .	33
2.8	Results extracted from the works presented in the 2008 MS Challenge . . .	39
2.9	Generated 3D volume with different lesion load.	42
3.1	ICBM452 population-based atlas	49
3.2	Advantages and drawbacks of the different atlas-based segmentation strategies	53
3.3	Diagram of how atlases are used	55
3.4	Internal structures of the brain	57
3.5	Example on fetus, neonates and elderly subjects	64

3.6	Focal tissue lesions	67
3.7	Space-occupying lesions	70
4.1	Preprocessing pipeline flowchart	78
4.2	Skull stripping example	79
4.3	Denoising and bias correction example	81
4.4	Example of atlas registration	83
4.5	Flowchart of the first pipeline used to segment tissues and lesions	85
4.6	Example of tissue segmentation	90
4.7	FLAIR histogram for the GM tissue	92
4.8	Example of the TISSUE-LOT pipeline segmentation	94
4.9	Flowchart of the second pipeline based on a trained classifier	95
4.10	Outlier map	96
4.11	Meta-features	97
4.12	Example of a 2D boosting training	99
4.13	Example of the BOOST pipeline segmentation	101
5.1	Examples of MS challenge 2008	107
5.2	Examples of images from Hospital Vall d’Hebron	109
5.3	Examples of images from Hospital Josep Trueta	110
5.4	Examples of images from Clínica Girona	111
5.5	Lesion load per hospital	112
5.6	Effect of γ during thresholding	117
5.7	Tissue labels for GT lesions and neighboring voxels	118
5.8	Regionwise TPF vs FPF for the TISSUE-LOT pipeline	119
5.9	Regionwise DSC for the TISSUE-LOT pipeline	120
5.10	Voxelwise TPF vs FPF for the TISSUE-LOT pipeline	121
5.11	Voxelwise DSC for the TISSUE-LOT pipeline	122
5.12	Average surface distance and average DSC for all lesions for TISSUE-LOT	123

5.13	Comparison between using or not using our outlier map as a feature	125
5.14	Chart of the boosting selected features	127
5.15	Regionwise TPF vs FPF for the BOOST pipeline	128
5.16	Regionwise DSC for the BOOST pipeline	129
5.17	Voxelwise TPF vs FPF for the BOOST pipeline	130
5.18	Voxelwise DSC for the BOOST pipeline	131
5.19	Average surface distance and average DSC for all lesions for BOOST	132
5.20	Comparison in terms of detection by hospital and method	133
5.21	Comparison in terms of segmentation by hospital and method	134
5.22	Comparison in terms of segmentation by hospital and method	135
5.23	Segmentation comparison for Hospital Vall d'Hebron	136
5.24	Segmentation comparison for Hospital Josep Trueta	137
5.25	Segmentation comparison for Clínica Girona	138
6.1	Examples of lesions in other brain diseases with GT	146
A.1	Previous FLAIR simulation	149
A.2	FLAIR simulation	151
A.3	Lesion segmentation methods	155
A.4	Evaluation measures	159
A.5	Tissue segmentation BW	160
A.6	Evaluation measures	161
A.7	Tissue segmentation example	162
A.8	Lesion in cortex example	163
A.9	Biased Challenge case	164
A.10	Anatomical atlas	168
A.11	Evaluation measures for GC	171
A.12	Evaluation measures for ICM	174
A.13	Evaluation measures for GC and ICM	175

A.14 GC drawbacks example 178

List of Tables

2.1	Summary of the supervised approaches based on atlas.	17
2.2	Summary of the supervised approaches based learning.	22
2.3	Summary of the unsupervised approaches based on tissue.	24
2.4	Summary of the unsupervised approaches based on lesion properties.	26
2.5	Common measures used in the evaluation of MS lesion segmentation methods.	34
2.6	Summary of the results presented in the articles analysed	36
3.1	Automated atlas-based segmentation methods for structures	56
3.2	Summary of brain structure segmentation results using the DSC metric	58
3.3	Automated atlas-based segmentation methods for healthy tissues	60
3.4	Summary of healthy tissue segmentation results with DSC metric	62
3.5	Automated atlas-based segmentation methods for challenging populations	63
3.6	Automated atlas-based segmentation methods for brains with lesions	65
3.7	Evaluation of reviewed methods for abnormal tissue segmentation	69
A.1	Pipeline summary	156
A.2	Data cost affinity matrix	169
A.3	Smoothness cost affinity matrix	169

Contents

1	Introduction	1
1.1	Multiple sclerosis (MS)	1
1.2	MS in magnetic resonance imaging (MRI)	3
1.2.1	Why use MRI?	3
1.2.2	What is MRI?	4
1.2.3	What are MR images of MS patients like?	6
1.2.4	Is MRI perfect?	7
1.2.5	What are we dealing with?	8
1.3	Scope of the research	9
1.4	Objectives	10
1.5	Document structure	11
2	A review on automated MS lesion segmentation	13
2.1	MS lesion segmentation	13
2.1.1	Features used for lesion segmentation	14
2.2	Classification of lesion segmentation approaches	15
2.2.1	Supervised strategies based on atlas	16
2.2.2	Supervised strategies based on learning from manual segmentation	20
2.2.3	Unsupervised strategies segmenting tissue	24
2.2.4	Unsupervised strategies segmenting only lesions	26
2.2.5	Summary of the strategies	27

2.3	Results	31
2.3.1	MS databases	31
2.3.2	Evaluation measures	33
2.3.3	Analysis of the results	35
2.4	Discussion	38
2.5	Conclusions	42
3	Brain atlases: concepts and application	45
3.1	Atlas based segmentation	45
3.2	Type of brain atlases	46
3.2.1	Topological atlases	47
3.2.2	Probabilistic atlases	48
3.3	Segmentation strategies	49
3.3.1	Label propagation	50
3.3.2	Multi-atlas label propagation	51
3.3.3	Probabilistic atlas-based segmentation	52
3.3.4	Summary	53
3.4	Segmentation methods and clinical targets	54
3.4.1	The brain and its internal structures	54
3.4.2	Brain tissues in healthy subjects	60
3.4.3	Brain tissues in challenging populations (fetus, neonates, and elderly subjects)	63
3.4.4	Brain tissues and lesions in pathological brains (MS, Alzheimer, tumors, etc.)	65
3.5	Discussion: How can we use an atlas to segment MS lesions?	71
3.6	Conclusions	72
4	MS lesion segmentation proposals	75
4.1	Overview	75

4.2	Preprocessing	76
4.2.1	Skull stripping	78
4.2.2	Bias correction and noise reduction	80
4.2.3	Inter-subject normalisation	81
4.2.4	Atlas registration	82
4.3	Tissue segmentation with lesion thresholding (TISSUE-LOT)	84
4.3.1	Tissue classification with expectation-maximisation (EM)	86
4.3.2	Lesion segmentation with FLAIR	90
4.4	Boosting with outliers and other spatial tools (BOOST)	93
4.4.1	Outlier map	95
4.4.2	Context meta-features	97
4.4.3	Boosting classifier	98
4.5	Summary	100
4.5.1	Preprocessing pipeline	101
4.5.2	TISSUE-LOT pipeline	102
4.5.3	BOOST pipeline	103
5	Experimental results	105
5.1	Evaluation	105
5.2	Databases	106
5.2.1	MS Lesion Segmentation Challenge 2008 database	106
5.2.2	SALEM database	107
5.3	Evaluation measures	112
5.3.1	Detection	113
5.3.2	Segmentation	114
5.4	TISSUE-LOT results	115
5.4.1	Parameters	115
5.4.2	Evaluation	118

5.5	BOOST results	124
5.5.1	Parameters	124
5.5.2	Evaluation	126
5.6	Strategy comparison	131
5.6.1	Quantitative comparison	132
5.6.2	Qualitative comparison	135
5.7	Discussion	139
6	Conclusions	141
6.1	Summary of the thesis	141
6.1.1	Contributions	142
6.2	Future work	144
6.2.1	Short-term proposal improvements	144
6.2.2	Future research lines	145
A	EPFL's internal report	147
A.1	Introduction	147
A.1.1	Goals	148
A.1.2	Structure	148
A.2	Study of MS segmentation approaches	149
A.2.1	Motivation	149
A.2.2	FLAIR simulation	150
A.2.3	Data sets	151
A.2.4	Methodology	152
A.2.5	Results	157
A.2.6	Conclusions	162
A.3	Markov random fields	164
A.3.1	MRF methods in MS segmentation	165
A.3.2	Motivation	166

A.3.3	Considerations	166
A.3.4	Graph cuts (GC)	170
A.3.5	Iterated conditional modes (ICM)	172
A.3.6	Discussion	173
A.4	Conclusions	178
A.4.1	Discussion	178
A.4.2	Impact on this PhD thesis	179
	Bibliography	181

INTRODUCTION

*Our first assignment is a documentary,
they're like real movies but with ugly people*
Abed, "Community (Season one)" (2009)

1.1 Multiple sclerosis (MS)

Multiple sclerosis (MS) is the most frequent non-traumatic neurological disease causing the most disability in young adults. It is relatively common in Europe, the United States, Canada, New Zealand, and parts of Australia, but rare in Asia, and the tropics and subtropics of all continents. In regions having a temperate climate, the incidence and prevalence of MS increase with latitude, both north and south of the equator. Multiple sclerosis is between two and three-times more common in women than in men; men have a tendency for a later disease onset with a poorer prognosis. The incidence of MS is low in childhood, increases rapidly after the age of 18, reaches a peak between 25 and 35, and then slowly declines, becoming rare at 50 and older. The world estimate is between 1.3 and 2.5 million cases of MS, with Western Europe having 350,000 [69]. Prevalence and incidence of MS is increasing around the world according to the latest epidemiological studies.

This disease is a chronic, persistent inflammatory-demyelinating and degenerative disease of the central nervous system (CNS), characterised pathologically by areas of inflammation, demyelination, axonal loss, and gliosis scattered throughout the CNS, often causing motor, sensorial, vision, coordination, deambulation, and cognitive impairment [50]. Relapses and progression are the two major clinical phenomena of prototypic MS. Relapses are considered the clinical expression of acute focal or multifocal inflammatory demyelination disseminated in the CNS. Remission of symptoms early in the disease is likely the

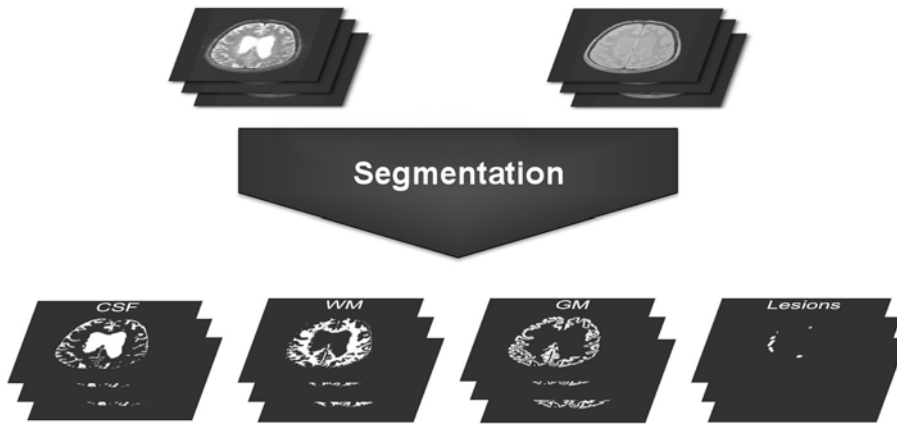


Figure 1.1: Tissue and lesion segmentation of two MRI images of the same patient.

result of remyelination, resolution of inflammation, and compensatory mechanisms such as redistribution of axolemmal sodium channels and cortical plasticity. These recovery mechanisms are less effective after recurrent attacks.

Conventional magnetic resonance imaging techniques (MRI) are highly sensitive for detecting MS plaques and can provide quantitative assessment of inflammatory activity and lesion load. MRI-derived metrics have become established as the most important paraclinical tool for diagnosing MS, for understanding the natural history of the disease and for monitoring the efficacy of experimental treatments [150]. This image modality also offers a high contrast between the main brain tissues, namely grey matter (GM), formed by neuron nuclei; white matter (WM), formed by neuronal axons; and the cerebrospinal fluid (CSF) protecting the brain.

For a quantitative analysis of focal lesions in both individual and temporal studies, manual or semi-automated segmentations of different MR images have been used to compute the total number of lesions and the total lesion volume (see figure 1.1). Manual delineation of MS lesions, however, is both challenging and time-consuming because of the large number of MRI slices for each patient that compose the three-dimensional information. Moreover, it is prone to intra-observer variability (the same study analysed by the same neuroradiologist at a different time) and inter-observer variability (the same study analysed by different neuroradiologists).

The development of fully automated MS segmentation methods, which can segment large amounts of MRI data and do not suffer from intra- or inter-observer variability, has become an active research field. Unfortunately, the results of these fully automated

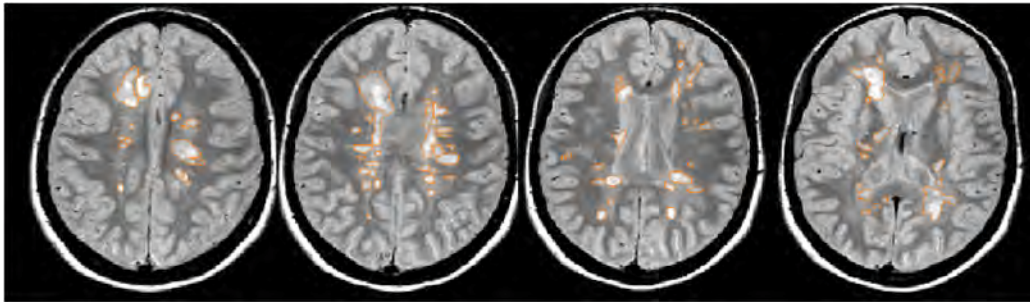


Figure 1.2: Lesion segmentation example. MS lesions, highlighted in orange, are small and their intensity usually overlaps the intensity distribution of other tissues.

methods show less agreement with manually segmented scans than those obtained with segmentations of independent observers due to the intensity overlap with other features. Moreover, when evaluating MS segmentation methods, there is a lack of an *in vivo* method to obtain a reliable ground truth data mainly because of the large intra- and inter-observer variability. Disagreements between segmentations may have a great influence on some evaluation measures due to the small volume of the lesions (see figure 1.2).

Automatic segmentation systems would, without doubt, help neuroradiologists to improve the diagnosis and follow-up in MS patients, both in clinical studies investigating future MS therapies and in every-day clinical practise. This may save time for the neuroradiologists and provide less subjective measures which will, in turn, provide better comparisons and analysis of the MS disease evolution.

1.2 MS in magnetic resonance imaging (MRI)

1.2.1 Why use MRI?

Magnetic resonance imaging is the most sensitive technique for the detection of demyelinating lesions in the central nervous system (CNS) in MS patients [77]. As a consequence of this high sensitivity, MRI has become an essential technique, not only in MS diagnosis, but also as a prognostic marker in the initial phase of the disease according to the frequency and gravity of future clinical recurrence, as well as future impairment rate [34, 120]. Moreover, MRI contributes to a better comprehension of its natural history and the evaluation of the effectiveness of new treatments [152, 80].

The new diagnostic criteria proposed by McDonald et al. [149, 171, 211, 170] highlight

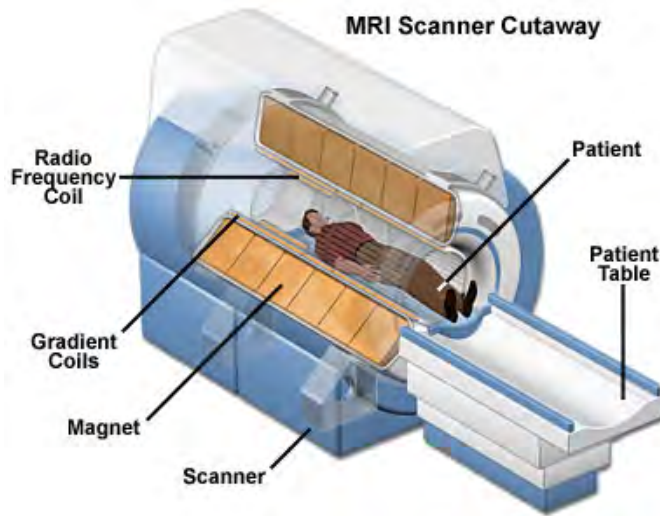


Figure 1.3: MRI scanner scheme with its main parts. Image extracted from <http://www.magnet.fsu.edu/education/tutorials/magnetacademy/mri/>.

MRI discoveries, by providing an MS diagnostic in patients with a single clinical episode when MRI proves the presence of spatially and temporally disseminated demyelinating lesions in the central nervous system [180, 181].

1.2.2 What is MRI?

An MRI system, as shown in figure 1.3, consists of the following components:

- A large magnet to generate the magnetic field and shim coils to make the magnetic field as homogeneous as possible. This magnetic field aligns the hydrogen nuclei of the brain.
- A radiofrequency (RF) coil to transmit a radio signal into the body part being imaged. This radio signal is applied after aligning the hydrogen nuclei with the high magnetic field.
- A receiver coil to detect the returning radio signals due to the nuclei relaxation.
- Gradient coils to provide spatial localisation of the signals applying additional magnetic fields. These additional magnetic fields can be used to only generate detectable signals from specific locations in the body (spatial excitation) and/or to make magnetisation at different spatial locations precess at different frequencies, which enables

k-space encoding of spatial information.

- A computer to reconstruct the radio signals into the final image (usually by means of Fourier transforms).

The voxel intensity of an MR image is determined by four basic parameters: proton density, T1 relaxation time, T2 relaxation time, and flow. Proton density is the concentration of protons in the tissue in the form of water and macromolecules (proteins, fat, etc). The T1 and T2 relaxation times define the way the protons revert back to their resting states after the initial RF pulse. The most common effect of flow is loss of signal from rapidly flowing arterial blood.

The contrast in an MR image can be manipulated by changing the pulse sequence parameters. A pulse sequence sets the specific number, strength, and timing of the RF and gradient pulses. The two most important parameters are the repetition time (TR) and the echo time (TE). The TR is the time between consecutive 90 degree RF pulses. The TE is the time between the initial 90 degree RF pulse and the echo.

The most common pulse sequences are the T1-w and T2-w spin-echo sequences. The T1-w sequence uses a short TR and a short TE ($TR < 1000\text{ ms}$, $TE < 30\text{ ms}$). The T2-w sequence uses a long TR and a long TE ($TR > 2000\text{ ms}$, $TE > 80\text{ ms}$). The T2-w sequence can be employed as a dual echo sequence. The first or shorter echo ($TE < 30\text{ ms}$) is proton density (PD) weighted or a mixture of T1 and T2. This image is very helpful for evaluating periventricular pathology such as multiple sclerosis, because the hyperintense plaques are contrasted against the lower signal CSF. More recently, the FLAIR (fluid attenuated inversion recovery) sequence has been introduced. FLAIR images are T2-w with the CSF signal suppressed, presenting a high contrast between tissue and lesions. As a consequence, these images have become a useful tool for every-day clinical practise.

However, the correlation between the burden of lesions observed in conventional MRI scans and the clinical manifestations of the disease remains weak. The discrepancy between clinical and conventional MRI findings in MS is explained, at least partially, by the limited ability of conventional MRI to characterise and quantify the heterogeneous features of the MS pathology. Other quantitative MR-based techniques, however, have the potential to overcome such a limitation of conventional MRI. Indeed, magnetisation transfer MRI, diffusion tensor MRI (DTI), proton MR spectroscopy, and functional MRI (fMRI) are nowadays contributing to elucidate the mechanisms that underlie injury, repair, and functional adaptation in patients with MS [79, 146].

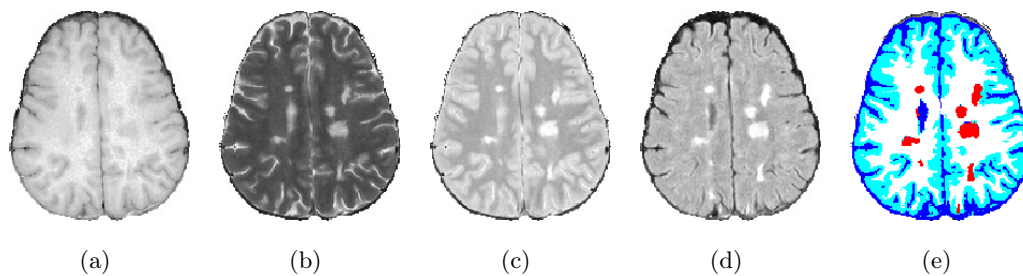


Figure 1.4: Different MRI images of the brain: a) T1-w image, b) T2-w image, c) PD-w image and d) FLAIR image and their e) tissue segmentation: CSF appears dark blue, GM appears blue, WM appears white and lesions appear red.

1.2.3 What are MR images of MS patients like?

Looking at figure 1.4, contrast differences between conventional MRI caused by different relaxation times are clear. For instance, the CSF appears dark in both T1 and FLAIR, while it's the brightest tissue in T2 and has similar intensities to GM in PD. On the other hand, WM is the brightest tissue in T1, has an intermediate grey level in FLAIR, similar to GM, and has the lowest signal in both PD and T2. Finally, GM also appears with an intermediate grey level in T2 and T1 in comparison with the other two brain tissues.

Both acute and chronic MS plaques appear as focal high signal intensity areas in T2-w images (including FLAIR), reflecting their increased tissue water content. The signal increase indicates edema, inflammation, demyelination, reactive gliosis and/or axonal loss in proportions that differ from lesion to lesion. They are typically discrete and focal in the early stages of the disease, but become confluent as the disease progresses.

Gadolinium-enhanced T1-w imaging (consisting of applying a contrast agent before acquiring the image) is highly sensitive in detecting inflammatory activity. This technique detects disease activity 5 to 10 times more frequently than clinical evaluation of relapses, suggesting that most of these **enhancing lesions (EL)** are clinically silent. Individual and temporal MRI studies have shown that the formation of new MS plaques is often associated with contrast enhancement, mainly in the acute and relapsing stages of the disease.

Approximately 10% to 20% of T2 **hyperintense lesions (HL)** are also visible as areas of low signal intensity compared with normal appearing white matter (WM) in T1-w images. These so-called T1 **black holes (BH)** have a different pathological substrate that depends, in part, on the lesion's age. Chronic black holes correlate pathologically

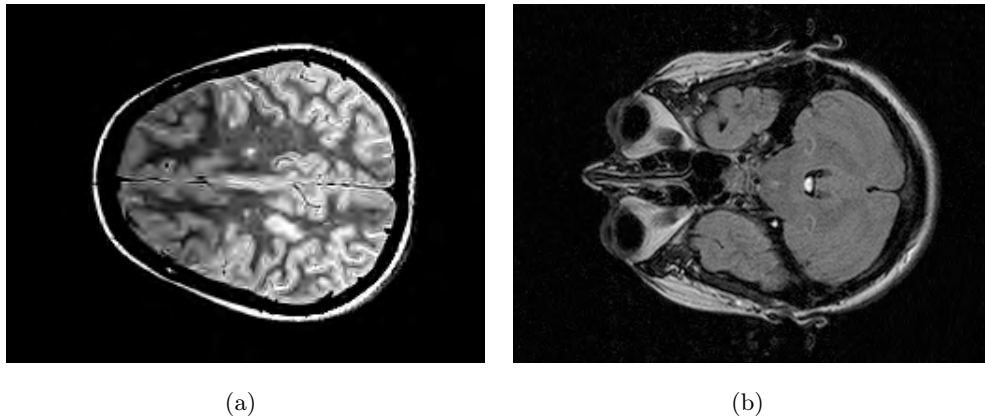


Figure 1.5: a) Slice of an MRI volume with clear intensity inhomogeneities (the leftmost part is darker than the right) and noise (mainly in the WM). b) Slice of an MRI volume where different non-brain structures appear (i.e. skull, eyes, fat).

with the most severe demyelination and axonal loss, indicating areas of irreversible tissue damage. T1-w images have a higher specificity than T2-w images for detecting lesions with irreversible tissue damage and may serve as surrogate markers of the disability progression in clinical trials.

1.2.4 Is MRI perfect?

Even though MRI has become an interesting imaging tool, several issues arise from the capture process causing undesired image artifacts. Some of these artifacts are usually dealt with during the scanning procedure. For instance, motion artifacts due to inadvertent head movement can be corrected using fast sequences and constraining the patient's head. Some others may require a calibration procedure or rescan. For example, peak artifacts due to a high signal value that darkens the whole image or aliasing caused by using a small field of view.

However, some artifacts are inherent to the capture system and can not be dealt with during acquisition. As with any capture system, signals are corrupted by noise often assumed additive and following a Rician distribution. In the case of MRI, this noise is caused by the gradient magnets (the ones responsible for voxel localisation). Inhomogeneities in the magnetic field as well as patient properties cause a smooth inhomogeneity field across the MRI (see figure 1.5(a)). This artifact is also known as bias field. Even though these two issues can not be avoided on acquisition, they can be reduced in order to increase the

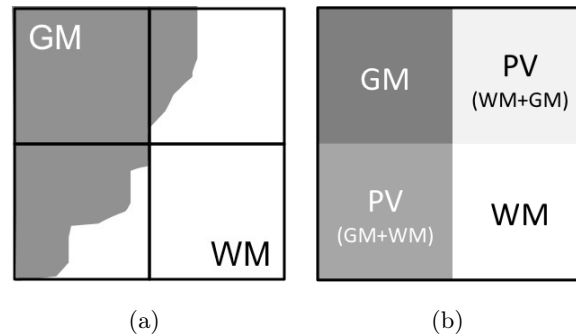


Figure 1.6: 2D synthetic example of the partial volume effect. a) Represents the real tissues and how they are sampled while b) shows the results of sampling zones with both tissues.

signal to noise ratio (SNR).

The gradient magnets are also partially responsible for the so-called partial volume effects (PVE). During localisation, the image space is discretised in volumes that may comprise more than one tissue. For instance, the cortex is folded in sulcus containing WM. These frontiers are usually scanned as a voxel, blurring the real edges between tissues (see figure 1.6).

Moreover, the presence of other non-brain tissues (see figure 1.5(b)) affects intensity distributions in the image. This is also inherent to the capture process but is not as much an artifact as it is an issue for automatic segmentation systems. It is not clear how the probability density function of each main tissue (GM, WM, and CSF) is altered by these external intensities, but segmentation results are usually improved when those voxels are masked out.

1.2.5 What are we dealing with?

At this point, we have seen why MRI has become a powerful technique in clinical practise for MS. Thanks to the presence of water molecules in the brain, and more precisely, hydrogen nuclei, MRI scanners can provide volumetric soft tissue information with a high contrast. Moreover, by tuning capture parameters, such as the pulse sequence or relaxation times, different volume sequences can be acquired. The most widely used images in MS trials are PD-w, T1-w, T2-w, and FLAIR.

Focusing on MS lesions, they usually appear as bright spots in PD-w and T2-w images (including FLAIR). Furthermore, all these lesions can be subdivided into 3 groups

depending on their pathology and properties in other images:

- Hyperintense lesions (HL): These are lesions that only appear in the abovementioned images and are the most frequent.
- Enhancing lesions (EL): These lesions appear as bright spots in T1-w images after applying a contrast agent, commonly Gadolinium, and represent current inflammatory activity.
- Black holes (BH): These lesions are named for their appearance, since they are characterised as “dark regions” in T1-w images. These lesions represent chronic and irreversible damage.

A reliable and robust automatic segmentation of these lesions could be used in clinical practise to diagnose the disease according to McDonald’s criteria. Following the current revision of these criteria from Polman et al. [170], a patient with just one attack (and thus only 1 imaging exploration) could be diagnosed with MS if the following criteria for dissemination in space (DIS) and time (DIT) are met:

- DIS: One lesion is found in at least 2 of the 4 MS-typical regions of the CNS (periventricular, juxtacortical, infratentorial, or spinal cord).
- DIT: Simultaneous presence of asymptomatic EL and HL lesions at any time.

As stated above, an automatic system to detect and segment MS lesions would prove helpful in the clinical practise to diagnose, as well as to evaluate a patient’s follow-up and the effect of drug therapy.

1.3 Scope of the research

The Computer Vision and Robotics group (VICOROB) of the University of Girona has been working on medical image analysis since 1996, mainly in segmentation and registration of mammographic images. Thanks to their previous knowledge acquired through several medical projects, the group recently started to focus their research on brain MRI analysis. This new line of research started with the segmentation of MS lesions and has expanded to other fields such as temporal analysis, registration (temporal and intersubject) or atrophy analysis.

All these studies have been carried out within the funded research projects PI09/91918 “SALEM: Segmentación Automática de Lesiones de Esclerosis Múltiple en imágenes de resonancia magnética” awarded by the Instituto Carlos III, and the VALTEC09-1-0025 “Salem: toolkit para la segmentación automática de lesiones de esclerosis múltiple en resonancia magnética” awarded in 2009 by the Generalitat de Catalunya within the “Projectes de valorització VALTEC”.

The goal of both projects is twofold: to create a novel dataset with imaging data from three different hospitals using three different scanning machines from different manufacturers and to study and develop techniques to segment MS lesions that can be transferred to experts for clinical use. Within these projects, in which this PhD was the starting point of the research, there has been a strong relationship with medical expert teams in the field of multiple sclerosis. Specifically:

- From the Hospital Vall d’Hebron: Dr. Rovira, who is the director of the “Unitat de Ressonància Magnètica-Centre Vall d’Hebron” (URMVH) and has participated in several research projects funded by public and private institutions in the last few years, as well as Dr. Pareto and technicians Huerga and Corral. This group is part of the MAGNIMS network, a European network of centers that share an interest in the MS study through MRI.
- From the Clínica Girona: Dr. Vilanova and Dr. Barceló are the codirectors of the “Unitat de ressonància magnética” at the Clínica Girona and are members of several national and international radiology societies.
- From the Hospital Josep Trueta: Dr. Ramió-Torrentà, who is the current coordinator of the “Unitat de Neuroimmunologia i Esclerosi Múltiple”, as well as Drs. Quiles and Valls, who work for the radiology unit. The relationship with this hospital arose from a previous collaboration with Dr. Gich within the project EM-Line that studied MS rehabilitation through interactive activities and games.

1.4 Objectives

As part of the SALEM framework, this PhD thesis’ main goal is

the proposal of a new pipeline capable of classifying brain tissues and detecting MS lesions in magnetic resonance imaging.

This objective refers to the segmentation of a set of images from a patient at a given stage and does not include the temporal evolution of that patient.

This general goal can actually be divided into several sub-goals focused on the different stages of this thesis. The first sub-goal is **to exhaustively analyse the state of the art of MS lesion segmentation techniques**. This objective aims to review the current MS lesion segmentation strategies to better understand their advantages and drawbacks. With this analysis, it was clear that images are corrupted by several artifacts and using prior knowledge can improve the results.

Following this notion, our second sub-goal is **the analysis of the state of the art in atlas-based segmentation applied to brain MRI**. Atlas-based segmentation is an important field of brain MRI analysis due to its strength to intrinsically encode spatial constraints (such as the location of certain structures) and the capability of adapting prior knowledge to a specific subject. Our main research **hypothesis** throughout this thesis is that the use of an atlas can overcome the limitations of the current detection and segmentation approaches.

The only thing left before the segmentation process is to reduce all the image artifacts. Thus, our third sub-goal is **to analyse different tools for MRI preprocessing**. These preprocessing steps can be divided into five main groups: noise reduction due to the capturing process, the correction of the bias field inherent to this image modality, inter-subject intensity normalisation by means of histogram matching, skull stripping to remove non-brain tissue that can bias segmentation results, and registration to adapt a healthy tissue atlas to the subject's space. Public software solutions will be tested as part of the whole preprocessing pipeline.

With the images corrected and the atlases prepared, the next step and fourth sub-goal is **to implement a new proposal exploiting the strengths of the best methods of the state-of-the-art and reducing their weaknesses**. This proposal will be atlas-based for the aforementioned reasons.

Finally, in order to test our results, the last sub-goal consists of **creating a novel database with imaging data from three different hospitals that can be used by other members of the group**.

1.5 Document structure

This thesis is structured as follows:

- **Chapter 1. Introduction.** This chapter presented the background, objectives and planning of this thesis project. In the following chapters, all details regarding MS segmentation are presented and extended in order to present the current techniques in this field and to introduce a new proposal.
- **Chapter 2. A review on automated MS lesion segmentation.** After stating the problem in chapter 1, we will review the most recent techniques dealing with this problem focusing on advantages and drawbacks. A classification of these techniques in supervised and unsupervised strategies will also be introduced, emphasising atlas-based segmentation methods. Finally, the reported results will be gathered, along with the most common evaluation measures and databases followed by the extracted conclusions.
- **Chapter 3. Brain atlases: concepts and application.** After the review in chapter 2, we will further investigate one of the most common and powerful strategies. Atlas-based segmentation has been a wide research topic in brain MRI due to its strength in using prior knowledge to segment either tissues or common structures. We provide an overview of the most important steps in these approaches: atlas creation, atlas registration and the final segmentation.
- **Chapter 4. MS lesion segmentation proposals.** Two new supervised pipelines (both based on an atlas) will be presented. First we introduce their mutual preprocessing steps and then we explain the two different approaches in detail. The first is based on the expectation-maximisation algorithm and tissue segmentation, while the second uses an ensemble classifier with state-of-the-art contextual features together with an outlier map.
- **Chapter 5. Experimental results.** The methods implemented will be tested and evaluated with real data using common similarity measures. In this chapter, we present our results, pointing out strengths and weaknesses as we also present and describe a novel database comprising imaging data from three different hospitals with different lesion loads.
- **Chapter 6. Conclusions.** In this final chapter, conclusions summarising the work developed are presented. Based on these conclusions, possible solutions are also introduced to be implemented as future works.

A REVIEW ON AUTOMATED MS LESION SEGMENTATION

What's that? A gun? I got a gun. He got a gun.

He got a gun... Everybody got guns!

Gyp Rosetti, "Boardwalk Empire (Season three)" (2012)

2.1 MS lesion segmentation

In image analysis, segmentation is defined as the process of delimiting a region with certain properties. For instance, such regions could be the sky in a landscape picture, a star in astronomical images or a certain brain tissue in MRI. Related to this concept, detection is defined as the process of finding an object in a certain image. For instance, we could detect a car in a street scene, a cluster of stars forming a galaxy in a satellite image or an MS lesion. Usually, when dealing with detection, a rough estimation of the region occupied by the object is enough. However, both segmentation and detection can be performed and evaluated together.

Another important concept of artificial intelligence is classification, which is defined as the process of assigning a certain category to a new observation using a set of observations whose category is known. This definition implies a two-step process where, first, a model is learned in the training set and, afterwards, the model is tested on the new observations to determine this category. When applied to image pixels (or voxels), the classification becomes a segmentation, since a label is assigned to each pixel delimiting a region.

In this chapter, the recent state-of-the-art of automated MS lesion segmentation is re-

viewed [141]. Firstly, the main image features used as input for the different segmentation algorithms are analysed. Afterwards, a classification of the different strategies is proposed and the most significant works in this field are described. Finally a discussion and conclusions are presented.

2.1.1 Features used for lesion segmentation

Based on the assumption that different brain tissues have different grey-level intensities, the most common feature used for lesion segmentation is the voxel intensity [15]. Furthermore, the appearance of the tissue and the lesions depends on the MR image used (see figure 1.3 from the previous chapter, which shows four examples of different MRI scans of a patient with MS lesions).

Analysing the literature, one may distinguish between single-channel or multi-channel approaches, i.e. approaches that use only one MR image or those that combine several images. Single-channel approaches are mainly used to segment the brain tissues. For instance, T1-w images are widely used for this purpose, since they show the best contrast between the three main brain tissues: WM, GM, and CSF. This initial tissue segmentation may then be used to help obtain the final lesion segmentation, and T2-w and PD-w are the classical images for detecting MS lesions. Another example of the single-channel approach is the segmentation of MS lesions using just the FLAIR sequence [121]. The multi-channel approaches, on the other hand, use at least two of the PD-w, T1-w, T2-w, and FLAIR images. One of the benefits of using more than one of the different MR images is that it increases the intensity feature space, producing a better discrimination between brain tissues. Furthermore, more than one kind of image may be required because MS lesions can appear independently in different images [244], depending on their subtype.

Additional features are used in some approaches to include spatial information in the algorithms. This is usually done using Markov Random Fields (MRF) to model neighbourhood interactions [228, 249]. If the parameters controlling the strength of the spatial interactions are properly selected, smoother structures are obtained. Alternatives to MRF that also include spatial information are the fuzzy connectedness (FC) segmentation methods [225] or the inclusion of a probabilistic atlas [125].

Moreover, most of the algorithms can be roughly divided into either global or local depending on the information they use (i.e. the feature extraction process). Global methods extract features from the whole image and then use this information to classify each voxel independently. In contrast, local methods use only local information to create, in many

cases, an undetermined number of local regions. The challenge that these local methods must overcome is to combine these local regions to build a global and meaningful segmentation.

2.2 Classification of lesion segmentation approaches

In this section, the MS lesion segmentation methods are described according to the classification shown in tables 2.1, 2.2, 2.3 and 2.4. Notice that the tables offer an overview of all these works with respect to the strategy, the type of MR images used, the type of lesions detected and the algorithms. The criteria used to select these methods are based on several aspects: 1) representative works for each of the identified strategies; 2) the reported experimental results and the evaluation measures used by the authors; and 3) the data sets used to perform the experiments (synthetic, real cases, and data from the MS Lesion Segmentation Challenge 2008 [208], which enables a quantitative comparison of the evaluation results).

The papers are divided between supervised and unsupervised segmentation strategies. Supervised approaches are those based on using some kind of a priori information or knowledge to perform the MS lesion segmentation. The group of supervised strategies is further subdivided into two sub-groups of approaches. In the first group all the approaches use atlas information and therefore require the application of a registration process to the analysed image to perform the segmentation. In the second group all the approaches perform an initial training step on features extracted from manually segmented images annotated by neuroradiologists. The methods in this second group employ the image intensities previously segmented by an expert to train a classifier that segments the tissues and lesions of the MR images.

With regard to the unsupervised strategies, where no prior knowledge is used, two different sub-groups can also be identified. A sub-group of methods that segment brain tissue to help lesion segmentation and another sub-group that use only the lesion properties for segmentation. In the first sub-group, there are methods that either segment the tissue first and then the MS lesions, or segment the tissue and the lesions at the same time. In the second sub-group, the methods directly segment the lesions according to their properties, without providing tissue segmentation. The advantage of segmenting the tissue is that neuroradiologists can also evaluate the GM tissue volumetry and monitor the progression of cerebral atrophy.

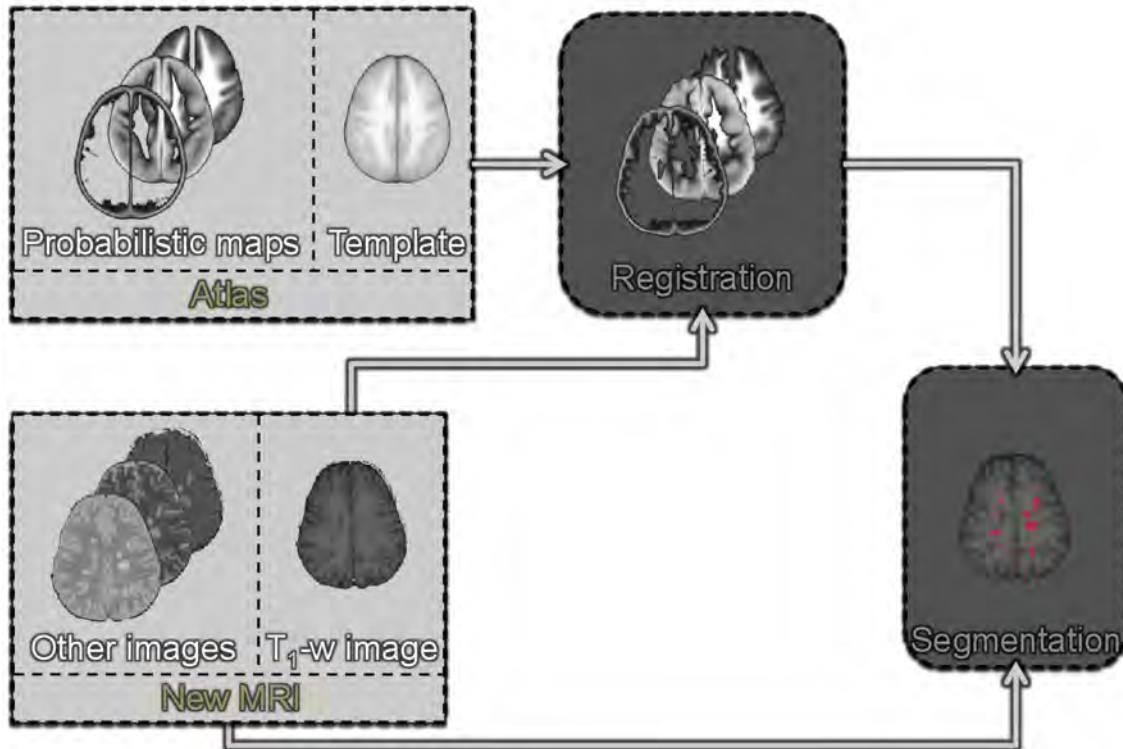


Figure 2.1: Flowchart of the supervised strategies based on atlas.

2.2.1 Supervised strategies based on atlas

Looking at the strategies based on atlas information, it is possible to distinguish between the use of both statistical and topological atlases. A statistical atlas provides the prior probability of each voxel to belong to a particular tissue class. This statistical atlas is built from a set of manual segmentations of the structures of interest, where the boundaries of each structure are used to make a smooth probability map and to account for anatomical variations beyond those present within the training set. Notice that the use of an atlas can be helpful in classifying tissues in the presence of noise or inhomogeneities [56] (an atlas takes spatial information into account), or in order to segment lesions as deviations from normal human brains. On the other hand, a topological atlas is a parcellation of the brain that is edited to encode a specific topology for each structure and group of structures. This topological atlas is usually used to preserve topology and to lower the influence of competing intensity clusters in regions that are spatially disconnected, while the statistical atlas affects the segmentation of adjacent structures that have similar intensity. As shown in the flowchart of figure 2.1, notice that in atlas-based segmentation methods, the analysed

Article	Algorithms	Images	Lesions
Kamber (1995) [119]	Different classifiers	PD, T1, T2	HL
Van Leemput (2001) [135]	EM + MRF	PD, T1, T2	HL
Zijdenbos (2002) [251]	ANN	PD, T1, T2	HL
Wu (2006) [244]	kNN	PD, T1c, T2	EL & BH & HL
Bricq (2008) [35]	FAST-TLE + HMC	T2, FLAIR	HL
Kroon (2008) [125]	PCA	T1, T2, FLAIR, DTI (FA, MD)	HL
Prastawa (2008) [174]	Region partitioning	T1, T2	HL
Shiee (2008) [198]	FCM	PD, T1, T2, FLAIR	HL
Shiee (2008) [197]	FCM	T1, T2, FLAIR	HL
Souplet (2008) [204]	EM + GMM	T1, T2, FLAIR	HL
Tomas (2009) [220]	Bayes	T1, T2, FLAIR	HL
Akselrod-Ballin (2009) [2]	FLD + Decision Forest	PD, T1, T2, FLAIR	HL
Shiee (2010) [199]	FCM	T1, T2, FLAIR	HL
Geremia (2011) [92]	Random forests	T1, T2, FLAIR	HL
Schmidt (2012) [187]	EM + Thresholding	T1, FLAIR	HL

Table 2.1: Summary of the supervised approaches based on atlas for MS lesion segmentation with respect to the sequences, algorithms and lesions. The acronyms for the algorithms stand for (in alphabetical order): Artificial neural networks (ANN), expectation-maximisation (EM), fast trimmed likelihood estimator (FAST-TLE), fuzzy c-means (FCM), fisher linear discriminant (FLD), hidden Markov chains (HMC), k-nearest neighbours (kNN), Markov random fields (MRF) and principal components analysis (PCA). The acronyms for the lesions and sequences stand for: Diffusion tensor imaging (DTI), fractional anisotropy (FA), mean diffusivity (MD), black holes (BH), enhanced lesions (EL), and hyperintense T2 lesions (HL).

MR image has to be registered with the atlas before the segmentation is done. Hence, the challenge for these atlas-based approaches is to align the atlas and the images, thereby converting the segmentation problem into a registration problem.

The first example of table 2.1 is the proposal by Kamber et al. [119], where the inputs for the training step are not voxel intensities but rather their probabilities of belonging to WM, GM, or CSF tissue categories. Using this prior information as a pattern, their approach trains and tests a set of different classifiers to segment MS lesions.

Van Leemput et al. [135] provide another example of an atlas-based approach for MS lesion segmentation¹. They proposed an intensity-based tissue classification using a stochas-

¹ This software is publicly available at <http://www.medicalimagecomputing.com/downloads/ems.php> as an add-on to the well-known SPM package (<http://www.fil.ion.ucl.ac.uk/spm/>).

tic model extracted from an expectation-maximisation (EM) algorithm [64], while simultaneously detecting MS lesions as outliers that were not well explained by the model. In this method, a prior classification is derived from a digital brain atlas that contains information about the expected location of WM, GM, and CSF. Their approach also corrects for MRI field inhomogeneities, estimates tissue-specific intensity models from the data itself, and incorporates contextual information into the classification using an MRF.

Following the outlier concept, Schmidt et al. [187] presented an automated tool based on outlier models. First, tissues were segmented using SPM on the T1-w image. Afterwards, tissue distributions were estimated on the FLAIR image using the segmentation results. These distributions were then used to estimate outlier maps for each tissue that were added. With this final outlier map, a threshold was estimated to obtain a mask that was refined using a region growing approach.

The k-nearest neighbour (kNN) [71] is probably the commonest of supervised classification methods. With this algorithm a test sample is classified by the majority class of its closest neighbours. kNN makes strong assumptions about the data, i.e. that there is no correlation among different multivariate channels and that all variances are the same. Based on this classifier, Wu et al. [244] proposed the automatic segmentation of MS lesions into three subtypes: enhancing lesions, black holes and hyperintense lesions. An intensity-based statistical kNN classifier is combined here with atlas segmentation to extract WM masks. Based on the assumption that lesions are only found within WM regions, the authors discard all the lesions outside the masks. Moreover, partial volume problems are corrected using morphological operators. Another supervised approach relying on atlas information is provided by Zijdenbos et al. [251]. This method² uses the probability - extracted from the atlas - that tissue is WM, GM, or CSF and uses the intensities from T1-w, T2-w and PD-w as inputs to an ANN classifier, which performs the lesion segmentation.

There are more examples of atlas-based approaches. The method proposed by Shiee et al. [197, 198, 199] segments brain tissues in an iterative way, interleaving a fuzzy segmentation and defining topologically consistent regions. MS lesions are identified as dark holes inside the WM. The authors use multi-channel images to segment the major structures of the whole brain. Basically, their method is an atlas-based segmentation technique employing a topological atlas and a statistical atlas, together with the well-known fuzzy c-means (FCM) algorithm [29] which performs the classification. As reported by Shiee et al., the advantage of using the topological atlas is that all segmented structures are spatially con-

² This software is publicly available at <http://www.bic.mni.mcgill.ca/>, as part of the INSECT application.

strained, thereby allowing subsequent processing to perform cortical reconstruction and cortical unfolding. The automated atlas-based segmentation method presented by Bricq et al. [35] performs tissue classification using an algorithm based on the trimmed likelihood estimation of a mixture model [102], while the lesions are outliers to the model. Neighbourhood information is encoded by the hidden Markov chain (HMC) model and also incorporates the use of a statistical atlas. Prastawa and Gerig [174] proposed a fully automated lesion segmentation method that combines outlier detection and region partitioning, and is based on using an atlas of healthy subjects to detect lesions as outliers. The algorithm is iterative and alternates between estimating the intensity probability density functions (PDF), computing voxelwise spatial probabilities, correcting for intensity inhomogeneities and partitioning the images into spatially coherent regions. Notice that the intensity PDF for the healthy brain tissue is computed using samples obtained from the atlas. The minimum covariance determinant robust estimation scheme is then applied to the intensity samples of healthy tissues to determine inliers and outliers. The inlier samples are used to form the PDF of the brain tissue intensities, while the outliers are assigned to the lesion class, similar to the approach of Van Leemput et al. [135]. Moreover, a watershed algorithm that takes the neighbourhood information into account is used to initialise the segmentation. Therefore, segmented regions are used rather than individual voxels. The authors argued that the use of regions helps to significantly reduce false positives inherently linked to conventional voxel-based classification.

Kroon et al. [125] introduced a method based on a local feature vector for automated lesion segmentation of multi-channel MRI data. Their local feature vector contains neighbourhood voxel intensities, histogram, and probabilistic atlas information. The histogram information is added to provide the model with low pass intensity information of a certain region, while the atlas probability allows it to exclude false MS voxel detection in areas where MS is less probable. Principal component analysis (PCA), with a log-likelihood ratio, is then used to classify each voxel. Using a different algorithm, Souplet et al. [204] proposed a method designed to detect hyperintense signal areas on a T2-FLAIR sequence³. They first apply an EM algorithm to perform the segmentation in T1-w and T2-w images. This segmentation, using an atlas, allows them to initialise, for each voxel, the probability of belonging to the different tissue intensities modelled as Gaussian mixture models. From the resulting tissue segmentations, the authors propose to automatically apply a threshold for the T2-FLAIR sequence to detect the most plausible lesions in the hyperintense signals. In this final step an enhanced FLAIR image is generated to allow a better lesion

³This software is publicly available at <http://www-sop.inria.fr/asclepios/software/SepINRIA>.

segmentation.

Following the idea of Prastawa and Gerig [174], Tomas et al. [220] used the samples extracted from an atlas to train a classifier. This procedure to extract healthy tissue training samples is based on the work of Scully et al. [188], who extended the intensity feature space to obtain better discrimination between tissue clusters. Subsequently, their method uses an atlas to create a distance map from which training samples were selected. With this training data, a Bayes classifier performed the MS lesion segmentation. This approach also introduced an atlas-based post-processing step to remove false positive lesions. A different classification approach, was presented by Geremia et al. [92]. This approach introduced the use of meta-features based on context and symmetry to further aid the segmentation. To compute these features, a fast technique based on image integrals is used. After their computation, these features combined with atlas probabilities and image intensities were used to create a huge feature pool to train a random forest algorithm. Finally, Akselrod-Ballin et al. [33] proposed a multi-scale approach that combines segmentation with classification to automatically segment MS lesions in multi-channel MR images. Their method uses segmentation to obtain a hierarchical decomposition of the MR images, which produces a rich set of features describing the regions in terms of intensity, shape, location, neighbourhood relations, and anatomical context. The atlas information is applied in this step to obtain statistical features. Afterwards, segmentation is performed using a decision forest along with Fisher linear discriminant analysis to deal with multiple features.

In conclusion, atlas-based approaches can be used to segment both the tissue and the lesions. Moreover, atlases make it possible to treat the lesions as outliers in the tissue, to introduce spatial information into the segmentation process and to reduce the false positive lesion segmentations. As a drawback, these approaches rely on building an atlas, which is not a simple task. In addition, they also introduce the registration problem into the MS lesion segmentation. Note that this registration step is even more difficult when dealing with cases with severe atrophy, large numbers of lesions, etc.

2.2.2 Supervised strategies based on learning from manual segmentation

The second group of supervised approaches uses manually-segmented images annotated by neuroradiologists to segment the MS lesions. Figure 2.2 shows the flowchart followed by these algorithms. Note that unlike atlas-based approaches these approaches do not need any registration process between the analysed images and the atlas. In contrast, these methods use mainly the image intensities of different MR images to train a classifier for

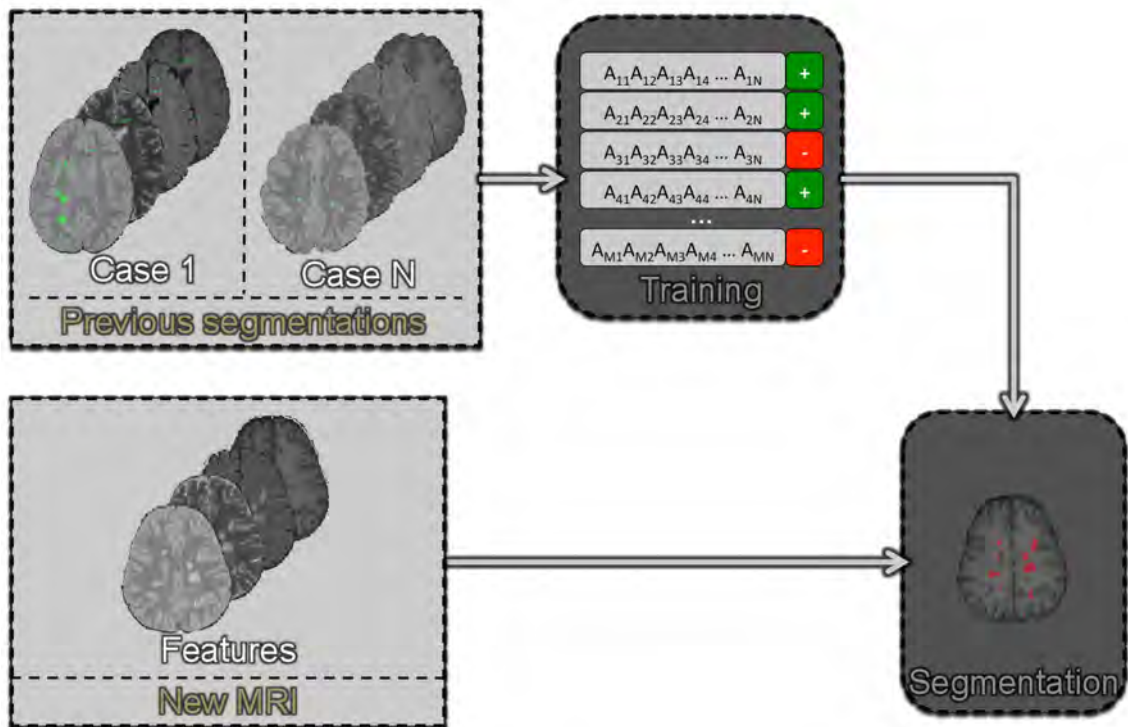


Figure 2.2: Flowchart of the supervised strategies based on learning from manual segmentation.

the segmentation purpose. As reported by several authors, the use of prior knowledge to guide the segmentation of MS lesions improves the robustness of the algorithms, thus reducing the volume of false positive lesions compared to purely data-driven segmentations. Note that some of the approaches classified in this category may be similar to strategies described in the previous section. However, the majority of the strategies included here rely on a training process performed using features extracted from manually-segmented MR images. Furthermore, some of these methods include the use of registration algorithms that focus on the intra-sequence and inter-sequence pre-processing registration steps. Table 2.2 shows that a large number of proposals have followed this strategy, most of them being multi-channel approaches. Furthermore, different classifiers or a combination of them, for example, ANN, kNN, AdaBoost, Bayesian classifiers, and decision trees, have been used to perform the segmentation.

Goldberg et al. [94] use local thresholding to select the brightest regions of the image. Afterwards, the lesions are segmented by looking for closed contours and using different morphological properties such as area, perimeter and shape. For segmentation, an ANN is

Article	Algorithms	Images	Lesions
Goldberg (1998) [94]	ANN	PD, T2, FLAIR	EL & HL
Alfano (2000) [3]	Spatial clustering	PD, T1, T2	HL
Anbeek (2004) [6]	kNN	PD, T1, T2, FLAIR, IR	HL
Anbeek (2005) [5]	kNN	PD, T1, T2, FLAIR, IR	HL
Sajja (2006) [186]	Parzen windows	PD, T2, FLAIR	HL
Datta (2006) [55]	Parzen windows + MGR	PD, T1, T2, FLAIR	BH
Anbeek (2008) [7]	kNN	FLAIR	HL
Scully (2008) [188]	Bayes	T1, T2, FLAIR	HL
Morra (2008) [156]	AdaBoost	T1, T2, FLAIR, DTI (FA, MD)	HL
Subbanna (2009) [209]	Simulated annealing + MRF	PD, T1, T2	BH & HL
Lecoeur (2009) [133]	Graph cuts	PD, T1, T2	HL

Table 2.2: Summary of the supervised approaches based on learning for MS lesion segmentation with respect to the sequences, algorithms, and lesions. The acronyms for the algorithms stand for (in alphabetical order): Artificial neural networks (ANN), k-nearest neighbours (kNN), morphological grayscale reconstruction (MGR) and Markov random fields (MRF). The acronyms for the lesions and sequences stand for: Diffusion tensor imaging (DTI), fractional anisotropy (FA), mean diffusivity (MD), black holes (BH), enhanced lesions (EL), and hyperintense T2 lesions (HL).

trained and used to classify the regions. Alfano et al. [3] also used previous segmentations of normal tissues to extract features used to train a spatial clustering algorithm. The authors stated in their experiments that their approach was also suitable for monitoring changes in the disease over time.

Anbeek et al. applied the kNN classifier in two studies. The aim of the first was exclusively to detect MS lesions [6] and spatial features were included in the classifier to achieve better lesion segmentations; in the second study, the aim was to model all brain tissues [5] and in this case, the kNN classifier was used to classify tissues and lesions simultaneously. It is important to mention that the authors tested their multi-channel approach using information from T1-w, inversion recovery (IR), PD-w, T2-w, and FLAIR images, concluding that the incorporation of the T1-w, PD-w or T2-w did not significantly improve the segmentation results of the different brain tissue types.

From a different viewpoint, Sajja et al. [186] proposed segmenting CSF and lesions using a Parzen windows classifier and then segmenting WM and GM using a parametric method. The assumption behind this approach is that GM and WM, but not lesions and CSF, follow a Gaussian distribution. Therefore, the authors first classify CSF and hyperintense T2-w

lesions using a Parzen classifier and then the remaining brain parenchyma - excluding CSF and lesions - is classified into GM and WM using the PD-w and T2-w images and an MRF together with an EM algorithm. This method also exploits contextual information by using a fuzzy-connectedness to minimise the false negative lesion classifications. In another work, the authors also proposed segmenting black holes (considered as regional minima) in MS using a similar strategy [55]. After applying the previous segmentation method, black hole segmentation is achieved by using morphological grey-scale reconstruction in T1-w images.

The work of Scully et al. [188] introduced a new parametric method to the field of MS lesion segmentation. This method uses a vector image joint histogram, built over a training set, as an explicit model of the feature vectors indicating lesion. This model is then used to generate samples to train a naive Bayesian classifier which proceeds to classify the vector image composed of the T1-w, T2-w, and FLAIR images. Using a different strategy, Morra et al. [156] proposed a framework to automatically segment sub-cortical structures in brain MR images. Their method uses an AdaBoost algorithm to learn a unified appearance and context model which is then used to perform the lesion segmentation. Their feature pool includes intensity, position, and neighbourhood features.

Subbanna et al. [209] presented a fully automated framework for identifying MS lesions in multi-channel MR images. Manual segmented images are used to extract intensity histograms of both tissue and lesions. From the histograms, multivariate Gaussian distributions are estimated and used in the MRF classification step, which incorporate local spatial variations and neighbourhood information. Finally, Lecoeur et al. [133] also presented an optimised supervised lesion segmentation method using multi-channel MR images. Their proposal creates an optimised spectral gradient colour space from single-channel images. Based on this transformation, they then apply a graph cuts segmentation. As argued by the authors, the graph cuts algorithm provides an optimal solution for the joint use of regional and border information in a way similar to how MRFs work.

To summarise, the use of manually annotated data allows expert knowledge to be incorporated into MS segmentation approaches. Moreover, as shown in table 2.2, having initial segmentations allows a large variety of classifiers (ANN, kNN, EM, Bayes, etc) to be applied. As in the atlas-based approaches, selecting a good initial MRI training set is an important step. Another issue with many lesion segmentation algorithms, particularly those that employ training data to model lesion intensity profiles, is that they are dependent on a specific acquisition sequence. These approaches must be modified or re-trained to process data acquired using alternative pulse sequences.

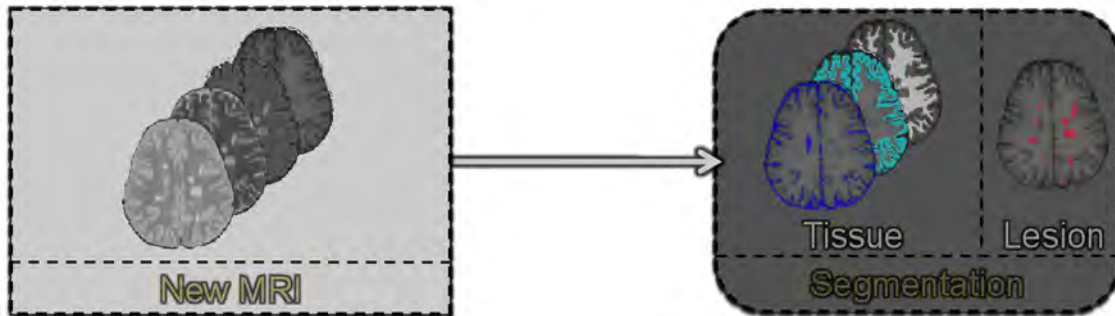


Figure 2.3: Flowchart of the unsupervised strategies based on tissue properties.

Article	Algorithms	Images	Lesions
Freifeld (2007) [83]	EM + CGMM	PD, T1, T2	HL
Khayati (2008) [121]	Bayes + AMM + MRF	FLAIR	HL
García-Lorenzo (2008) [88]	EM + MeS	T1, T2, PD	HL
García-Lorenzo (2008) [89, 87]	EM	PD, T1, T2	HL
García-Lorenzo (2008) [90]	EM	T1, T2, FLAIR	HL

Table 2.3: Summary of the unsupervised approaches based on tissue for MS lesion segmentation with respect to the sequences, algorithms and lesions. The acronyms for the algorithms stand for (in alphabetical order): Adaptive mixtures method (AMM), constrained Gaussian mixture models (CGMM), expectation-maximisation (EM), mean shift (MeS) and Markov random fields (MRF). The acronyms for the lesions and sequences stand for: Hyperintense T2 lesions (HL).

2.2.3 Unsupervised strategies segmenting tissue

There are several works in which the main principle consists of applying an unsupervised algorithm to segment the tissue and the lesions, as it is schematically shown in the flowchart of figure 2.3. These approaches usually detect lesions as outliers on each tissue rather than adding a new class to the classification problem. For instance, Freifeld et al. [83] first initialise their algorithm based on a pre-segmentation using k-means and its subsequent decomposition into a mixture of many spatially-oriented Gaussians per tissue (constrained GMM) in order to capture the spatial layout [97]. The intensity is considered as a global parameter and is constrained to be the same value for a set of related Gaussians per tissue. In order to detect the lesions, a set of rules that distinguishes between normal tissue regions and lesions is defined. Following initialisation, voxel-wise GMM parameters are learned via an EM algorithm. Finally, an active contour algorithm is used to delineate lesion

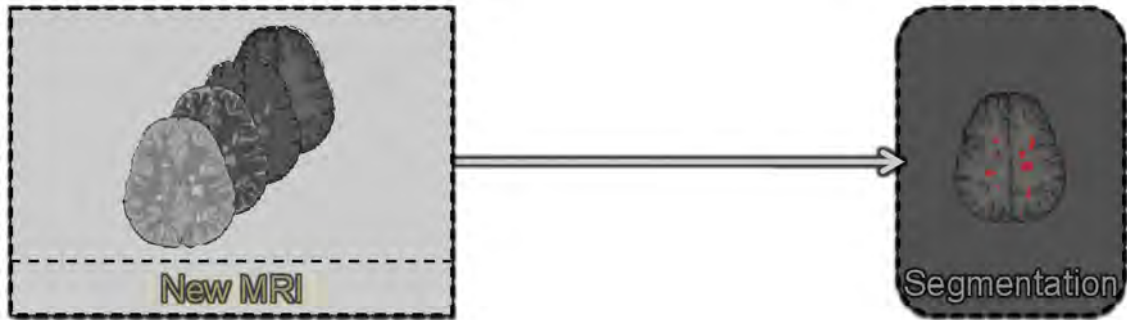


Figure 2.4: Flowchart of the unsupervised strategies based on lesion properties.

boundaries.

García-Lorenzo et al. [88] combined a modified version of the EM-based method (mEM) to the trimmed likelihood estimator with the mean shift algorithm [48] to segment MS lesions. A local segmentation approach (mean shift) is used to generate local regions in the images, while an EM variant is employed to classify the regions obtained into healthy tissue or lesions. In another work by García-Lorenzo et al. [90, 89, 87], the authors presented a modified version of the spatio temporal robust expectation-maximisation (STREM) [1] to perform the MS lesion segmentation. Their approach is based on three main processes: robust estimation of healthy tissues using the mEM introduced above, refinement of outlier detection, and application of lesion rules. The authors included in another work the possibility to semiautomatically improve the segmentation by using an interactive graph cuts approach [86]. Khayati et al. [121] combined an adaptive mixtures method (AMM), MRF and a Bayesian classifier to simultaneously classify the three main brain tissues and the MS lesions using only FLAIR images. In particular, they first propose to segment the brain into four classes: WM, GM, CSF and “others”. Afterwards, inside the “others” class, lesions are dealt with as outliers not correctly explained by the model.

In conclusion, this set of algorithms are based on classifying the outliers coming from a previous tissue segmentation in order to provide the lesion segmentation. Notice that the results of the lesion segmentation depend highly on the quality of the tissue segmentation. Furthermore, not all the tissue segmentation methods take abnormal cases (with severe lesions, atrophy, etc) into account.

Article	Algorithms	Images	Lesions
Bedell (1998) [26]	Threshold Subtraction	T1, T1c, AFFIRMATIVE	EL
Boudraa (2000) [33]	FCM	PD, T2	HL
He (2002) [106]	MGR	T1, T1c, AFFIRMATIVE	EL
Datta (2007) [54]	MGR	T1, T1c, T2, FLAIR	EL
Saha (2009) [185]	FCM	T2	HL

Table 2.4: Summary of the unsupervised approaches based on lesion properties for MS lesion segmentation with respect to the sequences, algorithms and lesions. The acronyms for the algorithms stand for (in alphabetical order): Fuzzy c-means (FCM) and morphological grayscale reconstruction (MGR). The acronyms for the lesions and sequences stand for: Attenuation of fluid by fast inversion recovery with magnetisation transfer imaging with variable echoes (AFFIRMATIVE), enhanced lesions (EL), and hyperintense T2 lesions (HL).

2.2.4 Unsupervised strategies segmenting only lesions

The last group of unsupervised approaches is based on using only the lesion properties to perform the MS lesion segmentation, avoiding hence the tissue segmentation step (figure 2.4 summarises the flowchart of this strategy). For instance, Bedell and Narayana [26] present an automated segmentation and quantification of contrast-enhanced lesions based on performing threshold subtraction to eliminate enhancing structures such as choroid plexus. The authors reported that all MS lesions larger than 5 mm^3 were successfully identified and the automated analysis produced no false positive or false negative lesions above this volume in 13 different patients. The method used by Boudraa et al. [33] performs a FCM algorithm two times to detect lesions. The first FCM algorithm has the goal of obtaining two clusters: one that groups together CSF and lesions, and another that groups together WM and GM. Afterwards, a second FCM algorithm is applied to distinguish between lesions and CSF. A final post-processing step based on anatomical knowledge is performed to remove extra segmented structures. This strategy assumes that, in the first step, all the lesions have been grouped together within a specific cluster and then, in the second step, this cluster is resegmented to distinguish between healthy tissue and lesions, taking spatial information into account. Saha et al. [185] automatically determine the number of clusters by introducing genetics into the algorithm. Membership values of points to different clusters are computed based on a point symmetry based distance rather than using the Euclidean distance. The chromosomes encode the centers of a number of clusters, whose value may vary.

He and Narayana [106] also proposed a method for the automated identification and segmentation of contrast-enhanced MS lesions in brain MR images. This method relies on an adaptive local segmentation derived from morphological grey-scale reconstruction operations to identify both lesion and non-lesion enhancements. Similarly, Datta et al. [54] developed a method for the identification and quantification of gadolinium (Gd) enhancements. This is also a multi-channel MRI approach and aims to identify enhancements using morphological operations. These enhancing lesions are further segmented based on fuzzy connectedness. Their experimental results show that accurate segmented Gd enhancements are obtained.

In regard to this strategy, one should note that unlike the methods presented in the previous section, these approaches do not rely on an initial WM, GM and CSF tissue segmentation. Moreover, they are useful for segmenting special lesions such as black holes and enhancing lesions. As a drawback, however, they have to specifically define the properties used in each image, which is not an easy task. It should be noted that some artefacts may share the same lesion properties.

2.2.5 Summary of the strategies

Four strategies used to deal with the automated MS lesion segmentation have been presented. In what follows a brief summary of these strategies, summarising the main advantages and drawbacks is provided (see the scheme in figure 2.5). Moreover, a more detailed description of a representative algorithm of each strategy is presented.

The approaches have been classified between supervised and unsupervised algorithms. The inherent advantage of supervised algorithms is that they can automatically learn the characteristics of both normal tissue and lesions. However, their main problem is that they rely on having a good training set, which may be difficult to obtain. Two supervised strategies have been identified according to the procedure the annotations are introduced into the algorithms: with or without using a registration step. The advantage of atlas-based approaches is that spatial information is inherently used, although registration is also a challenging task. On the other hand, training-based approaches allow to use real characteristics of the tissues and the lesions, but spatial information has to be imposed in a further step since it is not included in the training process. We have also seen that there is a group of techniques which are unsupervised and therefore do not depend on a training step, being more generalised algorithms. This group of unsupervised techniques has been subdivided into two different strategies according to the use of the tissue information. The

	Description	Strengths	Weaknesses
Supervised based on atlas	The algorithms use atlas information	<ul style="list-style-type: none"> •Includes local information •Probabilistic framework 	<ul style="list-style-type: none"> •Needs registration •Needs segmentations to create an atlas
Supervised based on training	The algorithms classify using manual training	<ul style="list-style-type: none"> •No registration needed •A wide list of classifiers can be used 	<ul style="list-style-type: none"> •Depends on previous segmentations and the acquisition sequences
Unsupervised based on tissue	The lesions are found as tissues outliers	<ul style="list-style-type: none"> •Tissue guides the lesion segmentation 	<ul style="list-style-type: none"> •Depends on the quality of the tissue segmentation
Unsupervised based on lesion	The lesions are found using empirical rules	<ul style="list-style-type: none"> •Special sublesions can be segmented 	<ul style="list-style-type: none"> •Specific rules for image and lesion type are required •Artifacts may share lesion properties

Figure 2.5: Advantages and drawbacks of the different reviewed MS lesion segmentation strategies.

advantage of using tissue information is that may help in localising the lesions. However, the correct segmentation of the tissue is critical in these approaches. On the other hand, defining rules according to the lesion features allows to identify special lesions, although the rules may change with the used modality and scanning machine.

An example of supervised MS lesion segmentation by using atlas registration is the method of Souplet et al. [204], which used the EM algorithm to maximise the log-likelihood between the MRI data (T1-w, T2-w, and FLAIR images) and a Gaussian model of 10 classes: WM, GM, CSF, 6 GM/CSF partial volume classes, and one outlier class representing mostly the vessels. In this approach, the probability of each voxel to belong to the different classes is first initialised thanks to the a priori registration with the atlas.

After the initialisation process, the two steps of the EM algorithm are iterated. In the maximisation step, the parameters of each class (mean and covariance) are computed from the voxel intensities and their probabilities of belonging to the different classes. The parameters of the partial volume classes are computed as a proportion of the pure tissue parameters, while the parameters of the outlier class are obtained as a fraction of the CSF parameters. Therefore, in the expectation step, the probability of each voxel to belong to the different classes is updated using the Gaussian function with the class parameters and the atlas values as prior probabilities. Finally, when the algorithm converges, a bi-dimensional Gaussian distribution is estimated for each class including GM, WM, CSF, and partial volumes as well as all the voxel probabilities related to those distributions.

After the EM segmentation, the obtained GM mask is used to compute the mean and standard deviation values of this region in the FLAIR image which in turn are used to determine a threshold that allows the lesion segmentation. The FLAIR image contrast is first enhanced using morphological operations and then thresholded for obtaining an initial lesion mask. In order to reduce false positives, this mask is further refined using the WM mask with the cavities filled (representing a “healthy WM” mask) and the CSF and WM masks coming from the EM. Therefore, the lesion voxels are defined as those that are present in the “healthy WM” but not present in the real WM nor CSF masks.

The approach of Subbanna et al. [209] is a clear example of a supervised framework for identifying MS lesions in multi-channel MRI using manually segmented images. These manual segmentations are used to extract intensity histograms of both tissue and lesions using the intensities from PD-w, T1-w and T2-w images. A multivariate Gaussian distribution is then fitted to those histograms. In the case of lesions, two different Gaussians are estimated: black holes (visible on both T1-w and T2-w images) and other lesions (only visible in T2-w images). These intensity distributions are then used in a MRF optimisation scheme, where the energy of a given segmentation is optimised using a simulated annealing approach. This energy, is defined by three different terms: a data term which is computed using the voxel intensities and the intensity distributions previously computed; a gradient term which is computed from the intensity distributions and the neighbouring voxel intensities; and finally a weighting term which is defined as the number of voxels that belong to a given class.

García-Lorenzo et al. [87] proposed an unsupervised method to classify each voxel of the brain as belonging to one of the four classes: MS lesions, WM, GM, and CSF, using only the intensities from PD-w, T1-w, and T2-w images. Firstly, the image intensities are modelled with a GMM, where each Gaussian represents one of the normal appearing brain tissues (GM, WM and CSF). In order to obtain the maximum likelihood estimation of the distribution parameters for each tissue (mean, covariance matrix, and mixing proportions), the EM algorithm is used. In this approach the authors proposed two contributions to improve the initialisation step of the EM and the sensibility to outliers when estimating the distribution parameters. First of all, instead of using an atlas as initialisation, several random initialisations of the distribution parameters are used, iterating the algorithm 50 steps using only the T1-w image. Afterwards, the classification with the highest likelihood is used to create a histogram for each tissue on each image to obtain its mean and variance. The covariance matrix for each tissue is then defined as a diagonal matrix with the variances of each image. In order to increase the robustness to outliers, the trimmed likelihood

estimator is used when computing the tissue parameters. Instead of using all of the voxels, a fraction of them with the lowest probability is rejected. In other words, the likelihood is only computed with the voxels most likely to belong to the model. Once the tissues have been segmented, the outliers with a Mahalanobis distance higher than a certain threshold are considered as possible lesions. This initial estimate is subsequently refined using heuristic rules. For instance, lesions should be hyperintense in PD-w and T2-w images, lesions smaller than a certain area (i.e. 9 mm^3) should be discarded, and lesions should be contiguous to WM and not to the brain border.

With a different strategy, Datta et al. [54] proposed an unsupervised method which does not rely on a tissue segmentation for identification and quantification of gadolinium enhancements using FLAIR images together with pre- and postcontrast T1-w images. The enhanced lesions on the postcontrast images are identified as regional maxima using an iterative elementary geodesic morphological reconstruction. A regional maximum on an image is defined as a group of connected voxels with signal intensities greater than the immediate neighbours. The original postcontrast T1-w image is eroded with a three-dimensional structuring element to obtain a T1'-w image. Afterwards, elementary geodesic dilation consisting on first dilating the image with the structuring element, followed by the point-wise minimum with the T1-w image is applied to the T1'-w image. This process is iterated until no change in the morphologically reconstructed image is found. Afterwards, regional maxima are obtained subtracting this image to the original postcontrast T1-w image. These hyperintense areas also include contrast passing within the vascular system and certain structures that lack a blood-brain barrier. In order to reduce these false positives lesions are also segmented in FLAIR using the method proposed by Sajja et al. [186]. Those postcontrast maxima that are not correlated with a FLAIR lesion are discarded. Moreover, the remaining false positives may be also discarded using a change ratio map obtained from the subtraction of pre- and postcontrast T1-w images. Finally, the lesion boundaries are refined using the fuzzy connectivity approach proposed by Udupa et al. [225].

Qualitative examples showing the result of automatic MS lesion segmentation algorithms are illustrated in figure 2.6. This figure provides visual examples from the representative methods of each strategy described in this section. The first row shows an example of the supervised approach based on atlas proposed by Souplet et al. [204], the second row corresponds to the supervised approach based on training from Subanna et al. [209], the third row illustrates the results from the unsupervised segmentation approach based on tissue of García-Lorenzo et al. [87], while the last row shows the result of the unsupervised

approach based on lesion properties from Datta et al. [54]. Note that this last algorithm was specifically designed to detect gadolinium-enhanced lesions in post-contrast T1 images.

2.3 Results

MS lesion segmentation approaches are usually evaluated using different quantitative measures and both synthetic and real MRI volumes. The most common data sets used in the works analysed and the typical measures computed for the evaluation are described in this section. Finally, a comparison and discussion of the results presented by the different approaches, highlighting the most interesting aspects, is also presented.

2.3.1 MS databases

As mentioned in the introduction, one of the main difficulties when comparing MS lesion segmentation approaches is the lack of a common database (with ground truth data) for training and testing the algorithms. Fortunately, this is becoming less of an issue with the appearance of new databases designed to meet this particular objective.

Synthetic databases are useful as an initial framework for testing the various approaches to segmentation. BrainWeb [47] is becoming a standard synthetic database for evaluating both tissue classification and MS lesion segmentation algorithms. BrainWeb contains simulated T1-w, T2-w and PD-w brain MRI data based on two anatomical models: a normal brain and a brain containing MS lesions. Moreover, this database provides different models according to parameters such as slice thicknesses, noise levels and levels of non-uniformity intensity. These simulated data sets are available in three orthogonal views - axial, sagittal, and coronal - although the majority of algorithms use only the axial view.

Even though synthetic data sets are useful for an initial evaluation, they are not as challenging as real data sets. In most cases, algorithms that are correctly tuned for synthetic data may not be successfully applied to a real in vivo acquired volume. Therefore, segmentation algorithms have to be tuned and tested with MRI volumes of real patients. A review of the work of recent years has highlighted a noticeable lack of a public MS MRI databases supplied by hospitals. However, the recent MS Lesion Segmentation Challenge (2008) [208] has provided a common framework for evaluating these algorithms, and for making comparisons among them. The MRI data for this competition was acquired separately by the Children's Hospital Boston (CHB) and the University of North Carolina (UNC). In total, 53 brain MRI volumes that were randomised into three different groups: 20 training

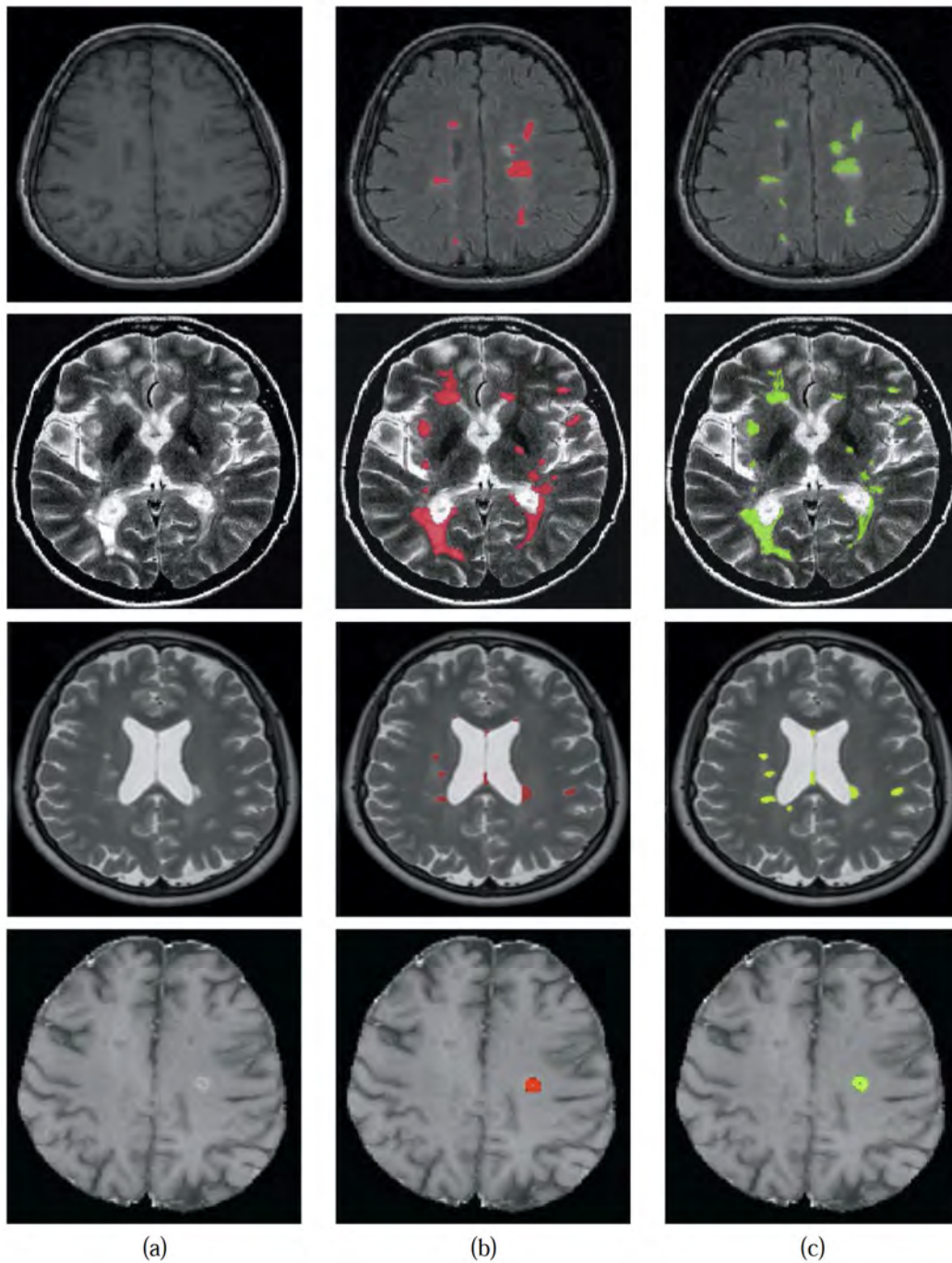


Figure 2.6: MS lesion segmentation examples. Each row shows the result of a representative algorithm of each reviewed strategy. From top to bottom: supervised approach based on atlas (Souplet et al. [204]), supervised approach based on training (Subanna et al. [209]), unsupervised approach based on tissue (García-Lorenzo et al. [87]), and unsupervised approach based on lesion (Datta et al. [54]). (a) shows the original images, (b) shows the results of the automatic MS lesion segmentation algorithms, and (c) the corresponding ground-truth annotations.

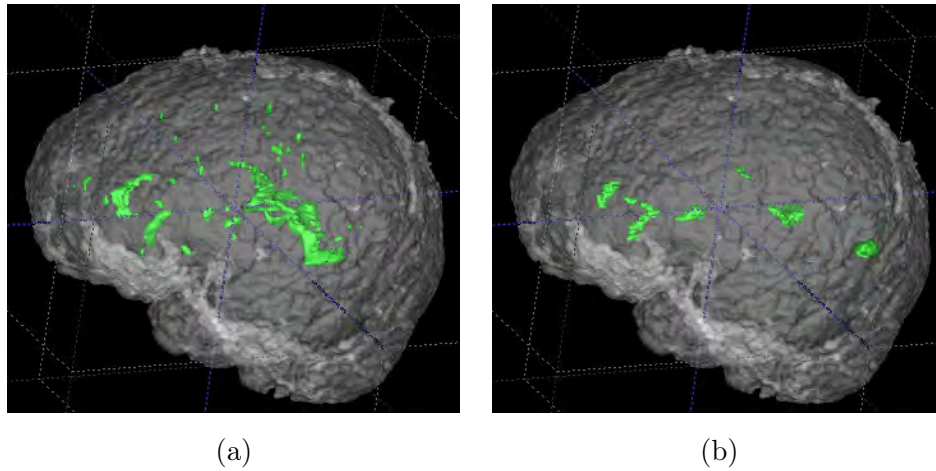


Figure 2.7: Generated 3D volume with MS lesions segmented by two different experts showing a large inter-rater variability. Note here the importance of using more than one manual annotation when evaluating the automatic algorithms (example extracted from the MICCAI challenge).

MRI volumes (including ground truth); 25 testing volumes (without ground truth) that were downloadable before the contest for training/validating the different algorithms; and 8 testing volumes used for the real contest from which the different participants extracted their on-site results.

2.3.2 Evaluation measures

The results are evaluated in different ways in the reviewed papers. However, all the measures are based on comparing the result of automated segmentations with the ground truth, which is usually annotated by an expert. In order to avoid intra- and inter-observer variability, segmentations from more than one expert should be used, since this provides a more consistent ground truth. Strategies such as the STAPLE algorithm [237] allows annotations from different experts to be fused. This is an important issue specially when considering the small volume of each lesion which may produce significant differences in the evaluation measures. Figure 2.7 shows an example of large discrepancy between two different expert annotations of the same patient (inter-rater variability).

Automated segmentations and ground truth can be compared by either comparing each voxel in each lesion (voxel-to-voxel), or using the whole detected lesion (lesion-to-lesion). Note that in both cases, voxels and lesions can be classified as a true positive (TP), false positive (FP), true negative (TN), or false negative (FN). Obviously, the objective is to

	Name	Computation
Hard	Accuracy Percentage agreement	$\frac{ TN + TP }{ TN + TP + FP + FN }$
	Dice similarity coefficient (DSC)	$\frac{2 \times TP }{2 \times TP + FP + FN }$
	Error rate	$\frac{ FP + FN }{ FP + FN + TP + TN }$
	Sensitivity Overlap Fraction Percentage of Correct Estimation True Positive Fraction/Rate	$\frac{ TP }{ TP + FN }$
	Specificity True Negative Fraction/Rate	$\frac{ TN }{ TN + FP }$
	False Positive Fraction/Rate	1 - Specificity
	Under estimation fraction (UEF) False Negative Volume Fraction	$\frac{ FN }{ TN + FN }$
	Over estimation fraction (OEF) Extra Fraction	$\frac{ FP }{ TN + FN }$
	Overlap objects fraction	$\frac{N_{obj}(TP)}{N_{obj}(Ref)}$
	Soft	Probabilistic similarity index
Probabilistic overlap fraction		$\frac{\sum P_{x,gs=1}}{\sum 1_{x,gs=1}}$
Probabilistic extra fraction		$\frac{\sum P_{x,gs=0}}{\sum 1_{x,gs=1}}$

Table 2.5: Common measures used in the evaluation of MS lesion segmentation methods.

obtain the maximum number of TPs and TNs, and at the same time reduce the number of FPs and FNs. However, in practice, one should find the best trade-off between these values, since increasing the number of TPs usually increases the number of FPs, while reducing the number of FNs also reduces the number of TNs. In fact, there is permanent debate about the best option: should we try to obtain more TPs or reduce FNs? One could argue that it is preferable to reduce the FNs at the expense of increasing FPs. However, increasing the number of FPs also leads to reduced confidence among neuroradiologists in computerised tools.

Table 2.5 summarises the most common measures used to evaluate the MS lesion segmentation algorithms. A distinction between two main groups has been done: those that evaluate a “hard” result (i.e. each voxel is assigned to only one tissue type) and those that provide a probabilistic result (i.e. each voxel has a membership value for belonging to the different tissue types). However, one should notice that all these measures are highly related. For instance, the probabilistic similarity index (PSI) is equivalent to the Dice

similarity coefficient (DSC) if the final segmentation is binary. Note also that different researchers used the same evaluation measure but under different names.

Other measures used that are not related to the ones described in the table are those based on distance measures (in voxels or in millimetres). The aim of these measures is to evaluate how far the boundaries of an obtained lesion segmentation are from those of the real one. Although in general, these measures are not common in most of the analysed works, they were used in the MS Lesion Segmentation Challenge [208].

In conclusion, although DSC has become something of a standard measure for evaluating MS lesion segmentation methods, none of the measures are perfect for this purpose. In fact, as stated by Cárdenes et al. [40], various different measures (i.e. measures based on intensity values, distances and connectivity) should be combined to obtain a more objective and reliable assessment.

2.3.3 Analysis of the results

This section provides a qualitative comparison of the results obtained by the analysed state-of-the-art approaches. Table 2.6 summarises the data, the evaluation measures, and the results obtained. As already mentioned, a quantitative evaluation among these approaches is a difficult task due to the variability in the data sets and in the measures used.

Table 2.6 clearly shows that the most commonly-used evaluation measures are the DSC and sensitivity (also known as the overlap fraction). The results obtained using real data range from 0.47 to 0.808 for the Dice coefficient. The highest Dice coefficients reported involving a large amount of data were obtained by Sajja et al. [186], who used 23 volumes from which they obtained a mean Dice coefficient of 0.78. The work presented by Van Leemput et al. [135] provides an example of how the variability of the ground truth affects the results obtained. Using the same automatically segmented data and comparing them with two different expert annotations, the DSC varies more than 10% (from 0.47 to 0.58). These results illustrate the usefulness of algorithms that combine the ground truth.

Several works also use sensitivity and specificity as evaluation measures (see table 2.6). Observe that the specificity reported by the different methods is always close to 1. This is because this measure evaluates the ratio between the number of voxels correctly classified as healthy divided by the number of voxels automatically classified as healthy. Therefore, considering that lesions are small spots within the whole volume, the specificity value always tends to be close to 1.

Article	Synthetic	Real (Database)	Measures	Results
Kamber (1995) [119]	-	12x56s (Montreal Neurological Inst.)	Error rate	0.02-0.04
Goldberg (1998) [94]	-	14x10s (Sheba Medical Center)	Sensitivity Specificity	0.87 0.96
Bedell (1998) [26]	-	13v (Univ. of Texas Medical School at Houston)	Qualitative	-
Boudraa (2000) [33]	-	10x22s (Hôpital d'Antiquaille)	Sensitivity	0.65
Alfano (2000) [3]	-	84x16s (Univ. Federico II)	Sensitivity	0.81
Van Leemput (2001) [135]	-	50v (BIOMORPH project)	DSC	0.47 & 0.58
Zijdenbos (2002) [251]	-	29v (Montreal Neurological Inst.)	DSC	0.60
He (2002) [106]	-	5v (Univ. of Texas Medical School at Houston)	Kappa	0.9
Anbeek (2004) [6]	-	20x38s (Univ. Medical Center Utrecht)	DSC Sensitivity OEF	0.80 0.79 0.19
Anbeek (2005) [5]	-	10x5s (Univ. Medical Center Utrecht)	DSC Sensitivity Specificity	0.808 0.815 0.999
Datta (2006) [55]	-	14v (Sanjay Gandhi Post-Graduate Inst. of Medical Sciences)	DSC Sensitivity OEF UEF	0.73 ± 0.11 0.72 ± 0.13 0.27 ± 0.21 0.28 ± 0.13
Sajja (2006) [186]	-	23v (Sanjay Gandhi Post-Graduate Inst. of Medical Sciences)	DSC Sensitivity OEF UEF	0.78 ± 0.12 0.88 ± 0.13 0.38 ± 0.27 0.11 ± 0.13
Wu (2006) [244]	-	6x2v (Slotervaart Hospital)	Sensitivity Specificity	0.70 (T2) 0.62 (BH) 0.98 (T2) 0.99 (BH)
Datta (2007) [54]	-	22v (Sanjay Gandhi Post-Graduate Inst. of Medical Sciences)	DSC Sensitivity OEF UEF	0.76 ± 0.18 0.74 ± 0.22 0.25 ± 0.62 0.26 ± 0.22
Khayati (2008) [121]	-	20x12~20s (Koorosh Diagnostics and Medical Imaging Center)	DSC Sensitivity OEF	0.7504 0.7402 0.2303
Garcia-Lorenzo (2008) [88]	-	7v (McConnel Brain Imaging Centre)	DSC	0.55 ± 0.05
Garcia-Lorenzo (2008) [90]	-	3v (MR and Image Analysis Research Centre)	DSC	0.56
Tomas (2009) [220]	-	9v (Boston Children's Hospital)	Sensitivity FPR	0.65 ± 0.29 0.71 ± 0.21
Subbanna (2009) [209]	-	10v (Montreal Neurological Inst.)	DSC FPR FNR	0.71 0.00 0.10
Akselrod-Ballin (2009) [2]	-	25+16vx24s (Scientific Inst. Ospedale San Raffaele)	DSC Sensitivity Specificity Accuracy	0.53 ± 0.1 0.55 ± 0.13 0.98 ± 0.01 0.97 ± 0.01
Shiee (2010) [199]	-	10v (National Multiple Sclerosis Society)	DSC	0.63
Freifeld (2007) [83]	BW (n3bf0) BW (n9bf0)	-	DSC	0.77 0.73
Garcia-Lorenzo (2008) [88]	BW (n3bf0) BW (n3bf20) BW (n3bf40)	-	DSC	0.87 0.85 0.63
Shiee (2008) [198]	BW (n3bf0) BW (n9bf0)	-	DSC	0.720 0.591
Shiee (2010) [199]	BW (n3bf0)	-	DSC	0.812

Table 2.6: Summary of the results presented in the articles analysed. For the simulated BrainWeb (BW) database, the noise and bias field of the tested volume ($n \times b f y$ means noise x and bias field y) are included. For the real data the number of slices (s) or volumes (v) and the origin of the database are shown. The two results for Van Leemput et al. are obtained by using the segmentation by two different experts as a ground-truth, while Wu et al. distinguish the results between T2 lesions (T2) and black holes (BH). DSC stands for Dice Similarity Coefficient and the $m \sim n$ means that there were between m and n slices.

The last four methods shown in table 2.6 used the simulated BrainWeb phantom to provide quantitative results. As already mentioned, the BrainWeb helps to compare the performance of different approaches more effectively. For example, each of these four methods included simulations with 3% added Gaussian noise and with zero bias field. The results show that the approach of García-Lorenzo et al. [88] clearly outperforms the others, obtaining a DSC of 0.87. The methods of Freifeld et al. [83] and the two by Shiee et al. [198, 199] obtained DSC values of 0.77, 0.72, and 0.81 respectively under the same conditions. Furthermore, García-Lorenzo et al. [88] also report the difference between using simulated data and using real MRI data when evaluating algorithms. Notice that when evaluating the same approach but using seven real volumes from the McConnel Brain Imaging Centre, the DSC drastically dropped to 0.55.

Finally, the quantitative results published during the 2008 MS Challenge⁴ [208] are presented in this section. The reported results are summarised in figure 2.8, which consists of three plots showing the results obtained when using three different evaluation measures: (a) the true positive rate (per lesion), (b) the false positive rate (per lesion), and (c) the average symmetric surface distance, which measures how far away the correctly segmented lesions are from the ground truth. The results in the plots have been ordered according to the final ranking in the on-site testing competition (from left to right). Notice that these plots illustrate slight differences in the results when using the ground truth from the UNC experts or from the CHB centre. Slightly better performances were obtained when using the annotations from the CHB expert during the training. As suggested by Styner et al. [208], this may be due to the fact that the UNC ground truth was obtained from two experts, while the CHB ground truth was obtained from only one expert⁴. In fact, the training and testing processes for the UNC can be performed using ground truth from different experts while this is not possible at the CHB centre. In general, figure 2.8(a) shows that there is room for improvement regarding the true positive rates. The best performances were obtained by Souplet et al. [204], Anbeek et al. [7] and Shiee et al. [197], each of whom had true positives rates of around 60%. There is also room for improvement regarding the false positive rates (see figure 2.8(b)), the best results being obtained by García-Lorenzo et al. [89], Bricq et al. [35], and Morra et al. [156], each with false positive rates of around 50%. This is to say, from each pair of marked lesions, only one was correctly placed. However, as stated in [186], this problem can be successfully overcome by applying an automated post-processing step for false positive reduction. Finally, figure 2.8(c) shows

⁴All this information can be found in: <http://grand-challenge2008.bigr.nl/> and <http://www.midasjournal.org/browse/publication/638>

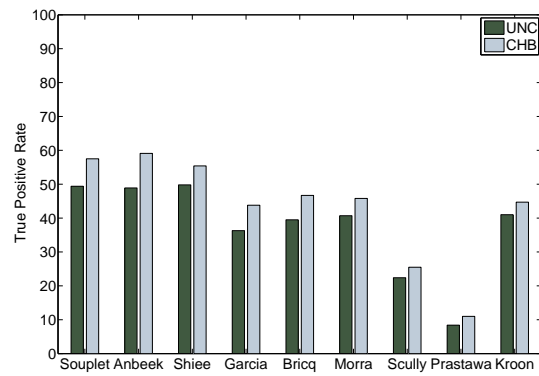
the average symmetric surface distance. Using this measure based on the distance from the ground truth to the lesion surface contour, the best results were obtained by Souplet et al. [204], Anbeek et al. [7], Shiee et al. [197], Bricq et al. [35] and Kroon et al. [125], with results of between 5 and 10 mm. Considering that the nominal diameter of an MS lesion is about 7mm [234], the results obtained, although promising, are still far from the requirements of a perfect automated volumetric tool.

In addition to the results obtained from the manual segmentations, the MS Challenge organisers computed a composite segmentation using the well-known STAPLE algorithm [237]. Specifically, the input for STAPLE included all the manual segmentations, as well as the segmentations provided by the workshop participants. Hence, it represented a composite of two human experts and nine automated segmentation methods. With respect to this evaluation experiment, the best sensitivities (which were from the works of Anbeek et al., Morra et al. and Kroon et al.) were around 0.5 while the specificities were close to 1, showing the ability of these algorithms to identify lesions correctly. These experiments using the STAPLE segmentation illustrate the improvement in reducing the false positive rate. In summary, the quantitative evaluation performed in the MS segmentation Challenge 2008 [208] has revealed both the challenges facing the participants as well as the need to develop new approaches to MS lesion segmentation.

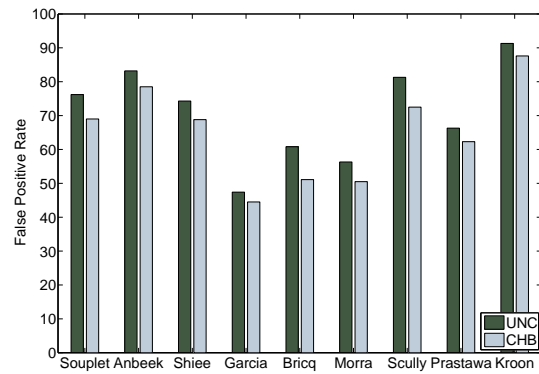
2.4 Discussion

As seen in previous sections, the most widely-used feature in all segmentation methods is voxel intensity, which is commonly employed with a multi-channel approach. In addition, features based on modelling the voxel neighbourhood are also used in some approaches to introduce (local) spatial information to the algorithms. Regarding the modalities, T1-w images are widely used for the tissue segmentation and also for the black holes and enhanced lesion segmentation. T2-w and PD-w images are typically used for detecting MS lesions. However, the major drawback of these images is the similarity in the intensities of lesions and CSF. Due to this similarity, the discrimination between ventricles and lesions may be difficult, especially when they are connected. Some approaches perform another segmentation step to solve this problem. FLAIR images also provide good discrimination between lesions and healthy tissue but, as some authors have pointed out, they have problems when dealing with sub-cortical structures.

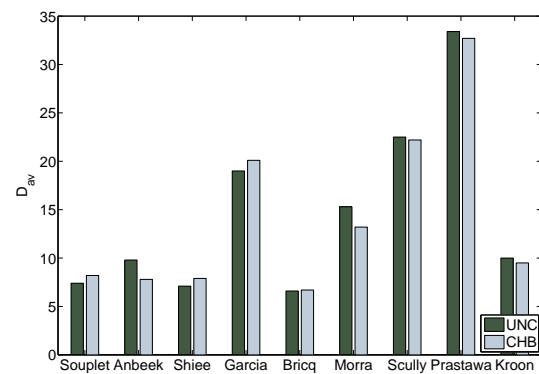
In regard to the segmentation algorithms shown in table 2.3, note that the vast majority



(a)



(b)



(c)

Figure 2.8: Results extracted from the works presented in the 2008 MS Challenge. Each plot details the results when using ground truth provided by either the UNC or the CHB experts. (a) shows the true positive rate per lesion, (b) the false positive rate per lesion, and (c) the average distance between obtained and ground truth segmentations. The results are ordered according to the final ranking in the on-site testing competition.

belong to the clustering family [85]. However, clustering algorithms do not naturally deal with the spatial information needed to obtain proper segmentations. This can be introduced using Markov random fields or fuzzy connectedness. The most widely-used clustering approaches are based on the FCM algorithm [29] and the EM algorithm [64]. The FCM is usually used to group the different tissues of the brain into three (WM, GM, CSF) or five (WM, GM, WM+GM, CSF, GM+CSF) different classes [15]. The aim of using five classes is to take partial volume effects into account, allowing a voxel to be composed of more than one tissue type. The EM algorithm also allows different models to be used for different tissues, which is indeed very useful since WM and GM can be assumed to follow Gaussian distributions while CSF does not follow any known distribution [186]. Despite the fact that EM methods are usually more accurate in the presence of noise than those of the C-Means family, they share with them the possibility of converging to a local maximum (or minimum). In a way similar to the EM, the AMM algorithm also allows distributions to be estimated. As argued by Khayati et al. [121] the advantages of the AMM over the EM are the non-requirement of initial knowledge for the number of terms and the initial estimate for the parameters. Moreover, as all the data are not used simultaneously to update the parameter estimation, less computational time is required. Although the clustering approaches introduced above are unsupervised algorithms, there are also supervised approaches which use these clustering algorithms to perform the classification. See for instance the works of Shiee et al. [198, 197] and of Souplet et al. [204].

While iterative approaches are inherently unsupervised approaches (even though they may use a priori information), supervised approaches have also been proposed for the purpose of MS lesion segmentation. Most methods rely on pattern recognition techniques to detect voxels that are either outliers to the tissue models or are similar to the lesion model (derived from a training set). In general, these methods tend to be more robust if a good training set is provided. However, collecting enough representative samples for a fully automated and accurate system may be difficult due to differences between scanning machines, patient anatomy, or expert reading. A large variety of supervised classifiers such as kNN, ANN, or Bayes are used for the MS lesion segmentation problem.

Atlas registration is another way of tackling MS lesion segmentation in brain MR images. These approaches are useful since the lesion intensities often overlap with the intensities of other structures in the brain. Hence, atlases might provide valuable contextual information to eliminate possible false positives. As a drawback, the registration process is a computationally intensive procedure. Moreover, the results obtained by these methods are affected

by the physiological variability of each subject and may lead to erroneous results in the case of diseased brains. This is because atlases are based on normal brains and lesions are very variable in size, shape and location, making it difficult to construct atlases from diseased brains (although this has been analysed in [125]).

An analogous problem to atlas registration is the temporal registration of MRI volumes. This procedure is necessary to track and compare the atrophy of the WM and GM, as well as to monitor the evolution of MS lesions. However, this registration can also be used to detect lesions, which are defined as those regions with apparent local volume variation [217, 177, 73]. Note that a registration step will be also needed for including information coming from other modality images [22] such as DTI [32].

Regarding the evaluation measures, the Dice coefficient and sensitivity are the two most commonly-used measures for lesion segmentation. Both measures take the number of true positives and false negatives into account, while the main difference is that the DSC also considers the number of false positives. On the other hand, specificity is usually high because this measure takes the number of true negatives into account and since lesions are small spots in the image, specificity has values close to 1 (the accuracy measure behaves in the same way as well).

Analysing the results, note that those methods that accomplished DSC values above 0.7 (which is considered a good segmentation result) use prior knowledge to help the detection and delimitation of the lesions. This prior knowledge is mostly defined by a set of manual segmentations that can aid to model a tissue intensity distribution or constrain it using either a classification scheme [6, 5, 209] or a registration based [55, 121] approach. However, this information may also be encoded as a set of pre-defined rules [55] such as that enhancing lesions should appear brighter in T1-w images with gadolinium enhancement than in T1-w images without enhancement.

We have also pointed out the difficulty of making quantitatively comparisons of results that are obtained using different databases, mainly because there is not a standardised evaluation measure nor a common public database. Note that even when using the same evaluation measures, cases with significant different lesion load like the ones shown in figure 2.9, can bias the results. Observe that it is likely to have a higher false positive rate in (b) than in (a) due to the small lesion volume of that patient. In respect to this difficulty of performing evaluations, the MS Challenge has helped to create a data set that can become a standard for evaluating the automated segmentation of MS lesions. In this competition, supervised strategies obtained the best results with the best ranked method

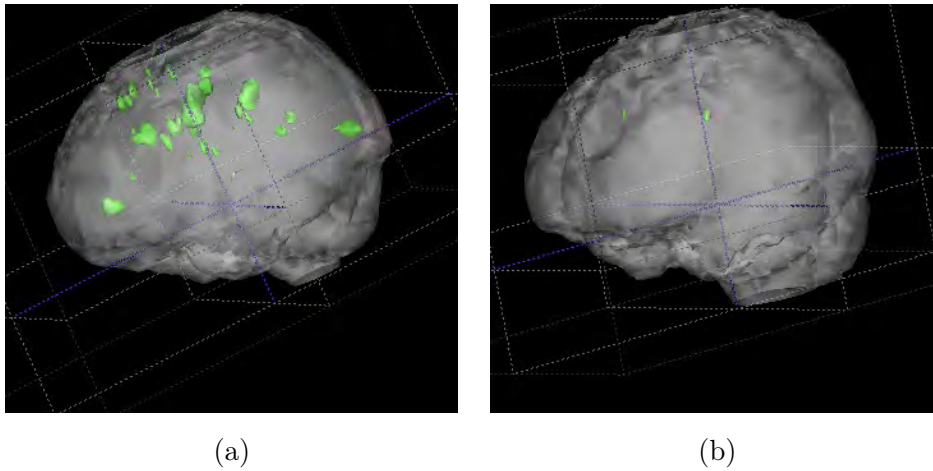


Figure 2.9: Generated 3D volume with different lesion load.

being an atlas-based strategy presented by Souplet et al. [204]. This method uses a tissue based strategy aided by a probabilistic atlas that provides prior knowledge during the tissue segmentation step. As pointed out in section 2.2.5, the advantage of the supervised algorithms is that they can automatically learn the characteristics of both normal tissue and lesions, while the algorithms which follow an atlas-based approach incorporate also spatial information to perform the segmentation task (as is done by Souplet et al. [204]).

2.5 Conclusions

This chapter has reviewed the automated approaches for MS lesion segmentation, classifying them according to the strategy used. In addition, the results obtained by these approaches have been summarised and compared, reviewing also the most common data sets and evaluation measures used in this field. We observed that the automated segmentation of different MS lesion types in MRI is a challenging task due to heterogeneous intensity values among the different MR images (enhancing lesions, black holes, and hyperintense lesions). Despite recent progress, there is not yet a specific automated lesion segmentation approach robust enough to emerge as a standard for clinical practice. The main reasons for this are the unsatisfactory results they produce, the high computational demand required, or their insufficient generalisation capability. Moreover, most of these approaches are unable to locate MS lesions inside the sub-cortical structures of the brain.

Finally, we would like to point out the importance of using prior knowledge to guide the lesion segmentation. Supervised approaches that rely on similar segmented cases usually

outperform unsupervised strategies. Moreover, atlas-based approaches (a subcategory of supervised), provide a good starting point for tissue segmentation and outlier detection.

BRAIN ATLASES: CONCEPTS AND APPLICATION

*There's two ways to deal with mystery:
uncover it, or eliminate it.*

Andrew Ryan, "Bioshock" (2007)

3.1 Atlas based segmentation

As presented in chapter 1, the automated segmentation of MR brain images is a challenging task due to image artifacts (such as intensity inhomogeneities and partial volume effects) and due to the fact that different anatomical structures may share the same tissue contrast. Hence, *a priori* anatomical information is essential for simplifying the segmentation task [39], as concluded in chapter 2. Prior information may be provided in different ways, for instance, as a set of predefined rules based on known tissue properties, or as a set of manual expert annotations. In this study, we focus on anatomical priors from an atlas to be matched to the target volume we wish to segment. Here, we consider an atlas as two image volumes: one intensity image (or *template*) and one segmented image (or *labelled* image). Note however that, as stated in [9], active shape models [52] or active appearance models [51] can also be considered as atlases since they bring spatial prior knowledge to the segmentation process.

At this point, the segmentation turns into a registration problem. Volumetric registration is often done in two steps. Firstly, a global registration (affine or rigid transformation) is performed to obtain an initial alignment at a low computational cost. Secondly, a local

registration is applied to adapt general models to a specific anatomy. Note that this local registration provides a better match between different brains at the expense of a high computational cost. Multi-resolution strategies may be used to reduce this computational cost [182].

Medical image registration has been widely reviewed in the literature [144, 137, 110, 167, 93]. These studies include reviews of registration techniques that can be used to align an atlas to an unseen MRI scan. However, in this chapter we will review the use of atlases for automatic segmentation of the brain, in general, to study its history and evolution to extract strengths and weaknesses that can be, later on, applied to MS lesion segmentation.

In particular, we first distinguish between three different ways of integrating the atlas information into the whole segmentation process: *Label propagation*, *Multi-atlas propagation* and *Probabilistic atlas-based segmentation*. We then go on to review the atlas-based segmentation methods according to their medical target: those that segment the brain and its internal structures (such as the amygdala or putamen), those that target brain tissues in healthy brains, those that target brain tissues in fetus, neonates and elderly subjects, and, finally, those that segment damaged brains with either focal lesions (such as multiple sclerosis lesions) or space-occupying lesions (like tumours).

For comparison purposes, atlas-based segmentation methods should be applied to a common database and quantitatively validated using ground truth. Unfortunately, only very few of such methods and databases are publicly available. Atlas-based segmentation methods also aim to segment different targets, such as, for instance, brain structures, brain tissues, or lesions. Our contribution is closely related to this idea, comparing atlas-based segmentation approaches qualitatively and quantitatively according to their strategy, target and accuracy reported in the literature.

The rest of this chapter is organised as follows. In section 3.2, we introduce the different types of public atlases, also describing their creation process. Then, in section 3.3, we present the strategies used to integrate atlases into the segmentation process. Finally, in section 3.4 we analyse and classify the reviewed methods according to the ultimate objectives of segmentation.

3.2 Type of brain atlases

The construction of a realistic anatomical brain atlas is a time-consuming task, especially when aimed at describing human data variability. Public atlas repositories have been

created to provide the research community with MRI data to go with the manual segmentations (or annotations) performed by expert radiologists. The contribution of these repositories is twofold: firstly, they allow the training of new segmentation algorithms, and secondly they allow evaluation data to be standardised for the developed algorithms.

3.2.1 Topological atlases

First attempts at atlas construction of the human brain were based on a single subject. In the literature these atlases are called *topological*, *single-subject* or *deterministic* atlases. The single-subject atlas is often a volume image that has been selected from a data set to be representative of the objects to be segmented in other images (average size, shapes or intensity).

In medicine, pioneering work was done with the Talairach atlas [213, 214], which was proposed to identify deep brain structures in stereotaxic coordinates. One of the first deterministic digital atlases was provided by the Visible Human Project of the National Library of Medicine [139]. The goal of this project was to create complete and detailed three-dimensional anatomical representations of normal male and female human bodies. These representations were obtained through the acquisition of transverse computed tomography (CT), MR, and high-resolution cryosection images of male and female cadavers. Also derived from a digitised cryosectioned human brain, the Karolinska Institute and Hospital, Stockholm, created a computerised brain atlas that was designed for display and analysis of tomographic brain images. The atlas includes the brain surface, the ventricular system and about 400 structures, and all Brodmann areas [98, 218].

Nowadays, the vast majority of deterministic atlases are generated from imaging acquisition. For instance, the Surgical Planning Laboratory's (SPL) digital brain atlas, developed by the Harvard Medical School [122], is based on a 3D MR atlas of the human brain to visualise spatially complex structures. For CT acquisitions, Bajcsy et al. [21, 20] created an artificial CT anatomical atlas based on the stained slices of a dead soldier's brain belonging to a 31 year-old normal male (from the so-called Yakovlev Collection). The McConell Brain Imaging Center [47] provides the research community with a digital brain phantom, based on 27 high-resolution scans from the same individual. Its average resulted in a high-resolution (1 mm isotropic voxels) brain atlas with an increased signal-to-noise ratio. This brain template is the reference data in the *BrainWeb* simulated brain database. Recently, twenty new normal anatomical models have become available as well as an anatomical model of a brain with MS lesions.

3.2.2 Probabilistic atlases

Atlases based on a single-subject are not constructed to represent the diversity of human anatomy. To better characterise the variability of anatomical structures, atlases have been constructed on the basis of populations. These atlases are often cited as *population-based*, *statistical* or *probabilistic* atlases. Such templates are in continuous evolution, as new images can easily be incorporated. Moreover, the population represented by a probabilistic atlas can be easily subdivided into groups according to specific criteria (age, sex, handedness, etc.). As for single-subject atlases, the first population-based atlases were based in Talairach space [111, 242]. Later, to compensate for the implicit limitations of Talairach space, such as poor resolution across slices (from 3 to 4 mm), population-based atlases from MR images were proposed. A composite MRI data set was constructed by Evans et al. [75] from several hundreds of normal subjects (239 males and 66 females of 23.4 ± 4.1 years old). All the scans were first individually registered in the Talairach coordinate system. Afterwards, they were intensity normalised and, finally all the scans were averaged voxel-by-voxel and probabilistic maps for brain tissue were created. The same procedure for constructing an average brain was later used by the International Consortium for Brain Mapping (ICBM) on 152 brains and later on 452 brains [129]. Figure 3.1 shows the tissue probabilities for one central slice of the ICBM452 template. The UCLA Laboratory Of Neuro Imaging (LONI), which is a member of the ICBM, provides also atlases for MR brain imaging contrasts, such as T2-weighted or diffusion tensor imaging (DTI) [130].

Another widely-used repository for MRI brain data is the Internet Brain Segmentation Repository (IBSR) [147]. The MRI studies contained in this database have also been used either to define a set of topological atlases for multi-atlas strategies (using the included manual segmentations) or to construct a probabilistic atlas after co-registering all the segmented cases to a standard space and computing the frequency of each voxel to belong to a specific structure [199, 25].

Interest in disease-based atlas construction [148, 74] has increased in recent years. For instance, the ICBM provides an Alzheimer's disease template. Disease atlases allow quantitative examination of the history and evolution (due to natural disease evolution or reaction to clinical treatment) of a specific disease.

Important questions arise when generating population-based atlases, such as selecting a reference space or the registration method for the data alignment. Many researchers have proposed new strategies to create unbiased average templates and multi-subject registration [99, 62, 118, 30, 142, 165, 253, 49].

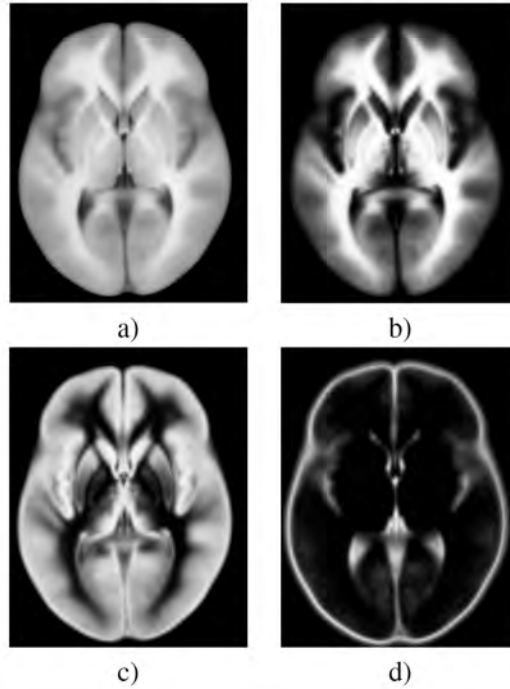


Figure 3.1: ICBM452 population-based atlas [129]: a) T1-weighted mean, b) white matter, c) grey matter, and d) cerebrospinal fluid.

3.3 Segmentation strategies

In formulating atlas-based segmentation, we can define the input image to be segmented, $I(x)$, as $I : x \in \mathbb{R}^3 \mapsto I(x) \in \mathbb{R}^N$, where x represents the 3D voxel coordinates, and N is the number of intensity values of multi-spectral MRI data for each voxel. Following this notation, when dealing with single-subject atlases, we can differentiate between the grey level volume, $\pi^I(x)$, defined as $\pi^I : x \in \mathbb{R}^3 \mapsto \pi^I(x) \in \mathbb{R}$, and the corresponding labelled volume, $\pi^L(x)$, defined as $\pi^L : x \in \mathbb{R}^3 \mapsto \pi^L(x) \in \mathcal{L}$, where $\mathcal{L} = \{1, \dots, C\}$ and C is the number of labels. If probabilistic atlases are available, for each class c the probabilistic volume $\pi_c^P(x)$ is defined as $\pi_c^P : x \in \mathbb{R}^3 \mapsto \pi_c^P(x) \in \mathbb{R}$ where $c \in \{1, \dots, C\}$.

By definition, volumes $\pi^I(x)$, $\pi^L(x)$ and $\pi_c^P(x)$ are on the same spatial coordinates, which we refer to as the *atlas space* \mathcal{X}_π . On the other hand, the input image usually lies in a different space, which we refer to as the *image space* \mathcal{X}_I . Therefore, in order to use the atlas information, a transformation $\tau : x_\pi \in \mathcal{X}_\pi \mapsto x_I \in \mathcal{X}_I$ must be found in the space of all possible transformations \mathcal{T} . The process of finding this transformation is commonly known as image registration and is defined in equation 3.1 as an optimisation problem,

where $\hat{\tau}$ is the estimated transformation and δ is a similarity metric used to compare the input image and the transformed intensity image volume of the atlas.

$$\hat{\tau} = \arg \max_{\tau \in \mathcal{T}} \delta(I(x), \pi^I(\tau(x))). \quad (3.1)$$

Finally, we can define the final segmentation, $S(x)$, as $S : x \in \mathbb{R}^3 \mapsto S(x) \in \mathcal{L}$.

3.3.1 Label propagation

Once the registration problem is solved, the easiest and fastest way to assign a label to each input image voxel is to propagate the atlas labels to image space, \mathcal{X}_I [21, 91, 154, 46, 41, 114, 57, 60, 19, 123, 244, 42]. This segmentation procedure is defined as:

$$S(x) = \pi^L(\hat{\tau}(x)). \quad (3.2)$$

With this strategy, the segmentation process relies on a registration process that aims to estimate the anatomical differences between the atlas and the input image volumes. Registration errors exist in all real-world applications but errors are larger if differences between two images are large. We assume that the atlas is *close* to the subject's anatomy. Otherwise, in cases where large anatomical differences exist, large registration errors may cause important segmentation errors.

Global rigid and affine transformations are usually enough when dealing with intra-subject medical applications, such as longitudinal studies of illness progression or multi-modal registration for radiotherapy planning. However, when dealing with inter-subject applications such as atlas matching, the anatomical variation between different subjects can only be captured using non-rigid algorithms.

Volume partitioning can be performed to account for these local deformations. In general, either the moving image (usually the atlas) or the target image or both image volumes are decomposed on smaller sub-volumes and these sub-volumes are then registered in a hierarchical manner using rigid and affine transformations [109, 8, 103].

Other partitioning approaches define a uniform grid, usually called the free-form deformation grid, and then apply a non-linear transform to each of the grid vertices. Depending on accuracy and time efficiency requirements, grid vertices can be defined as the voxels for the whole volume. Common non-linear transforms based on mathematic transformations are, for instance, cosine based functions [11], B-spline curves [182, 159] or level set partial

differential equations [229]. Other functions used to define a displacement field are based on the thermodynamics concept of demons [216, 231], optical flow models [230, 172] or elasticity properties [21, 91].

3.3.2 Multi-atlas label propagation

Label propagation has been extended to multiple atlases to better deal with the registration errors obtained when using a single atlas and also to better account for anatomical variability. With an atlas database, those voxels with low agreement between different label propagations can be discarded in order to minimise outliers. Due to its strengths over simple label propagation, this technique presents an improvement in accuracy when dealing with the segmentation of objects with well-defined shape that may present slight deformations between images.

There are two important considerations to take into account when dealing with a set of atlases. The first is related to the number of atlases to be used to segment a new patient and how to select them. We refer to this issue as the selection criteria problem. Different studies [124, 243, 4, 143] conclude that using more than one topological atlas improves accuracy, but that it is not necessary to use all the cases in a database.

Two principal methods exist for choosing the best matching cases: either using meta-information (which may not always be available), or using similarity metrics to compare the images. In order to use this second method, the new image must be aligned to all the manually segmented cases. One possible technique is to warp all atlases into a common space, and the subject to be segmented will then be matched in this space. This considerably reduces the number of registrations. However, with this strategy, there exists a strong dependency on the initially selected reference space. Therefore, new group-wise registration techniques [99, 62, 118, 30, 142, 165, 253, 49] may prove a better way of solving this issue. These techniques try to register all the subjects together constructing an average reference template at each step. It is also advisable to use a combination of similarity metrics to avoid bias from using the same metric as in the registration step.

The question of how atlases should be combined remains. Voting rules are commonly applied [124, 243, 4, 143, 107, 10]:

$$S(x) = \arg \max_c \sum_{i=1}^P w_i(x) \cdot f(\pi_i^L(\hat{\tau}_i(x)), c), \quad (3.3)$$

where c represents a class, P is the number of atlases, w_i is a weight function that may

vary for each atlas voxel and f is defined as:

$$f(\pi_i^L(\hat{\tau}_i(x)), c) = \begin{cases} 1 & : \pi_i^L(\hat{\tau}_i(x)) = c. \\ 0 & : \pi_i^L(\hat{\tau}_i(x)) \neq c. \end{cases} \quad (3.4)$$

This step can be seen as a specific case of classifier fusion. Within this voting procedure, the function $w_i(x)$ can either be defined as a constant value for all atlases to use a majority voting strategy [245, 178] ($w_i(x) = \frac{1}{P}$), a different constant value for each atlas to use globally-weighted voting [126] ($w_i(x) = K_i$) or a function for each voxel to use locally-weighted voting [10, 115] ($w_i(x) = f_i(x)$). Recently, a generative probabilistic model for this fusion step was presented by Sabuncu et al. [184].

Other combination strategies based on the Expectation Maximisation (EM) algorithm have also been presented [237, 179]. However, these methods usually obtain a lower accuracy when compared to local weighting methods [10]. Intuitively, these techniques weight each atlas according to estimated agreement or similarity with respect to the other atlases. However, in the case where the atlases with the highest agreement are not the best match for the subject being segmented, such a weighting procedure may bias the segmentation.

3.3.3 Probabilistic atlas-based segmentation

When probabilistic atlases are used, voxel probabilities can be integrated as part of a statistical Bayesian framework [103, 228, 135, 82, 14, 227, 35, 36, 204] as defined by:

$$S(x) = \arg \max_c p(I(x)|c) \cdot p(c), \quad (3.5)$$

where $p(I(x)|c)$ represents the conditional probability of the intensities, and $p(c)$ are the class priors. Probabilistic atlases can also be used within variational frameworks [199, 25, 145, 198, 197]:

$$S(x) = \arg \min_c (E_d + \lambda \cdot E_s), \quad (3.6)$$

where E_d is the data energy term, E_s is the smoothness energy term, and λ is a user-defined parameter.

Either parametric (for instance using Gaussian mixture models) or non-parametric approaches (for instance using Parzen windows) can be used to estimate $p(I(x)|c)$ and E_d . Initial estimates of such models often use propagation of the atlas probabilities [14, 35,

	Description	Strengths	Challenges
Label propagation (PA)	Atlas labels are directly applied in image space Application: ROI definition or contour initialisation	<ul style="list-style-type: none"> +Intuitive +Straightforward +Single registration 	<ul style="list-style-type: none"> •No anatomical variability •Atlas dependant •Registration dependant
Multi-atlas propagation (MA)	Multiple labels are combined in image space Application: Segmentation of structures (well-defined shape)	<ul style="list-style-type: none"> •Anatomical variability •Outlier minimisation 	<ul style="list-style-type: none"> •Atlas selection and combination •Multiple registrations
Probabilistic segmentation (PA)	Atlas values are used in a probabilistic framework Application: Segmentation of outliers or high-variability classes	<ul style="list-style-type: none"> •Anatomical variability •Single registration •Multiple input features 	<ul style="list-style-type: none"> •Atlas weighting •Complex model estimation

Figure 3.2: Advantages and drawbacks of the different reviewed atlas-based segmentation strategies.

36, 145]. Class priors ($p(c)$) and smoothness term (E_s) may also be encoded using atlas probabilities [204], sometimes in combination with other spatial priors [199, 25, 103, 228, 135, 82, 227, 198, 197] often modelled by Markov Random Models.

Some other methods [42, 119, 251, 125, 2] directly combine atlas probabilities with other image features such as voxel intensities or spatial coordinates (x) to train a classifier. These classifiers allow several features to be combined without the need to estimate a probability distribution in a high dimensional space.

The above-mentioned strategies use all the atlas probability values after registering the atlas with the patient. In order to reduce the effect of registration errors, some approaches select only a subset of voxel samples with high probability per class. These atlas samples can then be used to train a classifier [45, 233, 61, 220], to estimate class distributions [173, 174], or as initialisation points for a contour based segmentation [96].

3.3.4 Summary

In this section, we have presented three strategies for dealing with the information provided by an atlas (see the scheme in figure 3.2 for a summary of the main advantages and drawbacks for each strategy). The most straightforward technique is to assign atlas labels after registration of the intensity atlas volume with the subject we wish to segment. This label propagation technique is highly dependent on both the atlas image and the registration procedure, and it may not be desirable when dealing with subjects from very different populations. Nevertheless, label propagation is widely used as a segmentation method to define a region of interest (ROI) for further segmentation [244, 119, 220] or to initialise an

active contour strategy [19, 66].

Furthermore, several topological atlases can be taken into account to improve the capture of anatomical variability between different scans. This multi-atlas propagation is in fact an extension of label propagation. Therefore, these techniques are desirable when segmenting objects with a well-defined shape where there is low anatomical variability between different images.

Finally, when using probabilities (either from a probabilistic atlas or a combination of topological atlases), a probabilistic model of the input images can be estimated. This probabilistic model, which may be unknown, can be estimated by different methods (i.e. parametric, non-parametric, or trained classifiers) that apply outlier rules to segment the images into new classes not present in the atlas. Moreover, these models may also be learned from a subset of image voxels to reduce the effect of registration errors.

3.4 Segmentation methods and clinical targets

Tables 3.1, 3.3, 3.5 and 3.6 offer a compact overview of methods to segment the brain using atlas information. We have grouped the segmentation algorithms into four categories according to their medical target: 1) brain structures with well-defined shapes (such as the whole brain or the hippocampus), 2) brain tissues in healthy subjects (namely GM, WM and CSF), 3) brain tissues in challenging populations such as fetuses, neonates, and elderly subjects, and 4) damaged brains with either focal lesions (e.g. white matter lesions) or space-occupying lesions (e.g. tumours). Figure 3.3 depicts the relation of these different categories with the types of atlases used. The tables also specify the type of atlas used, the registration technique applied, and the corresponding atlas-based segmentation strategy.

In this section, we briefly describe these methods followed by a qualitative and quantitative evaluation of the results reported in the literature. We select the Dice similarity coefficient (*DSC*) as a measure for comparison since it is the most commonly used in the studies analysed.

3.4.1 The brain and its internal structures

The brain itself may be the first structure to be targeted. The procedure of removing non-brain tissue is a well-known pre-processing step in brain imaging. Several reviews and comparisons have been presented recently [134, 31, 78, 104], concluding that, among all

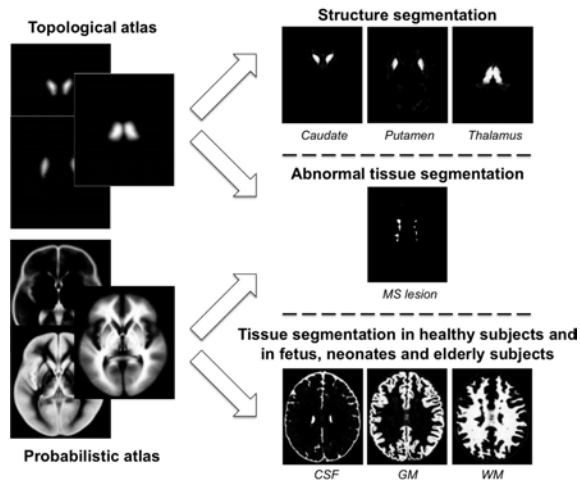


Figure 3.3: Diagram of how atlases are used to segment structures, healthy tissue (in both healthy and challenging population), and abnormal tissue and lesions. The MR brain data sets and their manual segmentations were provided by the Center for Morphometric Analysis at Massachusetts General Hospital and are available at <http://www.cma.mgh.harvard.edu/ibsr/>.

brain segmentation methods, atlas-based segmentation (mainly by label propagation of probabilistic atlases) is applied as an initial step, although further processing is needed to obtain a good brain segmentation [206, 190].

In this section, we focus on the segmentation of the internal structures of the brain such as the amygdala (AMY), accumbens (ACC), caudate (CAU), hippocampus (HIP), pallidum (PAL), putamen (PUT), or the thalamus (THA) (as shown in figure 3.4). Note that these structures present well-defined shapes that show some anatomical variability between subjects.

Due to the lack of clearly defined edges between some brain structures and substructures, approaches based solely on voxel intensities are expected to produce poor results [82]. Therefore, a priori spatial information for anatomical structures is required. Several studies have adopted label propagation strategies, focusing their contribution on developing new registration techniques based on, for instance, elasticity [21, 91, 154, 114, 57], deformation vectors [46], fluid mechanics [41], thermodynamics [60], optical flow [19], or hierarchical methods [123]. Propagated labels can later be extended to define a fuzzy label map (with higher values for voxels inside the label masks and lower values for voxels outside the masks) guiding a fuzzy controller. For instance, Ciofolo and Barillot [42] designed a fuzzy controller to guide a competitive level set approach initialised by cubes inside the brain

Article	Atlas type		Registration transforms		Strategy
	Statistical	Topological	Global transforms	Local transforms	
Bajcsy (1983) [21]	No	1 manual	Affine	Elastic	LP
Gee (1993) [91]	No	1 manual	Affine	Elastic	LP
Miller (1993) [154]	No	1 manual	Affine	Mechanoelastic	LP
Collins (1995) [46]	No	1 manual	Affine	Deformation vectors	LP
Christensen (1996) [41]	No	1 manual	Affine	Fluid mechanics	LP
Davatzikos (1997) [57]	No	1 manual	Affine	Elastic	LP
Iosifescu (1997) [114]	No	1 manual	Affine	Elastic	LP
Dawant (1999) [60]	No	1 manual	Affine	Demons	LP
Baillard (2001) [19]	No	1 manual	No	Optical flow	LP
Fischl (2002) [82]	12 manual	No	Affine	No	PA
Klein (2005) [123]	No	1 manual	Affine	Mindboggle	LP
Klein (2005) [124]	No	19 manual	Affine	Mindboggle	MA
Heckemann (2006) [107]	No	29 manual	Affine	B-splines	MA
Pohl (2006) [168]	80 manual	No	Affine	Hierarchical	PA
Han (2007) [103]	80 manual	No	Affine	Hierarchical	PA
Bazin (2008) [25]	18 manual	1 manual	Rigid	No	PA
van der Lijn (2008) [227]	19 manual	No	Affine	B-splines	PA
Aljabar (2009) [4]	No	275 manual	Affine	B-splines	MA
Artaechevarria (2009) [10]	No	18 manual	Affine	B-splines	MA
Ciofolo (2009) [42]	No	1 manual	Rigid	Hierarchical	LP
Lötjönen (2010) [143]	No	17 manual	Affine	Hierarchical	MA

Table 3.1: Classification of automated atlas-based segmentation methods for structures. Rows are the segmentation targets, while columns refer to: the type of atlas, the number of volumes used to build the atlas, the registration method, and the atlas strategy (Label Propagation (LP), Multi-Atlas Propagation (MA), and Probabilistic Atlas segmentation (PA)).

scan.

As mentioned in section 3.3, multi-atlas approaches have become a recent area of research on automated atlas-based segmentation to provide improved capture of anatomical variability. Heckemann et al. [107] proposed the combination of 29 topological atlases after registration via a vote rule decision fusion approach. This technique treats each propagation as a separate classifier and then applies the label with the maximum number of occurrences to each voxel. A similar study by Klein et al. [124] assigned a structure label to each voxel by applying the findings of their previous study [123] on 19 atlases.

When using a large database with previously segmented cases some of the segmentations can deviate greatly from the new subject and therefore bias the result and decrease segmentation accuracy. To resolve this issue, Aljabar et al. [4] presented a new approach based on selecting a subset of volumes instead of using all 275 cases (the total number of volumes in the database). The selection was based on taking both volume similarity and meta-information into account. Moreover, as pointed out by Artaechevarria et al. [10], even if the best atlases are chosen, their variability can affect the final result depending on the fusion technique applied. In this last study, the authors proposed different weight-

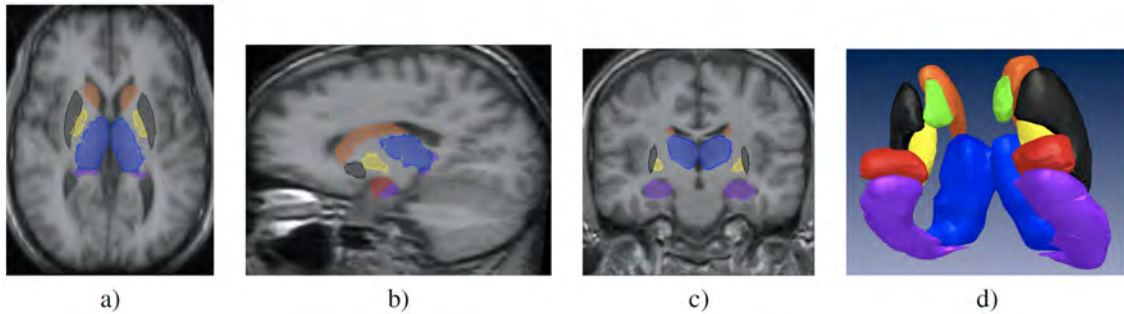


Figure 3.4: Internal structures of the brain. a) Axial plane, b) sagittal plane, c) coronal plane, and d) 3D representation of the following structures: thalamus (blue), putamen (black), pallidum (yellow), hippocampus (purple), caudate (orange), amygdala (red) and accumbens (green).

ing methods: either globally, using similarity measures for the whole volume, or locally, using a small neighbourhood area. In their study, local methods were found to improve the final segmentation when compared to global methods. The same conclusion was obtained by Lötjönen et al. [143], who reviewed selection and combination algorithms as well as non-rigid registration. Furthermore, they noticed that although atlas selection methods provided better segmentation results than random selection, there was still a clear difference with respect to the results obtained when knowing the optimal set of atlases.

While these studies create a statistical atlas and keep the labels with the highest probability per voxel, there are others that use atlas probabilities on a complex statistical framework (mainly based on Bayes theorem). For instance, van der Lijn et al. [227] created a probabilistic atlas by averaging different topological atlases and then incorporated those probabilities as an energy term in a graph-cuts segmentation approach.

Fischl et al. [82] proposed segmenting brain structures using the Iterated Conditional Modes (ICM) algorithm [28] after defining the transformation between a probabilistic atlas and the subject. This optimisation algorithm starts with initial estimates for the parameters, called conditional modes, and proceeds to update the segmentation sequentially until the optimum is found. In this case, the volume is defined as a Markov Random Field (MRF) to explicitly include spatial information. The ICM is initialised with the maximum a posteriori (MAP) estimate of the segmentation. In a subsequent work, Han and Fischl [103] proposed applying a pre-processing step to this framework, consisting in atlas intensity normalisation for enhancing the performance on different scanning platforms.

In a similar way, segmentation and registration can be combined in a probabilistic frame-

Article	Real data							Test data (MRI system)
	CAU	THA	PUT	PAL	HIP	AMY	ACC	
Dawant (1999) [60]	0.86	-	-	-	-	-	-	8v (Siemens 1.5T)
Fischl (2002) [82]	0.88	0.79	0.71	0.71	0.81	0.79	-	14v (GE 1.5T) + 13v (Siemens 1.5T)
Klein (2005) ⁺ [123]	-	-	-	-	-	-	-	10v (Siemens 3T)
Klein (2005) ⁺ [124]	-	-	-	-	-	-	-	10v (Siemens 3T) + 10v (GE 1.5T)
Heckemann (2006) [107]	0.90	0.90	0.90	0.80	0.81	0.80	0.70	30v (GE)
Pohl (2006) [168]	-	0.894	-	-	-	-	-	22v (GE 1.5T)
Han (2007) [103]	0.84	0.88	0.85	0.76	0.83	0.75	-	14v (GE 1.5T) + 13v (Siemens 1.5T)
van der Lijn (2008) [227]	-	-	-	-	0.858	-	-	20v (Siemens 1.5T)
Bazin (2008) [25]	0.781	0.773	0.817	-	-	-	-	18v (IBSR)
Aljabar (2009) [4]	0.881	0.908	0.898	0.818	0.834	0.777	0.758	275v (*)
Artaechevarria (2009) [10]	0.83	0.88	0.87	0.81	0.75	0.72	0.68	18v (IBSR)
Ciofalo (2009) [42]	0.60	0.77	0.66	0.56	-	-	-	18v (IBSR)
Lötjönen (2010) [143]	0.866	0.899	0.905	0.844	0.819	0.767	-	18v (IBSR)

Table 3.2: Summary of brain structure segmentation results using the DSC metric. Additional information on the number of volumes (v) and the MRI system is also given. Acronyms in alphabetical order: Amygdala (AMY), Accumbens (ACC), Caudate (CAU), Hippocampus (HIP), Pallidum (PAL), Putamen (PUT), and Thalamus (THA). +Results by Klein et al. [123, 124] are reported for the brain as a whole rather than providing values for each structure. *Aljabar et al. [4] do not specify the MRI system used.

work via an EM algorithm [64] or the Fuzzy C-Means (FCM) [29] algorithm. The EM algorithm computes the probability of each entry of the data set belonging to a certain distribution. It then estimates the hidden parameters of this distribution which maximise the previous expectation in an iterative manner until convergence is reached. Even though great convergence properties are found, this algorithm can lead to non-desired local minima, especially when only relying on the data itself. Therefore, proper initialisation and spatial information are introduced into the framework using atlas-based approaches. For instance, Pohl et al. [168] proposed using the EM algorithm to find the optimum parameters for registration while labelling each voxel to a brain structure. On the other hand, FCM inherently treats each class as a Gaussian with the same variance, since it only takes class centroid and membership values for each voxel into account. Note that no spatial information is encoded in the original algorithm. However, Bazin and Pham [25] proposed modifying the objective function to include spatial constraints as well as probabilities from a statistical atlas. Segmentation masks were then constrained by a topological atlas to ensure that topology was preserved. Finally, growing and thinning techniques were used to refine the final delineations of the brain structures.

Regarding the experimental evaluation of these works, the Dice Similarity scores for the central nuclei (amygdala, accumbens, caudate, hippocampus, pallidum, putamen, and thalamus) are summarised in table 3.2. This table includes information about the data

sets used for validation. Even if some of the approaches [25, 42, 143, 10] use the same data set (IBSR), a quantitative comparison of all methods is difficult due to the variability of the testing data. Furthermore, the number of cases used for testing varies, ranging from 10 to 30 volumes, with the exception of Aljabar et al. [4], who used a larger database of 275 subjects to show the influence of atlas selection.

We would like to point out that the results obtained by Klein et al. [123, 124] are not included since they provide similarity values for the whole brain instead of structure-based values. However, in their last study, the authors provided a comparison of the similarity values when increasing the number of atlases, showing also an increase in the Dice coefficient. This behaviour can also be observed in table 3.2, where the best Dice values for each structure are obtained from multi-atlas approaches [4, 143, 107], closely followed by statistical frameworks [103, 82, 168]. Notice that Aljabar et al. [4] obtained the best results on some structures (the thalamus, the hippocampus, and the accumbens) by applying selection strategies in their multi-atlas approach. Furthermore, among the approaches using IBSR data, the method of Lötjönen et al. [143] outperformed the others for all structures. Note that this method is also based on a multi-atlas strategy.

Only few studies focus on evaluating how algorithms are affected by image variability within different scanning devices. For instance, Klein et al. [123, 124] and Han and Fischl [103, 82] validated their methods with MRI data acquired from different machines. This later study also demonstrated the importance of normalising the atlas and the subject intensities when using different scanning machines, obtaining higher *DSC* values after atlas normalisation when using data sets from two different MRI scanners. Almost all of the studies in table 3.2 segment images from 1.5T devices. Only Klein et al. [123, 124] applied their method to MRI images acquired at higher magnetic fields (3T). The recent multi-site and multi-scanner database maintained by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) [160] will allow researchers to test the robustness of developed algorithms on different scanning devices.

Segmentations for central nuclei structures have a *DSC* of between 0.70 and 0.90, showing good agreement between manual and automatic segmentation. The best segmentation results are obtained for the caudate, the thalamus, and the putamen structures, with values of 0.90, while the segmentations of the amygdala, the hippocampus, and the accumbens achieve values near 0.80 when using multi-atlas strategies.

Article	Atlas type		Registration transforms		Strategy
	Statistical	Topological	Global transforms	Local transforms	
Van Leemput (1999) [228]	ICBM	No	Affine	No	PA
Marroquin (2002) [145]	ICBM	No	Affine	Level set	PA
Cocosco (2003) [45]	53 manual	No	Affine	No	PA
Grau (2004) [96]	29 manual	No	No	Demons	PA
Ashburner (2005) [13]	ICBM	No	Affine	Cosine	PA
Awate (2006) [14]	ICBM	No	Affine	No	PA
Vrooman (2007) [233]	22 manual	No	Affine	B-Splines	PA
Bricq (2008) [36]	ICBM	No	Affine	B-splines	PA

Table 3.3: Classification of automated atlas-based segmentation methods for healthy tissues. Rows are the segmentation targets, while columns refer to: the type of atlas, the number of volumes used to build the atlas, the registration method, and the atlas strategy (Label Propagation (LP), Multi-Atlas Propagation (MA), and Probabilistic Atlas segmentation (PA)).

3.4.2 Brain tissues in healthy subjects

MRI has become a standard modality for brain tissue segmentation due to its high effectiveness in contrasting between tissue types. However, some image artefacts like the partial volume effect, image noise and intensity non-uniformities (also known as bias field [232]) can considerably increase the difficulty of segmentation work. In addition to these artefacts, the large differences that may exist between sulci and gyri patterns for various subjects are the main issues when dealing with brain tissue segmentation.

Numerous approaches have been proposed for MR brain tissue segmentation [164, 43, 210] and only some use atlas priors. Atlas-based approaches for brain tissue segmentation are mostly within a statistical framework and make use of probabilistic atlas priors (see section 3.3.3).

Van Leemput et al. [228] proposed accounting for neighbouring relationships between voxels through an MRF model. Distribution parameters were then estimated using the EM algorithm using bias field estimation to aid the segmentation task. Due to the need for initialisation, the authors decided to use atlas probabilities as a prior classification and to constrain the classification process at each iteration by multiplying the E-step probability by the one given by each atlas voxel. In a similar way, Bricq et al. [36] proposed including Markovian properties to account for voxel relationships by means of hidden Markov chains. After applying this theory to MRI, model parameters were also estimated with EM using atlas probabilities during initialisation and each subsequent iteration step. In order to increase the robustness, Marroquin et al. [145] proposed a novel variant of the EM algorithm which also relies on atlas information for initialisation. This algorithm substitutes the

expectation step by using the MAP estimator to compute the MRF parameters followed by the maximiser of posterior marginal estimate instead of the maximisation step. Therefore, using the atlas values for initialisation, this algorithm iterates until convergence is achieved.

While most segmentation approaches using Markov theory rely on a parametric estimation, non-parametric methods are quite powerful themselves, as pointed out by Awate et al. [14]. In their study, the Parzen-window technique is used to estimate a Markov probability density function (PDF) without imposing strong parametric models on the data. This data-driven technique, which is also initialised using atlas probabilities, provides the ability to model and learn arbitrary PDFs. However, corrupted volumes may bias this estimation and cause misclassification errors.

All the above methods rely on an initial registration step and then apply a segmentation framework. However, both steps are closely related, since atlas registration can benefit from an initial segmentation, while brain tissue segmentation requires prior information such as atlas probabilities, which are obtained from the atlas registration process. Hence, by combining both steps, Ashburner and Friston [13] proposed a novel method for segmenting brain tissues while correcting the bias field and refining atlas registration. In their approach, the ICM algorithm is used to estimate the final segmentation, while Gaussian Mixture Model parameters are estimated via EM using atlas probabilities and volume intensity. A Levenberg-Marquardt algorithm is then applied to correct the bias field and refine atlas alignment.

Probabilistic atlases can also be translated to topological ones. By applying a threshold to the probabilities, topological masks may be obtained for each tissue. Once registered to the subject volume, these masks can be applied in order to select a set of representative tissue samples from the subject. These are the samples that are then used to train a classifier. This is the main idea of the approach by Cocosco et al. [45] and Vrooman et al. [233], who extracted tissue samples using an atlas and then used the minimum spanning tree algorithm to reduce outliers. These samples were used to train a k-Nearest Neighbour (kNN) classifier. This algorithm uses the distance in feature space between the current voxel and each training sample to assign the label with most votes from among its k closest neighbours. Although this algorithm proposes a simple implementation, its major drawback is its high computational cost. Similarly, Grau et al. [96] applied the watershed transform to segment tissues. The simple and intuitive foundation of this algorithm consists in simulating the flooding of a region while considering image intensities as heights. However, one of the major drawbacks associated with this approach is that the obtained results may

Article	BrainWeb		Real data		
	GM	WM	GM	WM	Test data (Scanner)
Van Leemput (1999) [228]	0.93	0.92	0.836	0.821	9v (Sigma 1.5T)
Marroquin (2002) [145]	0.891	0.892	0.797	0.812	20v (*)
Cocosco (2003) ⁺ [45]	-	-	-	-	43v (*)
Grau (2004) [96]	0.890	0.946	-	-	-
Ashburner (2005) [13]	0.932	0.961	-	-	-
Awate (2006) [14]	0.92	0.96	0.807	0.887	18v (*)
Cocosco (2007) [233]	-	-	0.915	0.937	12v (*)
Bricq (2008) [36]	0.975	0.980	0.799	0.865	18v (*)

Table 3.4: Summary of healthy tissue segmentation results with DSC metric. Results are given for both real and BrainWeb images, where available. Additional information is also given on number of volumes (v) and MRI system. ⁺Results by Cocosco et al. [45] are reported using the kappa similarity measure.

be over-segmented. The authors proposed setting some initial markers, corresponding to each tissue class, to overcome this issue. Skeletons are calculated by thresholding a statistical atlas and then removing outliers; the resulting skeletons are considered as the initial markers.

The lack of a gold standard for brain tissue classes in real images makes the quantitative evaluation of segmentation algorithms difficult. The BrainWeb phantom [47] provides a standard platform for comparing healthy brain approaches, but results on synthetic phantoms cannot be extrapolated to real conditions. Table 3.4 summarises the results for WM and GM on both synthetic and real data¹, including also information on data acquisition. The number of volumes used to test the methods ranges between 10 and 40, which does not significantly differ from the reviewed segmentation methods for brain structures.

Contrary to the methods presented in section 3.4.1 for deep brain structure segmentation, no tissue segmentation methods rely on label or multi-atlas propagation. Moreover, most brain tissue segmentation uses statistical atlases instead of topological ones. Tissue segmentation strategies are often based on probabilistic atlas segmentation. Furthermore, parametric estimation algorithms are most widely used, with *DSC* values over 0.90 on the BrainWeb phantom. For instance, values of 0.975 for GM and 0.980 for WM (both close to perfect agreement) are reported by Bricq et al. [36], who used the EM algorithm with atlas probabilities for initialisation and then include prior probabilities at each step.

Even though significant agreement is found for synthetic data, *DSC* values decrease when these methods are applied to real cases. Awate et al. [14], who obtained worse results than Bricq et al. [36] when evaluating the BrainWeb phantom, obtained better results when

¹CSF is not included since some methods do not include it in their segmentation.

Article	Atlas type		Registration transforms		Strategy
	Statistical	Topological	Global transforms	Local transforms	
Mortamet (2005) [157]	ICBM	No	Affine	No	PA
Prastawa (2005) [175]	3 manual	No	Affine	No	PA
Weisenfeld (2006) [238]	13 manual	No	Affine	No	PA
Xue (2007) [246]	No	3 manual	Affine	No	LP
Murgasova (2007) [158]	1 manual (37)	No	Affine	B-Splines	LP,PA
Smith (2007) [202]	141 manual	No	Affine	Cosine	PA
Habas (2008,2010) [100, 101]	14 manual	No	Affine	Elastic	PA
Weisenfeld (2009) [239]	15 manual	No	Affine	No	PA

Table 3.5: Classification of automated atlas-based segmentation methods for challenging populations. Rows are the segmentation targets, while columns refer to: the type of atlas, the number of volumes used to build the atlas, the registration method, and the atlas strategy (Label Propagation (LP), Multi-Atlas Propagation (MA), and Probabilistic Atlas segmentation (PA)).

testing with real data. These prove the need to use real data sets in conjunction with synthetic phantoms to evaluate and compare segmentation methods.

3.4.3 Brain tissues in challenging populations (fetus, neonates, and elderly subjects)

There are specific populations such as fetuses and newborns that are particularly challenging for atlas-based segmentation methods. Automated segmentation of these populations is a key tool for brain development studies. Their specific characteristics (such as spatial and temporal variations of the image contrast due to myelination and folding of the growing brain, and low signal to contrast ratio and high partial volume effects due to fast acquisition sequences to avoid motion) represent new challenges for brain tissue segmentation when compared to adult brain segmentation. Consequently, on the one hand, anatomical priors are needed to reduce the complexity of segmenting fetus and neonate brains, while on the other hand constructing these atlases is difficult due to the constantly evolving anatomy. Figures 3.5a and 3.5b show intra-utero fetal and newborn brains, respectively.

Pioneer studies on these populations were carried out on newborns and pre-term infants. Some groups proposed supervised classification methods [235, 238, 239, 175] using both MR signal (multi-spectral contrast in some cases) and spatial priors. Probabilistic atlases are either used to train the classifier [239] or to be included as features for classification [175]. Bayesian frameworks (see equation 3.5) for neonatal segmentation were suggested in [246]. Label propagation was used prior to the non-supervised statistical tissue segmentation to mask some structures. A label propagation approach [158] has also been presented to

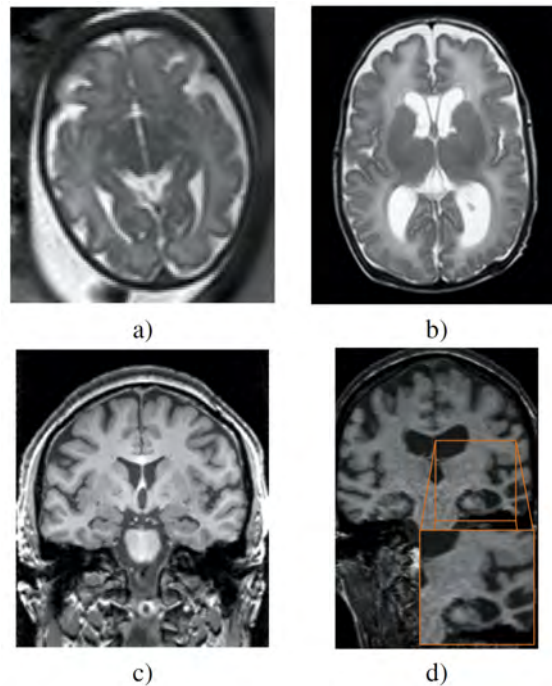


Figure 3.5: a) Fetus at 31 weeks of gestational age, b) premature newborn, c) 61 year-old woman and d) 66 year-old woman with dementia.

create a population-specific atlas for young children.

Very few studies exist related to the automated segmentation of fetal brain tissues [44, 100, 18, 101]. Indeed, only one research group has recently proposed a spatio-temporal probabilistic atlas and an atlas-based segmentation of developing brain tissues in young fetuses [100, 101].

Elderly subjects (see figure 3.5c) are another challenging population due to the loss of tissue volume related to ageing [157, 202, 219], also known as normal atrophy [68]. Still, the same registration methods are used as for young subjects. In this context, special attention needs to be paid to atlas selection and in cases where younger atlases are not representative of the elderly anatomy age and sex-specific atlases [136] can be used. It should be noted that the aging effect is also present in brains containing abnormal regions.

	Article	Atlas type		Registration transforms		Strategy
		Statistical	Topological	Global transforms	Local transforms	
Focal lesions	Kamber (1995) [119]	12 manual	No	Rigid	No	LP, PA
	Van Leemput (2001) [135]	ICBM	No	Affine	No	PA
	Zijdenbos (2002) [251]	53 manual	No	Rigid	No	PA
	Wu (2006) [244]	Yes	No	Affine	B-splines	LP
	Bricq (2008) [35]	31 manual	No	Affine	B-splines	PA
	Kroon (2008) [125]	ICBM	No	Affine	B-splines	PA
	Prastawa (2008) [174]	ICBM	No	Affine	No	PA
	Shiee (2008) [198, 197]	18 manual	1 manual	Rigid	No	PA
	Souplet (2008) [204]	ICBM	No	Affine	No	PA
	Akselrod-Ballin (2009) [2]	ICBM	No	Affine	Cosine	PA
	de Boer (2009) [61]	12 manual	No	Affine	B-splines	PA
	Tomas (2009) [220]	15 manual	15 manual	No	B-splines	LP, PA
	Shiee (2010) [199]	18 manual	1 manual	Rigid	No	PA
	Geremia (2011) [92]	ICBM	No	Affine	No	PA
Schmidt (2012) [187]	ICBM	No	Affine	Cosine	PA	
Space-occupying lesions	Kyriakou (1999) [128]	No	1 manual	Affine	Elastic	LP
	Dawant (1999) [58]	No	1 manual	Affine	Demons	LP
	Warfield (2000) [236]	No	1 manual	No	Elastic	LP
	Moon (2002) [155]	ICBM	No	Affine	No	PA
	Shen (2002) [194]	1 manual	1 manual	Affine	Hierarchical	LP
	Bach Cuadra (2004) [17]	No	SPL	Affine	Demons	LP
	Duay (2004) [70]	No	1 manual	No	Elastic	LP
	Liu (2004) [140]	No	1 manual	Affine	Hierarchical	LP
	Prastawa (2004) [173]	ICBM	No	Affine	No	PA
	Stefanescu (2004) [205]	No	1 manual	Rigid	Grid-based	LP
	Pollo (2005) [169]	No	SPL	Affine	B-Demons	LP
	Nowinski (2005) [161]	No	1 manual	No	Non-linear	LP
	Bach Cuadra (2006) [16]	No	SPL	Affine	Optical flow	LP
	Zacharaki (2008) [248, 247]	No	1 manual	Affine	Elastic	LP

Table 3.6: Classification of automated atlas-based segmentation methods for brains with lesions. Rows are the segmentation targets, while columns refer to: the type of atlas, the number of volumes used to build the atlas, the registration method, and the atlas strategy (Label Propagation (LP), Multi-Atlas Propagation (MA), and Probabilistic Atlas segmentation (PA)).

3.4.4 Brain tissues and lesions in pathological brains (MS, Alzheimer, tumors, etc.)

Abnormal atrophy (see figure 3.5d) is a common feature among neuro-degenerative brain disorders such as mild cognitive impairment, Alzheimer’s disease, or schizophrenia. This pathological tissue loss ² is not present on healthy atlases and is thus usually overlooked under the assumption that the registration step will capture it. As for healthy elderly subjects, the strategies presented in the previous sections are also used on these cases [107, 105, 63, 240, 241, 250]. Disease-specific atlases could be used to improve segmentation results [148, 74].

Damaged brains may contain more than subtle brain tissue loss, such as, for instance,

²Note that the tissue loss due to aging is present as well.

focal tissue lesions or large space-occupying lesions. Obviously, atlases do not contain such damaged areas as they vary greatly in size, shape, and location. Therefore, new strategies or extensions to existing methods are needed to deal with these pathological cases. Below, we distinguish between two kind of brain lesions. First, focal tissue lesions, which represent those produced by the loss and inflammation of tissue, as in strokes and multiple sclerosis. In these cases, registration error due to lesion areas not present in the atlas are neglected. Second, space-occupying lesions, like tumours, which induce a deformation on the patient's brain anatomy and where the deformation field caused by the lesion needs to be estimated.

Focal tissue lesions

At first glance, table 3.6, presents again the atlas-based papers discussed in chapter 2 and listed in table 2.1. However, here we present a brief summary of those methods from the atlas perspective and taking into account the concepts introduced in this chapter.

Segmentation methods that deal with focal tissue lesions usually rely on the use of healthy brain atlases to segment brain tissues and consider lesions as *outliers*. Figure 3.6 illustrates some examples of focal tissue lesions.

In most cases, following detection of the outliers a subsequent analysis is usually performed to ensure that outlier regions are actually lesions. These methods can be seen as an extension of healthy tissue and structure approaches that use a statistical frameworks, introducing a new class into the segmentation algorithm. For instance, Seghier et al. [189] extended the tissue segmentation approach of Ashburner and Friston [13] to detect stroke lesions, while Schmidt et al. [187] also used the SPM toolbox to create an outlier map to detect and segment MS lesions. On the other hand, Van Leemput et al. [135] extended their previous work [228] to detect multiple sclerosis lesions by searching for outliers that follow a set of user-defined rules, i.e. lesions should appear as hyper-intense on both PD-w and T2-w images. In a similar way, Bricq et al. [35] applied their hidden Markov chain approach [36] to detect multiple sclerosis lesions as outliers, while Shiee et al. [199, 198, 197] modified their fuzzy segmentation algorithm [25] to segment lesions inside WM tissue.

Similarly, Souplet et al. [204] extended a previous study by Dugas-Phocion et al. [72] based on segmenting tissues with an EM approach by including pure tissue and partial volume classes. Using pure tissue masks, normal appearing tissue parameters (mean and standard deviation) were estimated on the T2-FLAIR image to define a lesion threshold.

Furthermore, Shen et al. [195, 196] proposed to automatically detect stroke lesions by

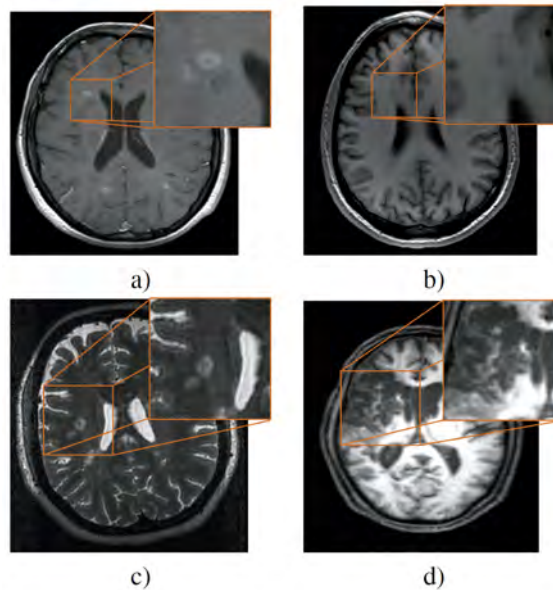


Figure 3.6: Focal tissue lesions. a) Enhancing MS lesions, b) black holes in MS, c) hyper-intense MS lesions, and d) Medial Cerebral Artery ischaemia.

comparing voxel-to-voxel the inconsistency between the result of applying an unsupervised tissue segmentation of the patient scans and the probability priors obtained within an atlas. Those regions where the inconsistency is large are assumed to be part of the lesion.

The unsupervised methods mentioned above can be biased by unhealthy tissues. Intuitively, supervised methods relying on tissue samples from lesion and non-lesion classes should perform better than unsupervised methods, since no intensity distribution models are assumed. For instance, Kamber et al. [119] compared three different classifiers. In their study, a probabilistic atlas was used to provide the classifiers with features as well as to constrain the search within the WM areas. Following the same idea, Wu et al. [244] implemented a kNN classifier trained on 20 voxels for each class. Following classification, a probabilistic atlas was used to relabel GM and multiple sclerosis lesions taking WM regions into account.

On the other hand, Zijdenbos et al. [251] used a probabilistic atlas as features to train an artificial neural network classifier. The input features included three MRI modalities and three spatial tissue priors from the atlas. Similarly, Geremia et al. [92] used a random forests approach where atlas probabilities were used as features and to create contextual meta-features that were used during the training. Furthermore, healthy atlases could be extended, as proposed by Kroon et al. [125]. In this study, segmented lesions were manually

warped to the ICBM atlas (from SPM) and this was used as a feature of a PCA-based classification framework which was trained with lesion and non-lesion samples. Developing a more general framework, Akselrod-Ballin et al. [2] proposed segmenting the volume in different regions using a graph-based algorithm. These regions were then characterised with a rich set of extracted features (comprising probabilities taken from an atlas) and classified using a decision forest along with the Fisher linear discriminant. This combination of segmentation and region classification helped to reduce misclassification at voxel level due to noisy data.

Atlases can also be used to sample healthy-looking tissue voxels. For instance, de Boer et al. [61] extended the healthy tissue classification from Cocosco et al. [45] to deal with white matter lesion segmentation. Similarly to Souplet et al. [204], after tissue segmentation a histogram of all GM voxels in T2-FLAIR is created and a threshold is defined to segment the lesions.

Atlases also provide a way to select abnormal tissue samples while estimating healthy ones. For example, Prastawa et al. [173, 174] proposed using the Minimum Covariance Determinant to estimate tissue PDFs using healthy samples. Outliers to this estimation are then considered as abnormal tissue.

Finally, a combination of training sample points and WM area refinement for multiple sclerosis segmentation was presented by Tomas and Warfield [220]. This approach used a set of topological atlases to define healthy tissue samples, obtained using the STAPLE algorithm [237], and to create a probabilistic atlas from the average of these manual segmentations. Multiple sclerosis samples were then defined as intensity outliers by comparing the reference group and the subject volumes. Subsequently, a Bayes classifier was trained to select lesion and non-lesion voxels. Since some of these voxels were misclassified as false positives, this classification was refined using WM regions extracted from the probabilistic atlas.

Different lesion sizes and locations make the comparison of segmentation methods even more challenging than the evaluation for segmenting healthy brain tissue and structures. Furthermore, evaluation measures would differ for different lesion sizes and types (multiple sclerosis, strokes or tumours). Public data sets for lesions along with ground truth segmentations are rare. As far as we know, only the BrainWeb site [47] provides a public synthetic phantom to validate multiple sclerosis lesions, and data for only 20 subjects from the training set of the 2008 Multiple Sclerosis Challenge are available.

Results with both synthetic and real data for white matter lesions are summarised in

Article	BrainWeb	Real data	
	Abnormal	Abnormal	Test data (Scanner)
Kamber (1995) ⁺ [119]	-	-	12v (Philips 1.5T)
Van Leemput (2001) [135]	-	0.49	23v (Philips 1.5T)
Zijdenbos (2002) [251]	-	0.60	29v (*)
Prastawa (2004) [173]	-	0.854	3v (*)
Wu (2006) [244]	-	0.6	12v (Siemens 1.5T)
Shiee (2008) [198]	0.677	0.531	10v (Philips 3T)
Akselrod-Ballin (2009) [2]	-	0.53, 0.55	25v (*) + 16v (*)
de Boer (2009) [61]	-	0.72	209v (GE 1.5T)
Tomas (2009) [220]	-	-	9v (GE 3T)
Shiee (2010) [199]	0.789	0.633	10v (Philips 3T)

Table 3.7: Evaluation of reviewed methods for abnormal tissue segmentation. DSC is presented for both real and BrainWeb simulated data, where available. Additional information is also given on number of volumes (v) and MRI acquisition devices. Studies presented in the MS Grand Challenge 2008 are not reported here since different evaluation measures were used and only off-site results are provided for all of them. +Results by Kamber et al. [119] are reported using error measures.

table 3.7, including information on data acquisition. The number of tested subjects ranges between 10 and 30 for these studies, with the exception of de Boer et al. [61], who used 209 volumes to validate their method. Similarly to methods used for structure and healthy brain tissue segmentation, abnormal tissue segmentation approaches are validated using a small number of cases. Validation with larger data sets would be desirable in order to assess their usability in clinical practice.

As with approaches used for healthy tissue segmentation, brain tissues and lesions are mostly segmented using a probabilistic atlas segmentation framework. This is because most lesion segmentation methods are extensions of approaches used for healthy tissue segmentation. Note also that white matter regions are often used to confine lesions to a region of interest in order to reduce false positives.

Finally, we would like to briefly recall here the quantitative results published during the 2008 Multiple Sclerosis Challenge³ [208]. A total of 53 volumes were separately acquired by the Children’s Hospital Boston (CHB) and the University of North Carolina (UNC) for this competition: 20 volumes for training, 25 public volumes before the contest for validation and 8 volumes for on-site testing. Five of the nine methodologies presented during the challenge used an atlas-based strategy [35, 204, 197, 125, 174] with the best results of the competition obtained by Souplet et al. [204], using an EM algorithm initialised with atlas probabilities.

³All this information can be found at: <http://grand-challenge2008.bigr.nl/> and the open journal www.midasjournal.org/browse/publication/638

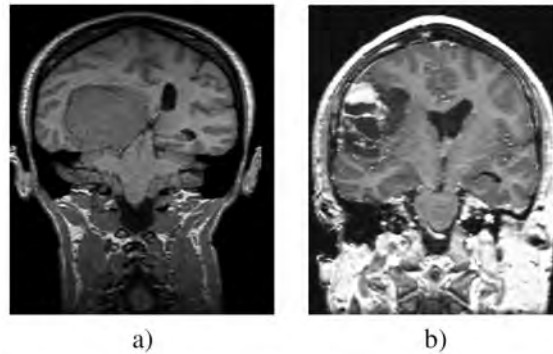


Figure 3.7: Space-occupying lesions: a) meningioma, and b) astrocytoma.

Space-occupying lesions

Examples of space-occupying lesions are shown in figure 3.7. These lesions induce large deformations and lack clear anatomical detail due to infiltration and edema [167], making the registration of diseased brains with normal atlases difficult. When space-occupying lesions are present, registration methods aim to capture not only the anatomical variability but also the deformations induced by the tumour. Thus the assumption of small and smooth deformations is clearly violated [16, 247].

The original works of applying atlas-based segmentation to the presence of tumours were by Kyriakou and Davatzikos [128] and Dawant et al. [58]. The aim of both approaches was to estimate how the presence of the lesion affected brain tissues and structures. The approach of Kyriakou and Davatzikos [128] modelled the soft tissue deformations induced by the tumour using a finite-element method, and subsequently, they registered the topological atlas with a transformed patient image from which the tumour was removed. As for Dawant et al. [58, 59], the patient (including the tumour) was registered with a seeded version of the atlas that included a region with the same intensity properties as the tumour, which was manually segmented previously. Bach Cuadra et al. [17] and Pollo et al. [169] improved this approach by avoiding registration of the full volumes, assuming a radial tumour growth from a single voxel seed. A similar assumption was taken into account by Nowinski and Belov [161], who proposed a non-linear tumour deformation after registration of the patient volume and the atlas using Talairach registration. All these methods required a precise pre-segmentation of the tumour, usually performed using semi-automatic algorithms [236]. More sophisticated models of lesion growth are proposed by the authors of the HAMMER [194] and the ORBIT [248, 247] frameworks.

Some other methods [70, 205, 140], instead of having a lesion growth model, apply

different rigidity constraints to the tumour area. For instance, Duay et al. [70] locally adapt the elasticity of the transformation, hence allowing large deformations around the tumour. Pre-segmentation of the tumour was not necessary. Stefanescu et al. [205] introduce specific rigidity parameters for the tumour, and Liu et al. [140] assume local rigidity by means of a Markov Random field-maximum a posteriori approach. However, both approaches again require a priori segmentation of the lesion.

Moon et al. [155] extended Van Leemput et al.'s [228] tissue segmentation approach for detecting brain tumours. The authors used the same EM approach but extended the number of classes with a tumour class. The prior spatial probabilities of the tumour location were introduced into the algorithm by multiplying the atlas probabilities by the difference image of the pre and post-contrast images. Prastawa et al. [173], prior to detecting MS lesions as seen in section 3.4.4, also proposed detecting tumours and edemas by using the Minimum Covariance Determinant to estimate tissue PDFs using healthy samples, and determining the diseased regions as the resulting outliers of the model.

Results for tumours [173] (*DSC* of 0.854 using only 3 cases) and general white matter lesions [61] (*DSC* of 0.72 using a large data set of 209 cases) show a good agreement in real data. Unfortunately, in the specific case of WM lesions corresponding to MS lesions, segmentation results show lower *DSC* values, like, for instance, the 0.633 obtained by Shiee et al. [199] using a statistical framework based on the fuzzy clustering algorithm. This indicates that there is a need for further development of MS lesion segmentation.

3.5 Discussion: How can we use an atlas to segment MS lesions?

As defined in chapter 1, MS lesions are abnormalities in the brain tissues. In our case, this lesions are WM abnormalities due to the demyelination of the neuronal axons. Clinically and statistically, MS lesions can be considered as tissue outliers. In fact, as we have seen in section 3.4.4, lesions behave as outliers in terms of image intensities and can be segmented using normal appearing tissue information to discard them.

Therefore, an atlas becomes a powerful tool for MS lesion segmentation within probabilistic frameworks. To be precise, the use of probabilistic atlases and outlier rules (i.e. a certain distance to a tissue distribution) are the most common strategies when dealing with this type of lesions (see table 3.6). As pointed out in section 3.4.4, focal lesions are small enough and rarely cause edema. Thus, when compared to the deformation intro-

duced by a large tumour, small misregistration errors can be neglected or treated as part of the probabilistic segmentation.

Another important aspect when detecting and segmenting lesions as outliers is the definition of the tissue model. Once again, looking at table 3.3 the most common approach is to use a probabilistic atlas within a probabilistic framework when dealing with healthy patients. As pointed out in section 3.4.4, most of these strategies can be extended to include MS lesions as outliers.

While probabilistic atlases can aid and guide tissue and lesion segmentation, topological atlases that delimit structure boundaries can also be used to impose certain spatial constraints to the segmentation as proposed by Shiee et al. [199]. Such constraints can be used to detect common lesions appearing around certain structures which can later be used to diagnose a MS patient according to the criteria described in chapter 1. For instance, periventricular lesions could be detected and characterised thanks to the segmentation of the ventricles.

To summarise, probabilistic frameworks are the most common approach to segment MR images of MS patients. On these cases, the importance of the atlas is twofold. On the one hand, tissue distributions and probabilities can be estimated. On the other hand, outliers of these distributions can be also estimated using a probabilistic atlas. Finally, label propagation and topological atlases can also be used to impose certain spatial constraints on the segmented lesions.

3.6 Conclusions

In this chapter we have presented the basics of atlas-based segmentation and we have reviewed its application in brain MRI. First, we described an atlas as a set of images including at least an intensity image considered a template and a segmented image and we presented a summary of the atlases' history in brain imaging [76]. We also differentiated between *topological atlases*, having only a labelled image, and *probabilistic atlases*, including a set of probabilistic maps.

Afterwards, we presented three different approaches to atlas-based segmentation: *label propagation*, where a single labelled image is used for segmentation after registration, *multi-atlas propagation*, where a set of labelled images are combined in the image space to obtain a segmentation, and *probabilistic atlas segmentation*, where probabilistic atlases are used as part of a probabilistic or variational framework.

Finally, we presented several works involving atlas-based segmentation divided into four categories according to their medical target: brain structures presenting well-defined shapes, brain tissues in healthy subjects, brain tissues in challenging populations and lesions damaged brains including the MS papers from chapter 2. For each category, we also presented the reported results for each strategy and a discussion.

MS LESION SEGMENTATION PROPOSALS

Sometimes your mistakes can surprise you.

My biggest mistake for instance, brought me here.

Harold Finch, “Person of Interest (Season two)” (2013)

4.1 Overview

In order to develop our proposal, we presented the state-of-the-art on MS lesion segmentation in chapter 2. After a thorough analysis, we concluded that supervised approaches are preferred to unsupervised ones thanks to the strength of using prior knowledge. Moreover, one of the most common techniques to introduce this prior knowledge was to use a so-called atlas. Therefore, in chapter 3, we introduced the atlas concept and defined various strategies to use them for segmentation. Furthermore, we presented all the possible clinical targets of the brain in order to point out strengths and weaknesses that could be applied to MS lesion segmentation.

In this chapter, we commence with a presentation of our preprocessing pipeline, with several steps to enhance our images and prepare the atlas, followed by two different atlas-based approaches: the first based on tissue segmentation, and the second fully supervised and based on training from manual expert annotations.

4.2 Preprocessing

In chapter 1 we explained how difficult it is to achieve a robust and automatic brain MRI segmentation due to numerous facts: variable imaging parameters, overlapping intensities, noise, partial volume effect, gradients, motion, echoes, blurred edges, normal anatomical variations, and susceptibility artefacts [191]. Therefore, before applying any approach to MS lesion segmentation, there are generally two preprocessing steps to be carried out: first, the minimisation of those image artefacts and second, the removal of nonbrain tissue, such as the skull, from the image. Other optional preprocessing steps such as the equalisation of soft brain tissues or registration between different MR images may be applied. In this section, these steps are briefly described.

As stated, the capture process itself corrupts MR images with various artefacts. This can lead to inaccurate segmentation. However, from the image processing viewpoint, it is common to simplify all these problems [201] by defining the observed intensities, \hat{I} , of each voxel x as:

$$\hat{I}(x) = \beta I(x) + \varepsilon \quad (4.1)$$

where $I(x)$ is the real intensity, β is a multiplicative smooth bias field that causes intensity inhomogeneities due to the sensitivity of the reception coil and ε represents additive noise. While it is often assumed that ε follows a Gaussian distribution, there have been different approaches to correct the bias field β . This is an important issue since voxels belonging to different tissues may be assigned with the same grey-level value when varying this term. Two different studies have reviewed different ways to overcome this preprocessing step problem [112, 232]. Both studies classify these methods into various groups: segmentation-based, filtering-based, surface fitting-based, histogram-based, and other specific techniques. However, as pointed out by Hou [112], none of these methods has been shown to be superior to the others and exclusively applicable.

Skull stripping is another important preprocessing step since fat, the skull, skin, and other nonbrain tissues may cause misclassifications in some approaches due to the intensity similarities with brain structures [53]. This is also the case of brain sub-cortical structures, where advanced analysis is not readily applicable since the intensity characteristics that overlap between different structures may reduce the reliability of automated segmentation methods. Two works have analysed and compared the state-of-the-art methods to extract brain regions from MR images. The first study, by Boesen et al. [31], compared

four systems: Statistical Parametric Mapping (SPM2) [12], the Brain Extraction Tool (BET) [203], the Brain Surface Extractor (BSE) [193], and their own Minneapolis Consensus Strip (McStrip) [176]. They validate these systems with three data sets of T1-w images. The second study, by Hartley et al. [104], compared only the accuracy of BET and BSE against 296 PD-w images. In both studies, manual segmentations were used as the “gold standard”. Boesen et al. [31] concluded that the McStrip, a hybrid algorithm incorporating intensity thresholding, nonlinear warping, and edge detection, consistently outperformed SPM2, BET, and BSE. Furthermore, other studies [117, 244] also suggest performing this region removal task before applying the inhomogeneity correction. In this way, the correction would be carried out only on those voxels belonging to the internal brain tissues.

As already mentioned, other preprocessing steps can be applied to improve the classification of tissues and the subsequent lesion segmentation. Examples of such preprocessing pipeline approaches can be found in studies by Zijdenbos et al. [251] and Hou and Huang [113]. The steps in their pipeline approach include registration between different MR images [23], intensity normalisation [192], and the transformation to a standard space to facilitate atlas-based segmentation (as presented in chapter 3). For example, the Talairach coordinate system is used to describe the location of brain structures that are independent from individual differences in the size and overall shape of the brain [214]. Some approaches also use intra-sequence and inter-sequence preprocessing registration steps. Intra-sequence registration enables the correction of the misregistration between the acquisition steps, since standard multi-slice acquisition sequences are acquired in multiple, interleaved passes (this step is only used when applicable to the actual acquisition sequence, such as a multi-slice dual-echo). On the other hand, the inter-sequence registration can help compensate for possible (and likely) patient motion between scans.

Ultimately, as in many image processing applications, the choice of all these preprocessing steps is a trade-off between algorithmic complexity (i.e. the number of tunable parameters), flexibility, and processing speed (although nowadays this is not a critical issue since a few hours of processing time for a large series of images is acceptable). Moreover, this selection may also depend on the demands made of the developed pipeline or on user-defined criteria in terms of accuracy, reproducibility, and consistency over time or across data sets acquired on different scanners. For our segmentation methods, we decided to follow the preprocessing pipeline presented in figure 4.1: firstly, the skull is removed; secondly, the noise and intensity inhomogeneities are reduced; afterwards, an atlas is registered to the corrected images; and, finally, the images are normalised for all patients.

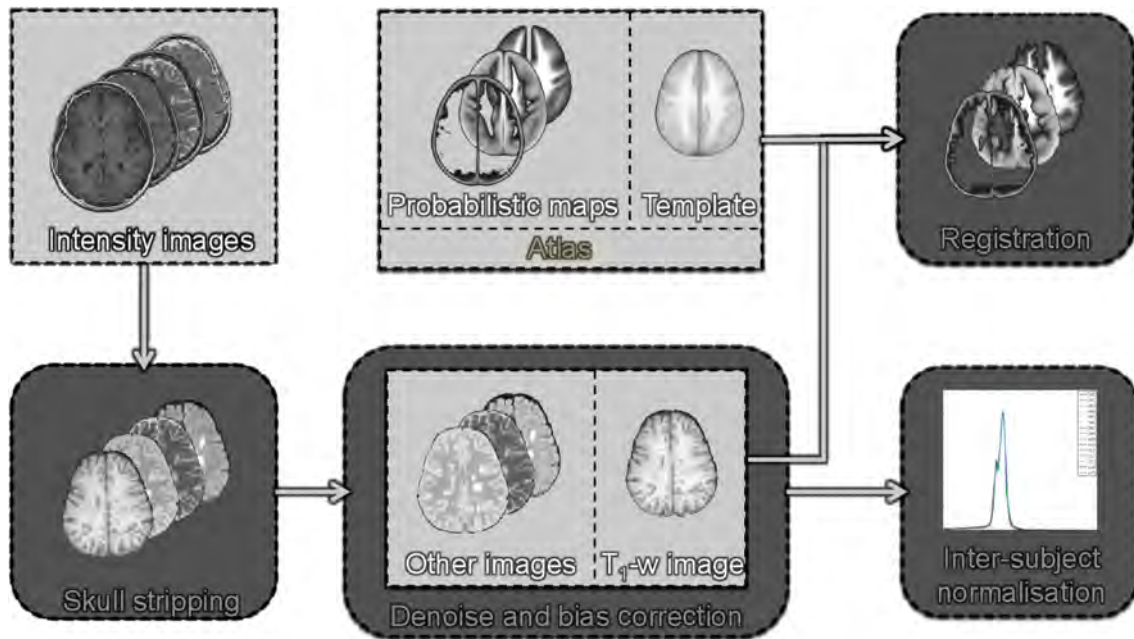


Figure 4.1: Flowchart of the preprocessing steps used for both segmentation methods.

4.2.1 Skull stripping

As pointed out before, Boesen et al. [31] concluded that their novel McStrip method [176] outperformed the others. This algorithm incorporates atlas-based segmentation with deformable registration, intensity thresholding with connectivity constraints, and edge detection with morphological operations. However, this method is not publicly available.

BET, a tool from the FSL toolbox [116], is a freely available algorithm based on T1-w images and is widely used in brain MRI processing. While the voxel misclassification is fairly superior when compared to the McStrip [31], the differences in accuracy are not significant when compared with other freely available tools such as BSE or SPM [104]. Therefore, we decided to use this tool to perform the skull stripping.

The BET algorithm uses a deformable model that evolves to fit the brain's surface by the application of a set of locally adaptive model forces. First, robust higher and lower intensity boundaries are estimated from the intensity histogram to obtain a rough threshold between the brain and other tissues. With this initial mask, the centre-of-gravity is found along with a rough size of the head. Afterwards, the initial deformable model is created inside the brain using a tessellation of a sphere's surface. Finally, the model is allowed to slowly deform following forces that preserve smoothness and keep the surface well-spaced while, at the same time, pushing to the brain's edges. This process is repeated until a

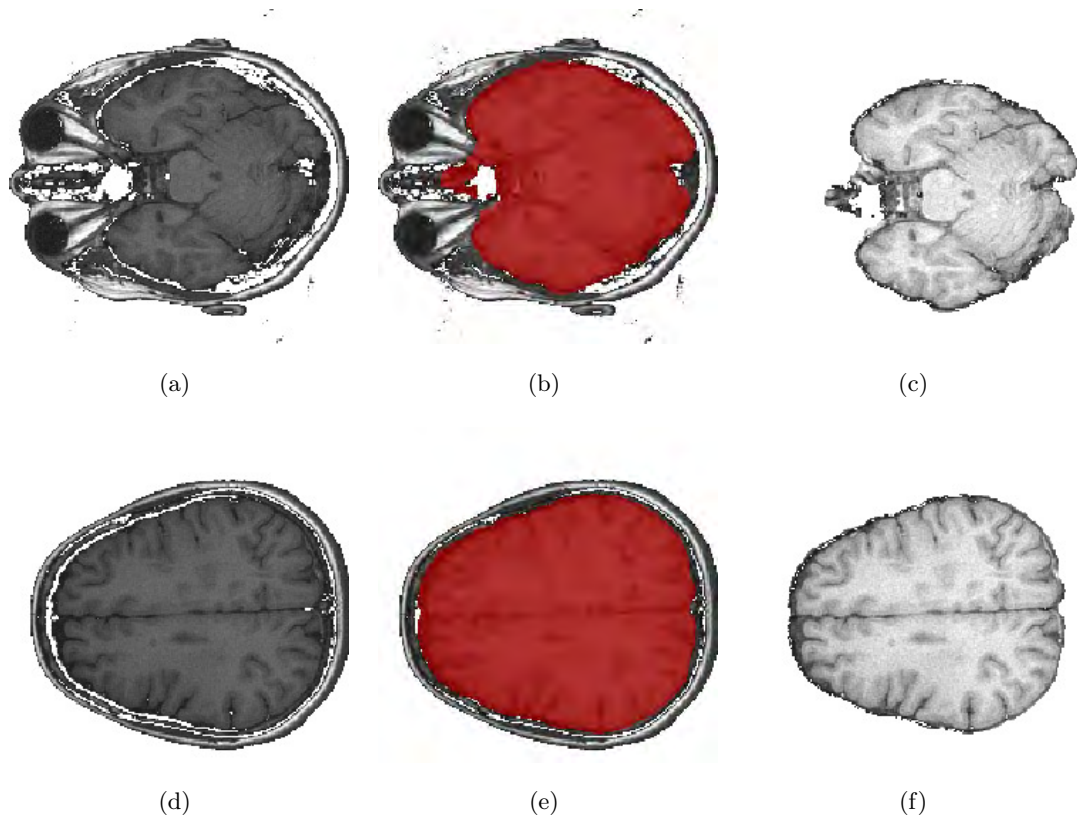


Figure 4.2: Skull stripping example: a,d) Original slices; b,e) the same slices with a superimposed mask obtained using BET, and c,f) the same slices after removing the nonbrain tissues.

suitable solution is obtained increasing the smoothness constraint. An example of skull stripping is shown in figure 4.2.

The BET's main parameter is the fractional intensity threshold. This parameter, which is by default 0.5, is used to compute the force component responsible for fitting the surface model to the real brain surface interacting with the image. Furthermore, BET also offers different options to interact with. For instance, there is an option to do a cleanup for the eyes and the neck, as well as options to do a robust brain centre estimation or to include a T2-w image.

In our preprocessing pipeline, we decided to apply the BET algorithm to the T1-w images.

4.2.2 Bias correction and noise reduction

An intensity in which where the observed intensities from the images are corrupted by an additive noise and a multiplicative bias field is defined by equation 4.1. Other models used in the literature are described by Vovk et al. [232]. However, this simplified model with additive noise and multiplicative bias is the most widely used in brain MRI.

Following this model, the first step to recover the real intensity values would be to eliminate, or reduce, the noise signal since the bias field is assumed to affect the original intensity values independently of the noise. Therefore, after the skull stripping, we perform a noise reduction step based on anisotropic diffusion [166].

Anisotropic diffusion algorithms are formulated in order to preserve specific image features while also reducing noise. For instance, light-dark transitions defining the edges between tissues are important and thus should be kept. Standard isotropic diffusion methods move and blur those boundaries causing artifacts similar to those observed as partial volume effects. In contrast, anisotropic diffusion methods are formulated to specifically preserve those edges.

In our proposal we use an anisotropic diffusion filter that tends to preserve edges over smoother regions according to the gradient magnitude at each point. This filter is publicly available as part of the Insight toolkit (ITK) software library and is implemented as an N-dimensional version of the classic Perona and Malik anisotropic diffusion equation for scalar images [166].

After noise reduction, the only issue from equation 4.1 that remains is the multiplicative bias field. Amongst the various bias correction algorithms that deal with this smooth multiplicative field, the well-known nonparametric nonuniform normalisation (N3) [201] has become a de facto standard for a variety of imaging acquisition strategies. The original method iteratively seeks a field, modeled using B-splines, that maximises the high frequency content of the distribution of tissue intensity. This algorithm is fully automatic and requires no prior knowledge contributing to its high popularity within the community.

Recently, Tustison et al [224] proposed two improvements to this method and a new implementation rebranded Nick's N3 (N4) based on the ITK library. These two major changes affect the B-splines modeling and the iterative optimisation scheme. The new B-spline approximation allows for smaller control point spacing, obviates the need of a regularisation parameter, and permits the specification of a weighted mask that could be used in a segmentation framework. Moreover, it allows for a multi-scale optimisation

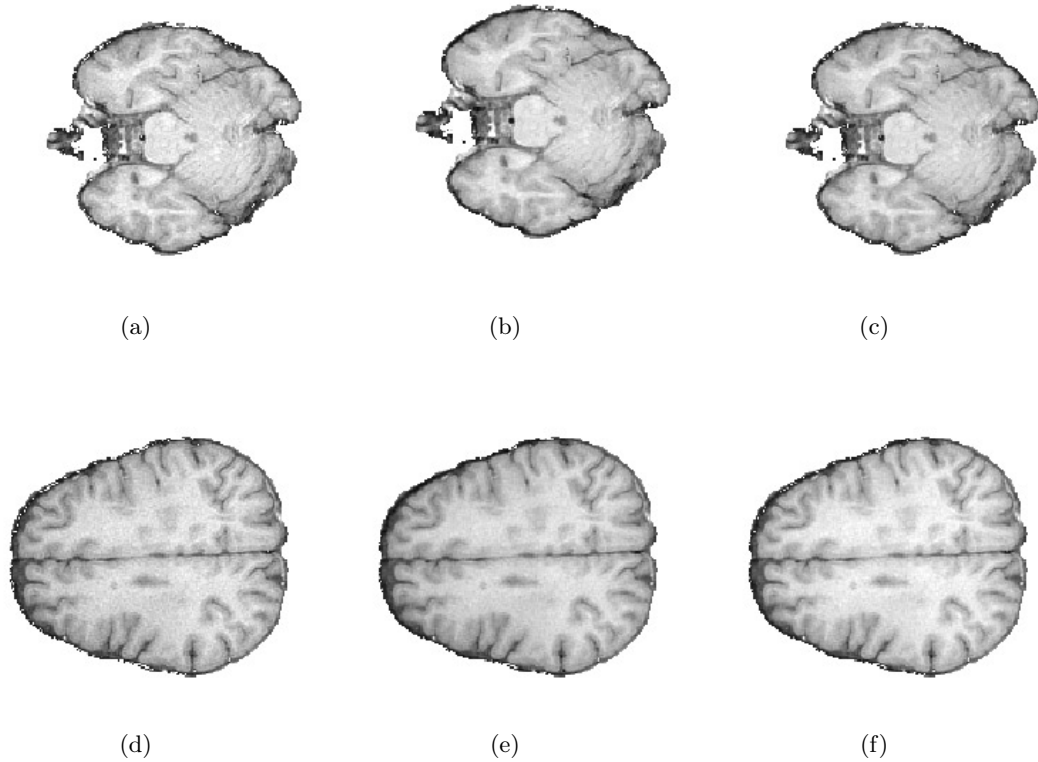


Figure 4.3: Denoising and bias correction example: a,d) Slices after skull stripping; b,e) the same slices after denoising, and c,f) the same slices after applying bias correction.

approach that reduces time and improves convergence.

Therefore, we decided to use the N4 bias correction filter provided as part of the ITK library, combined with the anisotropic diffusion filter to obtain the original intensity image $I(x)$. An example of this procedure is shown in figure 4.3.

4.2.3 Inter-subject normalisation

After dealing with the major artifacts and issues of brain MRI presented in chapter 1, no further preprocessing steps would be required when processing each image independently. However, if a supervised approach based on training is used, another major issue should be tackled. We are referring to inter-subject normalisation.

On every classification approach, the training and testing datasets should be normalised in order to keep the data in the same range. In our case, we have images from different

patients with initial intensity differences that are intensified after applying noise reduction and bias correction. These differences between intensity profiles may cause misclassification errors since lesions do not have the same intensity values for each patient.

Therefore, we propose applying an inter-subject normalisation algorithm focused on the intensity histogram of each patient. First, a model image is chosen (either randomly or arbitrarily). This image's histogram will then be used as a template for the other images in the same database. Afterwards, all other image distributions are matched to this histogram using the method proposed by Nyúl et al. [132] and implemented as part of the ITK library. This algorithm computes histogram landmarks (which can be either peaks and valleys, or concrete percentils) for the images to be matched and creates a look-up table (LUT) to transform the intensities from one image to another. While the brain's anatomy and lesions may change from patient to patient, the tissue percentages are mostly the same for each patient and the relative volume of the lesions compared to the whole brain is relatively small in comparison.

4.2.4 Atlas registration

Rigid registration is usually sufficient when dealing with intra-subject medical applications, such as temporal studies of the same subject. However, when dealing with inter-subject applications, such as atlas matching, nonrigid algorithms can explain local anatomical variations between the template and the subject brain that global methods fail to reproduce (i.e. GM sulci).

In our proposal, we have decided to implement with ITK one of the most common methods based on a multi-resolution approach [182] with a starting pre-registration using affine transformations to globally align the atlas template and the patient's scan, followed by a multi-resolution B-spline transformation. Within the registration framework, the mutual information metric is used as part of the optimisation process.

Similarity map

When the registration is completed, a similarity volume is computed comparing the moved atlas template and the patient's T1-w image. For both images, we create a patch for each voxel of an arbitrary size and compute a similarity metric between them. The normalised cross-correlation metric is chosen over other similarity metrics due to its invariance to linear intensity brightness, and also to decouple the similarity volume process from the

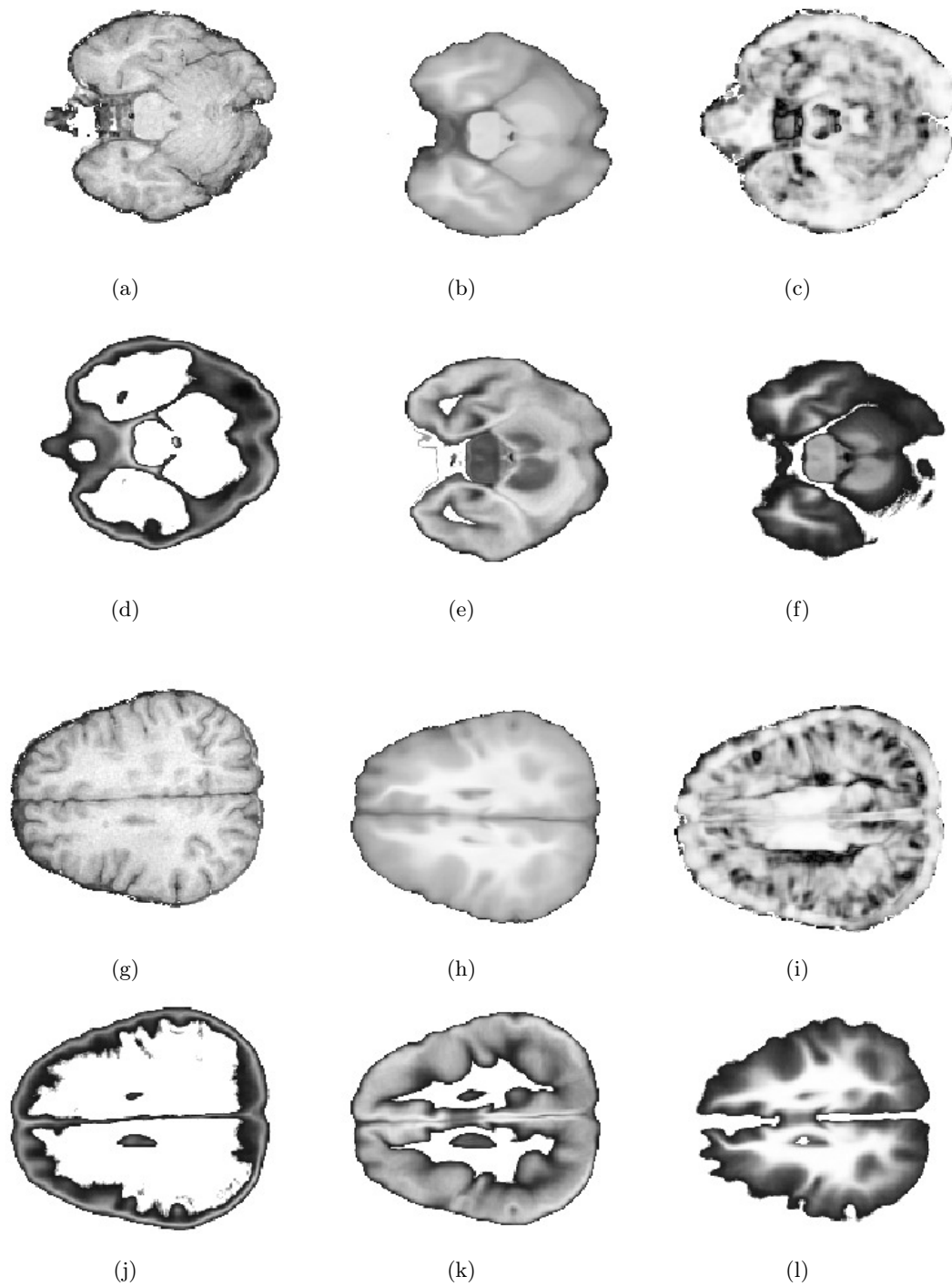


Figure 4.4: Example of atlas registration: a,g) T1-w image; b,h) atlas template; c,i) similarity map; d,j) CSF probability map; e,k) GM probability map and f,l) WM probability map.

optimisation one and reduce the bias introduced when using the same metrics for both processes. Therefore, the similarity map is defined by the following equation:

$$\phi(x) = \frac{\sum_{i \in [x, \mathcal{N}_x]} (I^{T_1}(i) - \bar{I}_x^{T_1})(\pi^I(i) - \bar{\pi}_x^I)}{\sqrt{\sum_{i \in [x, \mathcal{N}_x]} (I^{T_1}(i) - \bar{I}_x^{T_1})^2} \sqrt{\sum_{i \in [x, \mathcal{N}_x]} (\pi^I(i) - \bar{\pi}_x^I)^2}}, \quad (4.2)$$

$$\bar{I}_x^{T_1} = \frac{1}{|\mathcal{N}_x|} \sum_{i \in [x, \mathcal{N}_x]} I^{T_1}(i), \quad (4.3)$$

$$\bar{\pi}_x^I = \frac{1}{|\mathcal{N}_x|} \sum_{i \in [x, \mathcal{N}_x]} \pi^I(i), \quad (4.4)$$

where $I^{T_1}(x)$ is the T1-w image of the patient (used as the fixed image), $\pi^I(x)$ is the atlas template image (used as the moving image) and \mathcal{N}_x is the 3D neighboring area of pixel x .

The resulting similarity map expresses a local relation between the resulting transformation and the image we want to segment, highlighting those regions where the deformation adapts to the new case and darkening the regions where the optimisation process could not reach the desired accuracy. Therefore, this map can then be used as a weighting volume for the probabilistic atlases in a similar manner to the fusion step of multi-atlas propagation strategies [107, 10]. Figure 4.4 shows an example of atlas registration and the similarity map.

4.3 Tissue inference by statistical segmentation using expectation maximisation and lesion outlier thresholding (TISSUE-LOT)

The expectation-maximisation (EM) [64] algorithm is one of the most widely used in the literature for tissue segmentation. A quick look at the tables from chapter 2 corroborates this fact, that was reinforced in chapter 3 while describing probabilistic approaches based on probabilistic atlases. This algorithm interleaves the two interconnected main procedures that it is named after. During the expectation step, the probabilities for each datapoint pertaining to a given distribution are estimated, while in the maximisation step these distribution parameters are computed using the estimated probabilities. While being theoretically simple, it is capable of adapting to different statistical models, although it is mainly used to estimate Gaussian distributions.

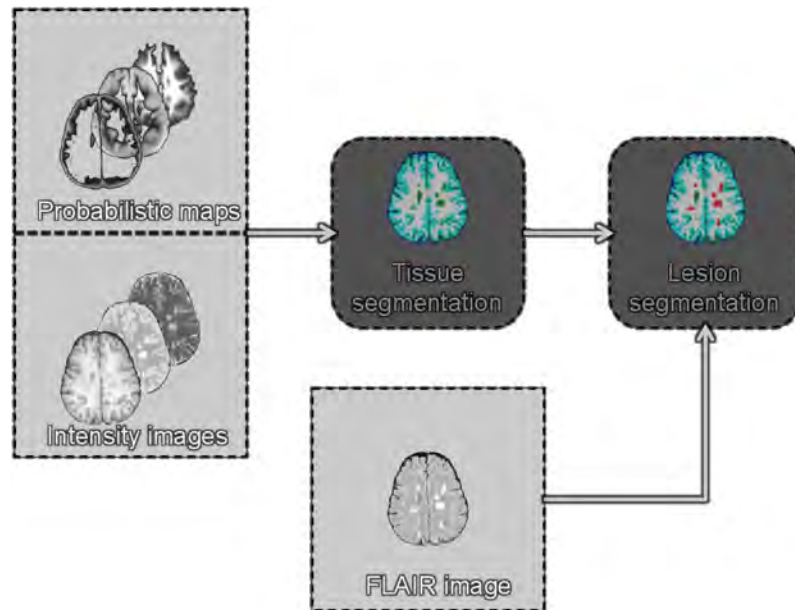


Figure 4.5: Flowchart of the first pipeline used to segment tissues and lesions.

For instance, Souplet et al. [204] adapted the EM algorithm to tackle partial volume (PV) effects. In their approach, which is an extension of the algorithm presented by Dugas-Phocion et al. [72], two main types of classes are defined: pure tissue classes, which are assumed to follow a Gaussian distribution and comprise CSF, GM, and WM; and partial volume classes, which are defined as proportions between GM and CSF. While other interactions may occur (such as WM and CSF around the ventricle boundaries), those are excluded due to their slight influence and the possibility of defining them using CSF and GM proportions.

After this initial tissue segmentation, where lesions are misclassified inside a tissue (mainly partial volumes and GM), a thresholding approach is applied followed by a refinement of this initial lesion mask. This thresholding step is applied exclusively to the FLAIR image and no correspondence between the other images is directly applied.

In this section, we will further explain the basics of this approach and present our contributions, introduced as part of our first lesion segmentation pipeline (see figure 4.5) based on atlas. All the steps in the proposed pipeline are applied to the images after the preprocessing. This means our initial images comprise the corrected T1-w, T2-w, PD-w and FLAIR images plus the CSF, GM and WM probabilistic maps in image space.

4.3.1 Tissue classification with expectation-maximisation (EM)

As stated earlier, the original method from Dugas-Phocion et al. [72] proposed to segment the brain introducing the concept of partial volume effects. To that extent, partial volume classes are introduced in what follows.

Within this framework, we define the pure tissue class set $PT = \{CSF, GM, WM\}$. For each class $c \in PT$, its intensity distribution is modeled as the following multidimensional Gaussian distribution:

$$y_c \sim N(\mu_c, \Sigma_c). \quad (4.5)$$

Given two pure tissue classes, c_1 and c_2 , a PV voxel x with a proportion α of class c_1 (and hence a proportion $1 - \alpha$ of class c_2), its intensity $I(x)$ is a random variable y_{PV} that can be computed as $y_{PV} = \alpha y_{c_1} + (1 - \alpha)y_{c_2}$. Considering all the PV voxels, α is uniformly distributed, however, if α is set as a constant, y_{PV} follows the subsequent Gaussian distribution:

$$y_{PV} \sim N(\alpha\mu_{c_1} + (1 - \alpha)\mu_{c_2}, \alpha\Sigma_{c_1} + (1 - \alpha)\Sigma_{c_2}) \quad (4.6)$$

Dugas-Phocion et al. [72] and Souplet et al. [204] decided to sample α at different constant values in the range $(0, 1)$ and set c_1 and c_2 as CSF and GM. A constant for the number of PV classes was defined in order to select those proportions. Furthermore, they introduced two different outlier classes to deal with imaging artifacts. For instance, Dugas-Phocion et al. defined an outlier class to differentiate between what the atlas considers as CSF but is actually a vessel, while Souplet et al. defined a pure outlier class based on the Mahalanobis distance.

While these two strategies reduce the variance of the PV and CSF classes, it creates small clusters that do not provide extra information and can sometimes collapse and become an empty set. For example, vessels may not be visible in low resolution images or may be represented by only a few voxels. Therefore, we simplified the partial volume approach fixing α to 0.5 and using only a PV class. In fact, we observed that most artifacts and lesions are classified inside this class, as well as the interfaces between CSF and GM.

Initialisation

Initialisation is a crucial aspect of the EM algorithm. Without proper initial parameters the algorithm may take longer to converge or, even worse, fall in a local maximum. Commonly, distribution parameters are estimated either semi-automatically or automatically, and the algorithm starts in the expectation step. However, probability maps (such as atlases) can also be used to estimate the initial tissue parameters as part of the maximisation step. This is the case in tissue segmentation approaches that rely on atlas-based segmentation. For instance, Souplet et al., used the registered atlas for pure tissue classes and, afterwards, PV class parameters were calculated following equation 4.6.

However, atlas registration may cause several errors close to the cortex region even when using deformable registration. This is due to the high spatial variability caused by the gyri and sulci of the brain. Moreover, lesions are usually masked inside WM due to their nonexistence in healthy tissue atlases, even though they do not share the same intensity distribution.

Therefore, initial estimates relying solely on the probabilistic maps after registration may be corrupted by those errors, causing biased initial estimates for the mean and a high variability for each tissue. While those issues can later be overcome with several iterations, reducing them decreases computation time and improves convergence. In our proposal, we combine the atlas probabilities with the similarity map computed during registration and set a threshold T^π to select only those atlas probabilities that are meaningful to the class. As a consequence, not all the voxels are used to initialise the model.

Expectation

In the expectation step, the probability of belonging to the different classes of each voxel is updated with the application of Bayes' rule:

$$p(c|I(x)) = p(I(x)|c)p(c). \quad (4.7)$$

As we explained before, each class (including the partial volume class) follows a Gaussian distribution with parameters $\theta_c = \{\mu_c, \Sigma_c\}$, where $c \in \{CSF, GM, WM, PV\}$. Therefore, $p(I(x)|c)$ can be computed using the following probability density function:

$$\hat{p}^i(I(x)|c, \hat{\theta}_c^{i-1}) = \frac{1}{\sqrt{(2\pi)^D |\Sigma_c|}} e^{-\frac{1}{2}(I(x) - \hat{\mu}_c^{i-1})^t (\hat{\Sigma}_c^{i-1})^{-1} (I(x) - \hat{\mu}_c^{i-1})}, \quad (4.8)$$

where $\hat{p}^i(I(x)|c, \hat{\theta}_c^{i-1})$ is the approximation of $p(I(x)|c)$ at iteration i with the previously estimated parameters $\hat{\theta}_c^{i-1}$ and D is the number of features. In our case $D = 3$, since we use T1-w, PD-w, and T2-w images, the most commonly used MR images for tissue and lesion segmentation, as stated in chapter 2.

To compute the priors for each class $p(c)$, Dugas-Phocion et al. proposed two different approaches. If probability atlases are available for all classes, $p(c) = \pi_c^P$. On the other hand, when no atlases are available, $p(c)$ is re-estimated at each iteration using the following equation:

$$\hat{p}^i(c) = \frac{\sum_{x=1}^{N_{voxels}} \hat{p}^{i-1}(c|I(x))}{N_{voxels}}, \quad (4.9)$$

where N_{voxels} is the total number of voxels. The first approach, introduces a local estimate for each voxel, imposing contextual and spatial constraints, while the second is a global estimate for all the voxels. However, no hybrid approach is proposed by either Dugas-Phocion et al. or Souplet et al. for when some probability atlases are present (i.e. pure tissue classes). Therefore, we propose to create a PV atlas defined as:

$$\pi_{PV}^P(x) = \frac{1}{2}(\pi_{CSF}^P(x) + \pi_{GM}^P(x)), \quad (4.10)$$

since we defined the class PV as an equal proportion of both CSF and GM . With this new atlas, we finally have a probability map for each class that can be used as a local prior probability. However, the probability maps must be normalised again since $\sum_c^C \pi_c^P(x) = 1$.

As pointed out in chapter 3, registration errors are common when registering a template to a new image. These errors arise from differences in anatomy that cannot be captured by the optimisation procedure and increase when dealing with an outlier class, such as MS lesions, that is not present in the atlas template. To take these errors into account, we compute a similarity map as explained in section 4.2.4 to weight the probability maps as priors. However, if we really want to reduce the effect of the atlas on those regions where anatomical differences are present, we must use another prior estimation since the weighting is applied equally to all classes for each voxel. Therefore, we propose the introduction of a simple neighbouring factor as expressed by the following equation:

$$\hat{p}^i(c) = \phi(x)\pi_c^P(x) + (1 - \phi(x)) \sum_{j \in \mathcal{N}_x} \frac{\hat{p}^{i-1}(c|I(j))}{|\mathcal{N}_x|}, \quad (4.11)$$

where $\phi(x)$ represents the similarity map, \mathcal{N}_x represents the 3D neighbouring voxels of x , and $\hat{p}^{i-1}(c|I(j))$ represents the probability of voxel j of pertaining to class c computed at the previous iteration.

Maximisation

In the maximisation step, the parameters of each class are computed from the voxels' intensities and their probabilities of belonging to the different classes. The partial volume class parameters are obtained as a proportion of the pure tissue parameters.

A common way to compute the mean and covariance matrix is to use the maximum likelihood estimator (MLE) [72, 204]:

$$\hat{\theta}_c = \arg \max_{\theta_c} \prod_{x=1}^{N_{\text{voxels}}} p(I(x)|\theta_c). \quad (4.12)$$

However, the MLE is sensitive to outliers and, as a consequence, one of the parameters can become arbitrarily large. García-Lorenzo et al. [87] proposed using a trimmed likelihood estimator to reduce the effect of those outliers. This new estimator uses the Mahalanobis distance to discard those voxels that could be considered outliers when computing the Gaussian parameters.

In our approach, we use an approximation of such estimator where a probability threshold is used instead of the Mahalanobis distance to compute the class parameters $\hat{\theta}_c^i$ as follows:

$$\hat{\mu}_c^i = \frac{\sum_{x \in \tilde{X}} \hat{p}^i(c|I(x))I(x)}{\sum_{x \in \tilde{X}} \hat{p}^i(c|I(x))}, \quad (4.13)$$

$$\hat{\Sigma}_c^i = \frac{\sum_{x \in \tilde{X}} \hat{p}^i(c|I(x))(I(x) - \hat{\mu}_c^i)^t(I(x) - \hat{\mu}_c^i)}{\sum_{x \in \tilde{X}} \hat{p}^i(c|I(x))}, \quad (4.14)$$

$$\tilde{X} = \{x : \hat{p}^i(c|I(x)) > T^\pi\}, \quad (4.15)$$

where T^π is the probability threshold used during the initialisation step.

Final tissue segmentation

Once the algorithm converges, meaning the probability maps $\hat{p}^i(c|I(x))$ obtained at the expectation step stay the same, or after a certain number of steps the iterative process is

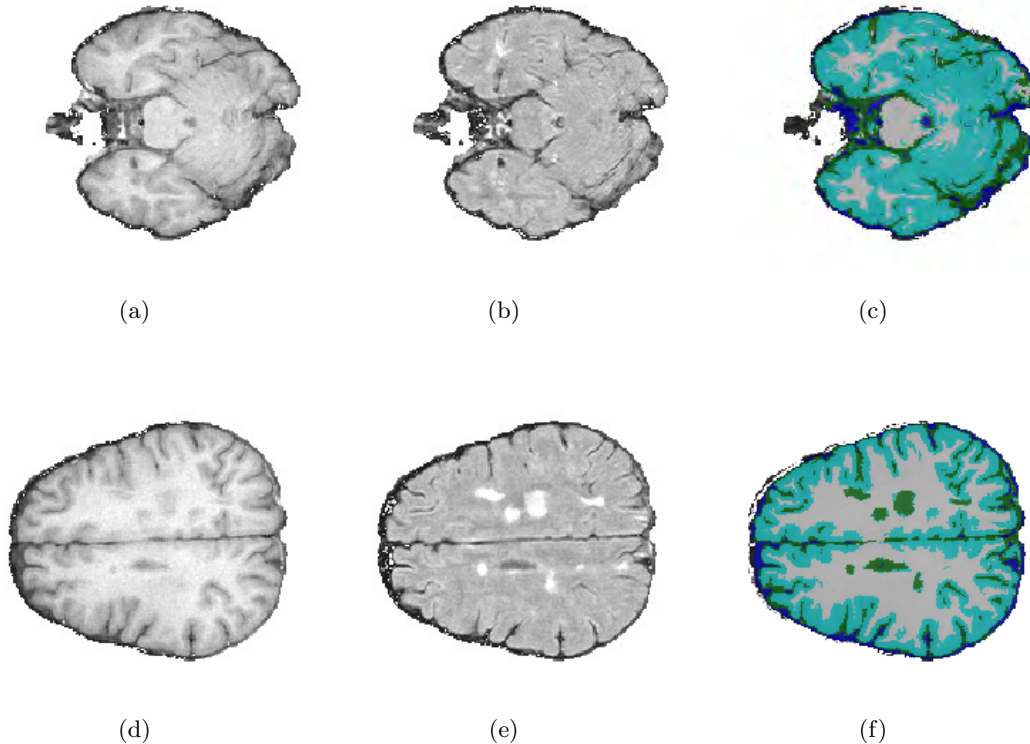


Figure 4.6: Example of tissue segmentation: a,d) T1-w image; b,e) FLAIR image; and c,f) final tissue segmentation $S(x)$.

halted and two separate outputs are obtained. On the one hand, we have an estimation of tissue Gaussian parameters $\hat{\theta}_c$, and, on the other hand, we have a set of probabilistic maps. Therefore, to obtain a final tissue segmentation, we apply the following equation, which we introduced in chapter 3:

$$S(x) = \arg \max_c \hat{p}^{end}(c|I(x)), \quad (4.16)$$

where $S(x)$ is the segmentation and $\hat{p}^{end}(c|I(x))$ represents the probability map for class c obtained at the last step of the EM algorithm. An example of this segmentation, alongside a T1-w image and a FLAIR image, used to segment the lesions, is illustrated in figure 4.6.

4.3.2 Lesion segmentation with FLAIR

As introduced in chapter 1, lesions appear hyperintense in FLAIR images. Moreover, GM voxels, the normal appearing tissue with the highest intensity in these images, have a

darker distribution than lesion voxels. As a consequence, there is a noticeable difference between normal appearing tissues and lesions in these images. Therefore, lesions that are misclassified by the EM algorithm are actually hyperintense outliers of the GM FLAIR intensity distribution. Moreover, those lesions should be either segmented as a GM-based class (pure GM or PV) or WM (black holes) and should always be surrounded by WM.

FLAIR thresholding

In order to find a threshold for the FLAIR image, we first assume that GM intensities follow a Gaussian distribution in the FLAIR image (as we did for the other MRI images to segment the tissues). We also assume that lesions are hyperintense outliers of this distribution. With these two assumptions, we can estimate a threshold T^F as follows:

$$T^F = \mu_{GM}^F + \gamma\sigma_{GM}^F, \quad (4.17)$$

where the μ_{GM}^F and σ_{GM}^F are the distribution parameters of GM on the FLAIR image and γ is an empirical parameter used to determine the outliers.

To estimate the tissue parameters, the GM mask, obtained in the previous step, is applied to the FLAIR image to compute a histogram (see figure 4.7). Since lesions are usually part of the GM mask (due to their intensity similarities in PD-w and T2-w images), the histogram is biased by the high lesion intensities. This bias does not affect the peak of the histogram, which is used to define the mean, however, it can affect the estimation of the standard deviation. Therefore, we use the full width at half maximum (FWHM) expression to compute it [61].

FWHM is an expression of the extent of a function, given by the difference between the two extreme values of the independent variable (FLAIR intensities) at which the dependent variable (histogram frequency) is equal to half of its maximum value. If the considered function is the normal distribution (as we assumed for GM) the relationship between FWHM and the standard deviation is:

$$\sigma_{GM}^F = \frac{FWHM}{2\sqrt{2\ln(2)}}, \quad (4.18)$$

where σ_{GM}^F is the standard deviation of GM on the FLAIR image we are looking for.

Upon obtaining μ_{GM}^F and σ_{GM}^F and using them to compute T^F by means of equation 4.17, we apply it to the FLAIR image to obtain an initial lesion mask where other hyperintense

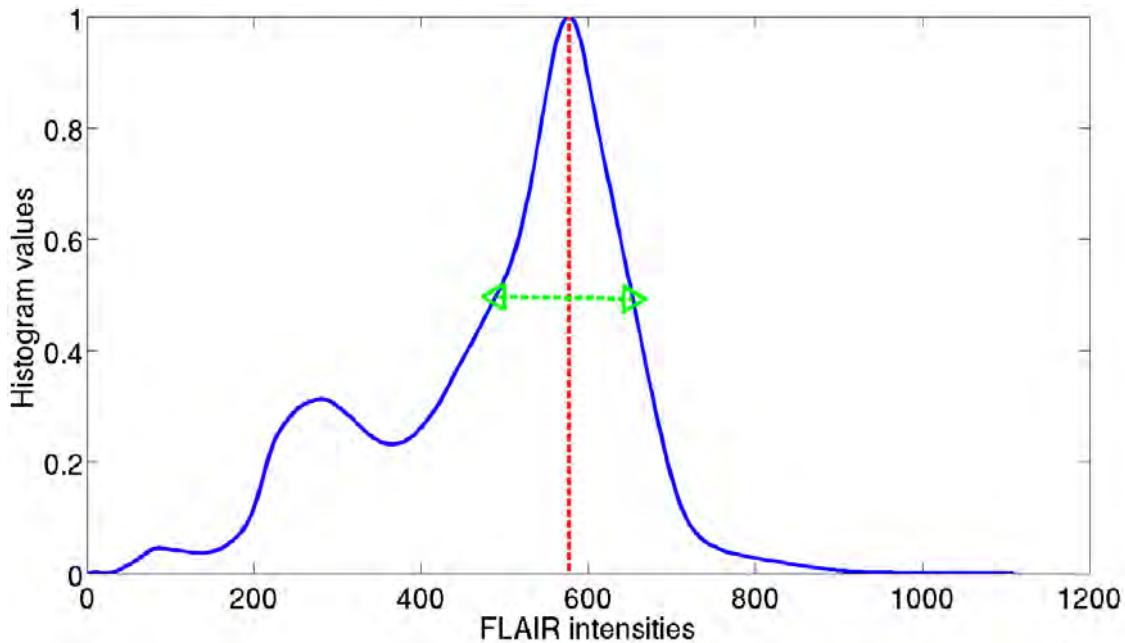


Figure 4.7: FLAIR histogram for the GM tissue. The red line represents the peak used to compute the mean, while the green line represents the full width at half maximum (FWHM) used to compute the standard deviation.

artifacts are also included. This mask must be refined, since a large number of false positive lesions are usually found.

Mask refinement

After thresholding, the mask is relabeled in order to divide it into different regions. A region is defined as a set of voxels that are connected using a 3D neighborhood cube of connectivity 26. That means that 2 voxels belong to the same region if they are directly connected in a 3D space. We decided to use a region-wise refinement step [174, 2, 61] to reduce FP in terms of detection due to artifacts.

In order to differentiate lesions from other regions, we define a set of rules that are true for WML:

- **Lesions are mostly classified as WM and GM-based classes (PV and pure GM).** In the case of WM, most lesions share a similar intensity profile to that tissue in T1-w images and have a high probability in the WM map (due to registration

limitations); while in the case of GM, lesions share the same intensity profile in T2-w and PD-w images. Furthermore, black holes share the same intensity profile in T1-w images and present low values in the similarity map (weighting down the atlas probabilities during segmentation). To determine if a region follows this rule, a positive mask (containing WM, PV, and pure GM) and a negative mask (containing the background and all other tissues) are defined and for each region the proportion between positively and negatively masked voxels is computed. If this proportion is higher than the user-defined threshold T^{ω_t} , the region is labeled as WML.

- **Lesions are surrounded by WM.** We are only focusing on WML, therefore all lesion regions should be surrounded (mostly) by WM voxels. To determine if a region follows this rule, a positive mask (containing only WM) and a negative mask (containing the background and all other tissues) are defined and for each region the proportion between all its positive and negative neighbouring voxels is computed. If this proportion is higher than a user-defined threshold T^{ω_N} , the region is labeled as WML.
- **Lesions should not be present between the ventricles.** As stated in chapter 1, one of the typical MS lesion locations is the surrounding of the ventricles. However, these lesions rarely appear between them and hyperintense regions in these areas are usually due to artifacts. Therefore, we remove all the regions that are close to the centre of the brain (which is usually between the ventricles).
- **Lesions should be of a minimum size.** Due to noise and inhomogeneities, some voxels may present random high intensities. These voxels are usually corrected during the preprocessing, however, some of them might still be considered as lesions after thresholding the FLAIR image. To remove these small outliers, we discard all regions that do not conform to a minimum size.

Finally, all the regions discarded according to one (or more) of these rules are reclassified according to the tissue segmentation. An example of the final segmentation (including lesions and tissues) is presented in figure 4.8.

4.4 Boosting with outliers and other spatial tools (BOOST)

In section 4.3, we have presented an atlas-based approach to introduce prior knowledge to the segmentation process. As introduced in chapter 2, other common supervised ap-

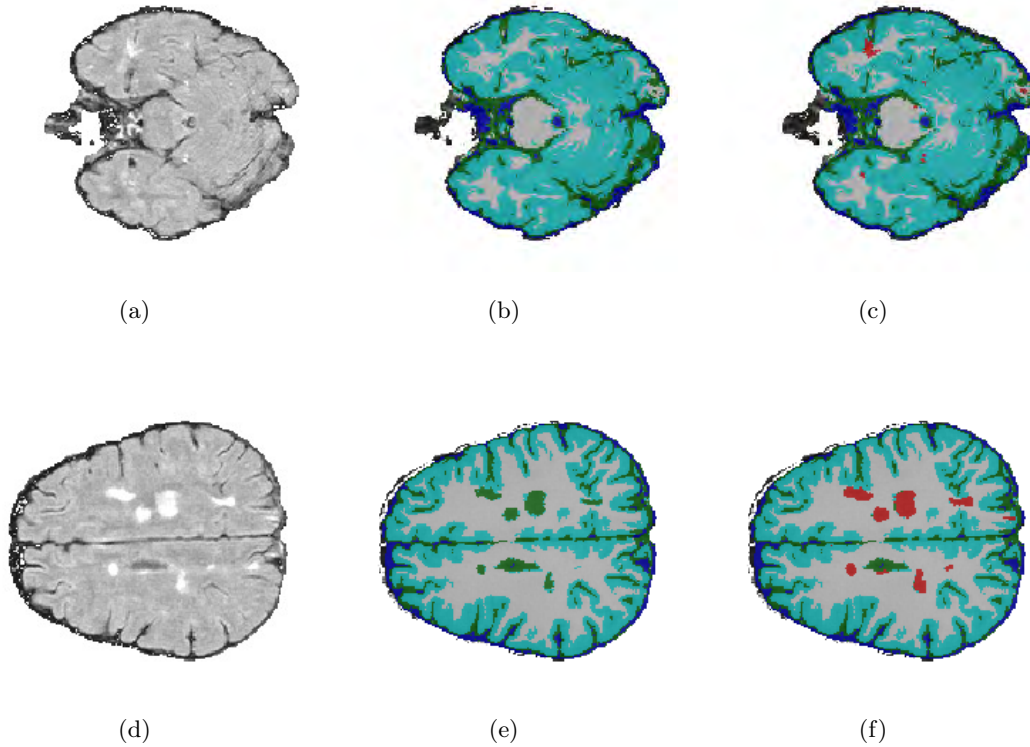


Figure 4.8: Example of the TISSUE-LOT pipeline segmentation: a,d) FLAIR image; b,e) tissue segmentation; and c,f) final segmentation including tissue and lesions.

proaches rely on the use of a classification algorithm. These algorithms usually involve two stages. In the first stage, a model is estimated on a training set composed of a set of features and their corresponding ground truth. This model can be as simple as a linear discriminator [119], or as complex as an artificial neural network [94]. In the second and final stage, this model is tested on a new dataset to classify it.

Most classification algorithms differentiate between lesion and not lesion (object and background), however, other algorithms can use multiple labels and produce a probabilistic tissue map as a result. These probabilistic maps are usually the result of using an ensemble of classifiers such as bagging [92] or boosting [156]. These strategies combine simple discriminative classifiers that focus on a small subset of a rich feature pool to classify a small subset of the training set. One of the advantages of this specialisation is the implicit feature selection applied to the dataset. These strategies give a stronger impact to the most discriminating features while underscoring or discarding the less discriminating ones.

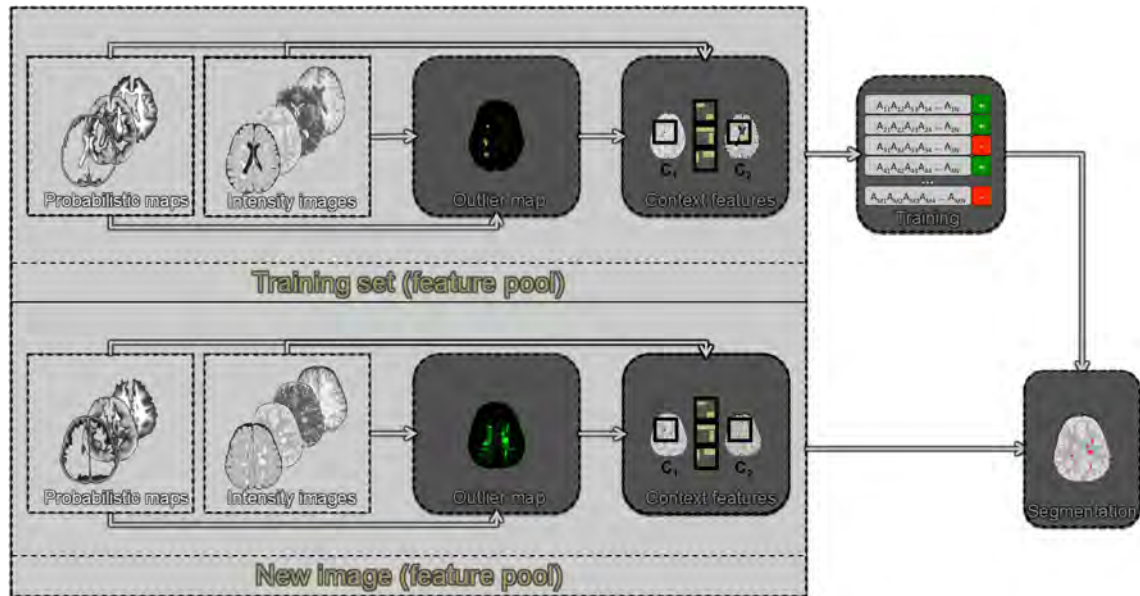


Figure 4.9: Flowchart of the second pipeline used to segment lesions using a trained classifier.

A rich pool of features is important when using ensemble classifiers. These features can comprise common features used for classification such as image intensities [119, 61], spatial locations [6], atlas probabilities [94] or more complex features such as meta-features that exploit contextual information [92]. These approaches have been extensively used in other imaging fields such as object detection [200] or astronomical source detection [223].

As a second pipeline (see figure 4.9), we proposed a classification approach based on an ensemble classifier of regression stumps with a rich feature pool that combines multi-channel image intensities, probability maps, contextual meta-features, and an outlier map based on image intensities.

4.4.1 Outlier map

Lesions are usually seen as hyperintense regions in PD-w, T2-w images, and FLAIR images. However, lesions may present a wide range of grey level values in T1-w images (ranging from normal appearing WM to CSF). Although FLAIR provides a good contrast between tissues and WML, it also presents some disadvantages and artifacts that may cause misclassifications and false positive (FP) detections [162]. Moreover, lesions should also appear in PD-w and T2-w images, while most FLAIR artifacts should not be present.

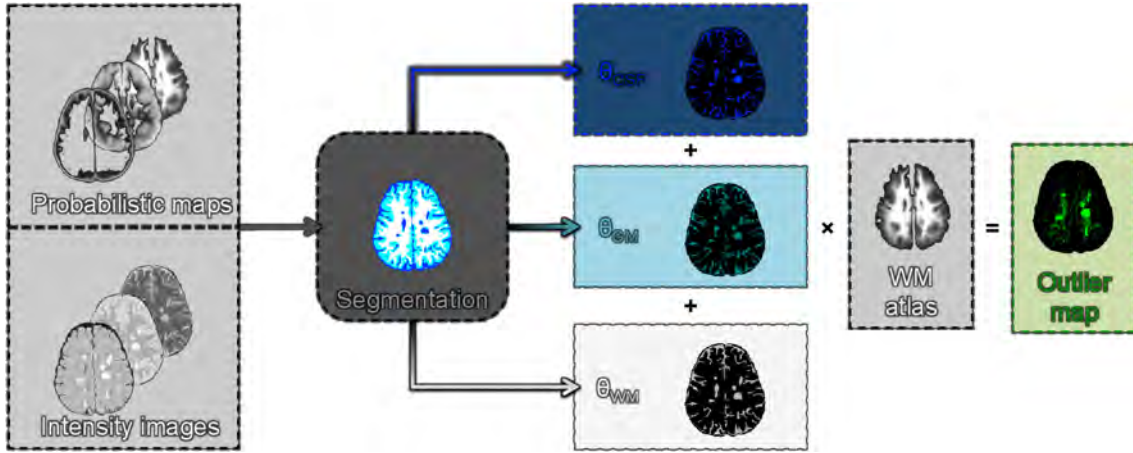


Figure 4.10: Scheme of the outlier map computation. θ_c represents the mean and covariance matrix of each tissue c used to compute the Mahalanobis distance to obtain the outlier maps for that tissue.

Therefore, we decided to extend the map presented by Schmidt et al. [187] to a multidimensional outlier map using PD-w and T2-w images to reduce the FP presence (see figure 4.10). Since WML have a higher intensity than all 3 tissues, the original approach consisted of assigning to each voxel the positive distance (0 otherwise) to the tissue distribution for WM, GM and CSF previously estimated using the EM algorithm presented in the previous section. We propose extending this approach to multi-channel using the Mahalanobis distance as follows:

$$O_c(x) = \begin{cases} d(x, \theta_c) & \text{if } \forall s : I^s(x) > \mu_c^s \\ 0 & \text{otherwise} \end{cases} \quad (4.19)$$

where x represents the voxel coordinates, $s \in \{T_2, PD, FLAIR\}$, $c \in \{CSF, GM, WM\}$, $\theta_c = \{\mu_c, \Sigma_c\}$, and $d(x, \theta_c)$ is computed using the following equation:

$$d(x, \theta_c) = \sqrt{(I(x) - \mu_c)^t \Sigma_c^{-1} (I(x) - \mu_c)}. \quad (4.20)$$

Afterwards, we sum the three tissue maps and weight them using the WM probabilistic atlas to reduce the effect of outliers outside the tissue. Finally, we relax this map using neighbouring information for each voxel as follows:

$$O(x) = O(x) \sum_{j \in \mathcal{N}_x} \frac{O(j) + \pi_{WM}^P(j) - O(j)\pi_{WM}^P(j)}{|\mathcal{N}_x|}, \quad (4.21)$$

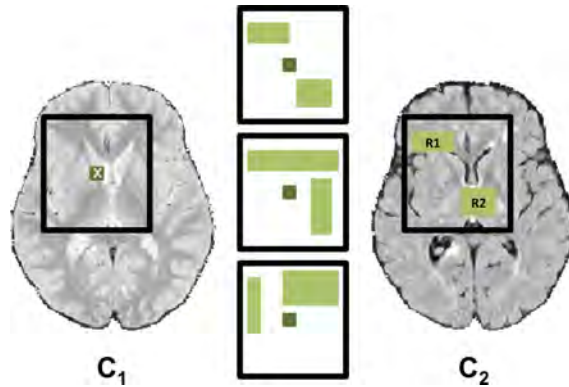


Figure 4.11: Scheme of the meta-features used. The 3D regions R_1 and R_2 are randomly sampled inside a fixed neighbourhood around the central voxel x , and their means are compared to that voxel.

$$O(x) = \pi_{WM}^P(x) T_1(x) \sum_c O_c(x). \quad (4.22)$$

With this smoothing, we want to reduce the effect of isolated spurious hyperintense voxels not surrounded by WM or other outlier voxels.

4.4.2 Context meta-features

Even though the outlier map can be used to detect lesions as hyperintense outliers, due to image artifacts, other regions may present similar properties. These false positive regions are usually located in unlikely places for WML (such as inside the cortex). Contextual information or spatial constraints can, therefore, aid automated strategies to differentiate between real lesions and artifacts present in the MRI scans.

To introduce this information, we propose using meta-features similar to those presented by Geremia et al. [92]. These meta-features (see figure 4.11) provide a comparison between the voxel on channel V_1 (which can be either a probabilistic map or an intensity image) and the mean of two different regions (R_1 and R_2) on channel V_2 as follows:

$$\lambda(x) = V_1(x) - \bar{V}_2(R_1) - \bar{V}_2(R_2) \quad (4.23)$$

where $V_1, V_2 \in \{I^d, \pi_c^P, \mathcal{O}\}$, I^d is the image $d \in \{T_1, PD, T_2, FLAIR\}$.

The three-dimensional regions R_1 and R_2 are randomly sampled (in size, shape and location) in a neighbouring cube around the voxel x , and the mean for them is efficiently computed using integral volume processing [200].

4.4.3 Boosting classifier

At this point, we have a rich pool of features characterised by 4 intensity channels (T1-w, PD-w, T2-w, and FLAIR), 3 probabilistic atlases (CSF, GM, and WM), an outlier map, a set of randomly sampled context features, and a set of training cases. From these training cases, we randomly sample a set of positive examples pertaining to WML and negative samples pertaining to any tissue with a ratio of 1 positive example for 3 negative ones (to account for the 3 possible tissues). These samples are the ones used to train a classifier.

We decided to use the Gentleboost algorithm [84, 221], which is based on the simple idea that the sum of weak classifiers can produce a strong classifier (see figure 4.12). This algorithm has been widely used for different segmentation applications within the VICOROB group [223, 222, 163] and offers a comprehensive description of the most discriminating features used. In the Gentleboost algorithm, the weak classifiers used at each round are simple regression stumps with one of the features:

$$h^i(p) = a \psi(p > T^{h^i}) + b, \quad (4.24)$$

where a and b are the parameters of the regression stump, being $a = 2$, $b = -1$ the values obtained in the perfect case, and $a = 0$ and $b = 0$ the values obtained in a random example, $\psi(\chi)$ refers to the Heaviside step function (0 if $\chi < 0$; 1 otherwise), and T^{h^i} is the threshold that determines if pattern p belongs to the object class. At each round, the values of the parameters a , b , and T^{h^i} are selected to minimise the error of the classifier:

$$e = \min \sum_{t=1}^{N_t} (z^i(y - h^i(v_t^f))^2), \quad (4.25)$$

where v_t^f is the value of feature $f \in \{I^d, \pi_c^P, \mathcal{O}, \lambda\}$ for the training data point t , y is the lesion label (being 1 or -1) and z^i the training data weights at round i . The minimisation is done by exhaustively looking for all the values of T^{h^i} since, assuming a fixed T^{h^i} , the values of a and b are automatically assigned. Hence, the value of b corresponds to the mean weighted label of the instances lower than the threshold, while a is the mean weighted label

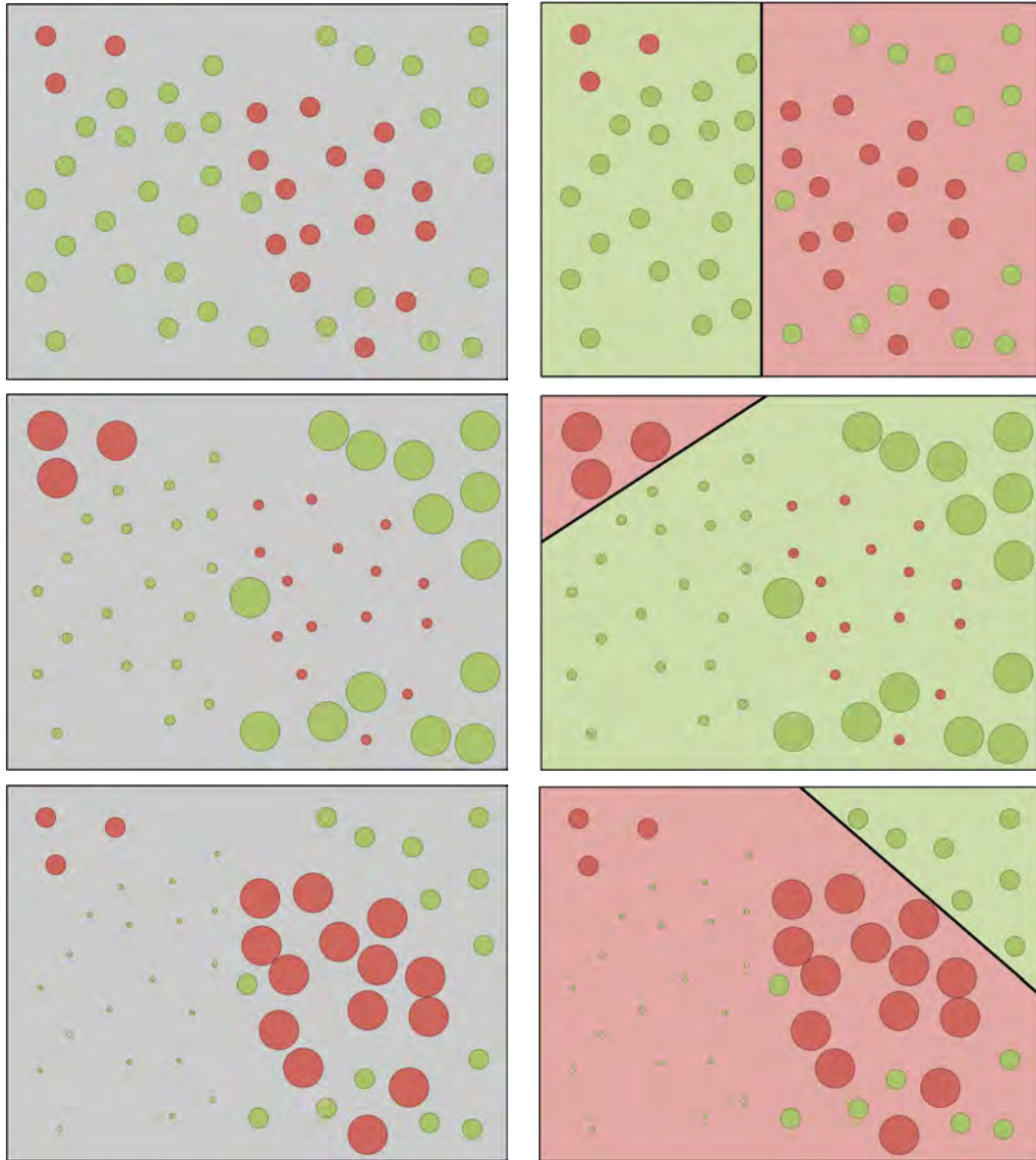


Figure 4.12: Example of a boosting training with linear classifiers in a 2D feature space. Green dots represent positive examples, red dots represent negative ones and the lines represent a linear classifier. After defining a classifier, those samples that are misclassified are given a higher weight while the correctly classified ones are given a lower weight in order to correctly classify the misclassified samples. The process is repeated until convergence or until a certain number of rounds has been achieved.

of the instances greater than the threshold (minus b to satisfy eq. 4.24):

$$b = \frac{\sum_{t \in \mathcal{T}_{\nabla}^f} y_t \cdot z_t^i}{\sum_{t \in \mathcal{T}_{\nabla}^f} z_t^i}, \quad (4.26)$$

$$a = \frac{\sum_{t \in \mathcal{T}_{\Delta}^f} y_t \cdot z_t^i}{\sum_{t \in \mathcal{T}_{\Delta}^f} z_t^i} - b, \quad (4.27)$$

where $\mathcal{T}_{\nabla}^f = \{t : v_t^f < T^{h^i}\}$ and $\mathcal{T}_{\Delta}^f = \{t : v_t^f \geq T^{h^i}\}$.

At each round, only the feature f^i that obtains the minimum error e is selected and used in the testing step to classify the new data. Moreover, at each round of the boosting the weights z_t are updated, increasing the possibility of correctly classifying the previous incorrectly classified instances in the following round. In the Gentleboost algorithm, the data weights are updated using:

$$z_t^{i+1} = z_t^i e^{y \cdot h^i(v_t^{f^i})}. \quad (4.28)$$

Hence, when testing new data, the final (strong) classifier is computed using the weak classifier created at each round of the boosting. Therefore, the testing data is classified according to the sign of the sum of weak classifiers:

$$H(v) = \sum_i^{N_{rounds}} h^i(v^{f^i}). \quad (4.29)$$

Therefore, voxels being part of a lesion should obtain positive values while the rest of the voxels should obtain negative values. Furthermore, the absolute value of $H(v)$ shows the confidence of the classified data. To obtain a final lesion mask, the output of $H(v)$ must be thresholded. An example of this thresholding and the outlier map is shown in figure 4.13. Moreover, the classifier also returns the f^i features selected. These features can be used more than once.

4.5 Summary

In this chapter, we have presented two different pipelines based on two different supervised approaches, both including an atlas. Moreover, these approaches share an initial preprocessing pipeline which comprises several steps to reduce the image artifacts and issues

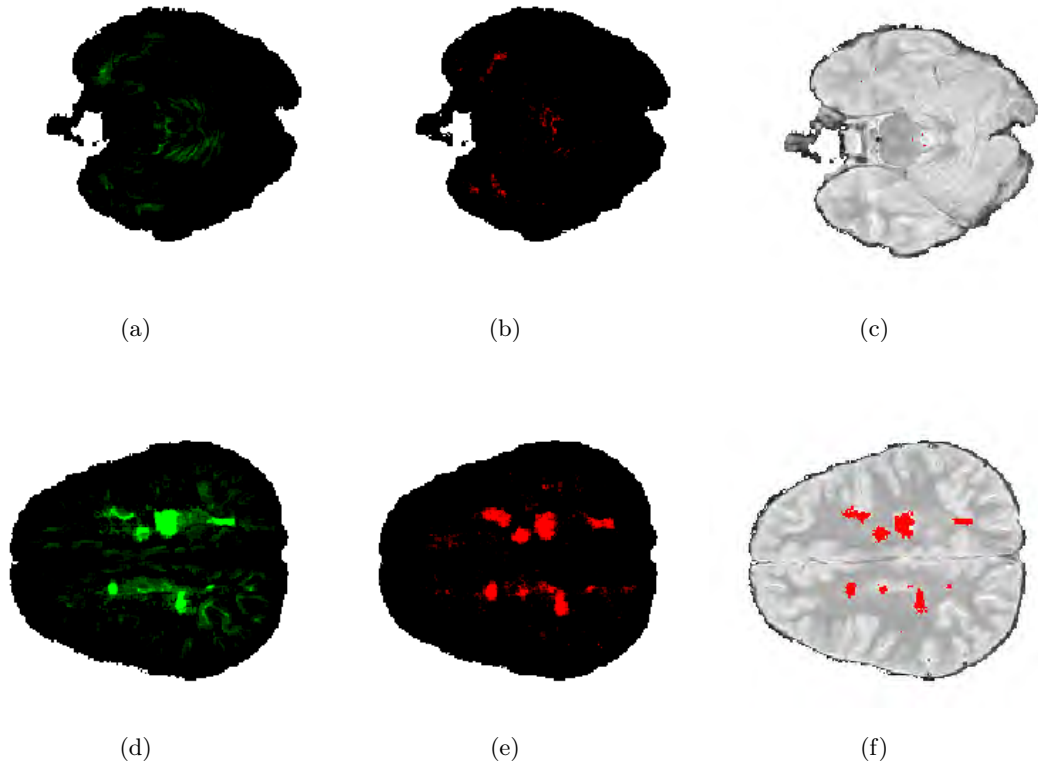


Figure 4.13: Example of the BOOST pipeline segmentation: a,d) outlier map; b,e) output of the classifier ($H(v)$); and c,f) final lesion segmentation.

introduced in chapter 1. In this section, we will summarise these pipelines, which will be evaluated in the next chapter, and highlight our contributions.

4.5.1 Preprocessing pipeline

Before segmentation, different preprocessing steps must be applied to prepare the images for further analysis. The first step consists of the removal of the skull and other nonbrain tissues. This step is applied by means of the publicly available BET tool [203], which is a part of the FSL toolbox. Afterwards, a gradient anisotropic diffusion filter [166] is applied, followed by the N4 ITK algorithm [224] to reduce image noise and intensity inhomogeneities.

Once the images are enhanced, we apply two other preprocessing steps needed by the subsequent segmentation pipelines. First, in order to improve the training step of our classification approach, we normalise the intensity images of the same dataset using a

histogram matching approach [132]. Finally, as part of any atlas-based segmentation, the atlas template and probabilistic maps provided by the ICBM [129] are aligned within a two-stage registration approach. Initially, a global affine transformation between the patient's T1-w image and the atlas' template is estimated using the mutual information metric and a gradient step descent optimiser, which thereafter is expanded and refined using a B-spline transformation to define local deformations [182]. When the atlas is finally warped into the image space, a similarity map representing the voxelwise normalised cross-correlation between the patient's T1-w image and the template is computed.

Due to its pipeline nature, each preprocessing step may propagate and produce new errors that may, in turn, affect both segmentation approaches. For instance, if the eyes are not cleanly removed, their intensity distribution may affect the bias processing and registration, and as a consequence bias the tissue distributions computed inside the TISSUE-LOT pipeline and the lesion model estimated by the BOOST pipeline.

4.5.2 TISSUE-LOT pipeline

Our first approach combines the atlas information with the intensity images to, first, produce a tissue segmentation estimate and, then, use this information as part of an outlier thresholding approach to segment lesions.

The initial tissue segmentation is conducted as part of a modified EM approach where 4 main classes are identified: CSF, GM, WM, and PV. This approach shares some similarities with Souplet et al.'s proposal [204]. However, we proposed a set of differences to improve convergence, increase the robustness of the tissue parameters estimates, and reduce the effect of missregistration errors. Our first contribution consists of creating a PV atlas that is used during the expectation step as a prior. Furthermore, we redefine the priors by introducing our similarity map computed during registration to weight the atlas and include neighbouring information in those areas where the registration failed to capture tissue deformations. With those changes, we assure a smooth segmentation (without small, random voxels) which is not absolutely driven by the atlas. Finally, in order to improve the tissue parameter estimation, we use a threshold and a trimmed likelihood estimator to compute the mean and covariance matrix during the maximisation step.

Once we obtain this initial tissue segmentation using T1-w, T2-w and PD-w images, we apply the GM mask to the FLAIR image, where lesions appear hyperintense in contrast with all the other tissues, to compute its distribution parameters from the histogram. In order to reduce the effect of misclassified lesion voxels, the mean is estimated using the

peak, while the standard deviation is computed using the FWHM. Afterwards, a threshold based on those two parameters and a user-defined one is applied to the image to obtain an initial lesion mask. This mask is further refined to reduce false positives by applying a set of rules based on MS lesion properties.

4.5.3 BOOST pipeline

Our second approach comprises an ensemble classifier approach trained with intensity images, atlas probabilities, an outlier map computed using the TISSUE segmentation algorithm, and a set of contextual meta-features.

Firstly, an outlier map is computed from the estimated parameters during the EM segmentation approach. This map is constructed in several steps. Initially, a positive tissue outlier map is estimated for CSF, GM, and WM using T2-w, PD-w, and FLAIR images, since lesions appear hyperintense on these images. Afterwards, these tissue outlier maps are summed and weighted using the WM probability map and the T1-w image. Finally, neighbouring information is applied to smooth the map.

Secondly, a set of contextual meta-features is computed using all the previous features. To compute these features, two cubes are defined around a neighbouring area of each voxel, and their mean is compared to the voxel's intensity.

Finally, a gentleboost classifier based on regression stumps is trained with all the features. Afterwards, the classifier is tested with the new image and a map with the confidence for each voxel is obtained. This map is, at last, thresholded to obtain a final lesion segmentation.

EXPERIMENTAL RESULTS

You cannot lose if you do not play.

Marla Daniels, “The Wire (Season one)” (2002)

5.1 Evaluation

Evaluation is a recurring topic when dealing with medical image analysis. Depending on the application, the process of validating a certain approach can be as complex as the approach itself due to different complications [226]. However, the two main key aspects to take into account in any medical image analysis application are the databases with which the approach will be tested and the evaluation measures that will be computed.

As introduced in chapter 2, one of the main difficulties in MS lesion segmentation is the lack of a common database to train and test the algorithms. Most strategies are tested with their own private database, difficulting the task of a fair and quantitative comparison among different methods. Moreover, there is no baseline method to which a new strategy can be compared, even though we pointed out some of the most promising approaches as part of the conclusions in chapter 2.

On the other hand, there is a clear tendency to use the same evaluation measures to validate MS lesion segmentation methods. In chapter 2, we also described all the different measures that have been discussed throughout the literature, with special emphasis on overlap measures. Specifically, the DSC has become a *de facto* standard for brain MRI segmentation in general as pointed out in chapter 3. However, as we will further discuss, using only a similarity index is not enough to capture all the strengths and weaknesses of a concrete segmentation approach.

Therefore, in this chapter we will, first, introduce the database we used to evaluate our two proposals described in chapter 4. Afterwards, we will present the measures we chose followed by the results for each pipeline. Finally, we will present a comparison and a discussion of the results obtained, pointing out the important aspects of each pipeline.

5.2 Databases

In chapter 2, we briefly introduced a database of real cases that was provided as part of the MICCAI conference’s MS lesion segmentation challenge 2008 [208]. Nowadays, the website for the challenge hosts the files for the 20 training cases (10 from the CHB and 10 from the UNC) with T1-w, T2-w and FLAIR images. In order to test the methods with the whole database, a petition has to be made through the website.

On the other hand, within the SALEM project’s framework, we created a novel database with images from 3 different scanning machines. Furthermore, for each hospital, patients with different lesion load were scanned.

5.2.1 MS Lesion Segmentation Challenge 2008 database

As part of MICCAI 2008, the program included a workshop to serve as a common framework to test and compare different approaches to segment MS lesions. To that extent, a public database was created with imaging data from two different centres, the University of North Carolina (UNC) and the Children’s Boston Hospital (CHB), comprising 53 different cases: 20 training volumes (with ground truth) and 25 testing volumes were readily available before the conference; and 8 final volumes were provided for on-site testing during the conference. Nowadays, only the training volumes can be downloaded from the workshop’s website. Further processing with other cases is available after petitioning the organisers at the expense of finally being ranked with all the other approaches that tested their methods with the database.

Challenge’s images were obtained using a 3T MRI scanning machine. While the higher magnetic field provides higher spatial resolution thus reducing partial volume effects, it also causes new artifacts or stronger artifacts than using a 1.5T scanning machine. For instance, intensity inhomogeneities are usually stronger and the volume presents a smooth but visible darkening through the z coordinate of the FLAIR image. Common intensity inhomogeneity algorithms do not deal with or can not accurately correct these issues. As a consequence, the estimated threshold is biased, producing a high number of FP inside the

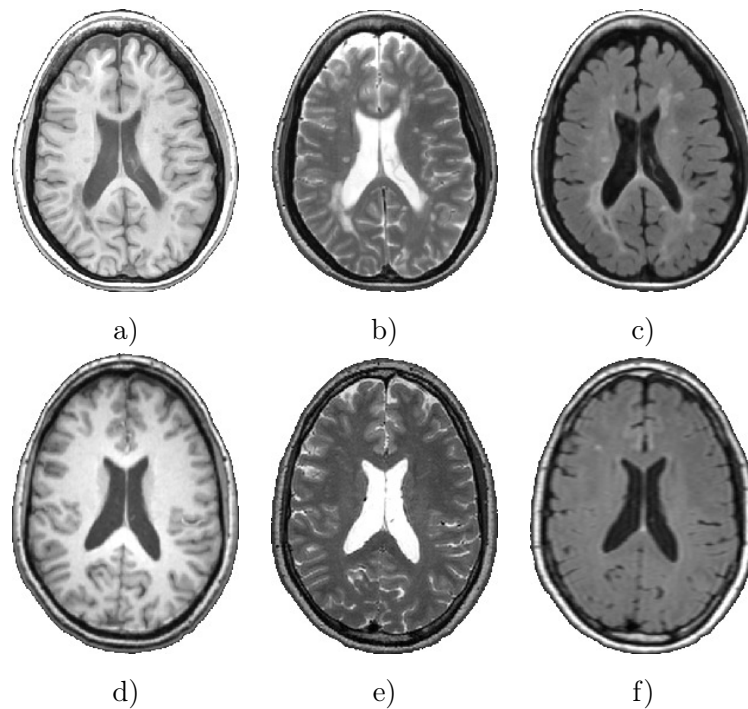


Figure 5.1: Examples of MS challenge 2008 for the CHB hospital, a) T1-w image, b) T2-w image, and c) FLAIR image; and the UNC hospital, d) T1-w image, e) T2-w image, and f) FLAIR image.

GM. In order to reduce these artifacts, a filtering approach based on MRF and neighboring information was used. However, this filtering is not necessary when dealing with 1.5T images, such as the ones that are part of the SALEM database, the focus of this PhD. In this chapter, we will present this database and evaluate the proposed pipelines with it. Also, a report with some initial results obtained using the challenge database is included in appendix A. Furthermore, this appendix includes results obtained using the BrainWeb synthetic phantom.

5.2.2 SALEM database

One of the main goals of the SALEM project was to create a novel database with images from three different hospitals to evaluate the effect of different scanning machines on the proposals.

For each hospital, 15 different patients were scanned with the same protocol (T1-w, T2-w, PD-w and FLAIR): from Hospital Vall d'Hebron, using a 1.5T Siemens Symphony

Quantum with 2D conventional spin-echo T1-w (TR 653 ms, TE 14 ms), dual echo PD T2-w (TR 2800 ms, TE 16 / 80 ms), and FLAIR (TR 8153 ms, TE 105 ms, and TI 2200 ms); from Hospital Josep Trueta, using a 1.5T Philips Intera (R12) with 2D conventional spin-echo T1-w (TR 450 ms, TE 17 ms), dual echo PD T2-w (TR 3750 ms, TE 14 / 86 ms), and FLAIR (TR 9000 ms, TE 114 ms, and TI 2500 ms); and from Clínica Girona, using a 1.5T GE Signa HDxt, with 3D fast spoiled gradient T1-w (TR 30 ms, TE 9 ms), fast spin echo T2-w (TR 5000-5600 ms, TE 74-77 ms), PD-w (TR 2700 ms, TE 11.9 ms), and FLAIR (TR 9002 ms, TE 80 ms, and TI 2250 ms). All the images were acquired in axial-view with a slice thickness of 3 mm ($1 \times 1 \times 3$ mm). Examples of the images from each hospital are presented in figures 5.2, 5.3 and 5.4;

From each hospital, patients with different lesion load were scanned as illustrated in figure 5.5. By looking at both the box and bar plots, there is a clear difference between the lesion load among the three hospitals.

Hospital Vall d'Hebron showed the lowest mean (3.1 cm^3) and median lesion load (1.62 cm^3) per patient. Besides, most of the cases have a lesion load inferior to the mean. On the other hand, Hospital Josep Trueta has the highest mean (18.03 cm^3) and median lesion load (10.54 cm^3) of all three hospitals. This hospital also has a high variability in terms of lesion load. For instance, the minimum and maximum lesion loads are 0.19 cm^3 and 52.48 cm^3 respectively. The last hospital, Clínica Girona, presents a mean (10.34 cm^3) and median (5.03 cm^3) value in-between the other two hospitals. Furthermore, the lesion load of these cases presents a similar variability to Vall d'Hebron (even though this hospital has the case with the highest lesion load of all three hospitals with 67.78 cm^3). Looking at all 45 patients, most have a lesion load lower than the mean lesion load for each hospital. This is due some of the cases having an exceptionally higher lesion load than the other cases handled at each hospital (as observed in the box plots).

Finally, all the lesions of each patient were accurately annotated by a trained technician and confirmed by expert radiologists from Hospital Vall d'Hebron. All the annotations were done on the PD-w images and semiautomatically delineated using JIM[©] software¹. These annotations will be used as the ground truth for comparison against our two proposals.

¹Xinapse Systems, JIM software webpage, <http://www.xinapse.com/home.php>.

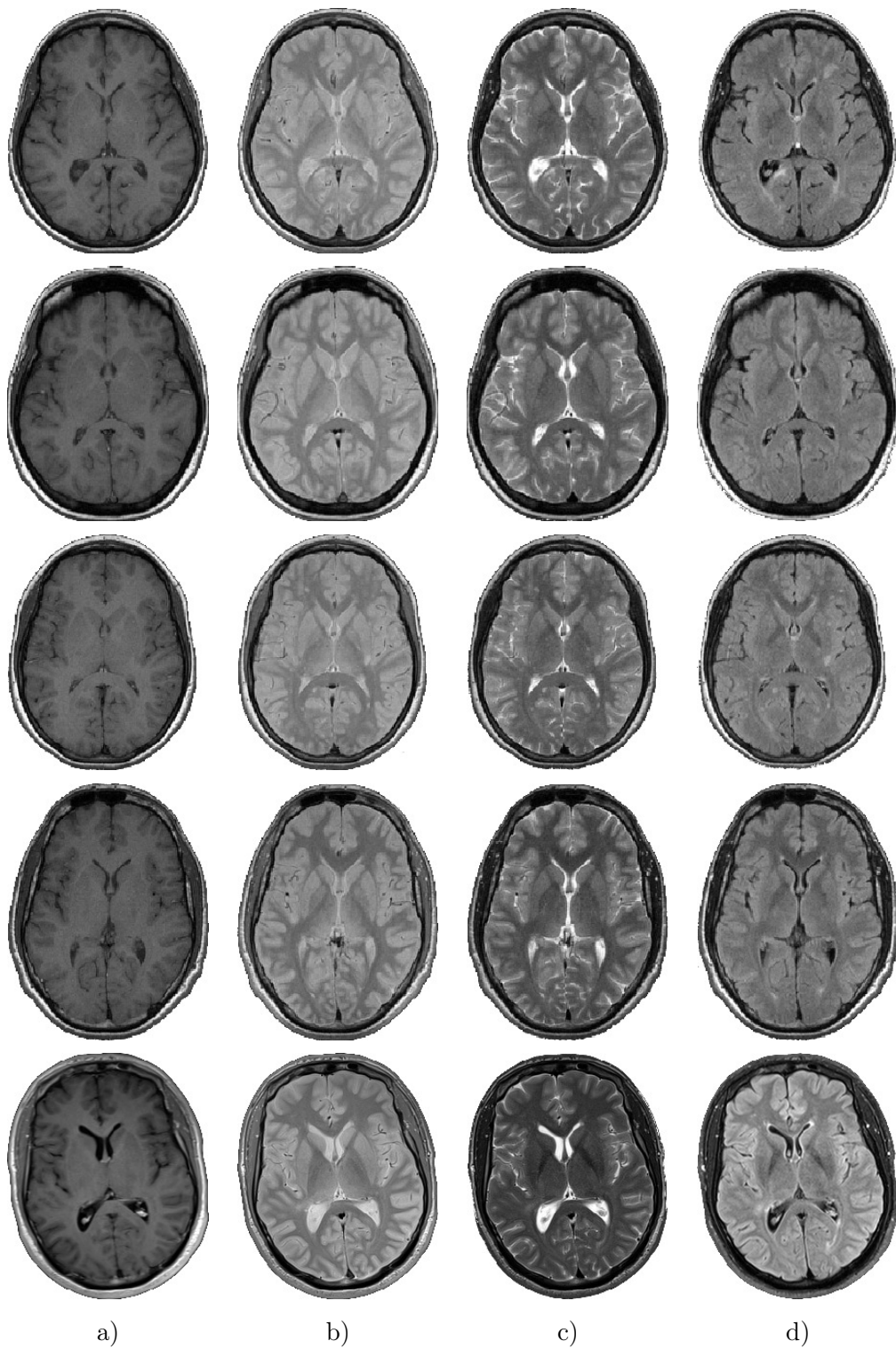


Figure 5.2: Examples of central slices from Hospital Vall d'Hebron: a) T1-w images; b) PD-w images; c) T2-w images; and d) FLAIR images.

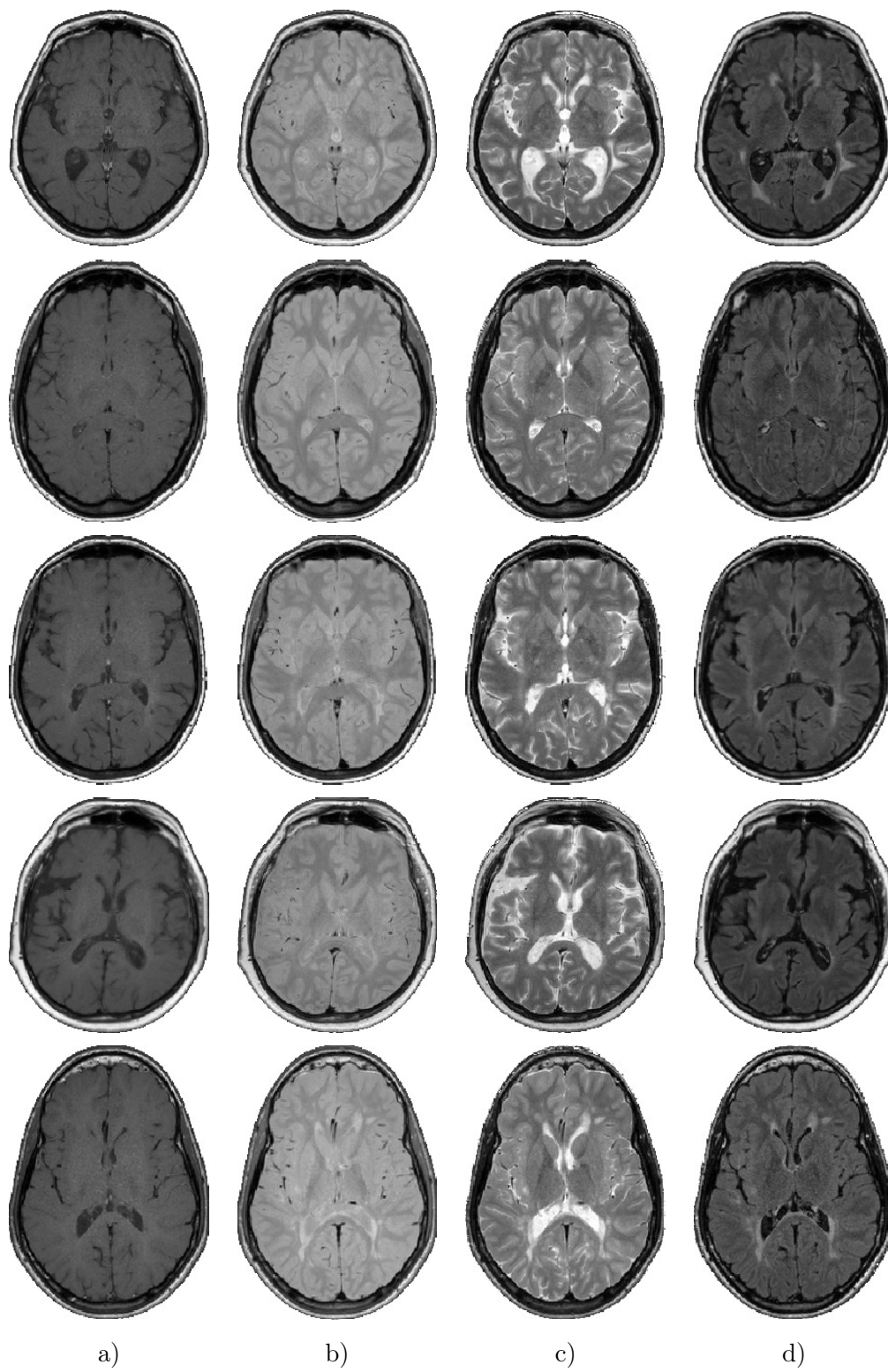


Figure 5.3: Examples of central slices from Hospital Josep Trueta: a) T1-w images; b) PD-w images; c) T2-w images; and d) FLAIR images.

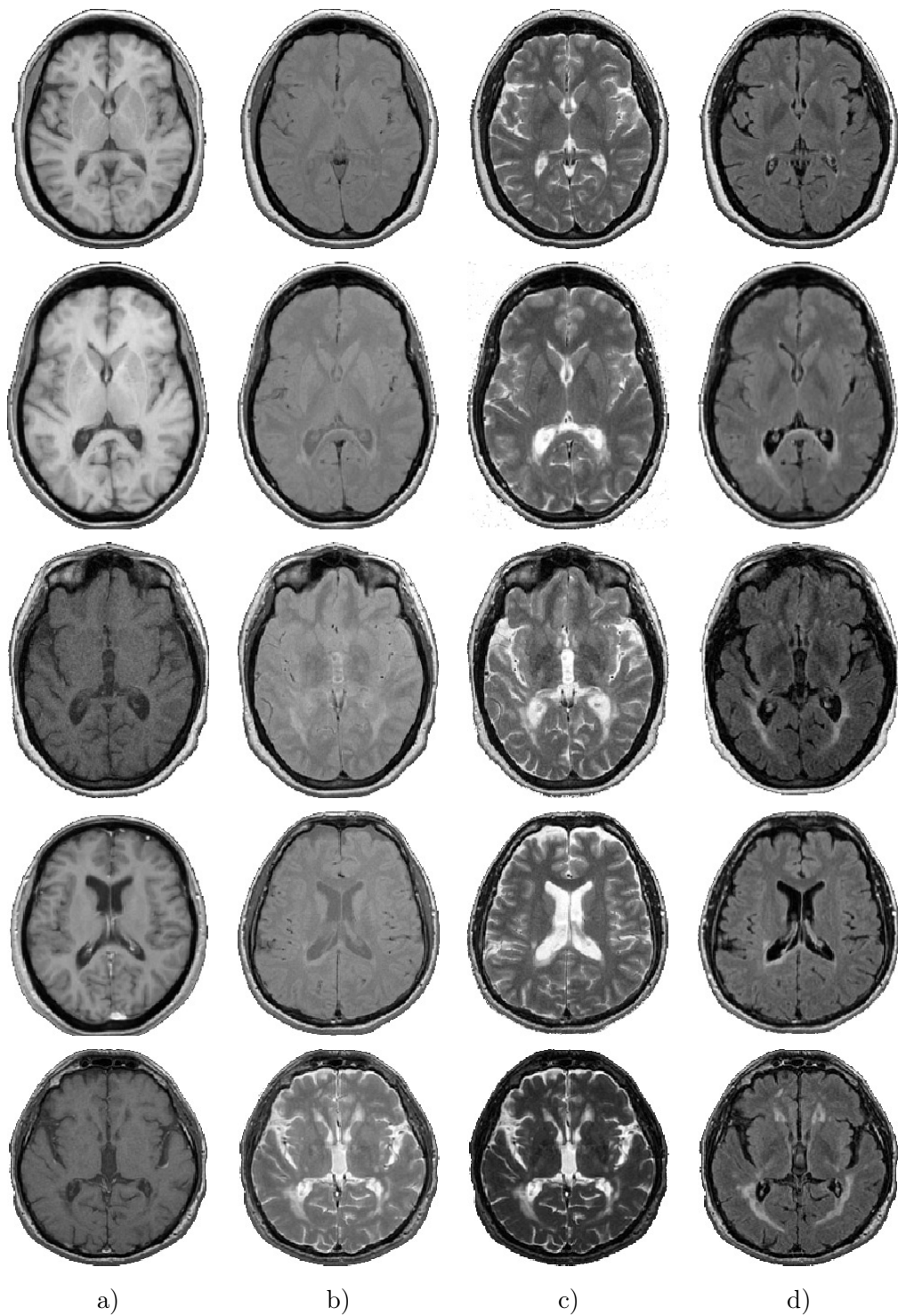


Figure 5.4: Examples of central slices from Clínica Girona: a) T1-w images; b) PD-w images; c) T2-w images; and d) FLAIR images.

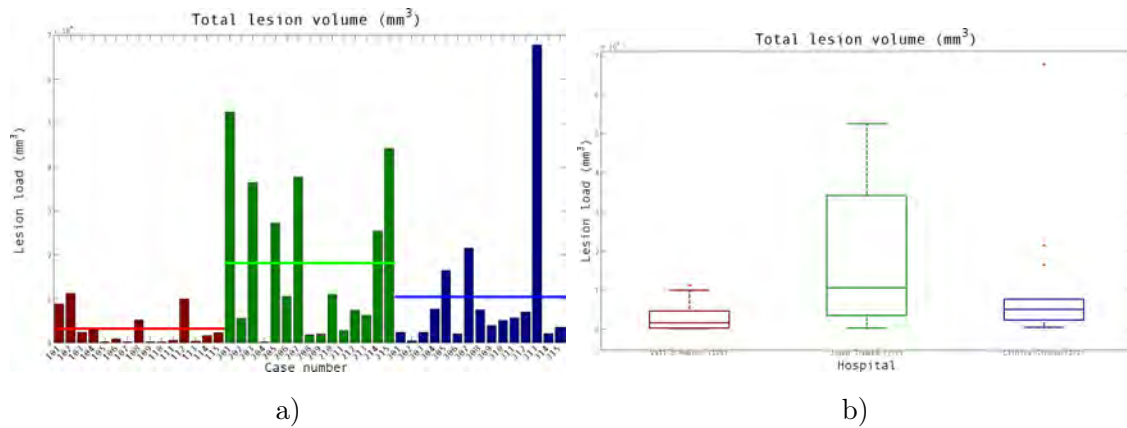


Figure 5.5: Lesion load per hospital: red represents Hospital Vall d’Hebron, green represents Hospital Josep Trueta and blue represents Clínica Girona. a) Bar plot for each patient with the mean lesion load per hospital represented by a line. b) Box plot for the three hospitals.

5.3 Evaluation measures

In chapter 2 we introduced the difference between detection and segmentation. As a reminder, detection is the process of finding a lesion (represented by the overlap of at least one voxel between the ground truth and the automatic segmentation [208]), while segmentation is the process of delimiting the volume of that region. In this section, we will separate the evaluation measures we used according to which of these processes is analysed. On the one hand, we have regionwise measures that evaluate the detection, while on the other hand, we have voxelwise measures that evaluate the segmentation. In this section we will describe some of the measures introduced in chapter 2 according to this distinction [208].

As stated in the previous sections, we assume that manual annotations from a single expert are the ground truth for all the measures presented. In consequence, when validating a segmentation, we are actually evaluating how close it is to the expert’s. However, as pointed out in chapter 2 (see figure 2.7), there might be huge differences if two different experts annotate the same patient. This inter-observer variability is important since some lesions might be missed or oversegmented by the expert.

5.3.1 Detection

First of all, we start describing the measures we use to evaluate the lesion detection. Detection is the most important process in clinical practice since radiologists usually count the number of lesions instead of looking at the total lesion volume. Therefore, it is important to first evaluate if a method can detect all the lesions of a given case. Consequently one common measure is the true positive fraction (TPF):

$$TPF_r = \frac{\#TP_r}{|M_r|}, \quad (5.1)$$

where $\#TP_r$ represents the number of lesions detected by the algorithm that overlap by at least one voxel with a manually annotated lesion, and $|M_r|$ is the total number of manually annotated lesions. This relative measure provides information about how many lesions are detected (and missed): a value of 0 indicates that no lesion was found, while a value of 1 indicates that all the lesions were correctly detected. However, it does not take into account how many of the lesions detected are actually lesions. Therefore, it is also common to estimate the false positive fraction (FPF):

$$FPF_r = \frac{\#FP_r}{|A_r|}, \quad (5.2)$$

where FP_r represents the lesions detected by the algorithm that do not overlap with any manually annotated lesion, and $|A_r|$ is the total number of detected lesions. Similarly to the TPF measure, the FPF is a relative measure. A value of 0 indicates that all the lesions detected are actually lesions, while a value of 1 indicates that all the regions detected are not lesions. In order to complement these two measures, we also compute the Dice similarity coefficient (DSC):

$$DSC_r = \frac{2 \cdot \#TP_r}{|A_r| + |M_r|}. \quad (5.3)$$

This measure gives a high value to the TP detections, while also taking into account missed and wrongly detected lesions, and its values vary between 0 - 1, where 1 denotes the exact similarity and 0 denotes that no lesion was correctly detected. To further ease the interpretation of the DSC, we can utilize the relationship between the DSC and the Kappa coefficient [81]. Zijdenbos et al. showed that under certain assumptions [252] (independent inter-raters and a higher number of TN than the sum of TP, FP and FN) the DSC is asymptotically equal to the Kappa coefficient. According to Landis et al. [131], the

Kappa coefficient values can be divided into six categories: less than 0, “No agreement”, 0-0.2, “Slight agreement”; 0.2-0.4, “Fair agreement”; 0.4-0.6, “Moderate agreement”; 0.6-0.8, “Substantial agreement”; 0.8-1.0, and “Almost perfect agreement”.

5.3.2 Segmentation

Even though we presented the previous measures to validate detection, they can also be used to evaluate the segmentation. The only difference is that instead of looking at detected regions and manually annotated lesions, we have to analyse the voxels labelled as lesion by the segmentation method and those voxels labelled by the expert. Therefore, the previous equations are redefined as follows:

$$TPF_v = \frac{\#TP_v}{|M_v|}, \quad (5.4)$$

$$FPF_v = \frac{\#FP_v}{|A_v|}, \quad (5.5)$$

$$DSC_v = \frac{2 \cdot \#TP_v}{|A_v| + |M_v|}, \quad (5.6)$$

where TP_v represents the voxels labelled as lesion by both the expert and the segmentation, FP_v represents the voxels mislabelled as lesion, $|A_v|$ represents the total number of voxels automatically segmented as lesion and $|M_v|$ represents the total number of lesion voxels in the ground truth.

All these measures provide a relative value of the detection and segmentation’s accuracy. This means that the effect of one misclassified voxel (or region) is quite different between those cases where the lesion load is small and those cases where the lesion load is high. Therefore, to obtain a fair comparison, cases should be grouped according to the total lesion volume.

The next measure we introduce was also used as part of the MS Challenge 2008 workshop. The average surface distance (ASD) provides an evaluation of the segmentation. For each lesion, the voxels in contact with the background (or another tissue) are selected as bordering voxels. This process is applied to both ground truth and automatically segmented lesions. Afterwards, for each bordering voxel of a segmentation (manual or automatic) the closest bordering voxel of the other segmentation is found and the distance between them is computed. Finally, all the distances are averaged to compute the final ASD. A value of 0 indicates a perfect overlap between the ground truth and the segmentation obtained. This

measure does not check if the lesion is detected in both segmentations; therefore, bordering voxels for FP and FN lesions will always obtain values higher than 0.

Finally, we compute a measure related to detection that only involves segmentation. That measure is once again computed using the DSC to evaluate segmentation, taking only TP detected lesions into account. Consequently, for each true lesion, we evaluate its segmentation independently from the others. Afterwards, we average these values for each patient to obtain a measure on how accurate the segmentation is when a lesion is found.

5.4 TISSUE-LOT results

The TISSUE-LOT pipeline presents 2 main steps aimed at two different goals. Initially, a tissue segmentation (including pure CSF, pure GM, pure WM, and PV) is obtained by means of a modified EM for Gaussian distributions guided by the ICBM atlas of 452 patients². Afterwards, a lesion segmentation step, based on thresholding and a posterior refinement, is applied. In this section, we will first discuss the parameters used as part of the lesion segmentation, followed by the analysis of the results obtained in terms of detection and segmentation using the SALEM database.

5.4.1 Parameters

In chapter 4 we introduced the two main steps of this proposal. For the initial tissue segmentation, a convergence parameter was introduced. This parameter, which is a probability threshold, discards voxels with low probabilities to reduce the variability of the estimated Gaussian distributions for pure tissues and improve convergence. Theoretically, a probability lower than 50% would indicate that the voxel has a higher chance of not being part of the tissue, and, therefore, it should not be included during the parameter estimation. Higher probability thresholds provide smaller distributions with high probabilities when the intensities are close to the mean. However, if this threshold is set too high, the real tissue variability might be missrepresented by a small set of the data. Therefore, we propose setting the T^π threshold to 75%, which is an intermediate value.

However, as pointed out previously, this initial tissue segmentation presents a missclassification of lesion voxels due to the lack of a lesion class. As a consequence, a second step to resegment these lesions proved essential. In our case, a thresholding followed by a

²Publicly available at http://www.loni.ucla.edu/ICBM/Downloads/Downloads_Atlases.shtml

refinement based on rules is used to obtain this lesion segmentation.

Thresholding step

In chapter 4, we defined lesions as hyperintense GM outliers in FLAIR. Moreover, we assumed that the pure GM tissue follows a Gaussian distribution and is the most intense, normal appearing tissue in FLAIR. In order to differentiate between lesions and tissues, we proposed using a thresholding technique based on these assumptions where the threshold is obtained using the mean and standard deviation of GM in FLAIR. The equation of this threshold depends on a positive empirical parameter γ .

According to the formal definition of a Gaussian distribution $G(\mu, \sigma)$ and a random variable $\mathcal{Y} \sim G(\mu, \sigma)$, there is a 50% probability that a random sample $y \in \mathcal{Y}$ will have a value higher than μ . This probability is reduced to 2.2% for a value higher than $\mu + 2\sigma$. Extrapolating this to our case, theoretically, 97.8% of GM voxels should have a value under $\mu_{GM}^F + 2\sigma_{GM}^F$, while outlier voxels (not belonging to GM) should present a higher intensity value. Therefore, theoretically, a value greater or equal to 2 could be used for γ .

In practice, not all the lesion voxels are actually outliers (due to partial volume effects or inhomogeneities in extreme slices) and not all the hyperintensities are actually lesions. To fix the second issue, we used the abovementioned parameters as part of a refinement step. On the other hand, to include as many lesion voxels as possible, σ must be set empirically (preferably between 2 and 3). Since the refinement step may mask the real effect of the thresholding parameter, in what follows, we present an exhaustive comparison of the result of varying γ between 0 and 3 without refinement.

Figure 5.6 offers a summary of the effect of varying γ to estimate the threshold in terms of detection and segmentation. For each γ value, a threshold is obtained in all 45 cases which is evaluated using the TPF and FPF measures in terms of detection and segmentation.

Looking at the detection plot, we can see how the FPF slowly decreases as the threshold increases. However, the TPF presents a different tendency: there is an initial increase (until $\gamma = 2$), and then it stabilises. This phenomenon is explained by how the TPF is computed. To estimate the number of detected TP, the automatically segmented lesions are counted only once. Therefore, if a big lesion contains 3 different GT lesions, only 1 TP will be counted. As the threshold increases, this big lesion tends to split into smaller lesions, increasing the detection rate.

On the other hand, looking at the segmentation plot, we can see how both FPF and

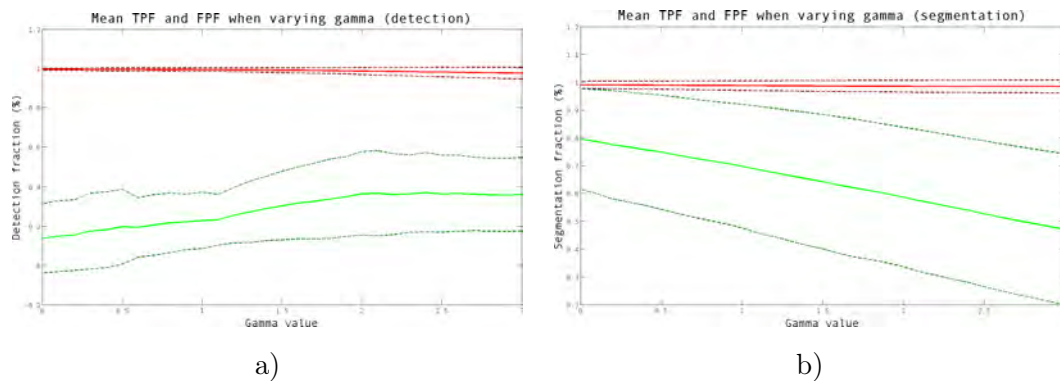


Figure 5.6: Effect of γ during thresholding in terms of a) detection and b) segmentation. For each γ value, the mean TPF (green) and FPF (red) for all cases is plotted with a continuous line, while a standard deviation under and over the mean are plotted in a dashed line. In terms of detection, there is an overall slight decrease of FP, with an initial increase in TP followed by a slight decrease as the threshold increases. However, in terms of segmentation there is an overall decrease of both TP and FP.

TPF decrease as the threshold increases, even though the TPF decreases at a higher rate. This phenomenon is caused by the decrease in number of voxels segmented by the higher threshold. This also proves how most artifacts segmented as lesions are also hyperintense, and hence, more information is needed to differentiate them from real lesion voxels.

Due to $\gamma = 2$ being the highest TPF point in terms of detection, we will set this value as default and reduce the elevated FPF (both in terms of detection and segmentation) using the refinement step.

Refinement step

As mentioned before, lesions are misclassified as a tissue during the first step of this approach. Due to their intensities, lesions tend to be classified as either WM or a GM-related class as demonstrated in figure 5.7. In other words, lesion voxels are rarely classified as CSF (on average only 1.42% of the voxels), as stated in chapter 4. Therefore, the threshold T^{ω_t} , according to the first rule stating that lesion areas should not contain CSF, can be restrictive and can be set to 0.9 to discard those lesions that are completely inside CSF. Moreover, a similar threshold is also used when checking the neighbouring voxels of the lesion areas. As observed in figure 5.7, most of the voxels surrounding lesions are, in fact, WM voxels (with an average of 63.98% and a minimum case of 36.70% of the voxels).

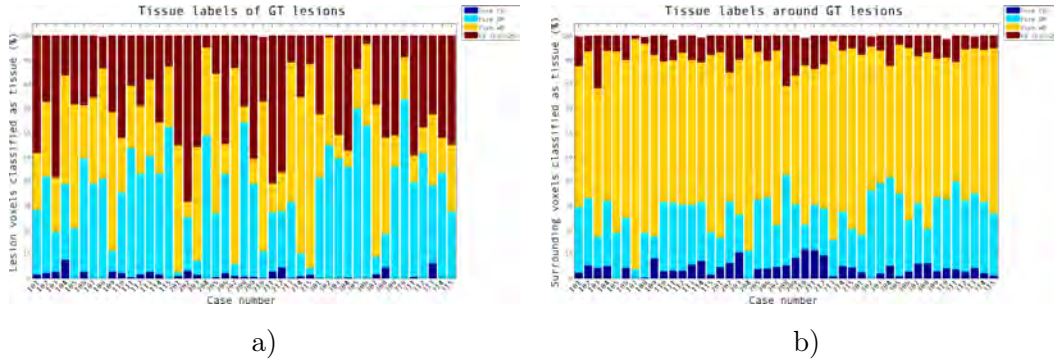


Figure 5.7: Tissue labels for a) GT lesions and b) neighbouring voxels. Each colour represents a tissue label assigned during the EM segmentation. In the first bar plot, we can observe how WML are rarely classified as CSF (with a mean of 1.42%); while in the second plot we observe how neighbouring voxels for GT lesions mostly belong to WM (with a mean of 63.98%).

Hence, a low $T^{\omega_{\mathcal{N}}}$ (0.6 in our experiments), would be adequate to discard voxels completely inside GM or CSF and keep lesions that are close to the ventricles or the cortex.

While the bar plots presented in figure 5.7 comprise only cases from our database, a clear tendency in both phenomena is observed, even though the cases are obtained from three different scanning machines with different lesion load. Consequently, it is expected that new cases should also present a similar behaviour.

The other parameter used as part of the refinement is the minimum size a lesion is allowed. In our experiments, we set this parameter to 10 voxels (representing 30 mm^3 approximately), which would approximately represent a cube with 3 mm edges [24].

5.4.2 Evaluation

To evaluate the method, we will use the parameter values presented in the previous section. Consequently, during the tissue segmentation step, the probability threshold T^{π} will be set to 75%. Afterwards, during the FLAIR thresholding, the parameter γ will be set to 2; thus, only the hyperintense voxels further than 2 deviations from the GM mean in the FLAIR image will be considered as lesions. Finally, the thresholds T^{ω_t} and $T^{\omega_{\mathcal{N}}}$ will be set to 0.9 and 0.6 respectively, meaning that all the lesions having less than 90% of the voxels previously classified as either WM or a GM-based class and are surrounded by less than 60% of WM voxels will be discarded, as well as those lesions with a voxel size lower

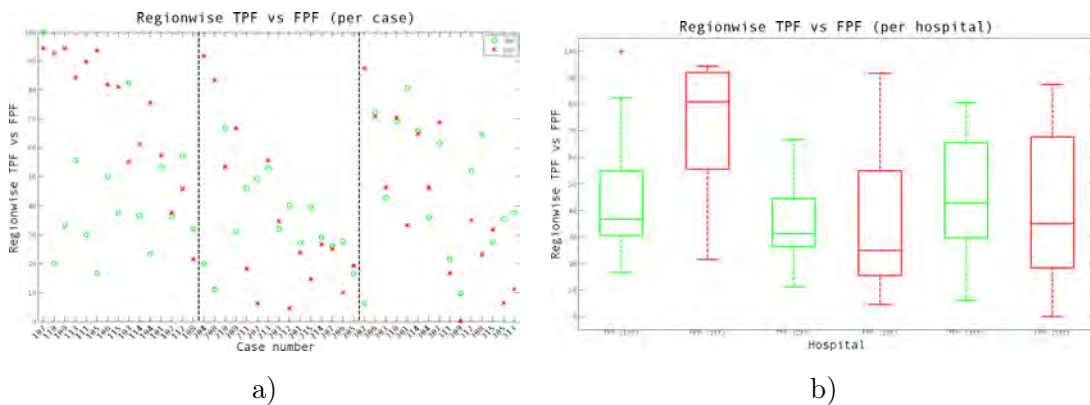


Figure 5.8: True positive fraction vs false positive fraction for the TISSUE-LOT pipeline. a) Comparison between TPF (green circles) and FPF (red crosses) separated by hospitals and ordered by number of GT lesions. b) Boxplot comparison between TPF (green) and FPF (red) separated by hospital.

than 10 (30 mm^3).

First, we will present the results for this pipeline in terms of detection, followed by the results in terms of segmentation with the measures described in section 5.3.

Detection

Even though the thresholding step is a voxelwise process, this pipeline has a strong regionwise component. After the initial segmentation, the refinement step is applied lesionwise in order to reduce FP in terms of detection. As a consequence, this improvement has a higher effect in detection than segmentation (since the segmentation of each lesion is not changed).

Figure 5.8 presents a regionwise comparison between the TPF and FPF in terms of detection for all the cases. The first plot, which is ordered by number of lesions in the GT, presents this comparison for each case separated by hospitals, while the box plot presents the comparison for each hospital.

For all three hospitals, there is a clear tendency for the FPF to decrease as the lesion load increases. On the one hand, this tendency is explained due to the highest probability of correctly detecting lesions when there is a higher number of GT lesions. On the other hand, this tendency is also explained by the number of the measure. By being a proportional measure, the effect of an FP lesion in the measure is higher when there is a small number of

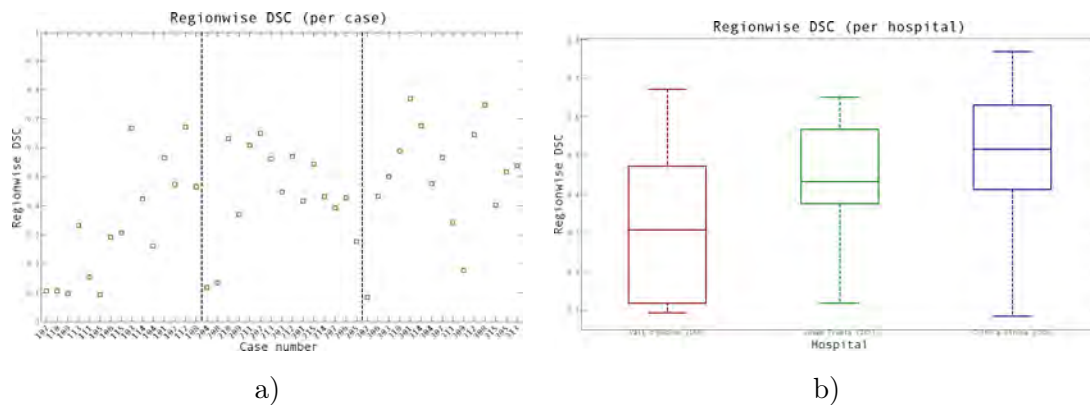


Figure 5.9: Regionwise DSC for the TISSUE-LOT pipeline. a) DSC values (yellow squares) separated by hospitals and ordered by number of GT lesions. b) Boxplot DSC values separated by hospital.

detected lesions (which is indirectly related to the number of real lesions). This tendency can also be observed in the boxplot for Hospital Vall d’Hebron, which has the lowest number of GT lesions (774 as opposed to 1514 and 926 for Hospital Josep Trueta and Clínica Girona respectively). In any case, these results also demonstrate the strength of the refinement step. Looking again at figure 5.6, we can clearly see how the mean FPF for any γ value exceeds 90%, while the mean FPF value for the final segmentation is 49.12%, (71.03%, 35.56% and 40.78% for Hospital Vall d’Hebron, Josep Trueta, and Clínica Girona respectively).

On the other hand, looking at the TPF, there is no clear tendency for this measure for any hospital. Furthermore, when looking at the boxplots for each hospital, the results in terms of TPF are similar for Hospital Vall d’Hebron and Clínica Girona, while Hospital Josep Trueta’s box is smaller, and this hospital obtains a lower mean (34.36% compared to 44.28% and 45.49%) and median (31.25% compared to 36.67% and 42.86%).

Finally, figure 5.9 shows the DSC in terms of detection for all cases. Once again, the first plot presents this measure for each case separated by hospitals and ordered by number of GT lesions, while the box plot presents the measure for each hospital. These two graphics, present a summary of the abovementioned characteristics. Due to the decreasing tendency for FPF as the number of GT lesions increases, as well as the TPF increasing tendency for Vall d’Hebron’s cases, the DSC also increases alongside the number of GT lesions. Moreover, due to the lack of a clear tendency in terms of TPF for the other two hospitals, no tendency is observed for the DSC measure either. Furthermore, Vall d’Hebron’s cases

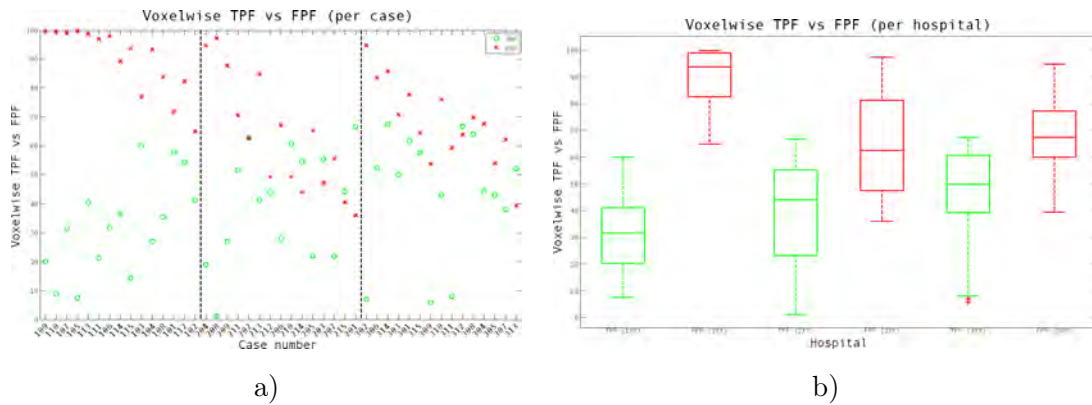


Figure 5.10: Voxelwise TPF vs FPF for the TISSUE-LOT pipeline. a) Comparison between TPF (green circles) and FPF (red crosses) separated by hospitals and ordered by the total lesion load. b) Boxplot comparison between TPF (green) and FPF (red) separated by hospital.

present the lowest DSC mean (0.33) and median (0.31), as well as a wider box, while the other two hospitals present a smaller box with Clínica Girona presenting the highest mean of the two (0.50 as opposed to 0.44) and median (0.52 as opposed to 0.43).

Segmentation

As stated in the detection analysis, this pipeline has a strong focus on detection. In fact, the whole refinement is regionwise and can cause FN lesions that directly affect segmentation measures and decrease accuracy measures, depending on the size of these lesions. Moreover, due to the lack of segmentation refinement for TP or the size of FP lesions, accuracy measures might be further decreased.

Similar to figure 5.8, figure 5.10 presents a voxelwise comparison between the TPF and FPF in terms of segmentation for all the cases. The first plot, which is ordered by the total lesion load instead of the number of GT lesions, presents this comparison for each case separated by hospitals, while the box plot presents the comparison for each hospital.

Looking at the plot for each case, a tendency to decrease of the voxelwise FPF is also globally observed according to the lesion load. This decrease is strongly related to the FPF decrease in terms of detection. By removing FP lesions, many FP voxels are also removed. However, the mean FPF segmentation percentages are higher than the detection percentages for each hospital: Hospital Vall d'Hebron, which obtained a 71.03% for detec-

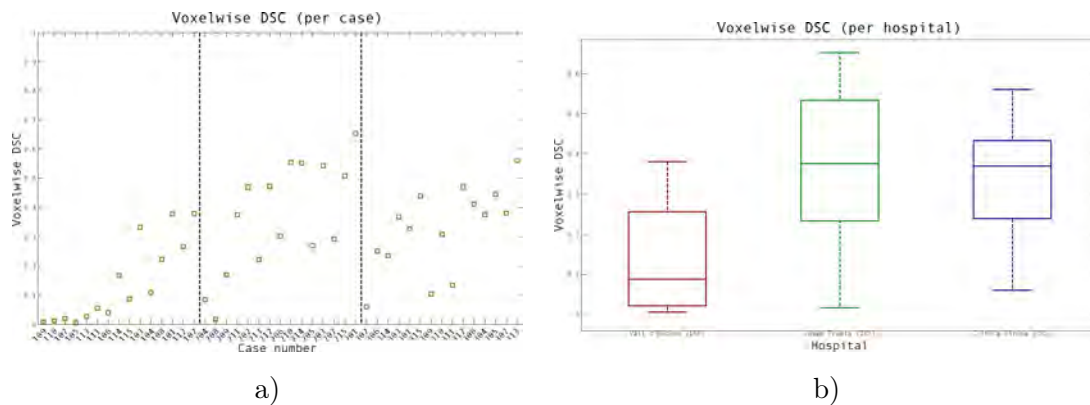


Figure 5.11: Voxewise DSC for the TISSUE-LOT pipeline. a) DSC values (yellow squares) separated by hospitals and ordered by total lesion load. b) Boxplot DSC values separated by hospital.

tion, presents a 89.78% for segmentation; Hospital Josep Trueta, which obtained a 35.56% for detection, presents a 63.40% for segmentation; and Clínica Girona, which obtained a 40.78% for detection, presents a 68.11%. In terms of TPF, only Vall d'Hebron presents a clear tendency to increase when the total lesion volume increases. This can be explained by the small lesion load of most of this hospital's cases. However, no tendency is observed for the other two hospitals.

According to these two segmentation measures, and looking at the box plot, we can see that the best ranking hospital in terms of TPF is Clínica Girona with the highest mean (44.05% compared to 32.52% and 39.97% for Vall d'Hebron and Josep Trueta respectively) and median (44.06% compared to 32.52% and 39.97%). On the other hand, the best ranking hospital in terms of FPF is Hospital Josep Trueta with the lowest mean and median (62.50% compared to 93.66% and 67.44% for Vall d'Hebron and Clínica Girona respectively). For both measures, Vall d'Hebron's cases obtain the lowest scores, partially due to the low lesion load.

Closely related to these two measures, the DSC values for each case, separated by hospital, are presented in figure 5.11. Similar to figure 5.9, the results observed in terms of voxelwise TPF and FPF can be summarised by this measure. For instance, we see there is a tendency for DSC to increase according to the lesion load for each hospital. This is partially due to the decrease in FPF, but is also strengthened for Hospital Vall d'Hebron by the increase in the TPF. Furthermore, we observe how the worst ranking hospital for both TPF and FPF also obtains the worst ranking in terms of DSC, while the other two

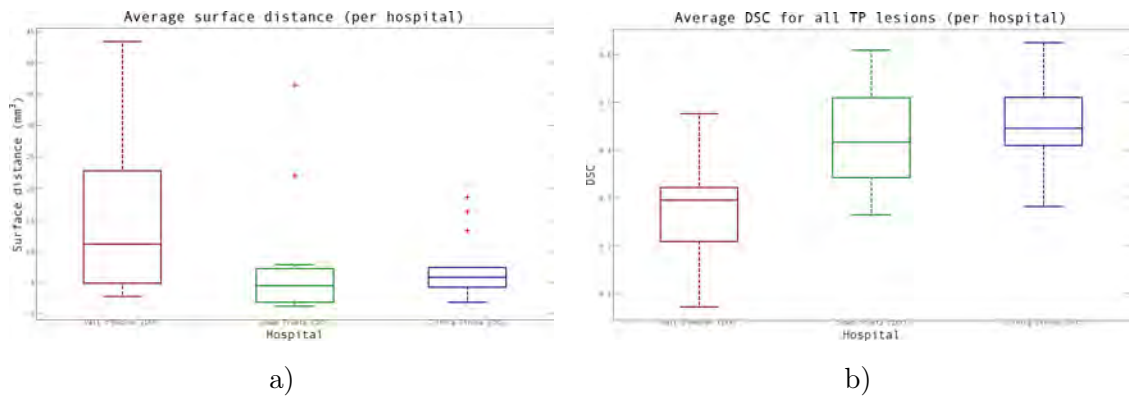


Figure 5.12: Average surface distance and average DSC for all lesions. a) Box plot for the average surface distance separated by hospitals and b) box plot for the average DSC for all TP lesions separated by hospital.

hospitals obtain similar results: Vall d'Hebron scoring a 0.14 mean and 0.08 median; Josep Trueta scoring a 0.36 mean and 0.38 median; and Clínica Girona scoring a 0.32 mean and 0.37 median. According to the DSC-Kappa relationship, these values represent a slight agreement between the TISSUE-LOT pipeline and the GT. Moreover, as observed in 5.11, for some cases, we obtain results that would be considered a fair and/or moderate agreement between both the manual and automatic segmentation (with a maximum value of 0.65).

Finally, figure 5.12 shows the box plots for the last two segmentation measures, loosely related to detection. These boxplots reinforce the results observed in the DSC segmentation values. On the one hand, Hospital Vall d'Hebron scores the highest average surface distance (with a wide box) and the lowest average DSC per TP lesion. On the other hand the boxes for the other two hospitals are similar to each other, with only slight differences in their mean and median for both measures.

Looking at each of these two measures in general, we observe how the average surface distance obtains a 10.10 mm mean and a 6.69 mm median. This difference is mainly caused by the bias of FP and FN detections having a higher impact in Vall d'Hebron's cases. On the other hand, looking at the DSC for the TP detection, we observe a mean DSC of 0.38 and a median value of 0.40. This measure reinforces the computation of the TPF when there is only 1 voxel overlap since, according to kappa, there is a fair agreement between the expert annotations and the TISSUE-LOT pipeline.

5.5 BOOST results

The BOOST pipeline presents a supervised strategy based on training a gentleboost ensemble classifier. This proposal also includes the computation of several novel features to characterise the lesions we want to segment. First, we will discuss the parameters and how we conducted the experiments, followed by the analysis of the results obtained in terms of detection and segmentation.

5.5.1 Parameters

In chapter 4, we introduced this proposal as a training-based approach with a rich pool of features. As features, we introduced a boosting map and a set of randomly sampled contextual meta-features. The number of computed features is one of the parameters to take into account and is related to the number of rounds used to train the classifier. Moreover, in order to train the classifier, a number of lesion voxels and non-lesion voxels must be randomly sampled. The number of voxels for both sets is also a parameter to optimise. The outlier of these parameters, as well as the effect of the optimisation map and the strategy used to evaluate the results, will be discussed in this section.

Once the training is finished, the new image is tested and a map defining the likeliness of each voxel to be a lesion is obtained. This map, which is not a binary mask, must be thresholded in order to obtain a final lesion mask. In this section, we will also study how we obtain this threshold.

Outlier map

In a previous work [38], we evaluated the effect of using our outlier map as a feature. To this effect, we performed 2 different experiments: one using only image intensities, atlas probabilities, and 800 context features, and a second one where we extended the feature pool by adding the outlier map as a feature and as part of the context feature computation described in chapter 4. For both experiments, we used a leave-one-out strategy for each hospital independently.

For this study, only the patients at Hospital Vall d'Hebron and Hospital Josep Trueta were used and the results obtained in terms of the DSC measure are summarised in figure 5.13. In this plot, the images are ordered in terms of lesion load in ascending order. Therefore, the last image is the one with the highest lesion load. On average, when using

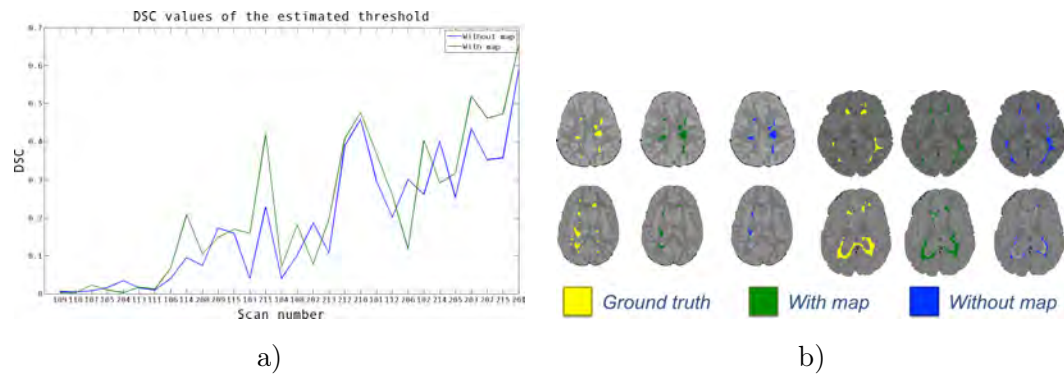


Figure 5.13: Comparison between using (green) or not using (blue) our outlier map as a feature. a) DSC values for the images from Hospital Vall d'Hebron and Hospital Josep Trueta, with the cases ordered by the total lesion load. b) Qualitative comparison between using the outlier map (green), training the classifier without the map (blue) and the GT (yellow).

the lesion map, DSC values were higher (with the exception of those cases with a low lesion load), pointing to a better training and classification process when this feature is included.

Training parameters

In [38], we also presented how to compute a threshold for the map obtained after using the trained classifier on a new case. To obtain this threshold, we tested the classifier with the training images and computed the optimal threshold for each one according to the Dice similarity coefficient (DSC) value. Afterwards, the mean of those thresholds was used on the new patient's output.

To optimise the rest of the parameters of our approach, we perform several tests on the number of contextual meta-features for the training, the number of boosting rounds, and the number of pixels selected for training. In order to evaluate each parameter individually, we fixed the values of the other parameters and repeated each experiment ten times. As expected, better results were obtained for all these parameters when we increased them, although we realised that from a certain point, this increment was almost inappreciable. However, in all the cases, it also produced a huge increment in the computational cost, so it is important to find the optimal parameters to maximise the performance without increasing the computational cost too much. After evaluating the results, we decided to set the number of contextual meta-features at 800 and select a total of 10.000 and

30.000 samples for lesion and not lesion from all the training images. It is also important to mention that the performance of the results also increased with a higher number of boosting rounds during the training process. However, after 400 rounds the improvement was almost inappreciable, while the computational cost increased drastically.

Experiments

In order to perform the experiments, we reproduced all the steps that involved a random sampling. During the feature extraction step, we created 5 different feature sets for each experiment (with random R_1 and R_2 regions) per hospital. 800 different contextual meta-features were computed for each case and feature set.

During the training step, for each training set containing 14 images from the same hospital, we randomly sampled 5 different sets of positive and negative examples. These samples were then used to train the classifier with a leave-one-out strategy. Finally, this classifier was used on the testing image to perform the segmentation. Therefore, for each patient we obtained 25 different classification maps that were thresholded to obtain the final segmentation. In the following sections, we will present the results obtained for each measure for these segmentations.

5.5.2 Evaluation

To evaluate the method, we used the parameter values presented in the previous section. Consequently, the gentleboost algorithm was trained with a rich feature pool of 4 images (T1-w, PD-w, T2-w, and FLAIR), 3 probabilistic maps (CSF, GM, WM), our outlier map, and 800 contextual meta-features (for a total of 808 features). This training was performed using a leave-one-out strategy with a random sampling of 10.000 positive examples and 30.000 negative examples for 400 rounds. Afterwards, this classifier was tested against these training images to compute an optimal threshold for each case's map and was also tested with the new image to obtain a lesion map. Finally, the optimal thresholds will be averaged to obtain a final threshold to compute our final lesion mask.

First, we will analyse the features chosen during the training step. Subsequently, we will present the results for this pipeline in terms of detection, followed by the results in terms of segmentation with the measures described in section 5.3.

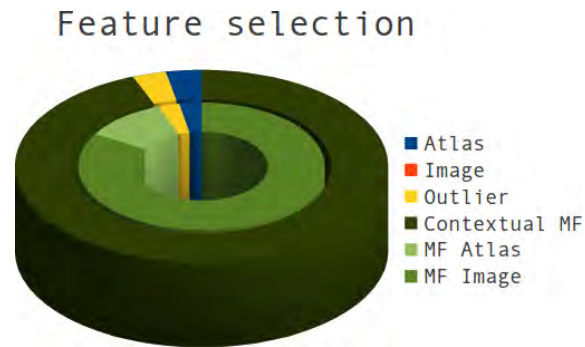


Figure 5.14: Pie chart of the boosting selected features. The outer circle presents the features according to their topology: atlas features (blue), image features (orange) and the outlier map (yellow). The inner circle presents the same features distinguishing between contextual features based on atlas (light green) or image (darker green).

Feature analysis

One of the advantages of using a gentleboost classifier is the possibility of obtaining an ordered list of the features selected. The order determines the importance of the selected feature. Accordingly, the first feature would be the most discriminative one, while the last one is usually focused on a small subset of all the voxels. In all the experiments, we observed that the first feature is always the outlier map. This statement reinforces the importance of the outlier map within the framework and emphasises its strength within the framework.

However, the outlier map as a feature is only used, on average, in 2.90%, 5.70% and 4.50% of the rounds for Vall d'Hebron, Josep Trueta and Clínica Girona respectively. Similar percentages are observed for probabilistic maps, while image features are rarely used and the majority of the features are contextual, as illustrated by figure 5.14. If we further study the use of contextual features, we observe how the most used ones are image-based, while atlas-based contextual features have a higher percentage than the sum of all the non-contextual features.

Finally, from the rich pool set of 808 features, only 250 unique features on average for all hospitals are selected during the 400 rounds. This fact, and the high usage percentages, emphasise the importance of using a rich feature pool and a large set of contextual meta-features.

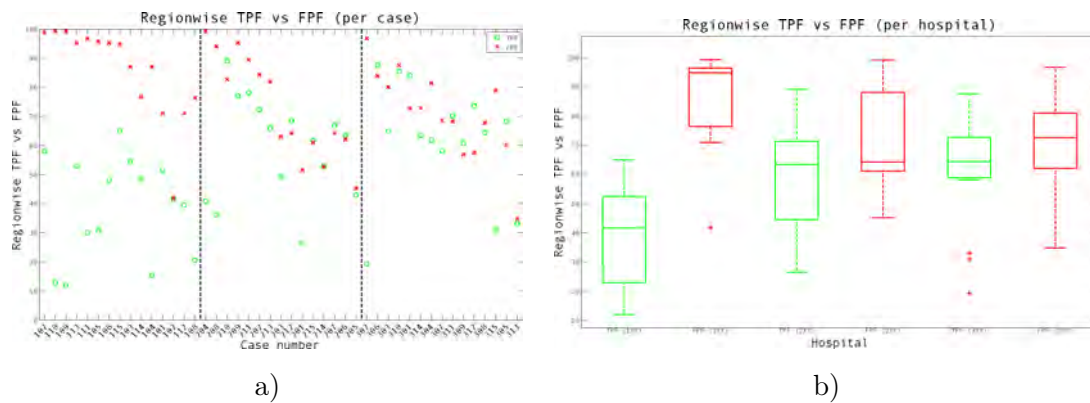


Figure 5.15: True positive fraction vs false positive fraction for the BOOST pipeline. a) Comparison between TPF (green circles) and FPF (red crosses) separated by hospitals and ordered by number of GT lesions. b) Boxplot comparison between TPF (green) and FPF (red) separated by hospital.

Detection

The BOOST pipeline is a supervised approach that relies solely on voxel features. Therefore, there is no regionwise analysis of any sort, which in turn affects the results in terms of detection. For instance, figure 5.15 presents a comparison between the regionwise FPF and TPF for each case, grouped by hospital. Looking at the first plot, we observe similar trends to those present in figure 5.8. On the one hand, there is a clear tendency for the FPF to decrease as the number of GT lesions increases, mainly for Hospital Josep Trueta and Clínica Girona. There is also a tendency to decrease for Hospital Vall d’Hebron, with a smaller incline due to the lower number of lesions for this hospital’s cases. Moreover, there is no clear tendency for the TPF according to the lesion load.

Looking at the box plot, there is also a clear difference between the boxes of the two last hospitals and Vall d’Hebron, again mainly due to the number of GT lesions. On the one hand, the TPF boxes for both Josep Trueta and Clínica Girona are arguably similar (when including the outliers) with a mean and median of 59.42% and 63.50% for the first hospital and 61.72% and 64.37% for the second, while Vall d’Hebron demonstrates a lower performance with a mean and median value of 38.73% and 41.60% respectively. On the other hand, we observe a high FPF for the three hospitals with means and medians of 85.72% and 94.78% for Vall d’Hebron; 72.69% and 64.27% for Josep Trueta; and 71.15% 72.66% for Clínica Girona respectively.

To complement these two measures, figure 5.16 illustrates the regionwise DSC for each

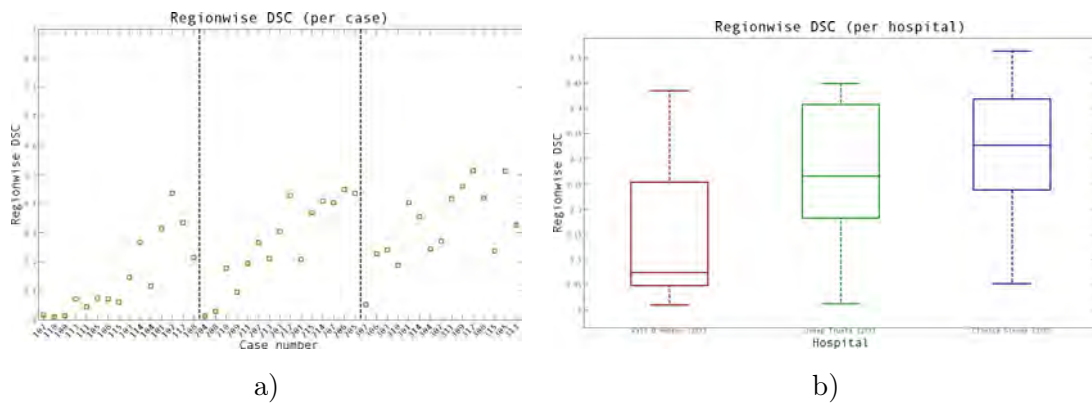


Figure 5.16: Regionwise DSC for the BOOST pipeline. a) DSC values (yellow squares) separated by hospitals and ordered by number of GT lesions. b) Boxplot DSC values separated by hospital.

patient and per hospital. As expected, there is a clear positive tendency in the DSC as the number of GT lesions increases in the first plot. This is a direct consequence of the decreasing tendency in the FPF. This also implies the elevated weight of the FP detections for the regionwise DSC. Looking at the bigger picture, we observe a similar tendency in the boxplots where the cases of Hospital Vall d'Hebron obtain the lowest results, with a mean and median of 0.15 and 0.07 respectively, and the cases of Josep Trueta, with a mean and median of 0.27, and Clínica Girona, with a mean and median of 0.32 and 0.33 respectively, both obtaining similar boxes, with fairly lower results for the first hospital, and a higher variability (illustrated by a bigger box). Taking the DSC-Kappa relationship into account, these results present a fair agreement for the two hospitals with the higher number of lesions.

Segmentation

As stated in the detection analysis, this pipeline is based purely on segmentation. Therefore, its strengths are highlighted in the segmentation analysis, even though the previous analysis demonstrated a fair agreement between the experts and the automatic detection.

First we start with the comparison between the voxelwise TPF and FPF illustrated in figure 5.17. As expected, there is a clear tendency to decrease for FPF as the lesion load increases, while the TPF values do not present a clear tendency for any hospital. Once again, this is due to the decreasing effect of small voxelwise errors as the total number of lesion voxels increases. This concept is summarised by the box plots for each hospital. On

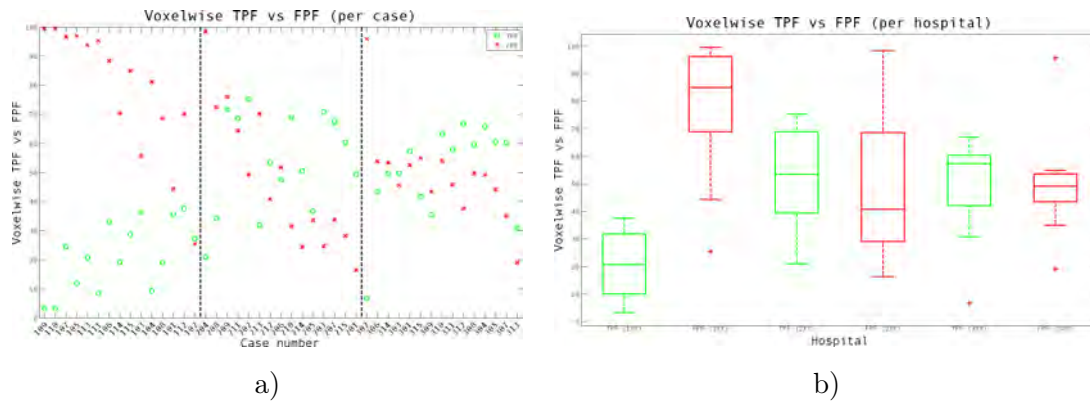


Figure 5.17: Voxelwise TPF vs FPF for the BOOST pipeline. a) Comparison between TPF (green circles) and FPF (red crosses) separated by hospitals and ordered by the total lesion load. b) Boxplot comparison between TPF (green) and FPF (red) separated by hospital.

the one hand, we observe how the TPF boxes for Josep Trueta, with a mean and median of 53.87% and 53.84% respectively, and Clínica Girona, with a mean and median of 49.90% and 51.92%, are similar (with a higher variability for the first hospital), while the box for Vall d'Hebron presents lower values, with a mean and median of 21.21% and 20.72%. On the other hand, the FPF boxes for Josep Trueta, with a mean and median of 47.70% and 40.79%, and Clínica Girona, with a mean and median of 48.89% and 49.15%, are similar (with the Josep Trueta's box, also presenting a higher variability) with lower values than the Vall d'Hebron's box, with a mean and median of 78.04% and 84.94%.

These behaviours are summarised with the complementary DSC plots from figure 5.18. Once more, a tendency to increase in respect to the lesion load can be observed in the first plot, closely related to the TPF trend. This phenomenon emphasises the decrease in difficulty when the lesion load increases. Furthermore, a closer look at the box plot for each hospital further reveals these assumptions. Yet again, the boxes for Josep Trueta, with a mean and median of 0.50 and 0.56, and Clínica Girona, with a mean and median of 0.49 and 0.52, present similar values (with a larger variability for the first hospital) that are higher than those presented by the Vall d'Hebron's box, with a mean and median of 0.19 and 0.17. According to the DSC-Kappa relationship, there is moderate agreement in terms of segmentation between the expert annotations and the BOOST masks for the two hospitals with the higher lesion loads.

Finally, a box plot analysis for the last two segmentation measures is presented in figure 5.19. Both box plots reinforce the ideas observed in the previous plots from another

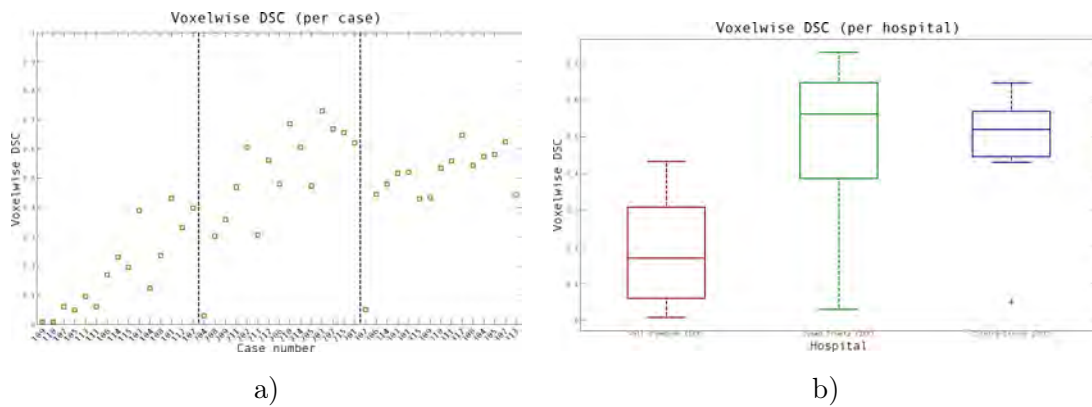


Figure 5.18: Voxelwise DSC for the BOOST pipeline. a) DSC values (yellow squares) separated by hospitals and ordered by the total lesion load. b) Boxplot DSC values separated by hospital.

perspective.

On the one hand, looking at the average surface distance for each hospital, we observe low values and thin boxes for Josep Trueta, with a mean and median of 4.26 and 1.72 mm respectively, and Clínica Girona, with a mean and median of 3.49 and 2.29 mm, respectively, and higher distances for Vall d'Hebron, with a mean and median of 11.47 and 8.65 mm respectively. According to the slice thickness of 3 mm, the surface difference of the two hospitals with the higher lesion loads is close to the size of a voxel in the z coordinate.

On the other hand, as stated in the detection analysis, there is only fair agreement in terms of detection between the expert annotations and our BOOST pipeline. However, looking at the segmentation of the TP detection, we observe how the values of the boxes for the three hospitals imply a moderate agreement, where there is a close tie between Josep Trueta's box, with a mean and median of 0.51 and 0.50 respectively, and Clínica Girona's box, with a mean and median of 0.49 and 0.50; and, once more, Vall d'Hebron's box presents lower values, with a mean and median of 0.33.

5.6 Strategy comparison

In the previous sections, we presented the results for each pipeline separately in order to analyse their accuracy, strengths and weaknesses independently. However, a further comparison is required to fully comprehend the strengths and weaknesses of both approaches

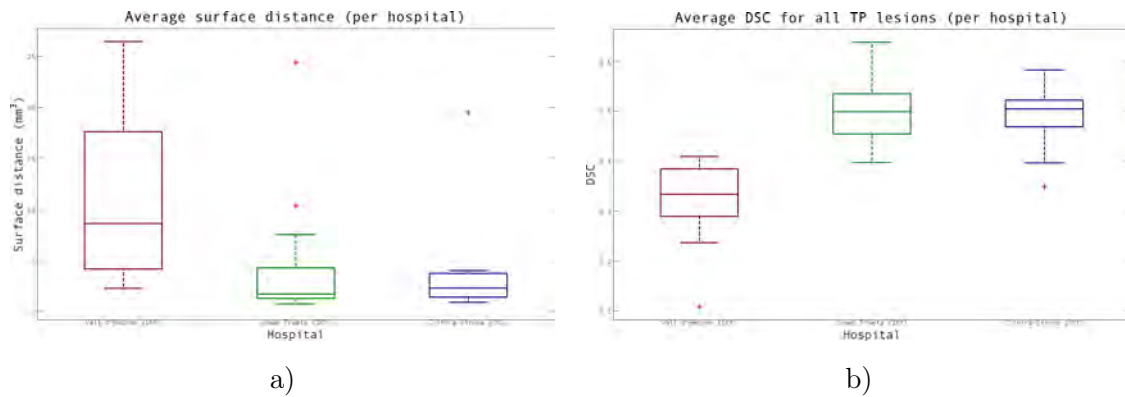


Figure 5.19: Average surface distance and average DSC for all lesions for the BOOST pipeline. a) Box plot for the average surface distance separated by hospitals and b) box plot for the average DSC for all TP lesions separated by hospital.

when choosing one over the other.

Furthermore, no baseline method has been used to validate any of these pipelines. As stated in chapter 2 and section 5.1 of this same chapter, there is no baseline method for MS lesion segmentation, even though we highlighted some promising approaches. One such approach was presented by Souplet et al. [204] as part of the 2008 segmentation challenge, and is described in detail in chapter 4 due to its similarity to our TISSUE-LOT pipeline. This approach was implemented as part of the SepINRIA software tool³, allowing us to provide a comparison with our two own pipelines.

Therefore, in this section we will present a comparison between the TISSUE-LOT pipeline, the BOOST pipeline and Souplet et al.'s [204] strategy applied to our SALEM database. This comparison will be presented quantitatively with the evaluation measures presented in this chapter, and qualitatively with commented visual examples of the three approaches.

5.6.1 Quantitative comparison

Following the structure of the previous sections and starting with the detection analysis followed by the segmentation, figure 5.20 summarises the regionwise results for the three approaches. As expected, the best results in terms of detection are obtained using the TISSUE-LOT pipeline, which has a regionwise refinement step. While the TPF is lower

³SepINRIA software webpage, <http://www-sop.inria.fr/asclepios/software/SepINRIA>.

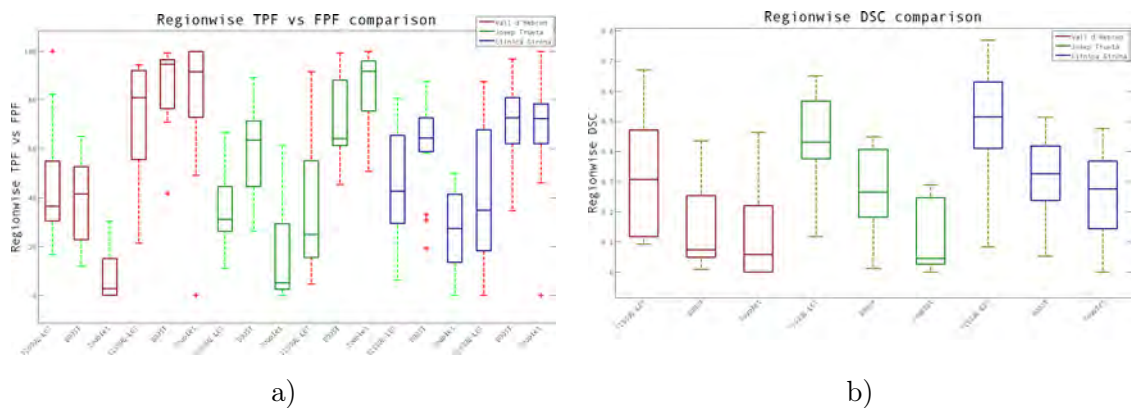


Figure 5.20: Comparison in terms of detection: a) TPF (green) vs FPF (red) plot, and b) DSC (yellow) plot. Each hospital's box is outlined with a different color: Hospital Vall d'Hebron in dark red, Hospital Josep Trueta in dark green and Clínica Girona in dark blue.

than the values obtained by the BOOST strategy, the FPF is also lower, causing a higher overall DSC for each hospital, as illustrated by the second plot. This is due to the impact of a large number of FPF on the DSC measure.

Looking at both images, it is also clear that Souplet et al.'s approach obtained the worst results for each measure, with low TPF values and high FPF ones. This approach is usually restrictive, causing low accuracy and a high number of FN. As a consequence, most of the detected lesions are actually artifacts, which in turn causes a high FPF. This issue is further accentuated when analysing the DSC measures. Part of this lower performance might also be explained by the fact that Souplet et al.'s strategy was optimised for 3T images.

Following with the segmentation results, figure 5.21 summarises the voxelwise measures for the three strategies. While a better detection was expected for the TISSUE-LOT pipeline, the plots from this figure present a higher accuracy for the BOOST pipeline in terms of segmentation. As stated, this pipeline is trained to find lesion voxels from a rich feature pool with no regionwise refinement. As a consequence, a higher TPF is obtained for the hospitals with higher lesion loads and a lower FPF in terms of segmentation. In turn, this is reflected in higher DSC values for all three hospitals for this approach.

A further analysis of these results also reveals that Souplet et al.'s approach obtains the lowest TPF results, even though the DSC median, minimum, and maximum are similar to those obtained by the TISSUE-LOT pipeline for Hospital Vall d'Hebron and Clínica Girona. In the case of the first hospital, this similarity is closely related to the lower

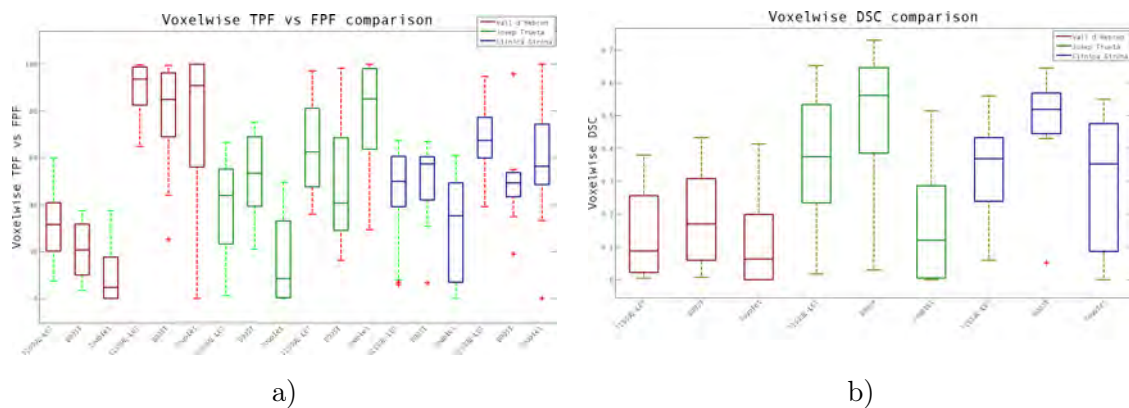


Figure 5.21: Comparison in terms of segmentation: a) TPF (green) vs FPF (red) plot, and b) DSC (yellow) plot. Each hospital's box is outlined with a different color: Hospital Vall d'Hebron in dark red, Hospital Josep Trueta in dark green and Clínica Girona in dark blue.

lesion load, since all three methods obtained scores below 0.2 for the DSC. However, when analysing the third hospital, there is a higher variability for the voxelwise DSC obtained by the SepINRIA toolbox, illustrated by the wider box.

Finally, to complement the previous results, figure 5.22 illustrates the comparison of the average surface distance and average DSC for all TP lesions. The first plot suggests a better segmentation for all TP lesions when using Souplet et al.'s approach with the images from Hospital Vall d'Hebron and Clínica Girona. These results demonstrate that, even though the method is restrictive (excluding a large number of lesions as FN), the lesions detected are accurately segmented. The second plot emphasises that idea since the average surface distance is higher than 0 for FN and FP detections by definition. Therefore, this approach obtains an elevated surface distance for the three hospitals. Furthermore, when looking at both plots, Souplet et al.'s boxes imply a high variability for both measures, illustrated by wide boxes.

Looking further at the difference between the two proposals described in chapter 4, both plots reinforce the idea that the BOOST pipeline obtains the best results in terms of segmentation. Not only do the boxes present a lower variability, but the DSC values are higher, and the surface distance is lower. As stated in the analysis of figure 5.21, this second pipeline is capable of segmenting the lesions better, even though the accuracy in terms of detection is lower.

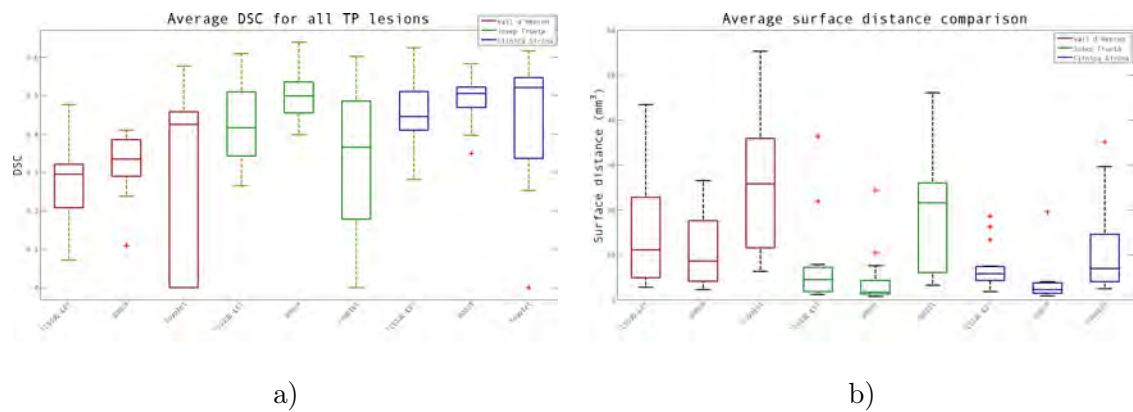


Figure 5.22: Comparison in terms of segmentation by hospital and method. a) Average DSC for all TP lesions, and b) average surface distance plot. Each hospital’s box is outlined with a different color: Hospital Vall d’Hebron in dark red, Hospital Josep Trueta in dark green and Clínica Girona in dark blue.

5.6.2 Qualitative comparison

After the quantitative comparison described in the previous section, and a state-of-the-art approach, figures 5.23, 5.24, and 5.25 illustrate the conclusions observed with a variety of visual examples comprising different lesion loads. In order to reinforce the analysis in the previous section, we will further comment these proposed examples.

One of the first ideas we introduced was the low detection accuracy obtained by Souplet et al.’s approach due to its restrictive tendency. This is generally observed for all three hospitals in the majority of the exposed examples. This approach has a lower number of detected lesions whether correctly detected or not. For some cases, no lesion is detected, as shown in figure 5.20 by the minimum values for this method.

Another general idea is the better detection from the TISSUE-LOT pipeline, which is closely related to the FP reduction. As observed in most of the images, the BOOST pipeline usually finds small spurious lesions that highly increase the FPF in terms of detection and have a small effect on terms of segmentation due to their size (see the fourth row of figure 5.23, the first row of figure 5.24, or the last row of figure 5.25). As a consequence, the TISSUE-LOT obtains better detection DSC values since, as stated, the detected TPF for Souplet et al.’s approach is low.

On the other hand, the small FP lesions detected by the BOOST do not have much impact on the overall segmentation (or the segmentation of TP, for that matter). Therefore,

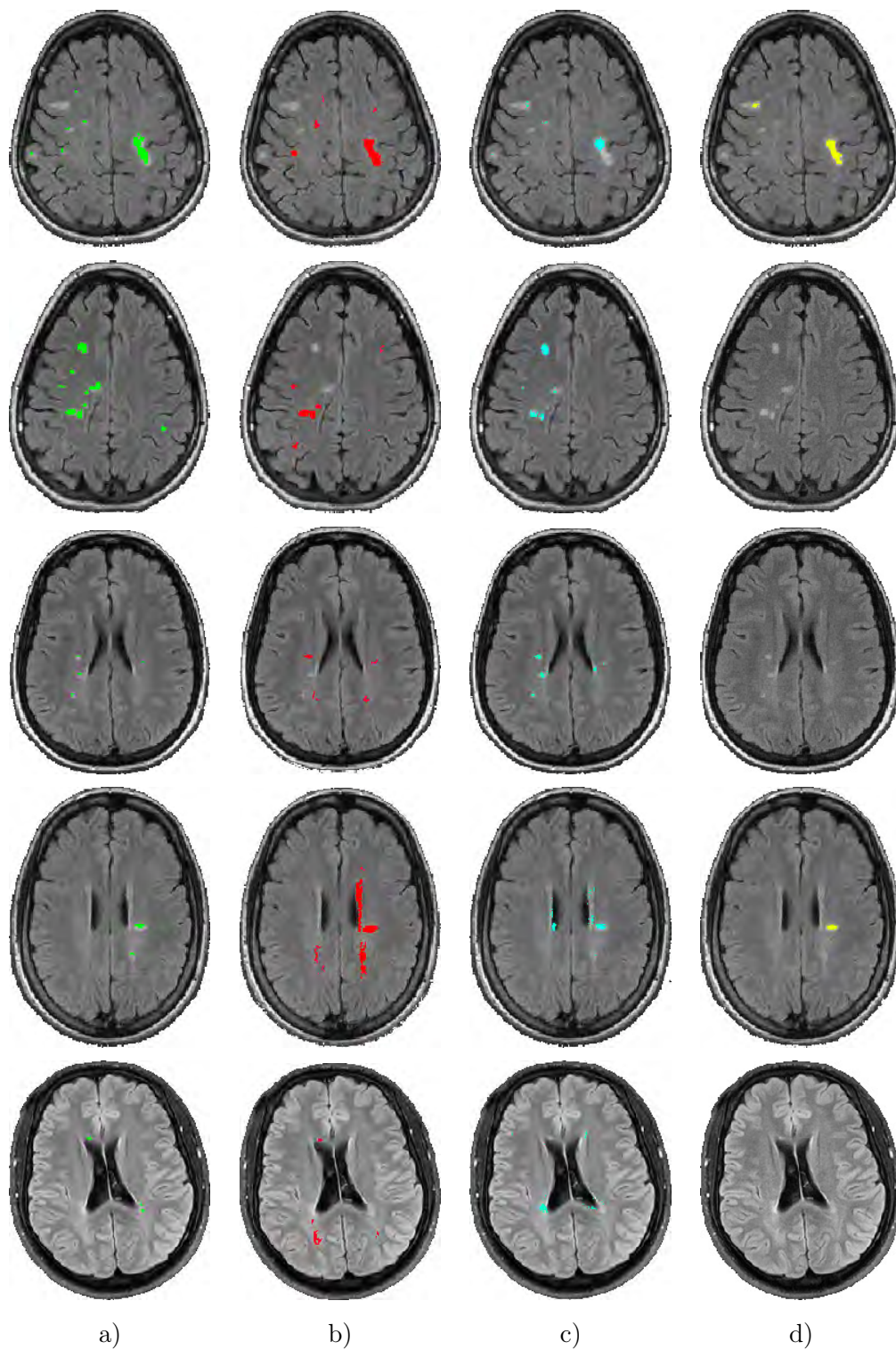


Figure 5.23: Segmentation comparison for Hospital Vall d'Hebron: a) GT (green); b) TISSUE-LOT pipeline (red); c) BOOST pipeline (cyan); and d) Souplet et al. [204] (yellow).

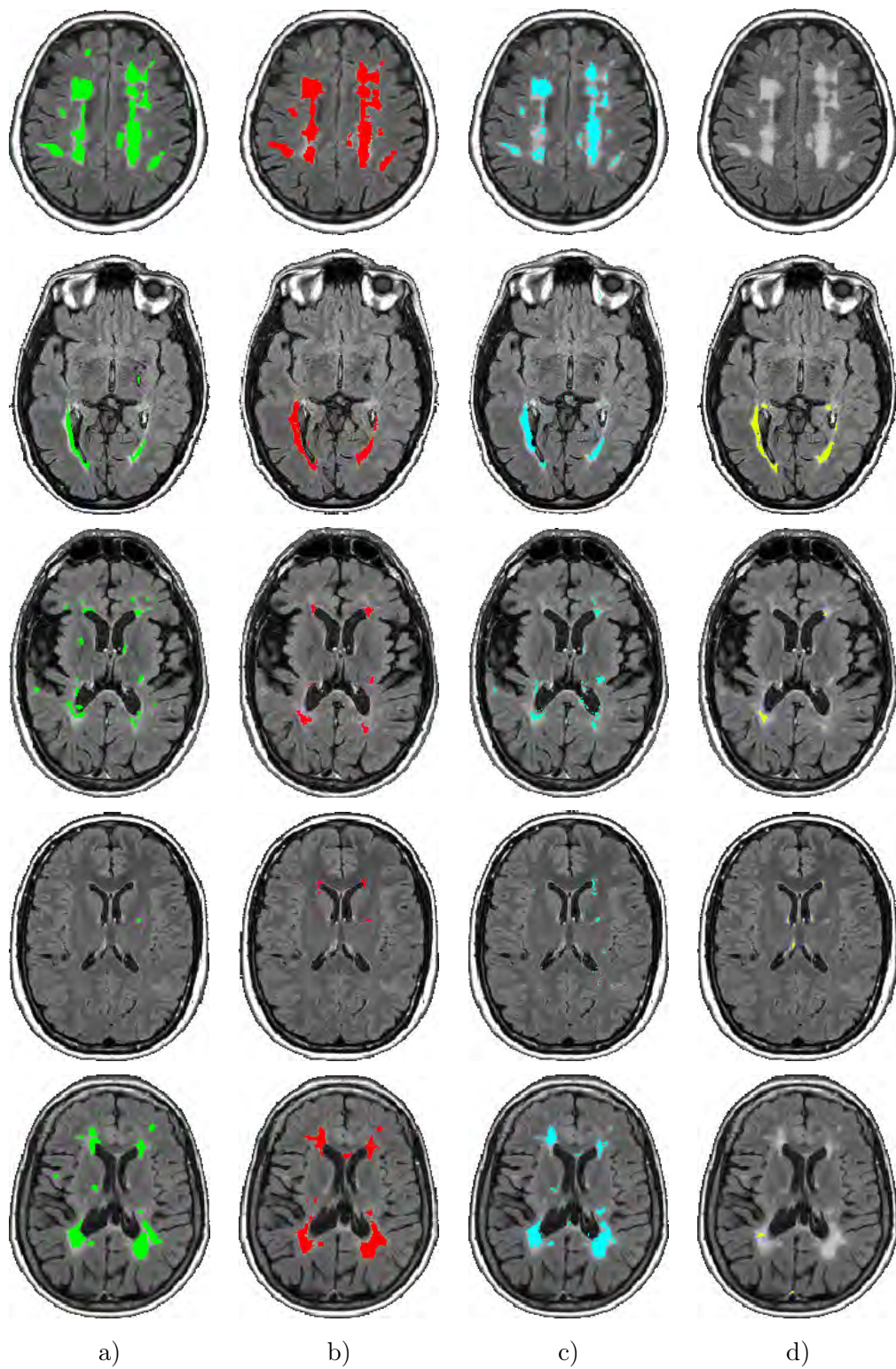


Figure 5.24: Segmentation comparison for Hospital Josep Trueta: a) GT (green); b) TISSUE-LOT pipeline (red); c) BOOST pipeline (cyan); and d) Souplet et al. [204] (yellow).

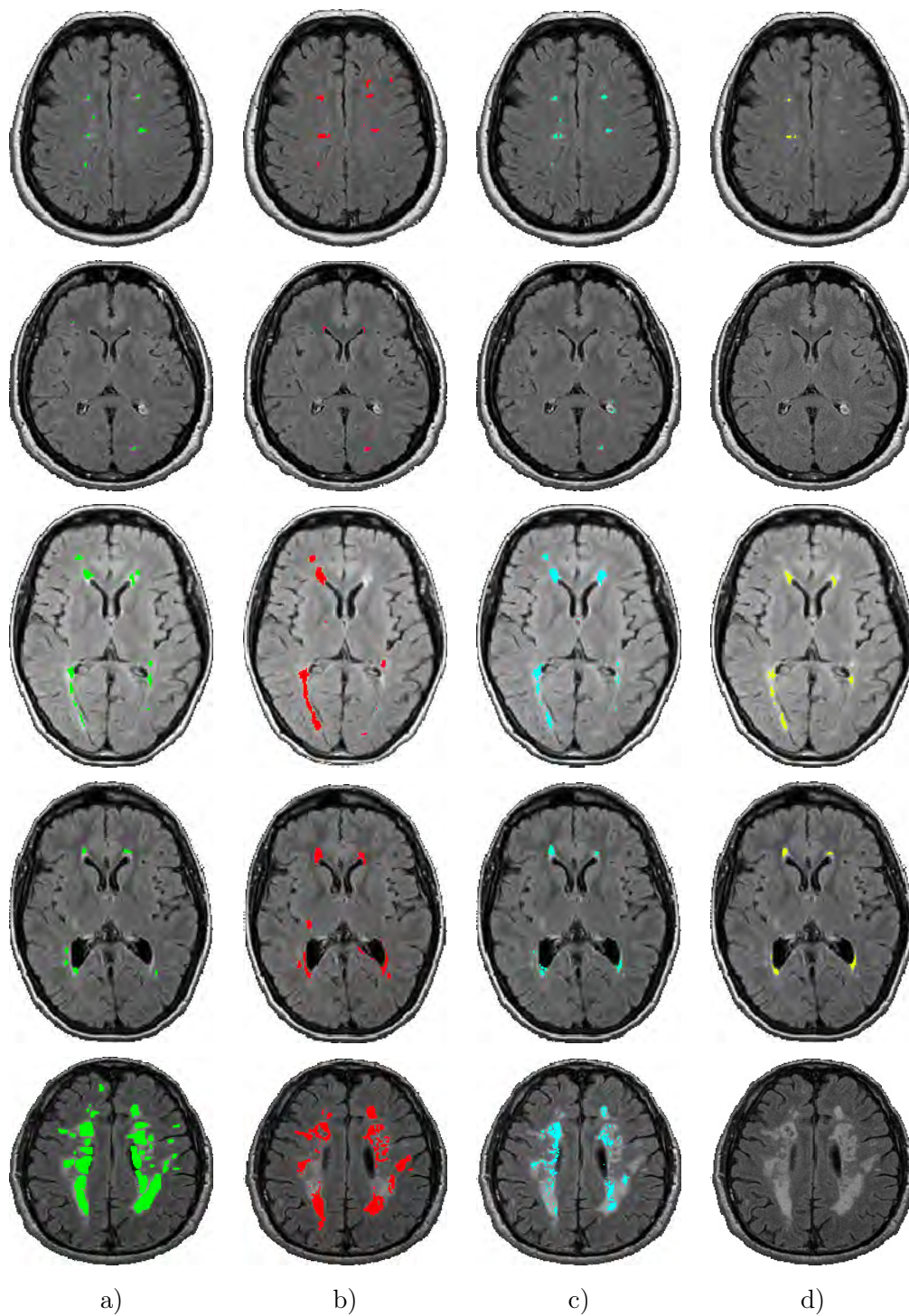


Figure 5.25: Segmentation comparison for Clínica Girona: a) GT (green); b) TISSUE-LOT pipeline (red); c) BOOST pipeline (cyan); and d) Souplet et al. [204] (yellow).

better segmentation results are obtained by the BOOST pipeline when compared with the other two approaches. The TISSUE-LOT has a tendency to oversegment the lesions (when compared to the GT) that can easily be observed in the fourth row of figure 5.23, the second and fifth row of figure 5.24, or the first, third and fourth row of figure 5.25; while SepINRIA usually undersegments all lesions with prominent examples in the second and third row of figure 5.23, the first and last row of figure 5.24, or the last row of figure 5.25.

However, it is always hard to choose one method over another even when looking at these visual examples, since lesions missed by one approach might be detected by another, or some lesions might be better defined by one method in a set of examples, but be incorrectly segmented in another set.

5.7 Discussion

In this chapter we have presented the results for the proposals described in chapter 4, both quantitatively and qualitatively in terms of detection and segmentation, and compared these approaches to a state-of-the-art approach. To do so, we started by describing the evaluation process and focused on the two key aspects of this process: the databases and the evaluation measures.

In terms of databases, we started with a brief description of the MS Grand Challenge 2008 data. This database comprises 3T images obtained from two different hospitals. This type of image presents some new issues not present in 1.5T images, while also correcting, or reducing part of the issues related with lower magnetic fields (such as partial volumes). Afterwards, we described our novel database for the SALEM project, which comprises 1.5T images from 3 different hospitals obtained with different scanning machines. Therefore, our proposals were tested only with our own database, even though some initial results with the MICCAI database are presented in appendix A.

In terms of evaluation measures, we separated them in two groups: detection measures, those that evaluate the number of lesions found and computed regionwise; and segmentation measures, those that evaluate the volume of the masks found and computed voxelwise. For both groups, we compute similar metrics related to the TP, FP, and FN detections or segmentations. Furthermore, we included two other measures in order to evaluate the overall segmentation per lesion in terms of distance and overlap.

Following these introductory sections, we presented the results for each pipeline independently. First, we analysed the results for the TISSUE-LOT pipeline with its strong focus

on detection. As expected, detection results were fairly higher than segmentation results for this pipeline due to its nature. Afterwards, we analysed the results for the BOOST pipeline, which is purely voxelwise. Also, as expected, the results for segmentation were fairly higher than the measures obtained in terms of detection. Both analysis were done in a quantitative manner.

Finally, in order to better understand the strengths and weaknesses of each pipeline and its contributions to the state-of-the-art, we presented a comparison with one of the most promising methods reviewed in chapter 2, Souplet et al.'s approach [204], which is similar to the TISSUE-LOT pipeline. After a thorough numerical analysis, the TISSUE-LOT pipeline presented better detection results when compared to the other two methods, while the BOOST pipeline obtained the best results in terms of segmentation. In both cases, Souplet et al., implemented as part of the SepINRIA toolbox, scored results below the other two methods. These quantitative results were also backed up in a qualitative manner with visual examples from each hospital.

CONCLUSIONS

I'll leave when I'm good and ready.

Lucille Bluth, “Arrested Development (Season two)” (2004)

6.1 Summary of the thesis

The aim of this thesis has been the proposal of a new pipeline capable of classifying brain tissues and detecting MS lesions in magnetic resonance imaging. Starting with an initial study of the state-of-the-art of MS lesion segmentation, we realised the importance of using prior knowledge. Medical images, specifically brain MRI, are usually corrupted with several artifacts and often include additional tissues and organs with similar intensities present in the region of interest. As a consequence, prior knowledge offers additional information that aids the segmentation process. Through this analysis, we also uncovered the relevance of atlas-based segmentation as a supervised strategy.

A further analysis into atlas-based brain MRI segmentation was also provided in order to highlight its strengths and weaknesses, as well as its applicability to our case of study: MS lesions. Starting with the basics, we defined an atlas as a set of images comprising at least one template image and one labelled image that define a segmentation. This atlas, which usually lies in its own space, must be warped to the case we want to segment through an optimisation process called registration. After a thorough study of the state-of-the-art on atlas-based segmentation strategies, we concluded that probabilistic atlases, consisting of a template image and a probabilistic map for each tissue, in conjunction with a statistical framework, are the most suitable option for random lesions (in terms of location, size and shape), due to their capability of estimating outlier classes not present in the atlas.

Therefore, we proposed a first pipeline, named TISSUE-LOT, to classify tissue and segment lesions. This first strategy uses a modified EM algorithm to estimate pure tissue classes (for CSF, GM, and WM), as well as a partial volume class containing CSF and GM. This algorithm provides an initial segmentation in which lesions are not included, thus a second step is needed to find them. This second step involves the thresholding of the FLAIR image (which shows high contrast between tissues and lesions), followed by a regionwise refinement of this thresholding. Consequently, this strategy heavily focuses on detection.

In order to further introduce prior information, we presented a second pipeline, called BOOST, focused only on the voxelwise segmentation of MS lesions. This fully supervised approach involves a classification strategy with a rich feature pool comprising intensity images, probabilistic maps from an atlas, a novel outlier map, and a rich set of contextual meta-features. The outlier map is heavily influenced by TISSUE-LOT and is based on the tissue estimation of this pipeline and the outlier properties of lesion voxels, while the contextual meta-features provide additional spatial information by comparing the voxels with randomly sampled regions around these voxels. This feature pool is then used to train a gentleboost algorithm, an ensemble classifier, capable of selecting the most discriminating features from the whole pool.

To evaluate these approaches, we used a novel database with 45 cases comprising 1.5T imaging data from three different hospitals with a variable lesion load per case. These images were manually annotated by experts in order to obtain GT segmentations that were used to validate the automatic segmentations with a set of similarity metrics for detection and segmentation. The results suggest that both approaches are complementary in the sense that while the first pipeline obtains better results in terms of detection, the second one presents better segmentation, presenting also a better segmentation of TP detections. Both approaches also outperform one of the best state-of-the-art approaches for our database. Moreover, the first approach also provides a tissue classification that can be refined to segment partial volume voxels between GM and CSF, as well as used to further refine results from the second pipeline.

6.1.1 Contributions

The goal of this thesis is to aid radiologists in day-to-day practise by assisting them in the challenging task of detecting and segmenting lesions. Idealistically, our proposal should accurately detect and segment all the lesions of a given patient. However, a more realistic

expectation is to allow experts to process a batch of patients off-line in order to accurately segment a majority of the lesions, reducing the experts' interaction with the images to the correction of some segmentations or the detection of a small number of missing lesions. By reducing the interaction we are also reducing the inter- and intra-observer variability. From this point of view, the main contributions of this thesis to both the scientific community and the hospitals are:

- An extensive survey of MS lesion segmentation algorithms, which are classified according to the use (or lack of) prior information into supervised and unsupervised algorithms. Analysing these works, we prove the importance of using prior information and spatial properties in order to accurately detect and segment lesions.
- An extensive survey and analysis of brain MRI atlas-based segmentation. We introduce some of the basic concepts of atlas-based segmentation, including atlas creation, registration, strategies, and applications; and we review the main methods presented. Moreover, analysing the results, we conclude that using a probabilistic framework with a probabilistic atlas is the best option to segment lesions.
- A novel database for the SALEM project with 1.5T imaging data for 45 cases with different lesion loads from three different hospitals; Hospital Vall d'Hebron, Hospital Josep Trueta, and Clínica Girona, acquired with three different scanning machines from different manufacturers; Siemens, Philips, and General Electrics respectively. These cases were manually annotated by experts to obtain GT segmentations.
- A new algorithm to classify tissue voxels and segment lesions according to this tissue segmentation. This pipeline approach, named TISSUE-LOT, is exhaustively tested using our SALEM database.
- A new algorithm to segment lesions using a rich feature pool including spatial information, an outlier map and an ensemble classifier. This pipeline approach, called BOOST, is exhaustively tested using our SALEM database.
- An experimental qualitative and quantitative comparison between our two proposed pipelines and a promising state-of-the-art approach in terms of both detection and segmentation.
- An early prototype to detect and segment lesions to be tested in the hospitals. This tool has been implemented in C++ and is currently available at Hospital Vall d'Hebron as a set of console functions and scripts.

6.2 Future work

The analysis of brain MRI images for MS patients is a complex topic involving several aspects and multiple research lines. This notion is exemplified in our case, by the modular design described in our two proposed pipelines. Several steps are involved in the image enhancement process prior to the segmentation itself. Furthermore, most of the concepts applied to MS lesion segmentation can be applied to other brain MRI imaging topics or can be studied further. Besides, other interesting topics arise from the needs of the current clinical practise for MS patients.

Hence, future directions are presented divided into two categories: those related to increasing the reliability of our proposal, and future research lines departing from this thesis.

6.2.1 Short-term proposal improvements

In this thesis, we presented two different approaches. The first included a tissue classification step (with tissue Gaussian parameters estimation) that served as the initial process to create an outlier map. The goal of this map was to highlight outliers on a multispectral basis to reduce the effect of hyperintense artifacts in FLAIR images. This map was in turn used as a feature in a classification approach that also produced a likelihood map for lesions. In order to further interconnect both pipelines and to take advantage of their strengths, one improvement to our proposals could be the combination of the outlier map and the classification map when using only intensity images, probabilistic atlas maps and contextual meta-features as part of the rich feature pool. Furthermore, the results of this combination could be refined with tissue information in a similar manner to the TISSUE-LOT refinement step.

Other improvements to the pipelines would be the study of other pre-processing algorithms. For instance, the skull stripping could be improved for those rare cases where BET is not capable of fully removing the eyes or other non-brain tissues. Moreover, a new algorithm for inter-slice inhomogeneities, rarely tackled by current bias correction algorithms, could be developed to aid the segmentation of lesions in the topmost slices. These lesions usually present a lower intensity profile that might difficult the task of global intensity methods.

Another possible closely related improvement could be the implementation of a novel atlas database with multiple topological atlases. These atlases would be selected and

combined for each new patient to create a new probabilistic atlas for each independent case in order to better adapt to it. Moreover, a synthetic atlas, using the simulation concepts presented in appendix A, could also be developed in order to better capture the morphology of a new case.

Other future work we propose as a short-term goal would be the expansion of the current SALEM database to further study the differences between scanning machines and the effect of the lesion load during the segmentation. At present, we have a database with a low number of cases per hospital that present a variable lesion load, which in turn hinders the task of separating the effect of the scanner and the lesion load.

Finally, we would like to test the prototype provided at Hospital Vall d'Hebron with new clinical data to evaluate its performance and usability in clinical practise, as well as to optimise, refine, and extend its functionality. For instance, we would like to provide a user-friendly GUI for doctors to interact with.

6.2.2 Future research lines

In the long term, there are several new research lines departing from this thesis that could be studied by the group. One such proposal, closely ties into the research presented in this thesis and can benefit from it. This other research line proposal involves the study of lesions in temporal cases. In fact, the SALEM database contains two different control points for each patient: a basal initial scan, and a follow-up taken between 6 to 12 months later. The study of the evolution of MS lesions is usually performed by subtraction or the analysis of the transformation between two images. Our segmentation methods could aid this process by providing some initial lesion masks in both the basal and follow-up images. In fact, atrophy markers based on MRI analysis are commonly used in clinical practise although they suffer from the present variability under the presence of MS lesions.

Further departing from this thesis and related to the previous topic, there is the study of atrophy in conjunction with lesions. In this thesis, we briefly address tissue segmentation for a single scan. However, the study of the evaluation of these segmentations at different time instants coupled with the lesion study could help to understand the relationship between both phenomena in the MS disease.

In this thesis, we have defined MS as a disease that presents hyperintense WML disseminated in time and space. Similar presentations can occur in patients having an infectious, neoplastic, congenital, metabolic or vascular disease, or non-MS idiopathic inflammatory

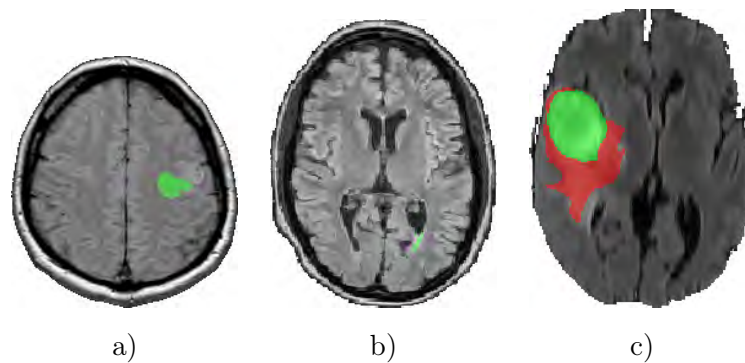


Figure 6.1: Examples of lesions in other brain diseases with GT: a) FLAIR image with a stroke lesion, b) FLAIR image with a lupus lesion, and c) FLAIR image with a tumor (green) and edema (red).

demyelinating disease [153]. These patients can even present symptoms similar to those in MS, with differences in their treatment and prognosis. From the image point-of-view, images of brains suffering these diseases will present the same properties as MS imaging with hyperintense focal lesions distributed along the brain. Such images could be also automatically segmented using the techniques presented in this thesis to aid radiologists.

On a separate note, the methods and concepts presented here could also be applied to the study of other lesions with similar properties. For instance, as illustrated in figure 6.1, lupus lesions appearing in WM are hyperintense and can appear near the ventricles; stroke lesions can also appear as hyperintense lesions of variable sizes in T2-w images; and tumors usually appear as large hyperintense lesions that deform the tissues surrounding them causing edema. Moreover, for MS lesions, we could also try to differentiate between the lesion subtypes presented in chapter 1 and develop new tools to segment them to help the study of their properties. This could also aid the MS diagnosis in those patients with a first clinical episode according to the McDonald's criteria.

Finally, the development of new MRI techniques with higher contrast and reduced artifacts and new tissue attenuated sequence relying on inversion recovery, or the introduction of functional MRI and diffusion tensor imaging (DTI), could be also studied to help uncover some unsolved questions for MS.

EPFL'S INTERNAL REPORT: MS LESION SEGMENTATION IN MRI USING STATISTICAL METHODS AND SPATIAL INFORMATION

A.1 Introduction

Accurate and robust brain tissue and lesion segmentation from MR images is a key issue in many applications of medical image analysis and, particularly, in the study of multiple sclerosis. Manual tracing by an expert on the three brain tissue types, white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF) as well as white matter lesions (WML), in MR images is far too time consuming as the data involved in most studies is large. Moreover, automated and reliable tissue classification is a demanding task as the intensity representation of the data normally does not allow a clear delimitation of the different tissue types present in a natural MRI. This is due to the partial volume (PV) effect (presence of more than one brain tissue type in a voxel), image noise and intensity non-uniformities caused by the inhomogeneities in the magnetic field of the MR scanner. Most approaches dealing with this issue rely on statistical segmentation strategies [135, 121, 204, 61, 209].

In [37], we analysed the state of the art on multiple sclerosis segmentation. In this study, we concluded that using a priori knowledge can improve the segmentation of brain tissues. Furthermore, one of the most common strategies relying on prior knowledge consists of registering a template (called atlas) with tissue priors to aid the segmentation.

Consequently, in [37], we decided to compare two of the most promising atlas-based segmentation techniques and propose a new lesion segmentation method. Those atlas-based methods encode anatomical spatial priors as atlas probabilities. Note that this information is in fact global spatial information. In order to also include local spatial information to improve our segmentation results, we decided to study the use of Markov random fields (MRF) during my research stay at the EPFL under the BE grant 2010 BE1 00828. This work was developed with the LTS5 group led by Dr. Jean-Philippe Thiran under the supervision of Dr. Meritxell Bach Cuadra.

A.1.1 Goals

In this report, we **investigate different automatic pipelines capable of classifying brain tissues and detecting MS lesions in magnetic resonance imaging (MRI). We also propose a MRF refinement** in order to improve the accuracy of those pipelines. The general goal of this research stay can be subdivided into the following objectives:

- Validate our previous work [37] with a more realistic fluid-attenuated inversion recovery (FLAIR) simulation and extend it to real data.
- Learn and understand the basics of MRF models.
- Improve the segmentation results using MRF by implementing a post-processing algorithm based on spatial relationships between tissues and inner brain structures.

A.1.2 Structure

The rest of this report is organized as follows: In section A.2 we present a comparative study, with both synthetic and real data, between two state of the art methods for MS lesion segmentation proposed by Souplet [204] and de Boer [61]. Our new refinement approach based on MRF theory is introduced in section A.3. Finally, we summarise the conclusions of our work and their impact on this PhD thesis in section A.4.

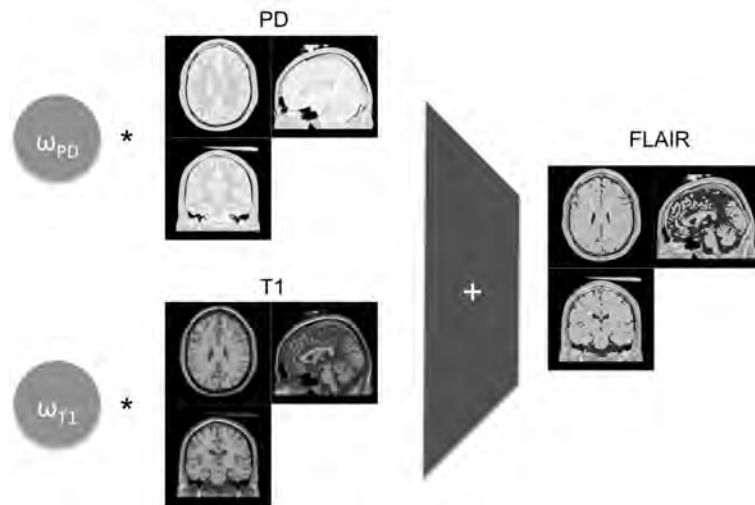


Figure A.1: Scheme of the previous FLAIR simulation pipeline. After the weighted sum, the CSF was darkened to simulate fluid attenuation.

A.2 Study of MS segmentation approaches

A.2.1 Motivation

In [37], we studied two different atlas-based strategies for MS lesion segmentation and, based on these, we proposed a new segmentation pipeline. These methods, with two segmentation steps for tissue classification and lesion segmentation, were evaluated quantitatively using synthetic data from the Brainweb and qualitatively by visual inspection using our own real database.

Brainweb is a publicly available synthetic phantom of the brain. Two versions of this phantom can be downloaded from the website: a healthy subject and the same subject with registered WML. For each case, T1-w, T2-w and PD-w images can be simulated with different bias fields, noise and slice thickness configurations. However, there is no option to simulate FLAIR images, difficulting the task of testing segmentation methods based on this imaging sequence. In our previous work [37], we proposed the simulation this image using the other available ones in a weighted sum with similar contrast to real FLAIR images, as illustrated in figure A.1. However, this method was far from realistic since it did not follow any physical principle. Therefore, our first objective was to better simulate this sequence in order to validate the methods with a more realistic synthetic case.

In [37], our methods were tested with a data set consisting of 6 cases with T1-w, T2-w,

PD-w and FLAIR images provided by the Hospital Vall d'Hebron in Barcelona (Spain). Those images were obtained with a SIEMENS scanner of 1.5T and the number of slices ranged between 44 and 46. Since no ground truth for those images was available at the time of that study, we conducted a qualitative evaluation. During this research study, we also wanted to validate those methods with a larger data set with ground truth to obtain a quantitative evaluation.

Finally, we propose new pipelines using the SPM8 [13] tissue segmentation method. The main goal of using this new segmentation method is to obtain better skull stripping. However, this new segmentation method also provides a tissue segmentation that we can use as input to the already developed lesion segmentation methods.

A.2.2 FLAIR simulation

In order to have a realistic simulation, one should recreate the different steps involved in the MRI scanning (from RF signal simulation, to the acquisition of the image in k-space to the final reconstruction of the MRI volume) [127]. This task requires great knowledge of MR physics and is difficult to implement. In order to simplify this simulation, another approach consisting of simulating the process in image space using an equation to approximate the voxel value according to the tissue and acquisition parameters [183, 151, 215], can be used:

$$S = M_0(1 - 2e^{-TI/T1} + e^{-TR-TElast/T1})e^{-effTE/T2}. \quad (\text{A.1})$$

We can define the FLAIR signal for each tissue according to its proton density (M_0), its $T1$ and $T2$ times and the following scanning parameters: inversion time (TI), recovery time (TR) and the last ($lastTE$) and effective echo time ($effTE$). The signal S is different for each tissue and is independent of its spatial position. To obtain this tissue signal, we used the tissue parameters M_0 , $T1$ and $T2$ provided by the Brainweb site at http://mouldy.bic.mni.mcgill.ca/brainweb/tissue_mr_parameters.txt/ while the acquisition parameters were defined with the following common FLAIR values for the human brain: $TI = 2800ms$, $TR = 14000ms$, $lastTE = 216ms$ and $effTE = 108ms$.

Once we obtain the tissue FLAIR signals, the next step is to map them in image space. In order to map the signals to their positions, we used the fuzzy membership of each voxel to a tissue type. These fuzzy membership maps are provided by the Brainweb site as ground truth for the segmentation evaluation. Therefore, the final intensity value for each voxel is defined as,

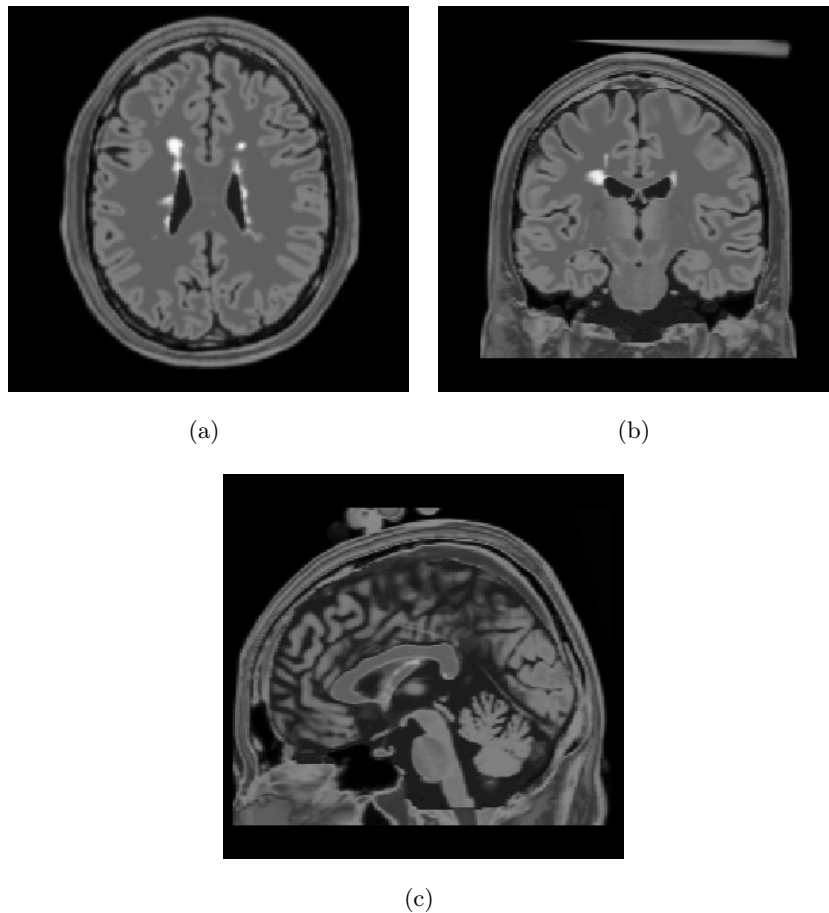


Figure A.2: FLAIR simulation with 0% noise.

$$I(x, y, z) = K \sum_{c=1}^C \pi_c(x, y, z) \cdot S_c, \quad (\text{A.2})$$

where (x, y, z) are the spatial coordinates of the voxel, C is the number of tissues (fat, muscle, skin, CSF, GM, WM, lesions, ...), π_c is the fuzzy membership for the tissue c , S_c is the tissue signal and K is a scaling constant. A slice example of this simulation is presented in figure A.2.

A.2.3 Data sets

In this section we summarise the data sets used for evaluation. First, we present the images we used from the synthetic Brainweb phantom and then briefly describe the real data set.

Brainweb

As mentioned earlier, the Brainweb is a publicly available synthetic phantom of the human brain. The PD-w, T1-w and T2-w images are publicly available for downloading in the website to test brain image analysis methods. Different noise and bias field levels can also be applied to the images. For this study, we provide results with the 3 images mentioned plus our new simulation for the FLAIR with different noise levels: 0%, 1%, 3%, 5% and 7%. All volumes have a size of $181 \times 217 \times 181$ and spatial resolution $1mm \times 1mm \times 1mm$.

Real data

In order to perform the evaluation with real data sets, we selected 16 training cases from the MS challenge workshop of the MICCAI conference [208]. Those cases were acquired at the University of North Carolina (UNC) and the Children's Hospital of Boston (CHB) using a 3T scanner. For all cases, the following high dimensional images were obtained: a T1-w scan, a T2-w scan and a FLAIR image. All volumes have a size in voxels of $512 \times 512 \times 512$ and spatial resolution of $0.5mm \times 0.5mm \times 0.5mm$.

A.2.4 Methodology

In this section we briefly summarise the methods we compared.

Tissue classification with kNN

Following the idea presented by de Boer et al. [61], CSF, GM and WM are segmented using an automatically trained kNN classifier, an extension of the work by Cocosco et al. [45]. Training samples for the kNN classifier are obtained from the subject himself by atlas-based registration [182]. These atlases are then thresholded in order to get candidate training samples with a predefined probability of belonging to a specific label. The threshold value is empirically defined in order to include variability in the CSF class.

For all three classes, candidate training samples are randomly taken from the spatial locations masked by the thresholded atlases. The features of the samples consist of the intensity values of the PD-w and T1-w images at the sample locations. Afterwards, a pruning step, based on a minimal spanning tree, is applied to this initial set of samples to remove outliers.

Finally, a kNN classifier based on a *fast nearest neighbor lookup library*¹ performs the final classification based on the pruned sample set. The parameters k and n were chosen to ensure the minimum possible error probability of a generic classifier (achieved when the classifier knows the true data distributions) [207, 65].

Tissue classification with EM

Similar to Souplet et al. [204], the algorithm presented in the proposal of Dugas-Phocion et al. [72] is applied to the T1-w and T2-w images. This algorithm uses the principle of the EM algorithm to maximise the log-likelihood between the MRI data and a Gaussian model of four classes: WM, GM, CSF, and the GM/CSF partial volume class.

First, the probability of each voxel of belonging to one of the different classes is initialised thanks to the a priori information of the atlas. Probabilities are taken directly from the atlases for pure classes (CSF, GM, WM), however, partial volume classes have no priors. To overcome this issue, a new atlas with a proportion of the CSF and GM atlases is created for the partial volume class.

After initialisation, two steps are iterated:

- In the maximisation step, the parameters (mean μ_k and covariance matrix S_k) of each class are computed from the voxel intensities and their probabilities of belonging to the different classes defined. The partial volume class parameters are obtained as a proportion of the pure tissue parameters.
- In the expectation step, the probability of each voxel of belonging to one of the different classes is updated depending on the classes parameters, using the Gaussian function and the atlas values as prior probabilities.

Finally, when the algorithm converges, a bidimensional Gaussian distribution is estimated for each class including GM, WM, CSF and partial volumes as well as all the voxel probabilities related to these distributions.

Tissue classification with SPM8

SPM is a Matlab toolbox containing different brain image analysis methods ranging from bias correction, to registration or the segmentation of different brain imaging sequences

¹This library can be downloaded from <http://www.cs.umd.edu/~mount/ANN/>

(PET, MRI, fMRI, etc.). The main segmentation method of this toolbox, presented by Ashburner et al. [13], is a hybrid algorithm that combines segmentation, atlas registration and bias correction. At each iteration, the segmentation and registration steps are interleaved in order to refine each other. Two variants of the algorithm are presented in the toolbox. One is a single channel approach with an atlas composed of the three main tissues (GM, WM and CSF) and the other is a multi-channel approach that uses 6 atlas: 3 for the main tissues, 2 for non-brain tissues and 1 for the background. We used the first method due to some segmentation problems with the eyes when using the second method combined with T2-w images.

Lesion segmentation 1 (LS1)

Figure A.3(a) illustrates the whole lesion segmentation method. After the tissue segmentation using the kNN algorithm, the GM mask is applied to the FLAIR image to compute a histogram. Using this histogram, the mean value (μ) is estimated from the peak centre location and the standard deviation value (σ) is calculated using the full width at half maximum. Afterwards, a threshold is defined as:

$$T = \mu + \alpha\sigma, \quad (\text{A.3})$$

where the parameter α is defined empirically. Upon applying this threshold to the FLAIR image, an initial lesion mask is obtained.

In order to reduce the number of false positives, for every lesion (considered as a connected component), the following WM fraction is computed:

$$WM_{fraction} = \frac{|WM_{neighbours}|}{|CSF_{neighbours}| + |GM_{neighbours}|}, \quad (\text{A.4})$$

using a 3D 18-neighbourhood relation. Afterwards, those lesions whose $WM_{fraction}$ is under an empirical threshold β are reclassified as GM.

Lesion segmentation 2 (LS2)

Again, figure A.3(b) illustrates the whole lesion segmentation method. After the tissue segmentation using the EM algorithm, the GM mask is applied to the FLAIR image to compute the mean and standard deviation values and a threshold is computed following equation A.3. Afterwards, the FLAIR image contrast is enhanced using morphological operations. First, both eroded and dilated FLAIR images are obtained. Then, a new

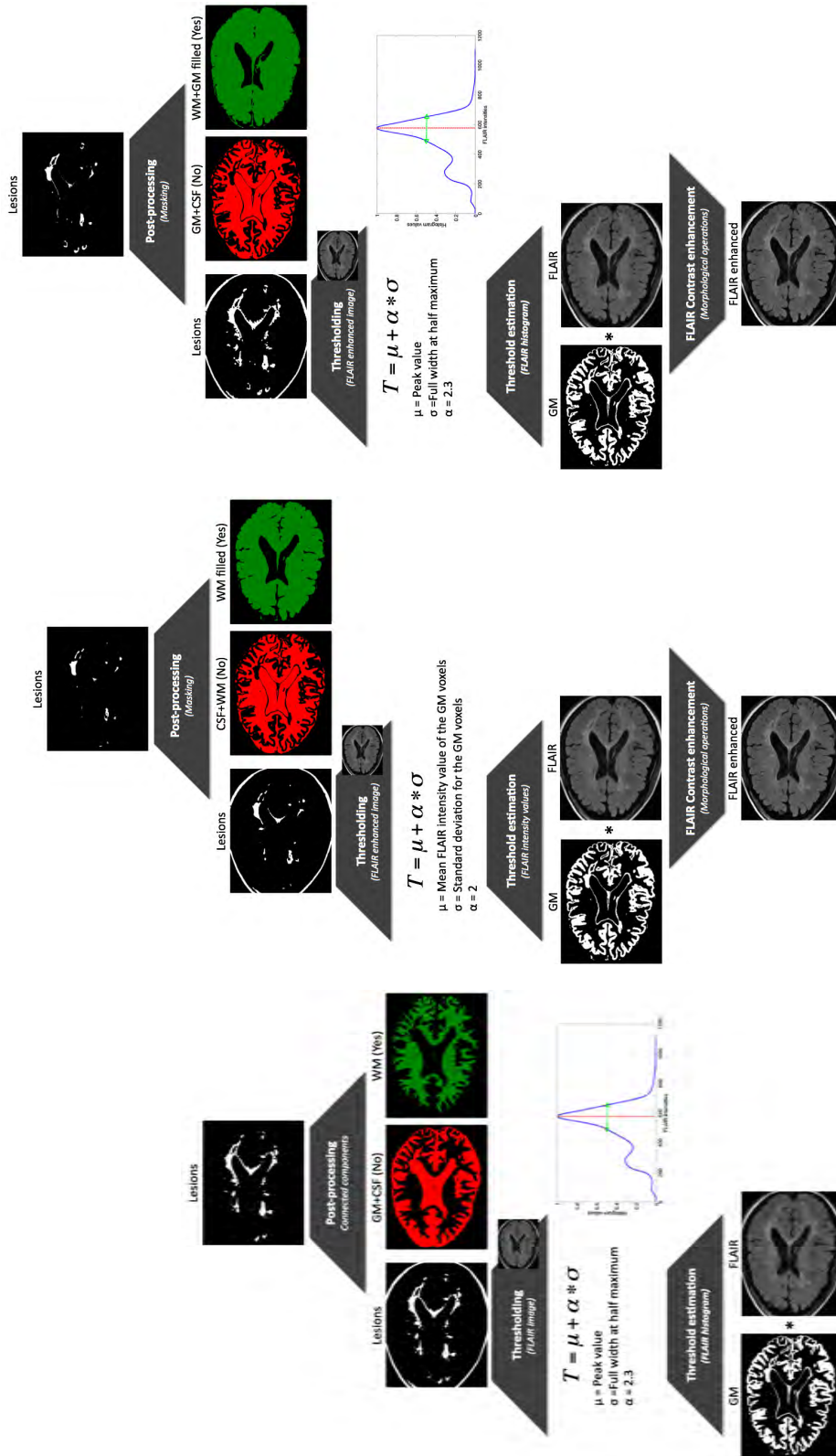


Figure A.3: Overview of the three different lesion segmentation methods.

Table A.1: Summary of the different pipelines studied.

	Tissue	Lesion
kNN-LS1 [61]	kNN	Lesion 1
EM-LS2 [204]	EM	Lesion 2
EM-LS3 [37]	EM	Lesion 3
kNN-LS3 [37]	kNN	Lesion 3
SPM-LS1	SPM	Lesion 1
SPM-LS2	SPM	Lesion 2
SPM-LS3	SPM	Lesion 3

enhanced image is obtained comparing voxelwise the eroded and dilated images to the original. The intensity value closer to the original FLAIR value is then assigned to the new enhanced voxel. This new enhanced image is thresholded to obtain an initial lesion mask.

In order to reduce false positives, this initial mask is further refined using the WM mask with the cavities filled and the CSF and WM masks coming from the EM. Lesion voxels should be on the “healthy WM”, represented by the filled WM mask, but they should not be on the real WM or CSF.

Lesion segmentation 3 (LS3)

We propose a new strategy as a combination of the two previous ones as shown in figure A.3(c). First, μ and σ are estimated from the histogram and the threshold is defined as in LS1. Afterwards, the FLAIR image contrast is enhanced using morphological operations as in LS2.

This new enhanced image is thresholded to obtain an initial lesion mask. Finally, similar to LS2, we consider as lesions those voxels inside the filled union of the GM and WM masks and discard the voxels inside the WM and CSF masks to obtain the final lesion mask.

Pipeline summary

In this study, we compare 7 different pipelines summarised in table A.1. Two of them are the implementation of the original proposals of Souplet et al. [204] and de Boer et al. [61], while the other pipelines are combinations of the tissue and lesion methods.

A.2.5 Results

In this section, we present the results for the 4 pipelines studied in [37]. First, we summarise the results obtained with the new FLAIR simulation for the Brainweb followed by the results obtained with the challenge data base.

Evaluation measures

In order to evaluate the lesion detection and segmentation, we decided to use 3 of the measures that were used in the challenge: the average surface distance, the true positive rate and the false positive rate.

The average surface distance computes the average distances between each automatically segmented lesion border voxel and the closest ground truth voxel in mm. This measure can give us a sense of overlapping. In order to have a perfect overlap, a distance value of 0 should be obtained.

The last two measures compare the number of true positives and false positives to the number of ground truth lesions and the number of segmented lesions respectively. Those two measures are pure detection measures. The first gives us an idea of the underestimation of the number of lesions while the second provides an idea of the overestimation. A value of 100 and 0, respectively, defines a perfect detection. Note also that a completely segmented image produces a 100% true positive rate while an empty image produces a 0% false positive rate.

Finally, we also present tissue results using the Dice similarity coefficient (DSC) [67],

$$DSC = \frac{2|T_A \cap T_{gt}|}{|T_A| + |T_{gt}|}, \quad (\text{A.5})$$

where T_A is the obtained tissue mask we want to evaluate and T_{gt} is the ground truth mask for the same tissue. This additional overlap measure can give us an idea of the accuracy of the tissue segmentation methods which may help to explain the final lesion results by introducing more information on the pipeline errors.

Synthetic

Figure A.4 summarises the results for the Brainweb phantom with the new FLAIR simulation. Looking closely at the true positive fraction, we can see how the new lesion segmentation method obtains values similar to the LS1 method. If we look at the values

when the images have no noise, we see low values for the original state-of-the-art methods, which increase radically when the noise is augmented to 1%. Moreover, all the methods are quite robust to any increase in the noise level, with a slight decrease on true positive values.

However, if we look at the false positive fraction, we can observe slightly better results with the kNN based methods, with the kNN-LS1 obtaining the lowest false positives. This method relies on a neighbourhood strategy to reduce false positives instead of masks. Moreover, this strategy outperforms the three others in terms of the average surface distance, suggesting a better segmentation of the detected lesions.

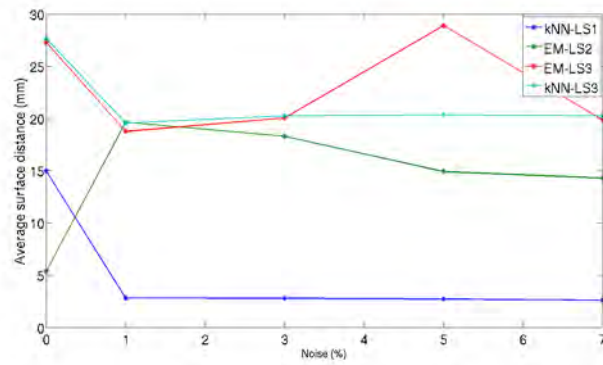
Looking at the overlap between the automatic tissue segmentation methods and the ground truth (see figure A.5), we can clearly see how EM obtains better results for the three main tissues. In order to introduce a fair evaluation, the PV class has been resegmented into GM and CSF according to the distance between the voxel and the tissue masks as well as the Mahalanobis distance between the intensity of the voxel and the tissue distributions. The EM method obtains fairly lower values than the kNN when there is no noise in the images. This kind of behaviour is due to the nature of the simulated images with high pikes and small standard deviations for each tissue. Therefore, the EM estimates wider Gaussian distributions. When noise is introduced, the higher variability of tissue intensities allows the EM to better fit Gaussians into the image.

In conclusion, we have seen how the EM seems to be the best tissue segmentation strategy, while the LS1 method optimises the TP/FP ratio and the segmentation boundaries. Moreover, we have seen how the EM outperforms the kNN when some noise is introduced into the image. Since real images present noise artifacts due to their acquisition, this improvement in accuracy for the tissue segmentation is also expected when evaluating the methods with real datasets.

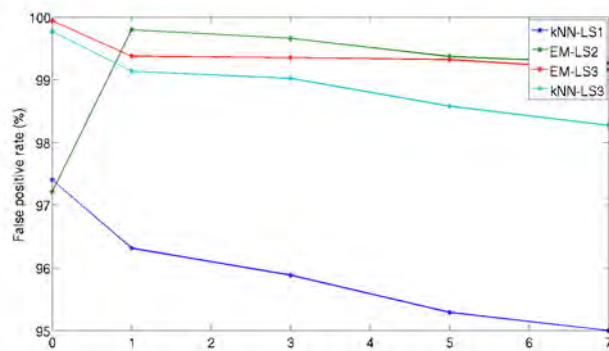
Real data

In this section, we compare the 7 different pipelines using the true positive fraction, false positive fraction and average surface distance measures.

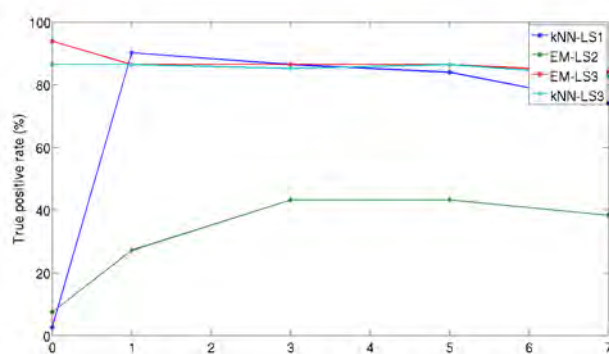
First of all, figure A.6(a) contains the boxplot for the average surface distance for all the methods. Looking at the boxplot, we can see how the SPM-LS1 pipeline obtains the lowest average surface distance (its mean value being $6.1658 \pm 4.6800mm$). Compared to the kNN-LS1 pipeline, there is an improvement in this measure due to a better tissue segmentation.



(a) Average surface distance



(b) False positive rate



(c) True positive rate

Figure A.4: The methods evaluated comprising the 4 original pipelines. (a) Average surface distance boxplot, (b) false positive fraction, and (c) true positive fraction.

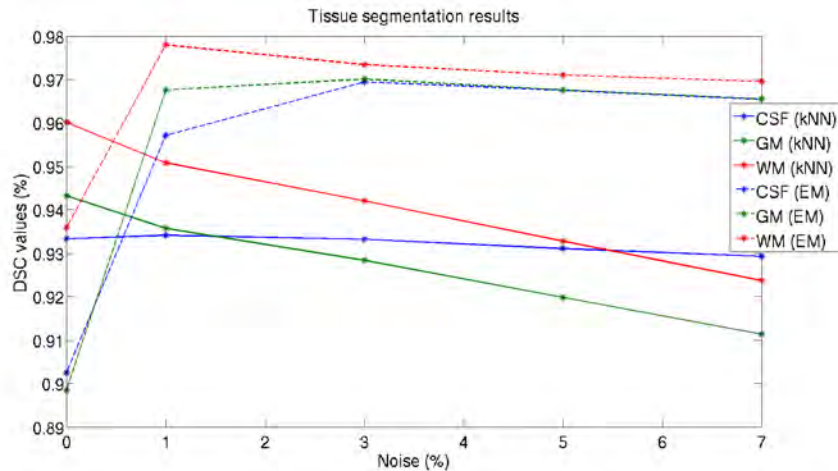
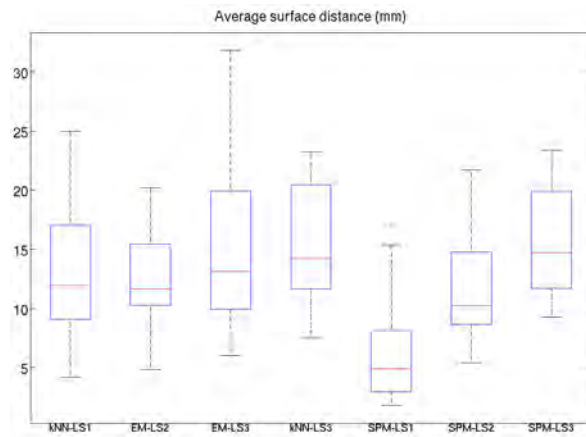


Figure A.5: DSC values for the 3 main brain tissues: CSF, GM and WM when using the EM and kNN tissue segmentation algorithms. PV voxels have been reclassified to CSF and GM.

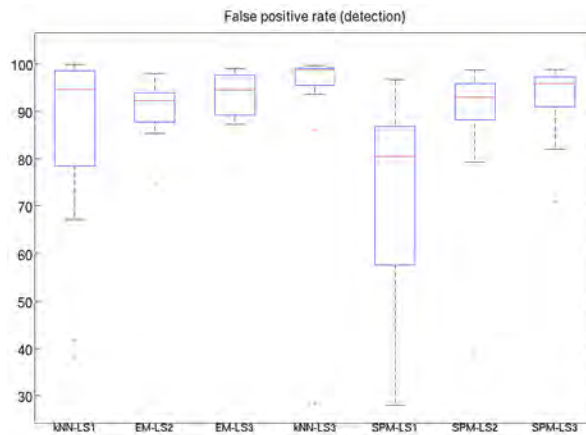
The kNN method usually misclassifies CSF as GM affecting the final threshold for the lesion segmentation. Therefore, applying a better segmentation (for instance, the one provided by the SPM) is expected to improve segmentation results. Note also how the new lesion segmentation method obtains the worst results applied to the tree tissue segmentation approaches even though there is not a significant difference between the three boxplots.

A similar effect can be observed when looking at the false positive fraction in figure A.6(b). The results of the LS1 lesion approach in terms of false positive are reduced when a better tissue segmentation approach is used (SPM in contrast with kNN). In this sense, SPM-L1 outperforms the results obtained using EM-LS2, going from a mean value of $85.4063 \pm 20.0952\%$ to $72.6828 \pm 24.4370\%$. However, the number of misclassified lesions still remains high with the false positive fraction over the 70%. Similar to the distance measure, results for the LS3 lesion segmentation method are higher (in terms of false positives) than the other methods being validated, pointing to a worse segmentation.

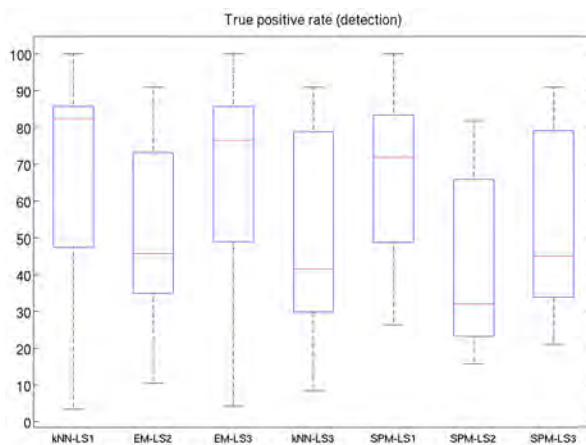
Finally, if we analyse the true positive fraction, we can see that using the SPM segmentation the results are lower. In fact, the methods that detect the most true positive are the original method from de Boer (kNN-LS1) and the EM-LS3. These best results can be explained by an oversegmentation of the lesions, hence the high false positive values or surface distances. In this sense, big blobs are considered as lesions and these blobs overlap the lesions.



(a) Average surface distance



(b) False positive rate



(c) True positive rate

Figure A.6: The evaluated methods comprising the 4 pipelines, plus the 3 lesion segmentation approaches applied after segmenting the tissues with SPM8. (a) Average surface distance boxplot, (b) false positive fraction, and (c) true positive fraction.

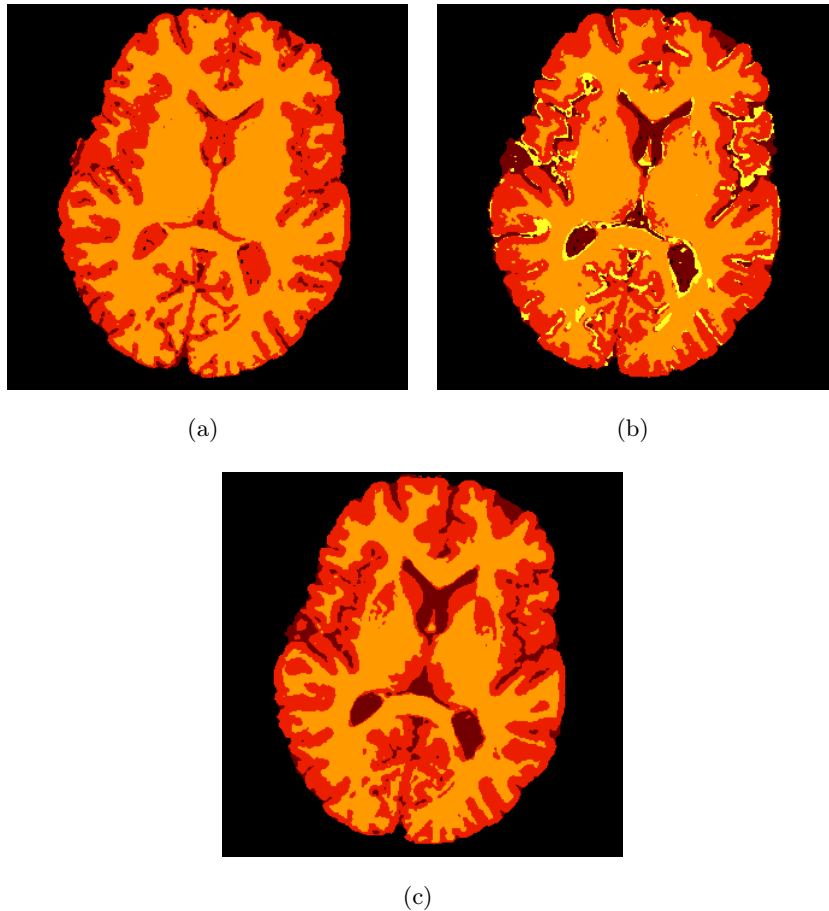


Figure A.7: Example of the (a) kNN tissue segmentation, (b) EM tissue segmentation and (c) SPM8 tissue segmentation for CSF (brown), GM (dark orange), WM (orange) and PV (yellow).

A.2.6 Conclusions

We have analysed 7 different pipelines, two of them being part of the state-of-the-art MS lesion segmentation methods based on atlas. All those pipelines have a thresholding step in common. The threshold is always computed from the previously segmented GM mask applied to the FLAIR image, where lesions are clearly hyperintense in comparison with the other tissues, and the other tissue masks are used on a post-processing step to perform the false positive reduction step. As a consequence, obtaining a good tissue segmentation becomes a crucial requisite. For all the pipelines, 3 tissue segmentation methods are used: an EM method with a partial volume class, a kNN method trained using an atlas and the SPM8 method.

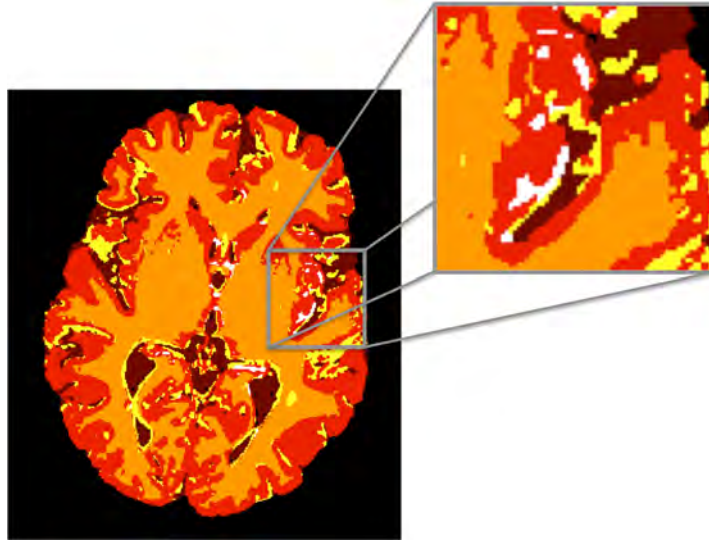


Figure A.8: Slice example of misclassified WML inside the GM. Tissues in the image are: CSF (brown), GM (dark orange), WM (orange), PV (yellow) and WML (white).

In general, the EM and SPM methods produce similar results for the CSF and GM (see figure A.7), while the kNN algorithm obtains worse results when segmenting the CSF, specially the ventricles. In general, this is the case for all cases with tissue segmentation, therefore, the best lesion segmentation results should be obtained when using either the EM or the SPM tissue masks. As we have seen in the previous section, this is actually the case.

We can also see that when using better tissue strategies, the best lesion segmentation approach is LS1, presented by de Boer et al. [61]. The main difference between this strategy and the others is the post-processing step based on the lesion neighbourhood. This suggests that masking methods are expected to produce a higher number of false positives since no spatial information is taken into account and no spatial coherence is actually preserved.

Another common issue among the cases analysed is the number of false positives located inside GM or close to external CSF and the cortex (as shown in figure A.8). These FP, which are detected due to a low FLAIR threshold, are usually kept during the post-processing due to their proximity to WM or because they are inside the GM mask. The low threshold values are obtained due to acquisition artifacts in the FLAIR images. These artifacts present a high bias field around the top and bottom slices of the images that usually corrupts the threshold computation. An example of these artifacts can be seen in figure A.9.

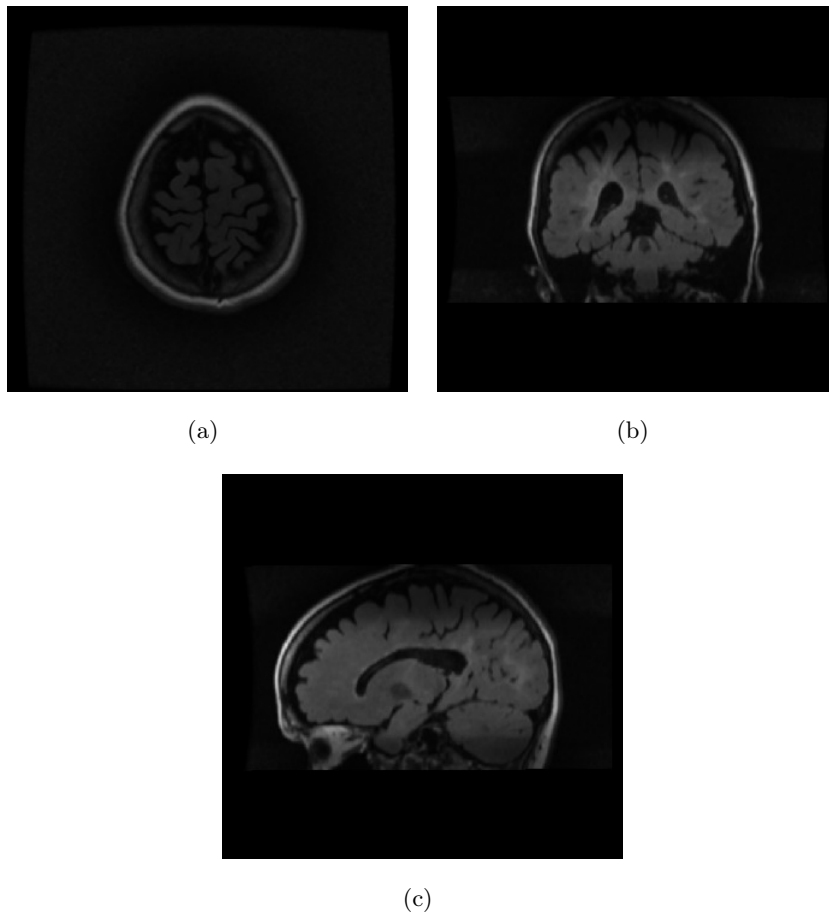


Figure A.9: Example of a volume with a high bias field in the top and bottom slices.

A.3 Markov random fields

Markov random fields (MRF) have been widely used in image processing in order to introduce neighbourhood and spatial constraints [138, 108]. Methods relying on MRF define the image as a graph in which each node represents a voxel connected to its neighbouring voxels. Each of these nodes belongs to a classe $c \in 1, \dots, K$, where K is the number of nodes. Following markovian properties, the probability of this class $p(c)$ is independent of the rest of the nodes given a neighbourhood set:

$$p(c_i | c_{S-\{i\}}) = p(c_i | c_{\mathfrak{N}_i}), \quad (\text{A.6})$$

where $j \in \mathfrak{N}_i \iff i \in \mathfrak{N}_j$ and S represents the collection of all the sites in the image.

According to Hammersley-Clifford's theorem [27], an MRF is equivalent to a Gibbs

distribution:

$$p(c) = \frac{1}{Z} e^{-U(c)}, \quad (\text{A.7})$$

where Z is a normalisation factor and $U(c)$ is the energy function that encodes the spatial constraints. This function is computed over the neighbours of each voxel and can be tuned in order to penalise or prioritise certain neighbourhood configurations in order to introduce *a priori* knowledge.

A.3.1 MRF methods in MS segmentation

Three different approaches for MS lesion segmentation modeling the image as an MRF have been presented in the literature [135, 121, 209]. These approaches can be divided into two large groups: those methods based on an iterative approach that estimates an statistical intensity model of the image and benefits from the MRF, re-estimated at each step, to obtain a smooth and spatially coherent segmentation [135, 121], and those methods that model the MRF once, defining a final segmentation [209].

Both approaches have the MRF estimation step in common, which is usually defined as an optimization where an energy term E is minimised. This energy term can be defined as

$$E = E_{data} + \lambda E_{smooth}, \quad (\text{A.8})$$

where E_{data} is an energy term that depends on the data but is completely independent for each voxel, E_{smooth} is the smoothness term that takes into account the relationship between neighboring voxels, and λ is a parameter to weight the degree of smoothness and can be obtained from equation A.7. If no smoothness term is provided, the optimization behaves like a typical segmentation algorithm, where a data model is already defined by the data, while, if there is no data term, the algorithm would tend to homogenise the segmentation, leaving a single region for the whole volume.

For the data term, a Gaussian mixture model (GMM) is usually assumed. For statistical iterative algorithms (such as the expectation maximisation [135] or the adaptive mixture method [121]), this model is estimated at each step and the new parameters are used to reestimate the MRF, while for the other approaches, an intensity model is first learned (based, for instance, on histograms [209]) and then the model is kept for the whole segmentation.

For the smoothness term, different neighboring configurations can be taken into account. For instance, a global affinity matrix defining possible connections between neighboring voxels (with a lower energy value) and impossible ones (with a higher penalizing value) can be defined. Local neighboring approaches with different affinity matrices depending on the region can also be applied in order to better encode *a priori* knowledge based on structures or tissues.

With these two terms defined, the total energy term can be optimised using different algorithms, for instance the iterated conditional modes (ICM) [135], the simulated annealing algorithm [209], or estimated using the segmentation obtained at each step [121]. In order to accomplish these goals, we decided to implement two MRF filtering strategies: a global one based on graph cuts (GC) [212, 95] and a local one based on the iterated conditional modes algorithm (ICM) [212].

A.3.2 Motivation

In the previous section, we have seen how the pipelines analysed present a high number of false positives with mean values over 80%. Moreover, we have also seen how the best lesion segmentation strategy relies on a post-processing step based on neighbourhood information. Furthermore, some of the false positives are located outside the WM or in the vicinity of highly unlikely tissues.

Alternatively, when using the EM tissue method, most of the WML are classified in a first stage as PV. Some of these PV blobs are kept as PV during the lesion segmentation step reducing the TP due to their contrast and surroundings. Usually, these regions are close to the ventricles or completely surrounded by WM, suggesting a high probability of actually being a WML.

One of the best options to improve the abovementioned results would be to implement a MRF strategy encoding those previous considerations to reduce the number of false positive and increase (when possible) the number of true positive.

A.3.3 Considerations

We have presented a generical MRF model optimization with two energy terms: the data term and the smoothness term. Moreover, we have presented an initial segmentation of the brain tissues and WML (see figure A.8) we want to segment. This segmentation can include PV, and some of the WML objects are in fact FP belonging to either GM or CSF.

In order to define our post-processing algorithm, some considerations must be taken into account inside this MRF model. In this section, we present them.

- **Ventricles:** Inside the brain, we can differentiate between two kinds of CSF: external and internal. Even though there are minor intensity differences between them, they are spatially and geometrically different. While external CSF appears surrounding the cortex and presents *random* sulci and gyri patterns, internal CSF appears inside the WM forming the ventricular system. These ventricles have a general common shape among subjects but present some anatomical deviations in individuals. Moreover, some WML appear next to these ventricles while having any lesion in contact with external CSF is impossible since there is no neighboring WM region. In order to better encode these differences, we decided to encode an anatomic atlas to differentiate between the CSF class and the ventricle class to better outline our MRF configuration, as shown in figure A.10. This atlas also includes the GM and WM labels and could be extended to introduce other structures.
- **Partial volume:** We defined PV voxels as a mixture of GM and CSF, therefore, these voxels should actually be the interface between these two tissues. As a result, PV should never appear inside WM. Most of these PV regions surrounded by WM are in fact lesions that share the intensity distribution with the PV class. With this knowledge, these regions could be easily reclassified as WML instead of PV using an MRF model, like the one presented here.
- **Data term:** In the previous subsections, we have seen how a data term is needed when optimizing a MRF model in order to prevent a full homogenisation of the initial segmentation. Including a real data term based on the intensities would cause a reclassification of the whole image that would not take into account this initial segmentation. However, with this post-processing step, we only want to reclassify an initial segmentation, keeping the normal appearing brain tissues intact. In order to encode these constraints, we define three possible data values: infinite, change and zero.

A value of *infinite* defines an impossible change. For instance, if we have a voxel labelled as CSF and we do not want it to change, the data cost of changing this voxel to the tissue GM would be infinite, thus preventing its change during the minimization. Since the value of infinite can not be represented with a computer, we chose a high enough value as infinite.

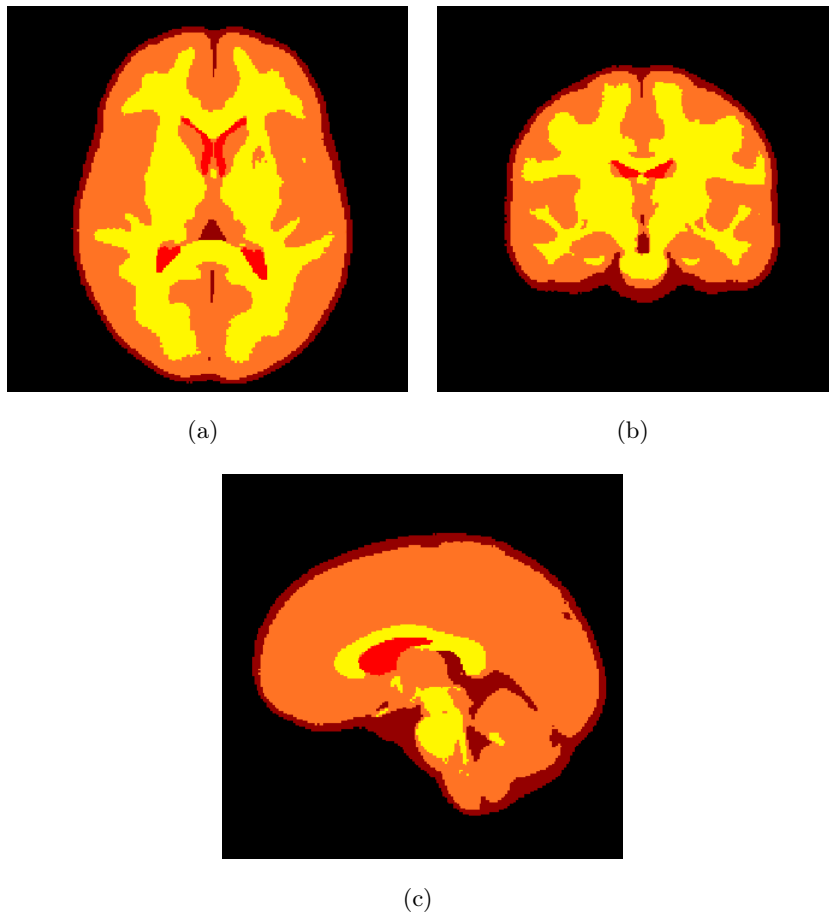


Figure A.10: Anatomical atlas with the following labels: GM (orange), WM (yellow), ventricular system (red) and external CSF (brown)

A value of *zero* is used to keep the same label. For all classes, the data cost of keeping the label will therefore be 0. With this value and the previous one, we can preserve the label of those voxels we do not want to change.

Finally, the value for *change* is slightly higher than zero. This value is closely related to the λ value and the neighbourhood configuration. For example, if we have a PV object, we might want to change its label to WML if it is surrounded by WM, but if it is surrounded by GM, we do not want to modify it. Using this *change* value to pass from PV to WML, we can prevent a change in the second case depending on the smoothness cost of PV as a neighbor of GM.

The final data cost configuration can be expressed as a matrix, shown in table A.2, where the rows represent the initial class of the voxel and the columns represent the

	BCK	CSF	VEN	GM	WM	PV	WML
BCK	0	∞	∞	∞	∞	∞	∞
CSF	∞	0	∞	∞	∞	∞	∞
VEN	∞	∞	0	∞	∞	∞	∞
GM	∞	∞	∞	0	∞	∞	∞
WM	∞	∞	∞	∞	0	∞	∞
PV	∞	∞	∞	∞	∞	0	<i>ch</i>
WML	∞	<i>ch</i>	∞	<i>ch</i>	∞	∞	0

Table A.2: Data cost affinity matrix. Rows represent the class of the voxel for the initial segmentation, and columns represent the new class. The classes are background (BCK), external fluid (CSF), the ventricular system (VEN), grey matter (GM), white matter (WM), partial volume (PV) and lesions (WML). The value *ch* represents the *change* value.

	BCK	CSF	VEN	GM	WM	PV	WML
BCK	0	0	1	1	1	1	1
CSF	0	0	1	1	1	0	1
VEN	1	1	0	1	1	1	0
GM	1	1	1	0	0	1	1
WM	1	1	1	0	0	1	0
PV	1	0	1	1	1	0	1
WML	1	1	0	1	0	1	0

Table A.3: Smoothness cost affinity matrix. The classes are background (BCK), external fluid (CSF), the ventricular system (VEN), grey matter (GM), white matter (WM), partial volume (PV) and lesions (WML).

cost of changing this voxel to a new class.

- **Smoothness term:** We model the smoothness term with a simple approach defined by two values: zero and one. The value zero defines a possible configuration, for instance, WML can be surrounded by WM. The value one defines a configuration we do not allow, for instance WML should not be in contact with external CSF. The different configuration values can be summarised by the affinity matrix shown in A.3.

Following this configuration, we want to reduce the number of FP lesions and increase the number of TP, reclassifying some missclassified PV. In order to penalise the unlikely configurations higher, the λ parameter must be tuned according to the *change* data cost value, however, this affinity matrix should not be modified.

All these considerations are applicable to any optimization algorithm, either global or local. In the next sections, we will present two different algorithms used to refine our initial lesion segmentation, both based on these assumptions. The main differences between these algorithms is that one estimates a global energy minima while the other falls into a local minimum. The idea of this comparison is to define the best solution to optimise our initial segmentation.

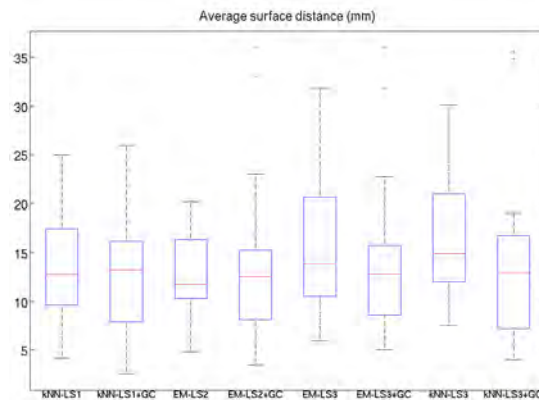
A.3.4 Graph cuts (GC)

The first approach we present is a global optimization algorithm based on graph cuts. In this approach, a graph is created with the voxels as intermediate nodes connected with their neighbours and the classes as terminal nodes. Therefore, the segmentation is defined by connecting each node to the terminal node corresponding to the final label. The links between voxels represent the smoothness costs, while the links between the voxels and the terminal nodes represent the data cost. The algorithm is called expansion-move and tries to expand each class at each iteration making expansion moves. The algorithm quickly converges to a local minimum (close to the global one) where no other expansion move can be made.

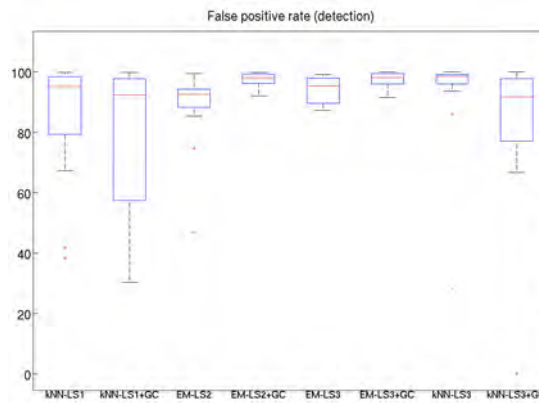
Implementation

The code for this filter is implemented in C++. The main routines for the graph cuts optimisation are implemented as a stand-alone library. However, its functionality is wrapped into the class `BrainSegmentation`, which is compatible with the ITK code. This class implements a method to apply the MRF filtering to a labelled image with tissues and lesions. The input parameters for this method are the labeled image, the λ value, the *change* value and the maximum number of iterations. Two other methods to define the smoothness array matrix and the data costs are also provided.

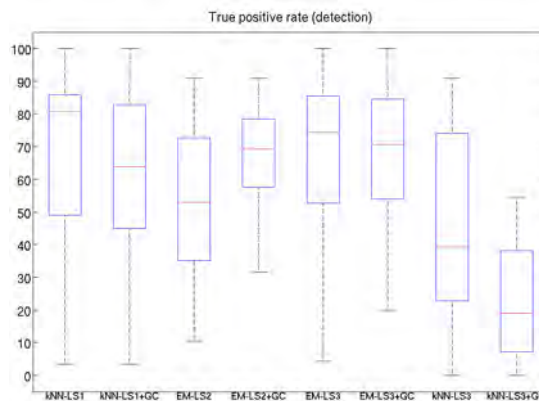
Generally, this algorithm takes only a few minutes to filter the labeled image. The initialisation steps to define the data and smoothness costs take only a few seconds. However, the main drawback of this algorithm is the high amount of memory required for its execution. Therefore, images of more than a hundred slices should be downsampled.



(a) Average surface distance



(b) False positive rate



(c) True positive rate

Figure A.11: The methods evaluated comprising the 4 original pipelines after the MRF filtering with GC optimisation. (a) Average surface distance boxplot, (b) false positive fraction, and (c) true positive fraction.

Results

Figure A.11 summarises the results obtained before and after the MRF filtering using the GC optimisation algorithm. We can differentiate between the results applied to the kNN-based pipelines and the EM-based ones since the behavior of the filtering is slightly different.

As expected, the kNN-based pipelines have a decreased value for the detection measures since there is no reclassification into WML. This points to a slight improvement in both segmentation and detection due to the FP reduction as unlikely lesions are reclassified into tissue, while sacrificing some of the previously segmented lesions. These errors are lesions close to the cortex and are reclassified into GM.

On the other hand, the EM-based pipelines present an increase in the detection measure values. This points to an increase of the number of lesions detected. These new lesions are reclassified PV inside WM that were discarded during the lesion segmentation step. As a drawback, some other PV regions close to the cortex are reclassified as lesions, increasing the FP rate. Note that these reclassified regions are higher in number than the number of previous FP in unlikely regions reclassified into tissue.

Finally, if we look at the distance measures, we can see how the new lesion segmentation method obtains lower values while the other two methods present a slight increase. This increase is due to extreme values in some images since the first and third quartils present lower values that suggest that the overall average surface distance has decreased.

A.3.5 Iterated conditional modes (ICM)

This second approach is a pure local optimization algorithm based on the iterated conditional modes. It starts with the initial segmentation and, at each step, gives the labeling with lower total cost to each voxel. This process is iterated until no further change is applied. This method is highly sensitive to the initial segmentation, however, this drawback can be an advantage since we actually want to refine this initial segmentation. We are looking, therefore, for a local minima based on our initial estimation.

Another major difference between this method and the GC is the possibility of defining a different membership kernel. This kernel is defined as a sphere with radius r . This difference allows us to take into account a higher number of neighbours in order to make a decision.

Implementation

The code for this filter is implemented in C++ and is based on ITK inside the class `BrainSegmentation` as well. This class implements a method to apply the MRF filtering to a labeled image with tissues and lesions. The input parameters for this method are the labeled image, the λ value, the *change* value and the radius r .

Generally, the code takes only a few minutes to filter the labeled image, however, the time needed for convergence highly depends on the number of voxels needing to be reclassified. No initialisation is needed since the cost is recalculated at each step, and only for those voxels whose neighbor has changed. In this sense, the first iterations require more time, since a higher number of voxels are expected to change class, than the last ones, where only a few costs are calculated.

Results

Figure A.12 summarises the results obtained before and after the MRF filtering using the ICM optimisation algorithm. Here we can also differentiate between the results obtained by the EM-based pipelines and the kNN-based ones in terms of FP and TP.

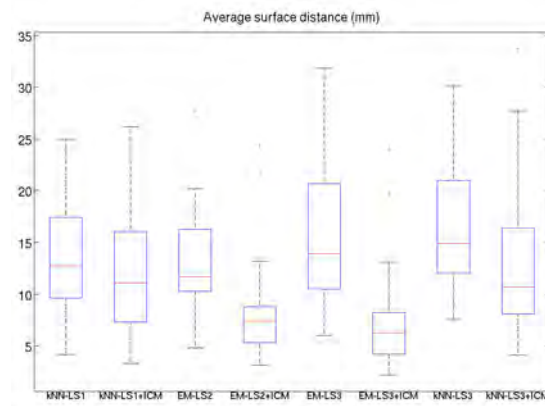
In a similar manner, the kNN-based pipelines present a decrease in the number of FP while decreasing in a similar percentage the TP rate. On the other hand, the EM methods obtain similar FP rates without a huge decrease in the number of TP. Moreover, the EM+LS3 pipeline decreases the mean FP rate value after the MRF filtering with the ICM optimisation.

Finally, all the pipelines present an improvement in lesion segmentation, reducing the average surface distance with a higher increase obtained by the EM-based pipelines.

A.3.6 Discussion

We have presented two different post-processing algorithms based on MRF. Moreover, the main difference between these two methods comes from the optimisation algorithm used to estimate the hidden MRF. While the GC algorithm obtains a global energy result, the ICM algorithm optimises the original segmentation locally and thus depends on the initial segmentation estimate.

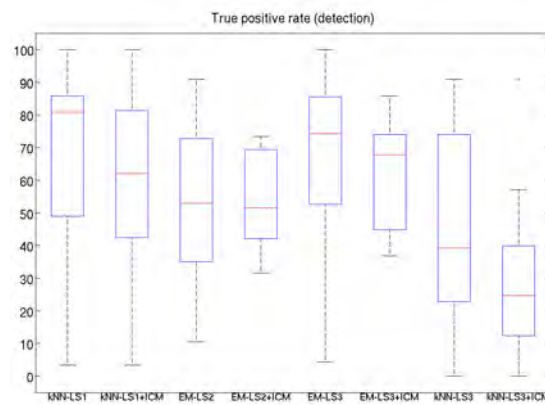
Figure A.13 offers a comparison in terms of evaluation measures between the two post-processing strategies. The local nature of the ICM algorithm allows the optimisation to



(a) Average surface distance

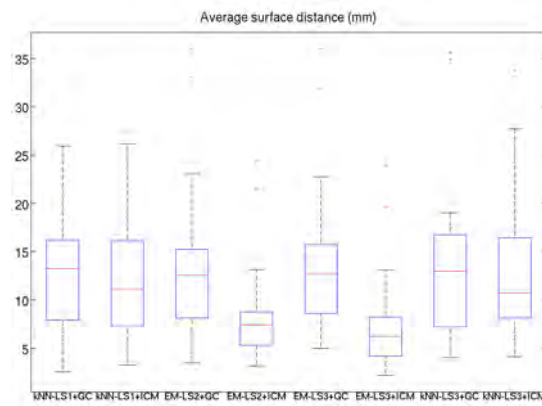


(b) False positive rate

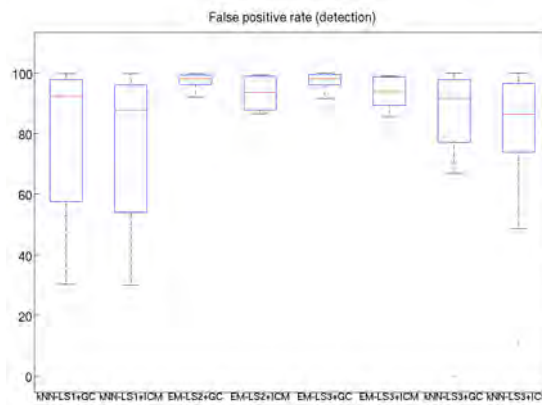


(c) True positive rate

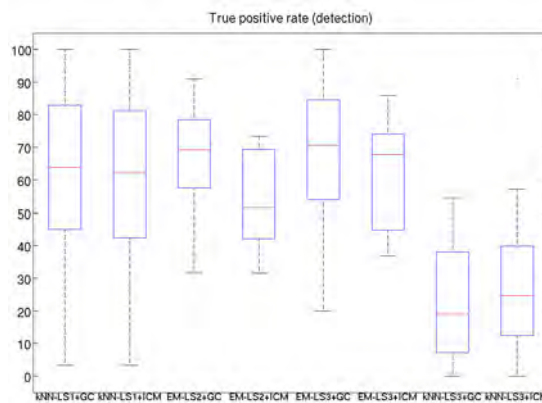
Figure A.12: The methods evaluated comprise the 4 original pipelines after the MRF filtering with ICM optimisation. (a) Average surface distance boxplot, (b) False positive fraction, (c) True positive fraction.



(a) Average surface distance



(b) False positive rate



(c) True positive rate

Figure A.13: Comparison of the methods evaluated comprising the 4 original pipelines after the MRF filtering with ICM and GC optimisation. (a) Average surface distance boxplot, (b) false positive fraction, and (c) true positive fraction.

reclassify a higher number of lesions, decreasing both the number of FP and TP for all the pipelines. Moreover, the lowest average surface values are also achieved using this second optimisation scheme. This can be explained by the need to reduce the energy and the configuration of both the smoothness and data term globally.

In order to reclassify a lesion, we need to change all its voxels to another tissue. This change has both a data cost value and a smoothness cost value. The new data cost is determined by the number of voxels that define the lesion (N_{area}) and the ch value:

$$E'_{data} = N_{area} \cdot ch, \quad (A.9)$$

while the old smoothness cost value depends on the neighboring voxels belonging to external CSF or GM ($\aleph_{\{GM,CSF\}}$):

$$E_{smooth} = \lambda \cdot |\aleph_{\{GM,CSF\}}|, \quad (A.10)$$

and the new smoothness cost value depends on the neighboring WM or ventricular voxels ($\aleph_{\{WM,VEN\}}$):

$$E'_{smooth} = \lambda \cdot |\aleph_{\{WM,VEN\}}|, \quad (A.11)$$

where $\aleph_{\{CSF,GM\}}$ and $\aleph_{\{WM\}}$ depend on the number of voxels on the lesion's perimeter (N_{border}).

In order for a lesion to be reclassified into CSF or GM, the following inequality must be true:

$$E'_{data} + E'_{smooth} < E_{smooth}, \quad (A.12)$$

since the old data cost is 0. In general, lesions in an unlikely spot (external CSF or cortical GM) have an E'_{smooth} close to 0 because the lesion is surrounded by either external CSF or GM, therefore $E'_{data} < E_{smooth}$. While E'_{data} depends on ch and N_{area} exclusively, E_{smooth} depends on λ and $\aleph_{\{GM,CSF\}}$. Assuming the same values for ch and λ , the inequality becomes: $N_{area} < \sum_i^{N_{border}} \aleph_{i\{GM,CSF\}}$, where $\aleph_{i\{GM,CSF\}}$ are the neighboring voxels of external CSF and GM for each lesion border voxel. Finally we can define:

$$\sum_i^{N_{border}} \aleph_{i\{GM,CSF\}} = \bar{\aleph}_{\{GM,CSF\}} \cdot N_{border}, \quad (A.13)$$

where $\bar{N}_{\{GM,CSF\}}$ is the average number of external CSF and GM voxels for each lesion border voxel. Therefore, the final inequality that must be true to reclassify a lesion region into tissue is the following:

$$N_{area} < \bar{N}_{\{GM,CSF\}} \cdot N_{border}. \quad (\text{A.14})$$

Taking into account this final inequality, we can conclude that large lesions in an unlikely spot with a small enough border may not be reclassified using the GC optimisation algorithm, since the global energy would increase when reclassifying this FP region. This is not the case for the ICM algorithm, as it would reclassify each voxel according to its energy change at each step. Therefore, the new energy value for a voxel of a lesion in an unlikely spot should be lower if the voxel is reclassified into tissue.

If a high λ value is given to avoid this issue when using the GC algorithm, this value will also affect lesions close to the cortex that are in fact WML. If the smoothness cost to keep the lesion is higher than the data cost to change the lesion region into tissue, the number of those lesions close to GM could be removed, decreasing the number of TP. Therefore, the λ and ch values should be defined according to each data set to ensure the highest accuracy.

Figure A.14 shows an example in which the global optimisation may not reclassify a FP lesion. In this 2D example, we have 24 neighboring GM voxels, which, counted as neighbours, amount to 36 voxels since some of these voxels are neighbours to more than 1 lesion voxel. On average, each border voxel has 1.5 GM neighboring voxels and the smoothness cost, if we keep the lesion, is $36 \cdot \lambda$, while this value decreases to 0 if we change the whole lesion to GM. On the other hand, we have 44 lesion voxels with 20 of them in the border, therefore, the global data cost, if we reclassify the lesion, would be $44 \cdot ch$ and 0 if we keep it as lesion. If $\lambda = ch$, this lesion would stay as lesion, even though it clearly is inside GM. Using an ICM strategy with the same parameters, the optimisation would slowly erode the border of the lesion until it is finally reclassified as GM. However, we could also have $\lambda > ch$ in order to remove the lesion using the GC optimisation, but this value increase may also affect other lesions close to the cortex.

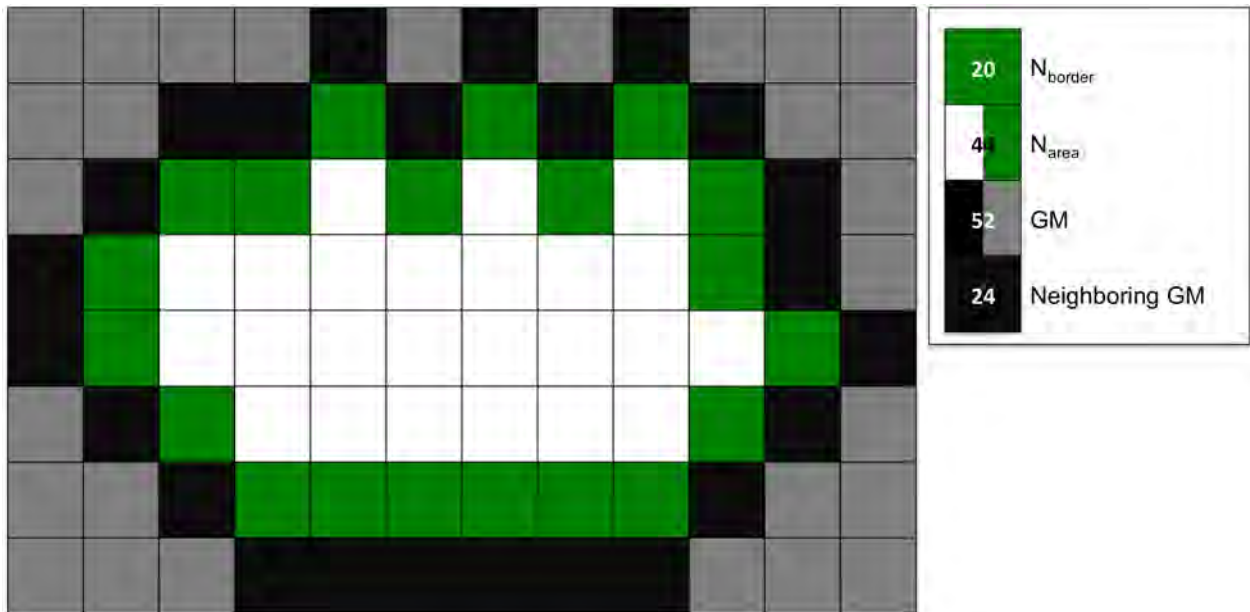


Figure A.14: Example of a 2D lesion inside GM. If $\lambda = ch$, this lesion will not be reclassified into GM, even though it is inside it. However, if $\lambda > ch$, the lesion will be reclassified, but other lesions close to the cortex might be reclassified as well.

A.4 Conclusions

A.4.1 Discussion

In this report, we presented and evaluated 7 different lesion segmentation pipelines based on 3 atlas-based tissue segmentation methods (EM, kNN and SPM) and 3 lesion segmentation approaches based on thresholding a FLAIR image. Four of these pipelines were already presented and tested in [37]. However, a quantitative evaluation was only provided for the Brainweb data with a nonrealistic FLAIR simulation, therefore, we extended this evaluation with a realistic FLAIR simulation and a quantitative analysis of real datasets from the MICCAI Challenge. Moreover, we compared those results to three new pipelines using the SPM tissue segmentation approach combined with the 3 lesion segmentation methods developed.

Thanks to this validation, we discovered several issues in the lesion segmentation that could be improved. These issues were a low number of TP due to some being misclassified as PV when using the EM approach, and a high number of FP, as some were located inside GM and/or external CSF. In order to improve these initial results, we presented two new

post-processing algorithms based on an MRF optimisation approach with a global and local optimisation algorithm. These filters refine the initial segmentation obtained with the lesion segmentation pipelines taking into account the label information and some MRF configurations defined in order to remove lesions in unlikely locations and reclassify PV regions inside WM or close to the ventricular system. We have seen how both methods present an improvement over the initial results with a slight reduction of the number of FP and the average surface distance between the ground truth and the automatically segmented lesions. However, the number of TP also presents a slight decrease due to some lesions being close to GM.

A.4.2 Impact on this PhD thesis

This initial validation of the proposals presented in my master thesis [37] with both synthetic and real data of 3T images proved essential to pinpoint weaknesses from our initial approaches and point out the importance of spatial information. In this report, all spatial information is encoded using a specific implementation of MRF. However, the MRF theory basically states that the voxel's probability of belonging to a certain distribution is only determined by the directly neighbouring voxels in a given 3D connectivity.

In our first proposal, we presented a lighter implementation of this concept by defining the prior probabilities according to a 3D connectivity and the probability maps of the previous step. This prior probabilities are combined to atlas probabilities using a similarity map as explained in chapter 4. In our second proposal, instead of using an MRF model, in which only direct neighbours are taken into account, we encoded the contextual information of randomly sampled regions inside a pre-defined 3D window around each voxel. Thus, we proposed including spatial information in both approaches to aid the segmentation.

Finally, the whole TISSUE-LOT proposal was implemented taking the conclusions in section A.2.6 into account. For instance, the tissue segmentation was improved in terms of convergence and smoothness, and the lesion segmentation step was defined as an extension of LS1, which obtained the best results in this report.

Bibliography

- [1] L.S. Ait-Ali, S. Prima, P. Hellier, B. Carsin, G. Edan, , and C. Barillot. STREM: a robust multidimensional parametric method to segment MS lesions in MRI. In *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, pages 409 – 416, 2005.
- [2] A. Akselrod-Ballin, M. Galun, J.M. Gomori, M. Filippi, P. Valsasina, R. Basri, and A. Brandt. Automatic segmentation and classification of multiple sclerosis in multi-channel MRI. *IEEE Trans. Biomed. Eng.*, 56(10):2461 – 2469, 2009.
- [3] B. Alfano, A. Brunetti, M. Larobina, M. Quarantelli, E. Tedeschi, A. Ciarmiello, E. M. Covelli, and M. Salvatore. Automated segmentation and measurement of global white matter lesion volume in patients with multiple sclerosis. *J. Magn. Reson. Imag.*, 12(6):799 – 807, 2000.
- [4] P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *NeuroImage*, 46(3):726 – 738, 2009.
- [5] P. Anbeek, K. L. Vincken, G. S. van Bochove, M. J.P. van Osch, and J. van der Grond. Probabilistic segmentation of brain tissue in MR imaging. *NeuroImage*, 27(4):795 – 804, 2005.
- [6] P. Anbeek, K. L. Vincken, M. J.P. van Osch, R. H.C. Bisschops, and J. van der Grond. Probabilistic segmentation of white matter lesions in MR imaging. *NeuroImage*, 21(3):1037 – 1044, 2004.
- [7] P. Anbeek, K. L. Vincken, and M. A. Viergever. Automated MS-lesion segmentation by k-nearest neighbor classification. In *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, 2008.

-
- [8] A. Andronache, M. von Siebenthal, G. Szekely, and Ph. Cattin. Non-rigid registration of multi-modal images using both mutual information and cross-correlation. *Med. Image Anal.*, 12(1):3 – 15, 2008.
- [9] E. Angelini, Y. Jin, and A. Laine. State of the Art of Level Set Methods in Segmentation and Registration of Medical Imaging Modalities. In *Handbook of Biomedical Image Analysis*, pages 47 – 101. Springer US, 2007.
- [10] X. Artaechevarria, A. Muñoz-Barrutia, and C. Ortiz-de Solorzano. Combination strategies in multi-atlas image segmentation: Application to brain MR data. *IEEE Trans. Med. Imag.*, 28(8):1266 – 1277, 2009.
- [11] J. Ashburner and K. J. Friston. Nonlinear spatial normalization using basis functions. *Hum. Brain Mapp.*, 7(4):209 – 217, 1999.
- [12] J. Ashburner and K. J. Friston. Voxel-based morphometry: the methods. *NeuroImage*, 11(6):805 – 821, 2000.
- [13] J. Ashburner and K. J. Friston. Unified segmentation. *NeuroImage*, 26(3):839 – 851, 2005.
- [14] S. P. Awate, T. Tasdizen, N. Foster, and R. T. Whitaker. Adaptive Markov modeling for mutual-information-based, unsupervised MRI brain-tissue classification. *Med. Image Anal.*, 10(5):726 – 739, 2006.
- [15] M. Bach-Cuadra, L. Cammoun, T. Butz, O. Cuisenaire, and J. P. Thiran. Comparison and validation of tissue modelization and statistical classification methods in T1-weighted MR brain images. *IEEE Trans. Med. Imag.*, 24(12):1548 – 1565, 2005.
- [16] M. Bach-Cuadra, M. De Craene, V. Duay, B. Macq, C. Pollo, and J. P. Thiran. Dense deformation field estimation for atlas-based segmentation of pathological MR brain images. *Comput. Meth. Prog. Biomed.*, 84(2 - 3):66 – 75, 2006.
- [17] M. Bach-Cuadra, C. Pollo, A. Bardera, O. Cuisenaire, J. G. Villemure, and J. P. Thiran. Atlas-based segmentation of pathological MR brain images using a model of lesion growth. *IEEE Trans. Med. Imag.*, 23(10):1301 – 1314, 2004.
- [18] M. Bach-Cuadra, M. Schaer, A. Andre, L. Guibaud, S. Eliez, and J. Ph. Thiran. Brain tissue segmentation of fetal MR images. In *Work. on Image Anal. Dev. Brain*, pages 1 – 9, 2009.

- [19] C. Baillard, P. Hellier, and C. Barillot. Segmentation of brain 3D MR images using level sets and dense registration. *Med. Image Anal.*, 5(3):185 – 194, 2001.
- [20] R. Bajcsy. Digital anatomy atlas and its registration to MRI, fMRI, PET: The past presents a future. In *Int. Workshop on Biomedical Image Registration*, pages 201 – 211, 2003.
- [21] R. Bajcsy, R. Lieberman, and M. Reivich. A computerized system for the elastic matching of deformed radiographic images to idealized atlas images. *J. Comp. Assist. Tomo.*, 7(4):618 – 625, 1983.
- [22] S. Banerjee, D. P. Mukherjee, and D. D. Majumdar. Fuzzy c-means approach to tissue classification in multimodal medical imaging. *Inform. Sciences*, 115(1 - 4):261 – 279, 1999.
- [23] A. Bardera, M. Feixas, I. Boada, and M. Sbert. Image registration by compression. *Inform. Sciences*, 180(7):1121 – 1133, 2010.
- [24] F. Barkhof, M. Filippi, D. H. Miller, P. Scheltens, A. Campi, C. H. Polman, G. Comi, H. J. Adèr, N. Losseff, and J. Valk. Comparison of MRI criteria at first presentation to predict conversion to clinically definite multiple sclerosis. *Brain*, 120(11):2059 – 2069, 1997.
- [25] P. L. Bazin and D. L. Pham. Homeomorphic brain image segmentation with topological and statistical atlases. *Med. Image Anal.*, 12(5):616 – 625, 2008.
- [26] B. J. Bedell and P. A. Narayana. Automatic segmentation of gadolinium-enhanced multiple sclerosis lesions. *Magn. Reson. Med.*, 39(6):935 – 940, 1998.
- [27] J. Besag. Spatial interaction and statistical analysis of lattice systems. *J. R. Stat. Soc. B*, 36(2):192 – 236, 1974.
- [28] J. Besag. On the statistical analysis of dirty pictures. *J. R. Stat. Soc. B*, 48(3):259 – 302, 1986.
- [29] J. C. Bezdek. *Pattern Recognition With Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981.
- [30] K. K. Bhatia, J. V. Hajnal, B. K. Puri, A. Edwards, and D. Rueckert. Consistent groupwise non-rigid registration for atlas construction. In *IEEE Int. Symp. Biomed. Imag.*, pages 201 – 211, 2003.

-
- [31] K. Boesen, K. Rehm, K. Schaper, S. Stoltzner, R. Woods, E. Lüders, and D. Rotenberg. Quantitative comparison of four brain extraction algorithms. *NeuroImage*, 22(3):1255 – 1261, 2004.
- [32] Hervé Boisgontier, Vincent Noblet, Fabrice Heitz, Lucien Rumbach, and Jean-Paul Armspach. Generalized likelihood ratio tests for change detection in diffusion tensor images: Application to multiple sclerosis. *Med. Image Anal.*, 16(1):325 – 338, 2012.
- [33] A.O. Boudraa, S. M. R. Dehak, Y.M. Zhu, C. Pachai, Y.G. Bao, and J. Grimaud. Automated segmentation of multiple sclerosis lesions in multispectral MR imaging using fuzzy clustering. *Comput. Biol. Med.*, 30(1):23 – 40, 2000.
- [34] P.A. Brex, O. Ciccarelli, J.I. O’Riordan, M. Sailer, A.J. Thompson, and D.H. Miller. A longitudinal study of abnormalities on MRI and disability from multiple sclerosis. *New Engl. J. Med.*, 346(3):158 – 164, 2002.
- [35] S. Bricq, C. Collet, and J. P. Armspach. MS lesion segmentation based on hidden Markov chains. In *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, 2008.
- [36] S. Bricq, Ch. Collet, and J. P. Armspach. Unifying framework for multimodal brain MRI segmentation based on Hidden Markov Chains. *Med. Image Anal.*, 12(6):639 – 652, 2008.
- [37] M. Cabezas. Atlas-based segmentation of brain mri: Application to multiple sclerosis. Master’s thesis, Universitat de Girona, Spain, 2010.
- [38] M. Cabezas, A. Oliver, J. Freixenet, and X. Lladó. A supervised approach for multiple sclerosis lesion segmentation using context features and an outlier map. In *Iberian Conf. Pattern Recog. Image Anal.*, pages 782 – 789, 2013.
- [39] M. Cabezas, A. Oliver, X. Lladó, J. Freixenet, and M. Bach Cuadra. A review of atlas-based segmentation for magnetic resonance brain images. *Comput. Meth. Prog. Biomed.*, 104(3):e158 – e177, 2011.
- [40] R. Cárdenes, R. de Luis-García, and M. Bach-Cuadra. A multidimensional segmentation evaluation for medical image data. *Comput. Meth. Prog. Biomed.*, 96(2):108 – 124, 2009.
- [41] G. E. Christensen, R. D. Rabbitt, and M. I. Miller. Deformable templates using large deformation kinematics. *IEEE Trans. Image Processing*, 5(10):1435 – 1447, 1996.

- [42] C. Ciofalo and C. Barillot. Atlas-based segmentation of 3D cerebral structures with competitive level sets and fuzzy control. *Med. Image Anal.*, 13(3):456 – 470, 2009.
- [43] L. P. Clarke, R. P. Velthuizen, M. A. Camacho, J. J. Heine, M. Vaidyanathan, L. O. Hall, R. W. Thatcher, and M. L. Silbiger. MRI segmentation: Methods and applications. *Magn. Reson. Imag.*, 13(3):343 – 368, 1995.
- [44] I. Claude, J. L. Daire, and G. Sebag. Fetal brain MRI: segmentation and biometric analysis of the posterior fossa. *IEEE Trans. Biomed. Eng.*, 51(4):617 – 626, 2004.
- [45] C. A. Cocosco, A. P. Zijdenbos, and A. C. Evans. A fully automatic and robust brain MRI tissue classification method. *Med. Image Anal.*, 7(4):513 – 527, 2003.
- [46] D. L. Collins, C. J. Holmes, T. M. Peters, and A. C. Evans. Automatic 3D model-based neuroanatomical segmentation. *Hum. Brain Mapp.*, 3(3):190 – 208, 1995.
- [47] D. L. Collins, A. P. Zijdenbos, V. Kollokian, and J. G. Sled. Design and construction of a realistic digital brain phantom. *IEEE Trans. Med. Imag.*, 17(3):463 – 468, 1998. <http://www.bic.mni.mcgill.ca/brainweb>.
- [48] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Machine Intell.*, 24(5):603 – 619, 2002.
- [49] O. Commowick and G. Malandain. Evaluation of atlas construction strategies in the context of radiotherapy planning. In *Statist. Atlases Personal. Models Work.*, pages 1 – 4, 2006. Held in conjunction with MICCAI.
- [50] A. Compston and A. Coles. Multiple sclerosis. *Lancet*, 359(9313):1221 – 1231, 2002.
- [51] T. Cootes, C. Beeston, G. Edwards, and C. Taylor. A unified framework for atlas matching using active appearance models. In *Int. Conf. Image Proc. Med. Ima.*, pages 322 – 333, 1999.
- [52] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models - Their training and application. *Comput. Vis. Image Underst.*, 61(1):38 – 59, 1995.
- [53] S. Datta and P. A. Narayana. Automated brain extraction from T2-weighted magnetic resonance images. *J. Magn. Reson. Imag.*, 33(4):822 – 829, 2011.
- [54] S. Datta, B. R. Sajja, R. He, R. K. Gupta, J. S. Wolinsky, and P. A. Narayana. Segmentation of gadolinium-enhanced lesions on MRI in multiple sclerosis. *J. Magn. Reson. Imag.*, 25(5):932 – 937, 2007.

- [55] S. Datta, B. R. Sajja, R. He, J. S. Wolinsky, R. K. Gupta, and P. A. Narayana. Segmentation and quantification of black holes in multiple sclerosis. *NeuroImage*, 29(2):467 – 474, 2006.
- [56] S. Datta, G. Tao, R. He, J.S. Wolinsky, and P. A. Narayana. Improved cerebellar tissue classification on magnetic resonance images of brain. *J. Magn. Reson.*, 29(5):1035 – 1042, 2009.
- [57] C. Davatzikos. Spatial transformation and registration of brain images using elastically deformable models. *Comput. Vis. Image Underst.*, 66(2):207 – 222, 1997.
- [58] B. M. Dawant, S. L. Hartmann, and S. Gadamsetty. Brain atlas deformation in the presence of large space-occupying tumors. In *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, pages 589 – 596, 1999.
- [59] B. M. Dawant, S. L. Hartmann, S. Pan, and S. Gadamsetty. Brain atlas deformation in the presence of small and large space-occupying tumors. *Comput. Aided Surgery*, 7(1):1 – 10, 2002.
- [60] B. M. Dawant, S. L. Hartmann, J. P. Thirion, F. Maes, D. Vandermeulen, and P. Demaerel. Automatic 3-D segmentation of internal structures of the head in MR images using a combination of similarity and free-form transformations. i. methodology and validation on normal subjects. *IEEE Trans. Med. Imag.*, 18(10):909 – 916, 1999.
- [61] R. de Boer, H. A. Vrooman, F. van der Lijn, M. W. Vernooij, M. A. Ikram, A. van der Lugt, M. M. B. Breteler, and W. J. Niessen. White matter lesion extension to automatic brain tissue segmentation on MRI. *NeuroImage*, 45(4):1151 – 1161, 2009.
- [62] M. De Craene, A. du Bois d’Aische, B. Macq, and S. K. Warfield. Multi-subject registration for unbiased statistical atlas construction. In *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, pages 655 – 662, 2004.
- [63] L. W. de Jong, K. van der Hiele, I. M. Veer, J. J. Houwing, R. G. J. Westendorp, E. L. E. M. Bollen, P. W. de Bruin, H. A. M. Middelkoop, M. A. van Buchem, and J. van der Grond. Strongly reduced volumes of putamen and thalamus in Alzheimer’s disease: an MRI study. *Brain*, 131(12):3277 – 3285, 2008.
- [64] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum-likelihood from incomplete data via EM algorithm. *J. R. Stat. Soc. B*, 39(1):1 – 38, 1977.

- [65] L. Devroye. On the almost everywhere convergence of nonparametric regression function estimates. *Ann. Stat.*, 9(6):1310 – 1319, 1981.
- [66] P. D’Haese, V. Duay, R. Li, A. du Bois d’Aische, T. Merchant, A. Cmelak, E. Donnelly, K. Niermann, B. Macq, and B. Dawant. Automatic Segmentation of Brain Structures for Radiation Therapy Planning. In *Medical Imaging: Image Processing*, ISCAS, pages 517 – 526. SPIE, 2003.
- [67] L. R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297 – 302, 1945.
- [68] B. Draganski, J. Ashburner, C. Hutton, F. Kherif, R.S.J. Frackowiak, G. Helms, and N. Weiskopf. Regional specificity of mri contrast parameter changes in normal ageing revealed by voxel-based quantification (vbq). *NeuroImage*, 5(4), 2011.
- [69] T. Dua and P. Rompani. *Atlas: Multiple sclerosis resources in the world 2008*. World Health Organization, 2008.
- [70] V. Duay, P. F. D’Haese, R. Li, and B. M. Dawant. Non-rigid registration algorithm with spatially varying stiffness properties. In *IEEE Int. Symp. Biomed. Imag.*, pages 408 – 411, 2004.
- [71] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, New York, 2 edition, 2001.
- [72] G. Dugas-Phocion, M.A.G. Ballester, G. Malandain, C. Lebrun, and N. Ayache. Improved EM-based tissue segmentation and partial volume effect quantification in multi-sequence brain MRI. In *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, pages 26 – 33, 2004.
- [73] C. Elliott, D. Collins, D. Arnold, and T. Arbel. Temporally consistent probabilistic detection of new multiple sclerosis lesions in brain MRI. *IEEE Trans. Med. Imag.*, PP(99):1–1, 2013.
- [74] M. M. Esiri and J. H. Morris. *The neuropathology of dementia*. Cambridge University Press, 2002.
- [75] A. Evans, D. Collins, P. Neelin, M. Kamber, and T. S. Marrett. Three-dimensional correlative imaging: applications in human brain mapping. In *Functional Neuroimaging: Technical Foundations*, pages 145 – 162. Academic Press, 1994.

- [76] Alan C. Evans, Andrew L. Janke, D. Louis Collins, and Sylvain Baillet. Brain templates and atlases. *NeuroImage*, 62(2):911 – 922, 2012.
- [77] F. Fazekas, F. Barkhof, M. Filippi, R. I. Grossman, D. K. B. Li, W. I. McDonald, H. F. McFarland, D. W. Paty, J. H. Simon, J. S. Wolinsky, and D. H. Miller. The contribution of magnetic resonance imaging to the diagnosis of multiple sclerosis. *Neurology*, 53(3):448 – 457, 1999.
- [78] C. Fennema-Notestine, I. B. Ozyurt, C. P. Clark, S. Morris, A. Bischoff-Grethe, M-W-Bondi, T. L. Jernigan, B. Fischl, F. Segonne, D. W. Shattuck, R. M. Leahy, D. E. Rex, A. W. Toga, K. H. Zou, and G. G. Brown. Quantitative evaluation of automated skull-stripping methods applied to contemporary and legacy images: effects of diagnosis, bias correction and slice location. *Hum. Brain Mapp.*, 27(2):99 – 113, 2006.
- [79] M. Filippi and F. Agosta. Imaging biomarkers in multiple sclerosis. *J. Magn. Reson. Imag.*, 31(4):770 – 788, 2010.
- [80] M. Filippi, V. Dousset, H.F. McFarland, D.H. Miller, and R.I. Grossman. Role of magnetic resonance imaging in the diagnosis and monitoring of multiple sclerosis: Consensus report of the white matter study group. *J. Magn. Reson. Imag.*, 15(5):499 – 504, 2002.
- [81] H. Finch. Comparison of distance measures in cluster analysis with dichotomous data. *J. Data Science*, 3(1):85 – 100, 2005.
- [82] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A. M. Dale. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341 – 355, 2002.
- [83] O. Freifeld, H. Greenspan, and J. Goldberger. Lesion detection in noisy MR brain images using constrained GMM and active contours. In *IEEE Int. Symp. Biomed. Imag.*, pages 596 – 599, 2007.
- [84] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. *Ann. Stat.*, 38(2):337 – 374, 2000.
- [85] K. S. Fu and J. K. Mui. A survey on image segmentation. *Pattern Recog.*, 13:3 – 16, 1981.

- [86] D. García-Lorenzo, J. Lecoœur, L. D. Arnold, D. L. Collins, and C. Barillot. Multiple Sclerosis lesion segmentation using an automatic multimodal Graph Cuts. In *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, pages 584 – 591. SpringerLink, 2009.
- [87] D. García-Lorenzo, S. Prima, D. Arnold, L. Collins, and C. Barillot. Trimmed-likelihood estimation for focal lesions and tissue segmentation in multi-sequence MRI for multiple sclerosis. *IEEE Trans. Med. Imag.*, 30(8):1455 – 1467, 2011.
- [88] D. García-Lorenzo, S. Prima, D. L. Collins, D. L. Arnold, S. P. Morrissey, and C. Barillot. Combining robust expectation maximization and mean shift algorithms for multiple sclerosis brain segmentation. In *Work. Med. Image Anal. Mult. Scler.*, pages 82 – 91, 2008.
- [89] D. García-Lorenzo, S. Prima, S. Morrissey, and C. Barillot. A robust expectation-maximization algorithm for multiple sclerosis lesion segmentation. In *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, 2008.
- [90] D. García-Lorenzo, S. Prima, L. Parkes, J. C. Ferré, S. P. Morrissey, and C. Barillot. The impact of processing workflow in performance of automatic white matter lesion segmentation in multiple sclerosis. In *Work. Med. Image Anal. Mult. Scler.*, pages 104 – 112, 2008.
- [91] Jim C. Gee, Martin Reivich, and Ruzena Bajcsy. Elastically deforming 3D atlas to match anatomical brain images. *J. Comp. Assist. Tomo.*, 17(2):225 – 236, 1993.
- [92] Ezequiel Geremia, Olivier Clatz, Bjoern H. Menze, Ender Konukoglu, Antonio Criminisi, and Nicholas Ayache. Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. *NeuroImage*, 57(2):378 – 390, 2011.
- [93] A. Gholipour, N. Kehtarnavaz, R. Briggs, M. Devous, and K. Gopinath. Brain functional localization: A survey of image registration techniques. *IEEE Trans. Med. Imag.*, 26(4):427 – 451, 2007.
- [94] D. Goldberg-Zimring, A. Achiron, S. Miron, M. Faibel, and H. Azhari. Automated detection and characterization of multiple sclerosis lesions in brain MR images. *Magn. Reson. Imag.*, 16(3):311 – 318, 1998.
- [95] S. Gorthi, J. P. Thiran, and M. Bach Cuadra. Comparison of energy minimization methods for 3-D brain tissue classification. In *IEEE Int. Conf. Image Proc.*, 2011.

-
- [96] V. Grau, A. U. J. Mewes, M. Alcaniz, R. Kikinis, and S. K. Warfield. Improved watershed transform for medical image segmentation using prior information. *IEEE Trans. Med. Imag.*, 23(4):447 – 458, 2004.
- [97] H. Greenspan, A. Ruf, and J. Goldberger. Constrained Gaussian mixture model framework for automatic segmentation of MR brain images. *IEEE Trans. Med. Imag.*, 25(9):1233 – 1245, 2006.
- [98] T. Greitz, C. Bohm, S. Holte, and L. Eriksson. A computerized brain atlas: construction, anatomical content and some applications. *J. Comp. Assist. Tomo.*, 15(1):26 – 38, 1991.
- [99] A. Guimond, J. Meunier, and J. Thirion. Average brain models: a convergence study. *Comput. Vis. Image Underst.*, 77(9):192 – 210, 2000.
- [100] P. A. Habas, K. Kim, F. Rousseau, O. A. Glenn, A. J. Barkovich, and C. Studholme. Atlas-based segmentation of the germinal matrix from in utero clinical MRI of the fetal brain. In *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, pages 351 – 358, 2008.
- [101] P. A. Habas, K. Kim, F. Rousseau, O. A. Glenn, A. J. Barkovich, and C. Studholme. Atlas-based segmentation of developing tissues in the human brain with quantitative validation in young fetuses. *Hum. Brain Mapp.*, 31(9):1348 – 1358, 2010.
- [102] Ali S. Hadi and Alberto Luceño. Maximum trimmed likelihood estimators: a unified approach, examples, and algorithms. *Comput. Stat. Data Anal.*, 25(3):251 – 272, 1997.
- [103] X. Han and B. Fischl. Atlas renormalization for improved brain MR image segmentation across scanner platforms. *IEEE Trans. Med. Imag.*, 26(4):479 – 486, 2007.
- [104] S. W. Hartley, A. I. Scher, E. S. C. Korf, L. R. White, and L. J. Launer. Analysis and validation of automated skull stripping tools: A validation study based on 296 MR images from the Honolulu Asia aging study. *NeuroImage*, 30(4):1179 – 1186, 2006.
- [105] S. L. Hartmann, J. P. Thirion, F. Maes, D. Vandermeulen, and P. Demaerel. Automatic 3-D segmentation of internal structures of the head in MR images using a combination of similarity and free-form transformations. ii. validation on severely atrophied brains. *IEEE Trans. Med. Imag.*, 18(10):917 – 926, 1999.

- [106] R. He and P. A. Narayana. Automatic delineation of Gd enhancements on magnetic resonance images in multiple sclerosis. *Med. Phys.*, 29(7):1536 – 1546, 2002.
- [107] R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage*, 33(1):115 – 126, 2006.
- [108] K. held, E. R. Kops, B. J. Krause, W. M. Wells, R. Kikinis, and H. W. Muller-Gartner. Markov random field segmentation of brain MR images. *IEEE Trans. Med. Imag.*, 16(6):878 – 886, 1997.
- [109] P. Hellier and C. Barillot. A hierarchical parametric algorithm for deformable multimodal image registration. *Comput. Meth. Prog. Biomed.*, 75(2):107 – 115, 2004.
- [110] Derek L. G. Hill, Philipp G. Batchelor, Mark Holden, and David J. Hawkes. Medical image registration. *Phys. Med. Biol.*, 46(3):R1 – R45, 2001.
- [111] K. Hohne, M. Bomans, M. Riemer, R. Schubert, U. Tiede, and W. Lierse. A volume based anatomical atlas. *IEEE Comput. Graphics Appl.*, 12(4):72 – 78, 1992.
- [112] Z. Hou. A review on MR image intensity inhomogeneity correction. *Int. J. Biomed. Imag.*, 2006(1):1 – 11, 2006.
- [113] Z. Hou and S. Huang. Characterization of a sequential pipeline approach to automatic tissue segmentation from brain MR images. *Int. J. Comput. Assist. Radiol. Surg.*, 2(5):305 – 316, 2008.
- [114] D. V. Iosifescu, M. E. Shenton, S. K. Warfield, R. kikinis, J. Dengler, F. A. Jolesz, and R. W. McCarley. An automated registration algorithm for measuring MRI subcortical brain structures. *NeuroImage*, 6(1):13 – 25, 1997.
- [115] I. Isgum, M. Staring, A. Rutten, M. Prokop, M. A. Viergever, and B. van Ginneken. Multi-atlas-based segmentation with local decision fusion - application to cardiac and aortic segmentation in CT scans. *IEEE Trans. Med. Imag.*, 28(7):1000 – 1010, 2004.
- [116] Mark Jenkinson, Christian F. Beckmann, Timothy E.J. Behrens, Mark W. Woolrich, and Stephen M. Smith. FSL. *NeuroImage*, 62(2):782 – 790, 2012.
- [117] B. Johnston, M. S. Atkins, B. Mackiewicz, and M. Anderson. Segmentation of multiple sclerosis lesions in intensity corrected multispectral MRI. *IEEE Trans. Med. Imag.*, 15(2):154 – 169, 1996.

-
- [118] S. Joshi, B. Davis, M. Jomier, and G. Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23(1):151 – 160, 2004.
- [119] M. Kamber, R. Shinghal, D. L. Collins, G. S. Francis, and A. C. Evans. Model-based 3-D segmentation of multiple sclerosis in magnetic resonance brain images. *IEEE Trans. Med. Imag.*, 4(3):442 – 453, 1995.
- [120] L. Kappos, D. Moeri, E. W. Radue, A. Schoetzau, K. Schweikert, F. Barkhof, D. Miller, C. R.G. Guttman, H.L. Weiner, C. Gasperini, and M. Filippi. Predictive value of gadolinium-enhanced magnetic resonance imaging for relapse rate and changes in disability or impairment in multiple sclerosis: a meta-analysis. *Neurology*, 353(9157):964 – 969, 1999.
- [121] R. Khayati, M. Vafadust, F. Towhidkhah, and S. M. Nabavi. Fully automatic segmentation of multiple sclerosis lesions in brain MR FLAIR images using adaptive mixtures method and Markov random field model. *Comput. Biol. Med.*, 38(3):379 – 390, 2008.
- [122] R. Kikinis, M. Shenton, D. Iosifescu, R. McCarley, P. Saiviroonporn, H. Hokama, A. Robantino, D. Metcalf, C. Wible, C. Portas, R. Donnino, and F. Jolesz. A digital brain atlas for surgical planning, model driven segmentation and teaching. *IEEE Trans. Visual. Comput. Graphics*, 2(3):232 – 241, 1996.
- [123] Arno Klein and Joy Hirsch. Mindboggle: a scatterbrained approach to automate brain labeling. *NeuroImage*, 24(2):261 – 280, 2005.
- [124] Arno Klein, Brett Mensh, Satrajit Ghosh, Jason Tourville, and Joy Hirsch. Mindboggle: Automated brain labeling with multiple atlases. *BMC Med. Imag.*, 5(1):7, 2005.
- [125] D.J. Kroon, E. van Oort, and K. Slump. Multiple sclerosis detection in multispectral magnetic resonance images with principal components analysis. In *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, 2008.
- [126] L. I. Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, 2004.
- [127] R. K. S. Kwan, A. C. Evans, and G. B. Pike. MRI simulation-based evaluation of image-processing and classification methods. *IEEE Trans. Med. Imag.*, 18(11):1085 – 1097, 1999.

- [128] S. Kyriakou and C. Davatzikos. Nonlinear elastic registration of brain images with tumor pathology using a biomechanical model. *IEEE Trans. Med. Imag.*, 18(7):580 – 592, 1999.
- [129] Laboratory of Neuro Imaging (UCLA). International Consortium for Brain Mapping. <http://www.loni.ucla.edu/ICBM/>. last visit: 29/12/2010.
- [130] Laboratory of Neuro Imaging (UCLA). International Consortium for Brain Mapping. <http://www.loni.ucla.edu/Atlases/>. last visit: 20/03/2010.
- [131] J.R. Landis and G.G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159 – 174, 1977.
- [132] Jayaram K. Udupa Lázló G. Nyúl and Xuan Xhang. New variants of a method of MRI scale standardization. *IEEE Trans. Med. Imag.*, 19(2):143 – 150, 2000.
- [133] J. Lecoœur, J. C. Ferré, and C. Barillot. Optimized supervised segmentation of MS lesions from multispectral MRIs. In *Work. Med. Image Anal. Mult. Scler.*, pages 5 – 14, 2009.
- [134] J. M. Lee, U. Yoon, S. H. Nam, J. H. Kim, I. Y. Kim, and S. I. Kim. Evaluation of automated and semi-automated skull-stripping algorithms using similarity index and segmentation error. *Comput. Biol. Med.*, 33(6):495 – 507, 2003.
- [135] K. Van Leemput, F. Maes, D. Vandermeulen, A. Colchester, and P. Suetens. Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE Trans. Med. Imag.*, 20(8):677 – 688, 2001.
- [136] Herve Lemaitre, Fabrice Crivello, Blandine Grassiot, Annick Alperovitch, Christophe Tzourio, and Bernard Mazoyer. Age- and sex-related effects on the neuroanatomy of healthy elderly. *NeuroImage*, 26(3):900 – 911, 2005.
- [137] Hava Lester and Simon R. Arridge. A survey of hierarchical non-linear medical image registration. *Pattern Recog.*, 32(1):129 – 149, 1999.
- [138] S. Z. Li. *Markov random field modeling in image analysis*. Springer, Tokyo, 2001.
- [139] National Library of Medicine. The visible human project. <http://www.nlm.nih.gov/research/visible/>. last visit: 29/12/2010.

- [140] T. Liu, D. Shen, and C. Davatzikos. Deformable registration of tumor-diseased brain images. In *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, pages 720 – 728, 2004.
- [141] X. Lladó, A. Oliver, M. Cabezas, J. Freixenet, J.C. Vilanova, A. Quiles, L. Valls, Ll. Ramió-Torrentà, and A. Rovira. Segmentation of multiple sclerosis lesions in brain MRI: a review of automated approaches. *Inform. Sciences*, 186(1):164 – 185, 2012.
- [142] P. Lorenzen, B. Davis, and S. Joshi. Unbiased atlas formation via large deformations metric mapping. In *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, pages 411 – 418, 2005.
- [143] J. MP. Lötjönen, R. Wolz, J. R. Koikkalainen, L. Thurfjell, G. Waldemar, H. Soininen, and D. Rueckert. Fast and robust multi-atlas segmentation of brain magnetic resonance images. *NeuroImage*, 49(3):2352 – 2365, 2010.
- [144] J. B. Antoine Maintz and Max A. Viergever. A survey of medical image registration. *Med. Image Anal.*, 2(1):1 – 36, 1998.
- [145] J. L. Marroquin, B. C. Vemuri, S. Botello, E. Calderón, and A. Fernández-Bouzas. An accurate and efficient Bayesian method for automatic segmentation of brain MRI. *IEEE Trans. Med. Imag.*, 21(8):934 – 945, 2002.
- [146] M. Martin-Fernandez, E. Muñoz-Moreno, L. Cammoun, J.-P. Thiran, C.-F. Westin, and C. Alberola-Lopez. Sequential anisotropic multichannel Wiener filtering with Rician bias correction applied to 3D regularization of DWI data. *Med. Image Anal.*, 13(1):19 – 35, 2009.
- [147] Massachusetts General Hospital. Internet Brain Segmentation Repository. <http://www.cma.mgh.harvard.edu/ibsr/>. last visit: 29/12/2010.
- [148] J. C. Mazziotta, A. W. Toga, and R. S. J. Frackowiak. *Brain Mapping: The Disorders*. Academic Press, 2000.
- [149] W. I. McDonald, A. Compston, G. Edan, D. Goodkin, H.-P. Hartung, F. D. Lublin, H. F. McFarland, D. W. Paty, C. H. Polman, S. C. Reingold, M. Sandberg-Wollheim, W. Sibley, A. Thompson, S. Van Den Noort, B. Y. Weinshenker, and J. S. Wolinsky. Recommended diagnostic criteria for multiple sclerosis: Guidelines from the international panel on the diagnosis of multiple sclerosis. *Ann. Neurol.*, 50(4):121 – 127, 2001.

- [150] H.F. McFarland, L.A. Stone, P.A. Calabresi, H Maloni, C.N. Bash, and J.A. Frank. MRI studies of multiple sclerosis: implications for the natural history of the disease and for monitoring effectiveness of experimental therapies. *Mult. Scler.*, 2(4):198 – 205, 1996.
- [151] E. R. Melhem and R. Itoh. Effect of T1 relaxation time on lesion contrast enhancement in FLAIR MR imaging: A study using computer-generated brain maps. *Am. J. Roentgenol.*, 176(2):537 – 539, 2001.
- [152] D. H. Miller, R. I. Grossman, S. C. Reingold, and H. F. McFarland. The role of magnetic resonance techniques in understanding and managing multiple sclerosis. *Brain*, 121(1):3 – 24, 1998.
- [153] DH Miller, BG Weinschenker, M Filippi, BL Banwell, JA Cohen, MS Freedman, SL Galetta, M Hutchinson, RT Johnson, L Kappos, J Kira, FD Lublin, HF McFarland, X Montalban, H Panitch, JR Richert, SC Reingold, and CH Polman. Differential diagnosis of suspected multiple sclerosis: a consensus approach. *Mult. Scler.*, 14(9):1157 – 1174, 2008.
- [154] M. I. Miller, G. E. Christensen, Y. Amit, and U. Grenander. Mathematical textbook of deformable neuroanatomies. *Proc. Natl. Acad. Sci. U.S.A.*, 90(24):11944 – 11948, 1993.
- [155] N. Moon, E. Bullitt, K. van Leemput, and G. Gerig. Automatic brain and tumor segmentation. In *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, pages 372 – 379, 2002.
- [156] J. H. Morra, Z. Tu, A. W. Toga, and P. M. Thompson. Automatic segmentation of MS lesions using a contextual model for the MICCAI grand challenge. In *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, 2008.
- [157] Benedicte Mortamet, Donglin Zeng, Guido Gerig, Marcel Prastawa, and Elizabeth Bullitt. Effects of healthy aging measured by intracranial compartment volumes using a designed MR brain database. In *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, pages 383 – 391, 2005.
- [158] M. Murgasova, L. Dyet, D. Edwards, M. Rutherford, J. Hajnal, and D. Rueckert. Segmentation of brain MRI in young children. *Acad. Radiol.*, 14(11):1350 – 1366, 2007.

- [159] V. Noblet, C. Heinrich, F. Heitz, and J. P. Armspach. 3-D deformable image registration: a topology preservation scheme based on hierarchical deformation models and interval analysis optimization. *IEEE Trans. Image Processing*, 14(5):553 – 566, 2005.
- [160] Northern California Institute for Research and Education (UCSF). Alzheimer’s Disease Neuroimaging Initiative (ADNI). <http://www.adni-info.org/>. last visit: 01/03/2011.
- [161] W. L. Nowinski and D. Belov. Toward atlas-assisted automatic interpretation of MRI morphological brain scans in the presence of tumor. *NeuroImage*, 12(8):1049 – 1057, 2005.
- [162] T. Okuda, Y. Korogi, Y. Shugematsu, T. Sugahara, T. Hirai, I. Ikushima, L. Liang, and M. Takahashi. Brain lesion: when should fluid-attenuated inversion recovery sequences be used in MR evaluation? *Radiology*, 212(3):793 – 798, 1999.
- [163] Arnau Oliver, Albert Torrent, Xavier Lladó, Meritxell Tortajada, Lidia Tortajada, Melcior Sentí’s, Jordi Freixenet, and Reyer Zwiggelaar. Automatic microcalcification and cluster detection for digital and digitised mammograms. *Knowl.-Based Syst.*, 28(0):68 – 75, 2012.
- [164] Nikhil R Pal and Sankar K Pal. A review on image segmentation techniques. *Pattern Recog.*, 26(9):1277 – 1294, 1993.
- [165] H. Park, P. Bland, A. Hero, and C. Meyer. Least biased target selection in probabilistic atlas construction. In *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, pages 419 – 426, 2005.
- [166] Pietro Perona and Jalhandra Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Machine Intell.*, 12(7):629 – 639, 1990.
- [167] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE Trans. Med. Imag.*, 22(8):986 – 1004, 2003.
- [168] Kilian M. Pohl, John Fisher, W. Eric L. Grimson, Ron Kikinis, and W. M. Wells III. A Bayesian model for joint segmentation and registration. *NeuroImage*, 31(1):228 – 239, 2006.

- [169] C. Pollo, M. Bach-Cuadra, O. Cuisenaire, J.-G. Villemure, and J. P. Thiran. Segmentation of brain structures in presence of a space-occupying lesion. *NeuroImage*, 24(4):990 – 996, 2005.
- [170] C. H. Polman, S. C. Reingold, B. Banwell, M. Clanet, J. A. Cohen, M. Filippi, K. Fujihara, E. Havrdova, M. Hutchinson, L. Kappos, F. D. Lublin, X. Montalban, P. O’Connor, M. Sandberg-Wollheim, A. J. Thompson, E. Waubant, B. Weinshenker, and J. S. Wolinsky. Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria. *Ann. Neurol.*, 69(2):292 – 302, 2011.
- [171] Chris H. Polman, Stephen C. Reingold, Gilles Edan, Massimo Filippi, Hans-Peter Hartung, Ludwig Kappos, Fred D. Lublin, Luanne M. Metz, Henry F. McFarland, Paul W. O’Connor, Magnhild Sandberg-Wollheim, Alan J. Thompson, Brian G. Weinshenker, and Jerry S. Wolinsky. Diagnostic criteria for multiple sclerosis: 2005 revisions to the McDonald criteria. *Ann. Neurol.*, 58(6):840 – 846, 2005.
- [172] G. Postelnicu, L. Zollei, and B. Fischl. Combined volumetric and surface registration. *IEEE Trans. Med. Imag.*, 28(4):508 – 522, 2009.
- [173] M. Prastawa, E. Bullitt, S. Ho, and G. Gerig. A brain tumor segmentation framework based on outlier detection. *Med. Image Anal.*, 8(3):275 – 283, 2004.
- [174] M. Prastawa and G. Gerig. Automatic MS lesion segmentation by outlier detection and information theoretic region partitioning. In *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, 2008.
- [175] M. Prastawa, J. H. Gilmore, S. Jiang, J. Allsop, L. Perkins, L. Srinivasan, T. Hayat, S. Kumar, and J. Hajnal. Automatic segmentation of MR images of the developing newborn brain. *Med. Image Anal.*, 9(5):457 – 466, 2005.
- [176] K. Rehm, K. Schaper, J. Anderson, R. Woods, S. Stoltzner, and D. Rottenberg. Putting our heads together: a consensus approach to brain/non-brain segmentation in T1-weighted MR volumes. *NeuroImage*, 22(3):1262 – 1270, 2004.
- [177] D. Rey, G. Subsol, H. Delingette, and N. Ayache. Automatic detection and segmentation of evolving processes in 3D medical images: application to multiple sclerosis. *Med. Image Anal.*, 6(2):163 – 179, 2002.
- [178] T. Rohlfing, R. Brandt, R. Menzel, and C. R. Maurer Jr. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *NeuroImage*, 21(4):1428 – 1442, 2004.

- [179] Torsten Rohlfing, D. B. Russakoff, and C. R. Maurer Jr. Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. *IEEE Trans. Med. Imag.*, 23(8):983 – 994, 2004.
- [180] A. Rovira and A. León. MR in the diagnosis and monitoring of multiple sclerosis: An overview. *Eur. J. Radiol.*, 67(3):409 – 414, 2008.
- [181] A. Rovira, J. Swanton, M. Tintoré, E. Huerga, F. Barkhof, M. Filippi, J. L. Frederiksen, A. Langkilde, K. Miszkiel, C. Polman, M. Rovaris, J. Sastre-Garriga, D. Miller, and X. Montalban. A single, early magnetic resonance imaging study in the diagnosis of multiple sclerosis. *Arch. Neurol.*, 66(5):587 – 592, 2009.
- [182] D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach, and D.J. Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans. Med. Imag.*, 18(8):712 – 721, 1999.
- [183] J. N. Rydberg, S. J. Riederer, C. H. Rydberg, and C. R. Jack. Contrast optimization of fluid-attenuated inversion recovery (FLAIR) imaging. *Magn. Reson. Imag.*, 34(6):868 – 877, 1995.
- [184] M. R. Sabuncu, B. T. T. Yeo, K. van Leemput, B. Fischl, and P. Golland. A generative model for image segmentation based on label fusion. *IEEE Trans. Med. Imag.*, 29(10):1714 – 1729, 2010.
- [185] S. Saha and S. Bandyopadhyay. A new point symmetry based fuzzy genetic clustering technique for automatic evolution of clusters. *Inform. Sciences*, 179(9):3230 – 3246, 2009.
- [186] B. R. Sajja, S. Datta, R. He, M. Mehta, R. K. Gupta, J. S. Wolinsky, and P. A. Narayana. Unified approach for multiple sclerosis lesion segmentation on brain MRI. *Ann. Biomed. Eng.*, 34(1):142 – 151, 2006.
- [187] Paul Schmidt, Christian Gaser, Milan Arsic, Dorothea Buck, Annette Forschler, Achim Berthele, Muna Hoshi, Rudiger Ilg, Volker J. Schmid, Claus Zimmer, Bernhard Hemmer, and Mark Mahlau. An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis. *NeuroImage*, 59(4):3774 – 3783, 2012.
- [188] M. Scully, V. Magnotta, C. Gasparovic, P. Pelligrino, D. Feis, and H. J. Bockholt. 3D segmentation in the clinic: A grand challenge II at MICCAI 2008 - MS lesion segmentation. In *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, 2008.

- [189] M. L. Seghier, A. Ramalackhansingha, A. P. Leffa J. Criniona, and C. J. Pricea. Lesion identification using unified segmentation-normalisation models and fuzzy clustering. *NeuroImage*, 41(4):1253 – 1266, 2008.
- [190] F. Segonne, A. M. Dale, E. Busa, M. Glessner, D. Salat, H. K. Hahn, and B. Fischl. A hybrid approach to the skull stripping problem in mri. *NeuroImage*, 22(3):1060 – 1075, 2004.
- [191] D. D. Sha and J. P. Sutton. Towards automated enhancement, segmentation and classification of digital brain images using networks of networks. *Inform. Sciences*, 138(1 - 4):45 – 77, 2001.
- [192] M. Shah, Y. Xiao, N. Subbanna, S. Francis, D. L. Arnold, D. L. Collins, and T. Arbel. Evaluating intensity normalization on MRIs of human brain with multiple sclerosis. *Med. Image Anal.*, 15(2):267 – 282, 2011.
- [193] D. W. Shattuck, S. R. Sandor-Leahy, K. A. Schaper, D. A. Rottenberg, and R. M. Leahy. Magnetic resonance image tissue classification using a partial volume model. *NeuroImage*, 13(5):856 – 876, 2001.
- [194] D. Shen and C. Davatzikos. HAMMER: Hierarchical attribute matching mechanism for elastic registration. *IEEE Trans. Med. Imag.*, 21(11):1421 – 1439, 2002.
- [195] S. Shen, A. J. Szameitat, and A. Sterr. Detection of infract lesions from single MRI modality using inconsistency between voxel intensity and spatial location - a 3D automatic approach. *IEEE Trans. Inform. Technol. Biomed.*, 12(4):532 – 540, 2008.
- [196] S. Shen, A. J. Szameitat, and A. Sterr. An improved lesion detection approach based on similarity measurement between fuzzy intensity segmentation and spatial probability maps. *Magn. Reson. Imag.*, 28(2):245 – 254, 2010.
- [197] N. Shiee, P. L. Bazin, and D. L. Pham. Multiple sclerosis lesions segmentation using statistical and topological atlases. In *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, 2008.
- [198] N. Shiee, P.L. Bazin, J. L. Cuzzocreo, D. S. Reich, P. A. Calabresi, and D. L. Pham. Topologically constrained segmentation of brain images with multiple sclerosis lesions. In *Work. Med. Image Anal. Mult. Scler.*, pages 71 – 81, 2008.

- [199] Navid Shiee, Pierre-Louis Bazin, Arzu Ozturk, Daniel S. Reich, Peter A. Calabresi, and Dzung L. Pham. A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. *NeuroImage*, 49(2):1524 – 1535, 2010.
- [200] J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost: multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *Int. J. Comp. Vis.*, 81(1):2 – 23, 2009.
- [201] J.G. Sled, A.P. Zijdenbos, and A.C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imag.*, 17(1):87 – 97, 1998.
- [202] Charles D. Smith, Himachandra Chebrolu, David R. Wekstein, Frederick A. Schmitt, and William R. Markesbery. Age and gender effects on human brain anatomy: A voxel-based morphometric study in healthy elderly. *Neurobiol. of Aging*, 28(7):1075 – 1087, 2007.
- [203] S. M. Smith. Fast robust automated brain extraction. *Hum. Brain Mapp.*, 17(3):143 – 155, 2002.
- [204] J. C. Souplet, C. Lebrun, N. Ayache, and G. Malandain. An automatic segmentation of T2-FLAIR multiple sclerosis lesions. In *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, pages 1 – 11, 2008.
- [205] R. Stefanescu, O. Commowick, G. Malandain, P.-Y. Bondiau, N. Ayache, and X. Pennec. Non-rigid atlas to subject registration with pathologies for conformal brain radiotherapy. In *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, pages 704 – 711, 2004.
- [206] M. Stella Atkins and B. T. Mackiewicz. Fully automatic segmentation of the brain in mri. *IEEE Trans. Med. Imag.*, 17(1):98 – 107, 1998.
- [207] C.J. Stone. Consistent nonparametric regression. *Ann. Stat.*, 5(4):595 – 620, 1977.
- [208] M. Styner, J. Lee, B. Chin, M.S. Chin, O. Commowick, H. Tran, V. Jewells, and S. Warfield. Editorial: 3D segmentation in the clinic: A grand challenge II: Ms lesion segmentation. In *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, pages 1 – 8, 2008.

- [209] N.K. Subbanna, M. Shah, S. J. Francis, S. Narayanan, D. L. Collins, D. L. Arnold, and T. Arbel. MS lesion segmentation using Markov Random Fields. In *Work. Med. Image Anal. Mult. Scler.*, pages 15 – 26, 2009.
- [210] Jasjit S. Suri, Sameer Singh, and Laura Reden. Computer Vision and Pattern Recognition Techniques for 2-D and 3-D MR Cerebral Cortical Segmentation (Part I): A State-of-the-Art Review. *Pattern Anal. Appl.*, 5:46 – 76, 2002.
- [211] J K Swanton, K Fernando, C M Dalton, K A Miszkiel, A J Thompson, G T Plant, and D H Miller. Modification of MRI criteria for multiple sclerosis in patients with clinically isolated syndromes. *J. Neurol. Neurosurg. Psychiatry*, 77(7):830–833, 2006.
- [212] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *IEEE Trans. Pattern Anal. Machine Intell.*, 30(6):1068 – 1080, 2008.
- [213] J. Talairach, M. David, P. Tournoux, H. Corredor, and T. Kvasina. *Atlas d' Anatomie Stéréotaxique des Noyaux Gris Centraux*. Masson, Paris, 1957.
- [214] J. Talairach and P. Tournoux. *Co-planar stereotaxic atlas of the human brain*. Mark Rayport, Trans. Thieme., Stuttgart, 1988.
- [215] T. Taoka, M. Fujioka, Y. Matsuo, M. Notoya, S. Iwasaki, A. Fukusumi, H. Nakagawa, M. Sakamoto, K. Kichikawa, and H. Ohishi. Signal characteristics of FLAIR related to water content: comparison with conventional spin echo imaging in infarcted rat brain. *Magn. Reson. Imag.*, 22(2):221 – 227, 2004.
- [216] J. P. Thirion. Image matching as a diffusion process: an analogy with Maxwell's demons. *Med. Image Anal.*, 2(3):243 – 260, 1998.
- [217] J. P. Thirion and G. Calmon. Deformation analysis to detect and quantify active lesions in three-dimensional medical image sequences. *IEEE Trans. Med. Imag.*, 18(5):429 – 441, 1999.
- [218] L. Thurjfell, C. Bohm, T. Greitz, and L. Eriksson. Transformations and algorithms in a computerized brain atlas. *IEEE Trans. Nucl. Sci.*, 40(4):1187 – 1191, 1993.
- [219] Arthur W. Toga and Paul M. Thompson. Temporal dynamics of brain anatomy. *Annu. Rev. Biomed. Eng.*, 5:119 – 145, 2003.

- [220] X. Tomas and S. K. Warfield. Fully-automatic generation of training points for automatic multiple sclerosis segmentation. In *Work. Med. Image Anal. Mult. Scler.*, pages 49 – 59, 2009.
- [221] A. Torralba, K.P. Murphy, and W.T. Freeman. Sharing visual features for multi-class and multiview object detection. *IEEE Trans. Pattern Anal. Machine Intell.*, 29(5):854 –869, 2007.
- [222] A. Torrent, X. Lladó, J.Freixenet, and A. Torralba. Simultaneous detection and segmentation for generic objects. In *IEEE Conf. Image Processing*, pages 653 – 656, 2011.
- [223] A. Torrent, M. Peracaula, X. Lladó, J. Freixenet, J.R. Sánchez-Sutil, J. Martí, and J.M. Paredes. Detecting faint compact sources using local features and a boosting approach. In *IEEE Conf. Pattern Recog.*, pages 4613 – 4616, 2010.
- [224] N.J. Tustison, B.B. Avants, P.A. Cook, Yuanjie Zheng, A. Egan, P.A. Yushkevich, and J.C. Gee. N4ITK: Improved N3 bias correction. *IEEE Trans. Med. Imag.*, 29(6):1310 –1320, 2010.
- [225] J. K. Udupa, L. Wei, S. Samarasekera, Y. Miki, M. A. van Buchem, and R. I. Grossman. Multiple sclerosis lesion quantification using fuzzy-connectedness principles. *IEEE Trans. Med. Imag.*, 16(5):598 – 609, 1997.
- [226] Jayaram K. Udupa, Vicki R. LeBlanc, Ying Zhuge, Celina Imielinska, Hilary Schmidt, Leanne M. Currie, Bruce E. Hirsch, and James Woodburn. A framework for evaluating image segmentation algorithms. *Comput. Med. Imag. Graphics*, 30(2):75 – 87, 2006.
- [227] F. van der Lijn, T. den Heijer, M. M. B. Breteler, and W.J. Niessen. Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts. *NeuroImage*, 43(4):708–720, 2008.
- [228] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automated model-based tissue classification of MR images of the brain. *IEEE Trans. Med. Imag.*, 18(10):897 – 908, 1999.
- [229] B. C. Vemuri, J. Ye, Y. Chen, and C. M. Leonard. A level-set based approach to image registration. In *IEEE Work. Math. Meth. Biomed. Imag. Anal.*, pages 86 – 93, 2000.

- [230] Baba C. Vemuri, Shuangying Huang, Sartaj Sahni, Christiana M. Leonard, Cecile Mohr, Robin Gilmore, and Jeffrey Fitzsimmons. An efficient motion estimator with application to medical image registration. *Med. Image Anal.*, 2(1):79 – 98, 1998.
- [231] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(Supplement 1):S61 – S72, 2009.
- [232] U. Vovk, F. Pernus, and B. Likar. A review of methods for correction of intensity inhomogeneity in MRI. *IEEE Trans. Med. Imag.*, 26(3):405 – 421, 2007.
- [233] H. A. Vrooman, C. A. Cocosco, F. van der Lijn, R. Stokking, M. A. Ikram, M. W. Vernooij, M. M. B. Breteler, and W. J. Niessen. Multi-spectral brain tissue segmentation using automatically trained k-nearest-neighbor classification. *NeuroImage*, 37(1):71 – 81, 2007.
- [234] L. Wang, H. Lai, A. Thompson, and D. Miller. Survey of the distribution of lesion size in multiple sclerosis: implication for the measurement of total lesion load. *J. Neurol. Neurosurg. Psychiatry*, 63(4):452 – 455, 1997.
- [235] S. K. Warfield. Fast k-NN classification for multichannel image data. *Pattern Recognit. Lett.*, 17(7):713 – 721, 1996.
- [236] S. K. Warfield, M. Kaus, F. A. Jolesz, and R. Kikinis. Adaptive, template moderated, spatially varying statistical classification. *Med. Image Anal.*, 4(1):43 – 55, 2000.
- [237] S. K. Warfield, K. H. Zou, and W. M. Wells. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans. Med. Imag.*, 23(7):903 – 921, 2004.
- [238] N. I. Weisenfeld, A. U. J. Mewes, and S. K. Warfield. Highly accurate segmentation of brain tissue and subcortical gray matter from newborn MRI. In *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, pages 199 – 206, 2006.
- [239] N. I. Weisenfeld and S. K. Warfield. Automatic segmentation of newborn brain MRI. *NeuroImage*, 47(2):564 – 572, 2009.
- [240] Jennifer L. Whitwell. Voxel-based morphometry: An automated technique for assessing structural changes in the brain. *J. Neurosci.*, 29(31):9661 – 9664, 2009.

- [241] Robin Wolz, Rolf A. Heckemann, Paul Aljabar, Joseph V. Hajnal, Alexander Hammers, Jyrki Lötjönen, and Daniel Rueckert. Measurement of hippocampal atrophy using 4D graph-cut segmentation: Application to adni. *NeuroImage*, 52(1):109 – 118, 2010.
- [242] R. Woods, M. Dapretto, N. Sicotte, A. Toga, and J. Mazziotta. Creation and use of a Talarach-compatible atlas for accurate, automated, nonlinear intersubject registration, and analysis of functional imaging. *Hum. Brain Mapp.*, 8(2-3):554 – 566, 1999.
- [243] Minjie Wu, Caterina Rosano, Pilar Lopez-Garcia, Cameron S. Carter, and Howard J. Aizenstein. Optimum template selection for atlas-based segmentation. *NeuroImage*, 34(4):1612 – 1618, 2007.
- [244] Y. Wu, S. K. Warfield, I. Leng Tan, W. M. Wells III, D. S. Meier, R. A. van Schijndel, F. Barkhof, and C. R. G. Guttmann. Automated segmentation of multiple sclerosis lesion subtypes with multichannel MRI. *NeuroImage*, 32(3):1205 – 1215, 2006.
- [245] Lei Xu, Adam Krzyzak, and Ching Y. Suen. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Trans. Syst., Man, Cybern.*, 22(3):418 – 435, 1992.
- [246] H. Xue, L. Srinivasan, S. Jiang, M. Rutherford, A. Edwards, D. Rueckert, and J. Hajnal. Automatic segmentation and reconstruction of the cortex from neonatal mri. *NeuroImage*, 38(3):461 – 477, 2007.
- [247] E. I. Zacharaki, C. S. Hoge, D. Shen, G. Biros, and C. Davatzikos. Non-diffeomorphic registration of brain tumor images by simulating tissue loss and tumor growth. *NeuroImage*, 46(3):762 – 774, 2009.
- [248] E. I. Zacharaki, D. Shen, S. K. Lee, and C. Davatzikos. ORBIT: a multiresolution framework for deformable registration of brain tumor images. *IEEE Trans. Med. Imag.*, 27(8):1003 – 1017, 2008.
- [249] Y. Zhang, M. Brady, and S. Smith. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imag.*, 20(1):45 – 57, 2001.
- [250] Lu Zhao, Ulla Ruotsalainen, Jussi Hirvonen, Jarmo Hietala, and Jussi Tohka. Automatic cerebral and cerebellar hemisphere segmentation in 3D MRI: Adaptive disconnection algorithm. *Med. Image Anal.*, 14(3):360 – 372, 2010.

-
- [251] A. P. Zijdenbos, R. Forghani, and A. C. Evans. Automatic “pipeline” analysis of 3-D MRI data for clinical trials: Application to multiple sclerosis. *IEEE Trans. Med. Imag.*, 21(10):1280 – 1291, 2002.
- [252] A.P. Zijdenbos, B.M. Dawant, R.A. Margolin, and A.C. Palmer. Morphometric analysis of white matter lesions in MR images: method and validation. *IEEE Trans. Med. Imag.*, 13(4):716 – 724, 1994.
- [253] L. Zollei, E. L. Miller, W. Grimson, and W. M. Wells III. Efficient population registration of 3D data. In *Comput. Vis. Biomed. Image Appl.*, pages 291 – 301, 2005.